Umweltforschungsplan
des Bundesministeriums für Umwelt,
Naturschutz und Reaktorsicherheit

Umweltchemikalien, Schadstoffwirkungen

Förderkennzeichen (UFOPLAN) 203 67 450

Initial Draft Report: OECD Validation Phase 1 (Optimisation) for the Amphibian Metamorphosis Assay
(Frosch-Metamorphose-Assay – Optimierung des Testprotokolls)

von
Robert Opitz *et al.*

Leibnitz-Institut für
Gewässerökologie und Binnenfischerei
IGB, Berlin

Leiter der Abteilung Binnenfischerei
Prof. Dr. Werner Kloas

IM AUFTRAG
DES UMWELTBUNDESAMTES

Mai 2004

Berichts-Kennblatt

1. Berichtsnummer UBA-FB	2. Umweltchemikalien	3. 203 67 450

4. Titel des Berichts
Frosch-Metamorphose-Assay - Optimierung des Testprotokolls
Initial Draft Report: OECD Validation Phase 1 (Optimisation) for the Amphibian Metamorphosis Assay

5. Autor(en), Name(n), Vorname(n)
Opitz, Robert; Braunbeck, Thomas; Tooi, Osamu;
Sagoe, Charles; Degitz, Sigmund; Tietge, Joe;
Kloas, Werner

6. Durchführende Institution (Name, Anschrift)

Leibniz-Institut für Gewässerökologie und Binnenfischerei
Prof. Dr. Werner Kloas
Müggelseedamm 310
12587 Berlin

7. Fördernde Institution (Name, Anschrift)

Umweltbundesamt, Postfach 33 00 22, D-14191 Berlin

8. Abschlußdatum
Mai 2004

9. Veröffentlichungsdatum
Juni 2004

10. UFOPLAN-Nr.
203 67 450

11. Seitenzahl

12. Literaturangaben

13. Tabellen und Diagramme

14. Abbildungen

15. Zusätzliche Angaben
Wichtigster Zweck des Projekts war, die Studienergebnisse direkt in den internationalen Arbeitsprozess auf OECD-Ebene zu integrieren. Dadurch sind die Ergebnisse von Labors aus anderen Mitgliedsländern von Anfang an in den Bericht integriert. Das Berichtsformat weicht vom üblichen Muster für UFOPLAN-Projekte ab, da die Gestaltung den Erfordernissen für OECD-Arbeitsdokumente folgen musste. Alle Berichtsdateien stehen elektronisch z. Verfgg.

16. Kurzfassung
Die aktuelle Studie ist Bestandteil von UBA- und OECD-koordinierten Anstrengungen zur Entwicklung eines Amphibien-Metamorphose-Tests zur Bewertung von Chemikalien mit potentieller Wirkung auf das Schilddrüsensystem. Ein zentrales Ziel der Studie war der Vergleich der Empfindlichkeit zweier Expositionsszenarien des Xenopus Metamorphose Tests (XEMA) um stimulierende als auch hemmende Wirkungen zu detektieren. Xenopus laevis Kaulquappen wurden vergleichend im Stadium 51 für 21 Tage und im Stadium 54 für 14 Tage exponiert. Als Testsubstanzen wurden Propylthiouracil (2,5; 5; 10, 20 mg/l PTU) und Thyroxin (0,25; 0,5; 1,0; 2,0 µg/l T4) verwendet. Entwicklungshemmende Wirkungen wurden für 10 und 20 mg/l PTU (Stadium 51) und für 20 mg/l PTU (Stadium 54) festgestellt. In Versuchen mit Stadium 51 Kaulquappen wurde eine Hemmung des Hinterbeinwachstums durch 10 und 20 mg/l PTU ermittelt. Histologische Veränderungen der Schilddrüse nach Behandlung mit 5, 10 und 20 mg/l PTU umfassten vergrösserte Follikel, eine Hypertrophie des Gesamtorgans sowie eine verstärkte Kolloidresorption. Die histologischen Effekte waren in Versuchen mit Stadium 51 Kaulquappen stärker ausgeprägt. Eine beschleunigte Metamorphose wurde im Stadium 51 nach Behandlung mit 1,0 und 2,0 µg/l T4 und im Stadium 54 nach Behandlung mit 2,0 µg/l T4 beobachtet. T4 beschleunigte das Hinterbeinwachstum bei Stadium 51 (0,5, 1,0 und 2,0 µg/l T4) und Stadium 54 Kaulquappen (0,25 und 2,0 µg/l T4). Veränderungen der Schilddrüsenhistologie waren in den T4 Studien weitaus weniger ausgeprägt als in den Versuchen mit PTU. Behandlung mit PTU erhöhte die Expression der Thyreotropin ß-Untereinheit in der Hypophyse, während die Expression von Schilddrüsenhormon-induzierten Genen im Gehirn durch PTU unterdrückt und durch T4 stimuliert wurde. Die Resultate der Studie belegen, dass eine Exposition von Stadium 51 Kaulquappen zu einer erhöhten Sensitivität des XEMA Tests führen.
Die vorliegende Studie bestätigt weiterhin, dass eine kombinierte Analyse von morphologischen, histologischen und molekularbiologischen Parametern einen sensitiven und diagnostischen Testansatz zur Erfassung agonistischer (T4) und antagonistischer (PTU) Wirkungen auf das Hormonsystem darstellt.

17. Schlagwörter
Chemikalien; Stoffe; Laboruntersuchung; Schadstoffwirkung; Ringversuch; Testsubstanz; Ökotoxikologische Bewertung; Chronische Toxizität; Versuchstier; Lurch; Schilddrüse; Subletale Konzentration; Tierversuch; Prüfverfahren; Exposition; Statistische Auswertung; Endokrines System; Testorganismus; Embryo; Metamorphose; Biologische Wirkung; UFOPLAN 2003; Hormonelle Wirkung; Krallenfrosch; Xenopus laevis; Endokrine Disruptoren

18. Preis 36.171,- €	19.	20.

Report Cover Sheet

1. Report No. UBA-FB	2. Umweltchemikalien	3. 203 67 450
4. Report Title Amphibian Metamorphosis Assay - Test Protocol Optimisation Initial draft report: OECD Validation Phase I (Optimisation) for the Amphibian Metamorphosis Assay		

5. Autor(s), Family Name(s), First Name(s) Opitz, Robert; Braunbeck, Thomas; Tooi, Osamu; Sagoe, Charles; Degitz, Sigmund; Tietge, Joe; Kloas, Werner	8. Report Date May 2004
6. Performing Organisation (Name, Address) Leibnitz-Institut für Gewässerökologie und Binnenfischerei Prof. Dr. Werner Kloas Müggelseedamm 310 12587 Berlin	9. Publication Date June 2004
	10. UFOPLAN-Ref. No. 203 67 450
	11. No. of Pages
7. Sponsoring Agency (Name, Address) Umweltbundesamt, Postfach 33 00 22, D-14191 Berlin	12. No. of Reference
	13. No. of Tables, Diagrams
	14. No. of Figures

15. Supplementary Notes
The most important aim of this study was a direct integration of research results into the international evaluation process coordinated at the OECD level. The reporting format of this report deviates from the normal scheme used for UFOPLAN-projects because data from other OECD member country activities were included in the report which had to meet formatting demands for OECD documents. All study reporting files are available in electronic form.

16. Abstract
This study was part of the concerted efforts of UBA and OECD to develop an amphibian metamorphosis assay as a screening tool to assess chemicals for their potential to disrupt normal thyroid function. The primary objective of the present study was to evaluate two different exposure scenarios of the Xenopus Metamorphosis Assay (XEMA) regarding their sensitivity to detect stimulating and inhibiting effects of thyroid system-disrupting chemicals. Xenopus laevis tadpoles at stage 51 were exposed for 21 days and for comparison, tadpoles at stage 54 were exposed for 14 days. The chemicals included in this testing were 6-propylthiouracil (PTU, 2.5, 5.0, 10 and 20 mg/L) and thyroxine (T4, 0.25, 0.5, 1.0, 2.0 µg/L). PTU inhibited metamorphic development at 10 and 20 mg/L in the stage 51 exposure study and at 20 mg/L in the stage 54 exposure study. Hind limb growth was reduced at 10 and 20 mg/L PTU in the stage 51 study. Thyroid gland histology revealed increases in thyroid gland size, follicle size and colloid resorption at 5, 10 and 20 mg/L PTU. The severity and prevalence of histological changes was increased in the stage 51 exposure study. T4 enhanced metamorphosis at 1.0 and 2.0 µg/L in the stage 51 exposure study and at 2.0 µg/L in the stage 54 exposure study. Hind limb growth was increased at 0.5, 1.0 and 2.0 µg/L T4 in the stage 51 exposure study and at 0.25 and 2.0 µg/L T4 in the stage 54 exposure study. Thyroid gland histology showed less marked changes induced by T4 compared to PTU.
Gene expression analysis revealed increased thyrotropin-ß subunit gene expression by PTU only. PTU treatment also caused reduced expression of thyroid hormone-regulated genes in brain tissue while T4 treatment led to increased expression levels.
This study confirmed that a combined analysis of the proposed morphological, histological and molecular endpoints provides a functional, sensitive and diagnostic approach for detection of agonistic (T4) and antagonistic effects (PTU) on the thyroid system in Xenopus tadpoles. The present results further indicate that a higher sensitivity of the XEMA assay was achieved when exposure of tadpoles was initiated at stage 51 compared to stage 54.

17. Keywords
chemicals; laboratory research; pollutants effects; ringtest; test substance; ecotoxicological assessment
chronic toxicity; test organism; frog; thyroid gland; sublethal concentration; substances
risk assessment; exposure; statistical analysis; endocrine system; embryo; larvae; metamorphosis;
UFOPLAN 2003; hormonal effects; clawed frog; Xenopus laevis; endocrine disrupters

18. Price 36,171 €	19.	20.

NB:

Dies ist die über Internet verfügbare Version des UBA Forschungsberichts 000758 zum UFOPLAN-Projekt „Frosch-Metamorphose-Assay – Optimierung des Testprotokolls“ mit der FKZ 203 67 450. Das Projekt lieferte einen Beitrag zum internationalen OECD Prüfrichtlinienprogramm. Der Forschungsbericht spiegelt das wider, indem er eine Reihe von Arbeitsdokumenten der beteiligten OECD-Gremien enthält, die unter Beteiligung dieses UFOPLAN-Projekts entstanden sind. Die Papierversion des Forschungsberichts kann in der Fachbibliothek des UBA entliehen werden (Signatur FB-000758) und entspricht dem Stand von Mai 2004.

Diese elektronisch verfügbare Version des Forschungsberichts ist im Januar 2005 fertiggestellt worden. Da der übergeordnete internationale Arbeitsprozess kontinuierlich fortschreitet, erschien es sinnvoll, neuere Versionen einiger OECD-Arbeitsdokumente (Stand Dezember 2004) zu integrieren. Deshalb sind Papier- und elektronische Version dieses Berichts nicht in allen Teilen identisch.

Aktuelle Informationen über das OECD Prüfrichtlinienprogramm und die Arbeiten zu hormonell wirksamen Stoffen finden sich im öffentlichen Internet-Auftritt der OECD unter der Adresse www.oecd.org/env/testguidelines, dort weiter bei *Endocrine Disrupters – Testing, Assessment*.

NB:
This is the internet available version of the UBA Research Report 000758 for R+D project “Amphibian Metamorphosis Assay – Test Protocol Optimisation”, project number (FKZ) 203 67 450. The project contributed to the international OECD Test Guidelines Programme. This is reflected by the Research Report comprising several working documents which have been drafted in responsible OECD expert groups under direct contribution from this UFOPLAN-project. The paper version of this report is available in the Scientific Library of UBA (German Federal Environmental Agency) with signatory FB-000758 and reflects the status of May 2004.

The present e-version of the report has been completed in January 2005. Since the international working process advances continuously, it appeared reasonable to incorporate updated versions of some OECD working documents (status of December 2004). Thus, paper and electronic version of this report are not fully identical.

Topical information about the OECD Test Guidelines Programme and ongoing work on endocrine disrupters is available on the public OECD websites: www.oecd.org/env/testguidelines, then *Endocrine Disrupters – Testing, Assessment*.

DRAFT

Report of the Validation of the Amphibian Metamorphosis Assay for the detection of thyroid active substances:

Phase 1 – Optimisation of the Test Protocol

September 2004

TABLE OF CONTENTS

1. Summary 7
2. Introduction 8
3. Objective of the Inter-Laboratory Comparison Study 9
4. Methods 9
  4.1. Overview of Test Conditions 9
  4.2. Exposure Experiments 10
  4.3. Statistical Analysis 10
5. Results 11
  5.1. Analytical Chemistry 11
    5.1.1. Analytical Chemistry Results (US lab) 11
    5.1.2. Analytical Chemistry Results (JPN lab) 12
  Comparison of Control Data 12
  5.2. Exposure Studies with Propylthiouracil (PTU) 17
    5.2.1. Stage 51 Exposure Studies with PTU (21 Day Assay) 17
    5.2.2. Stage 54 Exposure Studies with PTU (14 Day Assay) 29
  5.3. Exposure Studies with Thyroxine (T4) 39
    5.3.1. Stage 51 Exposure Studies with T4 (21 Day Assay) 39
    5.3.2. Stage 54 Exposure Studies with T4 (14 Day Assay) 50
6. Discussion 60
  6.1. Control Organism Performance 60
  6.2. Effects Pattern of PTU on Metamorphosis and Thyroid System 60
  6.3. Endpoint Sensitivity in PTU Studies 61
    6.3.1. Developmental Stage 61
    6.3.2. Hind Limb Length 63
    6.3.3. Thyroid Histopathology 63
    6.3.4. Gene Expression Analysis 66
    6.3.5. Body Length and Body Weight 67
  Effects Pattern of T4 on Metamorphosis and Thyroid System 67
  6.4. Comparison of Endpoint Sensitivities in T4 Studies 68
    6.4.1. Developmental Stage 68
    6.4.2. Hind Limb Length 70
    6.4.3. Thyroid Histopathology 70
    6.4.4. Gene Expression Analysis 72
    6.4.5. Body Length and Body Weight 73
  6.5. Critical Review of Results from Validation Phase 1 and Proposal for Phase 2 Validation Activities 73
7. References 77

List of Tables:

Table 1. Overview of exposure studies conducted by the three participating labs 10
Table 2. Measured PTU concentrations (mg/L) in the stage 51 study conducted for 21 days. Sample size for each replicate was 4. 11
Table 3. Measured PTU concentrations (mg/L) in the stage 54 study conducted for 14 days. Sample size for each replicate was 3. 11
Table 4. Measured T4 concentrations (µg/L) in the stage 51 study conducted for 21 days. Sample size for each replicate was 2. 11
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 11
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. 12
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 12
Table 8. Distribution of developmental stages reached by control tadpoles within 21 days during exposure studies initiated at stage 51 in three different labs. 14
Table 9. Distribution of developmental stages reached by control tadpoles within 14 days during exposure studies initiated at stage 54 in three different labs. 14
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. 16
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 16
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 ($p<0.05$). 18
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. 20
Table 14. Summary of results from developmental stage determination during the stage 51 study with PTU in the GER lab. 22
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. 23
Table 16. Summary of results from developmental stage determination during the stage 51 study with PTU in the JPN lab. 24
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. 25
Table 18. Summary of results from developmental stage determination during the stage 51 study with PTU in the US lab. 26
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 ($p<0.05$). 30
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. 31
Table 21. Summary of results from developmental stage determination during the stage 54 study with PTU in the GER lab. 32
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. 33
Table 23. Summary of results from developmental stage determination during the stage 54 study with PTU in the JPN lab. 34
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. 35

Table 25. Summary of results from developmental stage determination during the stage 54 study with PTU in the US lab. 36
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). 40
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. 42
Table 28. Summary of results from developmental stage determination during the stage 51 study with T4 in the GER lab. 43
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. 44
Table 30. Summary of results from developmental stage determination during the stage 51 study with T4 in the JPN lab. 45
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. 46
Table 32. Summary of results from developmental stage determination during the stage 51 study with T4 in the US lab. 47
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). 51
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. 52
Table 35. Summary of results from developmental stage determination during the stage 54 study with T4 in the GER lab. 53
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. 54
Table 37. Summary of results from developmental stage determination during the stage 54 study with T4 in the JPN lab. 55
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. 56
Table 39. Summary of results from developmental stage determination during the stage 54 study with T4 in the US lab. 57
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 < 0.05) for all analyses. Multiple comparisons between all treatment groups and the controls by Dunn's method are summarized in this table. 62
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 < 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. 62
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. 64
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. 65
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. 69
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. 69

Table 46. Summary of endpoint sensitivities as detected in stage 51 studies with T4 71
Table 47. Summary of endpoint sensitivities as detected in stage 54 studies with T4 72

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. 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. 15

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)... 28

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)... 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)... 49

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 (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)... 59

1. 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.

2. 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.

3. 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

4. METHODS

4.1. 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.

4.2. 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

4.3. 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). Dunnett's test was used to compare data from the control group to all other treatment groups.

13. Developmental stage data 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. Differences in growth and developmental indices between treatments were considered significant at the level of $p<0.05$.

14. Gene expression data 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 $p<0.05$.

5. RESULTS

5.1. Analytical Chemistry

5.1.1. Analytical Chemistry Results (US lab)

15. 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.

16. 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

17. 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

5.1.2.Analytical Chemistry Results (JPN lab)

18. 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

19. 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

20. 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.

21. 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

22. 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.

23. 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.

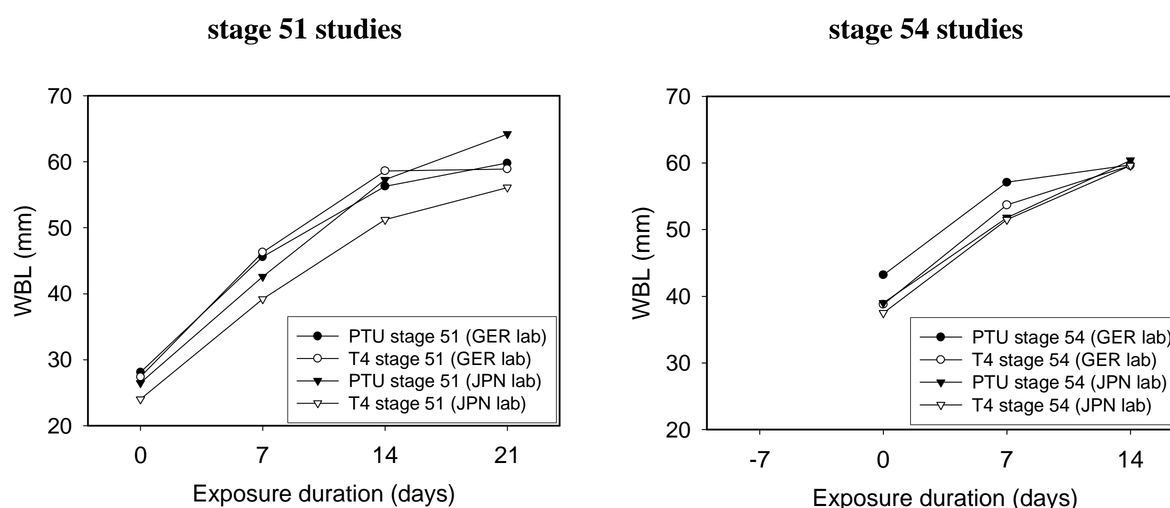


Figure 1. Temporal changes in whole body length (WBL) as determined for control tadpoles during the conduct of stage 51 (21 days) and stage 54 (14 days) exposure studies. Data from 8 independent experiments performed in the JPN and GER lab are shown. Mean values are shown while standard deviation bars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

24. 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

5.2. Exposure Studies with Propylthiouracil (PTU)

25. 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).

5.2.1. Stage 51 Exposure Studies with PTU (21 Day Assay)

Developmental stage

26. 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

27. 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

28. 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

29. 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 ± 150.1	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

30. 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

31. 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.

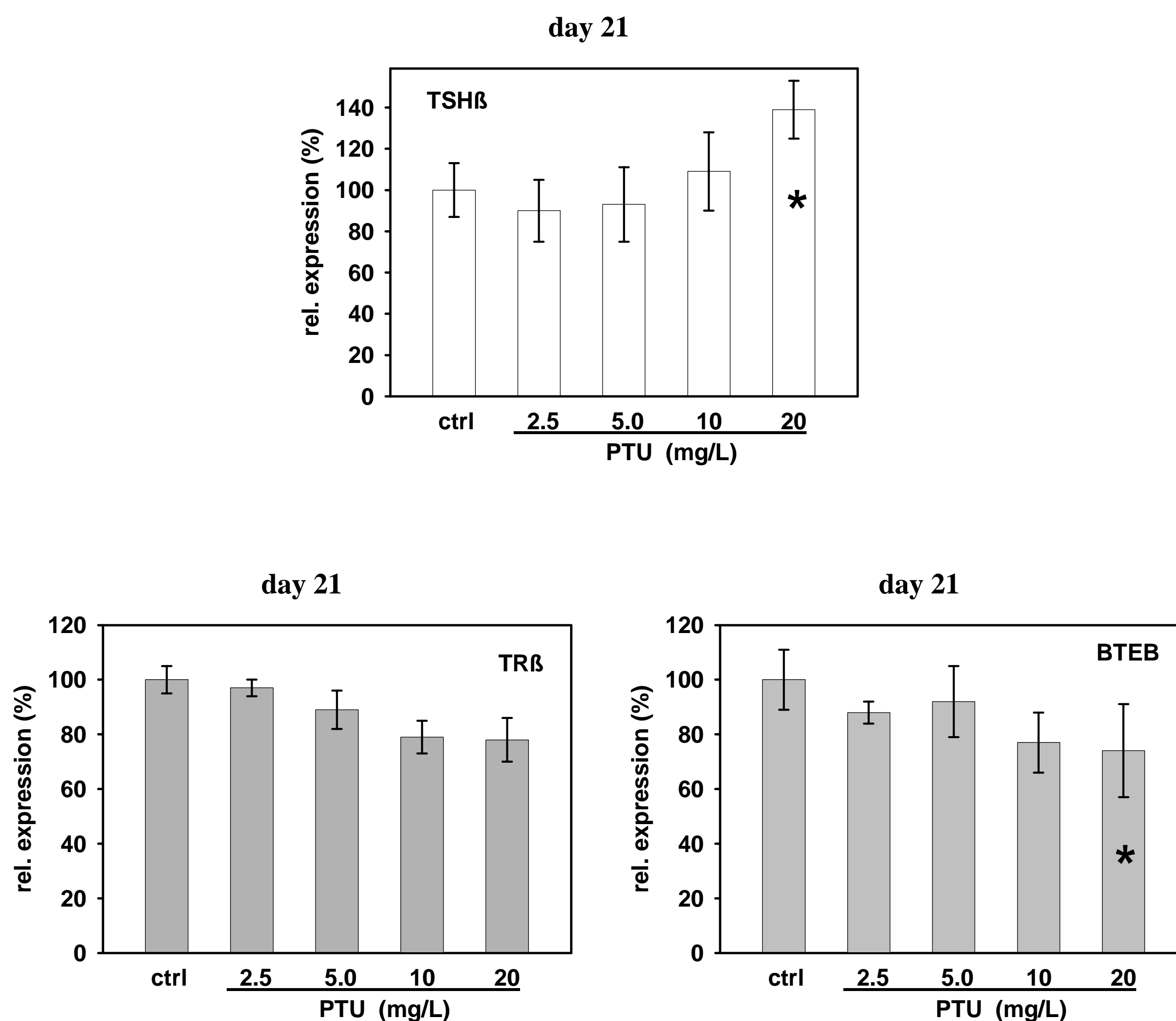


Figure 2. Effects of PTU on gene expression in tadpole brain. RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 51 exposure study with PTU in the GER lab. Semiquantitative RT-PCR analyses of mRNA expression of thyrotropin β -subunit (TSH β), thyroid hormone receptor 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).

5.2.2.Stage 54 Exposure Studies with PTU (14 Day Assay)

Developmental Stage

32. 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

33. 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

34. 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

35. 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 ± 102.7	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 ± 104.6	1089.8 ± 136.0	1062.2 ± 158.6	933.5 ± 176.1	1115.6 ± 106.1	
	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

36. 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

37. 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.

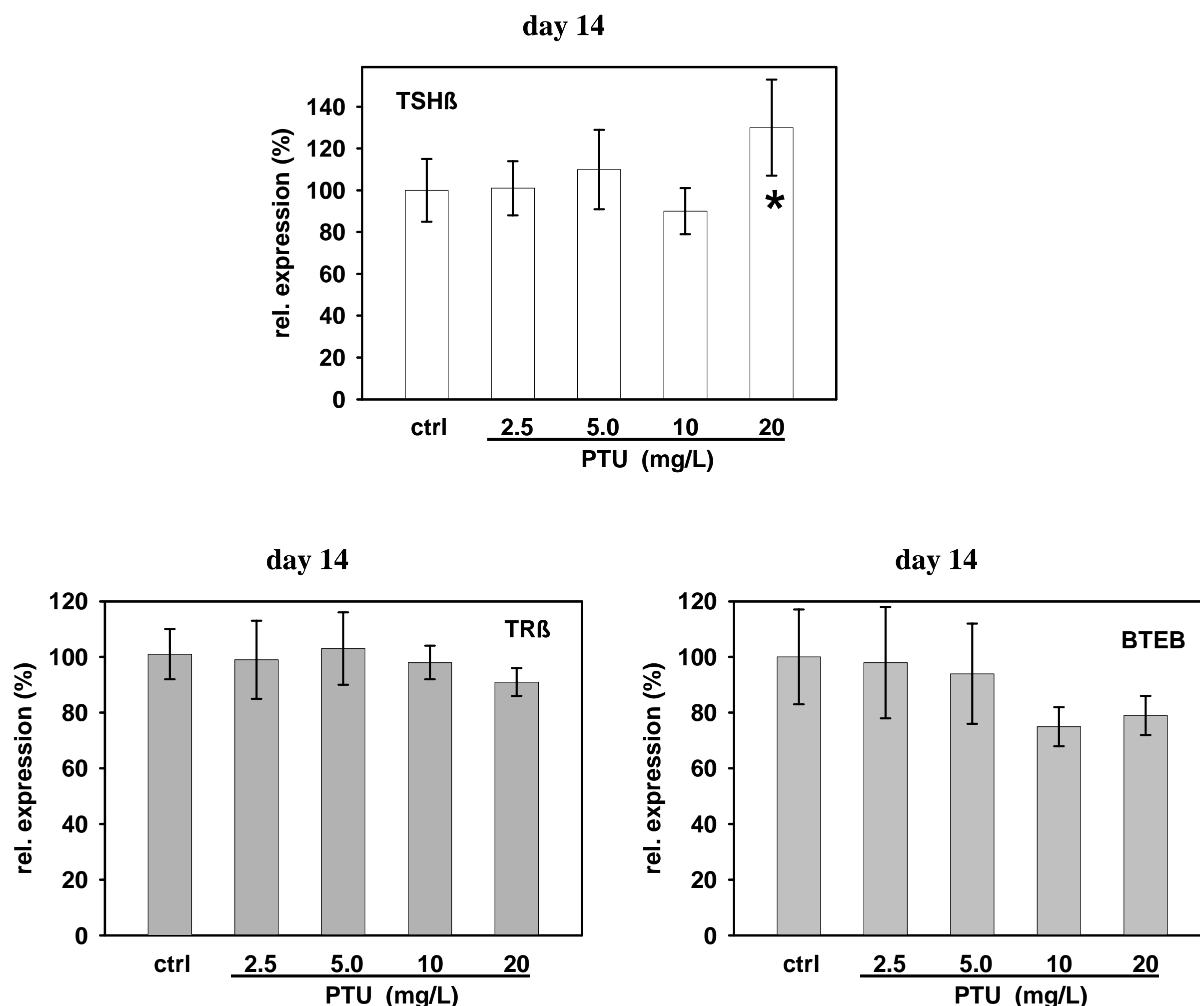


Figure 3. Effects of PTU on gene expression in tadpole brain. RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 54 exposure study with PTU in the GER lab. Semiquantitative RT-PCR analyses of mRNA expression of thyrotropin β -subunit (TSH β), thyroid hormone receptor 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).

5.3. Exposure Studies with Thyroxine (T4)

38. 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).

5.3.1. Stage 51 Exposure Studies with T4 (21 Day Assay)

Developmental Stage

39. 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.

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

Hind Limb Length

40. 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

41. 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

42. 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 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 ± 131.0	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 ± 107.0	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

43. 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

44. 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.

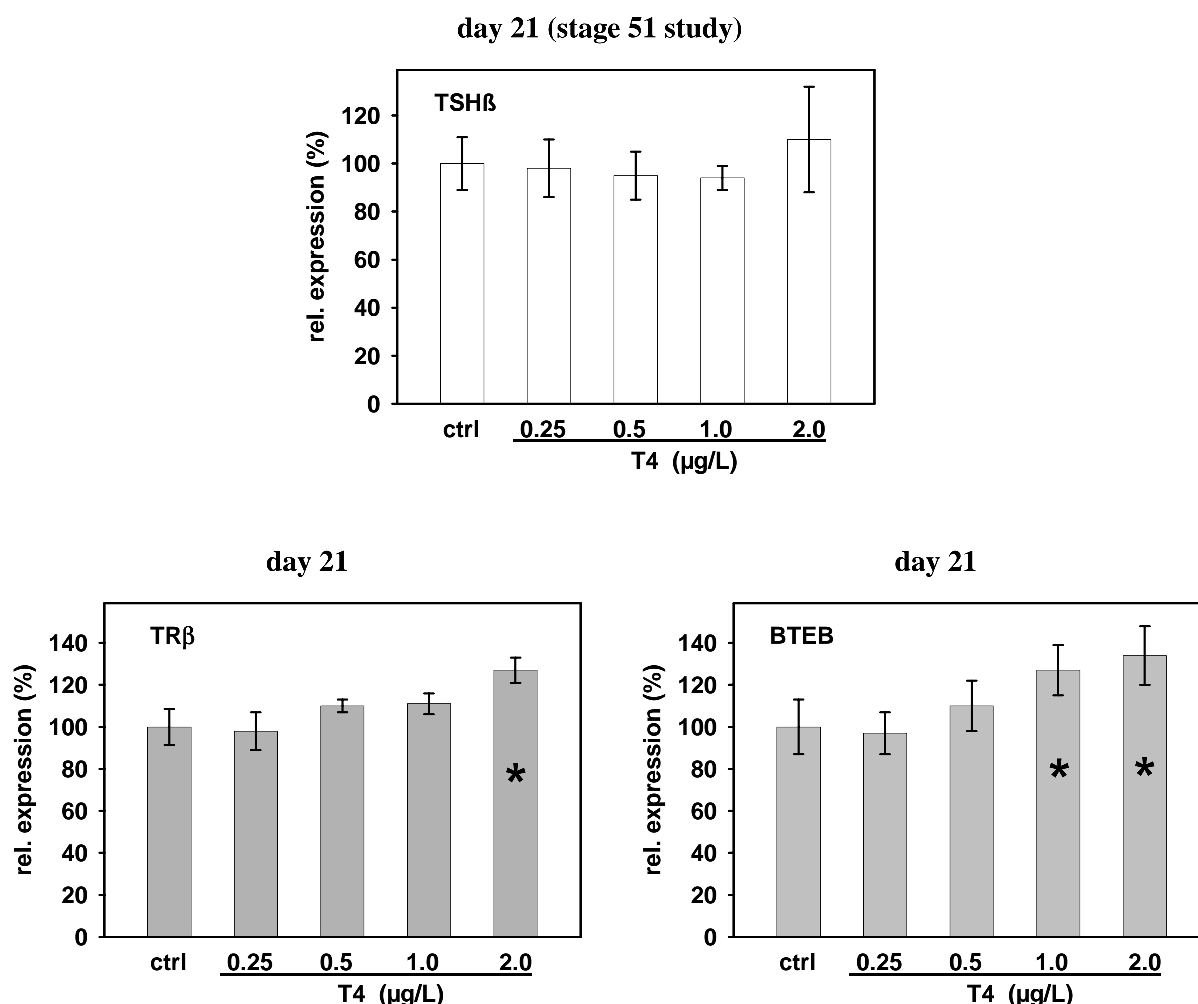


Figure 4. Effects of T4 on gene expression in tadpole brain. RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 51 exposure study with T4 in the GER lab. Semiquantitative RT-PCR analyses of mRNA expression of thyrotropin β -subunit (TSH β), thyroid hormone receptor 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).

5.3.2.Stage 54 Exposure Studies with T4 (14 Day Assay)

Developmental Stage

45. 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

46. 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

47. 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).

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 ± 170.6	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

48. 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

49. 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

50. 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.

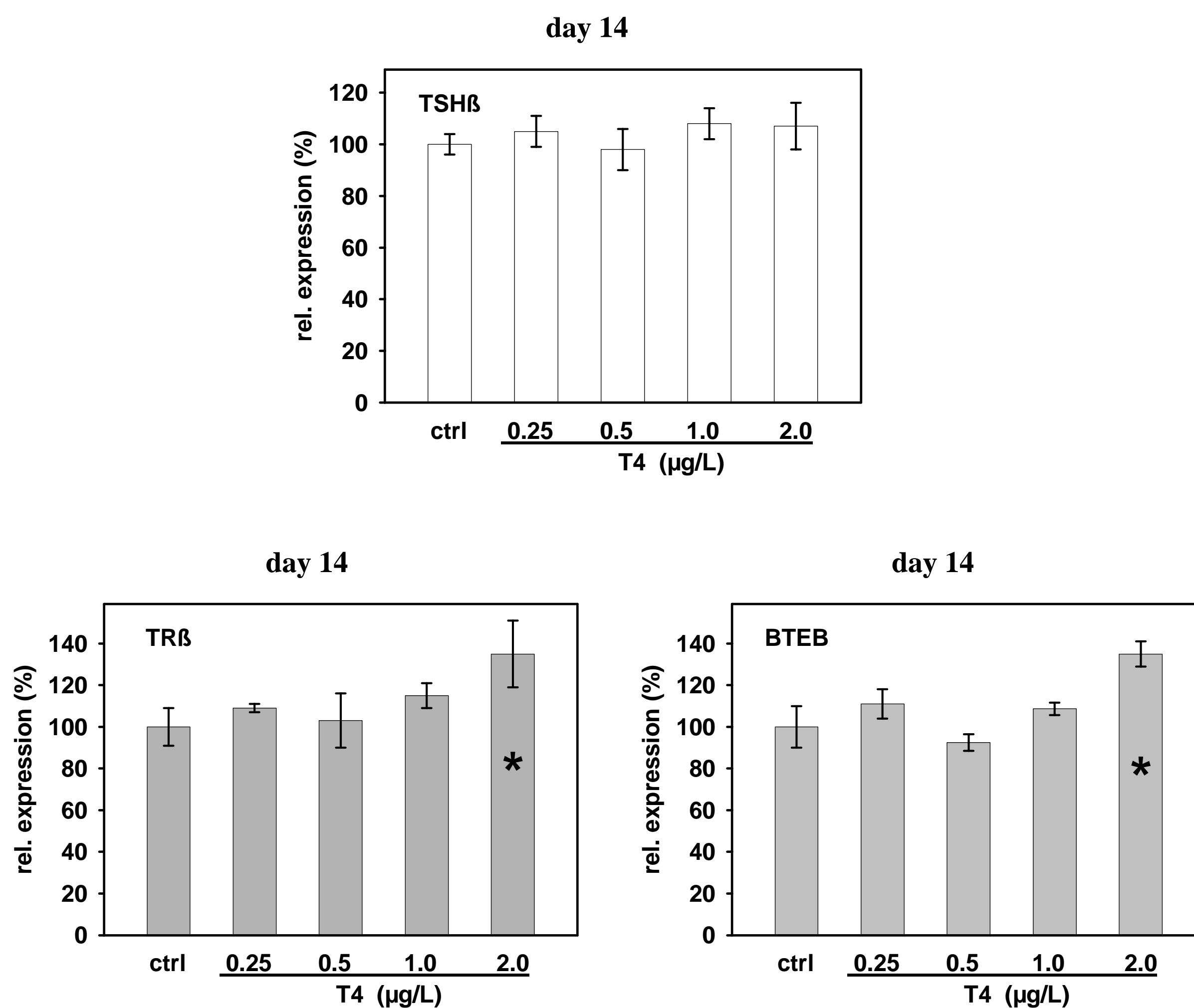


Figure 5. Effects of T4 on gene expression in tadpole brain. RNA was isolated from brain/pituitary tissue samples taken at test termination of the stage 54 exposure study with T4 in the GER lab. Semiquantitative RT-PCR analyses of mRNA expression of thyrotropin β-subunit (TSHβ), thyroid hormone receptor 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).

6. DISCUSSION

6.1. Control Organism Performance

51. 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.

6.2. Effects Pattern of PTU on Metamorphosis and Thyroid System

52. 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.

6.3. Endpoint Sensitivity in PTU Studies

6.3.1.Developmental Stage

53. 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

6.3.2.Hind Limb Length

54. 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.

6.3.3.Thyroid Histopathology

55. 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;

56. 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

57. 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.

58. 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.

59. 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.

60. 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.

6.3.4.Gene Expression Analysis

61. 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.

6.3.5.Body Length and Body Weight

62. 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.

63. 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

64. 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).

6.4. Comparison of Endpoint Sensitivities in T4 Studies

6.4.1. Developmental Stage

65. 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.

66. 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

6.4.2.Hind Limb Length

67. 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.

6.4.3.Thyroid Histopathology

68. 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

69. 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.

6.4.4.Gene Expression Analysis

70. 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.

71. 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.

6.4.5.Body Length and Body Weight

72. 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.

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

73. 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.

74. 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.

75. 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.

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

77. 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.

78. 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.

79. 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.

80. 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.

81. 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.

82. 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.

83. 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.

84. 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.

85. 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.

86. 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.

87. 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.

88. 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.

7. 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	

		Germany	Japan	United States
	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 5 per replicate	10 per treatment 5 per replicate	6 per treatment 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 1 of 3 sections used (at least 5 sections)	5 µm sections 2 serial sections at 5 steps 30 µm apart Total 10 sections	8 µm serial sections 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)**
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**

Appendix 3: Thyroid Histology Report (Germany)

Methods

Lower jaw tissue including the thyroid gland was taken from 10 randomly selected tadpoles per treatment group at test termination of each of the IGB studies. The methods used for histological analysis are described in detail in the corresponding SOP. Slides were prepared at the Department of Zoology at the University of Heidelberg (Germany). Tissues were sectioned (5 µm thick) in a transverse plane from dorsal to ventral and stained with hematoxylin and eosin. For each tadpole, 5 sections of the middle part of the right thyroid lobe were analyzed for exposure-related changes in the following qualitative parameters:

- Overall size of the thyroid gland (reduction, increase)
- Follicle size (reduction, increase)
- Follicle shape (regular, irregular, uniform)
- Colloid content (increase or reduction in colloid area; absence of colloid)
- Colloid density (homogeneous or heterogeneous tinctoral quality; pale, lacy, or granular colloid; peripheral vacuolation of colloid)
- Follicular cell shape (cuboidal, columnar, tall)
- Follicular cell height (increase, reduction)
- Structure of the epithelium (single cell layer or stratification resulting in multiple layers, papillary infoldings of the epithelial cell layer into the lumen)

In addition to a qualitative description, attempts were made to quantify selected changes in thyroid gland morphology by means of image analyzing techniques. The right lobe of 10 tadpoles per treatment group was analyzed regarding epithelial cell height and total cross section area. Epithelial cell height was measured for 3 randomly selected follicles per section. The height of 4 follicular cells per follicle was measured and a total of 5 different sections of the right lobe were analyzed for each animal (n=60 measurements per lobe). The maximum cross section area of the right lobe was determined by analyzing the total cross sectioon area for at least 5 sections of the middle part of the right lobe.

Effects of PTU Exposure on Thyroid Histology

In the stage 51 and the stage 54 exposure studies with PTU, the light histological appearance of thyroid glands from the 2.5 mg/L PTU treatment group did not differ from the control group (Figure 1). A single layer of cuboidal follicular epithelial cells was present in glands from these treatment groups. The colloid showed a homogeneous tinctoral quality and peripheral vacuolation of the colloid was minimal.

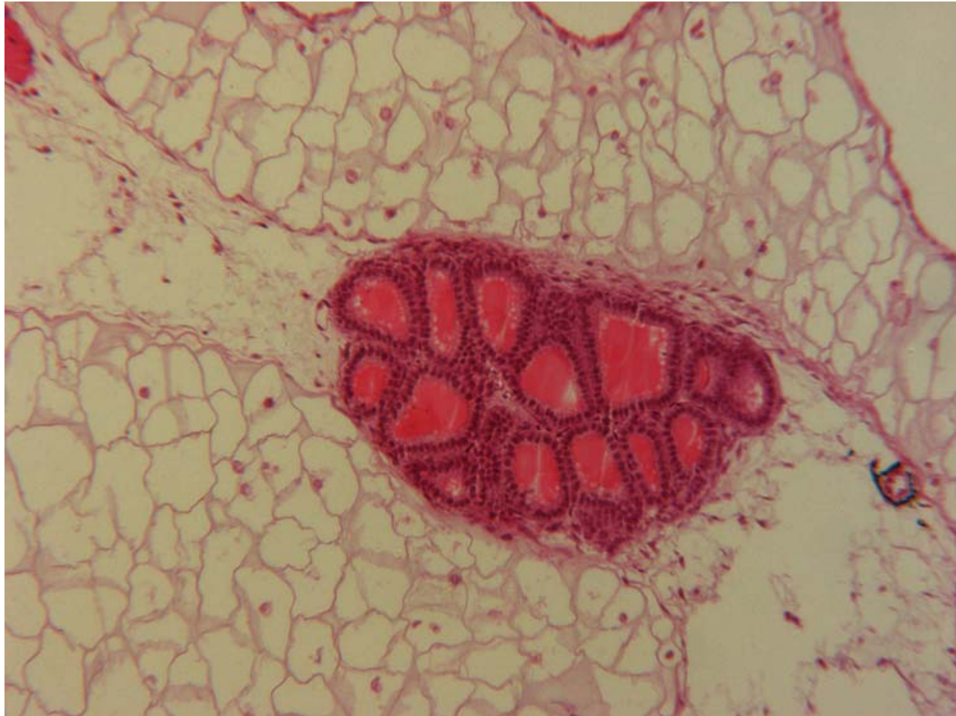
At PTU concentrations of 5, 10 and 20 mg/L, concentration-dependent increases in thyroid gland size, follicle size and the degree of colloid resorption were observed in both exposure studies with PTU. The prominent changes seen in the 5 mg/L PTU group included a minimal increase in follicle size and a resulting diffuse enlargement of the thyroid gland (Figures 2A, 5A). The prevalence of these changes was greater in tadpoles from the stage 51 exposure study compared to the stage 54 study. At 5 mg/L PTU, the follicular epithelium still consisted of a single layer of cuboidal follicular cells and as judged from the homogeneous tinctoral quality of the colloid, no changes were detectable for the density of the colloid. The selected changes observed in the 5 mg/L PTU group were also seen in tadpoles from the 10 mg/L PTU treatment group but increased in prevalence and severity (Figures 2B, 5B). In addition, the marked peripheral vacuolation of the colloid as observed in some glands of the 10 mg/L PTU group indicated a partial depletion of colloid content. Furthermore, a number of follicles in the glands from the 10 mg/L PTU group showed a change in shape of epithelial cells from cuboidal to columnar. The thyroid glands of tadpoles from the 20 mg/L PTU group were characterized by marked increases in follicle size and a pronounced enlargement of the gland (Figures 3, 6). A high prevalence of follicular cell hypertrophy and hyperplasia was observed together with markedly enhanced colloid depletion. Further, collapsed follicles devoid of colloid were noted in few glands from this treatment group.

Comparing the results from the qualitative analyses of thyroid gland histology between the two exposure studies, it was found that particularly at the lower PTU

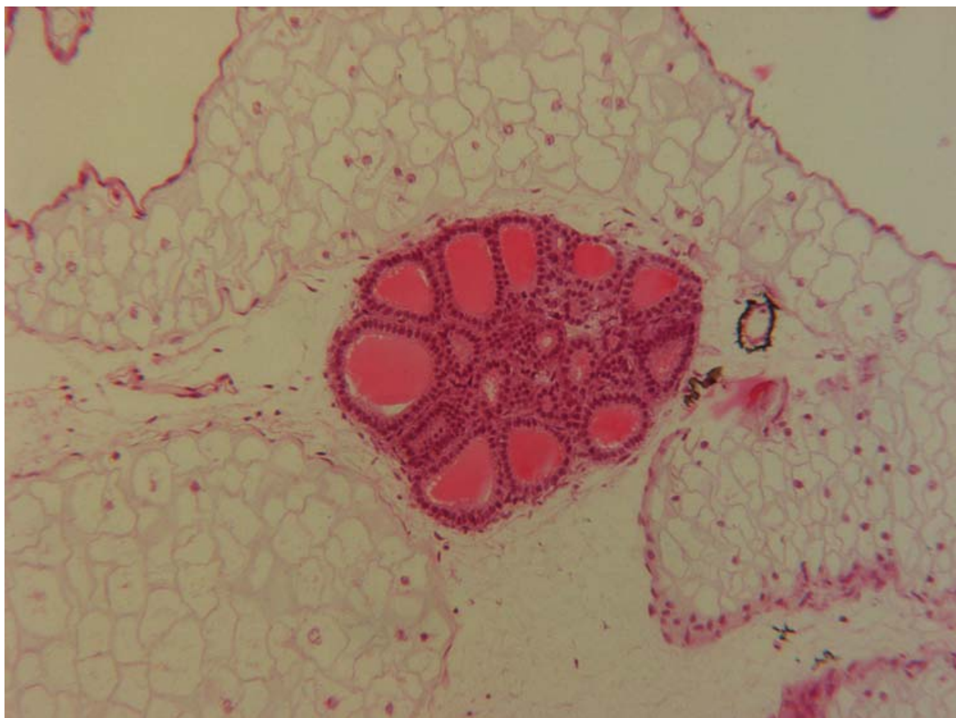
concentrations (5 and 10 mg/L), prevalence and severity of the aforementioned changes in thyroid gland histology were higher in tadpoles exposed from stage 51 for 21 days than in tadpoles exposed from stage 54 for 14 days.

Results from the quantificative measurements of epithelial cell heights are shown in Figure 7. In the stage 51 exposure study, increases in epithelial cell height were observed at the two highest PTU concentration (10 and 20 mg/L), but only the effect at 20 mg/L PTU was statistically significant. In the stage 54 exposure study, the epithelial cell height was significantly increased at 10 and 20 mg/L PTU. Thus, the results of these quantitative measurements confirm the qualitative description of the thyroid glands regarding the presence of follicular cell hypertrophy in tadpoles exposed to 10 and 20 mg/L PTU.

The maximum cross section area of the right thyroid lobe was determined in order to quantify the apparent enlargement of the thyroid glands in PTU-exposed tadpoles. In the stage 51 exposure study, enlargement of the thyroid glands following exposure to 5, 10 and 20 mg/L PTU could be confirmed by detection of increases in the maximum cross section area of the right lobe (Figure 8). This effect was statistically significant at PTU concentrations of 10 and 20 mg/L. The mean values of the maximum cross section area were approximately 2 and 3.5 times greater at these PTU concentrations than in untreated control. Statistically significant increases in the maximum cross section area of the right thyroid lobe were also observed in tadpoles exposed to 10 and 20 mg/L PTU during the stage 54 exposure study (Figure 8). In this study, the effect was less pronounced as evident from 1.7- and 2.3-fold increases in mean values of the maximum cross section area.

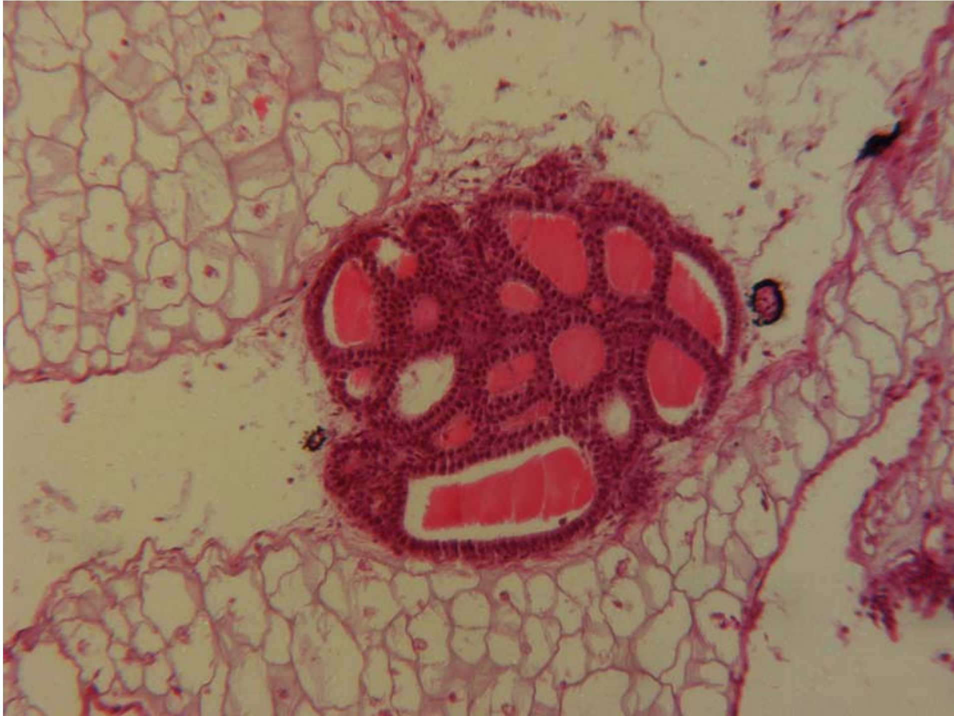


(A) control

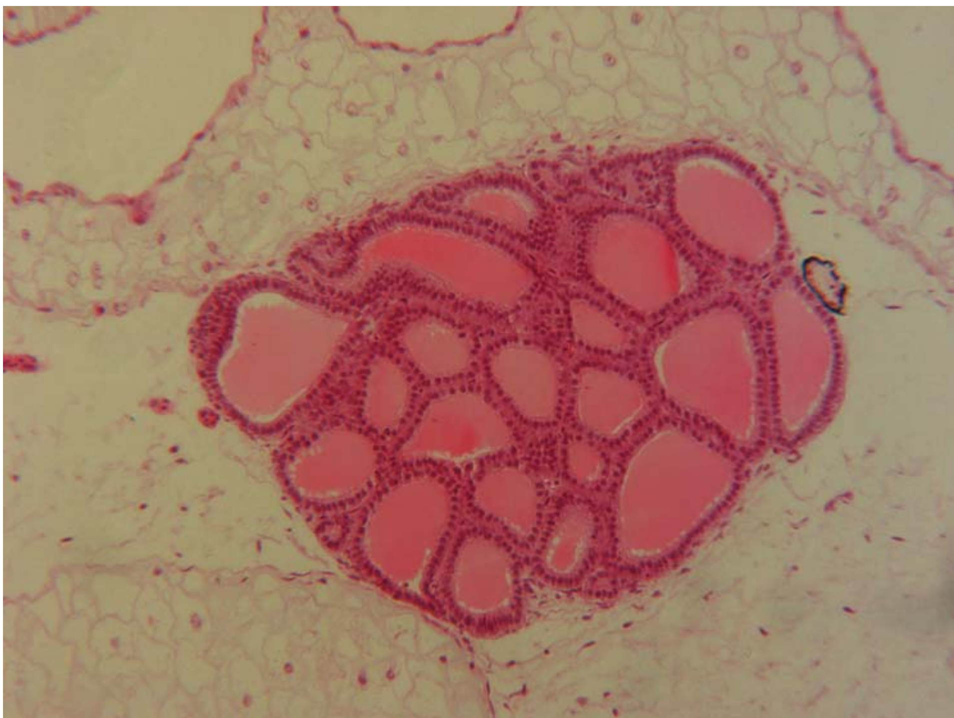


(B) 2.5 mg/L PTU

Figure 1. Light micrographs (10 x magnification) of histological sections of thyroid glands from *X. laevis* tadpoles after 21 days of exposure to PTU during the stage 51 exposure study. See text for details.

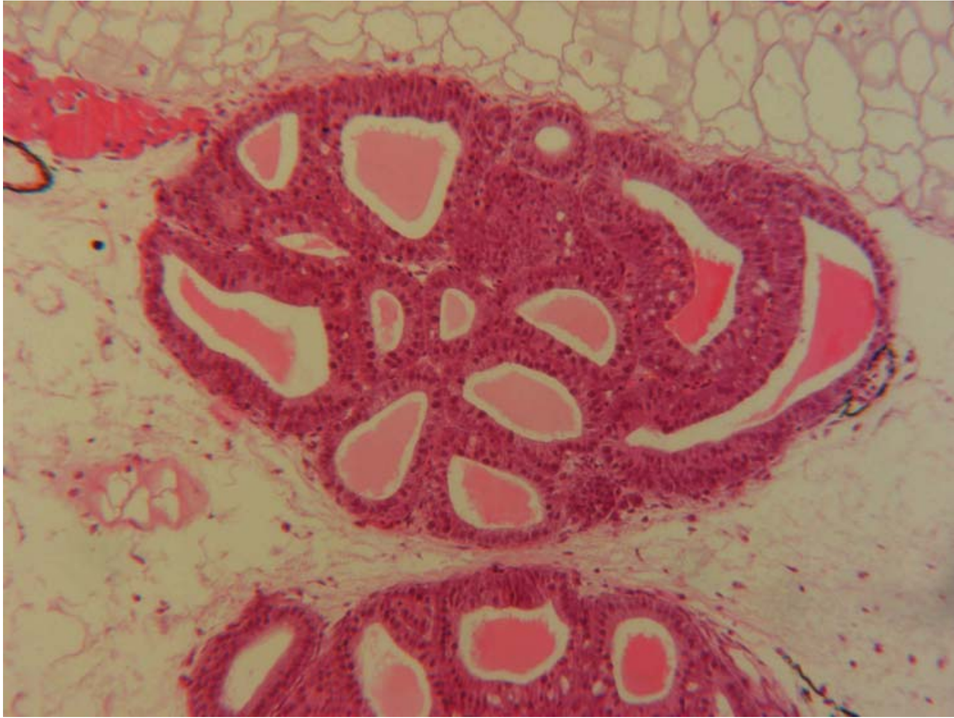


(A) 5.0 mg/L PTU



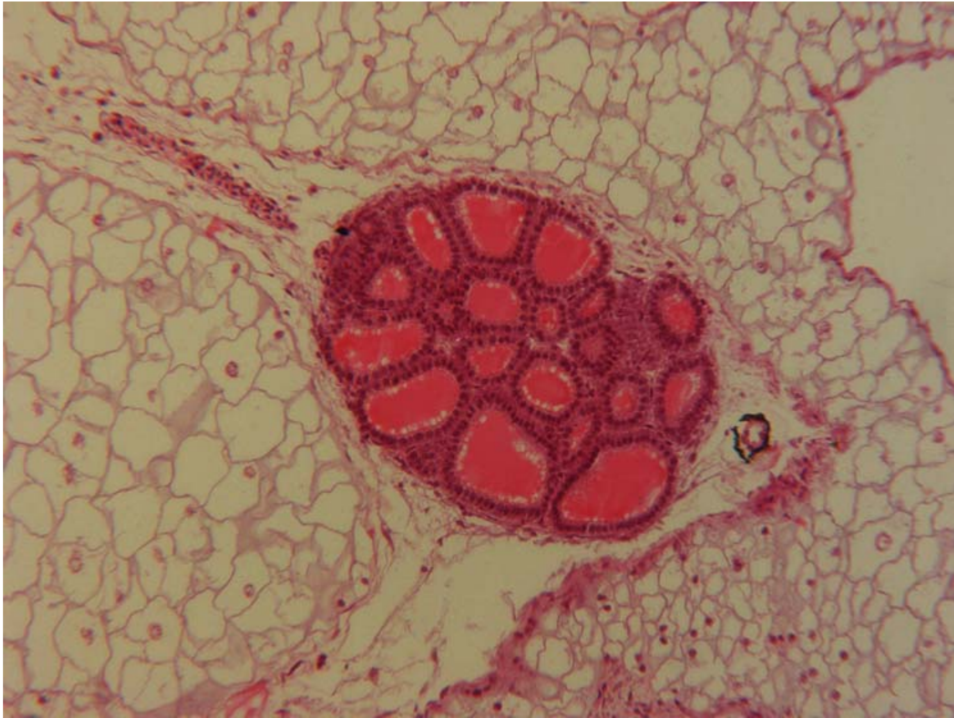
(B) 10 mg/L PTU

Figure 2. Light micrographs (10 x magnification) of histological sections of thyroid glands from *X. laevis* tadpoles after 21 days of exposure to PTU during the stage 51 exposure study. See text for details.

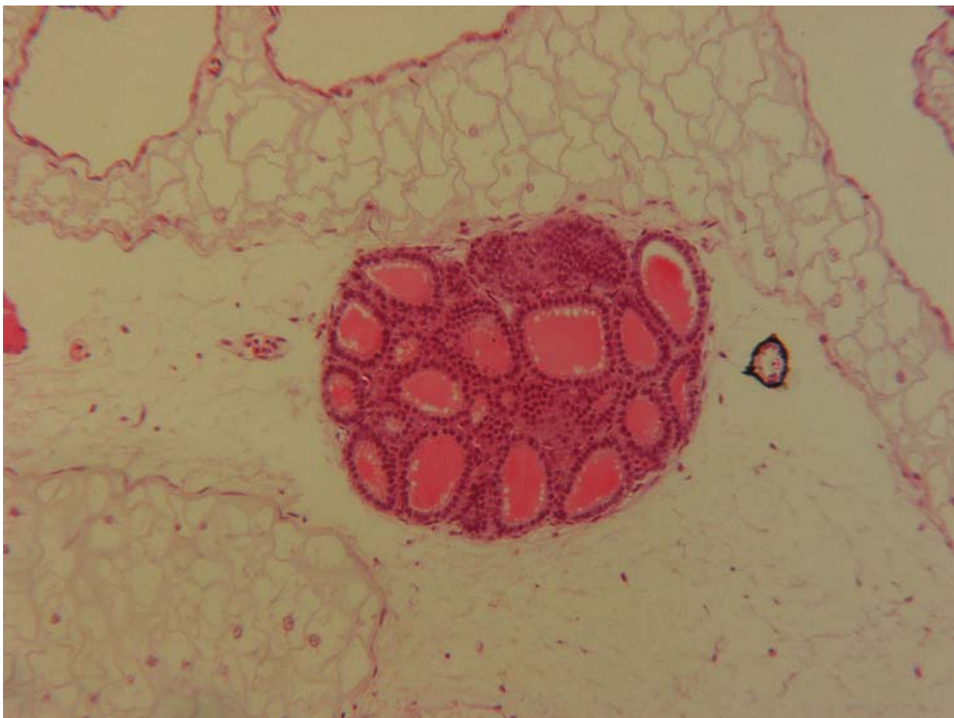


20 mg/L PTU

Figure 3. Light micrograph (10 x magnification) of a histological section of thyroid gland from *X. laevis* tadpole after 21 days of exposure to PTU during the stage 51 exposure study. See text for details.

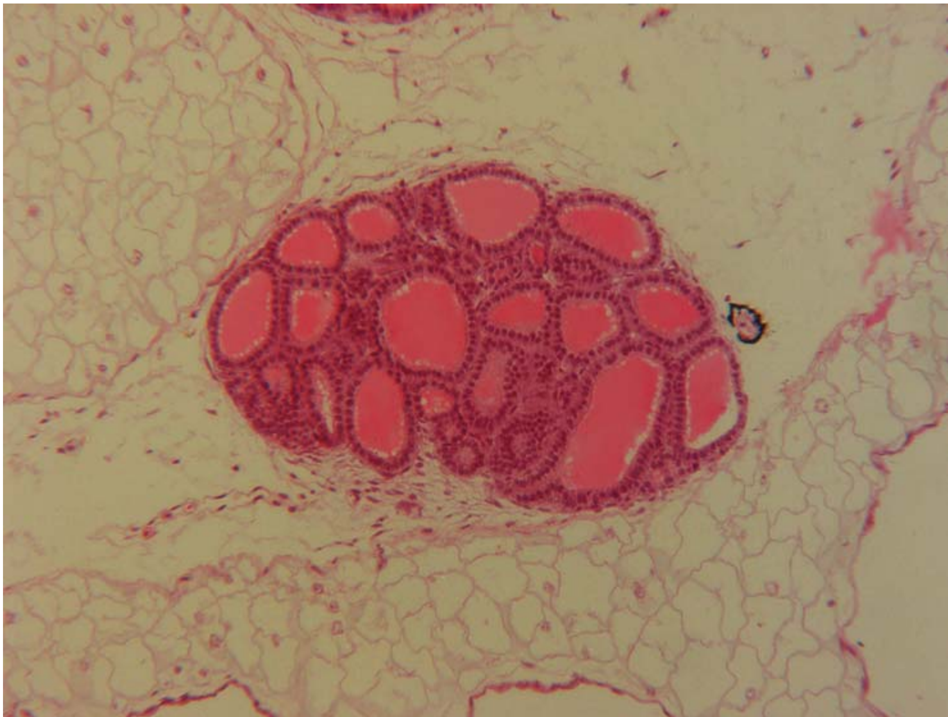


(A) control

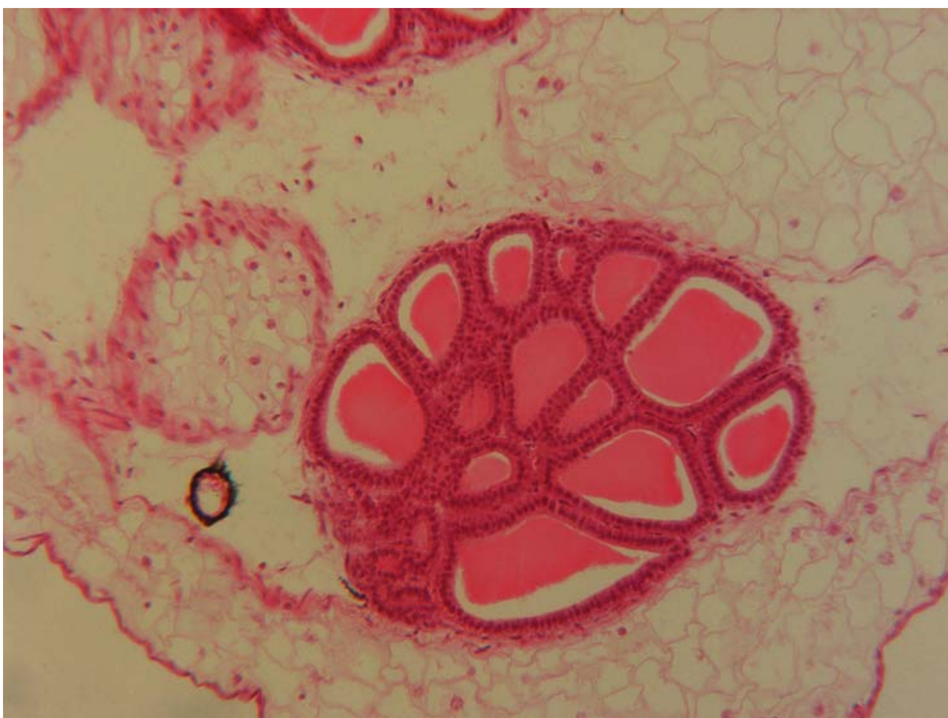


(B) 2.5 mg/L PTU

Figure 4. Light micrographs (10 x magnification) of histological sections of thyroid glands from *X. laevis* tadpoles after 14 days of exposure to PTU during the stage 54 exposure study. See text for details.

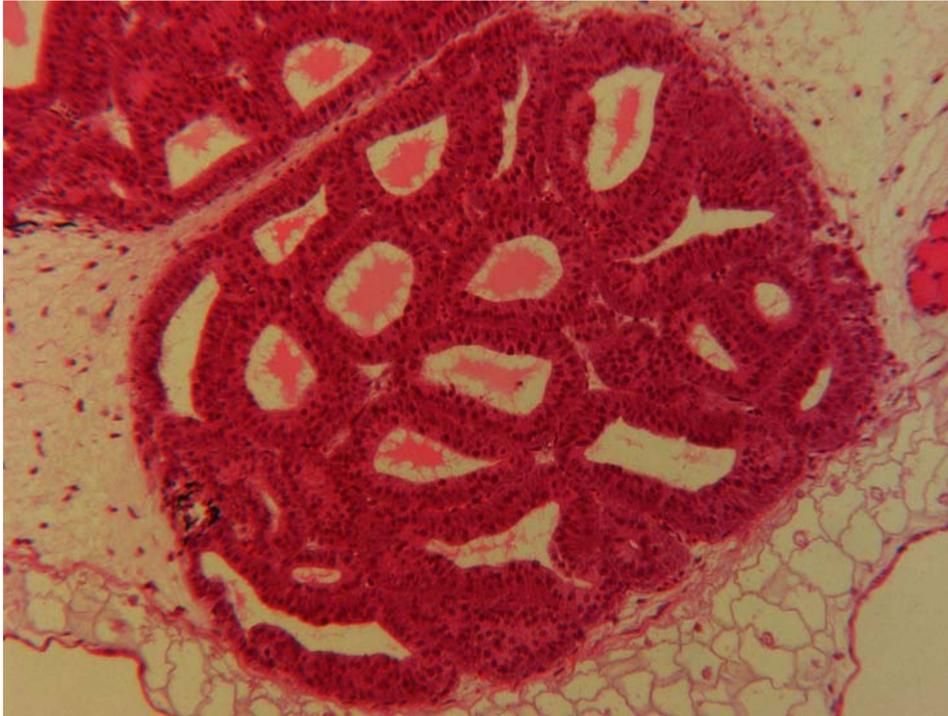


(A) 5.0 mg/L PTU



(B) 10 mg/L PTU

Figure 5. Light micrographs (10 x magnification) of histological sections of thyroid glands from *X. laevis* tadpoles after 14 days of exposure to PTU during the stage 54 exposure study. See text for details.



10 mg/L PTU

Figure 6. Light micrograph (10 x magnification) of a histological section of thyroid gland from *X. laevis* tadpole after 14 days of exposure to PTU during the stage 54 exposure study. See text for details.

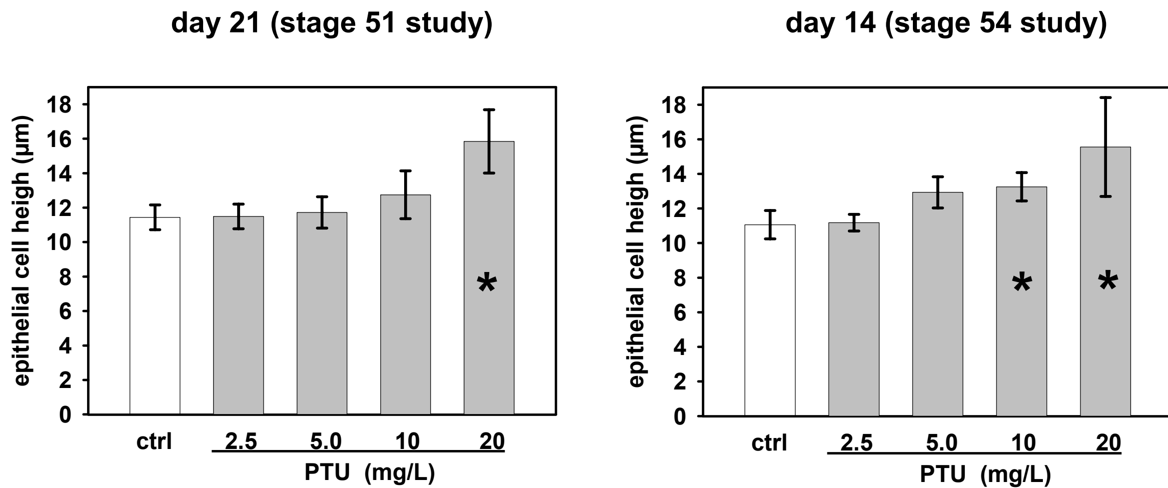


Figure 7. Effects of PTU exposure on epithelial cell height in thyroid glands of *X. laevis* tadpoles. Columns and bars represent mean values ± SD of 10 animals per treatment group. Significant differences from the control group (ctrl) are marked by asterisks (* $p<0.05$; Dunnett's test)

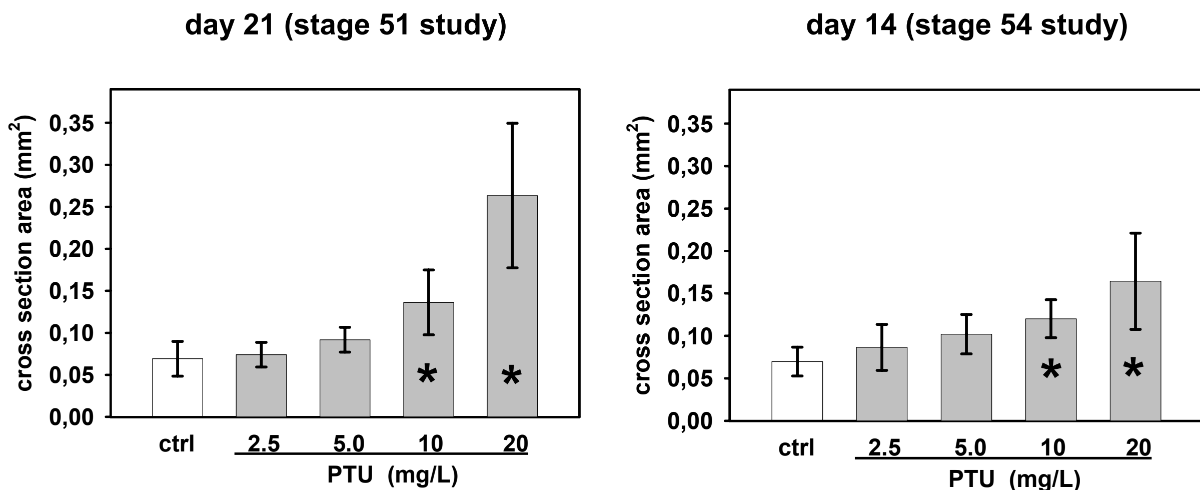
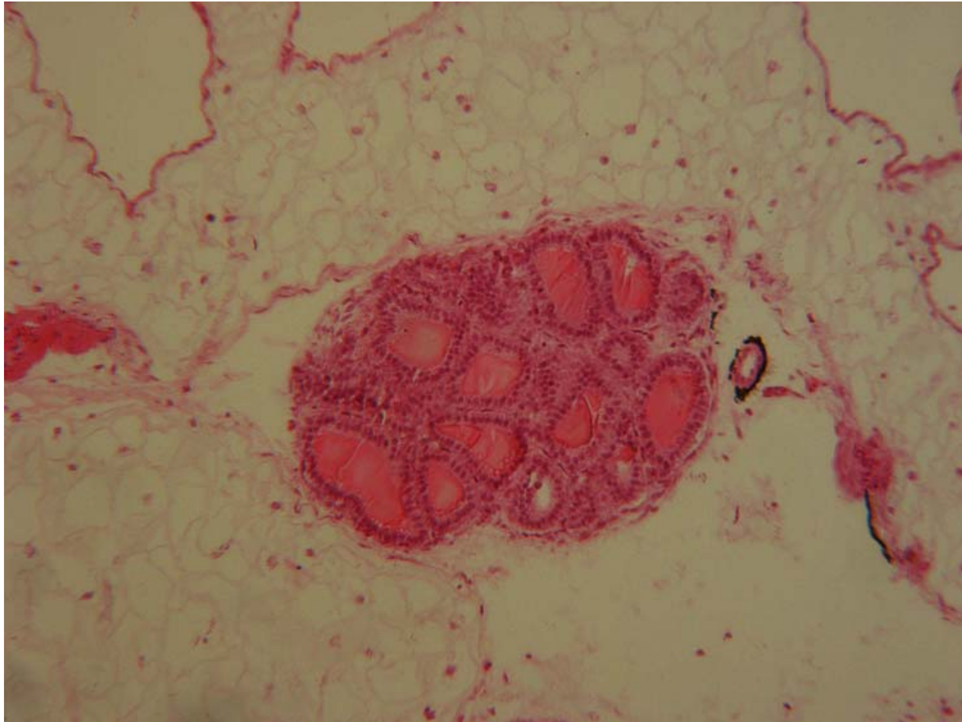


Figure 8. Effects of PTU exposure on maximum cross section area of thyroid glands of *X. laevis* tadpoles. Columns and bars represent mean values ± SD of 10 animals per treatment group. Significant differences from the control group (ctrl) are marked by asterisks (* $p<0.05$; Dunnett's test)

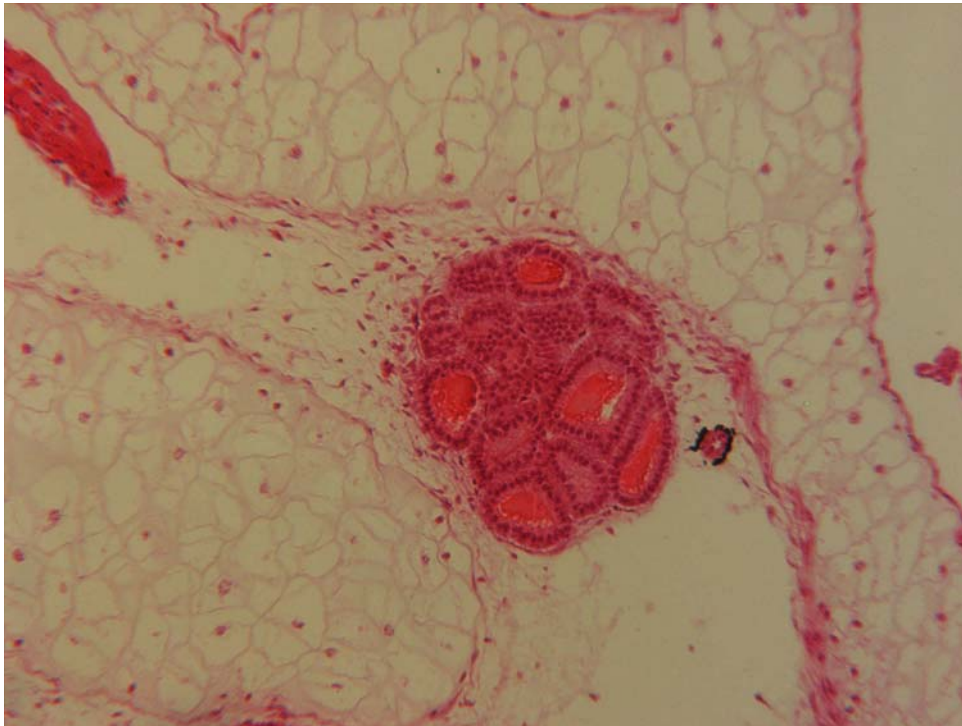
Effects of T4 Exposure on Thyroid Histology

In the stage 51 and stage 54 exposure studies with T4, the light histological appearance of thyroid glands from tadpoles exposed to T4 concentrations of 0.25, 0.5 and 1.0 µg/L did not differ from the control group (Figures 9A, 9B, 10A and 10B). A single layer of cuboidal follicular epithelial cells was present in glands from these treatment groups. The colloid showed a homogeneous tinctoral quality and peripheral vacuolation of the colloid was minimal. In tadpoles exposed to the highest T4 concentration (2.0 µg/L), a slight increase in the degree of peripheral vacuolation of the colloid was observed (Figure 11). Furthermore, an increased prevalence of follicles lined by columnar epithelial cells was noticed. These changes were observed in both exposure studies with T4.

A quantitative analysis of epithelial cell heights and maximum cross section area of the right thyroid lobe is currently in progress for tissue samples from both T4 exposure studies. Results from measurements of epithelial cell heights are shown in Figure 15. Statistically significant increases in epithelial cell height were detected at 2.0 µg/L (stage 51 exposure study) and 1.0 and 2.0 µg/L T4 (stage 54 exposure study).

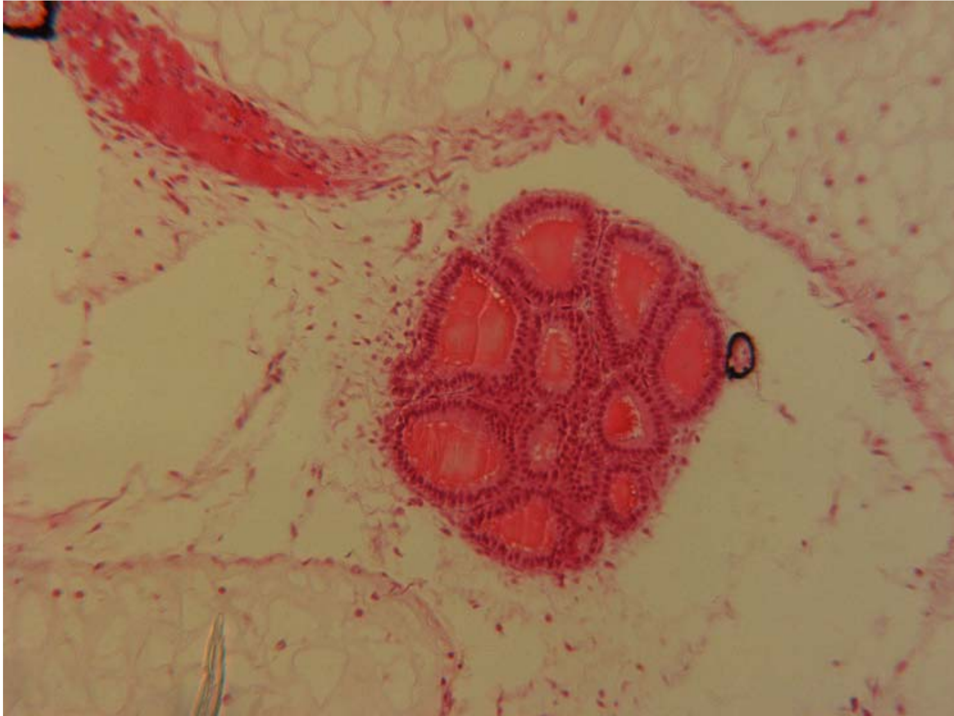


(A) control

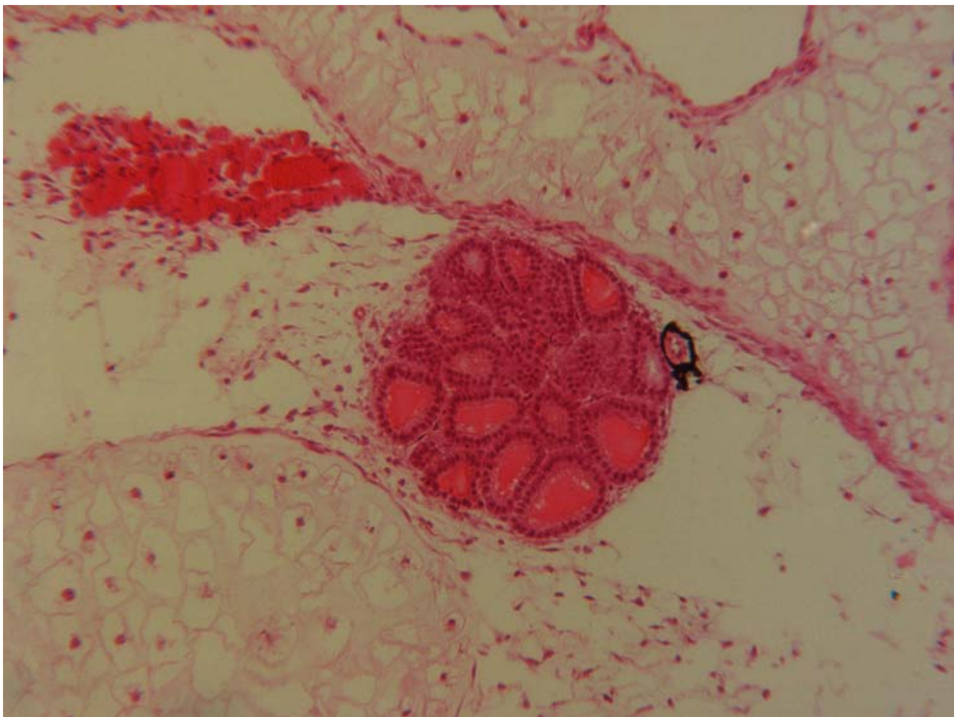


(B) 0.25 µg/L T4

Figure 9. Light micrographs (10 x magnification) of histological sections of thyroid glands from *X. laevis* tadpoles after 21 days of exposure to T4 during the stage 51 exposure study. See text for details.

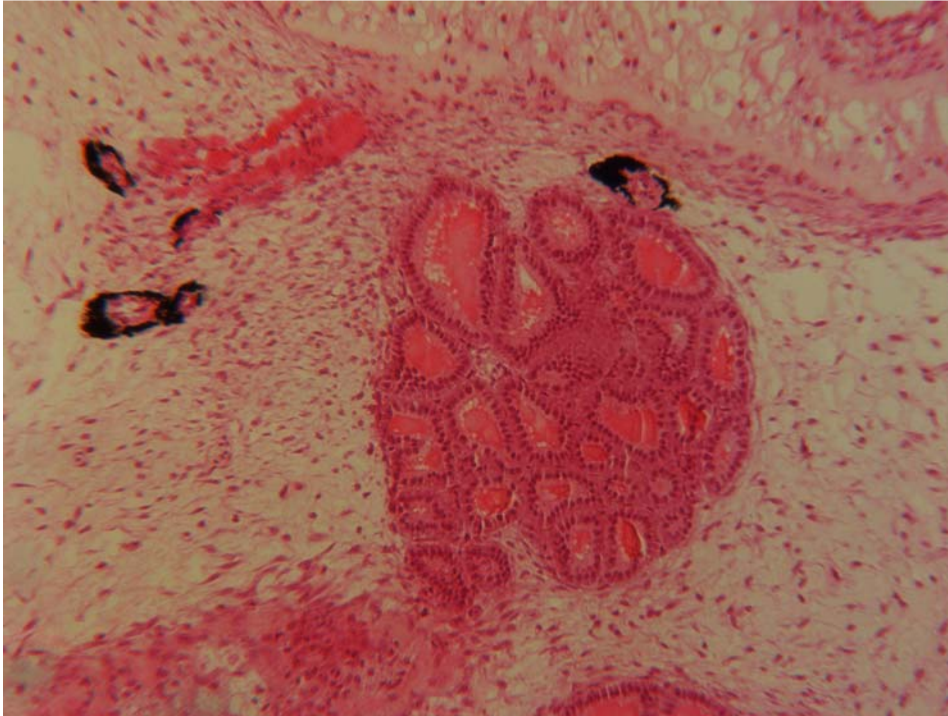


(A) 0.5 μg/L T4



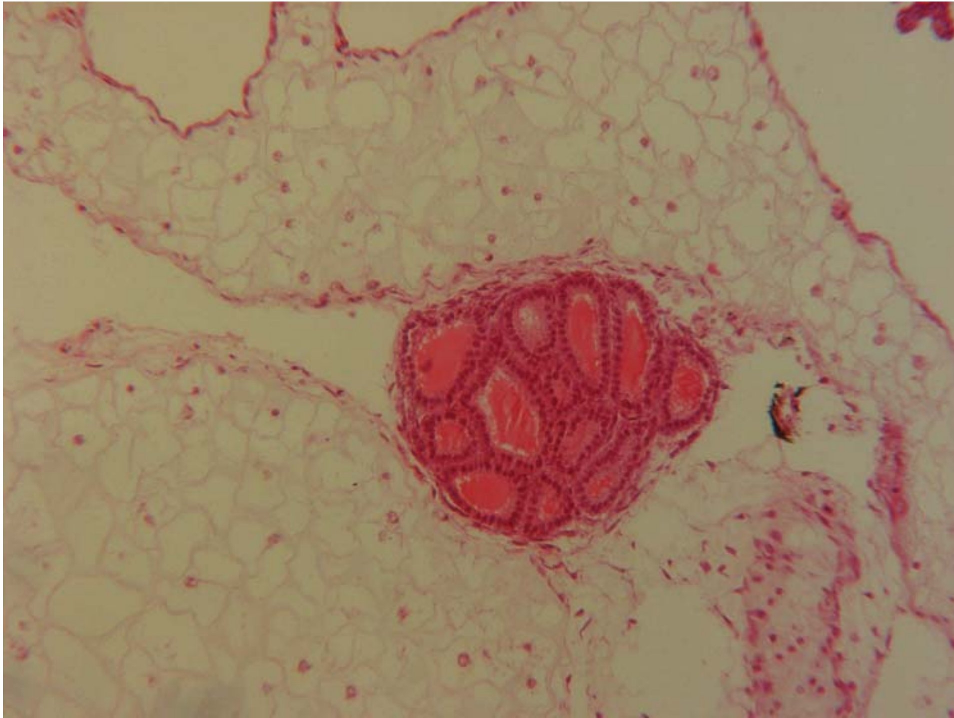
(B) 1.0 μg/L T4

Figure 10. Light micrographs (10 x magnification) of histological sections of thyroid glands from *X. laevis* tadpoles after 21 days of exposure to T4 during the stage 51 exposure study. See text for details.

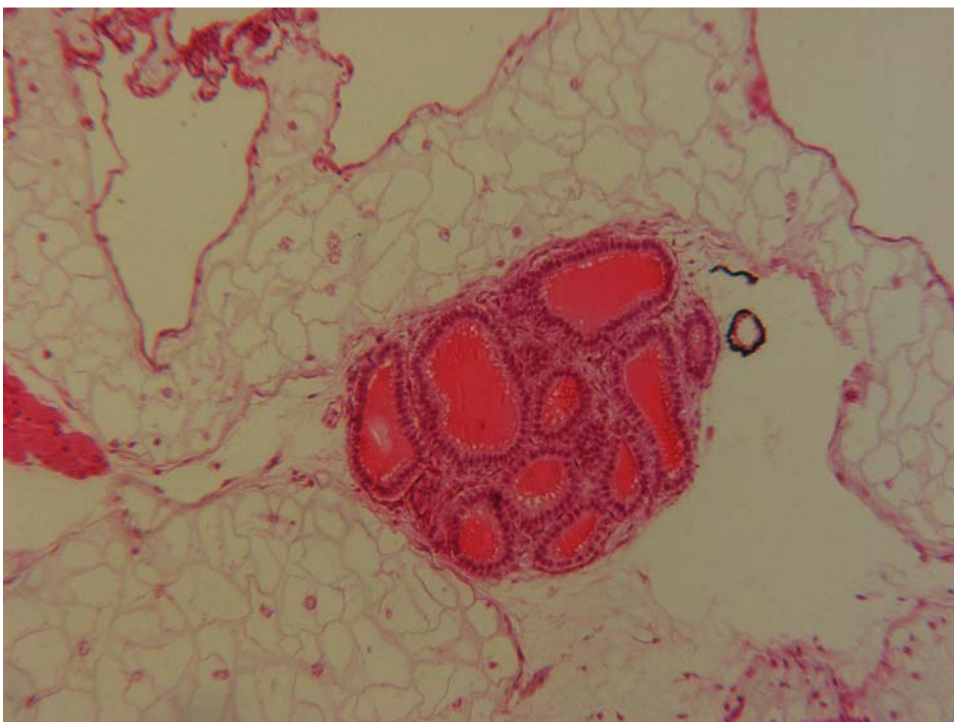


2.0 µg/L T4

Figure 11. Light micrograph (10 x magnification) of a histological section of thyroid gland from *X. laevis* tadpole after 21 days of exposure to T4 during the stage 51 exposure study. See text for details.

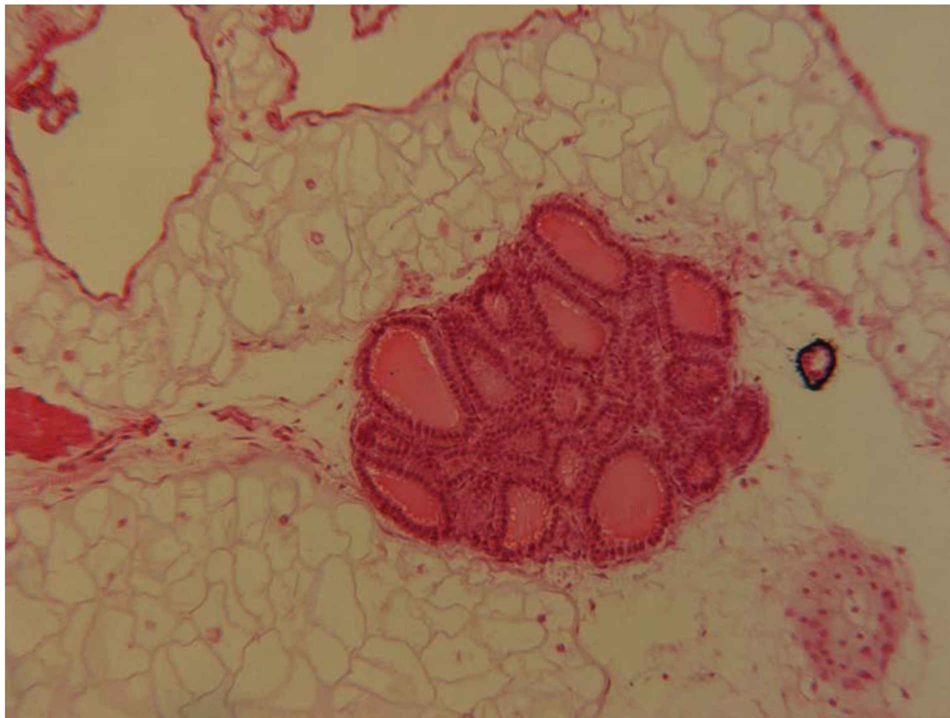


(A) control

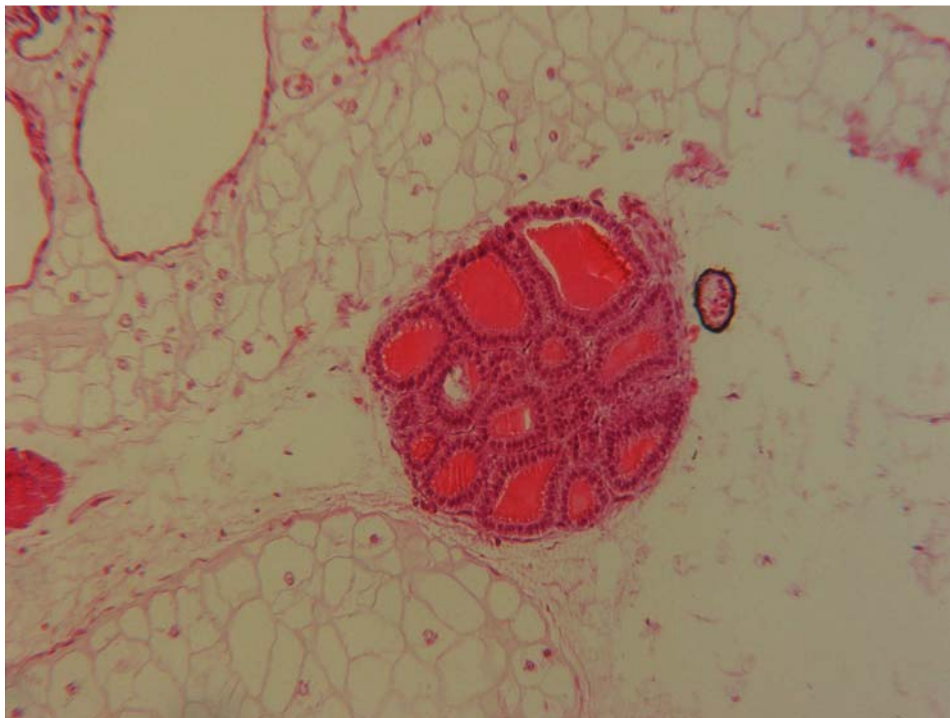


(B) 0.25 µg/L T4

Figure 12. Light micrographs (10 x magnification) of histological sections of thyroid glands from *X. laevis* tadpoles after 14 days of exposure to T4 during the stage 54 exposure study. See text for details.

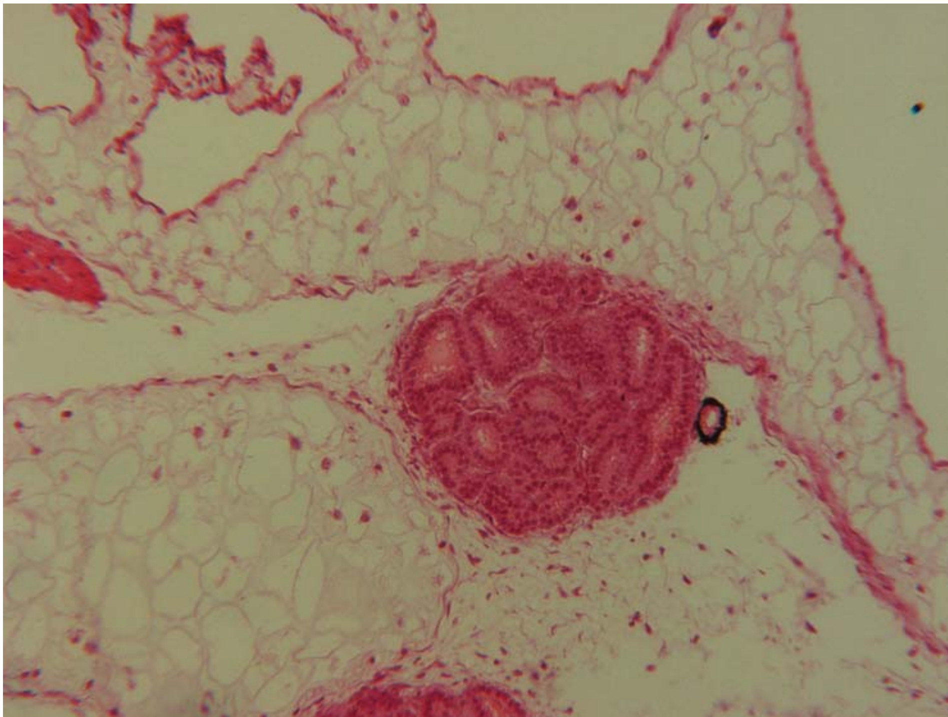


(A) 0.5 µg/L T4



(B) 1.0 µg/L T4

Figure 13. Light micrographs (10 x magnification) of histological sections of thyroid glands from *X. laevis* tadpoles after 14 days of exposure to T4 during the stage 54 exposure study. See text for details.



2.0 µg/L T4

Figure 14. Light micrograph (10 x magnification) of a histological section of thyroid gland from *X. laevis* tadpole after 14 days of exposure to T4 during the stage 54 exposure study. See text for details.

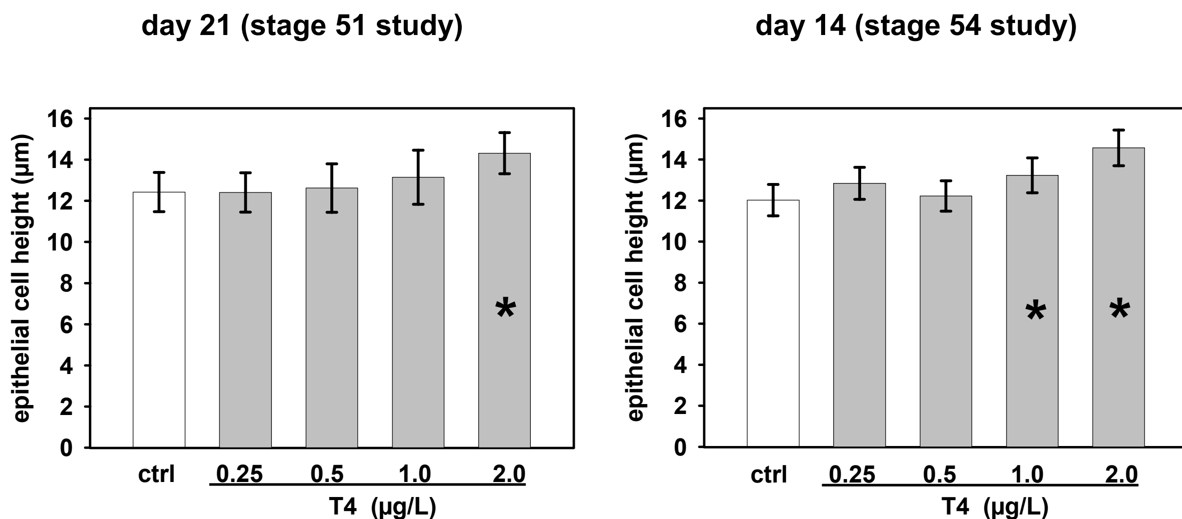


Figure 15. Effects of T4 exposure on epithelial cell height in thyroid glands of *X. laevis* tadpoles. Columns and bars represent mean values ± SD of 10 animals per treatment group. Significant differences from the control group (ctrl) are marked by asterisks (* $p<0.05$; Dunnett's test)

Proposal for Phase 1 of the Validation of the Amphibian Metamorphosis Assay
-rev.1, 6th October 2003-

TABLE OF CONTENTS

PURPOSE OF THE AMPHIBIAN METAMORPHOSIS ASSAY AND OBJECTIVES 2

Purpose and major characteristics of the Xenopus Metamorphosis Assay 2

Objective of the Validation Study 3

Key goals of Phase 1 3

ACTION PLAN AND TIMEFRAME 3

ANNEX 1: Standard operating procedures of the protocol for the Amphibian Metamorphosis Assay 5

ANNEX 2: Glossary 6

ANNEX 3: Schematic of experimental design for the Amphibian Metamorphosis Assay 7

ANNEX 4: Synopsis of test conditions applied in the three participating laboratories in Phase 1 of the Validation of the Amphibian Metamorphosis Assay 8

Proposal for Phase 1 of the Validation of the Amphibian Metamorphosis Assay

PURPOSE OF THE AMPHIBIAN METAMORPHOSIS ASSAY AND OBJECTIVES

Purpose and major characteristics of the Xenopus Metamorphosis Assay

1. The Xenopus Metamorphosis Assay is an *in vivo* screening assay for the identification of chemicals which interfere with the thyroid axis in *Xenopus laevis* exposed during a critical period of the metamorphosis (i.e. late premetamorphosis and/or prometamorphosis).

2. The development of this assay was supported by the EDTA Task Force for its potential to serve as a model for thyroid disruption in Vertebrates. Homology of genes exists between Amphibians and higher Vertebrates which allows to think that results obtained in this assay could be extrapolated to other organisms. At least two Detailed Review Papers are also being developed to document this assumption and support development of an amphibian metamorphosis assay. A major advantage of Xenopus among Amphibians is its well characterised development, especially during metamorphosis (staging criteria of Nieuwkoop and Faber), which makes it amenable to a quantitative assessment of changes observed.

3. The development of this proposal was made possible thanks to previous works in Germany, Japan and the United States. These works especially focused on the finding of a minimal but suitable exposure period, on the selection of good endpoints to measure thyroid effects in tadpoles, on the optimisation of rearing conditions and on suitable test species. A major initial piece of work was undertaken by Germany; it consisted in the ring-testing of a different version of a Xenopus Metamorphosis Assay (XEMA). Results obtained in the ring-test constitute the basis of the current proposal, with substantial input brought by study results from the USEPA on a similar assay.

4. Discussions held recently at the first OECD Expert Meeting on Amphibian Testing (June 2003, Duluth (US)) addressed the differences of opinions and experts found an area of agreement on most of the previously unresolved issues. They proposed to conduct a comparison study, as a first step in the Validation process.

5. In Phase 1 of the Validation, the assay will be initiated at two different stages of metamorphosis: stage 51 and stage 54. Thus tadpoles are exposed for 21 days (from stage 51 to stage 58-59) and for 14 days (from stage 54 to stage 58-59). The assay is terminated when animals reach stage 58-59 of the metamorphosis (staging criteria of Nieuwkoop and Faber). Mortality is checked daily. Development stage, body length, hind limb length are measured at weekly intervals. Histology of the thyroid gland is evaluated at termination of the assay on day 21 or on day 14.

6. Chemicals chosen for Phase 1 are T4 or Thyroxine (CAS NR. 300-30-1) as a thyroid agonist and PTU or Propylthiouracil (CAS NR. 51-52-5) as a thyroid antagonist. The following test concentrations will be used for PTU: 2.5, 5, 10 and 20 mg/l, and for T4: 0.25, 0.5, 1.0 and 2.0 µg/l.

Objective of the Validation Study

7. The objective of the Validation study is to experimentally establish the relevance of the assay by assessing its sensitivity to detect weak and strong (anti-) thyroid compounds and to assess the reproducibility of the assay by comparing results obtained by a variety of laboratories in diverse geographical locations.

8. Phase 1 of the Validation will focus on the demonstration of the relevance of the assay, i.e. its ability to detect (anti-)thyroid compounds through measurement of the selected endpoints.

9. The comparison of the 14-d and the 21-d protocols will help determine the optimum development stage at initiation of exposure, and thus the duration of the assay.

Key goals of Phase 1

- Obtain sufficient information on the relevance of endpoints selected for the assay and their ability to respond in a consistent way to Thyroxine and Propylthioruacil.
- Obtain information on the adequate development stage at initiation of exposure (stage 51 or stage 54);
- Obtain information on the optimal duration of exposure (14 days or 21 days);
- Optimise the rearing conditions in order to minimise the variation in the endpoints measured (e.g. larval density);
- Standardise the thyroid histology techniques and evaluation in order to validate the endpoint;
- Compare the performance of flow-through and static renewal regimes to determine whether the protocol can accommodate both or if it should be restricted to one water regime.

ACTION PLAN AND TIMEFRAME

10. In Phase 1 of the Validation, *Xenopus laevis* has been selected as the only species for which the historical database is sufficiently comprehensive to enable proper comparison of it with results obtained in the assay. Three laboratories, preferably with previous experience in the assay, are needed in this Phase 1. These laboratories have been provisionally identified during the OECD Meeting of the *Ad hoc* Expert Group on Amphibian Testing in June 2003, pending confirmation of funding availability. These laboratories are located in Germany, Japan and the United States.

11. Germany prepared the standard operating procedures of the protocol of the assay (Annex 1). These standard operating procedures were reviewed by other experts of the OECD Amphibian Expert Group and agreed upon. However, differences of practice will persist for some of the test conditions: the exposure regime (static renewal in one laboratory versus flow through in the other two), the larval density (number of tadpoles/volume of test solution) and the feeding regime and ration. All test conditions were previously used in two of the laboratories and gave comparable growth and development rates. Therefore, amphibian experts estimated that these parameters should not influence the main question to be addressed during Phase 1 which is the optimum

development stage at initiation of exposure. This assumption will be checked in Phase 1 and the outcome will determine whether the protocol developed will need to be restricted to just one exposure system or allowed to accommodate both the static-renewal and the flow-through system.

12. The three laboratories participating in Phase 1 will do both the 21-day exposure (stating at stage 51) and the 14-day exposure (starting at stage 54) using both chemicals: thyroxine and propylthiorucail.

13. The present proposal first needs approval from the VMG-eco before starting experimental work. This should be done through a written procedure, and if need be, a teleconference of the VMG-eco will be organised towards the end of September to agree on remaining issues.

14. Once agreement is reached with the VMG-eco, experimental work should start, co-ordinated by a lead laboratory among the three participating ones. It is expected that, by the end of 2003 or early 2004 Phase 1 will be completed. Results will be presented to the VMG-eco for discussion at its next meeting in Spring 2004.

15. The main outcome should be an agreement on whether a 14-day exposure (from stage 54 to stage 58-59) is sufficient or a 21-day exposure (from stage 51 to stage 58-59) is necessary to detect thyroid related effects. Phase 2 could then consist in a multi-chemical analysis in three or more laboratories, provided all other questions and issues are solved, in particular as regards standardisation of endpoint measurement and acceptable test conditions.

16. A glossary of important terms was prepared in Annex 2. A schematic of the experimental design of the Amphibian Metamorphosis Assay is in Annex 3 and a synopsis of assay conditions applied in the three laboratories can be found in Annex 4.

17. The supply of test substances will be co-ordinated through a chemical repository, Battelle, located in the United States. A standard order form has been prepared and laboratories will be requested to use it and submit it to the indicated contact point. Test substance from an identical lot number will be used by the three participating laboratories to ensure uniformity of compounds tested.

18. The lead laboratory will be responsible for the collection of test results from the other two laboratories on standardised Excel spreadsheets, for the statistical analysis and for the preparation of a draft report for Phase 1.

ANNEX 1: Standard operating procedures of the protocol for the Amphibian Metamorphosis Assay

(*See separate file: SOP Comparison study.pdf)*

ANNEX 2: Glossary

Larval density: number of tadpoles per volume of test solution.

Biological rate: gives an indication of how much of the chemical reaches the tadpole and is expressed in mg tadpole/mg test substance.

Exposure regime: type of exposure system used during the experiment. The semi-static system is referred to when the animals are transferred punctually (e.g. three times a week) to a freshly prepared test solution. The flow-through system means that a fixed flow of fresh test solution (e.g. 25 ml/min) is continuously dispensed to the test chamber; a fixed flow of existing test solution is also removed constantly; the whole system enables the quality of water to remain acceptable during the exposure period without having to move the animal.

ANNEX 3: Schematic of experimental design for the Amphibian Metamorphosis Assay

Propylthiouracil (PTU)	Thyroxine (T4)
Dilution water control	Dilution water control
2.5 mg/l	0.25 µg/l
5 mg/l	0.5 µg/l
10 mg/l	1.0 µg/l
20 mg/l	2.0 µg/l

FOR BOTH CHEMICALS AND BOTH REPLICATES:

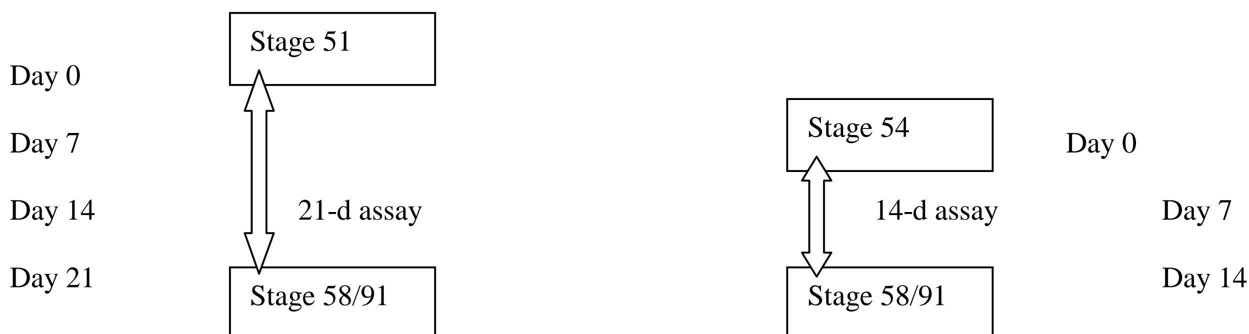
Measurements/observations made for all treatment levels and the control:

Development stage (Nieuwkoop and Faber)
Whole body length
} d-0, d-7, d-14, (d-21) on all 20 individuals/replicate

Hind limb length: d-7, d-14, (d-21) on all 20 individuals/replicate

Thyroid histology: at termination of the assay at d-14 or d-21 on 5 individuals per replicate.

Duration:



	Replicate	Control	Conc. 1	Conc. 2	Conc. 3	Conc. 4
PTU	A	20	20	20	20	20
	B	20	20	20	20	20
T4	A	20	20	20	20	20
	B	20	20	20	20	20

Number of tadpoles used in the study. Endpoints are measured on tadpoles weekly; tadpoles are then placed back in the test chamber each time.

ANNEX 4: Synopsis of test conditions applied in the three participating laboratories in Phase 1 of the Validation of the Amphibian Metamorphosis Assay

<table>
<tr><th colspan="2"></th><th colspan="2">Germany</th><th colspan="2">Japan</th><th colspan="2">United States</th></tr>
<tr><td colspan="2">Animal</td><td colspan="2">Xenopus laevis</td><td colspan="2">Xenopus laevis</td><td colspan="2">Xenopus laevis</td></tr>
<tr><td colspan="2" rowspan="2">Exposure period</td><td colspan="2">Exposure from stage 51: 21-days</td><td colspan="2">Exposure from stage 51: 21-days</td><td colspan="2">Exposure from stage 51: 21-days</td></tr>
<tr><td colspan="2">Exposure from stage 54: 14-days</td><td colspan="2">Exposure from stage 54: 14-days</td><td colspan="2">Exposure from stage 54: 14-days</td></tr>
<tr><td rowspan="2">Concentration of test substance</td><td>PTU</td><td colspan="2">2.5, 5, 10, 20 mg/l</td><td colspan="2">2.5, 5, 10, 20 mg/l</td><td colspan="2">2.5, 5, 10, 20 mg/l</td></tr>
<tr><td>T4</td><td colspan="2">0.25, 0.5, 1.0, 2.0 µg/l</td><td colspan="2">0.25, 0.5, 1.0, 2.0 µg/l</td><td colspan="2">0.25, 0.5, 1.0, 2.0 µg/l</td></tr>
<tr><td colspan="2">Control</td><td colspan="2">1 dilution water control</td><td colspan="2">1 dilution water control</td><td colspan="2">1 dilution water control</td></tr>
<tr><td colspan="2">Exposure regime</td><td colspan="2">Semi-static</td><td colspan="2">Flow-through (25 ml/min)</td><td colspan="2">Flow-through (25 ml/min)</td></tr>
<tr><td colspan="2">Larval density</td><td colspan="2">20 tadpoles/10 l tank</td><td colspan="2">20 tadpoles/ 4 l tank</td><td colspan="2">20 tadpoles/ 4 l tank</td></tr>
<tr><td colspan="2">Replication</td><td colspan="2">2 replicates for each test substance</td><td colspan="2">2 replicates for each test substance</td><td colspan="2">2 replicates for each test substance</td></tr>
<tr><td colspan="2" rowspan="5">Endpoints and determination days</td><td>Development stage</td><td rowspan="2">Day 0, 7, 14, 21</td><td>Development stage</td><td rowspan="2">Day 0, 7, 14, 21</td><td>Development stage</td><td rowspan="2">Day 0, 7, 14, 21</td></tr>
<tr><td>Whole body length</td><td>Whole body length</td><td>Whole body length</td></tr>
<tr><td>Hind limb length</td><td>Day 7, 14, 21</td><td>Hind limb length</td><td>Day 7, 14, 21</td><td>Hind limb length</td><td>Day 7, 14, 21</td></tr>
<tr><td>Wet weight</td><td>Day 0, final day</td><td>Wet weight</td><td>Day 0, final day</td><td>Wet weight</td><td>Day 0, final day</td></tr>
<tr><td>Mortality</td><td>Daily observation</td><td>Mortality</td><td>Daily observation</td><td>Mortality</td><td>Daily observation</td></tr>
<tr><td colspan="2">Acceptable mortality rate</td><td colspan="2"><5% in control</td><td colspan="2"><5% in control</td><td colspan="2"><5% in control</td></tr>
<tr><td rowspan="3">Fixation</td><td>Day</td><td colspan="2">Final day of exposure</td><td colspan="2">Final day of exposure</td><td colspan="2">Final day of exposure</td></tr>
<tr><td>Region</td><td colspan="2">Lower jaw; specific tissues (brain, tail) for gene expression analysis</td><td colspan="2">Whole body</td><td colspan="2">Whole body</td></tr>
<tr><td>Fixation fluid</td><td colspan="2">Bouin’s fluid or liquid nitrogen</td><td colspan="2">Bouin’s fluid</td><td colspan="2">Bouin’s fluid</td></tr>
<tr><td rowspan="3">Feeding</td><td>Food</td><td colspan="2">Sera micron</td><td colspan="2">Sera micron</td><td colspan="2">Mixture of TetraFin, Spirulina algae discs, Silver cup trout Starter, along with live brine shrimp</td></tr>
<tr><td>Amount</td><td colspan="2">300 mg/tank</td><td colspan="2">600 mg/tank</td><td colspan="2">To be determined</td></tr>
<tr><td>Frequency</td><td colspan="2">Once a day</td><td colspan="2">Twice/day on weekdays, once/day on weekends</td><td colspan="2">Twice/day on weekdays, once/day on weekends</td></tr>
<tr><td rowspan="2">Test medium</td><td>Component</td><td colspan="2">Commercial salt mixture Tropic Marine Meersalz to deionized distilled water</td><td colspan="2">Activated carbon processed water.</td><td colspan="2">Lake Superior Water (LSW)</td></tr>
<tr><td>Concentration</td><td colspan="2">0.025%</td><td colspan="2">0.025%</td><td colspan="2">-</td></tr>
</table>

		Germany	Japan	United States
Tank (L*W*H)		30*20*20 cm	22.5*14*16.5 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
Dissolved Oxygen concentration		Above 80% of the air saturation value	Above 80% of the air saturation value	To be confirmed
Room temperature		19-22ºC	19-22ºC	To be confirmed

XEMA - Standard Operation Procedure

Standard Operation Procedure

for the Conduct of the

Xenopus Metamorphosis Assay (XEMA)

(Comparison Study 2003)

Werner Kloas, Robert Opitz and Ilka Lutz

Department of Inland Fisheries, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin (Germany)

July 2003

Xenopus Metamorphosis Assay (XEMA)

1 Test Organism

The XEMA test uses tadpoles of the South African clawed frog *Xenopus laevis*. Other species of *Xenopus* can not be used. The test should be performed with animals bred in the laboratory and tadpoles purchased from commercial sources should not be used to conduct the XEMA test. The tadpoles used in this test should be derived from a healthy brood stock.

2 Objective of the Comparison Study

The primary objective of the comparison study is the comparative evaluation of the sensitivity of two different exposure scenarios to detect stimulating and inhibiting effects of chemicals on *X. laevis* metamorphosis. Exposure is initiated with *X. laevis* tadpoles at developmental stages 51 and 54, respectively. Exposure of stage 51 tadpoles is continued for a total of 21 days and exposure of stage 54 tadpoles is continued for a total of 14 days. Tadpoles are exposed to 4 different concentrations of the test substance (*n*= 2 replicates per concentration) and a control group (*n*= 2 replicates). Larval density is 20 tadpoles per 10 L treatment tank. All exposure experiments use an aqueous route of exposure. The test uses a static renewal exposure regime and test solutions are exchanged completely three times a week (Monday, Wednesday, Friday). The main endpoints in this comparison study are developmental stage, hind limb length and whole body length which are determined for each tadpole at exposure days 0, 7, 14 and 21. In addition, mortality is checked on a daily basis. Effects of the test chemicals are assessed by comparison with the control treatment.

3 Biological Endpoints

3.1 Developmental stage

The primary endpoint to assess possible effects of the test chemical on metamorphic development and thus, on thyroid function is the developmental stage of the tadpoles. The developmental stage of *X. laevis* tadpoles can easily be determined by using the staging criteria of Nieuwkoop and Faber (1994). The XEMA test is designed to detect both anti-thyroidal and thyroid hormone-mimicking effects of test substances. Anti-thyroidal compounds are expected to inhibit or slow down the rate of metamorphic development, causing lower developmental stages compared to untreated controls. On the other hand, thyroid hormone-mimicking substances cause enhanced metamorphic development leading to higher developmental stages compared to untreated controls.

3.2 Body Length and Hind Limb Length

Measurement of whole body length provides information about the growth rate of tadpoles. Whole body length is measured from the tip of the snout to the tail end (Figure 1). Differentiation and growth of the hind limbs is under control of thyroid hormones. Therefore, hind limb length is measured as an additional endpoint in order to evaluate the suitability of this endpoint to detect effects on the thyroid axis.

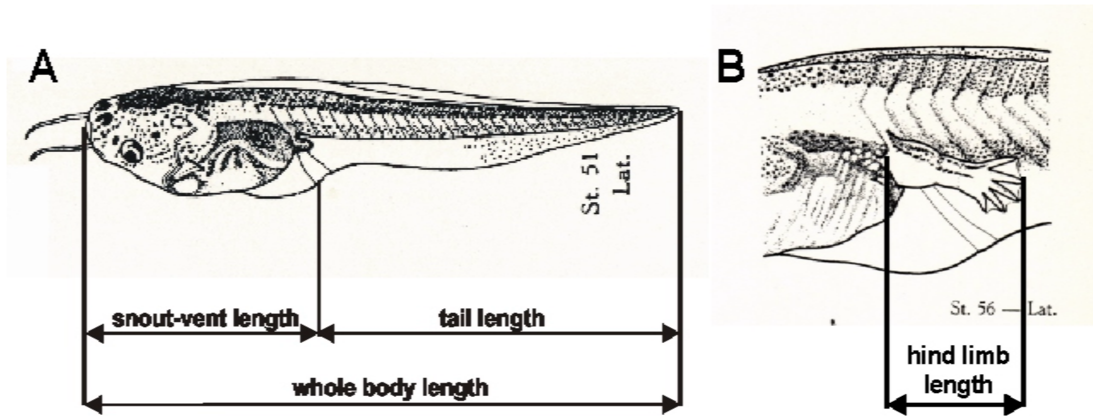


Figure 1 Measurement of whole body length (A) and hind limb length (B) in *Xenopus laevis* tadpoles (Nieuwkoop and Faber, 1994).

3.3 Additional Observations

Mortality in the test tanks is checked daily and the number of dead animals is recorded. Dead animals must be immediately removed from the test tanks. Under the conditions described here, mortality rates are normally less than 5 % in the control group. Increased mortality rates may indicate inappropriate test conditions or toxic effects of the test materials.

4 Equipment

The following equipment is needed:

- 50 L tanks (approx. 60 x 40 x 35 cm) with sturdy lids and spawn grates which fit into the tanks
- twenty 10 L glass aquaria (approx. 30 x 20 x 20 cm) with glass covers
- water bathes that accomodate for 20 glass aquaria
- air pump and airstones
- aquarium heaters (adjustable to 22±1°C)
- thermometer (preferably a recording thermometer measuring temperature minima and maxima)
- binocular dissection microscope
- transparent plastic chamber of an approximate size of 80 x 30 x 15 mm
- pH meter
- oxygen meter
- devices to scoop tadpoles out of the tanks (preferably a small strainer with a rather rigid mesh)
- different glass vessels (beakers, graduated flasks, graduated cylinders, Erlenmeyer flasks etc.)
- adjustable pipettors, disposable micropipettes (1000 μL), graduated pipettes (5 mL, 10 mL, 20 mL)

5 Breeding and Culture Conditions

5.1 Adult Breeding Animals

Details on holding the brood stock under satisfactory conditions can be found in the literature concerned with the performance of the FETAX test (Bantle, 1995; Bantle *et al.*, 1998) or other publications related to general culture and care of adult *X. laevis* (Dawson *et al.*, 1992; Hilken *et al.*, 1995; Hilken *et al.*, 1997; Kay and Peng, 1992).

5.2 Breeding

Spawning of adult *X. laevis* should be induced approximately 14 days prior to the intended initiation of the test. After fertilization, the exact time period for tadpoles to develop to stage 51 can be variable depending on laboratory conditions. Normally, larvae hatch until day 3 postfertilization (PF) and develop to stage 51 within 14-16 days after fertilization. Spawning of adult *X. laevis* is induced by injecting human chorionic gonadotropin (hCG; Chorionic Gonadotropin from Human Pregnancy Urine, Sigma-Aldrich) into the dorsal lymph sac of male and female adult frogs as described by Bantle *et al.* (1998). Adult frogs should not be bred more frequently than every 2 to 3 months. Breeding males and females should appear healthy and free of abnormalities or injuries. Only adult frogs that have not been treated for diseases during the past 4 weeks should be used for breeding.

The number of eggs per spawn under optimal laboratory conditions generally ranges from 500 to more than 2,000. Under the test conditions used during the comparison study, 400 tadpoles are required for each test trial. Because breeding success can be variable, it is recommended to simultaneously induce breeding in two adult pairs in order to ensure an appropriate number of fertilized eggs and hatched larvae. All tadpoles used in a test should be from the same brood.

For the induction of spawning, a hCG stock solution with a concentration of 200 U/100 µL is freshly prepared immediately before injection. A recommended schedule to induce mating is that two adult males are injected in the morning with 200 µL of hCG stock solution, respectively. After injection, the males are separated and placed in two 50 L plastic tanks filled with 40 L of the test medium at 22°C. These breeding tanks are aerated with airstones and covered with a sturdy lid to keep the animals under low light conditions and to prevent escaping of adult animals. In order to protect the laid eggs from the parent animals, a spawning grate, held a few centimeters off the bottom, is placed in the breeding tanks so that spawned eggs will fall through the mesh to the bottom of the tank where they are not available to the adult animals. In the evening of the same day, approximately 10-12 hours after the first injection, the males receive a second injection of 200 µL hCG stock solution. At the same time, two female adults are injected with 500 µL hCG stock solution, respectively. A hCG-primed male and female are placed together in a breeding tank to go into amplexus. The breeding pairs are left to breed overnight under low light conditions and any disturbance of the breeding pairs should be avoided during this time. Normally, the eggs are laid during the night. The next morning, the adult frogs are removed from the breeding tank.

5.3 Eggs and Tadpoles

Eggs and developing tadpoles are held in a synthetic test medium at 22±1°C and pH 7±0.5 during all phases of development. The synthetic test medium is a reconstituted tap water medium which is formulated by adding 2.5 g of the commercial salt mixture “Tropic Marine Meersalz” (Tagis Aquarium, 63271 Dreieich, Germany, Fon: +49 6103 64036) to 10 L deionized distilled water. This test medium is used as rearing water for eggs and tadpoles during the pre-exposure phase and as dilution water during the exposure phase. All vessels used for rearing of eggs and tadpoles are continuously aerated with airstones in order to maintain dissolved oxygen concentrations above 80 % of the air

saturation value. The light-dark cycle is 12:12 hours and the room temperature should be between 19 and 22°C.

Figure 2 Holding tanks for rearing of eggs and tadpoles during the pre-exposure phase

After spawning, eggs and developing tadpoles are reared in large holding tanks (e.g. 50 L plastic tanks) containing approximately 40 L of test medium (Figure 2). Tadpoles hatch until days 3-4 PF. Feeding of tadpoles starts at days 4-5 PF. During the pre-exposure period, tadpoles are fed daily with the commercial tadpole food sera micron (sera GmbH, D-52518 Heinsberg). The total daily food ration is dependend on the number of tadpoles per tank. The amount of food should be adjusted as needed so that the water becomes clear within 24 hours. Preferably, small portions of sera micron are given to the tadpoles 3-4 times per day.

At day 8 PF, 1000 tadpoles are removed from the large holding tanks and are transferred to smaller rearing tanks (e.g. 10 L glass aquaria) in order to reduce larval density. From day 8 PF onwards, the rearing density should be adjusted to approximately 45-50 tadpoles per 10 L test medium. The rearing tanks are placed in a thermoregulated water bath that is maintained at 22±1°C and each tank is aerated via an air stone (Figure 3). Once the tadpoles are transferred to

the smaller rearing tanks, they are fed 200 mg sera micron per day per tank (preferably, the total amount of 200 mg food per day is divided into 2 food rations given in the morning and afternoon, respectively). At days 11 and 14 PF, the rearing medium in the tanks is changed out completely. Tadpoles are reared under these conditions until they reach developmental stage 51 at days 14-16 PF. When the majority of tadpoles have reached stage 51, 200 larvae at stage 51 meeting the criteria described below (section 6.1.1) are used to initiate the 21-d exposure. The remaining tadpoles are placed in rearing tanks with fresh medium at a larval density of 35 tadpoles per 10 L tank. When the majority of the remaining tadpoles have reached stage 54 (approx. at day 20 PF), 200 larvae at stage 54 meeting the criteria described below (section 6.1.2) are used to initiate the 14-d exposure.

Figure 3 Test tanks (10 L glass aquaria) for exposure of tadpoles

6 Test Initiation and Conduct

6.1 Selection of Experimental Animals

6.1.1 21-d Exposure with Stage 51 Tadpoles

The 21-d exposure study is initiated with stage 51 tadpoles that have been reared at a larval density not exceeding 5 tadpoles/L from day 8 PF onwards. At day 0 of exposure, developmental stage and whole body length are determined for these tadpoles. The test phase is initiated with tadpoles that meet the following criteria:

- developmental stage: 51 (Nieuwkoop and Faber, 1994)
- whole body length: 27 - 32 mm

At developmental stage 51, whole body length of tadpoles normally ranges from 27 mm to 32 mm. Tadpoles that fulfill the criteria mentioned above but exhibit grossly visible malformations or damages (e.g. damage of the tail) are not usable in XEMA.

6.1.2 14-d Exposure with Stage 54 Tadpoles

The 14-d exposure study is initiated with stage 54 tadpoles that have been reared at a larval density not exceeding 3.5 tadpoles/L from day 14 PF onwards. At day 0 of exposure, developmental stage and whole body length are determined for these tadpoles. The test phase is initiated with tadpoles that meet the following criteria:

- developmental stage: 54 (Nieuwkoop and Faber, 1994)

Tadpoles that fulfill the criteria mentioned above but exhibit grossly visible malformations or damages (e.g. damage of the tail) are not usable in XEMA.

6.2 General Exposure Regime

The general experimental design is that tadpoles are exposed to 4 different concentrations of the test chemical (*n*=2 replicates per concentration) and a dilution water control (*n*=2 replicates). The test chemicals are the anti-thyroidal

substance 6-*n*-propyl-2-thiouracil (PTU) and the thyroid hormone thyroxine (T4). The test concentrations for PTU and T4 are as follows:

- PTU (mg/L): 2.5, 5, 10, 20
- T4 (µg/L) : 0.25, 0.5, 1.0, 2.0

Larval density is 20 tadpoles per test tank for all treatment groups. The resulting 10 treatment tanks are randomly assigned to a position in the water bath in order to account for possible variations in temperature and light intensity.

6.3 Preparation of Test Solutions

All exposure experiments according to the XEMA protocol use an aqueous route of exposure. Stock solutions of the test substances PTU and T4 are freshly prepared each week as described below (see section 6.4). Test solutions are formulated by dilution of the corresponding stock solutions in dilution water to obtain a total volume of test solution of 10 L per test tank. Test solutions are exchanged completely three times a week (Monday, Wednesday, Friday) in a static renewal design.

For the static renewal technique used in XEMA, two different renewal procedures may be followed. One possibility is that new test solutions are prepared in clean vessels and the tadpoles are gently transferred into the new vessels. Alternatively, the test animals can be transferred to an intermediate vessel (e.g. 1 L beaker) containing test solution and after the test solution in the test tanks is exchanged completely, the tadpoles are transferred back to their respective test tanks. In either case, it should be ensured that the temperature of the fresh test solution is 22±1°C.

Generally, the exchange of test solutions should be performed in a manner to minimize any stress for the test organisms. The following procedure is recommended. First, 5 to 6 L of old test solution are removed with a 1 L beaker to reduce the water level in the test tanks. The reduced water volume facilitates the removal of tadpoles which should be done by gentle scooping rather than by

collecting in nets in order to avoid any injury of the test organisms. There are reports of toxic effects arising from the use of latex gloves (Gutleb *et al.*, 2001) and it is recommended to use vinyl or nitril gloves during exchange of test solutions. Direct contact between gloves and the inner sides of the test tanks should be generally avoided. Before refilling the test tanks with fresh test solution, the inner tank sides, the bottom and the glass covers are cleaned with warm tap water and a sponge followed by thorough rinsing with deionized distilled water. To prevent cross contamination, separate sets of gloves, strainers, sponges and intermediate vessels are used for each treatment group during water exchange and tank cleaning.

6.4 Preparation of Stock Solutions of PTU and T4

Thyroxine (T4)

FW: 776.9 g/mol

Stock solution: 50 mg/L

Preparation of 100 mL stock solution:

Dissolve 5 mg T4 in 10 mL NaOH (1 M), shake the solution for 60 minutes to achieve complete dissolution. Then, add dilution water to a final volume of 100 mL and shake or stir the solution for further 10 minutes.

Test concentration (µg/L)	Added volumes of stock solution (µL / 10 L test medium)
0.25	50
0.5	100
1	200
2	400

Propylthiouracil (PTU) - anti-thyroid drug

Synonym: 6-n-propyl-2-thiouracil

FW: 107.2 g/mol

Water solubility : limited

Stock solution: 10 g/L

Preparation of 250 mL stock solution:

Dissolve 2.5 g PTU in 20 mL NaOH (1 M), shake vigorously for a few minutes to achieve complete dissolution; then, add dilution water to a final volume of 250 mL.

Test concentration (mg/L)	Added volumes of stock solution (mL / 10 L test medium)
2.5	2.5
5	5
10	10
20	20

Note: after addition of PTU stock solution to 10 L of test medium, adjustment of pH may require the addition of HCl to the test medium.

6.5 Feeding

During the entire test period of either 14 or 21 days, tadpoles are fed daily (including weekends) with sera micron according to the following schedule:

- 300 mg / day / test tank

The food is given to the tadpoles as dry food.

Alternatively, a stock solution of 10 g/L sera micron can be prepared in dilution water. Such a stock solution should be freshly prepared every other day. From this stock solution, tadpoles receive 30 mL / day / tank during the entire test period of either 14 or 21 days.

6.6 Determination of Biological Endpoints

For assessment of developmental stage, whole body length and hind limb length, tadpoles are singly removed from the test tanks by scooping with a small strainer and transferred to a transparent plastic chamber of an approximate size of 80 x 30 x 15 mm containing a few mL of test solution. It is not necessary to anaesthetize the tadpoles for this procedure. Experimental animals should be carefully handled during transfer from the test tanks to the measurement chamber in order to mimimize stress and to avoid any injury of the tadpoles.

The developmental stage of all test animals is determined according to the staging criteria of Nieuwkoop and Faber (1994) by using a binocular dissection microscope. For length measurements, the transparent measurement chamber is placed on a length scale that allows length measurements to the nearest mm. Alternatively, length measurements can be performed using an image analyzing system.

Developmental stage and whole body length are determined for each tadpole at the beginning of the exposure (day 0) and again on days 7, 14 and 21 of exposure. Hind limb length is measured for each tadpole at exposure days 7, 14 and 21. In addition, the test tanks should be checked daily for dead tadpoles and the numbers are recorded. Dead tadpoles should be removed from the test tanks as soon as observed. Under the conditions decribed here, mortality rates in the control groups are less than 5 %.

6.7 Test Termination

On exposure day 21 (initiation of exposure with stage 51 tadpoles) and on exposure day 14 (initiation of exposure with stage 54 tadpoles), tadpoles are anesthetized in MS-222 or on ice for determination of developmental stage, whole body length, hind limb length and wet weight. For possible histological analysis of the thyroid gland, the lawer jaw of tadpoles is preserved in bouins

solution for 12-24 hours. For analysis of gene expression levels, specific tissues (e.g. brain, tail) are readily dissected after anesthesia, snap-frozen in liqid nitrogen and stored at –80°C until RNA isolation.

6.7 Temperature and pH Measurements

Water temperature (22±1°C) and pH (7±0.5) in the test tanks must be checked at least before and after each exchange of test solutions.

7 Statistics

Data sets from whole body length, hind limb length and wet weight measurements of individual test tanks within treatments are analyzed for normal distribution (Kolmogorov-Smirnov-test) and homogeneity of variance (Levene-test). For normally distributed length data, analysis of variance (ANOVA) is used to determine tank-to-tank differences within treatments. Length and weight data which are not normally distributed, are analyzed for possible tank-to-tank differences within treatments by using the non-parametric Kruskal-Wallis test. For the following assessment of treatment-related effects on mean body length, mean hind limb length and mean wet weight, data from all tanks within each treatment group are pooled and in this analysis, the individual tadpoles are considered as the experimental units. Possible treatment effects on tadpole whole body length, hind limb length and wet weight are always assessed by comparison with the SC group. For normally distributed data, either Dunnett's test (homogeneity of variance) or Dunnett's T3 test (heterogeneity of variance) are used. In cases where length data are not normally distributed, pairwise comparisons between the SC group and each treatment group are made by the non-parametric Kruskal-Wallis test followed by the Mann-Whitney U-test.

Developmental stage data are usually not normally distributed and the nature of these data suggests that calculation of mean values is not a meaningful method. Therefore, only non-parametric methods are considered for statistical

analysis of developmental stage data. Developmental stage data are first analyzed using the non-parametric Kruskal-Wallis test for possible tank-to-tank differences in developmental stage composition within treatments. For the following assessment of treatment-related effects on the developmental stage composition, data from all tanks within a given treatment are pooled and in this analysis, the individual tadpoles are considered as the experimental units. The Kruskal-Wallis test is used to determine whether significant differences exist between the developmental stage composition of treatment groups and if indicated, the Mann-Whitney U-test is used for pairwise comparisons of developmental stage data between each treatment group and the SC group to determine which of the treatment groups are significantly different from the SC group. Differences in growth and developmental indices between treatments are considered significant at the level of $p<0.05$.

8 Documentation of Test Results

Data are summarized in tabular form in a data reporting form (Microsoft Excel format) containing appropriate data sheets, showing for each treatment and control group, the number of surviving tadpoles and the developmental stage, whole body length and hind limb length for all tadpoles at each observation.

Corresponding address:

Prof. Dr. Werner Kloas

Department of Inland Fisheries

Leibniz-Institute of Freshwater Ecology and Inland Fisheries

Müggelseedamm 310

12587 Berlin

tel.: +49(0)3064181630

fax: +49(0)3064181799

e-mail: werner.kloas@igb-berlin.de

8 References

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

Bantle, J.A. (1995): FETAX – a developmental toxicity assay using frog embryos. In: Fundamentals of Aquatic Toxicology. Rand, G.M. (ed.). Taylor and Francis, Washington DC, USA, 207-228.

Bantle, J.A., Dumont, J.N., Finch, R.A., Linder, G. and Fort, D.J. (1998): Atlas of Abnormalities. A Guide for the Performance of FETAX. Oklahoma State University Press, Stillwater, USA

Dawson, D.A., Schultz, D.W. and Shroeder, E.C. (1992): Laboratory care and breeding of the African clawed frog. Lab. Anim. 21, 3136

Hilken, G., Dimigen, J. and Iglauer, F. (1995): Growth of *Xenopus laevis* under different laboratory rearing conditions. Lab. Anim. 29, 152-152

Hilken, G., Iglauer, F. and Richter, H.-P. (1997): Der Krallenfrosch *Xenopus laevis* als Labortier. Biologie, Haltung, Zucht und experimentelle Nutzung. Ferdinand Enke Verlag, Stuttgart

Kay, B.K. and Peng, H.B. (eds.) (1992): Methods in cell biology Vol. 36, Academic Press, New York, USA

Gutleb, A.C., Bronkhorst, M., van den Berg, J.H.J. and Murk, A.J. (2001): Latex laboratory-gloves: an unexpected pitfall in amphibian toxicity assays with tadpoles. Environ. Toxicol. Pharmacol. 10, 119-121

MEETING TO DISCUSS PHASE I OF THE AMPHIBIAN METAMORPHOSIS ASSAY

10th-11th March 2004, Hiroshima, Japan

Meeting Report

BACKGROUND AND INTRODUCTION

The first OECD *Ad hoc* Expert Meeting on Amphibian Testing (June 2003, Duluth, U.S.A) proposed and agreed that Germany, Japan, and the United States should conduct a comparison study as the next step in the development of the Amphibian Metamorphosis Assay. The three laboratories volunteered to do a 21-day exposure (stating at stage 51) and a 14-day exposure (starting at stage 54) using thyroxine and propylthiorucail as the test chemicals.

Towa Kagaku Company Limited organised and hosted a meeting to review the status of the comparison studies. This meeting was held under the sponsorship of Environmental Safety Division, Japanese Environment Ministry. The meeting took place at the Garden Palace Hotel, Hiroshima, Japan. It was attended by 17 participants from Germany, Japan, and the United States. The participants list is attached to this report (Appendix I).

PURPOSE AND OBJECTIVES

The chairman of the meeting outlined the purpose and objectives of the meeting. The primary mandates of the meeting were to review the outcome and results of the comparison studies and to:

- Decide if the tests conducted by the three laboratories were robust enough to detect disruptions in the thyroid axis.
- Decide if the selected endpoints are relevant to the assay.
- Determine an appropriate development stage to start exposure experiment and the duration of the assay.
- Make proposals on rearing conditions such as larval density, feeding, evaluate the performance of the flow-through and static renewal regimes, and if possible standardize the thyroid histology techniques.
- Discuss a format to present Phase 1 data to the OECD Expert Group on Amphibian Testing.

AGENDA

The program and agenda of the meeting is attached to this report (Appendix II)

PRESENTATIONS

The three participating labs in the comparison studies, Germany, Japan, and the United States, gave their respective presentations.

- Phase I of Xenopus metamorphosis assay (XEMA) – German contribution: I. Evaluation of morphological endpoints II. Evaluation of histological and molecular biological endpoints. Werner Kloas and Robert Opitz, Germany.
- Phase I of Frog Metamorphosis Assay -Japan contribution. Osamu Tooi, Japan
- Progress Toward Developing An Amphibian Metamorphosis Assay for Thyroid Disruption. Joe Tietge, United States

In the ensuing discussions, the three participating laboratories agreed that, irrespective of the slightly variable conditions under which the tests were conducted between the different laboratories, the similarity in the results suggest that the *X. laevis* model is a relatively robust system to detect disruptions of the thyroid axis. The group agreed to the draft outline of an amphibian metamorphosis assay as shown in Table 1.

Table 1. Draft Proposal for Validation of Amphibian Metamorphosis Assay

Test animal		*Xenopus laevis*	
Exposure period		Exposure from stage 51 for 21 days	
Criteria for selecting test individuals		Primary criterion will be developmental stage, however exclusion criteria based on total length are to be determined	
Concentration of test substance	**PTU**	2.5, 5, 10, 20 mg/l	
	T_4	0.25, 0.5, 1.0, 2.0 µg/l	
Exposure regime		Flow-through design with 25 mL/min	
Endpoints and determination days		Developmental stage	Day 7 and 21
		Whole body length*	
		Hind limb length	Day 7 and 21
		Wet weight	Day 21
		Mortality	Daily observation
		Thyroid histology	Day 21
Control		One dilution water control	
Larval density		20 tadpoles per tank**	
Volume of test medium		4L with minimum water depth of 10 to 15 cm	
Test medium		Locally available and appropriate water demonstrated to promote normal growth and development	
Replication		2 replicates per test concentration	
Acceptable mortality rate		<5% in controls	
Fixation	Number	10 individuals per tank randomly selected	
	Region	Lower jaw or head (to be determined by each laboratory)***	
	Fixation fluid	Bouin's Fixative or alternative used for routine histological methods	
Feeding	Food	Sera micron or mixture of TetraFin, Spirulina algae discs, Silver cup trout Starter, along with live brine shrimp	
	Frequency/amount	Twice a day, quantity adjusted with age of tadpoles	
Lighting	Photoperiod	12 hr light:12 hr dark	
	Intensity	To be measured at water surface	
Water temperature		22±1ºC	
pH		6.5 - 8.0	
Dissolved oxygen concentration		Above 40% of the air saturation value	
Chemistry		Twice a week	

** The issue of body length was not fully resolved. The Japanese and German laboratories measured total length from the tip of the tail to the tip of the nose. The US lab measured from the tip of the nose to the vent. It therefore warrants further discussions.*

*** An alternative approach is to raise the initial loading to 30 organisms per tank, and then subsampling 10 per tank at day 7. The reason for this approach is the concern for the additional handling involved in the acquisition of apical endpoint data. By using a procedure that removes the organisms from the study, we do not have to be concerned with the additional stress of laying the organisms onto a glass plate, etc. This would permit anesthesia to be used for the sampled organisms. This is acceptable from a biological loading standpoint, as the organisms will not reach their maximal body weights until about stage 59, and at day 7 they should be at about stage 54. The only requirement here is that the sampling needs to be randomized.*

**** Requires further discussion to clarify the relative merits of the two different procedures.*

The group also agreed to the following:

- All three participating laboratories should present data using the same tabular and graphical formats.
- All three laboratories will share their respective data, and the US will carry out the statistical analysis after confirming the methodology with Germany.
- Germany will develop draft standard operating procedures (SOP) for the validation phase using the Xenopus Metamorphosis Assay (XEMA) protocol as a framework. This will be passed on to the US, and then Japan for the respective labs to make the necessary inputs. Each laboratory will include comments on the thyroid histology SOP, after consultation with the respective pathologist. Differences in the histological procedures will be reconciled. In addition, Japan will include an SOP for hind limb measurement, which was first presented at the meeting.
- Germany will then develop a final draft protocol for submission to the OECD by April ending.
- The US Representative to the OECD meeting in June will present the draft protocol on behalf of the three laboratories in an oral presentation.

Table 2. Timeframe for work sharing

Meeting to discuss Phase I of Amphibian Metamorphosis Assay	10 – 11 March 2004
Submission of meeting report to OECD	April 2004
Development of draft SOP for the validation phase	March to April 2004
Submission of “Straw man” guideline to OECD	1-14 June 2004
Meeting of Ad hoc Expert Group on Amphibian Testing	28-29 June 2004

APPENDIX I

LIST OF PARTICIPANTS TO THE MEETING TO DISCUSS PHASE I OF THE AMPHIBIAN METAMORPHOSIS ASSAY

10th-11th March 2004
Available to Government Representatives only

APPENDIX II

Meeting to Discuss Phase I of Frog Metamorphosis Assay
Hiroshima, Japan, on 10th-11th March 2004

Program

Wednesday, March 10, 2004

9:00 - 9:05	Welcome and Opening (**Dr. Akihiko Kashiwagi**)
9:05 - 9:10	Welcome (**Dr. Hiroaki Hori**)
9:10 - 9:15	Welcome (**Bungo Goda**)
9:15 - 9:30	Objective of the meeting **(Dr. Akihiko Kashiwagi)**
9:30 - 10:15	German presentation
10:15 – 10:25	Coffee break
10:25 - 10:55	Japanese presentation
11:00 - 11:45	U.S. presentation
12:00 - 13:00	Lunch break
13:00 - 17:00	**Discussion of data presented** **(Chair: Dr. Taisen Iguchi Co-Chair: Dr. Werner Kloas)** • What is the recommended stage of development at initiation of exposure? • What is the recommended exposure period? • Select the endpoints (stage, total length, hind limb length, wet weight, mortality) and measurement methods. • Test conditions (feeding regime and ratio, loading rate, flow through versus static renewal). • Others (thyroid histology, mRNA expression; measurement technique, measurement parameters).
17:00	Adjournment
18:00- 20:00	Welcome dinner

Thursday, March 11, 2004

9:00 - 10:30	**Discussion: Phase 1 data presentation format** **(Chair: Dr. Taisen Iguchi Co-Chair: Dr. Werner Kloas)** Draft report to amphibian expert group (Collection of the data in three labs, statistics methods)
10:40 - 11:45	**Discussion: Phase II draft proposal** **(Chair: Dr. Taisen Iguchi Co-Chair: Dr. Werner Kloas)** Preparation of a straw man guideline for amphibian metamorphosis assay

11:45 – 12:00 Closing remarks and adjournment **(Dr. Akihiko Kashiwagi)**

SECOND OECD MEETING OF THE *AD HOC* EXPERT GROUP ON AMPHIBIAN TESTING

28th-29th June 2004, Paris, France

Final Report of the Meeting

INTRODUCTION

1. The meeting was attended by 20 participants from 10 countries: Australia, Canada, France, Germany, Japan, Spain, Sweden, Switzerland, the United Kingdom, the United States and Industry (BIAC). The participants list is appended to this report (*Annex I*).

ELECTION OF THE CO-CHAIR

2. Following the principle of rotating chairmanship for this group, Dan Pickford from the United Kingdom who co-chaired the last meeting, became the chair to this meeting. Dr. Taisen Iguchi from Japan was elected as new co-chair.

ADOPTION OF THE AGENDA

3. The draft Agenda was adopted with minor changes to reflect the latest proposals from participants to give a presentation on their country activities on Amphibian testing. The final agenda is appended to the report (Annex II).

PURPOSE AND OBJECTIVES

4. The Secretariat made a brief presentation on the purpose and objectives of the meeting. The main objectives of the meeting were to:

- Review the outcome of Phase 1 of the validation and discuss additions to be made to the final Phase 1 report before its submission to the VMG-eco;
- Discuss and agree on the content of Phase 2 of the validation, in terms of standard operating procedures;
- Get organised to submit the proposal for Phase 2 to the VMG-eco at the beginning of September 2004, and agree on the sharing of responsibilities/activities;

DETAILED REVIEW PAPERS

DRP on Amphibian Metamorphosis Assays

5. The Expert Group took note of the last comments and amendments made on the Detailed Review Paper on Amphibian Metamorphosis Assays. This document is going to be published as an OECD document and posted on the OECD public website for free access this autumn (September-October 2004).

DRP on Thyroid Hormone Disruption Assays (in vivo and in vitro)

6. Les Touart indicated that a preliminary draft had been circulated in the United States in April this year and is expected to be submitted to the OECD at the end of August 2004 for circulation to National Co-ordinators.

DRP on Amphibian Growth and Reproduction

7. Les Touart further mentioned that a Detailed Review Paper on Amphibian Growth and Reproduction was in preparation in the US and similarly is expected to be submitted to the OECD at the end of August, with a view to have it circulated to the VMG-eco before its meeting in October 2004.

RESULTS OF PHASE 1 OF THE VALIDATION OF THE AMPHIBIAN METAMORPHOSIS ASSAY

Hiroshima Meeting Report

8. Following completion of Phase 1 experimental work, representatives from each of the three participating laboratories met on March 10th-11th, 2004 in Hiroshima, Japan to review their data, propose a protocol/SOP for Phase 2 of the validation work and identify issues for further discussion with the whole OECD Amphibian Expert Group.

9. Charles Sagoe from Japan summarized the outcome of the Hiroshima meeting. There was confidence that despite the differences in test conditions, the assay is robust enough to detect disruption in the thyroid axis. To the question of development stage at initiation of exposure, the laboratories suggested that stage 51 of metamorphosis should be preferred for the next step in validation. The current meeting further discussed the issue in preparing for Phase 2. Unresolved points from the Hiroshima meeting dealt with body length measurement (2 ways to do it), handling of individuals for the acquisition of apical endpoint data during the test, and dissection of tadpoles for thyroid histology. Otherwise, the participating laboratories organized themselves for the revision of the standard operating procedures, based on experience gained in Phase 1.

Presentation of the results from Phase 1

10. Joe Tietge from the United States presented the results from Phase 1. An analysis of the control data showed that development and growth (determined by weight and length) do not vary much across laboratories (maximum difference between labs for weight: 27%, for length: 6%) nor within laboratories (maximum within lab difference for weight: 18%, for length: 13%). One laboratory was in the process of getting familiar with the flow-through exposure system when they did Phase 1 experiments, and this might explain some of the within lab difference mentioned earlier. Overall, *Xenopus laevis* was thought to be robust despite differences in test conditions (exposure regime and food type).

11. The results of four tests were then explained: PTU (thyroid antagonist) 14-days study and 21-days study, T4 (thyroid agonist) 14-days study and 21-days study. The 14-days studies started at stage 54 of metamorphosis and the 21-days studies started at stage 51 of metamorphosis. The reasons for having both development stages and different exposure durations in Phase 1 were that:

i) pre-metamorphosis (from stage 51 to 54) seems to be a good period for the detection of thyroid agonism, as endogenous levels of thyroid hormone are relatively low at that period,

ii) it was not fully clear whether a 14-day test beginning with stage 54 was sensitive enough to detect thyroid agonism, and finally

iii) although it was generally accepted that a stage 54 study-based assay would adequately capture antagonism, experts suggested to compare its sensitivity relative to a longer assay beginning with stage 51.

12. Endpoints measured: whole body length, body weight, final development stage (or staging), hind limb length (not measured by the US lab), thyroid histology, biomolecular endpoints(measured in the German lab only): TSH-β, TR-β and Basic Transcription Element-binding Protein (BTEB) mRNAs.

PTU studies

13. Both the 14-d and the 21-d assays were able to detect significant effects. For staging, the level of statistical significance was reached at the highest concentration (20 mg/l PTU) in all labs; significance was also detected at 10 mg/l in the German lab (21-d study) and in the US lab (14-d study). For hind limb length, the significance level was reached at 20 mg/l in the German and Japanese labs at day 7 of exposure in the study starting at stage 51. In both assays (14-d and 21-d), thyroid histology was the most sensitive endpoint and effects could be detected at 5 mg/l in all laboratories. It was acknowledged that a uniform grading system with agreed histological parameters will need to be used in future to strengthen thyroid histology as an endpoint in the metamorphosis assay.

T4 studies

14. Both the 14-d and the 21-d assays were able to detect significant effects. Staging reached statistical significance at 2 µg/l in all laboratories in the study starting at stage 54 and at 1 µg/l in some studies starting at stage 51. Hind limb length measurement was significant at 1 µg/l in the stage 51 study after 7 days of exposure (2 µg/l in the stage 54 study).Thyroid histology, performed at termination of the experiments, was significant at 1 µg/l in the US lab and at 2µg/l in the German lab for both 14-d and 21-d studies.

15. It was concluded that thyroid histology (performed at termination of the assay) tends to be the most sensitive endpoint for antagonism; it was noted that follicular cell height is not as sensitive as the size of thyroid gland. Hind-limb length (measured at d-7 of exposure in the stage 51 study) and thyroid histology were the most sensitive endpoint to agonism.

16. The following things were learnt from Phase 1 Validation:

- *X. laevis* can be used successfully in different laboratories as a toxicological model for thyroid disruption;
- *X. laevis* is quite robust to different culture and testing conditions (when conditions are further standardized, inter-lab variation is expected to be minimal);
- Results of the PTU and T4 studies were remarkably similar among laboratories;
- The endpoints used in these studies were similarly sensitive and consistent with the modes of action of each chemical;
- For the chemicals tested, the 21-d assay beginning with NF stage 51 and the 14-day assay beginning with NF stage 54 both detected significant thyroid hormone disruption effects at equivalent sensitivities (5mg/l PTU and 1µg/l T4).

17. Participants to the meeting had no particular comments on the draft report of Phase 1 but it was agreed that the report will be checked by statisticians from Europe and the US to ensure that appropriate statistical tests were used, especially for non-parametric data (development stage). Tim Springer (US) and Clemens Schanné (Switzerland) will be consulted on this matter. The Secretariat reminded the need for full transparency regarding data and methodological details for future peer-review. Additionally, the report should contain a rationale on how Phase 1 results are used to build Phase 2. Robert Opitz from Germany made the commitment to finalize the report by the end of August 2004; Joe Tietge will work with him to that end. The histological evaluations still missing from Japan (T4 14-d and21-d studies and PTU 21-d study) will be submitted to Germany by the end of August for inclusion in the report. Analytical chemistry also needs to be included in the final report. The report will then be forwarded to the VMG-eco for

comments and approval in September-October; any issue arising from the report will be discussed at the VMG-eco 3 meeting on 21-22 October 2004.

PREPARATION FOR PHASE 2 OF THE VALIDATION

Stage at initiation of exposure

18. Extensive discussions on the stage at initiation of exposure took place before reaching an agreement for Phase 2. On the one hand, both the 14-d assay (starting at stage 54) and the 21-d assay (starting at stage 51) are able to detect thyroid agonism and antagonism, and therefore the shorter version should be preferred. On the other hand, the assay starting at stage 51 may be slightly more sensitive than the assay starting at stage 54 for the detection of agonistic effects (through hind limb measurement at d-7 of exposure). The group acknowledged that in the absence of data on an assay starting at stage 51 for a duration of 14 days, Phase 2 of the validation should be taken as an opportunity to clarify whether such an "stage 51/14-d" version would be able to combine the advantages of both the "stage 51/21-d"and "stage 54/14-d" versions.

Core endpoints and observation time points

19. In order to gather sufficient information in Phase 2 for further decision, it was agreed that three observation time points are needed in Phase 2:

- d-7 of exposure for hind limb measurement for the detection of agonistic effects
- d-14 and d-21 of exposure where all endpoints will be measured to be able to decide whether the shorter version is sensitive enough.

Core endpoints	d-0	d-7	d-14	d-21
Hind limb length	-	Yes	Yes	Yes
Staging	-	-	Yes	Yes
Whole body length	-	Yes	Yes	Yes
Body weight	-	-	Yes*	Yes
Thyroid size	-	-	Yes*	Yes
Thyroid histology	-	-	Yes*	Yes

Table 1: endpoints measured and time points.
*: endpoints measured at d-14 from a subset of test animals.

Exposure regime

20. Phase 1 experiments showed that the exposure regime (flow-through or static renewal) did not affect much the results. However, participating laboratories in Phase 1 indicated that in order to standardize test conditions, a flow-through system should be preferred in Phase 2.

Food type

21. Sera micron will be used by all participating laboratories in Phase 2; the amount needs to be agreed to keep tadpole development optimal under flow-through conditions.

Chemicals

22. The meeting participants proposed to use T4 as an agonist, IOP (iopanic acid) as an monodeiodinase inhibitor and perchlorate as an inhibitor of iodine uptake. These three substances will be tested in a range of 3 to 4 concentration levels. The issue of positive controls was addressed. Participants felt that at this stage of the validation, positive controls (e.g. one PTU and one T4 concentrations) were not needed nor useful as the selection of only one concentration of either of the chemicals would not contribute to assessing the sensitivity range of the assay.

23. T4 concentrations will be identical to concentrations used in Phase 1 (0.25, 0.5, 1, 2 µg/l). For IOP and Perchlorate, Joe Tietge and Robert Opitz will consult each other and make a proposal by the end of August when the SOPs for Phase 2 are completed. Les Touart will check availability of IOP from the chemical repository.

Whole body measurement

24. There are two options to take measures: snout-vent length and snout-tail length. It was recommended to take both measures as it does not represent extra work. The information gathered will then help in deciding which approach is best.

Thyroid histology

25. The fixative to be used is Davidson's; Joe Tietge will circulate the fixation protocol to Robert Opitz for inclusion in the SOPs for Phase 2. Dissection of the whole head as per the US protocol will be followed. Thomas Braunbeck will contact Japanese and American pathologists for an agreement on a standard grading system for the evaluation of thyroid histology endpoints; there exist a mammalian template which could be used as a basis for a guidance in thyroid histology. Thomas Braunbeck and Robert Opitz will ensure that the final SOPs also include sectioning techniques.

Lead laboratory

26. The USEPA Duluth laboratory will lead Phase 2 validation work.

Timeframe and action plan

Submission of standard operating procedures for Phase 2 to OECD Secretariat	end of August
Circulation to VMG-eco for comments and approval	September to mid October
Experimental part of Phase 2	November 2004 to June 2005
Thyroid histology	June to September 2005
Draft Report	October 2005

MEMBER COUNTRY ACTIVITIES [1]

Activities on the Amphibian Metamorphosis Assay

27. Australia has developed an assay using a native species: *Lymnodynastes peronii. Xenopus laevis* being an exotic species for Australia, there are restrictions to its import and maintenance in the laboratory. The "LYMA" (*Lymnodynastes* Metamorphosis Assay) system developed was tested using T4 and PTU in the same concentration range as in Phase 1. Its sensitivity appears to be similar to the metamorphosis assay using *Xenopus laevis*. However, if a Test Guideline was developed using *Xenopus laevis*, data generated by such Test Guideline would be accepted and used in Australia.

28. Sweden conducted a *Xenopus* metamorphosis assay using *Silurana tropicalis* (also known as *X. tropicalis*). The results show that the species is not so sensitive to PTU: development stage and hind limb length was significantly affected at 50 mg/l PTU under semi-static exposure conditions. Refinements to the assay, such as temperature adjustment and food availability, could enhance its sensitivity.

29. Japan informed the meeting of on-going activities on proteome analysis of several organs (liver, tail) from transgenic frogs exposed to thyroid hormones T3. Protein expression profile was also made for tadpoles exposed to bisphenol A (BPA). Results show that BPA inhibits T3-induced tadpole tail regression. Data on response of transgenic frogs exposed to T3, T4, TRIAC, and TETRAC was also presented.

30. The United States mentioned that they will also work with *S. tropicalis* in the future. Five chemicals are being or will be tested using the *Xenopus* metamorphosis assay: Phenobarbital, methimazole, pregnelone-16-alpha-carbonitrile, dexamethasone, PTU and T4. The results of this multi-chemical study are expected to be circulated to the expert group by December 2004. The US is also interested in protein expression profiles.

Other country activities on Amphibians

31. Australia will compare the sensitivity of fish and amphibians towards estrogenic substances (17-β estradiol) using a 60-d assay. Canada is trying to breed native species in captivity and plans higher tier-tests using *Rana pipiens*; the Canadian delegate also mentioned comparative work with *X. laevis* and *S. tropicalis*. France indicated the development of a micro-array on *S. tropicalis*. Japan is working on the following topics: 1) developing a partial life-cycle test using *Xenopus* sp., 2) developing sex reversal test using all ZZ-male tadpoles, 3) working on *Xenopus* receptor binding assay, 4) compiling an atlas database and gene analysis using *X. tropicalis* and *Rana sp*., 5) gene analysis of giant salamander has started, 6) Hiroshima University is now capable of providing *X tropicalis* for interested parties. The United Kingdom informed the meeting of on-going research/ED on UK native species *Bufo bufo* and *Rana temporaria*, with a view to develop a partial life-cycle test. The United States are currently drafting a Detailed Review Paper on Amphibian growth and development. Also in the US, experimental work is under way on *S. tropicalis* in a test starting with embryos and going until sexual maturity (6 months), flutamide is the test substance; a report is expected at the end of the year 2004.

WORKSHARE ACTIVITY

32. The Secretariat reminded the Meeting of the database for work sharing on endocrine disrupters. This database is located on the password protected website for the EDTA TF work. The database is updated regularly with country submissions on testing plans, test method development plans, test results

[1] All presentations given at the meeting are posted on the meeting website.

and assessment reports. For instance, the XEMA report submitted by Germany has been posted there. Participants are encouraged to submit documents that fit into one of the categories mentioned above, and which are related to OECD projects (e.g. amphibian metamorphosis assay).

CLOSURE OF THE MEETING

33. The meeting was adjourned at 15:00 on Tuesday. Dates for the next meeting will be proposed when sufficient progress in Phase 2 of the validation has been made. It is anticipated that it will take place in the fourth quarter of 2005.

ANNEX I

LIST OF PARTICIPANTS TO THE SECOND MEETING OF THE *AD HOC* EXPERT GROUP ON AMPHIBIAN TESTING

28th-29th June 2004

(available to government representatives only)

ANNEX II

MEETING OF THE *AD HOC* EXPERT GROUP ON AMPHIBIAN TESTING

Held in OECD Headquarters, Chateau de la Muette, Paris on 28th-29th June 2004

AGENDA

Monday, 28th JUNE 2004	
09h30	**OPENING AND INTRODUCTION TO THE MEETING**
09h30	Welcome and Opening
09h45	Election of a new Co-Chair
09h55	Adoption of the Agenda
10h00	Purpose and objective of the meeting (Secretariat)
10h20	**RESULTS OF VALIDATION PHASE 1**
10h20	**Presentations from speakers:** • Brief presentation of Hiroshima Meeting Report • Presentation of results from phase 1 • Overview of points for further discussion
11h00 - 11h30: Coffee Break	
11h30	**Discussion** • Needs for final phase 1 report • Preliminary conclusions for phase 2
13h15 - 14h30: Lunch Break	
14h30	**STATUS OF RELEVANT DETAILED REVIEW PAPERS**
14h30	**Status report on preparation of the DRP on Thyroid Hormone Disruption Assays, Covering *in vivo* and *in vitro* Methods and Mechanisms of Action**
15h00	**Final Draft of Detailed Review Paper on a Amphibian Metamorphosis Assay** • Final comments received and corresponding response / revisions; • Schedule for submission as an approved OECD DRP.
15h30 - 16h00: Coffee Break	
16h00	**CURRENT MEMBER COUNTRIES ACTIVITIES RELATED TO METAMORPHOSIS ASSAYS**
16h00	• Work with transgenic test organisms. • Update on Swedish work with *Xenopus tropicalis* and *Rana sp.* • Development of a metamorphosis assay using Australian native species. • Work sharing activity (*reference to the OECD ED WorkShare Database)* as contribution to co-ordination and avoidance of duplication.
18h00: Meeting adjourns for the day	

Tuesday, 29th JUNE 2004	
09h00	**PREPARATION OF VALIDATION PHASE 2** ***(incl. Coffee Break 10h15 – 10h45)***
09h00	• Further specification of protocol as needed • Approach for common histotechnical procedures and histopathological evaluation • Consideration of links to work in VMG-mammalian and VMG-non animal • Selection of test chemicals, arrangements with the chemicals repository • Interests and offers for participation • Action plan and time frame
12h15-13h30: Lunch Break	
13h30	**PERSPECTIVES, ACTIVITIES ON AMPHIBIAN ENDOCRINE DISRUPTERS TESTING BEYOND METAMORPHOSIS ASSAYS**
13h30	**Projects in member countries / on multi-national levels** • Projects/studies on amphibian testing for endocrine disrupters, including reproduction and development • Further endocrine disrupters research activities on amphibians in Japan (45 min) • …
15:00	**Perspectives for future work of OECD Expert Group on Amphibians Testing**
15h30 Meeting adjourns	

ANNEX III

List of action items agreed at the meeting and timelines for completion

Finalisation of Phase 1 report

- Japanese lab (Ch.Sagoe / O.Tooi) to submit remaining histological results to German lab (R. Opitz) **by end of August 04**; histological results will be appended to Phase 1 report;
- R. Opitz/C.Schanné and J.Tietge/T.Springer to check at the statistical analysis (especially for non-parametric data) in the **July-August timeperiod**;
- R. Opitz to consolidate Phase 1 report and submit it to the OECD Secretariat **by end of August 04**;
- OECD Secretariat (A. Gourmelon) to circulate the report to VMG-eco at the **beginning of September 2004** (VMG-eco meeting: 21-22 October 2004) for comments and approval;

Preparation of Phase 2 of the validation

- Th. Braunbeck to look into the thyroid histology grading system, and based on the existing template, propose a system to be used by pathologists in Phase 2; timeline: to be ready when Phase 2 starts (**November 2004**);
- R. Opitz and J. Tietge to check concentration range for IOP and Perchlorate in **July 2004**;
- L. Touart to ensure that sufficient amounts of test substances are available for Phase 2 laboratories (**October 2004,** once we know how many labs in Phase 2);
- J. Tietge to send the fixation protocol (Davidson's) to Robert Opitz for inclusion in the SOPs for Phase 2, **July 2004**;
- R. Opitz to revise the SOP for Phase 2 (IOP and Perchlorate concentrations and analytical method, frequency of analytical measurements; whole body measurement methods, thyroid fixation and sectioning) and to submit them to the Secretariat **by end of August 2004**;
- OECD Secretariat to prepare the proposal to the VMG-eco for Phase 2 (rationale, SOP, list of labs – call for additional labs, etc.) **at the beginning of September**, with a request for comments, if any, before the meeting in October.
- Lead laboratory (J. Tietge) to ensure that template data sheets reflect the required measurements before the start of Phase 2 (to be ready as soon as possible and **no later than end of October 2004**); Lead laboratory to distribute the SOP to participating laboratories and to address technical questions from participating laboratories during Phase 2 experimental work; lead laboratory to collect test results from participating labs and perform the statistical analysis and the drafting of preliminary report.

Note from the Secretariat:

This document contains the proposal for Phase 2 of the validation of the Amphibian Metamorphosis Assay. The proposal was elaborated on the basis of conclusions reached in Phase 1 and discussions held at the Second Meeting of the OECD Amphibian Expert Group, in June 2004.

The proposal was circulated for comments to the VMG-eco in September 2004. Comments received are listed in the table thereafter. Of particular interest are the comments that affect the proposed test design:

- Conflict on whether we should use a protocol starting at stage 51 or stage 54, based on the assumption that a stage 51 protocol might be more sensitive to thyroid agonism;
- it is suggested to increase replicate number from 2 to 5 - this reduces the sample size from 25 to 10 tadpoles per replicate, but does not affect the total number of animals required nor a randomized sampling scheme. The statistical tests proposed would be markedly improved in terms of power;
- Alternatively, it is suggested to ***i)*** either use more powerful step-wise tests (e.g., William's or Jonckheere-Terstra's step-down test) which take into account the expected direction of the effect, which are more powerful and biologically appropriate, instead of the Dunnett's multiple comparison test; ***ii)*** or to treat individual specimen as the units of analysis and ignore the replicate (can only be done for continuous data and under the condition that the within-tank normality of data is satisfied);
- More appropriate statistical methods are proposed, such as the Shapiro-Wilk's test for the normality, and binomial or Chi2 procedures for quantal /non-continuous data such as developmental stage.

The VMG is invited to consider these comments and others listed in the table and advise on changes needed to Phase 2 design before starting experimental work.

Table 1: Comments received on the proposal for Phase 2 of the validation of the Amphibian Metamorphosis Assay.

Source	**Comments**
Les Touart US EPA	Given my interpretation of the Phase 1 report, I have some difficulty endorsing the Phase 2 proposal to continue with the Stage 51 protocol. The key argument for beginning with Stage 51 is that it is "more sensitive" for detecting agonists, but there are only a very few known agonists (really no known environmental contaminant agonists, antagonists being more prevalent) and the data from Phase 1 do not support the argument but rather demonstrate that the Stage 54 protocol is quite capable of detecting agonists at a similar sensitivity to the Stage 51 protocol. The Phase 1 study demonstrates the robustness of the frog metamorphosis assay and the general reproducibility of the protocol. I would hope that we could effectively collect sufficient data in Phase 2 to allow completion of the validation process. I would therefore recommend we consider using the Stage 54 protocol for a Phase 2 interlaboratory validation exercise rather than a Stage 51 protocol based interlaboratory optimization exercise. The Phase 2 proposal addresses whether the Stage 51 protocol can be shortened to 14 days and provides comparable data to the 21 day version. It doesn't really answer the question of whether a 14 day assay would be better to start at Stage 51 or Stage 54. I am concerned that if Phase 2 proceeds as proposed, that we will be left with more questions than answers in selecting an optimal protocol to establish a harmonized test guideline. More specifically, to the proposal as it is. Quantitative histopathological measurements will be made to address follicular cell height, gland cross-sectional area, and follicle cross-sectional area. One area that is not addressed is colloid content/density which has been deemed a sensitive indicator of anti-thyroidal activity in mammals (O'Connor et al. 1999). The approach taken to measure colloid content requires use of periodic acid-Schiff stain. Differences in colloid staining are more apparent using PAS than H&E, and in conjunction with image analysis software, colloid density may be determined. I would recommend that PAS stain be considered in place or in addition to H&E stain. The proposal does not detail the rationale for the experimental design and sample sizes. It would be helpful to have a power analysis or discussion of the statistical approach to be taken provided in the section on experimental design.
John Green DuPont	**19**. Test Design: 2 tanks of 25 tadpoles at each of 4 test concentrations (0, .25. .5. 1. and 2 g/l) and a dilution water control. Flow-through design, 25 ml/min. Subsample of 5 tadpoles per tank at day 14, measured for wet weight and thyroid histology. Development stage, body length, hind limb length measured at days 7, 14, and 21. Wet weight and thyroid histology measured on day 21. Mortality measured daily.

	Comment: This means there will be two data points (tank mean or total) per concentration at each sampling time. For wet weight, body or hind limb length this will yield low power tests and limited information for curve fitting, resulting in uncertain shapes and wide confidence intervals. The design could and should be explored by computer simulation to determine the statistical properties before this study is conducted. There appear to be serious issue regarding what can be determined statistically from the proposed design. Some of these are discussed in detail below. **50, 51**. **Comment:** Other formal tests for normality should be allowed; specifically, the Shapiro-Wilk test. Also, no criteria are given as to what significance level is used to conclude non-normality or variance heterogeneity. Also, what data are used to explore normality and heterogeneity should be explained. In an ANOVA context, usually the residuals from the ANOVA are used for this purpose, pooling the residuals from all groups. The residuals from a regression models are used for this purpose where regression models are employed. If there are multiple variance components at each concentration, as is the case in this design, each with multiple subjects, then care must be taken to use the proper error term in the comparison of treatment means and assessment of normality or variance heterogeneity. Alternatively, the replicate is considered the sampling unit and the means for the two sampling units are the values analyzed, with an appropriate weighting scheme. Should some tadpoles die or are lost during the experiment, the replicate means should be weighted in the analysis. In fact, sample size alone is not the only consideration in determining the weight given to a replicate mean. Statistical judgment based on appropriate plots (scatterplots, histograms, QQ-plots, Stem-and-leaf plots) should be allowed to override or replace the results of formal tests. What is done if the data are judged non-normal or heterogeneous should be stated. Perhaps a statistical flow-chart would be helpful. Formal tests for normality based on two measurements for each of five concentrations (four positive concentrations plus control) are rather weak. Formal tests for variance heterogeneity are even weaker to the point of rarely being worthwhile. The plots are probably preferable, but only gross violations will be sufficiently clear to call. It would be far preferable to divide the 50 tadpoles per concentration into five replicates of 10 than two replicates of 25. All of the statistical properties would be improved. Dunnett's test makes no use of the dose-response nature of the experiments. It is designed more for experiments in which the treatment groups are unrelated, such as different chemicals or different mechanisms. Generally, increased concentrations are expected to give rise to increased effects. That is, the response is expected to be monotone with respect to concentration of the chemical and methods that incorporate the expected direction of effect are more powerful and more biologically appropriate. Examples of such approaches are Williams' test, Jonckheere-Terpsrta, Brown-Forsythe, and Welch, all applied in stepdown fashion. Given the paucity of data under the current scheme, an exact version of the Jonckheere-Terpstra test, such as available in SAS Proc Freq or StatXact would be preferable to the standard large sample version. All of the parametric tests, including Dunnett's, are likely to have low power for the given design and all would be markedly improved by the alternative design of five replicates of 10 tadpoles each at each concentration.. The use of the step-down tests named above based on ordered alternatives will help address the inconsistencies that pairwise methods often generate, such as a low or intermediate concentration mean significantly different from the control mean but means at higher

	concentrations are not significantly different from the control. The exact step-down Jonckheere-Terstra test is quite powerful even when applied to normally distributed data and reasonably powerful even for such small data sets as will be generated by this design. Dunn's test will be almost valueless for the design given, due to the small sample sizes. An exact Jonckheere-Terpstra test applied in stepdown manner will be more consistent with the experimental design and expected monotone dose-response. There is also no reason to require a significant Kruskal-Wallis test before applying Dunn's test, provided Dunn's test is applied only so as to compare each treatment mean to the control. This is analogous to the application of Dunnett's test in the normal case without requiring a significant F-test first. An alternative procedure that may be applicable under some conditions is to treat the individual specimen as the unit of analysis and ignore the replicate. This 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 serious 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. Some additional detail is needed about which developmental endpoints are to be analyzed by the methods indicated to be sure the assumptions underlying the non-parametric tests are met. **General Comment**: This is a non-statistical comment that applies to rat studies as well. A deficiency in Phase 2 is the lack of test chemicals with a broader range of mechanisms for thyroid effects. An earlier OECD meeting report noted that there is no data on whether this assay is sensitive to substances that affect liver metabolism in mammals (i.e. induction of UDP-GT) resulting in increased clearance of thyroid hormone and a compensatory response by the thyroid to produce thyroid hormone. I would recommend an inducer of UDP-GT be suggested (e.g., phenobarbital) for inclusion in the validation as this is a fairly common mechanism of thyroid hyperplasia in rats (males in particular), although well accepted not to be relevant to humans.
Toni Ratte Germany	**§25**: the statistical design is poor (2 replicates) for a study by which the determination of a NOEC is intended. It should be considered to increase the replication or - if not desirable for whatsoever reason – to shwitch to a regression approach where poor replication is less critical. **§50 and §51**: the statistical methods should be brought in line with the recommendation of the Draft OECD Statistical Guidance Document. Normality should be checked by the Shapiro-Wilks test. The NOEC should be determined using a more powerful step-wise test such as the Tamhane-Dunnett test or Williams test. Developmental data here are typical quantal data (a certain number out of the total

	introduced tadpoles shows a certain stage) and should be evaluated using methods particularly designed for quantal responses (binomial or Chi^2 procedures).
Victoria Pablos Spain	**Table 1**: Are there any recommended range of light intensity for Xenopus? **Paragraph 27**: the factor of 4 for PER concentrations is too large; the selection of this large factor increases the uncertainty to establish a NOAEC for this compound; for this particular case, it would be better to test five concentrations separated by a factor of two instead of four concentrations. **Paragraphs 29-30**: some of the hidrophobic compounds need to be dissolved on carried solvents, and in these cases, flow-through systems without solvents do not assure the dissolution of these compounds in the water. It would be interesting to have a list of potential solvents that could be used on Xenopus, as well as the concentration limits of these carriers that are not considered toxic for Xenopus. Does this information exist?. **Paragraph 31**: Although many test mediums can be used for the assay, a standarisation of this test would be important, at least for static exposures, due to the pattern of degradation of the tested compounds could be different in different exposure mediums. In these cases (static exposures), a detailed characterisation of the test medium should be included. **Table 2 (page 13):** determination of body length would be recommended at day 0 in order to compare the increases through the experiment.
Hans Rufli Switzerland	I have just a minor comment regarding the analytical sampling of the Proposal for Phase 2: a.. on **page 7**: test solution from each replicate tank should be sampled once per week b.. on **page 9** at bottom of table: "Chemistry: Once per week" c.. on **page 15**: "Samples of the exposure water of each tank should be taken twice a week..." Thus, **page 15** should be changed to "once per week". I do not recall any other guidelines which would ask for more then weekly sampling.

PROPOSAL FOR PHASE 2 OF THE VALIDATION OF THE AMPHIBIAN METAMORPHOSIS ASSAY FOR THE DETECTION OF THYROID ACTIVE SUBTANCES

September 2004

TABLE OF CONTENTS

SCOPE OF THE DOCUMENT **9**
TEST ORGANISM **9**
OBJECTIVE OF THE ASSAY **9**
BIOLOGICAL ENDPOINTS **10**
Apical Morphological Endpoints 10
Developmental stage 10
Hind Limb Length 10
Thyroid Gland Histology 10
Body Length and Wet Weight 11
Additional Observations 12
Analytical Chemistry 12
EQUIPMENT **12**
PRE-EXPOSURE PHASE **13**
TEST INITIAITION AND CONDUCT **13**
Selection of Experimental Animals 15
Selection Criteria 15
Animal Handling 15
General Experimental Design 15
Test Substances 16
Exposure System Characteristics 16
Test Medium Properties 17
Feeding 17
Determination of Biological Endpoints 17
Day 7 Measurements 18
Day 14 Measurements 18
Day 14 Subsampling of Test Animals 18
Day 21 Measurements (Test Termination) 19
Additional Observations 19
Histopathological Analysis of Thyroid Gland 20
Analytical Chemistry 20
Measurements of Physico-Chemical Parameters 20
REPORTING REQUIREMENTS **20**
STATISTICS **20**

REFERENCES..20

ANNEXES

Annex 1 **Breeding of *X. laevis* in the Laboratory**
Annex 2 **Recommendations for Analytical Chemistry Methods**
Annex 3 **Digital Length Measuring System**
Annex 4 **Histological Analysis of Thyroid Gland**

SCOPE OF THE DOCUMENT

1. This document contains a proposal for Phase 2 of the Validation of the Amphibian Metamorphosis Assay for the detection of thyroid active substances. It has been prepared by the participating laboratories in Phase 1, in conjunction with the *ad hoc* Expert Group on Amphibian Testing, following completion of Phase 1 and release of the draft report of Phase 1 (OECD, 2004a). The proposal outlines further experimental work, including standard operating procedures, to be conducted in Phase 2. The scientific basis of this assay and an extensive literature review are presented in a Detailed Review Paper on Amphibian Metamorphosis Assay (OECD, 2004b), thereby limiting the present document to the description of experimental procedures and possible outcome of Phase 2.

2. The Validation Management Group for Ecotoxicity Testing is invited to *i)* review the proposal, in particular the objective of the assay, the chemicals suggested as reference substances, the endpoints measurement and timepoints and the statistical approach, *ii)* make comments or approve the proposal as is, before 22 October 2004. Comments should also be made in light of prior discussion on the purpose of the bioassay (OECD, 2003a; OECD, 2003b, OECD, 2004c).

TIMELINES AND ACTION PLAN

3. Once this proposal is approved and any potential issue solved at the end of October 2004, participating laboratories will be able to start Phase 2 experimental work. The schedule of experimental work foresees completion in June 2005. Thyroid histology will be performed during the July-August period. A draft report should become available in October 2005.

4. The Secretariat canvassed participants to the *ad hoc* Expert Group on Amphibian Testing last June and it appears that 6 laboratories can join Phase 2. These laboratories are located in Germany (1), Japan (1), Switzerland (1), the United Kingdom (1) and the United States (2). A Lead laboratory has been identified in the United States.

TEST ORGANISM

5. The Metamorphosis Assay uses tadpoles of the South African clawed frog *Xenopus laevis*. The test can be performed with animals bred in the laboratory or with tadpoles obtained from commercial suppliers (e.g. Nasco in the USA). The tadpoles used in this test should be derived from a healthy brood stock.

OBJECTIVE OF THE ASSAY

6. The primary objective of the Amphibian Metamorphosis Assay is the evaluation of thyroid system-disrupting activities of individual test compound. The postembryonic development (metamorphosis) of *Xenopus laevis* and the regulatory role played by thyroid hormones (TH) during this process are well characterized. Metamorphic development is dependent on a functional hypothalmus-pituitary-thyroid axis. Inhibitory effects on TH synthesis and antagonistic effects on TH action lead to retarded metamorphic development. Moreover, metamorphosis can be precociously induced and/or accelerated by treatment with TH. The fact that *X. laevis* tadpoles respond differently to agonists and antagonists of the thyroid system provides the rationale for using this developmental process as a biological model to screen for thyroid system-disrupting chemicals. In the assay, exposure of *X. laevis* tadpoles is initiated at developmental stage 51 and is continued for a total of 21 days. A sub-sampling of 5 tadpoles per treatment tank is performed at exposure day 14 for collection of head tissue for thyroid gland histopathology.

7. Tadpoles are exposed to at least 4 different concentrations of a test substance and a dilution water control. During the exposure period, apical morphological endpoints (developmental stage, hind limb length,

whole body length) are assessed for treatment-related deviations from normal development and histological analysis of thyroid gland tissue is conducted with head tissue samples taken from test organisms. Chemical exposure is via the aqueous route achieved using a flow-through exposure regime.

TEST CHEMICALS

8. The proposed chemicals for the Phase 2 will be T4, sodium perchlorate (PER) and iopanoic acid (IOP). All three compounds are currently considered as potential reference compounds to be used in validation studies of endocrine-related tests where the thyroid system is a target. T4 is the native prohormone and is used as a reference compound for agonist activity on the thyroid system. T4 was already used for effects on metamorphosis and thyroid system function during Phase 1. PER is a well-known inhibitor of iodine uptake by the thyroid gland and has been shown to retard metamorphosis in *X. laevis* tadpoles (Goleman *et al.*, 2002a, 2002b; unpublished data from the US and GER lab). IOP is an inhibitor of all monodeiodinases, enzymes which catalyze the metabolism of TH in peripheral tissues (Becker *et al.*, 1997). Therefore, IOP is used as a reference compound for modulation of peripheral TH action. IOP is expected to enhance TH action in tissues that are normally protected from TH by expression of type III monodeiodinase (e.g., tail) while this compound may block TH action in tissues that require efficient conversion of T4 to T3 (e.g., hind limbs).

BIOLOGICAL ENDPOINTS

Apical Morphological Endpoints

9. The primary apical morphological endpoints to assess possible effects of the test substances on metamorphic development are the developmental stage of the tadpoles and the length of the hind limbs.

Developmental stage

10. The developmental stage of *X. laevis* tadpoles can easily be determined by using the staging criteria of Nieuwkoop and Faber (1994). The metamorphosis assay is designed to detect both agonistic and antagonistic effects on the thyroid system. Anti-thyroidal compounds are expected to retard metamorphic development, resulting in earlier developmental stages compared to untreated controls. On the other hand, TH agonists may cause enhanced metamorphic development resulting in more advanced developmental stages compared to untreated controls.

Hind Limb Length

11. Differentiation and growth of the hind limbs are under control of thyroid hormones and are major developmental landmarks already used in the determination of developmental stage (Section 4.1). THs exert stimulatory effects on hind limb morphogenesis, consistent with metamorphic development of other tissues. Hind limb development is used qualitatively in the determination of developmental stage, but is considered here as a quantitative endpoint. Therefore, hind limb length is measured as an additional endpoint to detect effects on the thyroid axis.

Thyroid Gland Histology

12. While developmental stage and hind limb length are important endpoints to evaluate exposure-related changes in metamorphic development, developmental delay cannot, by itself, be considered a diagnostic indicator of anti-thyroidal activity. This is based on the following two facts. First, some chemicals likely exist that may retard development through other modes of toxicity not related to perturbation of thyroid function. Second, recent studies demonstrate that normal development can still occur when a partial inhibition of TH synthesis is present. This suggests that some level of TH synthesis inhibition can be overcome by the compensatory mechanisms of the hypothalamic-pituitary-thyroid axis, resulting in effects

on the thyroid gland that are observable by routine histopathological analysis. These are not major considerations when working with model test chemicals with known modes of thyroid toxicity. However, when assessing the effects of chemicals during the screening process, it is important to consider these potential problems, as it is reasonable to assume that some chemicals exist which exhibit thyroid-disrupting potencies below those of the commonly used model chemicals (ie, methimazole, perchlorate, PTU). Therefore, histological analysis of the thyroid is a necessary endpoint that improves the diagnostic capability of the assay. The utility of thyroid histology is based, in part, on the well-established facts that inhibition of normal thyroid function (i.e., reduced circulating T3/T4 levels) results in an increase in TSH release by the pituitary. The thyroid gland responds to TSH stimulation by increasing synthetic activity, which is manifest at the histological level as follicular cell hypertrophy and hyperplasia and eventually thyroid gland hypertrophy. In addition to the compensatory responses of the thyroid gland to TSH, differential histopathological effects can also be observed which may be related to the mechanism of toxicity of the test chemical. Exposure to perchlorate, for example, an inhibitor of thyroidal iodide uptake, results in different histopathological effects than PTU, an inhibitor of thyroid peroxidase activity. These observations add to the potential diagnostic contribution of this endpoint.

Body Length and Wet Weight

13. Determinations of body length and wet weight are included in the test protocol to assess possible effects of test substances on the growth rate of tadpoles in comparison to the control group. Although measurements of tadpole wet weight may provide a more precise estimation of tadpole growth, weight determinations during the course of the experiments are technically much more difficult. Compared to length measurements, the removal of adherent water for weight determinations would cause more stressful conditions for tadpoles and may cause skin damage. Therefore, wet weight measurements are only performed at test termination (day 21) and for a subsample of test organisms terminally sampled at day 14.

14. There are two different approaches to assess tadpole growth by means of body length measurements: measurement of snout-vent length or whole body length (see Figure 1). The issue of which of these two body length measurements should be used was not fully resolved in the Phase 1 validation exercise. Therefore, for Phase 2 of test protocol development, it is proposed to determine both length parameters in order to produce a solid database for a comparison of both endpoints concerning their utility to reflect changes in tadpole growth.

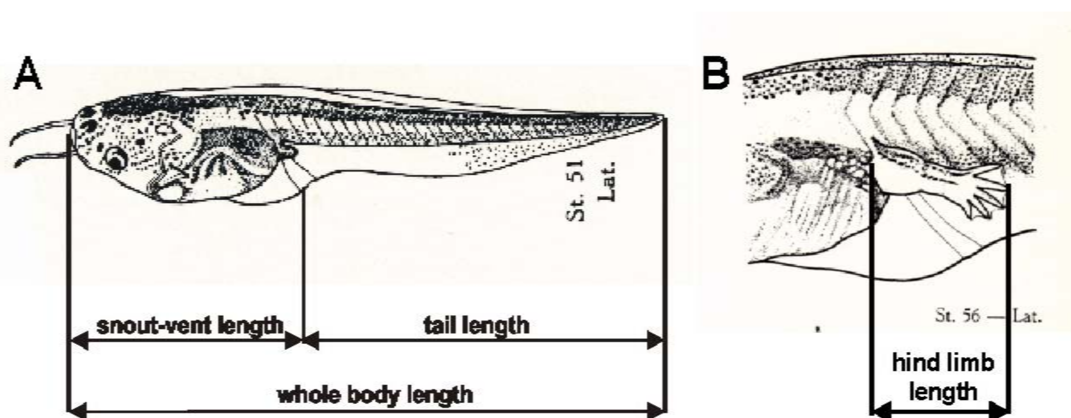


Figure 1 Snout-vent length (A), whole body length (A) and hind limb length (B) in *Xenopus laevis* tadpoles (Nieuwkoop and Faber, 1994).

15. Disruption of thyroid system function does not necessarily affect tadpole growth. Exposure to T4 hastens development which results in reduced organism weights through two mechanisms. First, the normal period of growth that occurs in pre- and pro-metamorphic larvae is reduced upon exposure to T4, so the ultimate weight of the larvae is typically reduced. Second, there is a normal loss of weight throughout

metamorphic climax in control animals and this weight loss is accelerated commensurate with the advanced development associated with exogenous T4 exposure. Exposure to TH synthesis inhibitors, on the other hand, prevents the larvae from experiencing the normal weight loss associated with climax and they continue to grow as prometamorphic animals and can reach extraordinary weights. Consequently, an analysis of growth by either length or weight requires some interpretation. For example, reduced growth rates in the absence of developmental effects provide an indication of toxic effects of a test compound.

Additional Observations

16. All test tanks are checked daily for dead tadpoles and the numbers are recorded. Mortality rates exceeding 5 % may indicate inappropriate test conditions or toxic effects of the test materials. Mortality was completely absent in untreated control groups during the experiments of validation Phase 1. Further measurements of non-specific toxicity could be incorporated into this test protocol, including the monitoring of tadpoles for abnormal behaviour (e.g. uncoordinated swimming, loss of equilibrium, hyperventilation, atypical quiescence, non feeding) or developmental anomalies (e.g. limb deformities, etc.). Such effects, although difficult to quantify can aid in the interpretation of test results by providing some information on general toxic effects of the test substances. However, at this stage of the evaluation process, no distinct criteria for a qualitative and quantitative assessment of possible morphological or behavioral abnormalities are provided.

Analytical Chemistry

17. Comparability among studies conducted by different laboratories is dependent on verifying test concentrations by analytical chemistry methods. Appropriate methods for the test substances mentioned above are recommended in Annex 2 and a sampling scheme must be developed based on the physicochemical properties of the compound. Test solution from each replicate tank at each concentration should be sampled for analytical chemistry once per week once the test is running. More frequent sampling is recommended during setting up the system (i.e., before day 0, at day 0 and within the first week of exposure). In addition, it is recommended that stock solutions be analyzed when they are changed, especially if the duration of the stock solution volume does not encompass the routine sampling. In the case of chemicals which cannot be detected at some or all of the concentrations used in the test, stock solutions must be measured and system flow rates must be recorded and used to calculate nominal concentrations. Also, prior to conducting a study, the stability of the compound on test should be evaluated using existing information on its degradability, volatility, etc.

EQUIPMENT

The following equipment is needed:

- flow-through equipment (description see below)
- glass vessels (description see below)
- aquarium heaters (adjustable to 22±1°C)
- thermometer (preferably a recording thermometer measuring temperature minima and maxima)
- binocular dissection microscope
- digital camera with at least 4 mega pixel resolution and micro function
- image analyzing software (description see Annex 3)
- transparent plastic chamber of an approximate size of 80 x 30 x 15 mm
- pH meter, oxygen meter, balance
- devices to scoop tadpoles out of the tanks (preferably a small strainer with a rather rigid mesh)
- various glassware (beakers, graduated flasks, graduated cylinders, Erlenmeyer flasks etc.)
- adjustable pipettes, disposable micropipettes (1000 µL), graduated pipettes (5 mL, 10 mL, 20 mL)

PRE-EXPOSURE PHASE

18. The pre-exposure phase should be started with tadpoles at developmental stages 44/45. If tadpoles are purchased from commercial suppliers, it is recommended to ship them at stages 44/45. If laboratory-bred animals are used for the assay, recommendations for raising tadpoles to stages 44/45 are summarized in Annex 1 of this document. It is recommended to raise approximately 400 tadpoles until test initiation, since at least 250 tadpoles of stage 51 are required for initiation of exposure. During the pre-exposure phase, tadpoles should be acclimated to the conditions of the actual exposure phase. This includes mainly the type of food, temperature, light-dark cycle and the culture medium. If a static culture system is used for maintaining tadpoles during the pre-exposure phase, the culture medium should be replaced completely at least twice per week. Crowding effects caused by high larval densities during the pre-exposure period should be avoided because such effects could markedly affect tadpole development during the subsequent testing phase. Therefore, the rearing density should be adjusted to approximately 40 tadpoles per 10 L culture medium (static culture system) or 30 tadpoles per 4 L culture medium (flow through culture system). Under these conditions, tadpoles should develop from stages 44/45 to stage 51 within 7-12 days. Representative tadpoles of this stock population should be inspected daily for developmental stage in order to estimate the developmental stage of the stock population and to identify the appropriate time point for initiation of exposure.

TEST INITIATION AND CONDUCT

19. The critical parameters of the test are summarized in Table 1 and are discussed in more detail below.

Table 1. Summary of test parameters for validation of Amphibian Metamorphosis Assay (Phase 2)

Test animal		*Xenopus laevis*	
Exposure period		Exposure from stage 51 for 21 days * * (subsampling of 5 tadpoles per replicate tank at day 14)	
Criteria for selecting test individuals		Primary criterion will be developmental stage, however further exclusion criteria based on total length are optional	
Concentration of test substance	**T4**	0.25, 0.5, 1.0, 2.0 g/l	
	PER	8, 32, 125, 500 µg/l (or 65, 125, 250, 500 µg/L)	
	IOP	0.75, 1.5, 3.0, 6.0 mg/l	
Exposure regime		Flow-through design with 25 mL/min	
Endpoints and determination days		Developmental stage	Day 0, 14, and 21
		Body length	Day 7, 14 and 21
		Hind limb length	Day 7, 14 and 21
		Wet weight	Day 21 * (and subsample at day 14)
		Mortality	Daily observation
		Thyroid histology	Day 21 * (and subsample at day 14)
Control		One dilution water control	
Larval density		25 tadpoles per tank * (subsampling of 5 tadpoles per tank at day 14)	
Volume of test medium		4 L with minimum water depth of 10 to 15 cm	
Test medium		Locally available and appropriate water demonstrated to promote normal growth and development	
Replication		2 replicates per test concentration	
Acceptable mortality rate		< 5% in controls	
Fixation	Number	5 individuals per tank (randomly selected)	
	Region	head	
	Fixation fluid	Davidson’s Fixative	
Feeding	Food	Sera micron	
	Frequency/amount	at least twice a day, quantity adjusted with age of tadpoles	
Lighting	Photoperiod	12 hr light:12 hr dark	
	Intensity	To be measured at water surface	
Water temperature		22±1°C	
pH		6.5 - 8.0	
Dissolved oxygen concentration		Above 40% of the air saturation value	
Chemistry		Once a week	

Selection of Experimental Animals

Selection Criteria

20. All tadpoles that are used as test organisms should be derived from the same spawn. The exposure phase of the test is initiated with stage 51 tadpoles that have been raised from stage 44/45 under the conditions described in the section "Pre-exposure conditions". The developmental stage is determined according to the staging criteria of Nieuwkoop and Faber (1994). In addition to the developmental stage selection, it is recommended to perform a size selection of the experimental animals. For this purpose, the whole body length should be measured at day 0 for a sub-sample of approximately 20 stage 51-tadpoles. After calculation of the mean whole body length for this group of animals, minimum and maximum limits for the whole body length of experimental animals can be set by allowing a range of the mean value ±3 mm (mean values of whole body length ranged between 24.0 and 28.1 mm for stage 51 tadpoles in experiments conducted during Phase 1 of the validation). Tadpoles that exhibit grossly visible malformations or damages (e.g. damage of the tail) are excluded from the assay.

Animal Handling

21. The exposure is initiated when a sufficient number of tadpoles in the stock population have reached developmental stage 51 according to Nieuwkoop and Faber (1994). For selection of test animals, all healthy and overtly normal looking tadpoles of the stock population should be pooled in a single vessel containing an appropriate volume of dilution water. For developmental stage determination, tadpoles are then individually removed from the pooling tank by scooping with a small strainer (rigid mesh) and transferred to a transparent measurement chamber (approximate size 80 x 30 x 15 mm) containing a few mL of dilution water. The experience gained during previous experimental work indicate that it is not required to anesthetize the tadpoles for stage determination but animals should be carefully handled during this transfer in order to minimize handling stress and to avoid any injury.

22. The developmental stage of the animals is determined by using a binocular dissection microscope. In order to reduce the ultimate variability in developmental stage, it is important that this staging is conducted as accurately as possible. According to Nieuwkoop and Faber (1994), the primary developmental landmark for selecting stage 51 organisms is hind limb morphology. The morphological characteristics of the hind limbs should be examined at 15 to 20-fold magnification. Because the morphological appearance of the hind limbs at stage 51 differs markedly from the limb morphology at stages 50 and 52, only little experience is needed to correctly distinguish the different stages of the tadpoles.

23. If an additional size selection based on whole body length is performed, stage 51 tadpoles are individually placed on a glass board (150 x 150 mm) and a dorsal view image showing the entire tadpole is captured using a digital camera (for further details see Annex 3 of this document).

24. Tadpoles that meet the stage and length criteria described above are collected in groups of five animals in small vessels (e.g., glass pots or beakers containing 1 L of dilution water), that have been randomly assigned to test vessels and labeled accordingly. Groups of five animals are repeatedly transferred to all glass pots until each vessel contains 25 tadpoles. Each pot is then inspected for animals with abnormal appearance (injuries or abnormal swimming behavior). Overtly unhealthy looking tadpoles are removed from the pots and replaced with larvae that are newly selected from the pooling tank. The exposure phase of the test is initiated by gently decanting the contents of each glass pot (containing a nominal of 25 tadpoles) into each of the test vessels.

General Experimental Design

25. The general experimental design is that tadpoles are exposed to at least 4 different concentrations of the test chemical (n=2 replicates per concentration) and a dilution water control (n=2 replicates). Larval density at test initiation is 25 tadpoles per test tank for all treatment groups. The treatment tanks are

randomly assigned to a position in the flow-through exposure system in order to account for possible variations in temperature and light intensity.

Test Substances

26. The substances used in testing during validation Phase 2 include thyroxine (T4; CAS # 51-48-9), sodium perchlorate (PER; CAS # 7791-07-3) and iopanoic acid (IOP; CAS # 96-83-3). The rationale for the selection of these substances is explained in the final report of Phase 1 validation work. Test concentrations were selected according to the following considerations:

- test concentrations must not reflect concentrations likely to be encountered in the environment (lower tier screening assay);
- at least three of the four test concentrations should represent sublethal concentrations (based on results from preliminary work conducted in laboratories participating in validation Phase 1);
- concentrations should span a concentration range that would allow to observe graded responses of the selected endpoints in order to validate endpoint measurements and to compare endpoint sensitivities (based on results from preliminary work conducted in laboratories participating in validation Phase 1);

27. The test concentrations proposed for validation Phase 2 are as follows:

- T4 (µg/L): 0.25, 0.5, 1.0, 2.0
- PER (µg/L): 8, 32, 128, 512
- IOP (µg/L): 750, 1500, 3000, 6000

28. Note that test concentrations for T4 and IOP are separated by a factor of 2 whereas test concentrations for PER are separated by a factor of 4. The higher dilution factor proposed for use in PER exposure studies is based on results from previous work performed in the US lab. The separation factor of 4 was selected to incorporate treatment levels which are expected to produce significant effects on all endpoints while ensuring that at least one treatment level is below the no-observed adverse effect concentration (NOAEC).

Exposure System Characteristics

29. All exposures shall be conducted using flow-through methods. There are two reasons for using flow-through methods. The first is that water quality is easier to maintain within a range that promotes normal growth and development. While *X. laevis* is more adaptable to varying water quality than other species and they are known to thrive in static conditions, flow-through systems are still valuable to eliminate waste products and replenish dissolved oxygen concentrations depleted by chemical and/or biological oxygen demand. More importantly, however, flow-through systems ensure that an efficacious exposure is administered. Several factors can conspire to prevent an efficacious exposure, including chemical losses to hydrolysis, photolysis, volatility, adsorption to surfaces, and biological degradation. Furthermore, relatively hydrophobic chemicals may be mass-limited in a static renewal system. Most of these problems can be overcome by use of flow through-methods, although knowledge of the physicochemical properties of the compound is important in judging the success of an exposure.

30. Numerous methods have been used to achieve flow-through exposures and several laboratories have systems that are custom built. Therefore, it is not useful to proscribe the system itself (ie, hardware), but to specify the performance of the system. In general, such systems should have water-contact components of glass, stainless steel, and Teflon. However, suitable plastics can be utilized if they do not compromise the study. Exposure tanks should be glass aquaria (with approximate measurements of 22.5 x 14.0 x 16.5 cm deep) equipped with standpipes that result in an actual tank volume of 4.0 L. The flow rate to each tank shall be 25 ml/minute. Fluorescent lamps should be used to provide a photoperiod of 12 hr light: 12 hr dark. Stock solutions should be dissolved in dilution water without the use of carrier solvents and dispensed to the water distribution system in a manner that meets the range of concentrations stipulated for the test.

Test Medium Properties

31. At the current state of knowledge, it is not possible to exactly define the range of appropriate physico-chemical characteristics of the water used for preparation of test solutions. Any water that is locally available (e.g. springwater or charcoal-filtered tap water) and permits normal growth and development of *X. laevis* tadpoles could be used. Because local water quality can differ substantially from one area to another, analysis of water quality should be undertaken, particularly, if historical data on the utility of the water for raising *Xenopus* is not available. Special attention should be given that the water is free of copper, chlorine and chloramines, all of which are toxic to frogs and tadpoles. It is further recommended to analyze the water concerning background levels of iodine, perchlorate and chlorate (by-product of drinking water disinfection) as all of these anions are substrates of the iodine transporter of the thyroid gland and elevated levels of each of these anions may confound the study outcome.

Feeding

32. Tadpoles are efficient filter feeders and a variety of diets have been used for *Xenopus* tadpoles raised in the laboratory (Dawson *et al.*, 1992; Hilken *et al.*, 1995; Kay and Peng, 1992; Schultz and Dawson, 2003). Although different types of food have been successfully used in experiments of validation Phase 1, a standardization of tadpole feeding regimes is proposed to minimize inter-laboratory differences in basal tadpole growth and development during validation Phase 2.

33. During the entire test period of 21 days, tadpoles are fed daily with Sera Micron® (Sera GmbH, Heinsberg, Germany), a commercially available tadpole food that has been shown to support proper growth and development of *X. laevis* tadpoles. Sera Micron is a fine particulate that stays suspended in the water column for a long period of time and hence, it is subject to washing out with the flow. Therefore, the total daily amount of food should be divided into smaller portions, which are given 2 to 3 times per day. A total daily food ration of 600 mg Sera Micron per tank is recommended for the first days of exposure. During the course of the experiments, the total daily food ration should be increased to accommodate tadpole growth. Feeding rates are recorded to allow for a further optimization of feeding regimes in subsequent validation studies.

34. The food can be given to the tanks as dry food. Alternatively, a stock solution of Sera Micron (density of 10 g/L) can be prepared in dilution water. Such a stock solution should be freshly prepared every other day.

Determination of Biological Endpoints

35. During the exposure phase of 21 days, determination of selected endpoints is performed on days 7, 14 and 21 of exposure. These time points were selected to collect data about the relationship between exposure duration and endpoint sensitivity. Such data are required to determine the minimum test duration that allows for a sensitive detection of the effects of the three test substances which represent different modes of action. Table 2 provides an overview on the proposed measurement endpoints and the corresponding observation time points.

Table 2. Observation time points for selected endpoints

	day 0	day 7	day 14	day 21
apical endpoints				
- stage	+		+	+
- hind limb length		+	+	+
- body length	(+)[a]	+	+	+
- body weight			+*	+
thyroid gland size			+*	+
thyroid gland histology			+*	+

Note: [a] determination of body length is optional for size selection of test animals
* endpoints are only measured for a subset of 5 animals per treatment tank

Day 7 Measurements

36. For day 7 determinations of hind limb length and whole body length, all tadpoles are removed from each test tank by scooping with a small strainer and transferred into a small vessel (e.g., beaker or glass pot, filled with 1 L of the corresponding test solution) that has been assigned to each test tank and labeled accordingly. Experimental animals should be carefully handled during this transfer in order to minimize stress and to avoid any injury of the tadpoles. It is not necessary to anaesthetize the tadpoles for the following measurement procedures. Tadpoles are then individually placed on a glass board (150 x 150 mm) and a ventral view image showing the entire tadpole is captured using a digital camera (for further details see Annex 3 of this document). After these measurements, tadpoles are gently decanted from the glass board back into the test tanks.

Day 14 Measurements

37. For day 14 determinations of developmental stage, all tadpoles are removed from each test tank by scooping with a small strainer and transferred into a small vessel (e.g., beaker or glass pot, filled with 1 L of the corresponding test solution) that has been assigned to each test tank and labeled accordingly. Tadpoles are then individually transferred into a transparent chamber of an approximate size of 80 x 30 x 15 mm containing a few mL of test solution. Experimental animals should be carefully handled during this transfer in order to minimize stress and to avoid any injury of the tadpoles. It is not necessary to anaesthetize the tadpoles for the following measurement procedures.

38. The developmental stage of all test animals is determined according to the staging criteria of Nieuwkoop and Faber (1994) by using a binocular dissection microscope (15 to 20-fold magnification). For day 14 determinations of developmental stage, special emphasis should be given to the morphological characteristics of the hind and fore limbs.

39. For snout-vent, whole body and hind limb length measurements, the tadpole is placed on a glass board (150 x 150 mm) and a dorsal view image showing the entire tadpole is captured using a digital camera (for further details see Annex 3 of this document). After these measurements, tadpoles are gently decanted from the glass board back into the test tanks.

Day 14 Subsampling of Test Animals

40. During the apical endpoint measurements at day 14, a total of 5 tadpoles are randomly selected from each treatment tank for subsequent histopathological analysis of the thyroid gland. For randomization, a random number generator should be used to select 5 numbers from the numbers 1-25. The tadpoles corresponding to these 5 numbers are then selected for sub-sampling during performing the endpoint measurements.

41. The selected tadpoles are anesthetized in 2000 mg/L MS-222 (tricaine methanesulfonate). Tadpoles are rinsed in water and blotted dry followed by body weight determination to the nearest milligram. Thereafter, tadpoles are decapitated behind the eyes. Dissected head tissue samples containing the lower jaw are placed in Davidson's fixative for 48 hours. Preservation and processing of tissue samples for histology are described in Annex 4.

Day 21 Measurements (Test Termination)

42. At test termination (day 21), all tadpoles are removed from the test tanks and are anesthetized in 2000 mg/L MS-222 (tricaine methanesulfonate). Tadpoles are rinsed in water and blotted dry followed by body weight determination to the nearest milligram. Developmental stage, body and hind limb lengths are determined for each tadpole as described above for day 14 measurements.

43. During the apical endpoint measurements at day 21, a total of 5 tadpoles are randomly selected from each treatment tank for subsequent histopathological analysis of the thyroid gland. For randomization, a random number generator should be used to select 5 numbers from the numbers 1-20. The tadpoles corresponding to these 5 numbers are then selected for head tissue sampling for thyroid gland histopathology.

44. Dissected head tissue samples containing the lower jaw are placed in Davidson's fixative for 48 hours. Preservation and processing of tissue samples for histology are described in Annex 4.

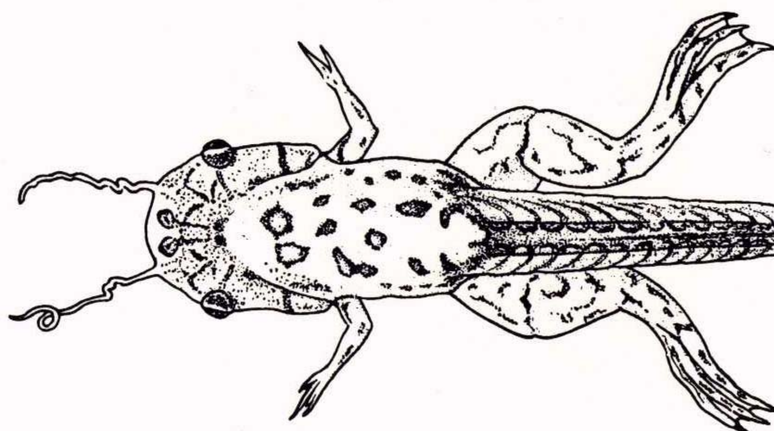


Figure 2. Schematic diagram showing the rostrocaudal level of the disconnection line for dissection of head tissue.

Additional Observations

45. All test tanks are checked daily for dead tadpoles and the numbers are recorded for each tank. Dead animals are removed from the test tank as soon as observed. Further cases of abnormal behaviour (e.g. uncoordinated swimming, hyperventilation, atypical quiescence, non feeding) and grossly visible malformation (e.g. limb deformities, etc.) should be recorded.

Histopathological Analysis of Thyroid Gland

46. Tissue samples for thyroid gland histopathology are collected from test animals at day 14 (subsampling) and day 21 (test termination). At both time points, 5 tadpoles are randomly selected from each test tank and head tissue is dissected. Tissue samples are placed in Davidson's fixative for 48 hours. The methodological aspects of dissection, preservation and processing of tissue samples for histopathology are explained in Annex 4.

Analytical Chemistry

47. Samples of the exposure water of each test tank should be taken twice a week for verification of test concentrations by analytical chemistry methods. Recommendations for analytical methods to be used during validation Phase 2 are provided in Annex 2 of this document.

Measurements of Physico-Chemical Parameters

48. Water temperature (22±1°C), pH (6.5-8.0) and dissolved oxygen concentration (> 40% of the air saturation value) in the test tanks should be checked at least every other day.

REPORTING REQUIREMENTS

49. Each participating laboratory should record and provide the raw data to the lead laboratory for centralised statistical analysis. A standard Excel spreadsheet will be provided by the lead laboratory for the reporting of results.

STATISTICS

50. Data sets from body length, hind limb length, and wet weight measurements are analyzed for normal distribution (Kolmogorov-Smirnov-test) and homogeneity of variance (Levene-test). For normally distributed data, Dunnett's test should be used to compare data from the control group to all other treatment groups. Data that are not normally distributed, should be 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.

51. Calculation of mean values is not a meaningful method for developmental stage data. Therefore, only non-parametric methods should be considered for statistical analysis of these data. The Kruskal-Wallis test should be used to determine whether significant differences exist between the developmental stage compositions of the different treatment groups and, if indicated, Dunn's multiple comparison test should be used for comparisons with the control group. Differences in growth and developmental indices between treatments are considered significant at the level of $p<0.05$.

REFERENCES

Dawson, D.A., Schultz, D.W. and Shroeder, E.C. (1992): Laboratory care and breeding of the African clawed frog. Lab. Anim. 21, 3136

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

Hilken, G., Dimigen, J. and Iglauer, F. (1995): Growth of *Xenopus laevis* under different laboratory rearing conditions. Lab. Anim. 29, 152-152

Kay, B.K. and Peng, H.B. (eds.) (1992): Methods in Cell Biology Vol. 36, Academic Press, New York, USA

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

OECD (2003a). Report of the First Meeting of the OECD *ad hoc* Expert Group on Amphibian Testing, 28-29 June 2003. Duluth, United States.

OECD (2003b). Report of the seventh Meeting of the Task Force on Endocrine Disrupters Testing and Assessment, 26-28 May 2003. OECD, Paris.

OECD (2004a). Draft Report of Phase 1 of the Validation of the Amphibian Metamorphosis Assay for the detection of Thyroid Active Substances.

OECD (2004b). Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances. OECD Environment Health and Safety Publications. Series on Testing and Assessment. No. 46 Paris, June 2004

OECD (2004c). Report of the Second Meeting of the OECD *ad hoc* Expert Group on Amphibian Testing, 28-29 June 2004. OECD, Paris.

Schultz, T.W. and Dawson, D.A. (2003). Housing and husbandry of Xenopus for oocyte production. Lab Anim. 32, 34-39

Annex 1: Breeding of *Xenopus laevis* in the Laboratory

Adult Breeding Animals

Details on holding the brood stock under satisfactory conditions can be found in the literature concerned with the performance of the FETAX test (Bantle, 1995; Bantle *et al.*, 1998) or other publications related to general culture and care of adult *X. laevis* (Dawson *et al.*, 1992; Hilken *et al.*, 1995; Hilken *et al.*, 1997; Kay and Peng, 1992; Schultz and Dawson, 2003).

Induction of Spawning

Spawning of adult *X. laevis* should be induced approximately 14 days prior to the intended initiation of the assay. After fertilization, the exact time period for tadpoles to develop to stage 51 can be variable depending on laboratory conditions. Under controlled animal husbandry, larvae typically hatch within 3 days and develop to stage 51 within 14 to 16 days of spawning. Spawning of adult *X. laevis* is induced by injecting human chorionic gonadotropin (hCG) into the dorsal lymph sac of male and female adult frogs as described by Bantle *et al.* (1998). For the induction of spawning, a hCG stock solution with a concentration of 200 U/100 µL is freshly prepared immediately before injection. A recommended schedule to induce mating is that two adult males are injected in the morning with 200 µL of hCG stock solution. After injection, the males are separated and placed in two 60 L tanks filled with 40 L of culture medium at 22±1°C. These breeding tanks are aerated with airstones and covered with a sturdy lid to keep the animals under low light conditions. In order to protect the laid eggs from the parent animals, a spawning grate, held a few centimeters off the bottom, is placed into the breeding tanks so that spawned eggs will fall through the mesh to the bottom of the tank. Approximately 10 to 12 hours after the first injection, the males receive a second injection of 200 µL hCG stock solution. At the same time, two female adults are injected with 500 µL hCG stock solution. One male and one female are placed together in a breeding tank to go into amplexus. The breeding pairs are left to breed overnight under low light conditions and any disturbance of the breeding pairs should be avoided during this time. Normally, the eggs are laid during the night. The next morning, the adult frogs are removed from the breeding tank. Adult frogs should not be bred more frequently than every 3 months. Breeding males and females should appear healthy and free of abnormalities or injuries. Only adult frogs that have not been treated for diseases during the past 4 weeks should be used for breeding. The number of eggs per spawn under optimal laboratory conditions generally ranges from 500 to more than 2,000. Under the test conditions described here, 200 tadpoles are required for each test trial. However, it is recommended to simultaneously induce breeding in two adult pairs in order to ensure an appropriate number of fertilized eggs and hatched larvae. All tadpoles used in a test should be from the same spawn.

Raising of *X. laevis* Larvae in the Laboratory to Stages 44/45

Eggs and developing tadpoles are reared at 22±1°C and pH 6.5-8.0 during all phases of development. The light-dark cycle is 12:12 hours. The same water should be used as culture medium for eggs and tadpoles during the pre-exposure phase and as dilution water for preparation of test medium during the exposure phase. After spawning, eggs and developing tadpoles can be maintained in large holding tanks (e.g. 60 L tanks) containing approximately 40 L of culture medium. If static culture conditions are used during the pre-exposure phase, aeration of the tanks by airstones is required. Tadpoles hatch within 3 days of spawning. Feeding of tadpoles starts when they reach the free-swimming state at day 5 post-fertilization. Tadpoles are fed daily. The total daily food ration depends on the culture system (static renewal or flow-through) used for raising *X. laevis* tadpoles during the pre-exposure phase. If a static culture system (rearing density:

40 tadpoles per 10 L culture medium) is used, the total daily food ration should be approximately 300 mg / tank. Under flow through conditions, higher daily feeding rates of SeraMicron are needed because 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, a total daily food ration of 400 to 600 mg SeraMicron per tank is recommended for flow-through systems. It is recommended that small portions of food are added to the tanks 2 or 3 times per day.

References

Bantle, J.A. (1995): FETAX – a developmental toxicity assay using frog embryos. In: Fundamentals of Aquatic Toxicology. Rand, G.M. (ed.). Taylor and Francis, Washington DC, USA, 207-228.

Bantle, J.A., Dumont, J.N., Finch, R.A., Linder, G. and Fort, D.J. (1998): Atlas of Abnormalities. A Guide for the Performance of FETAX. Oklahoma State University Press, Stillwater, USA

Dawson, D.A., Schultz, D.W. and Shroeder, E.C. (1992): Laboratory care and breeding of the African clawed frog. Lab. Anim. 21, 3136

Hilken, G., Dimigen, J. and Iglauer, F. (1995): Growth of *Xenopus laevis* under different laboratory rearing conditions. Lab. Anim. 29, 152-152

Hilken, G., Iglauer, F. and Richter, H.-P. (1997): Der Krallenfrosch *Xenopus laevis* als Labortier. Biologie, Haltung, Zucht und experimentelle Nutzung. Ferdinand Enke Verlag, Stuttgart, Germany (*in German*)

Kay, B.K. and Peng, H.B. (eds.) (1992): Methods in Cell Biology Vol. 36, Academic Press, New York, USA

Schultz, T.W. and Dawson, D.A. (2003). Housing and husbandry of *Xenopus* for oocyte production. Lab Anim. 32, 34-39

Annex 2 Digital Length Measuring System

Introduction

The conventional method of length measurements in tadpoles is by calipers. However, this methodology could be inaccurate and particularly for hind limb length almost impossible, especially at premetamorphic stages. Computer software analysis of digital images has been successfully applied in diverse biological research (Johnson et al., 2001; Machado-Silva et al., 2000; Merris et al., 2003), and is fairly accurate. This SOP describes how to measure tadpole body length and hind limb length using digital systems.

Equipment

- Digital camera with micro function (e.g. Canon, EOS 10D, 4 mega pixels)
- Remote switch (e.g. Canon, RS-80N3)
- Angle finder (e.g. Canon, ANGLE FINDER C)
- Lens (e.g. Canon, Macro Photo Lens MP-E 65 mm F2.8 1-5x)
- Lighting system (e.g. LEICA, CLS 150X)
- Glass board (15 cm x 15 cm)
- Image analysis software (e.g. Image Pro Plus Version 4.0, Media Cybernetics, U.S.A.)

Fig.1. Camera setup

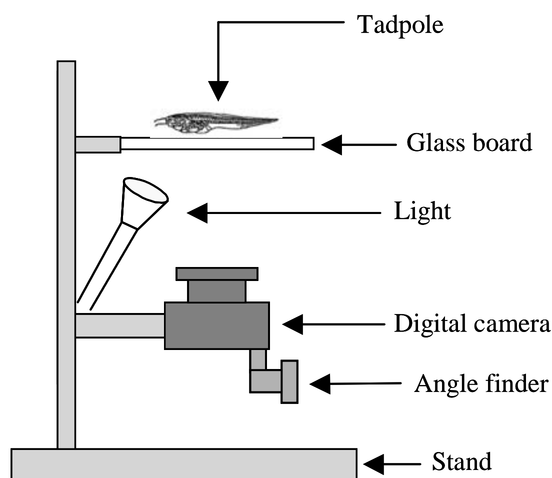


Fig.2. System schematic

Setup and photographing

1. Attach the glass board to the stand.
2. Attach the camera to the stand such that the lens faces upward and under the glass board (see Figure 2).

3. Adjust the light under the glass board.
4. Attach the remote switch and the angle finder to the camera (Figure 1).
5. Wet the glass board slightly and place carefully the tadpole on the wet glass such that the ventral side faces the camera lens.
6. Switch on lamp so that it illuminates the tadpole from below.
7. Adjust the focus such that the hind limbs are in full view and take a series of photographs.

Image analysis

1. Download the images from the camera to a computer. Use an appropriate image-analyzing software for analysis of the digital images (e.g. Image-Pro Plus).
2. Select the best images for measurements of the body length and the hind limb length, respectively.
3. For body length measurements, locate the tip of the snout, the end of the vent and the tip of the tail. Draw two lines, one from the tip of the snout to the end of the vent (snout-vent length) and another from the tip of the snout to the tip of the tail (whole body length).
4. For hind limb length measurements, locate the shining gold-colored part of the abdomen where the limb attaches to the abdomen, and draw a line right to the tip of the limb (Figure 3A).
5. If the limb is bent draw the line from that part of the abdomen to the knee joint (Figure 3B). Click and create a node at the knee joint, and continue drawing the line all the way to the tip of the limb (Figure 3B).
6. Calculate the limb length accordingly.

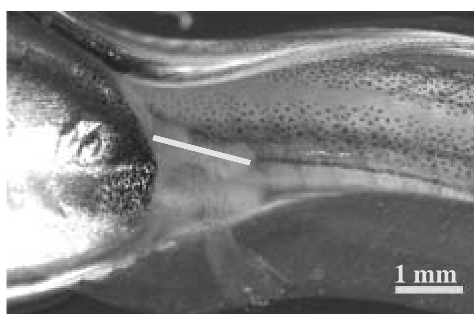


Fig. 3A

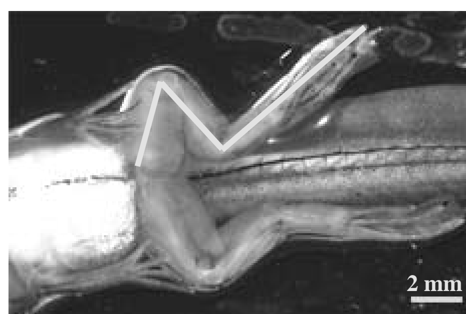


Fig. 3B

References

Johnson, I., Harman, M., Farrow, D. and Norris, M. (2001): An assessment of the feasibility of using image analysis in the oyster embryo-larval development test. Environ. Toxicol. 16, 68-77.

Machado-Silva, J.R., da Silva, C.H., Pereira, M.J., de Oliveira, R.M., Pinto, R.M. and Gomes, D.C. (2000): Differences in Brazilian strains of *Schistosoma mansoni* evaluated by means of morphometric analysis of cercariae of both sexes. Mem. Inst. Oswaldo Cruz. 95, 839-842

Merris, M., Wadsworth, W.G., Khamrai, U., Bittman, R., Chtwood, D.J. and Lenard, J. (2003): Sterol effects and sites of sterol accumulation in *Caenorhabditis elegans*: developmental requirement for 4-methyl sterols. J. Lipid Res. 44, 172-181

Annex 3 Analytical Methods for Perchlorate, Iopanoic Acid and Thyroxine (Joe, could you please check the symbols used for concentrations and flow rates, I think the original symbols got lost or have been replaced by other curious symbols)

The following sections describe the methodology for analyses of water samples as used in previous experimental work by the US lab. There are certainly other viable methods for detection of perchlorate (PER), thyroxine (T4) and iopanoic acid (IOP) in aqueous samples. A brief list summarizing various recent publications dealing with analytical methods for the three substances is added to this section. For perchlorate, the method of choice is certainly ion chromatography because of its sensitivity (Ellington and Evans, 2000; Jackson *et al.*, 2000). Information about suitable methods for IOP analysis is scarce (Andeejani *et al.*, 1994; Jones, 1994). Analysis of low T4 concentrations in water samples may require solid phase extraction followed by HPLC, LC-MS/MS, GC-MS or direct measurements of T4 by means of radioimmunoassays (Burman *et al.*, 1981; De Brabandere *et al.*, 1998; Tai *et al.*, 2002; Thienpont *et al.*, 1994; Van Blerk *et al.*, 1996).

Recommendations for Perchlorate Water Analyses (US EPA)

Perchlorate was analyzed in aqueous samples using a sensitive ion chromatographic method (Jackson et al. 2000; U.S. EPA 1999). Water samples (20 ml) were collected from the exposure chambers, placed into pre-cleaned 30 ml HDPE bottles and stored at 4 C (holding time, 28 d). At the time of analysis, water samples were warmed to room temperature, diluted if necessary with deionized water, placed into 5 ml polyvials and closed with filter-caps. Water samples were analyzed using a Dionex ion chromatography system (Sunnyvale, CA). Anion separation was performed by injecting an aliquot of sample (1000 l for 10 to 400 g/l or 100 l for 250 to 5000 g/l) into a Dionex IonPac AS16 (250 x 4.0 mm) analytical column equipped with an AG16 (50 x 4.0 mm) guard column. The column was eluted isocratically with 50 mM NaOH solution at a flow rate of 1.0 ml/min. Perchlorate was detected by suppressed conductivity detection using a Dionex ASRS®-ULTRA suppressor, operated at 300 mA in the external water mode. The run time for each injection was 20 min. and the perchlorate retention time was 13.3 min. Data were collected with Dionex Peaknet software and perchlorate concentrations were determined using the external standard method of quantification with a 7 point linear calibration curve. Routine quality assurance analyses (e.g., blanks, matrix spikes and duplicate samples) were conducted with each sample set.

References

Jackson, P.E., S. Gokhale, T. Streib, J.S. Rohrer and C.A. Pohl. 2000. Improved method for the determination of trace perchlorate in ground and drinking waters by ion chromatography. J. Chromatogr. A. 888:151-158.

U.S. EPA. Method 314.0. Determination of perchlorate in drinking water using ion chromatography, U.S. EPA, Cincinnati, OH, November, 1999.

Iopanoic Acid (IOP) Water Analysis

An iopanoic acid (IOP) analytical stock solution was prepared by dissolving 10 mg of IOP into 10 ml of ethanol (i.e., IOP stock concentration of 1000 g/ml). Analytical calibration standards (0.01, 0.05, 0.1, 0.5, 1.0, 5.0 and 10.0 g IOP/ml) were prepared by diluting the stock solution into 50% acetronitrile:50%

water.

Water samples (1 ml) were removed from the exposure chambers (0, 0.023, 0.094, 0.375, 1.5 and 6 g IOP/ml), placed into HPLC vials and crimp-sealed. The samples were immediately analyzed using a Beckman HPLC system. An aliquot of sample (100 l) was injected directly into a Hypersil ODS, C18, 5 (250 x 4.6 mm) analytical column (Alltech Associates, Deerfield, IL, USA). The column was eluted at 1.5 ml/min with mobile phase A (10% acetonitrile:90% water) and mobile phase B (80% acetonitrile:20% water) under isocratic conditions (10%A:90%B, equivalent to 73% acetonitrile). IOP was detected by diode-array detection at 231 nm. Data were collected using Beckman HPLC software and IOP concentrations were determined using the external standard method of quantification with a 7 point linear calibration curve. Routine quality assurance analyses (i.e., blanks, matrix spikes, duplicate samples) were conducted with each sample set.

Thyroxine (T4) Water Analysis by HPLC-ICP-MS

Water samples (i.e., 25 or 50 ml) were collected from the exposure chambers (0, 0.25, 0.5, 1.0, 2.0, 4.0 ng T4/ml) and placed into amber bottles. The samples were acidified with concentrated phosphoric acid (i.e., 0.5 ml H_3PO_4 added to 25 ml H_2O) and immediately concentrated by processing through solid phase extraction (SPE). Briefly, 3-ml (60 mg) Strata X SPE columns (Phenomenex, Torrance, CA) fitted with 50-ml reservoirs, were conditioned with 2- ml methanol and rinsed with 5-ml deionized water. The sample was then applied to the column under vacuum (5 mm Hg). Following sample addition, the column was washed with 2- ml of 5 % methanol in phosphate buffer (10 mM, pH 3) and dried for 30 s under full vacuum. T4 was slowly eluted from the column under vacuum (1 mm Hg) with six 0.5-ml aliquots of methanol. The eluate was evaporated, using nitrogen and a heated water bath (37 C), to a known volume (i.e., 0.5 or 1.0 ml). The concentrated extracts were mixed with a vortex mixer and then sonicated in a water bath for 10 min. The methanol extracts were stored in the refrigerator for approximately 3 to 4 weeks prior to instrumental analysis.

At the time of T4 analysis, the SPE methanol extracts were removed from the refrigerator and sonicated for 5 min. An aliquot (100 l) of the methanol extract was removed and placed into an amber HPLC vial containing a microinsert. An equal volume (100 l) of phosphate buffer (10 mM, pH 3) was added to the vial and the contents were mixed. The samples were then analyzed by HPLC-ICP-MS. For initial analyte separation, an Agilent 1100 series HPLC system (Wilmington, DE) was used. The HPLC system was equipped with a thermostatically controlled autosampler set at 4 C. An aliquot of sample (125 l) was injected directly into a Synergi Hydro RP, 4 , (50 x 2.0 mm) column (Phenomenex, Torrance, CA) that was maintained at 23 C. T4 was separated at a flow rate of 0.25 ml/min using a binary gradient program (Table 1) of mobile phase A, phosphate buffer (10 mM, pH 3) and methanol (50:50 v/v) and mobile phase B, phosphate buffer (10 mM, pH 3) and methanol (10:90 v/v). The total run time per sample was 30 min. The HPLC column effluent was connected to a high pressure micro-splitter valve (UpChruch Scientific, Oak Harbor, WA) where it was mixed with a post column reagent (10 % HNO_3:90 % H_2O) delivered from an isocratic pump at 0.55 ml/min. The post column reagent was used to reduce the methanol concentration introduced into the ICP-MS (Michalke et al. 2000). The combined HPLC effluent and post column reagent (combined flow of 0.80 ml/min) was connected directly to the nebulizer of a Varian UltraMass ICP-MS system (Mulgrave, Victoria, Australia) for subsequent iodine detection.

The ICP-MS was equipped with a Sturman-Masters spray chamber and a demountable torch with a small bore (0.8 mm i.d.) injector tube. The plasma, auxiliary and nebulizer argon gas flow rates were 15.5 L/min, 1.20 L/min and 0.80 L/min, respectively. The ICP-MS was operated in the time resolved mode and iodine was determined at m/z of 127 with a scan time of 100 msec. The time resolved data was imported into GRAMS/32 AI (version 6.00) chromatography software (Thermo Galactic, Salem, NH) for chromatogram generation, peak smoothing and integration. Microsoft Excel 97 (Redmond, WA) was then used to determine iodine concentrations (i.e., ng I/ml) for T4 using the external standard method of quantification with a 6 point linear calibration curve (i.e., 2.5, 5, 10, 20, 40, 60 ng I/ml). Iodine concentrations were converted and reported as concentrations of T4 (ngT4/ml). The recovery of T4 in the spiked water samples

was 83.9 % 3.0 (n = 4).

Table 1. T4 HPLC Gradient Conditions

Time (min)	%A[1]	%B[2]	Flow (ml/min)
0.0	99	1	0.25
0.3	99	1	0.25
0.31	1	99	0.25
16.0	1	99	0.25
16.01	99	1	0.25
30.0	99	1	0.25

[1] Mobile phase A, phosphate buffer (10 mM, pH 3) and methanol (50:50 v/v)
[2] Mobile phase B, phosphate buffer (10 mM, pH 3) and methanol (10:90 v/v)

Reference

Michalke, B., P. Schramel and H. Witte. 2000. Method developments for iodine speciation by reversed-phase liquid chromatography-ICP-MS Spectrometry. *Biol. Trace Element Res*. 78, 67-79.

Additional T4 Information

L-Thyroxine (T4) free acid form
$C_{15}H_{11}I_4NO_4$
CAS Number: [51-48-9]
Vendor: Sigma
Order Number: T2376
FW of T4: 776.9
FW of 4I: 507.5

Prepare T4 analytical superstock at 153 g T4/ml (100 g I/ml) in 50 mM Na_2CO_3 (pH 11)

Prepare T4 analytical secondary stock at 15.3 g T4/ml (10 g I/ml) in 50 mM Na_2CO_3 (pH 11)

Store T4 analytical stocks in teflon bottles and keep refrigerated (stocks stable for several months)

On the day of HPLC-ICP-MS analysis, prepare HPLC calibration standards (2.5 to 60 ng I/ml) by diluting T4 secondary stock with a solution of phosphate buffer (10 mM, pH 3) and methanol (50:50 v/v).

Thyroxine (T4) Analysis by HPLC-DAD

Mobile Phase A: phosphate buffer (10 mM, pH 3) and methanol (99:1, v/v)
Mobile Phase B: phosphate buffer (10 mM, pH 3) and methanol (50:50, v/v)

Gradient Time Table

Time (min)	%B	Flow (ml/min)
0.00	15	0.45
1.00	25	0.45
8.00	100	0.45
35.00	100	0.45

40.00	15	0.45
45.00	15	0.45

Diode-array wave length: 210 nm
Spectral data: store all in peak (collect from 190 nm to 400 nm)
Analytical column: Synergi Hydro RP, 4 (50 x 3.0 mm)
Injection volume: 110 l
Column Temperature: 23 C
Autosampler Temperature: 4 C
T4 Calibration Range: 230 ng/ml to 930 ng/ml in phosphate buffer (10 mM, pH 3) and methanol (92:8, v/v)
T4 retention time: 28.4 min

Additional References

Andeejani, A M; Hughes, H; Feuchuk, D M; Aboul-Enein, H Y (1994). Rapid assay of iopanoic acid in dog plasma using a Hisep column. Biomed. Chromatogr. 8(1), 26-28

Anderson, T.A. and Wu, T.H., (2002). Extraction, cleanup, and analysis of the perchlorate anion in tissue samples. Bull. Environ. Contam. Toxicol. 68, 684-691

L. Bhavana , V. J. Ajimon , S. L. Radhika , M. Sindhu and C. S. P. Iyer (2004): Ion chromatographic determination of thyroxine in urine. J. Chromatogr. B 803(2), 363-366

Burman KD, Bongiovanni R, Garis RK, Wartofsky L, Boehm TM. (1981): Measurement of serum T4 concentration by high performance liquid chromatography. J Clin Endocrinol Metab. 53(5):909-912.

De Brabandere VI, Hou P, Stockl D, Thienpont LM, De Leenheer AP. (1998): Isotope dilution-liquid chromatography/electro-spray ionization-tandem mass spectrometry for the determination of serum thyroxine as a potential reference method. Rapid Commun Mass Spectrom. 12(16), 1099-1103

Ellington, J.J. and Evans, J.J., (2000): Determination of perchlorate at parts-per-billion levels in plants by ion chromatography. J. Chromatogr. A 898, 193-199

Jackson, P.E., M. Laikhtman and J.S. Rohrer. (1999): Determination of trace level perchlorate in drinking water and ground water by ion chromatography. J. Chromatogr. A 850, 131-135

Jackson, P.E., S. Gokhale, T. Streib, J.S. Roher and C.A. Pohl. (2000): Improved method for the determination of trace perchlorate in ground and drinking waters by ion chromatography. J. Chromatogr. A 888, 151-158

Jones, A (1994): High-performance liquid chromatographic determination of iophenoxic acid in serum (IOP). J. Chromatogr. B, Biomedical Applications, 654(2), 293-296

Liu, Y. and Shifen Mou (2003): Determination of trace levels of haloacetic acids and perchlorate in drinking water by ion chromatography with direct injection. J. Chromatogr. A 997, (1-2), 225-235

Roehl, R., R. Slingsby, N. Avdalovic and P. E. Jackson (2002): Applications of ion chromatography with electrospray mass spectrometric detection to the determination of environmental contaminants in water. J. Chromatogr. A 956(1-2), 245-254

Smith, P.N., Theodorakis, C.W., Anderson, T.A. and Kendall, R.J., (2001): Preliminary assessment of perchlorate in ecological receptors at the Longhorn Army Ammunition Plant (LHAAP), Karnack, Texas. Ecotoxicology 10, 305-313

Tai, S.S., Sniegoski LT, Welch MJ. (2002): Candidate reference method for total thyroxine in human serum: use of isotope-dilution liquid chromatography-mass spectrometry with electrospray ionization. Clin. Chem. 48(4), 637-642

Thienpont, L.M., De Brabandere VI, Stockl D, De Leenheer AP. (1994): Development of a new method for the determination of thyroxine in serum based on isotope dilution gas hromatography mass spectrometry. Biol Mass Spectrom. 23(8), 475-482

Thienpont LM, Fierens C, De Leenheer AP, Przywara L. (1999): Isotope dilution-gas chromatography/mass spectrometry and liquid chromatography/electrospray ionization-tandem mass spectrometry for the determination of triiodo-L-thyronine in serum. Rapid Commun Mass Spectrom. 13(19), 1924–1931

Tian, K., Dasgupta, P.K. and Anderson, T.A., (2003): Determination of trace perchlorate in high-salinity water samples by ion chromatography with on-line preconcentration and preelution. Analytical Chemistry 75, 701-706

Urbansky, E.T., (1998): Perchlorate chemistry: implications for analysis and remediation. Bioremediation Journal 2, 81-95

Urbansky, E.T., M.L. Magnuson, D. Freeman and C. Jelks. (1999): Quantitation of perchlorate ion by electrospray ionization mass spectrometry (ESI-MS) using stable association complexes with organic cations and bases to enhance selectivity. J. Anal. At. Spectrom. 14, 1861-1866

Urbansky, E.T. (2000): Quantitation of Perchlorate Ion: Practices and Advances Applied to the Analysis of Common Matrices. Crit. Rev. Anal. Chem. 30(4), 311-343

Van Blerk, M; Smitz, J; Rozenski, E; Mees, M; Roelandt, P; Laermans, L; Callewaert, M; Van Steirteghem, A C (1996): Four radioisotopic immunoassays of free thyroxine compared. Ann. Clin. Biochem. 33(4), 335-343

Wagner, H.P., F. Xavier Suarez, Barry V. Pepich, Daniel P. Hautman and David J. Munch (2004): Challenges encountered in extending the sensitivity of US Environmental Protection Agency Method 314.0 for perchlorate in drinking water. J. Chromatogr. A 1039(1-2), 97-104

Annex 4 Histological Analysis of Thyroid Gland

Results from validation Phase 1 showed that histological analysis of the thyroid gland represents one of the most sensitive endpoints of the assay for the detection of anti-thyroidal activities of test chemicals. Different protocols for histological analysis have been used by the participating laboratories during validation Phase 1. For Phase 2 of validation, tissue processing procedures are standardized and a structured assessment scheme will be developed to ensure sensitivity and consistency of histopathological evaluations among different laboratories. In addition, morphometric techniques (optional measurements) will be further optimized. In particular, the endpoint selection for quantitative measurements was refined based on the effects pattern observed in validation Phase 1.

Tissue Sampling

Tissue samples for histopathological analysis of the thyroid gland are collected at two time points during the exposure experiments. At day 14 of exposure, a subsample of 5 randomly selected tadpoles per treatment tank will be analyzed for histological changes in the thyroid gland. At test termination (day 21), another 5 tadpoles from each treatment tank will be randomly selected for thyroid gland histopathology. At both exposure time points, selected tadpoles are anesthetized in 2000 mg/L MS-222. Tadpoles are then rinsed in water and blotted dry for determination of body weight, developmental stage, body length and hind limb length.

After apical endpoint measurements, head tissue including the lower jaw will be dissected as follows. Joe, could you please describe your methodology of decapitation and the location of the cut. You can also use the Figure 2 which I provided in the main SOP portion to locate the site where the cut will be made.

Fixation and preservation of tissues for histological analysis

The dissected tissue samples are placed immediately in Davidson's fixative solution for 48 hours. After fixation, tissue samples are briefly rinsed in water and are stored in neutral buffered formalin (4% formaldehyde) until further processing.

Dehydration and embedding of tissues

For the preparation of transversal sections of head tissue for light microscopy, head tissue samples are placed into tissue prep cassettes appropriate to the size of the sample. Tissues are then dehydrated in a graded alcohol series and are embedded in paraffin. The tissue samples will be embedded in the frontal plane in a manner that allows the caudal surface of the tissues to be microtomed first. A routine protocol for dehydration and embedding is provided below.

1.	80% EtOH	60 min	room temperature
2.	80% EtOH	60 min	room temperature
3.	90% EtOH	60 min	room temperature
4.	90% EtOH	60 min	room temperature
5.	96% EtOH	60 min	room temperature
6.	96% EtOH	60 min	room temperature
7.	100% isopropanol	60 min	room temperature
8.	100% isopropanol	60 min	room temperature

9.	methyl benzoate	60 min	room temperature
10.	methyl benzoate	12 hours	room temperature
11.	methyl benzoate	3 hours	room temperature
12.	paraffin I	3 hours	57°C
13.	paraffin II	12 hours	57°C

14. Samples are placed into embedding molds containing fresh molten paraplast and arranged in proper position
15. Samples are placed onto a cold plate (ensure that the tissue is kept in right position when the wax solidifies). After embedding in paraffin, the samples can be stored at room temperature until microtoming.

Microtoming and Staining

The head samples are sectioned in a transverse plane from caudal to rostral to obtain sections of 5 or 7 µm thickness. Initially, five step sections (30 microns apart) are taken from each block, and two serial sections of each step are placed on each of five slides. These slides are examined to assure that a sufficient amount of thyroid tissue is present bilaterally in at least two of the step sections. If necessary to meet these criteria, additional step sections should be cut from the block. When thyroid tissue is verified in each slide preparation, the slides are stained with hematoxylin and eosin and the sections are covered with glass cover slips. A routine protocol for staining is provided below.

Harris' Hematoxylin and Eosin Y staining

1.	orange terpene	3 x 5 min
2.	100% isopropanole	2 x 5 min
3.	96% EtOH	3 x 3 min
4.	90% EtOH	3 x 3 min
5.	80% EtOH	3 x 3 min
6.	70% EtOH	3 x 3 min
7.	H_2O	2 min
8.	hematoxylin	15 min
9.	running H_2O	10 min
10.	0.1% eosin Y	3 min
11.	H_2O	short
12.	70% EtOH	short
13.	80% EtOH	short
14.	90% EtOH	short
15.	96% EtOH	1 min
16.	100% isopropanole	2 x 5 min
17.	orange terpene	3 x 5 min

18. To coverslip, add a few drops Permount (DePeX, Serva) per slide and then carefully place cover slip before the Permount dries. Allow the Permount to harden overnight.

Histological Analysis

Qualitative Analysis

Two of the step sections from each animal are selected for evaluation of the following parameters:

- Overall size of the thyroid gland (reduction, increase)
- Follicle size (reduction, increase)
- Follicle shape (regular, irregular, uniform)
- Colloid content (increase or reduction in colloid area; absence of colloid)
- Colloid density (homogeneous or heterogeneous tinctoral quality; pale, lacy, or granular colloid; peripheral vacuolation of colloid)

- Follicular cell shape (cuboidal, columnar, tall)
- Follicular cell height (increase, reduction)
- Structure of the epithelium (single cell layer or stratification resulting in multiple layers, papillary infoldings of the epithelial cell layer into the lumen)
- Follicular cell hyperplasia

The changes are graded from 1 to 5 depending upon severity (1=minimal, 2=slight/mild, 3=moderate, 4=moderately severe, 5=severe/high). Non-gradable changes are designated as present (P).

A standardized and structured assessment scheme based on the endpoints mentioned above is currently developed and will be provided to all participants before the experimental portion of Phase 2 validation work begins.

Quantitative Measurements (Optional Endpoint Measurements)

Epithelial cell height

Photographs of sections from the right thyroid gland should be taken by high-power microscopes (magnification: 200x – 1000x) equipped with a digital camera. The height of the epithelial cells from base to apex should be measured on the digitized images (obtained at a magnification of at least x 250) using an image analyzing software after calibration of the recording system by a calibration slide.

- Epithelial cell height measurements are taken from 10 randomly chosen individuals per treatment group.
- For each specimen, 5 sections are analyzed.
- For each section, 3 follicles of the right thyroid gland are randomly selected. However, measurements should be taken from follicles that are predominated by a single epithelial cell layer in order to avoid measurement errors due to compression of epithelial cells in stratified regions.
- For each follicle, the height of 4 epithelial cells is recorded. For a selected follicle, it is recommended to take measurements on the opposite ends of two diameters intersecting at an angle of 90°.
- This measurement strategy leads to 60 values of the epithelial cell height per individual. The mean value for epithelial cell height is calculated for each specimen and is used in statistical evaluation of differences among treatment groups (n=10).

Thyroid gland – cross section area

For each treatment group, 10 tadpoles (5 tadpoles per replicate tank) are selected. Locate and select the largest cross section, usually found around the middle of section series. The area is compared with the adjacent sections to confirm the largest section. The outline of the largest section is then mapped, and the cross section area calculated using an image analyzing software after calibration of the recording system by a calibration slide. The maximum value of thyroid gland cross section area measured for each specimen is used in statistical evaluation of differences among treatment groups (n=10).

Follicle – cross section area

For each treatment group, 10 tadpoles (5 tadpoles per replicate tank) are selected. Using the largest section selected from above as the reference point, select the five serial sections before and after the largest section. This gives a total of 11 sections. The cross section area of each follicle in all the selected sections is then measured. For each section, the maximum follicle cross section area is determined. The mean value for maximum follicle cross section area is calculated for each specimen and this value is used in statistical evaluation of differences among treatment groups (n=10).