Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

31-Aug-2007

English - Or. English

ENV/JM/MONO(2007)23
Unclassified

ENVIRONMENT DIRECTORATE
**JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY**

**SERIES ON TESTING AND ASSESSMENT**
**Number 76**

# FINAL REPORT OF THE VALIDATION OF THE AMPHIBIAN METAMORPHOSIS ASSAY FOR THE DETECTION OF THYROID ACTIVE SUBSTANCES: PHASE 1 – OPTIMISATION OF THE TEST PROTOCOL

JT03231273



English - Or. English

**OECD Environment, Health and Safety Publications**

**Series on Testing and Assessment**

**No. 76**

# FINAL REPORT OF THE VALIDATION OF THE AMPHIBIAN METAMORPHOSIS ASSAY FOR THE DETECTION OF THYROID ACTIVE SUBSTANCES: PHASE 1 – OPTIMISATION OF THE TEST PROTOCOL

**Environment Directorate**

**Organisation for Economic Co-operation and Development**

**2007**

**Also published in the Series on Testing and Assessment:**

No. 1, *Guidance Document for the Development of OECD Guidelines for Testing of Chemicals (1993; reformatted 1995, revised 2006)*

No. 2, *Detailed Review Paper on Biodegradability Testing (1995)*

No. 3, *Guidance Document for Aquatic Effects Assessment (1995)*

No. 4, *Report of the OECD Workshop on Environmental Hazard/Risk Assessment (1995)*

No. 5, *Report of the SETAC/OECD Workshop on Avian Toxicity Testing (1996)*

No. 6, *Report of the Final Ring-test of the Daphnia magna Reproduction Test (1997)*

No. 7, *Guidance Document on Direct Phototransformation of Chemicals in Water (1997)*

No. 8, *Report of the OECD Workshop on Sharing Information about New Industrial Chemicals Assessment (1997)*

No. 9, *Guidance Document for the Conduct of Studies of Occupational Exposure to Pesticides during Agricultural Application (1997)*

No. 10, *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data (1998)*

No. 11, *Detailed Review Paper on Aquatic Testing Methods for Pesticides and industrial Chemicals (1998)*

No. 12, *Detailed Review Document on Classification Systems for Germ Cell Mutagenicity in OECD Member Countries (1998)*

No. 13, *Detailed Review Document on Classification Systems for Sensitising Substances in OECD Member Countries 1998)*

No. 14, *Detailed Review Document on Classification Systems for Eye Irritation/Corrosion in OECD Member Countries (1998)*

No. 15, *Detailed Review Document on Classification Systems for Reproductive Toxicity in OECD Member Countries (1998)*

No. 16, *Detailed Review Document on Classification Systems for Skin Irritation/Corrosion in OECD Member Countries (1998)*

No. 17, *Environmental Exposure Assessment Strategies for Existing Industrial Chemicals in OECD Member Countries (1999)*

No. 18, *Report of the OECD Workshop on Improving the Use of Monitoring Data in the Exposure Assessment of Industrial Chemicals (2000)*

No. 19, *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (1999)*

No. 20, *Revised Draft Guidance Document for Neurotoxicity Testing (2004)*

No. 21, *Detailed Review Paper: Appraisal of Test Methods for Sex Hormone Disrupting Chemicals (2000)*

No. 22, *Guidance Document for the Performance of Out-door Monolith Lysimeter Studies (2000)*

No. 23, *Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (2000)*

No. 24, *Guidance Document on Acute Oral Toxicity Testing (2001)*

No. 25, *Detailed Review Document on Hazard Classification Systems for Specifics Target Organ Systemic Toxicity Repeated Exposure in OECD Member Countries (2001)*

No. 26, *Revised Analysis of Responses Received from Member Countries to the Questionnaire on Regulatory Acute Toxicity Data Needs (2001)*

No 27, *Guidance Document on the Use of the Harmonised System for the Classification of Chemicals Which are Hazardous for the Aquatic Environment (2001)*

No 28, *Guidance Document for the Conduct of Skin Absorption Studies (2004)*

No 29, *Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media (2001)*

No 30, *Detailed Review Document on Hazard Classification Systems for Mixtures (2001)*

No 31, *Detailed Review Paper on Non-Genotoxic Carcinogens Detection: The Performance of In-Vitro Cell Transformation Assays (2007)*

No. 32, *Guidance Notes for Analysis and Evaluation of Repeat-Dose Toxicity Studies (2000)*

No. 33, *Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures (2001)*

No. 34, *Guidance Document on the Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment (2005)*

No. 35, *Guidance notes for analysis and evaluation of chronic toxicity and carcinogenicity studies (2002)*

No. 36, *Report of the OECD/UNEP Workshop on the use of Multimedia Models for estimating overall Environmental Persistence and long range Transport in the context of PBTS/POPS Assessment (2002)*

No. 37, *Detailed Review Document on Classification Systems for Substances Which Pose an Aspiration Hazard (2002)*

No. 38, *Detailed Background Review of the Uterotrophic Assay Summary of the Available Literature in Support of the Project of the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA) to Standardise and Validate the Uterotrophic Assay (2003)*

No. 39, *Guidance Document on Acute Inhalation Toxicity Testing (in preparation)*

No. 40, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures Which Cause Respiratory Tract Irritation and Corrosion (2003)*

No. 41, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures which in Contact with Water Release Toxic Gases (2003)*

No. 42, *Guidance Document on Reporting Summary Information on Environmental, Occupational and Consumer Exposure (2003)*

No. 43, *Draft Guidance Document on Reproductive Toxicity Testing and Assessment (in preparation)*

No. 44, *Description of Selected Key Generic Terms Used in Chemical Hazard/Risk Assessment (2003)*

No. 45, *Guidance Document on the Use of Multimedia Models for Estimating Overall Environmental Persistence and Long-range Transport (2004)*

No. 46, *Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances (2004)*

No. 47, *Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances (2004)*

No. 48, *New Chemical Assessment Comparisons and Implications for Work sharing (2004)*

No. 49, *Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs (2004)*

No. 50, *Report of the OECD/IPCS Workshop on Toxicogenomics (2005)*

No. 51, *Approaches to Exposure Assessment in OECD Member Countries: Report from the Policy Dialogue on Exposure Assessment in June 2005 (2006)*

No. 52, *Comparison of emission estimation methods used in Pollutant Release and Transfer Registers (PRTRs) and Emission Scenario Documents (ESDs): Case study of pulp and paper and textile sectors (2006)*

No. 53, *Guidance Document on Simulated Freshwater Lentic Field Tests (Outdoor Microcosms and Mesocosms) (2006)*

No. 54, *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application (2006)*

No. 55, *Detailed Review Paper on Aquatic Arthropods in Life Cycle Toxicity Tests with an Emphasis on Developmental, Reproductive and Endocrine Disruptive Effects (2006)*

No. 56, *Guidance Document on the Breakdown of Organic Matter in Litter Bags (2006)*

No. 57, *Detailed Review Paper on Thyroid Hormone Disruption Assays (2006)*

No. 58, *Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals (2006)*

No. 59, *Report of the Validation of the Updated Test Guideline 407: Repeat Dose 28-Day Oral Toxicity Study in Laboratory Rats (2006)*

No. 60, *Report of the Initial Work Towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1A) (2006)*

No. 61, *Report of the Validation of the 21-Day Fish Screening Assay fort he Detection of Endocrine Active Substances (Phase 1B) (2006)*

No. 62, *Final OECD Report of the Initial Work Towards the Validation of the Rat Hershberger Assay: Phase-1, Androgenic Response to Testosterone Propionate, and Anti-Androgenic Effects of Flutamide (2006)*

No. 63, *Guidance Document on the Definition of Residue (2006)*

No. 64, *Guidance Document on Overview of Residue Chemistry Studies (2006)*

No. 65, *OECD Report of the Initial Work Towards the Validation of the Rodent Uterotrophic Assay – Phase One (2006)*

No. 66, *OECD Report of the Validation of the Rodent Uterotrophic Bioassay - Phase 2: Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories (2006)*

No. 67, *Additional Data Supporting the Test Guideline on the Uterotrophic Bioassay in Rodents (2007)*

No. 68, *Summary Report of the Uterotrophic Bioassay Peer Review Panel, including Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2006)*

No 69, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models (2007)*

No. 70, *Report on Preparation of GHS Implementation by the OECD Countries (2007)*

No. 71, *Guidance Document on the Uterotrophic Bioassay – Procedure to Test for Antioestrogenicity (2007)*

No. 72, *Guidance Document on Pesticide Residue analytical Methods (2007)*

No. 73, *Report of the Validation of the Rat Hershberger Assay: Phase 3: Coded testing of Androgen Agonists, Androgen Antagonists and Negative Reference Chemicals by Multiple Laboratories. Surgical Castrate Model Protocol (2007)*

No. 74, *Detailed Review Paper for Avian Two-generation Testing (2007)*

No. 75, *Guidance Document on the Honeybee (*Apis Mellifera L.*) Brood Test under Semi-field Conditions (2007)*

No. 76, *Final Report of the Validation of the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances: Phase 1 – Optimisation of the Test Protocol (2007)*

## ABOUT THE OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 30 industrialised countries in North America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in ten different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides and Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and the Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/ehs/).

**This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.**

**The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The participating organisations are FAO, ILO, OECD, UNEP, UNIDO, UNITAR and WHO. The World Bank and UNDP are observers. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.**

## FOREWORD

This document presents the report of the Phase 1 of the validation of the amphibian metamorphosis assay. The amphibian metamorphosis assay was included in the OECD conceptual framework for the testing and assessment of endocrine disrupting chemicals for the detection of thyroid active substances in aquatic vertebrates.

An OECD expert group on amphibian testing was established in 2002 to work on the development and validation of the amphibian metamorphosis assay. Following the first meeting of the group (Duluth, United States, 2003), a proposal was developed by Germany, Japan and the United States for the Phase 1 (optimization phase) of the validation of the assay, and approved by the Validation Management Group for Ecotoxicity Testing (VMG-eco). The Phase 1 experimental work was undertaken in 2003-2004 in three laboratories. A first draft of the validation report was prepared in September 2004 and presented to the VMG-eco for comments in December 2004. The initial draft was revised in March 2005, taking into account comments received. The Phase 2 of the validation was undertaken in 2005-2006, and a separate report is available. The Task Force on Endocrine Disrupters Testing and Assessment and the Working Group of National Coordinators of the Test Guidelines Programme endorsed both reports with minor modifications.

The Phase 1 and the Phase 2 reports have to be considered together, in a logical step-wise approach towards the validation of the assay.

This document is published on the responsibility of the Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology.

**TABLE OF CONTENTS**

FOREWORD ..... 12
LIST OF TABLES: ..... 14
ACKNOWLEDGEMENT ..... 17
SUMMARY ..... 18
INTRODUCTION ..... 20
OBJECTIVE OF THE INTER-LABORATORY COMPARISON STUDY ..... 21
METHODS ..... 22
Overview of Test Conditions ..... 22
Exposure Experiments ..... 22
Statistical Analysis ..... 23
RESULTS ..... 23
Analytical Chemistry ..... 23
Analytical Chemistry Results (US lab) ..... 23
Analytical Chemistry Results (JPN lab) ..... 24
Comparison of Control Data ..... 25
Exposure Studies with Propylthiouracil (PTU) ..... 28
Stage 51 Exposure Studies with PTU (21 Day Assay) ..... 28
Stage 54 Exposure Studies with PTU (14 Day Assay) ..... 39
Exposure Studies with Thyroxine (T4) ..... 49
Stage 51 Exposure Studies with T4 (21 Day Assay) ..... 49
Stage 54 Exposure Studies with T4 (14 Day Assay) ..... 57
DISCUSSION ..... 68
Control Organism Performance ..... 68
Effects Pattern of PTU on Metamorphosis and Thyroid System ..... 68
Endpoint Sensitivity in PTU Studies ..... 69
Developmental Stage ..... 69
Hind Limb Length ..... 71
Thyroid Histopathology ..... 71
Gene Expression Analysis ..... 74
Body Length and Body Weight ..... 75
Effects Pattern of T4 on Metamorphosis and Thyroid System ..... 75
Comparison of Endpoint Sensitivities in T4 Studies ..... 76
Developmental Stage ..... 76
Hind Limb Length ..... 78
Thyroid Histopathology ..... 78
Gene Expression Analysis ..... 80
Body Length and Body Weight ..... 80
Critical Review of Results from Validation Phase 1 and Proposal for Phase 2 Validation Activities ..... 80

REFERENCES ..... 84
ANNEXES ..... 86

**LIST OF TABLES:**

Table 1. Overview of exposure studies conducted by the three participating labs ..... 22
Table 2. Measured PTU concentrations (mg/L) in the stage 51 study conducted for 21 days. Sample size for each replicate was 4. ..... 24
Table 3. Measured PTU concentrations (mg/L) in the stage 54 study conducted for 14 days. Sample size for each replicate was 3. ..... 24
Table 4. Measured T4 concentrations (µg/L) in the stage 51 study conducted for 21 days. Sample size for each replicate was 2 ..... 24
Table 5. Measured T4 concentrations (µg/L) in the stage 54 study conducted for 14 days. Sample size for replicate 1 was 2 and for replicate 2 was 1 ..... 24
Table 6. Measured PTU concentrations (mg/L) in the stage 51 study and the stage 54 study. Sample sizes of the stage 51 study and the stage 54 study were 4 and 3, respectively ..... 25
Table 7. Measured T4 concentrations (mg/L) in the stage 51 study and the stage 54 study. Sample size of the stage 51 study was 2 or 3, and sample size of the stage 54 study was 2. ..... 25
Table 8. Distribution of developmental stages reached by control tadpoles within 21 days during exposure studies initiated at stage 51 in three different labs. ..... 26
Table 9. Distribution of developmental stages reached by control tadpoles within 14 days during exposure studies initiated at stage 54 in three different labs. ..... 26
Table 10. Results from body length (mm) measurements of control tadpoles at test termination of 12 different tests ..... 27
Table 11. Results from body weight (mg) measurements of control tadpoles at test termination of 12 different tests ..... 27
Table 12. Distribution of developmental stages of initial stage 51 *X. laevis* tadpoles exposed to PTU for 7, 14, and 21 days. ..... 29
Table 13. Summary of results from whole body length, hind limb length and body weight measurements during the stage 51 study with PTU in the GER lab. ..... 31
Table 14. Summary of results from developmental stage determination during the stage 51 study with PTU in the GER lab. ..... 32
Table 15. Summary of results from whole body length, hind limb length and body weight measurements during the stage 51 study with PTU in the JPN lab. ..... 33
Table 16. Summary of results from developmental stage determination during the stage 51 study with PTU in the JPN lab. ..... 34
Table 17. Summary of results from snout-to-vent length and body weight measurements during the stage 51 study with PTU in the US lab. ..... 35
Table 18. Summary of results from developmental stage determination during the stage 51 study with PTU in the US lab. ..... 36
Table 19. Distribution of developmental stages of initial stage 54 *X. laevis* tadpoles exposed to PTU for 7 and 14 days.. ..... 40
Table 20. Summary of results from whole body length, hind limb length and body weight measurements during the stage 54 study with PTU in the GER lab. ..... 41

Table 21. Summary of results from developmental stage determination during the stage 54 study with PTU in the GER lab. ..... 42
Table 22. Summary of results from whole body length, hind limb length and body weight measurements during the stage 54 study with PTU in the JPN lab. ..... 43
Table 23. Summary of results from developmental stage determination during the stage 54 study with PTU in the JPN lab. ..... 44
Table 24. Summary of results from snout-to-vent length and body weight measurements during the stage 54 study with PTU in the US lab. ..... 45
Table 25. Summary of results from developmental stage determination during the stage 54 study with PTU in the US lab. ..... 46
Table 26. Distribution of developmental stages of initial stage 51 *X. laevis* tadpoles exposed to T4 for 7, 14, and 21 days.. ..... 50
Table 27. Summary of results from whole body length, hind limb length and body weight measurements during the stage 51 study with T4 in the GER lab. ..... 51
Table 28. Summary of results from developmental stage determination during the stage 51 study with T4 in the GER lab. ..... 52
Table 29. Summary of results from whole body length, hind limb length and body weight measurements during the stage 51 study with T4 in the JPN lab. ..... 53
Table 30. Summary of results from developmental stage determination during the stage 51 study with T4 in the JPN lab. ..... 54
Table 31. Summary of results from snout-to-vent length and body weight measurements during the stage 51 study with T4 in the US lab. ..... 55
Table 32. Summary of results from developmental stage determination during the stage 51 study with T4 in the US lab. ..... 56
Table 33. Distribution of developmental stages of initial stage 54 *X. laevis* tadpoles exposed to T4 for 7 and 14 days.. ..... 59
Table 34. Summary of results from whole body length, hind limb length and body weight measurements during the stage 54 study with T4 in the GER lab. ..... 60
Table 35. Summary of results from developmental stage determination during the stage 54 study with T4 in the GER lab. ..... 61
Table 36. Summary of results from whole body length, hind limb length and body weight measurements during the stage 54 study with T4 in the JPN lab. ..... 62
Table 37. Summary of results from developmental stage determination during the stage 54 study with T4 in the JPN lab. ..... 63
Table 38. Summary of results from snout-to-vent length and body weight measurements during the stage 54 study with T4 in the US lab. ..... 64
Table 39. Summary of results from developmental stage determination during the stage 54 study with T4 in the US lab. ..... 65
Table 40. Comparison of statistical results from the three participating laboratories on the developmental stage of initial stage 51 *X. laevis* larvae exposed to PTU for 7, 14, and 21 days. ..... 70
Table 41. Comparison of statistical results from the three participating laboratories on the developmental stage of initial stage 54 *X. laevis* larvae exposed to PTU for 7 and 14 days. ..... 70
Table 42. Summary of endpoint sensitivities as detected in stage 51 studies with PTU. LOEC values (mg/L) as determined by the different endpoint measurements are shown. ..... 72
Table 43. Summary of endpoint sensitivities as detected in stage 54 studies with PTU. LOEC values (mg/L) as determined by the different endpoint measurements are shown. ..... 73
Table 44. Comparison of statistical results from the three participating laboratories on the developmental stage of initial stage 51 *X. laevis* larvae exposed to T4 for 7, 14, and 21 days. ..... 77
Table 45. Comparison of statistical results from the three participating laboratories on the developmental stage of initial stage 54 *X. laevis* larvae exposed to T4 for 7 and 14 days. ..... 77
Table 46. Summary of endpoint sensitivities as detected in stage 51 studies with T4 ..... 79

Table 47. Summary of endpoint sensitivities as detected in stage 54 studies with T4 ..... 79

**LIST OF FIGURES:**

Figure 1. Temporal changes in whole body length (WBL) as determined for control tadpoles during the conduct of stage 51 (21 days) and stage 54 (14 days) exposure studies.. ..... 27

Figure 2. Effects of PTU on gene expression in tadpole brain. RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 51 exposure study with PTU in the GER lab..... 38

Figure 3. Effects of PTU on gene expression in tadpole brain. RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 54 exposure study with PTU in the GER lab..... 48

Figure 4. Effects of T4 on gene expression in tadpole brain. RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 51 exposure study with T4 in the GER lab..... 58

Figure 5. Effects of T4 on gene expression in tadpole brain. RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 54 exposure study with T4 in the GER lab..... 67

**ACKNOWLEDGEMENT**

This report is the fruit of collaboration between three laboratories: Leibniz Institute of Freshwater Ecology and Inland Fisheries (Germany), the Mid-Continent Ecology Division, National Health and Environment Effects Research Laboratory (US EPA), Towa Kagaku Co., Research and Development Institute (Japan), and has been prepared mainly by Dr. Robert Opitz, Leibniz Institute of Freshwater Ecology and Inland Fisheries (Germany).

## SUMMARY

*i)* The purpose of this document is to summarize the results of an international round of testing and research aimed at developing and optimizing an amphibian-based screening assay for thyroid axis disruption. This effort originated at a meeting of the Amphibian Expert Group, an advisory group to the Validation Management Group, in June 2003 at a meeting hosted by the US Environmental Protection Agency in Duluth, MN, USA.

*ii)* Endocrine disruption by environmental chemicals is an international toxicological concern. As such, the OECD has been working with member countries on the validation and harmonization of testing methods for detecting chemicals that interfere with estrogen, androgen, and thyroid pathways. One such method, an amphibian-based screening assay for thyroid axis disruption, has received considerable attention due to the potential of this method to be a cost-effective screen for thyroid active chemicals. The basis for the assay is that amphibian metamorphosis is primarily under the control of thyroid hormone and that morphological changes typical of metamorphosis would be modulated by agonists and antagonists of the thyroid system. Before an amphibian-based metamorphosis assay can be adapted as a screen, several important issues need to be resolved, including: selection of an appropriate developmental period during metamorphosis to test, development of thyroid-related endpoints, establishment of optimal assay conditions, execution of the assay using chemicals with different modes of thyroid, and demonstrated repeatability among different laboratories. This document reports substantial progress on resolving these issues and culminates with a proposal for additional international research.

*iii)* Three laboratories participated in this research effort: IGB, Japan, and MED. All three laboratories had been working on development of a metamorphosis assay using slightly different methods with the African clawed frog, *Xenopus laevis*, as the model species. The objective of this work was to evaluate the methodologies of these different laboratories following exposure to identical compounds and to use the outcomes of these studies to guide the development of a proposed research direction that would lead to standardization and validation of an acceptable method.

*iv)* Prior to this collaborative research, little or no work had been conducted using identical protocols and chemicals among different laboratories. Therefore, there was insufficient information to eliminate or select one methodology over the other. Consequently, it was decided that the three participating labs would each use their specific methods to test the anti-thyroid compound, 6-propylthiouracil (PTU), and the receptor agonist, T4, at comparable exposure concentrations. These studies initiated the exposure at two different developmental stages and were conducted for either 2 or 3 weeks. The primary endpoints were final developmental stage, thyroid histology, and limb length.

*v)* In summary, these studies resulted in remarkably similar outcomes among the different laboratories, despite minor methodological differences. PTU inhibited and T4 accelerated metamorphic development, each in a concentration-dependent manner in experiments conducted in all three labs. The effective concentrations of these chemicals were essentially identical when similar endpoints were considered. These results suggest that the assay is relatively insensitive to minor methodological differences and constitutes a relatively robust system with potential for use in screening chemicals for thyroid axis disruption.

*vi)* Finally, this report proposes the next phase in the validation process. This phase will differ from the previous work in that the experiments will be conducted using identical methods. This approach should reduce the already minimal variance in the results and provide a more comparable data set to evaluate inter-laboratory variations. The chemicals selected for this phase include the thyroid receptor agonist, T4, a thyroid hormone synthesis inhibitor, sodium perchlorate, and a deiodinase inhibitor, iopanoic acid. The latter two chemicals were chosen because they represent different mechanisms of action and will expand our understanding of the responsiveness of this assay.

## INTRODUCTION

1. This report summarizes the results from experimental work conducted in three laboratories during Phase I of the validation of the Amphibian Metamorphosis Assay. The Amphibian Metamorphosis Assay was selected by the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA) as an *in vivo* assay for identification of substances with the potential to disrupt functions of the thyroid system. The need for the development and validation of an *in vivo* assay for detection of thyroid system-disrupting substances arises from concern that a considerable number of compounds have the potential to interact with different aspects of thyroid system function and thyroid hormone (TH) action (reviewed in Brucker-Davis, 1998; Zoeller, 2003). TH regulates a wide range of biological processes associated with development, somatic growth, metabolism, energy provision and reproduction in vertebrates and thus, exogenous substances that can interfere with thyroid system function could pose a significant hazard to human health and wildlife (Colborn, 2002; Zoeller, 2003).

2. The biological principle of the assay is that the process of postembryonic development (metamorphosis) in anuran amphibians is dependent on a functional hypothalamus-pituitary-thyroid gland axis and undisturbed action of TH in peripheral tissues. The South African clawed toad *Xenopus laevis* was selected as test organism for the assay because metamorphic development and the regulatory role played by TH during this process are well characterized in this species. Previous work by the laboratories participating in Phase I showed that in *X. laevis* tadpoles, metamorphic development can be precociously induced and/or accelerated by TH agonists whereas anti-thyroidal agents inhibit metamorphic development.

3. The first OECD *ad hoc* Expert Meeting on Amphibian Testing (June 26 – 27, 2003, Duluth, USA) reviewed and discussed existing testing approaches and protocols and agreed on an action plan for Phase I validation work (OECD, 2003). The two main outcomes of this Expert Meeting were that (I) *X. laevis* represents the primary candidate for a test species to be used in the Amphibian Metamorphosis Assay, (II) an exposure phase covering pre- and prometamorphic development but not metamorphic climax offers considerable potential for the development of a sensitive test protocol.

4. Experience from previous studies conducted in various laboratories indicated that a test protocol which includes exposure of tadpoles from late premetamorphic stages (e.g., stage 51) throughout late prometamorphic stages 58/59 would require a test duration of approximately 21 days. However, the Expert group also acknowledged the need for a short exposure duration due to the intended use of the assay for screening purposes. Therefore, an alternative 14-d test protocol was proposed involving exposure of tadpoles from early prometamorphic stage 54 throughout late prometamorphic stages 58/59.

5. Accordingly, the primary objective of validation Phase I was a comparative evaluation of the utility and sensitivity of the two proposed exposure scenarios to detect stimulating and inhibiting effects of thyroid system-disrupting substances on *X. laevis* metamorphosis. For this purpose, exposures were initiated with *X. laevis* tadpoles at developmental stages 51 and 54, respectively. Exposure of stage 51 tadpoles was continued for a total of 21 days and exposure of stage 54 tadpoles was continued for a total of 14 days. Tadpoles were exposed to 4 different concentrations of the test substance ($n$= 2 replicates per concentration) and a dilution water control group ($n$= 2 replicates). All exposure experiments used an aqueous route of exposure. The chemicals included in this testing were 6-propylthiouracil (PTU) and

thyroxine (T4). PTU is a well studied chemical known to inhibit thyroid hormone synthesis and T4 is the native prohormone. Test concentrations for both compounds were selected based on the experience of the three participating laboratories in conducting related work with *X. laevis*.

6. Participants of the first OECD *ad hoc* Expert meeting further agreed on a set of morphological, histological and molecular biological endpoints that should be evaluated during Phase I work with regard to their relevance, sensitivity and diagnostic value for detection of thyroid system-related effects caused by the test chemicals. In addition, efforts towards the standardization of endpoint measurements were regarded as another major objective of the Phase I validation studies.

7. To that end, determination of the developmental stage of test organisms according to the staging criteria of Nieuwkoop and Faber (1994) and qualitative histological analysis of the thyroid gland were used as core endpoints in Phase I studies. Optional endpoints included hind limb length measurements, quantitative morphometric analysis of thyroid gland histology and gene expression analysis in different tissues. Monitoring of tadpole growth and survival was considered as a means to identify possible toxic side effects of test compounds and therefore, determination of body length and body weight of the test organisms as well as daily recordings of mortality rates were included as endpoints in Phase I experimental work.

## OBJECTIVE OF THE INTER-LABORATORY COMPARISON STUDY

8. The overall goal of the Phase I validation study was to comparatively assess the utility and sensitivity of two different exposure scenarios to detect changes in metamorphic development and thyroid system function in response to substances considered to act as potent agonists (T4) and antagonists (PTU) of thyroid system function. The two testing scenarios tested 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.

9. Further objectives of Phase I validation work were to:

- evaluate the intra- and inter-laboratory variability of developmental and growth rates of control animals
- appraise the robustness of the assay when applied in slightly different experimental conditions
- obtain data on the intra- and inter-laboratory variability and reproducibility among the selected core endpoints of the assay
- compare the different endpoints with view to their relevance, sensitivity and diagnostic value
- compare the two testing scenarios (stage 51 and stage 54 studies) in terms of sensitivity
- identify general protocol changes and/or refinements to enhance reproducibility, sensitivity and diagnostic value of the assay
- provide a proposal for a testing protocol to be used in Phase II of the validation process

## METHODS

### Overview of Test Conditions

10. Prior to initiation of the exposure studies, standard operating procedures (SOP; see Annex 5) for the conduct of the experimental part of the studies were developed and approved by the Validation Management Group for Ecotoxicity Testing (VMG-eco) for use in Phase I validation activities. An overview of the test conditions and techniques applied in the three participating laboratories is given in Annex 1 of this document. The main difference between the labs was that the US lab and the JPN lab used flow-through conditions whereas the GER lab used a static-renewal exposure system. Further differences between the labs included:

- **type of dilution water**; the JPN lab used activated carbon filtered, UV-irradiated tap water, the US lab used filtered, UV sterilized Lake Superior water, and the GER lab used a synthetic test medium.
- **the type of diet**; SeraMicron was used in the GER and JPN lab; the US lab used a customized mixture of trout starter, algae, TetraFin and brine shrimp.
- **number of test animals per replicate**; 20 tadpoles were used per replicate tank in the GER and JPN lab; the US lab used 25 tadpoles per replicate tank in the PTU studies.
- **developmental stage determination**; stages were recorded in the GER and JPN lab at day 0, 7, 14 and 21 of exposure; the US lab determined this endpoints only at test termination.
- **body length measurement**; the GER and JPN lab determined whole body length (WBL) from the snout to the tip of the tail; the US lab determined this endpoint as snout-to-vent length (SVL).
- **thyroid gland histology**; the GER analyzed transverse sections of the lower jaw from dorsal to ventral; the JPN lab analyzed sagittal sections of the whole body from left to right and the US lab analyzed transverse sections of the head from caudal to rostral.

### Exposure Experiments

11. Each lab conducted a total of four exposure studies according to the scheme depicted in Table 1.

**Table 1. Overview of exposure studies conducted by the three participating labs**

| Test chemical | Concentrations* | Stage at test initiation | Duration |
|---|---|---|---|
| PTU | 2.5, 5, 10, 20 mg/L | 51 | 21 days |
| PTU | 2.5, 5, 10, 20 mg/L | 54 | 14 days |
| T4 | 0.25, 0.5, 1.0, 2.0 µg/L | 51 | 21 days |
| T4 | 0.25, 0.5, 1.0, 2.0 µg/L | 54 | 14 days |

* In addition to the concentrations listed above, the US lab also tested a PTU concentration of 1.25 mg/L and a T4 concentration of 4.0 µg/L

**Statistical Analysis**

12. Data sets for body length, hind limb length and body weight measurements were analyzed for normal distribution (Kolmogorov-Smirnov-test) and homogeneity of variance (Levene-test). Differences were considered significant at $p<0.05$. Upon verification of normality and variance homogeneity, two-sided Dunnett's test was used to compare data from the control group to all other treatment groups.

13. Developmental stage data come from a ranking process, and thus do not lend themselves to analysis via parametric methods. They were analyzed by using the non-parametric Kruskal-Wallis test followed by Dunn's multiple comparison test to compare data from the control group to all other treatment groups. The same approach was followed for data that did not satisfy the conditions of normality and variance heterogeneity. Differences in developmental indices between treatments were considered significant at the level of $p<0.05$.

14. Gene expression data were presumed to follow a log-normal distribution and thus were log-transformed to satisfy the criteria of normality and homogeneity of variance. Dunnett's test was used to compare control data to all other treatment groups. Differences were considered significant at $p<0.05$.

15. All statistical tests were two-sided. Statistical analyses were performed on pooled data from both replicates (i.e. based on individual tadpoles) rather than replicate means. However, statisticians commented that this approach should only be considered if there is reason to believe that the within-replicate correlations are small and the between-replicate variance is small. Otherwise, the variability of the responses can be seriously misrepresented. Such an approach can be explored for continuous responses, since multiple variance components can be estimated from the data, provided the conditions for within-tank normality are satisfied. Such an approach will be more difficult to justify for developmental and histological endpoints, since these are not continuous responses and standard variance components are not applicable.

## RESULTS

**Analytical Chemistry**

***Analytical Chemistry Results (US lab)***

16. PTU was measured four times for each replicate in the 21 day study and three times for each replicate in the 14 day study. Overall, the actual test concentrations were very close to the nominal concentrations as detailed in the following two tables.

17. Measured PTU concentrations (mg/L) in the NF51 study conducted for 21 days. Sample size for each replicate was 4.

**Table 2. Measured PTU concentrations (mg/L) in the stage 51 study conducted for 21 days. Sample size for each replicate was 4.**

| | Replicate 1 | | Replicate 2 | | Combined | |
|---|---|---|---|---|---|---|
| Nominal | mean | std | mean | std | mean | std |
| Control | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 1.25 | 1.54 | 0.04 | 1.51 | 0.05 | 1.52 | 0.05 |
| 2.50 | 2.77 | 0.04 | 2.81 | 0.03 | 2.79 | 0.04 |
| 5.00 | 5.54 | 0.05 | 5.56 | 0.05 | 5.55 | 0.05 |
| 10.00 | 10.78 | 0.11 | 10.84 | 0.10 | 10.81 | 0.10 |
| 20.00 | 21.21 | 0.22 | 21.19 | 0.17 | 21.20 | 0.19 |

**Table 3. Measured PTU concentrations (mg/L) in the stage 54 study conducted for 14 days. Sample size for each replicate was 3.**

| | Replicate 1 | | Replicate 2 | | Combined | |
|---|---|---|---|---|---|---|
| Nominal | mean | std | mean | std | mean | std |
| Control | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 1.25 | 1.54 | 0.02 | 1.56 | 0.01 | 1.55 | 0.02 |
| 2.50 | 2.86 | 0.03 | 2.86 | 0.03 | 2.86 | 0.03 |
| 5.00 | 5.48 | 0.04 | 5.39 | 0.04 | 5.43 | 0.06 |
| 10.00 | 10.63 | 0.07 | 10.79 | 0.07 | 10.71 | 0.11 |
| 20.00 | 20.82 | 0.11 | 20.93 | 0.25 | 20.88 | 0.18 |

18. T4 was measured two times for each replicate in the 21 day study and one or two times for each replicate in the 14 day study. Overall, the actual test concentrations were very close to the nominal concentrations as detailed in the following two tables.

**Table 4. Measured T4 concentrations (µg/L) in the stage 51 study conducted for 21 days. Sample size for each replicate was 2.**

| | Replicate 1 | | Replicate 2 | | Combined | |
|---|---|---|---|---|---|---|
| Nominal | mean | std | mean | std | mean | std |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.25 | 0.28 | 0.04 | 0.27 | 0.03 | 0.27 | 0.03 |
| 0.50 | 0.56 | 0.10 | 0.49 | 0.01 | 0.52 | 0.07 |
| 1.00 | 1.11 | 0.11 | 0.96 | 0.00 | 1.04 | 0.11 |
| 2.00 | 2.41 | 0.68 | 1.84 | 0.04 | 2.13 | 0.51 |
| 4.00 | 3.86 | 0.12 | 3.50 | 0.19 | 3.68 | 0.25 |

**Table 5. Measured T4 concentrations (µg/L) in the stage 54 study conducted for 14 days. Sample size for replicate 1 was 2 and for replicate 2 was 1.**

| | Replicate 1 | | Replicate 2 | | Combined | |
|---|---|---|---|---|---|---|
| Nominal | mean | std | mean | std | mean | std |
| 0.00 | 0.00 | 0.00 | 0.00 | na | 0.00 | 0.00 |
| 0.25 | 0.21 | 0.01 | 0.21 | na | 0.21 | 0.01 |
| 0.50 | 0.45 | 0.06 | 0.51 | na | 0.47 | 0.05 |
| 1.00 | 0.77 | 0.14 | 0.94 | na | 0.83 | 0.14 |
| 2.00 | 1.65 | 0.11 | 1.83 | na | 1.71 | 0.13 |
| 4.00 | 3.48 | 0.37 | 3.57 | na | 3.51 | 0.27 |

### *Analytical Chemistry Results (JPN lab)*

19. PTU was measured four times in the stage 51 study and three times in the stage 54 study. Only one set of replicate tanks was measured.

**Table 6. Measured PTU concentrations (mg/L) in the stage 51 study and the stage 54 study. Sample sizes of the stage 51 study and the stage 54 study were 4 and 3, respectively.**

| | stage 51 study | | stage 54 study | |
|---|---|---|---|---|
| Nominal | mean | SD | mean | SD |
| 2.50 | 2.03 | 0.78 | 2.93 | 0.75 |
| 5.00 | 4.35 | 2.17 | 5.15 | 0.28 |
| 10.00 | 11.20 | 1.91 | 11.17 | 0.37 |
| 20.00 | 27.62 | 5.98 | 24.19 | 0.42 |

20. T4 was measured four times in the stage 51 study and three times in the stage 54 study. Only one set of replicate tanks was measured.

**Table 7. Measured T4 concentrations (mg/L) in the stage 51 study and the stage 54 study. Sample size of the stage 51 study was 2 or 3, and sample size of the stage 54 study was 2.**

| | stage 51 study | | stage 54 study | |
|---|---|---|---|---|
| Nominal | mean | SD | mean | SD |
| 0.25 | 0.32 | 0.11 | 0.35 | 0.08 |
| 0.50 | 0.48 | 0.01 | 0.65 | 0.05 |
| 1.00 | 1.22 | 0.11 | 1.06 | 0.23 |
| 2.00 | 2.74 | 0.42 | 2.00 | 0.84 |

## Comparison of Control Data

21. The baseline performance of test organisms in different laboratories is an important consideration for evaluating the robustness of any standardized protocol intended for broad use. Analysis of inter-laboratory variability in control treatment groups included a comparison of the parameters mortality, developmental rate and growth rate.

22. Mortality was absent in the control groups of all 12 exposure studies conducted during Phase I validation indicating that the general rearing conditions selected for the assay allow for high survival rates of tadpoles. A first inspection of control data showed that developmental rates of untreated tadpoles were generally within the expected range as control tadpoles reached late prometamorphic stages within the exposure periods of 21 days (stage 51 study) and 14 days (stage 54 study). The inter-laboratory differences in the developmental rate of the control treatments are highlighted in Table 8 for the stage 51 studies and Table 9 in for the stage 54 studies. In three of the four studies conducted in the JPN lab, (PTU stage 54; T4 stage 51; and T4 stage 54), tadpole development was slightly retarded in comparison to the US and GER labs which demonstrated very similar developmental rates despite different conditions used in the studies. While there may be several factors involved, reduced availability of food in the JPN studies is likely the most important factor. This conclusion is based on unpublished data from the US lab where the two different diets were evaluated using the related species, *X. tropicalis*. In that study, different concentrations of SeraMicron 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 SeraMicron feeding rates. The interpretation of these data are that SeraMicron is a fine particulate that stays suspended in the water column for a long period of time and is subject to washing out with the flow, resulting in reduced availability. Therefore, if SeraMicron is chosen as the preferred diet under flow through conditions, feeding rates higher than those used in the JPN lab in this exercise should be used.

**Table 8. Distribution of developmental stages reached by control tadpoles within 21 days during exposure studies initiated at stage 51 in three different labs.**

| | | Developmental stage | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | median | range[a] |
| PTU[b] | US | | 2 | 13 | 18 | 4 | 7 | 5 | 1 | 59 | 7 |
| | JPN | | 16 | 13 | 8 | 2 | 1 | | | 58 | 5 |
| | GER | 2 | 17 | 4 | 14 | 2 | 1 | | | 58 | 6 |
| | | | | | | | | | | | |
| T4[b] | US | | 6 | 4 | 15 | 8 | 4 | 3 | | 59 | 6 |
| | JPN | 2 | 3 | 16 | 12 | 4 | 3 | | | 58 | 6 |
| | GER | | | 3 | 15 | 11 | 4 | 7 | | 60 | 5 |

Note: [a] value indicates the number of different stages determined for the control group

[b] test chemical used in the exposure study for which control data are shown

**Table 9. Distribution of developmental stages reached by control tadpoles within 14 days during exposure studies initiated at stage 54 in three different labs.**

| | | Developmental stage | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | median | range[a] |
| PTU[b] | US | | 11 | 19 | 18 | 2 | | | | 58 | 4 |
| | JPN | 6 | 32 | 2 | | | | | | 57 | 3 |
| | GER | 1 | 14 | 14 | 9 | 2 | | | | 58 | 5 |
| | | | | | | | | | | | |
| T4[b] | US | 1 | 14 | 11 | 13 | 1 | | | | 58 | 5 |
| | JPN | | 25 | 6 | 7 | 2 | | | | 57 | 4 |
| | GER | | 6 | 15 | 17 | 2 | | | | 58 | 4 |

Note: [a] value indicates the number of different stages determined for the control group

[b] test chemical used in the exposure study for which control data are shown

23. Growth of tadpoles was assessed by means of whole body length (WBL) measurements throughout the exposure periods of 21 and 14 days, respectively. Body weight was only determined at test termination. Temporal changes in mean WBL of the control groups as determined in the JPN and GER labs are illustrated in Figure 1. At test initiation of stage 51 exposure experiments, mean WBL ranged between 24.0 and 28.1 mm. In the stage 54 exposure experiments, mean WBL ranged between 37.5 and 43.2 mm at test initiation. In all experiments, mean WBL values increased from day 0 of exposure to test termination. There were only slight differences in tadpole growth rates between the JPN and GER labs. The only exception was the stage 51 exposure study with T4 conducted in the JPN lab where lower growth rates were determined for the control group.

24. In the stage 51 exposure experiments, increases in WBL were relatively constant during the initial 14 days of exposure. However, particularly in the GER experiments, it was also observed that growth of tadpoles was reduced between exposure days 14 and 21. This was most likely due to the fact that tadpole growth ceased at stages 58/59 followed by a reduction in WBL due to reshaping of the head region and initiation of tail resorption.

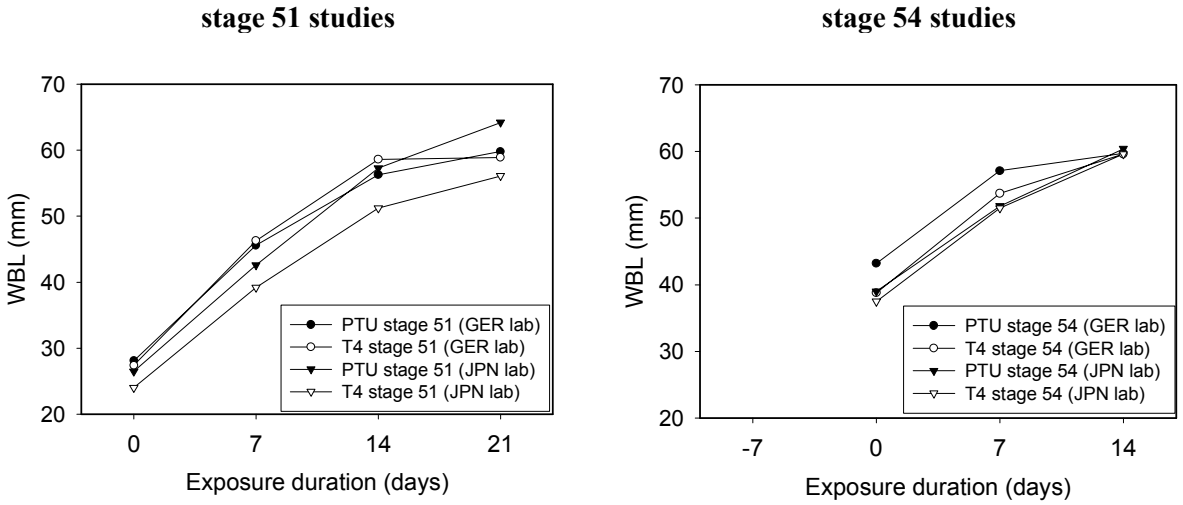


**Figure 1. Temporal changes in whole body length (WBL) as determined for control tadpoles during the conduct of stage 51 (21 days) and stage 54 (14 days) exposure studies. Data from 8 independent experiments performed in the JPN and GER lab are shown. Mean values are shown while standard deviation bars are omitted to reduce clutter.**

25. Table 10 and Table 11 summarize the results from measurements of WBL, snout-to-vent length (SVL) and body weight at test termination for all studies conducted during Phase I validation work. Although results from WBL and body weight measurements during the stage 51 exposure studies conducted in the JPN lab indicated an increased intra-laboratory variability in tadpole growth, the overall comparison of tadpole growth rates showed a high similarity between laboratories during Phase I validation work.

**Table 10. Results from body length (mm) measurements of control tadpoles at test termination of 12 different tests. Means and standard deviations were calculated for pooled control data of each exposure study. Note that the US lab measured body length as snout-to-vent length (SVL) whereas the JPN and GER labs measured whole body length (WBL) from the tip of the snout to the tip of the tail.**

| | | US lab (SVL) | GER lab (WBL) | JPN lab (WBL) | mean (WBL) | CV[a] (WBL) |
|---|---|---|---|---|---|---|
| 21 d[b] | PTU[c] | 19.5 ± 1.9 | 59.8 ± 2.6 | 64.2 ± 3.9 | | |
| 21 d | T4 | 19.5 ± 1.7 | 58.9 ± 4.2 | 56.1 ± 3.4 | 59.7 ± 2.9 | 4.8 % |
| | | | | | | |
| 14 d | PTU | 19.9 ± 1.2 | 59.7 ± 2.9 | 60.4 ± 2.9 | | |
| 14 d | T4 | 19.2 ± 0.9 | 59.6 ± 2.4 | 59.6 ± 3.2 | 59.8 ± 0.3 | 0.6 % |

Note: [a] coefficient of variation; [b] test duration, [c] test chemical used in the exposure study for which control data are shown.

**Table 11. Results from body weight (mg) measurements of control tadpoles at test termination of 12 different tests. Means and standard deviations were calculated for pooled control data of each exposure study.**

| | | US lab | GER lab | JPN lab | mean | CV[a] |
|---|---|---|---|---|---|---|
| 21 d[b] | PTU[c] | 1047 ± 230 | 919 ± 162 | 1065 ± 153 | | |
| 21 d | T4 | 1042 ± 202 | 824 ± 169 | 761 ± 127 | 943 ± 118 | 12.5 % |
| | | | | | | |
| 14 d | PTU | 1069 ± 161 | 957 ± 139 | 896 ± 113 | | |
| 14 d | T4 | 943 ± 157 | 861 ± 117 | 882 ± 131 | 934 ± 69 | 7.4 % |

Note: [a] coefficient of variation; [b] test duration, [c] test chemical used in the exposure study for which control data are shown

**Exposure Studies with Propylthiouracil (PTU)**

26. PTU was used as a model test compound with reported anti-thyroidal activities in mammals (Capen, 1996) as well as in anuran tadpoles (Goos *et al.*, 1968). Effects of PTU on the thyroid system of *Xenopus* tadpoles were comparatively assessed in a 21 day exposure study initiated with stage 51 tadpoles (stage 51 study) and in a 14 day exposure study initiated with stage 54 tadpoles (stage 54 study). Both exposure studies were performed in parallel with tadpoles from the same spawn. In both exposure scenarios, PTU treatment comprised nominal concentrations of 2.5, 5, 10 and 20 mg/L (the US lab also tested 1.25 mg/L PTU). Mortality was negligible in all exposure studies with PTU (data not shown).

***Stage 51 Exposure Studies with PTU (21 Day Assay)***

*Developmental stage*

27. Results from developmental stage determination in the different labs during the stage 51 exposure studies with PTU are summarized in Table 12. No significant effects on developmental stage were observed following 7 days of exposure to PTU in either the JPN or GER study. The US lab did not assess apical endpoints prior to test termination at day 21. Significant developmental retardation was observed at 5.0 and 20 mg/L PTU following 14 days of exposure in the JPN lab but not in the GER lab. All three labs showed significant differences following 21 days of exposure to 20 mg/L. In addition, the GER study showed a significant retardation following 21 days of exposure to 10 mg/L. The significant difference at 5 mg/L after 14 days of exposure in JPN 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 JPN lab does not persist at 21 days, suggesting that this observation is not real.

**Table 12. Distribution of developmental stages of initial stage 51 *X. laevis* tadpoles exposed to PTU for 7, 14, and 21 days. Open boxes highlight the normal development of the controls. Shaded boxes indicate statistical difference compared to controls by the Dunn's method (p<0.05).**

| | PTU Conc. mg/L | Stage at 7 days | | | | | Stage at 14 days | | | | | | Stage at 21 days | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 51 | 52 | 53 | 54 | 55 | 52 | 53 | 54 | 55 | 56 | 57 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 |
| Japan | 0.0 | | | 6 | 32 | 2 | | | | 2 | 18 | 20 | | | | | 16 | 13 | 8 | 2 | 1 | | |
| | 2.5 | | 1 | 2 | 37 | | | | | 4 | 22 | 13 | | | | 2 | 7 | 17 | 5 | 3 | 4 | 1 | |
| | 5 | | 1 | 8 | 30 | 1 | | | 3 | 10 | 23 | 3 | | | 2 | 2 | 17 | 14 | 3 | 0 | 1 | | |
| | 10 | | 1 | 2 | 35 | 2 | | | 2 | 9 | 18 | 11 | | | 1 | 4 | 7 | 11 | 12 | 5 | | | |
| | 20 | 2 | 3 | 8 | 27 | | 1 | 4 | 21 | 10 | 2 | 1 | | 14 | 10 | 5 | 6 | 2 | | | | | |
| | | **51** | **52** | **53** | **54** | **55** | **52** | **53** | **54** | **55** | **56** | **57** | **53** | **54** | **55** | **56** | **57** | **58** | **59** | **60** | **61** | **62** | **63** |
| Germany | 0.0 | | | 2 | 32 | 6 | | | | 4 | 34 | 2 | | | | 2 | 17 | 4 | 14 | 2 | 1 | | |
| | 2.5 | | | 3 | 37 | | | | | 9 | 30 | 1 | | | | 3 | 20 | 9 | 7 | 0 | 1 | | |
| | 5 | | | 1 | 34 | 5 | | | | 5 | 32 | 3 | | | | | 24 | 9 | 7 | | | | |
| | 10 | | | 1 | 38 | 1 | | | | 9 | 31 | | | | | 3 | 25 | 9 | 3 | | | | |
| | 20 | | | 10 | 29 | 1 | | | 1 | 13 | 26 | | | | 1 | 4 | 24 | 8 | 3 | | | | |
| | | | | | | | | | | | | | **53** | **54** | **55** | **56** | **57** | **58** | **59** | **60** | **61** | **62** | **63** |
| US | 0.0 | | | | | | | | | | | | | | | | 2 | 13 | 18 | 4 | 7 | 5 | 1 |
| | 1.25 | | | | | | | | | | | | | | | | 4 | 21 | 14 | 6 | 2 | 3 | |
| | 2.5 | | | | | | | | | | | | | | | | 9 | 5 | 23 | 8 | 1 | 3 | 1 |
| | 5 | | | | | | | | | | | | | | | | 10 | 12 | 14 | 7 | 2 | 5 | |
| | 10 | | | | | | | | | | | | | | | 1 | 7 | 12 | 22 | 2 | 4 | 2 | |
| | 20 | | | | | | | | | | | | 2 | 20 | 13 | 4 | 5 | 3 | 2 | | | | |

*Hind Limb Length*

28. In addition to stage determinations, measurements of hind limb length were used in the GER (Table 13) and JPN study (Table 15) to assess effects of PTU on hind limb morphogenesis. In both studies, exposure to the highest PTU concentration (20 mg/L) resulted in a significant retardation of hind limb growth at all timepoints studied (7, 14, 21 days). In the GER study, 10 mg/L PTU also caused a significant reduction in hind limb growth following 14 and 21 days of exposure. In the JPN study, 5.0 mg/L but not 10 mg/L PTU caused a significant reduction in hind limb length following 14 and 21 days of exposure.

*Body Length*

29. Effects of PTU on tadpole growth were examined in the GER and JPN study by WBL measurements (Table 13 ; Table 15). In the GER study, weak growth-retarding effects were observed following 7 and 14 days of exposure to 10 and 20 mg/L PTU. Effects of PTU on tadpole growth were more variable in the JPN study. Similar to the GER study, 20 mg/L PTU caused a significant reduction in WBL following 7 and 14 days of exposure. 10 mg/L PTU caused a significant increase in WBL after7 days whereas 5 mg/L PTU caused a significant reduction in WBL after 14 days. At test termination (day 21), no significant effects of PTU treatment on WBL were detectable. The US lab measured SVL at test termination only and determined a significantly increased SVL at 20 mg/L PTU.

*Body Weight*

30. Effects of PTU on tadpole growth were further examined in all labs by body weight measurements at test termination. No significant effects of PTU treatment on tadpole body weight were detectable in the GER study (Table 13). In the JPN study, PTU treatment at 5, 10 and 20 mg/L significantly decreased body weight (Table 15). In contrast, mean body weight of tadpoles exposed to 20 mg/L PTU was significantly increased in the US study (Table 17).

**Table 13. Summary of results from whole body length, hind limb length and body weight measurements during the stage 51 study with PTU in the GER lab.**

| | | | Test substance: PTU | | | | Lab: GER |
|---|---|---|---|---|---|---|---|
| | | **control** | **2.5 mg/L** | **5.0 mg/L** | **10 mg/L** | **20 mg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(mg/L)** |
| | | **whole body length (mm)** | | | | | |
| **0** | A | 28.2 ± 1.0 | 28.2 ± 0.9 | 27.8 ± 1.0 | 27.5 ± 0.8 | 28.0 ± 1.0 | |
| | **B** | 28.0 ± 1.0 | 28.0 ± 1.1 | 28.1 ± 1.2 | 27.7 ± 0.8 | 27.9 ± 0.9 | |
| | **Pool** | 28.1 ± 1.0 | 28.1 ± 1.0 | 28.0 ± 1.0 | 27.6 ± 0.8 | 27.9 ± 1.0 | |
| **7** | A | 46.1 ± 2.5 | 43.8 ± 2.0 | 44.0 ± 2.1 | 42.5 ± 2.2 | 44.1 ± 2.6 | |
| | **B** | 45.2 ± 2.6 | 44.7 ± 3.0 | 45.3 ± 2.0 | 43.9 ± 2.0 | 43.7 ± 2.6 | |
| | **Pool** | 45.6 ± 2.5 | 44.2 ± 2.6 * | 44.6 ± 2.1 | 43.2 ± 2.2 * | 43.9 ± 2.6 * | **10 #** |
| **14** | **A** | 56.9 ± 3.6 | 54.1 ± 3.0 | 54.2 ± 2.6 | 53.5 ± 3.0 | 54.4 ± 3.4 | |
| | **B** | 55.7 ± 3.1 | 55.3 ± 4.1 | 55.9 ± 2.5 | 54.4 ± 3.5 | 54.7 ± 3.1 | |
| | **Pool** | 56.3 ± 3.3 | 54.7 ± 3.6 | 55.1 ± 2.6 | 53.9 ± 3.2 * | 54.5 ± 3.1 * | **10** |
| **21** | **A** | 59.6 ± 2.7 | 58.1 ± 3.0 | 58.4 ± 2.6 | 58.0 ± 2.3 | 58.5 ± 3.9 | |
| | **B** | 60.0 ± 2.6 | 58.8 ± 3.2 | 59.3 ± 2.1 | 58.6 ± 3.3 | 58.9 ± 3.2 | |
| | **Pool** | 59.8 ± 2.6 | 58.5 ± 3.0 | 58.8 ± 2.4 | 58.3 ± 2.8 | 58.7 ± 3.5 | **ns** |
| | | **hind limb length (mm)** | | | | | |
| **7** | A | 2.3 ± 0.3 | 2.1 ± 0.2 | 2.2 ± 0.3 | 2.1 ± 0.2 | 2.1 ± 0.2 | |
| | **B** | 2.3 ± 0.3 | 2.0 ± 0.3 | 2.2 ± 0.3 | 2.1 ± 0.3 | 2.0 ± 0.3 | |
| | **Pool** | 2.3 ± 0.3 | 2.1 ± 0.2 * | 2.2 ± 0.3 | 2.1 ± 0.2 | 2.0 ± 0.2 * | **20 #** |
| **14** | **A** | 5.3 ± 1.0 | 4.4 ± 0.8 | 4.5 ± 0.6 | 4.3 ± 0.9 | 4.0 ± 0.8 | |
| | **B** | 5.0 ± 0.9 | 5.1 ± 0.9 | 5.0 ± 0.8 | 4.4 ± 0.5 | 4.5 ± 0.8 | |
| | **Pool** | 5.1 ± 0.9 | 4.7 ± 0.9 | 4.8 ± 0.7 | 4.3 ± 0.7 * | 4.2 ± 0.8 * | **10** |
| **21** | **A** | 12.1 ± 3.5 | 8.8 ± 2.5 | 9.3 ± 2.0 | 8.9 ± 2.5 | 8.0 ± 2.7 | |
| | **B** | 10.3 ± 3.1 | 10.6 ± 3.3 | 10.3 ± 2.1 | 8.8 ± 2.2 | 8.8 ± 2.6 | |
| | **Pool** | 11.2 ± 3.3 | 9.7 ± 3.0 | 9.8 ± 2.0 | 8.9 ± 2.3 * | 8.4 ± 2.6 * | **10** |
| | | **body weight (mg)** | | | | | |
| **21** | A | 878.5 ± 139.2 | 853.9 ± 163.7 | 824.8 ± 116.7 | 827.5 ± 112.3 | 838.4 ± 191.0 | |
| | **B** | 959.2 ± 179.9 | 875.9 ± 183.9 | 895.0 ± 130.2 | 861.9 ± 188.5 | 844.8 ± 167.2 | |
| | **Pool** | 918.8 ± 161.9 | 864.9 ± 170.1 | 859.9 ± 125.5 | 844.7 ± 152.2 | 841.6 ± 175.0 | **ns** |

ns no significant effects; asterisks denote significant differences from the control group ($p < 0.05$; Dunnett's Test)

# no concentration response relationship

**Table 14. Summary of results from developmental stage determination during the stage 51 study with PTU in the GER lab.**

| | | | | Test substance: PTU | | | | | | | | Lab: GER |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 2.5 mg/L | | 5.0 mg/L | | 10 mg/L | | 20 mg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | mg/L |
| **7** | **A** | 54 | 54-55 | 54 | 53-54 | 54 | 53-55 | 54 | 53-54 | 54 | 53-54 | |
| | **B** | 54 | 53-55 | 54 | 53-54 | 54 | 53-54 | 54 | 54-55 | 54 | 53-55 | |
| | **Pool** | **54** | **53-55** | **54** | **53-54** | **54** | **53-55** | **54** | **53-55** | **54** | **53-55** | **ns** |
| **14** | **A** | 56 | 55-57 | 56 | 55-56 | 56 | 55-57 | 56 | 55-56 | 55 | 54-56 | |
| | **B** | 56 | 55-56 | 56 | 55-57 | 56 | 55-57 | 56 | 55-56 | 56 | 55-56 | |
| | **Pool** | **56** | **55-57** | **56** | **55-57** | **56** | **55-57** | **56** | **55-56** | **56** | **54-56** | **ns** |
| **21** | **A** | 59 | 57-61 | 57 | 56-59 | 57 | 57-59 | 57 | 57-59 | 57 | 55-59 | |
| | **B** | 57 | 56-59 | 57.5 | 56-61 | 57 | 57-59 | 57 | 56-59 | 57 | 56-59 | |
| | **Pool** | **58** | **56-61** | **57** | **56-61** | **57** | **57-59** | **57** | **56-59** | **57** | **55-59** | **10** |

ns no significant effects; shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

**Table 15. Summary of results from whole body length, hind limb length and body weight measurements during the stage 51 study with PTU in the JPN lab.**

| | | | Test substance: PTU | | | | Lab: JPN |
|---|---|---|---|---|---|---|---|
| | | **control** | **2.5 mg/L** | **5.0 mg/L** | **10 mg/L** | **20 mg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(mg/L)** |
| | | **whole body length (mm)** | | | | | |
| **0** | A | 26.7 ± 1.6 | 26.3 ± 1.7 | 26.2 ± 1.7 | 26.4 ± 2.0 | 26.7 ± 1.3 | |
| | **B** | 26.4 ± 1.9 | 26.1 ± 1.8 | 26.4 ± 2.1 | 26.8 ± 2.4 | 26.5 ± 1.7 | |
| | **Pool** | 26.5 ± 1.7 | 26.2 ± 1.7 | 26.3 ± 1.9 | 26.6 ± 2.1 | 26.6 ± 1.5 | |
| **7** | A | 42.9 ± 3.0 | 44.7 ± 3.7 | 42.7 ± 3.0 | 44.8 ± 3.9 | 36.8 ± 5.5 | |
| | **B** | 42.2 ± 3.4 | 44.4 ± 3.1 | 39.7 ± 4.3 | 46.9 ± 3.9 | 43.1 ± 2.4 | |
| | **Pool** | 42.6 ± 3.1 | 44.6 ± 3.3 | 41.2 ± 3.9 | 45.9 ± 3.9 * | 39.9 ± 5.2 * | **10 #** |
| **14** | **A** | 56.9 ± 5.3 | 58.1 ± 4.4 | 55.9 ± 3.7 | 56.9 ± 4.5 | 49.5 ± 10.8 | |
| | **B** | 57.7 ± 3.6 | 57.8 ± 2.7 | 50.5 ± 5.1 | 57.5 ± 5.2 | 57.3 ± 3.6 | |
| | **Pool** | 57.3 ± 4.4 | 57.9 ± 3.5 | 53.3 ± 5.1 * | 57.2 ± 4.7 | 53.5 ± 8.7 * | **20 #** |
| **21** | **A** | 63.2 ± 4.6 | 64.5 ± 3.4 | 63.8 ± 3.1 | 63.8 ± 4.1 | 61.1 ± 8.7 | |
| | **B** | 65.3 ± 2.7 | 64.1 ± 2.7 | 60.7 ± 5.2 | 63.0 ± 4.6 | 65.7 ± 3.7 | |
| | **Pool** | 64.2 ± 3.9 | 64.3 ± 3.0 | 62.3 ± 4.4 | 63.4 ± 4.3 | 63.6 ± 6.7 | **ns** |
| | | **hind limb length (mm)** | | | | | |
| **7** | A | 2.2 ± 0.4 | 2.3 ± 0.4 | 2.2 ± 0.4 | 2.2 ± 0.4 | 1.8 ± 0.5 | |
| | **B** | 2.2 ± 0.4 | 2.4 ± 0.5 | 2.0 ± 0.4 | 2.4 ± 0.5 | 2.0 ± 0.3 | |
| | **Pool** | 2.2 ± 0.4 | 2.3 ± 0.4 | 2.1 ± 0.4 | 2.3 ± 0.4 | 1.9 ± 0.4 * | **20** |
| **14** | **A** | 5.6 ± 1.1 | 5.8 ± 1.4 | 5.4 ± 1.0 | 5.4 ± 1.1 | 3.0 ± 1.0 | |
| | **B** | 5.7 ± 1.1 | 6.0 ± 1.3 | 4.3 ± 1.1 | 5.7 ± 1.4 | 3.2 ± 1.1 | |
| | **Pool** | 5.7 ± 1.1 | 5.9 ± 1.3 | 4.8 ± 1.2 * | 5.6 ± 1.2 | 3.1 ± 1.1 * | **20 #** |
| **21** | **A** | 12.4 ± 3.2 | 13.7 ± 4.0 | 12.1 ± 3.1 | 12.1 ± 3.2 | 4.6 ± 2.2 | |
| | **B** | 13.3 ± 2.9 | 13.6 ± 3.6 | 9.4 ± 2.7 | 12.9 ± 4.3 | 5.6 ± 3.2 | |
| | **Pool** | 12.8 ± 3.0 | 13.7 ± 3.7 | 10.8 ± 3.1 * | 12.5 ± 3.7 | 5.1 ± 2.7 * | **20 #** |
| | | **body weight (mg)** | | | | | |
| **21** | A | 1028 ± 150.2 | 1053 ± 159.1 | 948 ± 125.2 | 969 ± 170.7 | 819 ± 261.0 | |
| | **B** | 1101 ± 154.1 | 996 ± 119.1 | 775 ± 171.9 | 965 ± 150.8 | 995 ± 167.9 | |
| | **Pool** | 1065 ± 152.7 | 1023 ± 139.3 | 864 ± 169.6 * | 967 ± 157.0 * | 914 ± 227.2 * | **5.0** |

ns no significant effects; asterisks denote significant differences from the control group ($p < 0.05$; Dunnett's Test)

# no concentration response relationship

**Table 16. Summary of results from developmental stage determination during the stage 51 study with PTU in the JPN lab.**

| | | | | Test substance: PTU | | | | | | | | Lab: JPN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 2.5 mg/L | | 5.0 mg/L | | 10 mg/L | | 20 mg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | mg/L |
| **7** | **A** | 54 | 53-55 | 54 | 52-54 | 54 | 53-54 | 54 | 52-54 | 54 | 51-54 | |
| | **B** | 54 | 53-54 | 54 | 53-54 | 54 | 52-55 | 54 | 53-55 | 54 | 52-54 | |
| | **Pool** | **54** | **53-55** | **54** | **52-54** | **54** | **52-55** | **54** | **52-55** | **54** | **51-54** | **ns** |
| **14** | **A** | 56 | 55-57 | 56 | 55-57 | 56 | 55-57 | 56 | 54-57 | 54 | 52-55 | |
| | **B** | 57 | 55-57 | 56 | 55-57 | 55 | 54-57 | 56 | 54-57 | 54 | 53-57 | |
| | **Pool** | **56.5** | **55-57** | **56** | **55-57** | **56** | **54-57** | **56** | **54-57** | **54** | **52-57** | **20 #** |
| **21** | **A** | 57.5 | 57-61 | 58 | 56-61 | 58 | 57-61 | 58 | 55-60 | 55 | 54-57 | |
| | **B** | 58 | 57-60 | 58 | 56-62 | 57 | 55-59 | 58 | 56-60 | 55 | 54-58 | |
| | **Pool** | **58** | **57-61** | **58** | **56-62** | **57** | **55-61** | **58** | **55-60** | **55** | **54-58** | **20** |

ns no significant effects; # no concentration response relationship

shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

**Table 17. Summary of results from snout-to-vent length and body weight measurements during the stage 51 study with PTU in the US lab.**

| | | | Test substance: PTU | | | | Lab: US |
|---|---|---|---|---|---|---|---|
| | | **control** | **2.5 mg/L** | **5.0 mg/L** | **10 mg/L** | **20 mg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(mg/L)** |
| | | **snout-to-vent length (mm)** | | | | | |
| **21** | **A** | 19.7 ± 1.6 | 19.2 ± 1.3 | 19.6 ± 2.0 | 19.8 ± 1.6 | 21.6 ± 1.4 | |
| | **B** | 19.5 ± 2.2 | 19.6 ± 2.0 | 20.5 ± 1.7 | 20.1 ± 1.6 | 22.0 ± 1.3 | |
| | **Pool** | 19.5 ± 1.9 | 19.4 ± 1.5 | 20.0 ± 1.9 | 19.7 ± 1.6 | 21.8 ± 1.3 * | **20** |
| | | **body weight (mg)** | | | | | |
| **21** | A | 1084.4 ± 194.5 | 967.2 ± 145.8 | 1043.3 ± 220.9 | 1084.3 ± 174.1 | 1152.1 ± 208.1 | |
| | **B** | 1009.5 ± 263.8 | 1040.1 ± 179.9 | 1119.8 ± 212.5 | 1066.6 ± 198.8 | 1272.6 ± 190.5 | |
| | **Pool** | 1047.0 ± 230.2 | 1003.7 ± 164.5 | 1081.5 ± 215.8 | 1075.5 ± 183.3 | 1211.1 ± 204.6 * | **20** |

asterisks denote significant differences from the control group (p< 0.05; Dunnett's Test)

**Table 18. Summary of results from developmental stage determination during the stage 51 study with PTU in the US lab.**

| | | | | Test substance: PTU | | | | | | | | Lab: US |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 2.5 mg/L | | 5.0 mg/L | | 10 mg/L | | 20 mg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | mg/L |
| **21** | **A** | 59 | 57-62 | 59 | 57-62 | 59 | 57-62 | 59 | 57-62 | 55 | 54-59 | |
| | **B** | 59 | 58-63 | 59 | 57-63 | 58 | 57-62 | 59 | 56-62 | 55 | 53-59 | |
| | **Pool** | **59** | **57-63** | **59** | **57-63** | **59** | **57-62** | **59** | **56-62** | **55** | **53-59** | **20** |

shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

*Histopathology*

31. The detailed reports of the histopathological analyses conducted by the US lab, GER lab and JPN lab are presented in Annex 2, Annex 3, and Annex 4, respectively. In summary, these analyses revealed exposure-related changes in the thyroid gland, which included distension of thyroid follicles, diffuse enlargement of the thyroid glands, colloid depletion and follicular cell hyperplasia. A low incidence of changes in the thyroid gland (minimal distension of thyroid follicles) was observed for tadpoles exposed to the lower PTU concentrations of 1.25 mg/L (only in the US study) and 2.5 mg/L. However, at higher PTU concentrations, follicular distension accompanied by diffuse enlargement of the thyroid glands increased in prevalence and severity in a concentration-dependent manner. An increase in the thickness of the epithelial cell layer (follicular cell hypertrophy) was observed at the two highest PTU concentrations of 10 and 20 mg/L. Hyperplasia of follicular cells was prominent in the 20 mg/L PTU treatment group but was also evident in some thyroid glands from the 10 mg/L PTU treatment. The degree of colloid depletion was also enhanced at the two highest PTU concentration and collapsed follicles devoid of colloid were detected in some glands from the 20 mg/L PTU treatment. Quantitative methods (morphometric analysis) were successfully applied to confirm the increase in epithelial cell height and the enlargement of the thyroid gland.

*Gene Expression in Brain/Pituitary*

32. Samples of whole brain tissue including the pituitary (brain/pituitary) taken at test termination of the GER study were analysed for changes in gene expression by means of semi-quantitative RT-PCR. Results from these RT-PCR analyses revealed increased expression of the β-subunit of thyrotropin (TSHβ) at 20 mg/L PTU (Figure 2). In addition, mRNA expression levels of several genes that are positively regulated by thyroid hormones were also analysed in brain/pituitary tissue. Thyroid hormone receptor β (TRβ) mRNA expression was slighly reduced at 10 and 20 mg/L PTU though these differences were not statistically significant. Reduced mRNA expression levels were also observed at 10 and 20 mg/L PTU for other genes including basic transcription element binding protein (BTEB), b/ZIP, prolactin, and type III monodeiodinase. However, due to the relatively small sample number that has been analysed so far (*n*=3-5 per treatment group), effects did not always reach statistical significance.

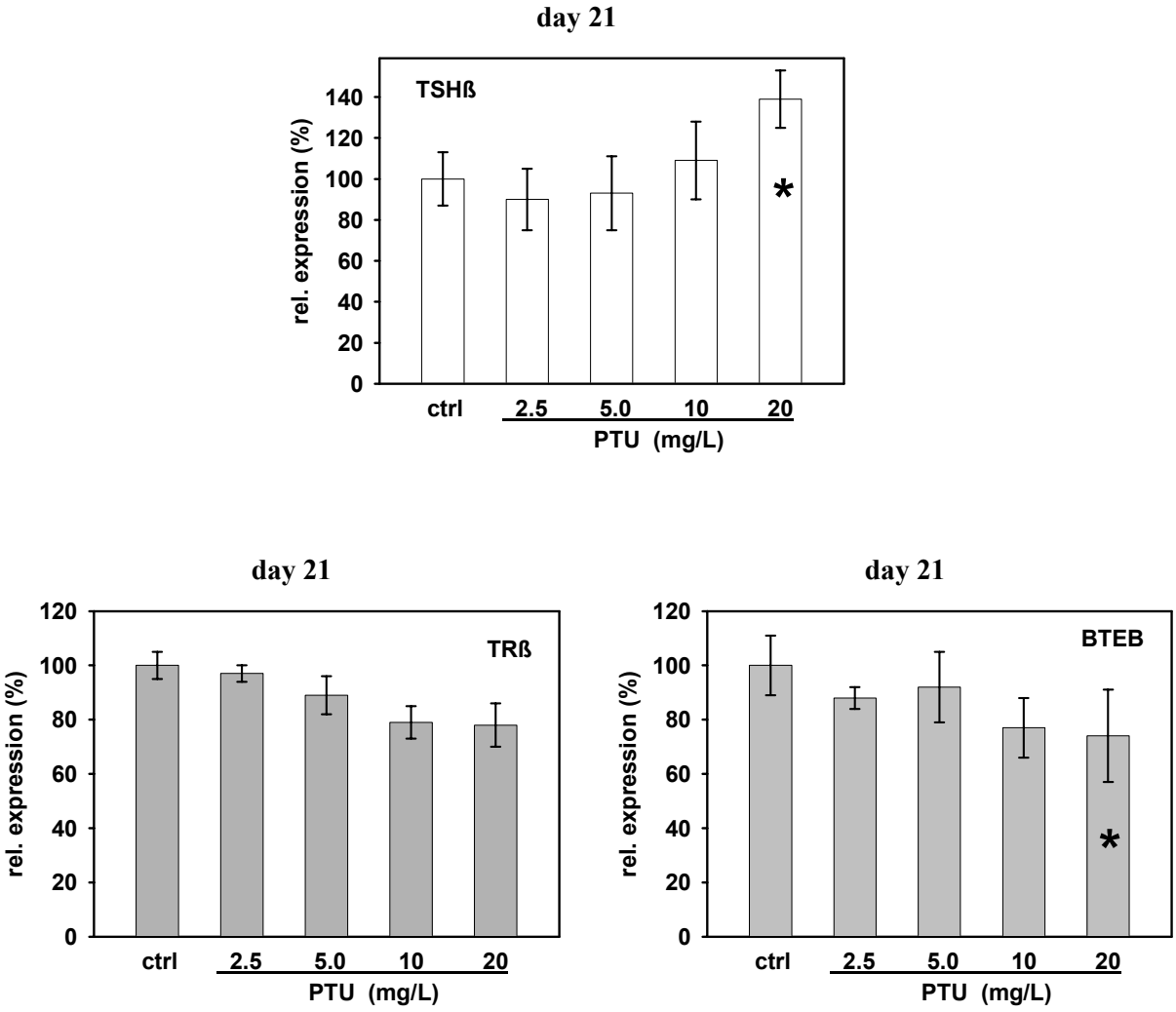


**Figure 2. Effects of PTU on gene expression in tadpole brain.**

RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 51 exposure study with PTU in the GER lab. Semiquantitative RT-PCR analyses of mRNA expression of thyrotropin β -subunit (TSH β), thyroid hormone receptor β (TR β) and basic transcription element binding protein (BTEB) were performed and results from densitometric analysis of scanned agarose gels are shown. Results were expressed relative to the control group (ctrl). Columns and bars represent mean values ± SEM of triplicate analyses. Statistically significant differences from the control are marked by asterisks (* $p<0.05$; Dunnett's test).

***Stage 54 Exposure Studies with PTU (14 Day Assay)***

*Developmental Stage*

33. Results from developmental stage determination in the different labs during the stage 54 exposure studies with PTU are summarized in Table 19. No significant effects on stage development were observed following 7 days of exposure to PTU in either the JPN or GER study. The US lab did not analyze apical endpoints prior to test termination at day 14. However, after 14 days of exposure, all three labs demonstrated significant retardation of development at 20 mg/L PTU. The US study also was significant at 10 mg/L.

*Hind Limb Length*

34. In addition to stage determinations, measurements of hind limb length were used in the JPN and the GER study to assess effects of PTU on development (Table 20; Table 22). No significant differences in hind limb length were observed following 7 days of exposure to PTU. At day 14, hind limb length was reduced in the 20 mg/L PTU treatment group but the effect was significant only in the JPN study.

*Body Length*

35. Effects of PTU on tadpole growth were examined in the JPN and the GER study by WBL measurements (Table 20; Table 22). Significant reductions in mean WBL were observed in the GER study for 5 and 20 mg/L PTU at day 7. At test termination, no significant effects of PTU treatment on WBL were detectable. The US lab measured SVL at test termination only and determined a significantly increased SVL at 20 mg/L PTU.

*Body Weight*

36. Effects of PTU on tadpole growth were further examined in all labs by body weight measurements at test termination (Table 20; Table 22; Table 24). No significant effects of PTU treatment on tadpole body weight were detectable at day 14.

**Table 19. Distribution of developmental stages of initial stage 54 *X. laevis* tadpoles exposed to PTU for 7 and 14 days. Open boxes highlight the normal development of the controls. Shaded boxes indicate statistical difference compared to controls by the Dunn's method (p<0.05).**

| | PTU Conc. mg/L | Stage at 7 days | | | | Stage at 14 days | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **54** | **55** | **56** | **57** | **55** | **56** | **57** | **58** | **59** | **60** |
| **Japan** | **0.0** | | 26 | 14 | | | 6 | 32 | 2 | | |
| | **2.5** | 3 | 26 | 11 | | | 12 | 25 | 1 | 2 | |
| | **5** | 1 | 33 | 6 | | | 7 | 32 | 1 | | |
| | **10** | 1 | 29 | 10 | | | 5 | 32 | 1 | 2 | |
| | **20** | 1 | 34 | 5 | | | 19 | 20 | 0 | 1 | |
| | | **54** | **55** | **56** | **57** | **55** | **56** | **57** | **58** | **59** | **60** |
| **Germany** | **0.0** | | 1 | 29 | 10 | | 1 | 14 | 14 | 9 | 2 |
| | **2.5** | | | 34 | 6 | 1 | 0 | 21 | 11 | 6 | 1 |
| | **5** | | 3 | 31 | 6 | | 1 | 25 | 6 | 8 | |
| | **10** | | | 34 | 6 | | | 25 | 12 | 2 | 1 |
| | **20** | | 3 | 33 | 4 | | 2 | 24 | 11 | 3 | |
| | | | | | | **55** | **56** | **57** | **58** | **59** | **60** |
| **US** | **0.0** | | | | | | | 11 | 19 | 18 | 2 |
| | **1.25** | | | | | | | 16 | 17 | 17 | |
| | **2.5** | | | | | | | 16 | 13 | 21 | |
| | **5** | | | | | | | 22 | 14 | 14 | |
| | **10** | | | | | | | 28 | 13 | 9 | |
| | **20** | | | | | 11 | 30 | 7 | 0 | 2 | |

**Table 20. Summary of results from whole body length, hind limb length and body weight measurements during the stage 54 study with PTU in the GER lab.**

| | | | Test substance: PTU | | | | Lab: GER |
|---|---|---|---|---|---|---|---|
| | | **control** | **2.5 mg/L** | **5.0 mg/L** | **10 mg/L** | **20 mg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(mg/L)** |
| | | **whole body length (mm)** | | | | | |
| **0** | A | 43.4 ± 1.3 | 42.6 ± 1.6 | 42.7 ± 1.9 | 42.6 ± 1.6 | 43.1 ± 1.4 | |
| | **B** | 43.1 ± 1.6 | 42.7 ± 1.6 | 42.5 ± 1.3 | 42.8 ± 1.6 | 42.7 ± 1.3 | |
| | **Pool** | 43.2 ± 1.4 | 42.6 ± 1.6 | 42.6 ± 1.6 | 42.7 ± 1.5 | 42.9 ± 1.4 | |
| **7** | A | 57.0 ± 2.8 | 56.0 ± 3.4 | 55.6 ± 2.3 | 55.5 ± 2.5 | 55.8 ± 2.5 | |
| | **B** | 57.2 ± 3.3 | 56.3 ± 2.4 | 55.5 ± 2.1 | 55.9 ± 2.0 | 54.8 ± 2.6 | |
| | **Pool** | 57.1 ± 3.0 | 56.1 ± 2.9 | 55.5 ± 2.2 * | 55.7 ± 2.2 | 55.3 ± 2.5 * | 20 # |
| **14** | **A** | 59.7 ± 3.1 | 58.5 ± 3.6 | 58.8 ± 1.7 | 58.4 ± 2.7 | 58.6 ± 2.3 | |
| | **B** | 59.8 ± 2.7 | 59.3 ± 2.1 | 58.0 ± 2.6 | 58.5 ± 2.0 | 58.0 ± 2.7 | |
| | **Pool** | 59.7 ± 2.8 | 58.9 ± 2.9 | 58.4 ± 2.2 | 58.5 ± 2.3 | 58.3 ± 2.5 | ns |
| | | **hind limb length (mm)** | | | | | |
| **7** | A | 5.7 ± 0.9 | 5.5 ± 0.9 | 5.0 ± 1.0 | 5.5 ± 0.9 | 5.2 ± 0.9 | |
| | **B** | 5.6 ± 0.9 | 5.3 ± 0.7 | 5.5 ± 0.8 | 5.3 ± 0.8 | 5.1 ± 0.7 | |
| | **Pool** | 5.6 ± 0.9 | 5.4 ± 0.8 | 5.3 ± 0.9 | 5.4 ± 0.8 | 5.2 ± 0.8 | ns |
| **14** | **A** | 11.0 ± 2.9 | 10.0 ± 2.6 | 9.3 ± 2.4 | 10.1 ± 2.3 | 9.5 ± 2.3 | |
| | **B** | 10.8 ± 2.8 | 9.6 ± 2.5 | 10.2 ± 2.7 | 9.6 ± 1.9 | 9.1 ± 2.0 | |
| | **Pool** | 10.9 ± 2.8 | 9.8 ± 2.5 | 9.8 ± 2.5 | 9.9 ± 2.1 | 9.3 ± 2.1 | ns |
| | | **body weight (mg)** | | | | | |
| **14** | A | 944.1 ± 126.3 | 925.9 ± 192.7 | 928.7 ± 112.2 | 906.5 ± 130.9 | 878.2 ± 135.5 | |
| | **B** | 969.5 ± 156.1 | 922.6 ± 122.9 | 917.6 ± 149.6 | 926.4 ± 95.5 | 879.0 ± 138.5 | |
| | **Pool** | 956.8 ± 139.0 | 924.2 ± 157.5 | 923.1 ± 129.0 | 916.4 ± 112.1 | 878.6 ± 133.6 | ns |

ns no significant effects; asterisks denote significant differences from the control group ($p < 0.05$; Dunnett's Test)

# no concentration response relationship

**Table 21. Summary of results from developmental stage determination during the stage 54 study with PTU in the GER lab.**

| | | | | Test substance: PTU | | | | | | | | Lab: GER |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 2.5 mg/L | | 5.0 mg/L | | 10 mg/L | | 20 mg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | mg/L |
| **7** | **A** | 56 | 56-57 | 56 | 56-57 | 56 | 55-57 | 56 | 56-57 | 56 | 55-57 | |
| | **B** | 56 | 55-57 | 56 | 56-57 | 56 | 56-57 | 56 | 56-57 | 56 | 55-57 | |
| | **Pool** | **56** | **55-57** | **56** | **56-57** | **56** | **55-57** | **56** | **56-57** | **56** | **55-57** | **ns** |
| **14** | **A** | 58 | 56-60 | 58 | 57-59 | 57 | 56-59 | 57 | 57-60 | 57 | 56-59 | |
| | **B** | 58 | 57-59 | 57 | 55-60 | 57 | 57-59 | 57 | 57-59 | 57 | 57-59 | |
| | **Pool** | **58** | **56-60** | **57** | **55-60** | **57** | **56-59** | **57** | **57-60** | **57** | **56-59** | 20 |

ns no significant effects; shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

**Table 22. Summary of results from whole body length, hind limb length and body weight measurements during the stage 54 study with PTU in the JPN lab.**

| | | | Test substance: PTU | | | | Lab: JPN |
|---|---|---|---|---|---|---|---|
| | | **control** | **2.5 mg/L** | **5.0 mg/L** | **10 mg/L** | **20 mg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(mg/L)** |
| | | **whole body length (mm)** | | | | | |
| **0** | A | 39.4 ± 2.3 | 38.7 ± 2.0 | 38.1 ± 2.0 | 39.5 ± 2.3 | 38.6 ± 1.9 | |
| | **B** | 38.7 ± 1.9 | 38.5 ± 4.1 | 38.9 ± 2.4 | 39.3 ± 2.4 | 38.2 ± 2.3 | |
| | **Pool** | **39.0 ± 2.1** | **38.6 ± 3.2** | **38.5 ± 2.2** | **39.4 ± 2.3** | **38.4 ± 2.1** | |
| **7** | A | 51.6 ± 2.5 | 48.9 ± 1.9 | 49.7 ± 2.6 | 51.9 ± 2.1 | 51.9 ± 2.8 | |
| | **B** | 51.9 ± 2.5 | 51.8 ± 3.3 | 51.6 ± 2.7 | 51.9 ± 3.2 | 52.6 ± 2.7 | |
| | **Pool** | **51.8 ± 2.4** | **50.4 ± 3.0** | **50.7 ± 2.7** | **51.9 ± 2.6** | **52.2 ± 2.7** | ns |
| **14** | **A** | 59.7 ± 3.0 | 59.3 ± 2.0 | 59.2 ± 2.6 | 60.3 ± 1.9 | 60.3 ± 3.8 | |
| | **B** | 61.2 ± 2.6 | 61.2 ± 3.1 | 59.0 ± 2.5 | 60.1 ± 2.6 | 61.5 ± 3.3 | |
| | **Pool** | **60.4 ± 2.9** | **60.2 ± 2.7** | **59.1 ± 2.5** | **60.2 ± 2.2** | **60.9 ± 3.5** | ns |
| | | **hind limb length (mm)** | | | | | |
| **7** | A | 4.2 ± 0.7 | 3.7 ± 0.7 | 4.0 ± 0.5 | 4.4 ± 0.6 | 4.1 ± 0.8 | |
| | **B** | 4.3 ± 0.6 | 4.3 ± 0.9 | 3.9 ± 0.6 | 4.0 ± 0.7 | 3.8 ± 0.6 | |
| | **Pool** | **4.2 ± 0.6** | **4.0 ± 0.8** | **3.9 ± 0.5** | **4.2 ± 0.7** | **3.9 ± 0.7** | ns |
| **14** | **A** | 9.1 ± 1.8 | 7.8 ± 1.4 | 8.8 ± 1.6 | 9.4 ± 1.8 | 7.8 ± 1.9 | |
| | **B** | 8.8 ± 1.4 | 9.9 ± 2.2 | 8.4 ± 1.7 | 8.4 ± 2.1 | 7.2 ± 1.6 | |
| | **Pool** | **8.9 ± 1.6** | **8.8 ± 2.1** | **8.6 ± 1.6** | **8.9 ± 1.9** | **7.5 ± 1.7 *** | 20 |
| | | **body weight (mg)** | | | | | |
| **14** | A | 868.0 ± 114.9 | 842.6 ± 73.2 | 848.4 ± 126.3 | 869.6 ± 88.3 | 845.6 ± 130.4 | |
| | **B** | 924.0 ± 110.4 | 919.3 ± 141.7 | 827.6 ± 112.8 | 835.7 ± 111.7 | 877.3 ± 133.0 | |
| | **Pool** | **896.0 ± 113.4** | **881.0 ± 116.4** | **838.0 ± 117.2** | **852.6 ± 99.6** | **861.4 ± 129.3** | ns |

ns no significant effects; asterisks denote significant differences from the control group ($p < 0.05$; Dunnett's Test)

**Table 23. Summary of results from developmental stage determination during the stage 54 study with PTU in the JPN lab.**

| | | | | Test substance: PTU | | | | | | | | Lab: JPN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 2.5 mg/L | | 5.0 mg/L | | 10 mg/L | | 20 mg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | mg/L |
| **7** | **A** | 55 | 55-56 | 55 | 54-56 | 55 | 54-56 | 55 | 55-56 | 55 | 55-56 | |
| | **B** | 55 | 55-56 | 55 | 55-56 | 55 | 55-56 | 55 | 54-56 | 55 | 54-56 | |
| | **Pool** | **55** | **55-56** | **55** | **54-56** | **55** | **54-56** | **55** | **54-56** | **55** | **54-56** | **ns** |
| **14** | **A** | 57 | 56-58 | 57 | 56-57 | 57 | 56-57 | 57 | 57-59 | 57 | 56-59 | |
| | **B** | 57 | 56-57 | 57 | 56-59 | 57 | 56-58 | 57 | 56-59 | 56 | 56-57 | |
| | **Pool** | **57** | **56-58** | **57** | **56-59** | **57** | **56-58** | **57** | **56-59** | **57** | **56-59** | 20 |

ns no significant effects; shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

**Table 24. Summary of results from snout-to-vent length and body weight measurements during the stage 54 study with PTU in the US lab.**

| | | | Test substance: PTU | | | | Lab: US |
|---|---|---|---|---|---|---|---|
| | | **control** | **2.5 mg/L** | **5.0 mg/L** | **10 mg/L** | **20 mg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(mg/L)** |
| | | **snout vent length (mm)** | | | | | |
| **14** | **A** | 19.9 ± 1.4 | 20.2 ± 1.2 | 20.0 ± 1.1 | 19.1 ± 1.2 | 20.9 ± 1.2 | |
| | **B** | 20.0 ± 1.0 | 19.4 ± 1.3 | 20.0 ± 0.9 | 20.5 ± 1.2 | 21.2 ± 1.0 | |
| | **Pool** | 19.9 ± 1.2 | 19.8 ± 1.3 | 20.0 ± 1.0 | 19.8 ± 1.4 | 21.1 ± 1.1 * | **20** |
| | | **body weight (mg)** | | | | | |
| **14** | A | 1052.5 ± 194.6 | 1089.8 ± 136.0 | 1062.2 ± 158.6 | 933.5 ± 176.1 | 1115.6 ± 196.1 | |
| | **B** | 1085.1 ± 123.7 | 996.4 ± 154.6 | 1055.0 ± 112.9 | 1101.6 ± 164.1 | 1162.7 ± 146.0 | |
| | **Pool** | 1068.8 ± 160.6 | 1043.1 ± 150.1 | 1058.6 ± 134.9 | 1017.6 ± 186.7 | 1139.2 ± 171.0 | ns |

ns no significant effects; asterisks denote significant differences from the control group ($p < 0.05$; Dunnett's Test)

**Table 25. Summary of results from developmental stage determination during the stage 54 study with PTU in the US lab.**

| | | | | Test substance: PTU | | | | | | | | Lab: US |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 2.5 mg/L | | 5.0 mg/L | | 10 mg/L | | 20 mg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | mg/L |
| **14** | **A** | 59 | 57-60 | 58 | 57-59 | 58 | 57-59 | 58 | 57-59 | 56 | 55-59 | |
| | **B** | 58 | 57-60 | 58 | 57-59 | 58 | 57-59 | 57 | 57-59 | 56 | 55-59 | |
| | **Pool** | **58** | **57-60** | **58** | **57-59** | **58** | **57-59** | **57** | **57-59** | **56** | **55-59** | 10 |

shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

*Histopathology*

37. The detailed reports of the histopathological analyses conducted by the US lab, GER lab and JPN lab are presented in Annex 2, Annex 3, and Annex 4, respectively. In summary, these analyses revealed exposure-related changes in the thyroid gland, which included distension of thyroid follicles, diffuse enlargement of the thyroid glands, colloid depletion and follicular cell hyperplasia. A low incidence of changes in the thyroid gland (minimal distension of thyroid follicles) was observed for tadpoles exposed to PTU concentrations of 1.25 mg/L (only in the US study) 2.5 and 5.0 mg/L. At 10 and 20 mg/L PTU, follicular distension accompanied by diffuse enlargement of the thyroid glands increased in prevalence and severity in a concentration-dependent manner. An increase in the thickness of the epithelial cell layer (follicular cell hypertrophy) was observed at the two highest PTU concentrations of 10 and 20 mg/L whereas hyperplasia of follicular cells was only observed in the 20 mg/L PTU treatment group. Colloid depletion was also observed at the two highest PTU concentration and collapsed follicles devoid of colloid were detected in some glands from the 20 mg/L PTU treatment. Quantitative methods (morphometric analysis) were successfully applied to confirm the increase in epithelial cell height and the enlargement of the thyroid gland.

*Gene Expression in Brain/Pituitary*

38. Samples of whole brain tissue including the pituitary (brain/pituitary) taken at test termination of the GER study were analysed for changes in gene expression by means of semi-quantitative RT-PCR (Figure 3). Results from RT-PCR revealed increased mRNA expression of TSH β at 20 mg/L PTU. In addition, reduced mRNA expression levels were also observed at 10 and 20 mg/L PTU for other genes including BTEB, b/ZIP, prolactin, and type III monodeiodinase. However, due to the relatively small sample number that has been analysed so far ($n$=3-5 per treatment group), effects did not always reach statistical significance.

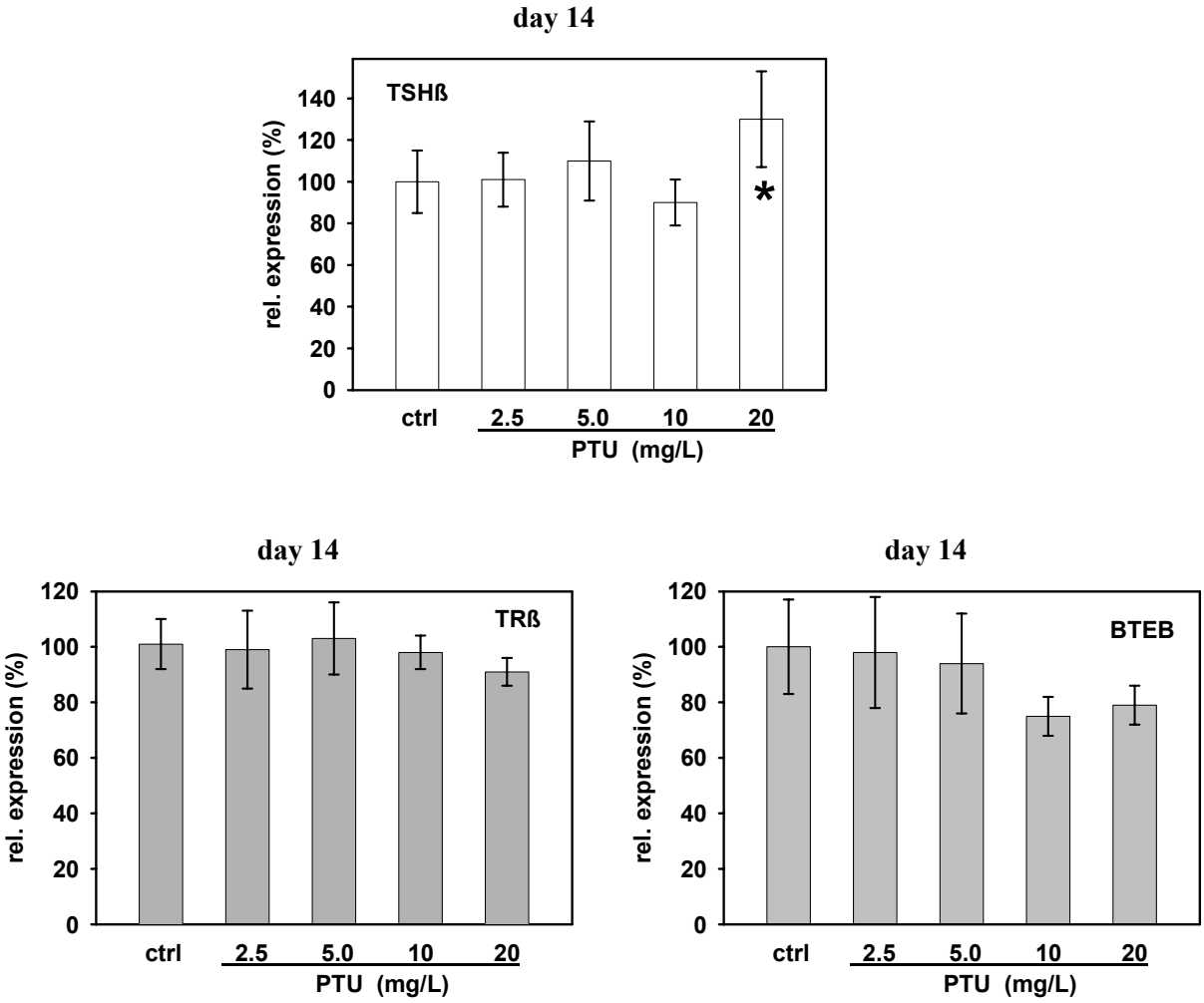


**Figure 3. Effects of PTU on gene expression in tadpole brain.**

RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 54 exposure study with PTU in the GER lab. Semiquantitative RT-PCR analyses of mRNA expression of thyrotropin β -subunit (TSH β), thyroid hormone receptor β (TR β) and basic transcription element binding protein (BTEB) were performed and results from densitometric analysis of scanned agarose gels are shown. Results were expressed relative to the control group (ctrl). Columns and bars represent mean values ± SEM of triplicate analyses. Statistically significant differences from the control are marked by asterisks (* $p<0.05$; Dunnett's test).

**Exposure Studies with Thyroxine (T4)**

39. T4 was used as a model test compound because this thyroid hormone has been shown to stimulate metamorphosis in anuran tadpoles at low concentrations without causing toxic side effects (Opitz *et al.*, in press). Effects of T4 on *Xenopus* metamorphosis were comparatively assessed in a 21 day exposure study initiated with stage 51 tadpoles (stage 51 study) and in a 14 day exposure study initiated with stage 54 tadpoles (stage 54 study). Both exposure studies were performed in parallel with tadpoles from the same spawn. In both exposure scenarios, T4 treatment comprised nominal concentrations of 0.25, 0.5, 1.0 and 2.0 µg/L (the US lab also tested 4.0 µg/L T4). Mortality was negligible in all exposure studies with T4 (data not shown).

***Stage 51 Exposure Studies with T4 (21 Day Assay)***

*Developmental Stage*

40. Results from developmental stage determination in the different labs during the stage 51 exposure study with T4 are summarized in Table 26. Significant acceleration of stage development was observed in the JPN study at 1.0 and 2.0 µg/L T4 after 7, 14, and 21 days of exposure. The GER study showed significant acceleration at 2.0 µg/L T4 after 7 and 14 days, as well as at 1.0 µg/L after 21 days of exposure. The US study was significant at 2.0 µg/L after 21 days of exposure (the US lab did not analyze the stages prior to test termination at day 21). The US lab showed further developmental acceleration in the 4.0 µg/L T4 treatment after 21 days of exposure.

*Hind Limb Length*

41. In addition to stage determinations, measurements of hind limb length were used to assess stimulatory effects of T4 on hind limb morphogenesis in the GER and JPN study (Table 27; Table 29). In the GER study, significant acceleration of hind limb morphogenesis was observed for T4 concentrations of 0.5, 1.0 and 2.0 µg/L at exposure day 7. At later time points of the GER study (day 14 and 21), hind limb length at T4 concentrations of 1.0 and 2.0 µg/L was significantly greater than in untreated controls. Significant acceleration of hind limb morphogenesis was also observed in the JPN study for 1.0 and 2.0 µg/L T4 at days 7, 14 and 21.

*Body Length*

42. Effects of T4 on tadpole growth were examined in the GER and JPN study by WBL measurements (Table 27; Table 29). None of the tested T4 concentrations affected tadpole growth during the initial 7 days of exposure. Treatment with 2.0 µg/L T4 caused a significant reduction in mean WBL at day 14 (JPN study) and at day 21 (GER and JPN study). The US lab measured SVL at test termination only and observed significantly reduced SVL at 1.0, 2.0 and 4.0 µg/L T4 (Table 31).

*Body Weight*

43. Effects of T4 on tadpole growth were further examined in all labs by body weight measurements at test termination (Table 27; Table 29; Table 31). Significant reductions in body weight were observed at 1.0 and 2.0 µg/L T4 in the GER and US studies. In the US study, a strong and significant reduction in body weight was also observed after treatment with 4.0 µg/L T4. Effects of T4 on body weight at day 21 occurred in a concentration-dependent manner in the GER and US labs. The effects detected in the JPN study were more difficult to interpret. In the JPN study, T4 treatment at 0.25 µg/L significantly increased body weight at day 21 whereas 2.0 µg/L T4 caused a significant reduction in body weight at day 21.

**Table 26. Distribution of developmental stages of initial stage 51 *X. laevis* tadpoles exposed to T4 for 7, 14, and 21 days. Open boxes highlight the normal development of the controls. Shaded boxes indicate statistical difference compared to controls by the Dunn's method (p<0.05).**

| | T4 Conc. (µg/L) | Stage at 7 days | | | | Stage at 14 days | | | | | Stage at 21 days | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 53 | 54 | 55 | 56 | 54 | 55 | 56 | 57 | 58 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 |
| Japan | 0.0 | 11 | 29 | | | 3 | 6 | 31 | | | 2 | 3 | 16 | 12 | 4 | 3 | | | | |
| | 0.25 | 2 | 38 | | | | 11 | 21 | 8 | | | 7 | 14 | 13 | 4 | 0 | 2 | | | |
| | 0.5 | | 34 | 6 | | | 1 | 29 | 10 | | | 3 | 15 | 13 | 2 | 5 | 2 | | | |
| | 1.0 | | 14 | 26 | | | | 29 | 11 | | | | 6 | 19 | 9 | 1 | 5 | | | |
| | 2.0 | | 2 | 38 | | | | | | 40 | | | 1 | 8 | 15 | 6 | 10 | | | |
| | | 53 | 54 | 55 | 56 | 54 | 55 | 56 | 57 | 58 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 |
| Germany | 0.0 | | 12 | 28 | | | | 7 | 33 | | | | 3 | 15 | 11 | 4 | 7 | | | |
| | 0.25 | | 6 | 34 | | | | 6 | 34 | | | 1 | 1 | 8 | 13 | 15 | 2 | | | |
| | 0.5 | | 2 | 38 | | | | 3 | 37 | | | | 2 | 11 | 13 | 7 | 7 | | | |
| | 1.0 | | | 40 | | | | | 38 | 2 | | | | 3 | 18 | 7 | 12 | | | |
| | 2.0 | | | 38 | 2 | | | | 7 | 33 | | | | | | 13 | 25 | 1 | 0 | 1 |
| | | | | | | | | | | | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 |
| US | 0.0 | | | | | | | | | | | 6 | 4 | 15 | 8 | 4 | 3 | | | |
| | 0.25 | | | | | | | | | | | 3 | 3 | 22 | 7 | 1 | 4 | | | |
| | 0.5 | | | | | | | | | | | 1 | 2 | 20 | 7 | 5 | 5 | | | |
| | 1.0 | | | | | | | | | | | | | 18 | 10 | 5 | 7 | | | |
| | 2.0 | | | | | | | | | | | | | 3 | 8 | 11 | 16 | 2 | | |
| | 4.0 | | | | | | | | | | | | | | | | | 11 | 15 | 8 |

**Table 27. Summary of results from whole body length, hind limb length and body weight measurements during the stage 51 study with T4 in the GER lab.**

| | | | Test substance: T4 | | | | Lab: GER |
|---|---|---|---|---|---|---|---|
| | | **control** | **0.25 µg/L** | **0.5 µg/L** | **1.0 µg/L** | **2.0 µg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(µg/L)** |
| | | **whole body length (mm)** | | | | | |
| **0** | A | 27.4 ± 0.5 | 27.5 ± 0.5 | 27.5 ± 0.5 | 27.4 ± 0.5 | 27.4 ± 0.5 | |
| | **B** | 27.4 ± 0.5 | 27.5 ± 0.5 | 27.4 ± 0.5 | 27.4 ± 0.5 | 27.5 ± 0.5 | |
| | **Pool** | 27.4 ± 0.5 | 27.5 ± 0.5 | 27.5 ± 0.5 | 27.4 ± 0.5 | 27.4 ± 0.5 | |
| **7** | A | 46.2 ± 1.8 | 46.6 ± 2.5 | 47.3 ± 2.6 | 47.1 ± 3.1 | 45.7 ± 2.5 | |
| | **B** | 46.5 ± 2.8 | 46.3 ± 2.9 | 46.9 ± 3.2 | 46.6 ± 3.0 | 45.2 ± 2.4 | |
| | **Pool** | 46.3 ± 2.3 | 46.5 ± 2.6 | 47.1 ± 2.8 | 46.9 ± 3.0 | 45.5 ± 2.4 | **ns** |
| **14** | **A** | 58.5 ± 2.0 | 59.1 ± 3.1 | 59.0 ± 3.1 | 59.2 ± 4.2 | 57.5 ± 3.3 | |
| | **B** | 58.7 ± 2.7 | 58.7 ± 4.0 | 58.1 ± 3.7 | 57.9 ± 3.1 | 56.4 ± 2.5 | |
| | **Pool** | 58.6 ± 2.3 | 58.9 ± 3.5 | 58.6 ± 3.3 | 58.6 ± 3.6 | 57.0 ± 2.9 | **ns** |
| **21** | **A** | 58.8 ± 3.2 | 58.7 ± 3.6 | 59.1 ± 2.9 | 57.1 ± 4.3 | 53.6 ± 4.7 | |
| | **B** | 59.0 ± 5.2 | 59.9 ± 3.4 | 59.1 ± 2.9 | 57.1 ± 4.2 | 46.8 ± 10.4 | |
| | **Pool** | 58.9 ± 4.2 | 59.3 ± 3.4 | 59.1 ± 2.9 | 57.1 ± 4.1 | 50.2 ± 8.6 * | **2.0** |
| | | **hind limb length (mm)** | | | | | |
| **7** | A | 2.4 ± 0.3 | 2.8 ± 0.3 | 2.9 ± 0.5 | 3.1 ± 0.3 | 3.7 ± 0.3 | |
| | **B** | 2.4 ± 0.3 | 2.6 ± 0.3 | 2.9 ± 0.4 | 3.1 ± 0.3 | 3.8 ± 0.3 | |
| | **Pool** | 2.4 ± 0.3 | 2.7 ± 0.3 | 2.9 ± 0.4 * | 3.1 ± 0.3 * | 3.7 ± 0.3 * | **0.5** |
| **14** | **A** | 7.0 ± 0.7 | 7.3 ± 0.7 | 7.3 ± 0.8 | 7.7 ± 1.0 | 8.7 ± 1.2 | |
| | **B** | 6.8 ± 1.0 | 6.8 ± 1.0 | 7.0 ± 0.9 | 7.4 ± 1.0 | 9.4 ± 0.8 | |
| | **Pool** | 6.9 ± 0.8 | 7.0 ± 0.9 | 7.1 ± 0.8 | 7.6 ± 1.0 * | 9.1 ± 1.0 * | **1.0** |
| **21** | **A** | 17.2 ± 2.2 | 18.0 ± 1.7 | 17.4 ± 2.5 | 18.3 ± 2.6 | 19.2 ± 1.6 | |
| | **B** | 16.2 ± 3.0 | 16.0 ± 3.5 | 16.8 ± 3.0 | 18.2 ± 1.8 | 19.6 ± 1.1 | |
| | **Pool** | 16.7 ± 2.6 | 17.0 ± 2.9 | 17.1 ± 2.7 | 18.2 ± 2.2 | 19.4 ± 1.4 * | **2.0** |
| | | **body weight (mg)** | | | | | |
| **21** | A | 827.4 ± 161.3 | 715.0 ± 103.7 | 727.8 ± 129.6 | 700.2 ± 125.3 | 590.0 ± 93.4 | |
| | **B** | 819.7 ± 184.0 | 808.8 ± 145.7 | 782.6 ± 129.6 | 682.7 ± 136.9 | 509.2 ± 101.6 | |
| | **Pool** | 823.5 ± 168.7 | 761.9 ± 131.9 | 755.9 ± 129.1 | 691.5 ± 128.2* | 549.6 ± 103.4 * | **1.0** |

ns no significant effects; asterisks denote significant differences from the control group ($p < 0.05$; Dunnett's Test)

**Table 28. Summary of results from developmental stage determination during the stage 51 study with T4 in the GER lab.**

| | | | | Test substance: T4 | | | | | | | | Lab: GER |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 0.25 µg/L | | 0.5 µg/L | | 1.0 µg/L | | 2.0 µg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | µg/L |
| **7** | **A** | 55 | 54-55 | 55 | 54-55 | 55 | 54-55 | 55 | 55-55 | 55 | 55-56 | |
| | **B** | 55 | 54-55 | 55 | 54-55 | 55 | 55-55 | 55 | 55-55 | 55 | 55-56 | |
| | **Pool** | **55** | **54-55** | **55** | **54-55** | **55** | **54-55** | **55** | **55-55** | **55** | **55-56** | **2.0** |
| **14** | **A** | 57 | 56-57 | 57 | 56-57 | 57 | 56-57 | 57 | 57-58 | 58 | 57-58 | |
| | **B** | 57 | 56-57 | 57 | 56-57 | 57 | 56-57 | 57 | 57-58 | 58 | 57-58 | |
| | **Pool** | **57** | **56-57** | **57** | **56-57** | **57** | **56-57** | **57** | **57-58** | **58** | **57-58** | **2.0** |
| **21** | **A** | 60 | 58-62 | 61 | 59-62 | 60 | 59-62 | 60 | 59-62 | 61.5 | 61-62 | |
| | **B** | 59 | 58-62 | 60 | 57-61 | 60 | 58-62 | 60 | 60-62 | 62 | 61-65 | |
| | **Pool** | **60** | **58-62** | **60** | **57-62** | **60** | **58-62** | **60** | **59-62** | **62** | **61-65** | **1.0** |

ns no significant effects; shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

**Table 29. Summary of results from whole body length, hind limb length and body weight measurements during the stage 51 study with T4 in the JPN lab.**

| | | | Test substance: T4 | | | | Lab: JPN |
|---|---|---|---|---|---|---|---|
| | | **control** | **0.25 µg/L** | **0.5 µg/L** | **1.0 µg/L** | **2.0 µg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(µg/L)** |
| | | **whole body length (mm)** | | | | | |
| **0** | A | 24.0 ± 1.7 | 23.2 ± 1.4 | 23.7 ± 1.6 | 23.7 ± 1.5 | 23.8 ± 1.2 | |
| | **B** | 23.8 ± 1.3 | 23.8 ± 1.1 | 23.9 ± 1.7 | 23.6 ± 2.0 | 23.3 ± 1.7 | |
| | **Pool** | 23.9 ± 1.4 | 23.5 ± 1.3 | 23.8 ± 1.6 | 23.6 ± 1.7 | 23.6 ± 1.5 | |
| **7** | A | 37.3 ± 2.4 | 42.6 ± 2.6 | 36.8 ± 2.9 | 39.0 ± 2.6 | 38.0 ± 3.1 | |
| | **B** | 41.1 ± 2.6 | 37.1 ± 3.3 | 41.9 ± 3.2 | 39.0 ± 3.1 | 38.8 ± 3.1 | |
| | **Pool** | 39.2 ± 3.1 | 39.8 ± 4.0 | 39.3 ± 3.9 | 39.0 ± 2.8 | 38.4 ± 3.0 | **ns** |
| **14** | **A** | 49.0 ± 3.4 | 54.8 ± 3.2 | 49.3 ± 3.6 | 51.8 ± 3.2 | 49.7 ± 3.6 | |
| | **B** | 53.4 ± 3.0 | 48.7 ± 4.7 | 54.3 ± 3.1 | 52.2 ± 3.4 | 47.7 ± 3.2 | |
| | **Pool** | 51.2 ± 3.8 | 51.8 ± 5.0 | 51.8 ± 4.1 | 52.0 ± 3.2 | 48.7 ± 3.4 * | **2.0** |
| **21** | **A** | 54.2 ± 3.4 | 58.5 ± 3.1 | 53.4 ± 2.8 | 55.8 ± 2.9 | 52.1 ± 5.0 | |
| | **B** | 57.9 ± 2.4 | 57.1 ± 3.6 | 58.2 ± 2.7 | 56.6 ± 2.6 | 50.2 ± 3.6 | |
| | **Pool** | 56.1 ± 3.4 | 57.8 ± 3.4 | 55.8 ± 3.6 | 56.2 ± 2.7 | 51.1 ± 4.4 * | **2.0** |
| | | **hind limb length (mm)** | | | | | |
| **7** | A | 1.9 ± 0.2 | 2.3 ± 0.3 | 2.1 ± 0.2 | 2.6 ± 0.3 | 2.9 ± 0.4 | |
| | **B** | 2.1 ± 0.3 | 2.0 ± 0.2 | 2.3 ± 0.3 | 2.7 ± 0.8 | 2.8 ± 0.4 | |
| | **Pool** | 2.0 ± 0.3 | 2.2 ± 0.3 | 2.2 ± 0.3 | 2.6 ± 0.6 * | 2.9 ± 0.4 * | **1.0** |
| **14** | **A** | 4.4 ± 1.0 | 5.6 ± 0.9 | 4.8 ± 0.9 | 6.0 ± 0.8 | 7.4 ± 1.0 | |
| | **B** | 5.1 ± 0.8 | 4.2 ± 0.6 | 5.5 ± 1.2 | 5.9 ± 0.9 | 7.4 ± 0.8 | |
| | **Pool** | 4.8 ± 1.0 | 4.9 ± 1.0 | 5.2 ± 1.1 | 5.9 ± 0.8 * | 7.4 ± 0.9 * | **1.0** |
| **21** | **A** | 10.3 ± 3.0 | 13.2 ± 2.2 | 11.0 ± 2.7 | 14.0 ± 1.8 | 14.8 ± 1.8 | |
| | **B** | 12.3 ± 1.7 | 10.1 ± 2.4 | 13.0 ± 2.7 | 14.3 ± 2.1 | 14.4 ± 1.5 | |
| | **Pool** | 11.3 ± 2.6 | 11.7 ± 2.7 | 12.0 ± 2.8 | 14.1 ± 1.9 * | 14.6 ± 1.6 * | **1.0** |
| | | **body weight (mg)** | | | | | |
| **21** | A | 709.0 ± 131.1 | 851.2 ± 118.7 | 680.0 ± 107.9 | 703.2 ± 127.4 | 587.8 ± 123.7 | |
| | **B** | 813.5 ± 106.1 | 838.8 ± 121.0 | 824.3 ± 134.1 | 767.1 ± 83.4 | 525.2 ± 79.6 | |
| | **Pool** | 761.2 ± 127.4 | 845.0 ± 117.0* | 752.1 ± 138.8 | 735.2 ± 109.7 | 556.5 ± 106.1 * | **2.0 #** |

ns no significant effects; asterisks denote significant differences from the control group ($p < 0.05$; Dunnett's Test)

# no concentration response relationship

**Table 30. Summary of results from developmental stage determination during the stage 51 study with T4 in the JPN lab.**

| | | | | Test substance: T4 | | | | | | | | Lab: JPN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 0.25 µg/L | | 0.5 µg/L | | 1.0 µg/L | | 2.0 µg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | µg/L |
| **7** | **A** | 54 | 53-54 | 54 | 54-54 | 54 | 54-55 | 55 | 54-55 | 55 | 54-55 | |
| | **B** | 54 | 53-54 | 54 | 53-54 | 54 | 54-55 | 55 | 54-55 | 55 | 54-55 | |
| | **Pool** | **54** | **53-54** | **54** | **53-54** | **54** | **54-55** | **55** | **54-55** | **55** | **54-55** | **1.0** |
| **14** | **A** | 56 | 54-56 | 56 | 56-57 | 56 | 55-57 | 56 | 56-57 | 58 | 58-58 | |
| | **B** | 56 | 55-56 | 55 | 55-56 | 56 | 56-57 | 56 | 56-57 | 58 | 58-58 | |
| | **Pool** | **56** | **54-56** | **56** | **55-57** | **56** | **55-57** | **56** | **56-57** | **58** | **58-58** | **1.0** |
| **21** | **A** | 58 | 56-61 | 59 | 58-62 | 59 | 57-61 | 59 | 58-62 | 60 | 58-62 | |
| | **B** | 59 | 58-61 | 58 | 57-60 | 58 | 57-62 | 59 | 58-62 | 60 | 59-62 | |
| | **Pool** | **58** | **56-61** | **58** | **57-62** | **59** | **57-62** | **59** | **58-62** | **60** | **58-62** | **1.0** |

ns no significant effects; shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

**Table 31. Summary of results from snout-to-vent length and body weight measurements during the stage 51 study with T4 in the US lab.**

| | | | Test substance: T4 | | | | Lab: US |
|---|---|---|---|---|---|---|---|
| | | **control** | **0.25 µg/L** | **0.5 µg/L** | **1.0 µg/L** | **2.0 µg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(µg/L)** |
| | | **snout vent length (mm)** | | | | | |
| **21** | **A** | 19.1 ± 1.7 | 19.2 ± 1.5 | 18.9 ± 1.3 | 18.1 ± 1.9 | 15.9 ± 1.6 | |
| | **B** | 19.8 ± 1.6 | 19.6 ± 1.5 | 18.9 ± 1.6 | 17.5 ± 1.8 | 16.8 ± 2.2 | |
| | **Pool** | 19.5 ± 1.7 | 19.4 ± 1.5 | 18.9 ± 1.4 | 17.8 ± 1.8 * | 16.4 ± 1.9 * | **1.0** |
| | | **body weight (mg)** | | | | | |
| **21** | A | 983.8 ± 191.6 | 1006.2 ± 198.5 | 1005.1 ± 169.2 | 876.1 ± 193.1 | 597.2 ± 146.7 | |
| | **B** | 1099.7 ± 205.1 | 1097.3 ± 192.7 | 990.1 ± 143.0 | 813.5 ± 154.9 | 682.2 ± 272.1 | |
| | **Pool** | 1041.8 ± 201.9 | 1051.7 ± 196.0 | 997.6 ± 152.9 | 844.8 ± 173.5* | 639.7 ± 217.3 * | **1.0** |

asterisks denote significant differences from the control group ($p < 0.05$; Dunnett's Test)

**Table 32. Summary of results from developmental stage determination during the stage 51 study with T4 in the US lab.**

| | | | | Test substance: T4 | | | | | | | | Lab: US |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 0.25 µg/L | | 0.5 µg/L | | 1.0 µg/L | | 2.0 µg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | µg/L |
| **21** | **A** | 59 | 57-62 | 59 | 57-62 | 59 | 57-62 | 59 | 59-62 | 61 | 59-62 | |
| | **B** | 59 | 57-61 | 59 | 57-62 | 59.5 | 58-62 | 60 | 59-62 | 61.5 | 59-63 | |
| | **Pool** | **59** | **57-62** | **59** | **57-62** | **59** | **57-62** | **60** | **59-62** | **61** | **59-63** | **2.0** |

Shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

*Histopathology*

44. The detailed reports of the histopathological analyses conducted by the US lab, GER lab and JPN lab are presented in Annex 2, Annex 3, and Annex 4, respectively. In summary, the light microscopical appearance of thyroid glands from tadpoles exposed to T4 concentrations of 0.25 and 0.5 µg/L did not differ markedly from the control group. The US lab mainly noted changes in colloid content and colloid density at the two highest T4 concentrations including collapsed follicles and reduced or absent colloid in these follicles. In the GER study, an increase in the number of follicles lined by columnar epithelial cells was observed at 1.0 and 2.0 µg/L T4 and minimal increases in the degree of peripheral vacuolation of the colloid were noted in tadpoles exposed to the highest T4 concentration (2.0 µg/L). Quantitative analysis as conducted by the JPN lab revealed no significant change in follicular lumen area and thyroid gland area.

*Gene Expression in Brain/Pituitary*

45. Samples of whole brain tissue including the pituitary (brain/pituitary) taken at test termination of the GER study were analysed for changes in gene expression by means of semi-quantitative RT-PCR (Figure 4). Results from RT-PCR revealed no significant changes in mRNA expression of TSH β and TSH β mRNA following T4 treatment. There was, however, a slight trend towards elevated TSH β mRNA expression at the highest T4 concentration. TR β mRNA expression was significantly increased at 2.0 µg/L T4. Elevated mRNA expression levels were also observed at 1.0 and 2.0 µg/L T4 for other genes including BTEB, b/ZIP, prolactin, and type III monodeiodinase.

***Stage 54 Exposure Studies with T4 (14 Day Assay)***

*Developmental Stage*

46. Results from developmental stage determination in the different labs during the stage 54 exposure study with T4 are summarized in Table 33. Significant acceleration of developmental stage was observed in all studies following 7 and 14 days of expoure to T4 concentrations of 2.0 µg/L. There were no differences between the laboratories. The degree of acceleration was enhanced in the 4 µg/L treatment used by the US lab. The other labs did not test this concentration.

*Hind Limb Length*

47. In addition to stage determinations, measurements of hind limb length were used to assess stimulatory effects of T4 on hind limb morphogenesis in the GER and JPN study (Table 34; Table 36). In the GER study, significant acceleration of hind limb morphogenesis was observed for 0.25 and 2.0 µg/L T4 at exposure day 7 and for 2.0 µg/L T4 at exposure day 14. Effects of T4 on hind limb growth were more variable in the JPN study. Following 7 days of exposure, a strong stimulating effect was observed for 2.0 µg/L T4. A slight increase in mean hind limb length was still detectable at day 14 in this treatment group but the effect was not significant. In contrast, a significant reduction in hind limb length was detected for 0.25 and 1.0 µg/L T4 at day 14.

*Whole Body Length*

48. Effects of T4 on tadpole growth were examined in the GER and JPN study by whole body length measurements (Table 34; Table 36). The highest T4 concentration (2.0 µg/L) reduced WBL in the GER study (day 14) and the JPN study (day 7 and day 14). In the JPN study, WBL was also significantly reduced following treatment with 0.25 and 1.0 µg/L T4 at day 7 and 0.25, 0.5 and 1.0 µg/L T4 at day 14. The US lab measured SVL at test termination only and observed significantly reduced SVL at 2.0 and 4.0 µg/L T4 (Table 38).

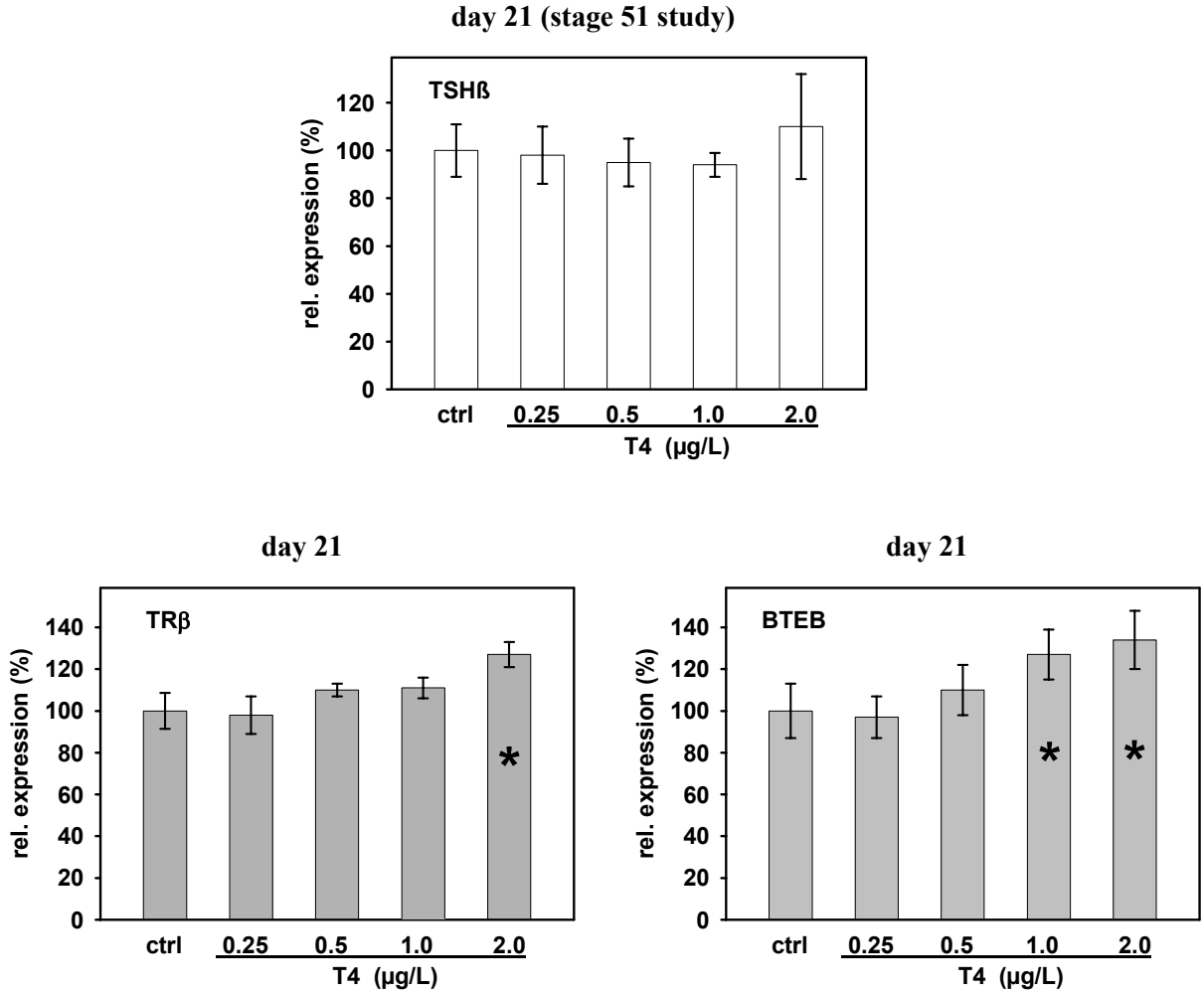


**Figure 4. Effects of T4 on gene expression in tadpole brain.**

RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 51 exposure study with T4 in the GER lab. Semiquantitative RT-PCR analyses of mRNA expression of thyrotropin β -subunit (TSH β), thyroid hormone receptor β (TR β) and basic transcription element binding protein (BTEB) were performed and results from densitometric analysis of scanned agarose gels are shown. Results were expressed relative to the control group (ctrl). Columns and bars represent mean values ± SEM of triplicate analyses. Statistically significant differences from the control are marked by asterisks (* $p<0.05$; Dunnett's test).

**Table 33. Distribution of developmental stages of initial stage 54 *X. laevis* tadpoles exposed to T4 for 7 and 14 days. Open boxes highlight the normal development of the controls. Shaded boxes indicate statistical difference compared to controls by the Dunn's method (p<0.05).**

| | T4 Conc. (µg/L) | Stage at 7 days | | | | Stage at 14 days | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 55 | 56 | 57 | 58 | 56 | 57 | 58 | 59 | 60 | 61 | 62 |
| Japan | 0.0 | 21 | 17 | 2 | | | 25 | 6 | 7 | 2 | | |
| | 0.25 | 25 | 12 | 3 | | 1 | 25 | 6 | 5 | 3 | | |
| | 0.5 | 22 | 17 | 1 | | | 22 | 14 | 4 | | | |
| | 1.0 | 21 | 19 | | | | 27 | 10 | 3 | | | |
| | 2.0 | | 26 | 14 | | | | 12 | 24 | 4 | | |
| | | 55 | 56 | 57 | 58 | 56 | 57 | 58 | 59 | 60 | 61 | 62 |
| Germany | 0.0 | 4 | 35 | 1 | | | 6 | 15 | 17 | 2 | | |
| | 0.25 | 3 | 31 | 6 | | | 3 | 12 | 18 | 4 | 2 | 1 |
| | 0.5 | 2 | 36 | 2 | | | 4 | 18 | 16 | 1 | | |
| | 1.0 | | 35 | 4 | | | 2 | 15 | 17 | 4 | 0 | 1 |
| | 2.0 | | 12 | 27 | 1 | | | 2 | 20 | 12 | 6 | |
| | | | | | | 56 | 57 | 58 | 59 | 60 | 61 | 62 |
| US | 0.0 | | | | | 1 | 14 | 11 | 13 | 1 | | |
| | 0.25 | | | | | | 16 | 6 | 15 | 3 | | |
| | 0.5 | | | | | 1 | 10 | 10 | 15 | 3 | 1 | |
| | 1.0 | | | | | | | 23 | 17 | | | |
| | 2.0 | | | | | | | 1 | 32 | 5 | 0 | 1 |
| | 4.0 | | | | | | | | | 13 | 13 | 14 |

**Table 34. Summary of results from whole body length, hind limb length and body weight measurements during the stage 54 study with T4 in the GER lab.**

| | | | Test substance: T4 | | | | Lab: GER |
|---|---|---|---|---|---|---|---|
| | | **control** | **0.25 µg/L** | **0.5 µg/L** | **1.0 µg/L** | **2.0 µg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(µg/L)** |
| | | **whole body length (mm)** | | | | | |
| **0** | A | 38.8 ± 0.9 | 39.4 ± 1.3 | 38.7 ± 0.7 | 38.8 ± 0.8 | 38.8 ± 1.0 | |
| | **B** | 38.7 ± 0.9 | 39.1 ± 1.1 | 38.9 ± 0.9 | 38.8 ± 1.0 | 38.8 ± 1.1 | |
| | **Pool** | 38.8 ± 0.9 | 39.2 ± 1.2 | 38.8 ± 0.8 | 38.8 ± 0.9 | 38.8 ± 1.0 | |
| **7** | A | 54.2 ± 2.2 | 54.7 ± 2.6 | 54.2 ± 2.2 | 53.4 ± 2.0 | 52.9 ± 1.7 | |
| | **B** | 53.2 ± 2.3 | 54.1 ± 2.1 | 53.9 ± 2.0 | 54.2 ± 2.3 | 52.7 ± 2.7 | |
| | **Pool** | 53.7 ± 2.3 | 54.4 ± 2.3 | 54.1 ± 2.1 | 53.7 ± 2.1 | 52.8 ± 2.2 | ns |
| **14** | **A** | 59.8 ± 2.2 | 59.0 ± 3.2 | 60.1 ± 2.8 | 58.6 ± 2.4 | 57.3 ± 3.0 | |
| | **B** | 59.5 ± 2.6 | 59.6 ± 2.0 | 59.2 ± 3.6 | 59.7 ± 3.6 | 55.7 ± 3.7 | |
| | **Pool** | 59.6 ± 2.4 | 59.3 ± 2.6 | 59.6 ± 3.2 | 59.2 ± 3.0 | 56.5 ± 3.4 * | 2.0 |
| | | **hind limb length (mm)** | | | | | |
| **7** | A | 5.0 ± 0.8 | 5.7 ± 0.7 | 5.3 ± 0.5 | 5.5 ± 0.6 | 6.6 ± 0.5 | |
| | **B** | 5.0 ± 0.7 | 5.3 ± 0.8 | 5.1 ± 0.6 | 5.4 ± 0.7 | 6.5 ± 0.6 | |
| | **Pool** | 5.0 ± 0.7 | 5.5 ± 0.7 * | 5.2 ± 0.6 | 5.5 ± 0.6 | 6.5 ± 0.5 * | 2.0 # |
| **14** | **A** | 12.4 ± 2.5 | 13.0 ± 2.6 | 12.4 ± 2.1 | 13.1 ± 2.2 | 14.6 ± 1.3 | |
| | **B** | 11.9 ± 2.7 | 13.3 ± 2.7 | 11.8 ± 2.1 | 13.2 ± 2.3 | 15.2 ± 1.9 | |
| | **Pool** | 12.2 ± 2.6 | 13.1 ± 2.6 | 12.1 ± 2.1 | 13.1 ± 2.2 | 14.9 ± 1.6 * | 2.0 |
| | | **body weight (mg)** | | | | | |
| **14** | A | 858.2 ± 90.2 | 875.9 ± 109.2 | 881.2 ± 157.5 | 860.9 ± 148.3 | 753.9 ± 132.2 | |
| | **B** | 863.3 ± 143.1 | 865.3 ± 116.3 | 871.0 ± 127.2 | 855.8 ± 165.9 | 716.3 ± 120.9 | |
| | **Pool** | 860.7 ± 116.6 | 870.6 ± 110.1 | 875.9 ± 139.1 | 858.4 ± 153.0 | 735.1 ± 124.9 * | 2.0 |

ns no significant effects; asterisks denote significant differences from the control group ($p < 0.05$; Dunnett's Test)

# no concentration response relationship

**Table 35. Summary of results from developmental stage determination during the stage 54 study with T4 in the GER lab.**

| | | | | Test substance: T4 | | | | | | | | Lab: GER |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 0.25 µg/L | | 0.5 µg/L | | 1.0 µg/L | | 2.0 µg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | µg/L |
| **7** | **A** | 56 | 55-56 | 56 | 55-57 | 56 | 55-57 | 56 | 56-57 | 57 | 56-57 | |
| | **B** | 56 | 55-57 | 56 | 55-57 | 56 | 55-57 | 56 | 56-57 | 57 | 56-58 | |
| | **Pool** | **56** | **55-57** | **56** | **55-57** | **56** | **55-57** | **56** | **56-57** | **57** | **56-58** | **2.0** |
| **14** | **A** | 58 | 57-60 | 59 | 57-60 | 59 | 57-59 | 59 | 57-60 | 59 | 58-61 | |
| | **B** | 58.5 | 57-60 | 59 | 57-62 | 58 | 57-60 | 59 | 57-62 | 59.5 | 59-61 | |
| | **Pool** | **58** | **57-60** | **59** | **57-62** | **58** | **57-60** | **59** | **57-62** | **59** | **58-61** | 2.0 |

ns no significant effects; shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

**Table 36. Summary of results from whole body length, hind limb length and body weight measurements during the stage 54 study with T4 in the JPN lab.**

| | | | Test substance: T4 | | | | Lab: JPN |
|---|---|---|---|---|---|---|---|
| | | **control** | **0.25 µg/L** | **0.5 µg/L** | **1.0 µg/L** | **2.0 µg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(µg/L)** |
| | | **whole body length (mm)** | | | | | |
| **0** | A | 37.8 ± 2.9 | 37.4 ± 2.0 | 37.5 ± 2.1 | 37.2 ± 2.0 | 37.2 ± 2.3 | |
| | **B** | 37.3 ± 2.6 | 37.9 ± 3.0 | 37.2 ± 1.9 | 37.7 ± 2.4 | 37.4 ± 2.3 | |
| | **Pool** | 37.5 ± 2.7 | 37.6 ± 2.5 | 37.3 ± 2.0 | 37.4 ± 2.2 | 37.3 ± 2.3 | |
| **7** | A | 52.6 ± 3.1 | 47.6 ± 3.0 | 50.6 ± 2.8 | 42.7 ± 2.3 | 43.8 ± 3.2 | |
| | **B** | 50.4 ± 3.4 | 48.3 ± 3.7 | 49.1 ± 3.2 | 47.4 ± 2.6 | 48.4 ± 3.1 | |
| | **Pool** | 51.5 ± 3.4 | 47.9 ± 3.3 * | 49.8 ± 3.1 | 45.0 ± 3.4 * | 46.1 ± 3.8 * | 1.0 # |
| **14** | **A** | 60.1 ± 3.1 | 54.0 ± 2.2 | 60.3 ± 2.6 | 48.0 ± 2.2 | 49.3 ± 2.8 | |
| | **B** | 59.1 ± 3.4 | 54.9 ± 3.2 | 54.1 ± 3.0 | 56.5 ± 2.9 | 55.1 ± 3.1 | |
| | **Pool** | 59.6 ± 3.2 | 54.5 ± 2.7 * | 57.2 ± 4.1 * | 52.3 ± 5.0 * | 52.2 ± 4.1 * | 0.25 |
| | | **hind limb length (mm)** | | | | | |
| **7** | A | 4.2 ± 1.1 | 4.0 ± 0.7 | 4.3 ± 0.6 | 3.9 ± 0.4 | 4.9 ± 0.5 | |
| | **B** | 4.2 ± 0.8 | 4.1 ± 1.1 | 4.2 ± 0.6 | 4.0 ± 0.5 | 5.1 ± 0.5 | |
| | **Pool** | 4.2 ± 0.9 | 4.1 ± 0.9 | 4.2 ± 0.6 | 4.0 ± 0.4 | 5.0 ± 0.5 * | 2.0 |
| **14** | **A** | 10.2 ± 2.9 | 8.6 ± 2.2 | 9.9 ± 1.7 | 7.7 ± 1.2 | 10.5 ± 1.6 | |
| | **B** | 10.1 ± 2.4 | 9.2 ± 2.6 | 8.7 ± 1.5 | 9.1 ± 1.4 | 11.2 ± 1.1 | |
| | **Pool** | 10.1 ± 2.6 | 8.9 ± 2.4 * | 9.3 ± 1.7 | 8.4 ± 1.4 * | 10.8 ± 1.4 | # |
| | | **body weight (mg)** | | | | | |
| **14** | A | 913.2 ± 103.6 | 667.9 ± 88.5 | 941.1 ± 107.9 | 470.2 ± 73.9 | 517.4 ± 66.5 | |
| | **B** | 850.9 ± 152.2 | 710.4 ± 105.6 | 680.2 ± 93.0 | 737.3 ± 112.6 | 689.9 ± 116.0 | |
| | **Pool** | 882.0 ± 130.7 | 689.1 ± 97.4 * | 810.6 ± 163.3 | 603.7 ± 162.6* | 603.6 ± 126.2 * | 1.0 # |

ns no significant effects; asterisks denote significant differences from the control group ($p < 0.05$; Dunnett's Test)

# no concentration response relationship

**Table 37. Summary of results from developmental stage determination during the stage 54 study with T4 in the JPN lab.**

| | | | | Test substance: T4 | | | | | | | | Lab: JPN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | control | | 0.25 µg/L | | 0.5 µg/L | | 1.0 µg/L | | 2.0 µg/L | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | µg/L |
| **7** | **A** | 55 | 55-57 | 55 | 55-57 | 55 | 55-56 | 56 | 55-56 | 56 | 56-57 | |
| | **B** | 55.5 | 55-56 | 55 | 55-57 | 55.5 | 55-57 | 55 | 55-56 | 56 | 56-57 | |
| | **Pool** | **55** | **55-57** | **55** | **55-57** | **55** | **55-57** | **55** | **55-56** | **56** | **56-57** | 2.0 |
| **14** | **A** | 57 | 57-60 | 57 | 57-60 | 58 | 57-59 | 57 | 57-59 | 59 | 58-60 | |
| | **B** | 57 | 57-59 | 57 | 57-60 | 57 | 57-59 | 57 | 57-59 | 59 | 59-60 | |
| | **Pool** | **57** | **57-60** | **57** | **57-60** | **57** | **57-59** | **57** | **57-59** | **59** | **58-60** | 2.0 |

ns no significant effects; shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

**Table 38. Summary of results from snout-to-vent length and body weight measurements during the stage 54 study with T4 in the US lab.**

| | | | Test substance: T4 | | | | Lab: US |
|---|---|---|---|---|---|---|---|
| | | **control** | **0.25 µg/L** | **0.5 µg/L** | **1.0 µg/L** | **2.0 µg/L** | **LOEC** |
| day | tank | mean ± SD | mean ± SD | mean ± SD | mean ± SD | mean ± SD | **(µg/L)** |
| | | **snout vent length (mm)** | | | | | |
| **14** | **A** | 19.4 ± 0.9 | 19.5 ± 1.1 | 18.8 ± 0.9 | 19.1 ± 1.0 | 18.1 ± 1.1 | |
| | **B** | 19.1 ± 0.9 | 19.1 ± 1.1 | 19.2 ± 1.1 | 19.1 ± 0.9 | 18.1 ± 1.0 | |
| | **Pool** | **19.2 ± 0.9** | **19.3 ± 1.1** | **19.0 ± 1.0** | **19.1 ± 0.9** | **18.1 ± 1.0 *** | 2.0 |
| | | **body weight (mg)** | | | | | |
| **14** | A | 932.7 ± 164.6 | 998.3 ± 179.6 | 912.4 ± 119.9 | 922.2 ± 143.1 | 825.2 ± 144.9 | |
| | **B** | 952.6 ± 156.0 | 969.8 ± 169.3 | 950.6 ± 146.2 | 926.2 ± 140.1 | 802.6 ± 130.4 | |
| | **Pool** | **942.7 ± 156.6** | **984.1 ± 170.7** | **931.5 ± 131.7** | **924.2 ± 138.0** | **813.6 ± 134.6 *** | 2.0 |

asterisks denote significant differences from the control group (p< 0.05; Dunnett's Test)

**Table 39. Summary of results from developmental stage determination during the stage 54 study with T4 in the US lab.**

| | | | | **Test substance: T4** | | | | | | | | **Lab: US** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **control** | | **0.25 µg/L** | | **0.5 µg/L** | | **1.0 µg/L** | | **2.0 µg/L** | | LOEC |
| day | tank | median | range | median | range | median | range | median | range | median | range | µg/L |
| **14** | **A** | 58 | 56-59 | 58 | 57-60 | 59 | 57-61 | 58 | 58-59 | 59 | 58-62 | |
| | **B** | 58 | 57-60 | 58.5 | 57-60 | 58 | 56-60 | 59 | 58-59 | 59 | 59-60 | |
| | **Pool** | **58** | **56-60** | **58** | **57-60** | **58** | **56-61** | **58** | **58-59** | **59** | **58-62** | 2.0 |

Shaded cells indicate significant differences from the control group ($p < 0.05$; Dunn's Test)

*Body Weight*

49. Effects of T4 on tadpole growth were further examined in all labs by body weight measurements at test termination (Table 34, Table 36, Table 38). Significant reductions in body weight were observed at 2.0 µg/L T4 in the GER and US studies. In the US study, a strong and significant reduction in body weight was also observed after treatment with 4.0 µg/L T4. In the JPN study, T4 treatment at 0.25, 1.0 and 2.0 µg/L significantly reduced body weight at day 21.

*Histopathology*

50. The detailed reports of the histopathological analyses conducted by the US lab, GER lab and JPN lab are presented in Annex 2, Annex 3, and Annex 4, respectively. In summary, the light microscopical appearance of thyroid glands from tadpoles exposed to T4 concentrations of 0.25 and 0.5 µg/L did not differ markedly from the control group. The US lab mainly noted changes in colloid content and colloid density at the two highest T4 concentrations including collapsed follicles and reduced or absent colloid in these follicles. In the GER study, an increase in the number of follicles lined by columnar epithelial cells was observed at 1.0 and 2.0 µg/L T4 and minimal increases in the degree of peripheral vacuolation of the colloid were noted in tadpoles exposed to the highest T4 concentration (2.0 µg/L). Quantitative analysis as conducted by the JPN lab revealed a significant reduction in follicular lumen area and thyroid gland area at 1.0 and 2.0 µg/L T4.

*Gene Expression in Brain/Pituitary*

51. Samples of whole brain tissue including the pituitary (brain/pituitary) taken at test termination of the GER study were analysed for changes in gene expression by means of semi-quantitative RT-PCR (Figure 5). Results from RT-PCR revealed no significant changes in mRNA expression of TSHβ and TSHβ mRNA following T4 treatment. TRβ mRNA expression was significantly increased at 2.0 µg/L T4. Elevated mRNA expression levels were also observed at 2.0 µg/L T4 for other genes including BTEB, b/ZIP, prolactin, and type III monodeiodinase.

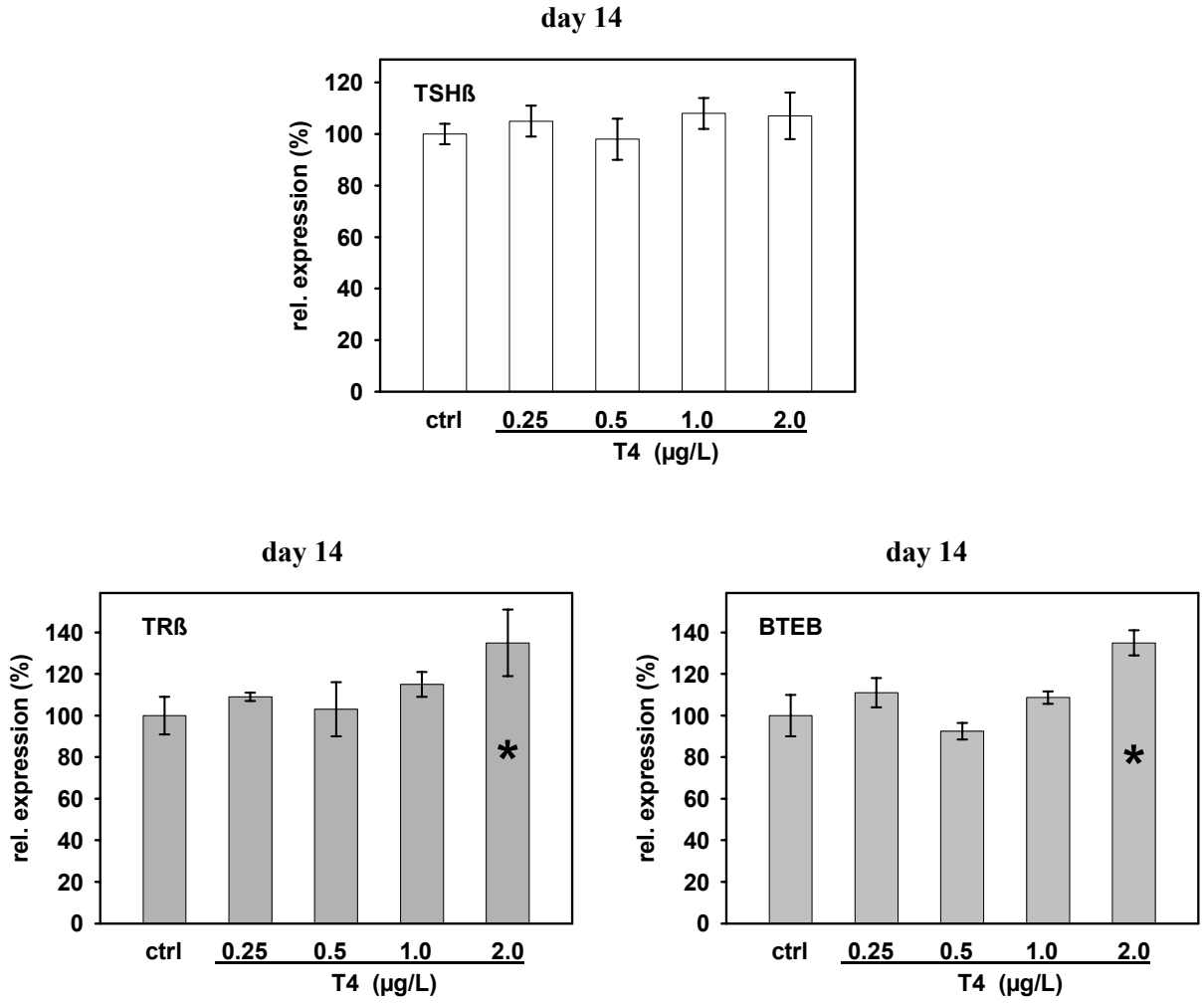


**Figure 5. Effects of T4 on gene expression in tadpole brain.**

RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 54 exposure study with T4 in the GER lab. Semiquantitative RT-PCR analyses of mRNA expression of thyrotropin β-subunit (TSHβ), thyroid hormone receptor β (TRβ) and basic transcription element binding protein (BTEB) were performed and results from densitometric analysis of scanned agarose gels are shown. Results were expressed relative to the control group (ctrl). Columns and bars represent mean values ± SEM of triplicate analyses. Statistically significant differences from the control are marked by asterisks (* $p<0.05$; Dunnett's test).

# DISCUSSION

## Control Organism Performance

52. A total of 12 exposure experiments with *X. laevis* tadpoles were conducted during the first validation phase of the Amphibian Metamorphosis Assay. Despite differences in testing conditions between the three laboratories, the overall performance of control organisms, as judged from mortality, developmental and growth rates was remarkably similar. Slightly lower rates of development were observed in the control group during some of the experiments conducted in the JPN lab. This lab had to cope with some technical problems of the newly established flow-through exposure system, which may explain some of the slight differences in control animal performance. Further potential factors (e.g., reduced food availability) which may be responsible for these deviations have also been identified and it is expected that further optimization and standardization of test conditions during the ongoing validation work will minimize inter-laboratory variation in future studies. Moreover, since the effects caused by a given test chemical are evaluated on a relative basis, slight differences in performance of control organisms do not represent a major concern for the robustness of the assay. In particular, the high degree of similarity of developmental and growth rates between the GER lab (static renewal system) and US lab (flow-through system) indicates that *X. laevis* development is relatively insensitive to different exposure conditions if water quality and food supply are optimized. This fact strongly supports the robustness and practicability of the *Xenopus* metamorphosis model with regard to a broad use in different laboratories.

## Effects Pattern of PTU on Metamorphosis and Thyroid System

53. PTU was used as a test substance during Phase I validation work because of its well investigated anti-thyroidal activity in mammals (Capen, 1996) and *X. laevis* tadpoles (Goos *et al.*, 1968). In mammals, PTU inhibits the synthesis of TH in the thyroid gland by inactivating the critically important enzyme thyroid peroxidase (Cooper *et al.*, 1983) and the same mode of action is expected to occur in amphibian tadpoles. The test concentrations of PTU used in the present study were selected based on experience of the participating laboratories in conducting related work with *X. laevis*. All labs could clearly identify inhibiting effects of PTU on metamorphic development in *X. laevis* tadpoles based on determination of developmental stages of test organisms. Hind limb growth and differentiation represent the most obvious morphological modifications that occur during the developmental phase covered by the exposure protocols used in this study. Morphometric analysis of hind limb growth as performed in two labs confirmed the inhibition of metamorphic development by PTU. The observation of anti-metamorphic effects by PTU is consistent with the fact that normal development of tadpoles is dependent on sufficient supply of TH by the thyroid gland and the anticipated inhibitory activity of PTU on TH synthesis in the thyroid gland. Moreover, results from thyroid gland histology and analysis of TSH- β gene expression in the pituitary provided further confirmation that PTU caused alterations in function of the pituitary-thyroid gland axis. Histopathological analysis of thyroid gland sections revealed distension of thyroid follicles, diffuse enlargement of the thyroid gland, depletion of colloid and follicular cell hyperplasia, while expression of TSH- β mRNA was increased at the highest PTU concentrations in response to the inhibition of TH production by PTU. These findings strongly suggest that PTU acts as an anti-thyroidal agent in *X. laevis* tadpoles.

**Endpoint Sensitivity in PTU Studies**

***Developmental Stage***

54. Determination of the developmental stage of test organisms was the primary endpoint to detect effects of PTU on metamorphic development. All laboratories were able to detect significant retardation of tadpole development at the highest test concentration of PTU (20 mg/L). This was independent of whether exposure of tadpoles was initiated at stage 51 or at stage 54. Results from the GER and JPN studies showed that exposure of stage 51 and 54 organisms with 20 mg/L PTU for 7 days was not sufficient to detect statistically significant deviations from the control group, although trends towards developmental retardation were evident. The increased sensitivity of the assay at later time points can be explained in that development continues in the control organisms and the degree of developmental stage separation from the treated (i.e., inhibited) organisms continues to increase during the exposure period. Different observations were made in the three laboratories with regard to the sensitivity of the two exposure protocols for detection of inhibitory effects on developmental stage (Table 40; Table 41). No sensitivity differences were observed in the JPN lab (LOEC: 20 mg/L PTU for both protocols). Results from the GER lab showed a higher sensitivity of the stage 51 exposure protocol (LOECs: 10 mg/L PTU for stage 51 and 20 mg/L PTU for stage 54) whereas the US lab observed a higher sensitivity when the stage 54 exposure protocol was used (LOECs: 20 mg/L PTU for stage 51 and 10 mg/L PTU for stage 54). Thus, based on developmental stage data, it was not possible to conclude whether a stage 51 exposure for 21 days or a stage 54 exposure for 14 days represents the more sensitive testing approach. Overall, the LOEC values determined from this endpoint differed by not more than a factor of two between the laboratories and between different exposure protocols within a lab.

**Table 40. Comparison of statistical results from the three participating laboratories on the developmental stage of initial stage 51 *X. laevis* larvae exposed to PTU for 7, 14, and 21 days.**

Initial nonparametric analysis by Kruskal-Wallis test was significant ($p \leq 0.05$) for all analyses. Multiple comparisons between all treatment groups and the controls by Dunn's method are summarized in this table.

| Comparison | Laboratory | | | | | | |
|---|---|---|---|---|---|---|---|
| | German | | | Japan | | | US |
| Control vs. | 7 d | 14 d | 21 d | 7 d | 14 d | 21 d | 21 d |
| 1.25 mg/L | -- | -- | -- | -- | -- | -- | ns |
| 2.5 mg/L | ns | ns | ns | ns | ns | ns | ns |
| 5.0 mg/L | ns | ns | ns | ns | * | ns | ns |
| 10 mg/L | ns | ns | * | ns | ns | ns | ns |
| 20 mg/L | ns | ns | * | ns | * | * | * |

* $p \leq 0.05$

- - not tested ns not significant

**Table 41. Comparison of statistical results from the three participating laboratories on the developmental stage of initial stage 54 *X. laevis* larvae exposed to PTU for 7 and 14 days.**

Initial nonparametric analysis by Kruskal-Wallis test was significant ($p \leq 0.05$) for all three laboratories at 14 days, but not significant at 7 days. Multiple comparisons between all treatment groups and the controls by Dunn's method are summarized in this table.

| Comparison | Laboratory | | | | |
|---|---|---|---|---|---|
| | German | | Japan | | US |
| Control vs. | 7 d[1] | 14 d | 7 d[1] | 14 d | 14 d |
| 1.25 mg/L | -- | -- | -- | -- | ns |
| 2.5 mg/L | na | ns | na | ns | ns |
| 5.0 mg/L | na | ns | na | ns | ns |
| 10 mg/L | na | ns | na | ns | * |
| 20 mg/L | na | * | na | * | * |

* $p \leq 0.05$

[1]KW not significant

- - not tested

na not analyzed ns not significant

***Hind Limb Length***

55. Hind limb length was used in the JPN and GER studies as an additional endpoint to detect exposure-related alterations in development. Data from both laboratories indicated a sensitivity difference of this parameter between the two exposure protocols (Table 42; Table 43). In the stage 51 experiments, inhibitory effects of 20 mg/L PTU on hind limb growth were already detectable at day 7 and persisted until test termination. In the GER stage 51 study, 10 and 20 mg/L PTU caused a significant retardation in hind limb growth at day 14 and day 21 whereas stage determination could only detect significant retardation of development at test termination (day 21). Thus, developmental delay caused by PTU (10 and 20 mg/L) was more readily detected by hind limb length compared to stage measurements in the stage 51 exposure studies. In contrast, no significant effects on hind limb growth were detectable in the GER stage 54 exposure study and significant retardation of hind limb growth was only detected at day 14 (20 mg/L PTU) in the JPN stage 54 exposure study. Together, these data indicate that hind limb length measurements in a stage 51 exposure protocol could provide a valuable and sensitive endpoint to more rapidly detect developmental retardation caused by anti-thyroidal substances.

***Thyroid Histopathology***

56. Developmental stage determination and hind limb length measurements represent apical endpoints that can be used to evaluate exposure-related changes in metamorphic development. However, both endpoints suffer from a lack of diagnostic value regarding confirmation of thyroid system-related modes of action. Thyroid gland histopathology was proposed as a core endpoint for the Amphibian Metamorphosis Assay in order to enhance the assay's specificity for thyroid system-related mechanisms of action. Thyroid histopathology represents a classical approach to identify anti-thyroidal activities of chemicals (Capen and Martin, 1989), but little was known about changes of histological endpoints in the thyroid gland of anuran tadpoles in response to anti-thyroidal substances (Goleman *et al.*, 2002b). This lack of information made it difficult to provide a standardized protocol to guide histological evaluation of thyroid glands in the different laboratories and different approaches were used in different laboratories during validation Phase I. The overall aim was to collect information about the responsiveness of a set of different histological parameters that may bear a potential to indicate changes in the functional state of the thyroid gland.

**Table 42. Summary of endpoint sensitivities as detected in stage 51 studies with PTU. LOEC values (mg/L) as determined by the different endpoint measurements are shown.**

| lab | | GER | | | JPN | | | US |
|---|---|---|---|---|---|---|---|---|
| day | | 7 | 14 | 21 | 7 | 14 | 21 | 21 |
| morphology | stage | ns | ns | 10 (-) | ns | 20 (-)* | 20 (-) | 20 (-) |
| | HL | 20 (-)* | 10 (-) | 10 (-) | 20 (-) | 20 (-)* | 20 (-)* | -- |
| histology (qualitative) | | -- | -- | 5.0 | -- | -- | 2.5 | 5.0 |
| histology (quantitative) | cell height | -- | -- | 20 (+) | -- | -- | -- | -- |
| | gland area | -- | -- | 10 (+) | -- | -- | 20 (+) | -- |
| | lumen area | -- | -- | -- | -- | -- | ns | -- |
| molecular biology | TSH β | -- | -- | 20 (+) | -- | -- | -- | -- |
| | TR β | -- | -- | ns | -- | -- | -- | -- |
| | BTEB | -- | -- | 20 (-) | -- | -- | -- | -- |
| body length | WBL | 10 (-)* | 10 (-) | ns | 10 (-) | 20 (-)* | ns | -- |
| | SVL | -- | -- | -- | -- | -- | -- | 20 (+) |
| body weight | | -- | -- | ns | -- | -- | ns | 20 (+) |

ns not significant; -- not determined; (-) significant reduction; (+) significant increase;

* transient effects at lower concentrations or no dose response

BTEB basic transcription element binding protein; HL hind limb length; SVL snout-to-vent length; TR β thyroid hormone receptor β; THS β thyrotropin β -subunit; WBL whole body length;

57. This information was needed to identify evaluation criteria which provide the basis for a more structured histopathological assessment scheme to be used in future studies. The consultation of recognized experts in thyroid histopathology (Environmental Pathology Laboratories, Inc.) by the US lab helped to develop a catalogue of candidate evaluation criteria to be used in a standard grading system for consistent evaluation of changes in thyroid gland histology. Further orientation is given by a recent publication which reports about the successful application of an enhanced grading system to evaluate changes in rat thyroid glands following treatment with anti-thyroidal substances (Hooth *et al.*, 2001). This approach has also been applied to the evaluation of *X. laevis* thyroids following exposure to sodium perchlorate (Tietge *et al.*, 2004).

**Table 43. Summary of endpoint sensitivities as detected in stage 54 studies with PTU. LOEC values (mg/L) as determined by the different endpoint measurements are shown.**

| lab | | GER | | | JPN | | | US |
|---|---|---|---|---|---|---|---|---|
| day | | 7 | 14 | | 7 | 14 | | 14 |
| morphology | stage | ns | 20 (-) | | ns | 20 (-) | | 20 (-) |
| | HL | ns | ns | | ns | 20 (-) | | -- |
| histology (qualitative) | | -- | 5.0 | | -- | 2.5 | | 5.0 |
| histology (quantitative) | cell height | -- | 10 (+) | | -- | -- | | -- |
| | gland area | -- | 10 (+) | | -- | 20 (+) | | -- |
| | lumen area | -- | -- | | -- | 5.0 (+)* | | -- |
| molecular biology | TSH β | -- | 20 (+) | | -- | -- | | -- |
| | TR β | -- | ns | | -- | -- | | -- |
| | BTEB | -- | ns | | -- | -- | | -- |
| body length | WBL | 20 (-)* | ns | | ns | ns | | -- |
| | SVL | -- | -- | | -- | -- | | 20 (+) |
| body weight | | -- | ns | | -- | ns | | ns |

For footnotes see after Table 42

58. 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 microscopical 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. Prevalence and severity of these changes increased in a concentration-dependent manner in all experiments. 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. These changes are consistent with previous studies in which *X. laevis* tadpoles were treated with relatively high concentrations of PTU (Degitz *et al.*, 2004; Goos *et al.*, 1968; unpublished data from the GER lab). The foamy appearance of the colloid in thyroid glands that are inhibited in hormone production by 20 mg/L PTU may reflect the loss of TH as it is being utilized but not replaced. The observed effect pattern at the high PTU concentration suggests an increased stimulation of the thyroid glands by TSH (confirmed through gene expression analysis of TSHβ), consistent with the anticipated disruption of negative feedback signalling between the thyroid gland and the pituitary by PTU. Together, results from the qualitative histological evaluation of thyroid glands were relatively consistent among the laboratories and provided strong confirmation for anti-thyroidal activity of PTU in *X. laevis* tadpoles.

59. The JPN and GER laboratories also performed morphometric analyses in order to quantify changes in selected histological endpoints. By using image analysis techniques, measurements of epithelial cell heights, follicular lumen area and thyroid gland area could confirm the presence of follicular cell hypertrophy, diffuse enlargement of follicles and diffuse thyroid gland enlargement at high PTU concentrations. However, a high variability in measured values for the selected endpoints prevented the detection of statistically significant differences at low PTU concentrations. Determination of changes in thyroid follicular epithelial cell heights is a classical approach used in many mammalian studies to evaluate alterations in the functional state of the thyroid gland (Delverdier *et al.*, 1991; Herrmann *et al.*, 1989). In the present study, it was found that changes in this particular parameter provided a less sensitive endpoint compared to follicle size and thyroid gland size.

60. As noted above, a direct inter-laboratory comparison of thyroid histology results is difficult due to the different evaluation approaches that were used in the three laboratories. The qualitative analyses conducted by the US and GER laboratories were consistent in that both laboratories observed very little changes at the lowest PTU concentrations (1.25 and 2.5 mg/L PTU). Histopathology reports from both laboratories noted that the effect patterns caused by PTU were the same for both exposure protocols whereas prevalence and severity of selected changes were increased at lower PTU concentrations in the stage 51 exposure studies compared to the stage 54 exposure studies. Assuming that the detection of some histopathological changes may be enhanced if the changes are more severe and prevalent, the stage 51 exposure protocol could be considered a slightly more sensitive approach to detect changes in thyroid histology. However, no conclusion can be drawn as to whether the increased severity of histopathological changes is due to the earlier stage at which exposure was initiated in the stage 51 studies or due to the longer exposure period.

61. In summary, thyroid histology greatly enhanced the diagnostic value of the assay for detection of thyroid system-related modes of action. Qualitative analysis of histopathological changes provided a more sensitive approach to detect exposure-related alterations than quantification of changes by image analysis. For future studies, development of a more standardized and structured assessment scheme is necessary to ensure sensitivity and consistency of histopathological evaluations among different laboratories. In addition, quantitative techniques should be further optimized. In particular, the endpoint selection for quantitative measurements requires refinement in order to accommodate the effects pattern observed in the present study.

***Gene Expression Analysis***

62. Analysis of gene expression profiles is currently considered as an optional endpoint for the Amphibian Metamorphosis Assay. A potential advantage of gene expression analysis is that it may provide mechanistic information about modes of action of a test compound in *X. laevis* tadpoles. Selection of appropriate target tissues and marker genes is essential to a successful application of molecular techniques. During validation Phase I, gene expression analysis was only conducted by the GER lab with the general aim to identify and evaluate potential marker genes in brain/pituitary of tadpoles. Results from semi-quantitative RT-PCR experiments provided further confirmation for the anticipated mode of PTU action. The increased expression of TSH β mRNA detected in tadpoles exposed to the highest PTU concentration (20 mg/L) suggests disruption of negative feedback between the thyroid gland and the pituitary by PTU. This effect was consistent with histological findings in that marked hypertrophic changes in the epithelial cell layer were also observed at 20 mg/L PTU. Expression of several TH-responsive genes (e.g. TR β, BTEB) was repressed in brain/pituitary of tadpoles showing retarded development due to PTU exposure. Brain/pituitary tissue samples were collected during all four studies performed in the GER lab. However, gene expression analysis has not yet been completed and so far only a limited number of samples (approximately 50% of the samples taken at test termination of the 4 exposure experiments) have been analyzed. Low sample numbers prevented in many cases the detection of statistically significant effects.

Therefore, results from molecular biological analyses could not contribute data for the sensitivity comparison of the two exposure protocols. The preliminary data indicate, however, that gene expression analysis offers the potential to provide mechanistic information (e.g., increased TSH β expression following PTU treatment) and a further enhancement of molecular analyses by using real-time quantitative PCR is proposed for future studies.

***Body Length and Body Weight***

63. The distinction between thyroid system-related and unrelated mechanisms altering metamorphic development is an important problem that needs to be thoroughly investigated in order to ensure the specificity of any metamorphosis assay for thyroid system-related effects. The use of thyroid system-related endpoints such as thyroid gland histology represents one approach taken during test protocol development. In addition, this issue was further addressed by assessing the utility of growth parameters (body length and body weight) to serve as indicators of non-specific mechanisms affecting tadpole development.

64. Measurements of whole body length (from the tip of the snout to the tip of the tail) at early time points during the exposure phase indicated the presence of weak growth-retarding effects at higher PTU concentrations. The observation of a weak inhibition of tadpole growth due to PTU treatment in the present experiments is consistent with results from previous exposure studies using PTU concentrations from 50 to 100 mg/L (Opitz *et al.*, in press). Results from experiments with other anti-thyroidal compounds (e.g. ethylenethiourea, amitrole, perchlorate) suggest that even complete blockage of TH synthesis and thus, complete inhibition of metamorphosis, does not necessarily inhibit growth of tadpoles (Kloas *et al.*, 2003; Goleman *et al.*, 2002a; Opitz *et al.*, in press). Because PTU has been shown to produce biological effects through various extrathyroidal modes of action (Bandyopadhyay *et al.*, 2002), the reduced growth rates as observed in this study may be the result of PTU mechanisms that are not related to the thyroid system. However, further studies are necessary to investigate the relationship between tadpole growth and disrupted functions of the thyroid system in *X. laevis*.

***Effects Pattern of T4 on Metamorphosis and Thyroid System***

65. T4 was used as a reference substance with agonist properties during Phase I validation work. T4 is the native prohormone synthetized by the thyroid gland of all vertebrates. Acceleration of amphibian metamorphosis by T4 is a well-documented phenomenon. Previous studies conducted by the three participating laboratories showed that low T4 concentrations stimulate metamorphosis without disrupting the normal sequence of morphological changes or causing overt toxicity. During the present study, all laboratories could clearly identify accelerating effects of T4 on metamorphic development in *X. laevis* tadpoles based on determination of developmental stages of test organisms. Hind limb growth and differentiation are the first visible changes in morphology that occur under the influence of the still very low endogenous concentrations of TH during early prometamorphosis. Morphometric analyses of hind limb growth as performed by two labs confirmed the stimulating effect of T4 on metamorphic development. The observation of accelerated development due to T4 treatment is consistent with the fact that rate of metamorphic development is dependent on circulating TH concentrations. Histopathological analysis of thyroid gland sections revealed less marked changes compared to the PTU studies. Exposure-related changes in the thyroid gland included reduced follicular lumen area, collapsed follicles and reduced or absent colloid in these follicles. Furthermore, an increased prevalence of follicles lined by columnar epithelial cell (hypertrophic cells) was observed at the two highest concentrations (1.0 and 2.0 µg/L T4).

**Comparison of Endpoint Sensitivities in T4 Studies**

***Developmental Stage***

66. Determination of the developmental stage of test organisms was the primary endpoint to detect effects of T4 on metamorphic development. All laboratories were able to detect significant acceleration of tadpole development at the highest test concentration of T4 (2.0 µg/L). This was independent of whether exposure of tadpoles was initiated at stage 51 or at stage 54. Results from the GER and JPN studies showed that exposure of stage 51 and 54 organisms with 2.0 µg/L T4 for 7 days was sufficient to detect statistically significant deviations from the control group. The rapid detection of stimulating effects of T4 on development can be explained in that endogenous TH concentrations are very low during premetamorphosis and early prometamorphosis (Leloup and Buscaglia, 1977). Thus, exogenous addition of even low concentrations of T4 may produce a biologically relevant increase in circulating T4 levels during this early developmental phase. In this respect, it should be noted that a previously proposed amphibian testing protocol (tail-resorption-assay) using tadpoles at later developmental stages (stage 60) was limited in its ability to detect stimulating effects at even higher T4 concentrations (US lab unpublished data). The failure of TH treatment to enhance development relative to the untreated controls in the tail resorption assay was most likely due to the high endogenous level of TH already present in tadpoles at late stages and thus, additional exogenous TH could not significantly affect the system. Both testing protocols employed in the current validation study clearly displayed an increased sensitivity to detect agonistic activities compared to the tail resorption assay.

67. Different observations were made in the three laboratories with regard to the sensitivity of the two exposure protocols for detection of stimulatory effects of T4 on development (Table 44; Table 45). No sensitivity differences were observed in the US lab (LOEC: 2.0 µg/L T4 for both protocols). Results from the GER and JPN labs showed a higher sensitivity of the stage 51 exposure protocol (LOECs: 1.0 µg/L T4 for stage 51 and 2.0 µg/L T4 for stage 54). Thus, based on developmental stage data from at least two labs, it appeared that a stage 51 exposure for 21 days represents the more sensitive testing approach than a stage 54 exposure for 14 days. Overall, the LOEC values determined from stage data differed by not more than a factor of two between the laboratories and between different exposure protocols within a lab.

**Table 44. Comparison of statistical results from the three participating laboratories on the developmental stage of initial stage 51 *X. laevis* larvae exposed to T4 for 7, 14, and 21 days.**

Initial nonparametric analysis by Kruskal-Wallis test was significant for all analyses. Multiple comparisons between all treatment groups and the controls by Dunn's method are summarized in this table.

| Comparison | Laboratory | | | | | | |
|---|---|---|---|---|---|---|---|
| | German | | | Japan | | | US |
| Control vs. | 7 d | 14 d | 21 d | 7 d | 14 d | 21 d | 21 d |
| 0.25 μg/L | ns | ns | ns | ns | ns | ns | ns |
| 0.5 μg/L | ns | ns | ns | ns | ns | ns | ns |
| 1.0 μg/L | ns | ns | * | * | * | * | ns |
| 2.0 μg/L | * | * | * | * | * | * | * |
| 4.0 μg/L | -- | -- | -- | -- | -- | -- | * |

* $p \leq 0.05$

- - not tested

ns not significant

**Table 45. Comparison of statistical results from the three participating laboratories on the developmental stage of initial stage 54 *X. laevis* larvae exposed to T4 for 7 and 14 days.**

Initial nonparametric analysis by Kruskal-Wallis test was significant for all analyses. Multiple comparisons between all treatment groups and the controls by Dunn's method are summarized in this table.

| Comparison | Laboratory | | | | |
|---|---|---|---|---|---|
| | German | | Japan | | US |
| Control vs. | 7 d | 14 d | 7 d | 14 d | 14 d |
| 0.25 μg/L | ns | ns | ns | ns | ns |
| 0.5 μg/L | ns | ns | ns | ns | ns |
| 1.0 μg/L | ns | ns | ns | ns | ns |
| 2.0 μg/L | * | * | * | * | * |
| 4.0 μg/L | -- | -- | -- | -- | * |

* $p \leq 0.05$

- - not tested

ns not significant

***Hind Limb Length***

68. Hind limb length was used in the JPN and GER studies 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 (Table 46; Table 47). In the stage 51 experiments, hind limb length measurements at day 7 provided the most sensitive endpoint to detect the stimulatory effects of T4 (LOEC: 0.5 µg/L T4 in the GER lab and 1.0 µg/L in the JPN lab). In the stage 54 experiments, hind limb length measurements were as sensitive as developmental stage determinations. During spontaneous metamorphosis, the hind limbs are the first tissues that undergo obvious morphological changes (growth and differentiation) under the influence of relatively low endogenous TH concentrations. It should be noted that the sensitivity of hind limb length measurements to detect T4 effects was diminished at later time points in the GER stage 51 exposure study. Together, these data indicate that hind limb length measurements at day 7 in a stage 51 exposure protocol could provide a very sensitive endpoint to detect agonist activities. One factor that may contribute the different sensitivities of the endpoints hind limb length and developmental stage may be that parametric statistical methods were used for analysis of hind limb length data whereas non-parametric approaches were used for evaluation of developmental stage data. Future studies should further address a comparative evaluation of the sensitivities of these two apical endpoints.

***Thyroid Histopathology***

69. Results from histopathological analyses of thyroid glands of T4-treated tadpoles were less consistent between the laboratories and are more difficult to interpret compared to the effect pattern seen in the PTU studies. The US lab mainly noted changes in colloid content and colloid density. At the two highest T4 concentrations, prevalent changes included collapsed follicles and reduced or absent colloid in these follicles. Similar effects on follicular colloid content were only rarely observed at the highest T4 concentration (2.0 µg/L) in the GER lab. Quantitative analysis as conducted by the JPN lab 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 stage 54 experiment but not in the stage 51 experiment. The most obvious change determined in the GER experiments was an increase in the number of follicles lined by columnar epithelial cells at the two higher T4 concentrations.

**Table 46. Summary of endpoint sensitivities as detected in stage 51 studies with T4**

| lab | | GER | | | JPN | | | US |
|---|---|---|---|---|---|---|---|---|
| day | | 7 | 14 | 21 | 7 | 14 | 21 | 14 |
| morphology | stage | 2.0 (+) | 2.0 (+) | 1.0 (+) | 1.0 (+) | 1.0 (+) | 1.0 (+) | 2.0 (+) |
| | HL | 0.5 (+) | 1.0 (+) | 2.0 (+) | 1.0 (+) | 1.0 (+) | 1.0 (+) | -- |
| histology (qualitative) | | -- | -- | 2.0 | -- | -- | ns | 1.0 |
| histology (quantitative) | cell height | -- | -- | 2.0 (+) | -- | -- | -- | -- |
| | gland area | -- | -- | -- | -- | -- | ns | -- |
| | lumen area | -- | -- | -- | -- | -- | ns | -- |
| molecular biology | TSH β | -- | -- | ns | -- | -- | -- | -- |
| | TR β | -- | -- | 2.0 (+) | -- | -- | -- | -- |
| | BTEB | -- | -- | 1.0 (+) | -- | -- | -- | -- |
| body length | WBL | ns | ns | 2.0 (-) | ns | 2.0 (-) | 2.0 (-) | -- |
| | SVL | -- | -- | -- | -- | -- | -- | 1.0 (-) |
| body weight | | -- | -- | 1.0 (-) | -- | -- | 2.0 (-) | 1.0 (-) |

ns not significant; -- not determined; (-) significant reduction; (+) significant increase;

* transient effects at lower concentrations or no dose response

BTEB basic transcription element binding protein; HL hind limb length; SVL snout-to-vent length; TR β thyroid hormone receptor β; THS β thyrotropin β -subunit; WBL whole body length;

**Table 47. Summary of endpoint sensitivities as detected in stage 54 studies with T4**

| lab | | GER | | | JPN | | | US |
|---|---|---|---|---|---|---|---|---|
| day | | 7 | 14 | | 7 | 14 | | 14 |
| morphology | stage | 2.0 (+) | 2.0 (+) | | 2.0 (+) | 2.0 (+) | | 2.0 (+) |
| | HL | 2.0 (+)* | 2.0 (+) | | 2.0 (+) | * | | -- |
| histology (qualitative) | | -- | 2.0 | | -- | 1.0 | | 1.0 |
| histology (quantitative) | cell height | -- | 1.0 (+) | | -- | -- | | -- |
| | gland area | -- | -- | | -- | 1.0 (-) | | -- |
| | lumen area | -- | -- | | -- | 1.0 (-) | | -- |
| molecular biology | TSH β | -- | ns | | -- | -- | | -- |
| | TR β | -- | 2.0 (+) | | -- | -- | | -- |
| | BTEB | -- | 2.0 (+) | | -- | -- | | -- |
| body length | WBL | ns | 2.0 (-) | | 1.0 (-)* | 0.25 (-) | | -- |
| | SVL | -- | -- | | -- | -- | | 2.0 (-) |
| body weight | | -- | 2.0 (-) | | -- | 1.0 (-) | | 2.0 (-) |

For footnotes see after Table 46

70. Depletion of colloid stores and increases in epithelial cell height are known to occur at climax stages during normal development (Regard, 1978) when TSH synthesis and release by the pituitary (Buckbinder and Brown, 1993; Manzon and Denver, 2004) and T4 synthesis and secretion by the thyroid gland (Leloup and Buscaglia, 1977; Regard, 1978) reach maximum levels. Early studies by Saxen *et al.* (1957) reported marked colloid depletion and also collapse of follicles in *X. laevis* tadpoles at climax

stages. Similarly, results from an unpublished study by the GER lab showed that the degree of colloid resorption and the thickness of the epithelial cell layer were maximal at stages 60 to 62 in *X. laevis* tadpoles during spontaneous metamorphosis. Therefore, the challenge in interpreting the histological findings of the present study is to distinguish whether the selected changes occurred in response to T4-induced alterations of the functional state of the hypothalamus-pituitary-thyroid gland axis or merely reflect the advanced stage of the tadpoles in the corresponding T4 treatment groups.

***Gene Expression Analysis***

71. Analysis of mRNA expression was performed with tissue samples taken at test termination of the T4 studies. Results from semi-quantitative RT-PCR experiments did not indicate differences in pituitary expression of TSH β mRNA between control and T4-treated tadpoles. TSH β mRNA expression was slightly elevated at the highest T4 concentration. Thus, by means of semi-quantitative RT-PCR analysis, there was no indication for negative feedback effects of the tested T4 concentrations on TSH β mRNA expression. However, tadpoles were randomly selected at test termination for tissue dissection and subsequent RNA extraction and were therefore not necessarily at the same developmental stage. Because TSH β mRNA, like many other genes, shows a developmental expression profile with peak expression at climax stages, the slightly elevated expression level observed at 2.0 µg/L T4 may reflect the advanced stage of the tadpoles in the corresponding T4 treatment group. Further, the absence of a down-regulation of pituitary TSH β mRNA expression by T4 (when analyzed at test termination) does not preclude the possibility that negative feedback by T4 may occur at earlier exposure time points.

72. Expression of several TH-responsive genes in brain (e.g. TR β, BTEB) was elevated in tadpoles showing accelerated development due to T4 exposure. However, the observed differences in TR β and BTEB mRNA expression between control and T4-exposed tadpoles correlate closely with the stage-dependent changes in expression that occur during spontaneous development.

***Body Length and Body Weight***

73. Measurements of whole body length (from the tip of the snout to the tip of the tail) at early time points during T4 exposure did not indicate effects of T4 on initial tadpole growth (Table 46; Table 47). At termination, however, reduced mean values of whole body length, snout-to-vent length and body weight were consistently determined at higher T4 concentrations in the GER and US labs. In the JPN studies with T4, heterogeneous effects on tadpole growth were observed for the different T4 concentrations. Overall, the results indicate that the tested T4 concentrations did not affect initial tadpole growth but that accelerated development occurring at high T4 concentrations (which was accompanied by resorption of gill and tail tissue) led to decreased mean values of whole body length, snout-to-vent length and body weight at test termination. Similar to the PTU studies, evaluation of growth-related endpoints at early time points during the exposure period may provide relevant information to identify possible toxic effects.

**Critical Review of Results from Validation Phase 1 and Proposal for Phase 2 Validation Activities**

74. The data presented in this report have been initially reviewed and discussed by representatives from all three participating laboratories at a status meeting held in Hiroshima, Japan, in March 2004 (see Annex 6). As an outcome of this meeting, a draft report presenting a compilation and initial discussion of Phase 1 study results as well as a protocol proposal for validation Phase 2 were prepared and submitted to OECD. These documents provided the basis for further discussion of Phase 1 study results at the second OECD meeting of the *ad hoc* Expert Group on Amphibian Testing in Paris, France, in June 2004 (OECD 2004). In this section, the main outcome of the two meetings is summarized in order to provide the rationale for the test protocol proposal that will be developed for validation Phase 2.

75. The analysis of control data from the studies performed during Phase 1 showed that mortality, development and growth of *X. laevis* tadpoles was very similar across the three participating laboratories despite differences in several aspects of test conditions and test performance. Further, all labs could clearly detect inhibitory effects of the anti-thyroidal model compound PTU and stimulatory effects of the agonist model compound T4 on metamorphic development. Analysis of histological and molecular biological endpoints provided diagnostic power for thyroid system-related mechanisms of action. Together, these results suggest that *X. laevis* represents a suitable test organism and strongly confirm the practicability, robustness, and specificity of the Amphibian Metamorphosis Assay for detection of thyroid system-disrupting activities.

76. For further standardization and optimization of the assay in validation Phase 2, the following changes in general testing conditions were agreed upon:

- only flow-through systems should be used for exposure experiments,
- Sera micron, a widely available commercial tadpole and fish fry food, should be used as a standard food type.

77. These protocol modifications are expected to further minimize the slight variations in control organism performance.

78. A major aim of validation Phase 1 work was to compare two previously proposed exposure protocols with regard to their utility and sensitivity to detect the agonistic and antagonistic effects of T4 and PTU on TH-dependent metamorphic development. The two protocols differed in the developmental stage at which exposure of tadpoles was initiated and in their duration. In one protocol, premetamorphic stage 51 tadpoles were exposed for 21 days to the test compounds whereas in the alternative protocol, exposure was initiated at stage 54 and lasted for only 14 days. The selection of stages at test initiation and the corresponding exposure periods was based on previous experience of the participating laboratories in order to allow control tadpoles to develop throughout prometamorphosis to stages 58/59 during the exposure period of either test protocol.

79. Results from validation Phase 1 confirmed the ability of both exposure protocols to detect the effects of T4 and PTU on metamorphosis and thyroid system function. The advantage of the stage 54 test protocol was the shorter test duration (14 days). However, the stage 51 test protocol (21 days) showed a higher sensitivity for detection of the agonistic activities of T4 while the sensitivity to detect the anti-thyroidal activity of PTU was similar for both test protocols.

80. The most sensitive endpoint to detect anti-thyroidal effects of PTU was thyroid histology. Histopathological effect patterns of PTU were the same in both test protocols, though the prevalence and severity of histopathological changes was increased in the stage 51 protocol. However, it is not known whether the latter finding was due to the early stage at exposure initiation or the longer test duration used in the stage 51 protocol. Another observation favouring the stage 51 protocol was that morphological effects, in particular a delay in hind limb growth (observed at day 7), were more rapidly detectable in the stage 51 protocol.

81. The most sensitive endpoint to detect agonistic effects of T4 was hind limb length when measured at early time points (day 7) during the exposure phase of the stage 51 protocol. From an endocrinological perspective, the observation of accelerating effects on metamorphic development, particularly if they occur during late premetamorphic and early prometamorphic development, may be considered a diagnostic finding for TH agonist activity by itself. This interpretation is based on results from a large number of studies showing that only THs are able to induce precocious development in

premetamorphic tadpoles (Kikuyama *et al.*, 1993; Shi, 1999). During later development, other endocrine factors such as corticosteroid hormones may modulate (e.g., enhance) the action of endogenous TH causing accelerated development. Therefore, consideration of premetamorphic stage 51 tadpoles for initiation of exposure may also increase the diagnostic value of the assay in cases where acceleration effects on development are observed.

82. Although several observations from Phase 1 and general endocrinological considerations may support a stage 51 test protocol lasting for 21 days, the longer test duration represents a disadvantage, given that the assay is currently envisaged as a screening assay. Therefore, possible modifications of the test protocol to combine the sensitivity advantage of the stage 51 test protocol with the need for a shortened exposure period were discussed. Unpublished data from *X. laevis* exposure studies with anti-thyroidal compounds in the US lab show that there was no sensitivity difference between two 14 day exposure protocols in which exposure was initiated with stage 51 and stage 54 tadpoles. These data suggest that a 14 day test protocol using stage 51 *X. laevis* tadpoles may provide a means to retain the sensitivity of the assay for detection of antagonistic and agonistic effects while reducing the exposure duration to 14 days.

83. Results from validation Phase 1 experiments showed that stage 51 control tadpoles develop within 14 days to prometamorphic stage 56. Accordingly, a 14 day test initiated with stage 51 tadpoles will be terminated at an earlier stage compared to the tests performed during Phase 1 studies. In order to reliably assess the sensitivity of the shortened 14 days-stage 51 test protocol with the original 21 days-stage 51 test protocol, the following test protocol modifications were agreed upon for validation Phase 2 studies.

84. Exposure will be initiated with stage 51 tadpoles in order to ensure the high sensitivity of the assay for detection of agonistic effects. Based on results from the first validation phase, determination of hind limb length and whole body length at exposure day 7 are mandatory endpoints because day 7 was the most sensitive time point to identify accelerating effects on metamorphosis and possible growth-retarding effects of the test substances.

85. At exposure day 14, developmental stage, hind limb length and whole body length are determined because day 14 would be an alternative time point of test termination. In addition, 10 tadpoles are randomly selected at day 14 within each treatment group (5 tadpoles per replicate tank) for a subsampling to collect tissue for thyroid gland histopathology. For all specimens in this subsample of test organism, body weight is determined.

86. The test is terminated after 21 days. At test termination, developmental stage, hind limb length, whole body length and body weight are determined for all remaining tadpoles. 10 tadpoles per treatment group are used for thyroid histopathology at day 21. The data collected from this exercise using the modified testing protocol will contribute to a decision after validation Phase 2 whether the 14 day or 21 day protocol offer any advantages.

87. Another rationale for the complex sampling scheme proposed for validation Phase 2 is to address for the first time the utility and sensitivity of the metamorphosis assay to detect effects of a well-known monodeiodinase inhibitor, iopanoic acid (IOP). The inclusion of IOP as a test compound during validation Phase 2 is based on the following consideration. From an endocrinological perspective, the main effect of T4 and PTU was to increase (T4) or decrease (PTU) the amount of circulating TH in the test organisms, conditions which can be described as general hyperthyroidism and hypothyroidism. Results from validation Phase 1 work showed that the net result was an acceleration or retardation of metamorphic development, which still proceeded in a coordinated manner. IOP inhibits all monodeiodinases thereby enhancing TH action in tissues which are normally protected from TH by expression of type III

monodeiodinase (e.g., tail) while blocking TH action in tissues which require efficient conversion of T4 to T3 (e.g., hind limbs). The conditions caused by IOP can be described a combination of local hyperthyroidism and local hypothyroidism. Results from preliminary studies performed in the US and GER lab indicate that IOP disrupts the normal sequence of morphological changes in developing tadpoles. It was considered important to thoroughly analyze the expected alternative effect patterns caused by IOP on metamorphic development and thyroid system function during the next validation phase to include this information in the subsequent decision process for a final test protocol.

88. Hind limb length and thyroid histopathology were the most sensitive endpoints to detect the agonist activity of T4 and the anti-thyroidal activity of PTU, respectively. Therefore, standardization of these endpoint measurements was considered important to enhance sensitivity and reproducibility of the assay. For hind limb length determination, a digital length measuring system was proposed for use in validation Phase 2 based on successful application of this system in the JPN lab during validation Phase 1. While different approaches were used for histopathological analysis of thyroid gland sections during Phase 1, the information gathered provided the basis for the development of a more structured and standardized histopathological assessment scheme to be used in Phase 2. Recent publications describing the methodology of an enhanced grading system to sensitively diagnose histological changes in rat thyroid glands following treatment with anti-thyroidal substances (Hooth *et al.*, 2001) and in *X. laevis* larvae exposed to sodium perchlorate (Tietge *et al.*, 2004) will be used as the basis for an assessment scheme for these studies.

89. The test substances proposed for use in validation Phase 2 include T4, IOP and sodium perchlorate (PER). T4 should be used as a reference compound for agonist activity on the thyroid system and was already assessed for effects on metamorphosis and thyroid system function during Phase 1. PER should be used as a reference compound for anti-thyroidal activity on the thyroid system. Similar to PTU, PER inhibits the synthesis of TH in the thyroid gland. However, the actual mode of PER action is different to PTU. PER is a well-known inhibitor of iodine uptake and has been shown to retard metamorphosis in *X. laevis* tadpoles (Goleman *et al.*, 2002a, 2002b; Tietge *et al*, 2004; unpublished data from the GER lab). The proposal to use PER instead of PTU in Phase 2 is based on unpublished observations from the US lab indicating that the histopathological effects pattern of PER differs from PTU and hence, validation Phase 2 should be used to investigate the possible diagnostic value of thyroid gland histopathology to differentiate between different inhibitory modes of action on TH synthesis. The rationale for including IOP as a test substance during Phase 2 has already been outlined in this section. IOP can be regarded as a reference compound for modulation of TH action in peripheral tissues and thus, it is expected that effects pattern caused by IOP may be more complex than those observed following T4 and PTU treatment during Phase 1. The specific value of using IOP as a test substance during Phase 2 is to assess the utility and sensitivity of the so far established endpoints of the assay for detection of the more complex effects that may be caused by compounds which target peripheral TH action.

## REFERENCES

Bandyopadhyay, U., Biswas, K. and Banerjee, R.K. (2002): Extrathyroidal actions of antithyroid thionamides. Toxicol. Lett. 128, 117-127

Buckbinder, L. and Brown, D.D. (1993): Expression of the *Xenopus laevis* prolactin and thyrotropin genes during metamorphosis. Proc. Natl. Acad. Sci. USA 90, 3820-3824

Cai, L. and Brown, D.D. (2004): Expression of type II iodothyronine deiodinase marks the time that a tissue responds to thyroid hormone-induced metamorphosis in *Xenopus laevis*. Dev. Biol. 266, 87-95

Capen, C.C. (1996): Toxic responses of the endocrine system. In: Casarett and Doull's Toxicology: The Basic Science of Poisons. Klaassen, C.D., Amdur, M.O. and Doull, J. (eds.), Pergamon Press, New York, US, 617-640

Capen, C.C. and Martin, S.L. (1989): The effects of xenobiotics on the structure and function of thyroid follicular and C-cells. Toxicol. Pathol. 17, 266-293

Delverdier, M., Cabanie, P., Roome, N., Enjalbert, F. and van Haverbeke, G. (1991): Quantitative histology of the rat thyroid. Influence of histologic techniques on the morphometric data. Anal. Quant. Cytol. Histol. 13, 110-114

Colborn, T. (2002): Clues from wildlife to create an assay for thyroid system disruption. Environ. Health Perspect. 110 (Suppl. 3), 363-367

Degitz, S.J., Holcombe, G.W., Flynn, K.M., Kosian, P.A., Korte, J.J. and Tietge, J.E. (2004): Thyroid axis inhibition in *Xenopus laevis*: Development of an amphibian-based screening assay. Toxicol. Sci. (in press)

Goleman, W.L., Urquidi, L.J., Anderson, T.A., Smith, E.E., Kendall, R.J. and Carr, J.A. (2002a): Environmentally relevant concentrations of ammonium perchlorate inhibit development and metamorphosis in *Xenopus laevis*. Environ. Toxicol. Chem. 21, 424-430

Goleman, W.L., Carr, J.A. and Anderson, T.A. (2002b): Environmentally relevant concentrations of ammonium perchlorate inhibit thyroid function and alter sex ratios in developing *Xenopus laevis.* Environ. Toxicol. Chem. 21, 590-597

Goos, H.J., Zwanenbeek, H.C. and van Oordt, P.G. (1968): Hypothalamic neurosecretion and metamorphosis in *Xenopus laevis*. II. The effect of thyroxine following treatment with propylthiouracil. Arch. Anat. Histol. Embryol. 51, 267-274

Herrmann, F., Hambsch, K., Wolf, T., Rother, P. and Muller, P. (1989): Comparison of three histometric methods for the comprehension of stimulating effects on the rat thyroid gland. Gegenbaurs. Morphol. Jahrb. 135, 679-688

Hooth, M.J., Deangelo, A.B., George, M.H., Gaillard, E.T., Travlos, G.S., Boorman, G.A. and Wolf, D.C. (2001): Subchronic sodium chlorate exposure in drinking water results in a concentration-dependent increase in rat thyroid follicular cell hyperplasia. Toxicol. Pathol. 29(2), 250-259

Kikuyama, S., Kawamura, K., Tanaka, S. and Yamamoto, K. (1993): Aspects of amphibian metamorphosis. Hormonal control. Int. Rev. Cytol. 145, 105-148

Kloas, W., Opitz, R. and Lutz, I. 2003. Ringtest: effects of pesticides and other chemicals on thyroid system in the amphibian *Xenopus laevis*. Federal Environmental Agency Research Report 200 67 409, Berlin, Germany.

Leloup, J., and Buscaglia, M. (1977): La triiodothyronine, hormone de la metamorphose des amphibiens. C.R. Acad. Sci. Hebd. Seances Acad. Sci. D. 284, 2261-2263

Manzon, R.G. and Denver, R.J. (2004): Regulation of pituitary thyrotropin gene expression during *Xenopus* metamorphosis: negative feedback is functional throughout metamorphosis. J. Endocrinol. 182, 273-285

Nieuwkoop, P.D. and Faber, J. (1994): Normal Table of *Xenopus laevis* (Daudin), Garland, New York, US

OECD, 2003. Final Report of the 1st Meeting of the OECD *Ad hoc* Expert Group on Amphibian Testing, Duluth, 28-29 June 2003. (ENV/JM/TG/EDTA (2004).

OECD, 2004. Final Report of the 2nd Meeting of the OECD *Ad hoc* Expert Group on Amphibian Testing, Paris, 28-29 June 2004. (ENV/JM/TG/EDTA (2004).

Opitz, R., Braunbeck, T., Bögi, C., Pickford, D.B., Nentwig, G., Oehlmann, J., Tooi, O., Lutz, I. and Kloas, W. (2004): Description and initial evaluation of a *Xenopus* Metamorphosis Assay (XEMA) for detection of thyroid system-disrupting activities of environmental compounds. Environ. Toxicol. Chem. (in press).

Regard, E. (1978): Cytophysiology of the amphibian thyroid gland through larval development and metamorphosis. Int. Rev. Cytol. 52, 81-118

Saxen, L., Saxen, E., Toivonen, S. and Salimaki, K. (1957): Quantitative investigation on the anterior pituitary-thyroid mechanism during frog metamorphosis. Endocrinol. 61, 35-44

Shi, Y.-B. (1999): Amphibian Metamorphosis. From Morphology to Molecular Biology. John Wiley & Sons, New York, US, 1999

Tietge, J.E., Holcombe, G.W., Flynn, K.M., Kosian, P.A., Korte, J.J., Anderson, L.E., Wolf, D.C. and Degitz, S.J. (2004): Metamorphic inhibition of *Xenopus laevis* by sodium perchlorate: effects on development and thyroid histology. Environ Toxicol Chem (*in press*).

Wong, J., and Shi, Y.-B. (1995): Coordinated regulation of and transcriptional activation by *Xenopus* thyroid hormone and retinoid X receptors. J. Biol. Chem. 270, 18479-18483

Zoeller, R.T. (2003): Challenges confronting risk analysis of potential thyroid toxicants. Risk Anal. 23, 143-162

# ANNEXES

**ANNEX 1:**
**OVERVIEW OF TEST CONDITIONS AND HISTOLOGICAL METHODS USED THE PARTICIPATING LABORATORIES**

**Table 1. Test conditions applied in the three participating laboratories in Phase 1 of the Validation of the Amphibian Metamorphosis Assay**

| | | | Germany | | Japan | | United States | |
|---|---|---|---|---|---|---|---|---|
| **Animal** | | | *Xenopus laevis* | | *Xenopus laevis* | | *Xenopus laevis* | |
| **Exposure period** | | | Exposure from stage 51: 21-days | | Exposure from stage 51: 21-days | | Exposure from stage 51: 21-days | |
| | | | Exposure from stage 54: 14-days | | Exposure from stage 54: 14-days | | Exposure from stage 54: 14-days | |
| **Concentration of test substance** | | **PTU** | 2.5, 5, 10, 20 mg/L | | 2.5, 5, 10, 20 mg/L | | 1.25, 2.5, 5, 10, 20 mg/L | |
| | | **T4** | 0.25, 0.5, 1.0, 2.0 µg/L | | 0.25, 0.5, 1.0, 2.0 µg/L | | 0.25, 0.5, 1.0, 2.0 4.0 µg/L | |
| **Control** | | | 1 dilution water control | | 1 dilution water control | | 1 dilution water control | |
| **Exposure regime** | | | Semi-static | | Flow-through (25 ml/min) | | Flow-through (25 ml/min) | |
| **Larval density** | | | 20 tadpoles/10 L tank | | 20 tadpoles/ 4 L tank | | 25 tadpoles/ 4 L tank | |
| **Replication** | | | 2 replicates for each treatment | | 2 replicates for each treatment | | 2 replicates for each treatment | |
| **Endpoints and determination days** | | | Development stage | Day 0, 7, 14, 21 | Development stage | Day 0, 7, 14, 21 | Development stage | Day 0, 21 |
| | | | Whole body length | | Whole body length | | Whole body length | |
| | | | Hind limb length | Day 7, 14, 21 | Hind limb length | Day 0, 7, 14, 21 | Hind limb length | |
| | | | Wet weight | Day 0, final day | Wet weight | Day 0, final day | Wet weight | Day 0, final day |
| | | | Mortality | Daily observation | Mortality | Daily observation | Mortality | Daily observation |
| **Acceptable mortality rate** | | | <5% in control | | <5% in control | | <5% in control | |
| **Fixation** | **Day** | | Final day of exposure | | Final day of exposure | | Final day of exposure | |
| | **Region** | | Lower jaw; specific tissues (brain, tail) for gene expression analysis | | Whole body | | Head | |
| | **Fixation fluid** | | Bouin's fluid or liquid nitrogen | | Bouin's fluid | | Bouin's fluid | |
| **Feeding** | **Food** | | Sera micron | | Sera micron | | Mixture of TetraFin, Spirulina algae discs, Silver cup trout Starter, along with live brine shrimp | |
| | **Amount** | | 300 mg/tank | | 600 mg/tank | | To be determined | |
| | **Frequency** | | Twice/day | | Twice/day on weekdays, once/day on weekends | | Twice/day on weekdays, once/day on weekends | |
| **Test medium** | **Component** | | Commercial salt mixture Tropic Marine Meersalz to deionized distilled water | | Activated carbon processed water. | | Lake Superior Water (LSW) | |
| | **Concentration** | | 0.025% | | | | - | |
| **Tank (L*W*H)** | | | 30*20*20 cm | | 30.5*15.2*20.3 cm | | 22.5*14*16.5 cm | |
| **Lighting** | **Photoperiod** | | 12 hr light: 12 hr dark | | 12 hr light: 12 hr dark | | 12 hr light: 12 hr dark | |
| | **Intensity** | | To be measured | | To be measured | | Range from 61 to 139 lumens at the water surface | |
| **Water temperature** | | | 22±1ºC | | 22±1ºC | | 21ºC | |
| **pH** | | | 7±0.5 | | 7±0.5 | | 7±0.5 | |
| **Room temperature** | | | 19-22ºC | | 24±1°C | | To be confirmed | |

**Table 2. Comparison of histological methods used by German, US, and Japan laboratories to analyze exposure-related effects of PTU and T4 on thyroid gland.**

| | Germany | US | Japan |
|---|---|---|---|
| Organism sampling | 10 per treatment<br>5 per replicate | 10 per treatment<br>5 per replicate | 6 per treatment<br>3 per replicate |
| Tissue sampling | Lower jaw | Transverse section of head caudal to eyes | Whole body |
| Fixation | Bouin's for 12-24 h | Bouin's for 48h | Bouin's for 12-24 h |
| Storage until processing | at 4°C in 70% EtOH | Neutral buffered formalin (4% formaldehyde) | at 4°C in 70% EtOH |
| Dehydration | Graded alcohol series | Graded alcohol series | Graded alcohol series |
| Embedment | Paraffin | Paraffin | Paraffin |
| Sectioning | 5 µm sections<br>1 of 3 sections used<br>(at least 5 sections) | 5 µm sections<br>2 serial sections at 5 steps 30 µm apart<br>Total 10 sections | 8 µm serial sections<br>11 sections used |
| Staining | Harris's H&E | H&E | Harris's H&E |
| Tissues analyzed | Right lobe only | Right and left lobe | Left lobe only |
| | | | |
| **Parameter** | **Germany** | **US** | **Japan** |
| Qualitative analysis | | | |
| Histological endpoints | Overall size of thyroid gland | Overall size of thyroid gland | Overall size of thyroid gland |
| | Follicle size | Follicle size | Follicular lumen area |
| | Follicle shape | Follicle shape | |
| | Colloid content | Colloid content | |
| | Colloid density | Colloid density | |
| | Follicular cell shape | Follicular cell shape | |
| | Follicular cell height | Follicular cell height | |
| | Epithelial structure | Epithelial structure | |
| | | Follicular cell hyperplasia | |
| Quantitative analysis | | | |
| Histological endpoints | Follicular cell height | none | Overall size of thyroid gland |
| | Stereological analysis of glandular components | | Follicular lumen area |
| | Cross section area | | |

**Annex 2:** **Thyroid Histology Report (US EPA)**
See separate document: **Histology Report US.pdf**

**Annex 3:** **Thyroid Histology Report (Germany)**
See separate document: **Histology Report Germany.pdf**

**Annex 4:** **Thyroid Histology Report (Japan)**
See separate document: **Histology Report Japan.pdf**

**Annex 5:** **SOP Phase 1**

**Annex 6:** **Meeting Report Hiroshima**