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

31-Aug-2007

English - Or. English

ENV/JM/MONO(2007)24
Unclassified

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

SERIES ON TESTING AND ASSESSMENT
Number 77

FINAL REPORT OF THE VALIDATION OF THE AMPHIBIAN METAMORPHOSIS ASSAY: PHASE 2 – MULTI-CHEMICAL INTERLABORATORY STUDY

JT03231274

English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 77

FINAL REPORT OF THE VALIDATION OF THE AMPHIBIAN METAMORPHASIS ASSAY: PHASE 2 – MULTI-CHEMICAL INTERLABORATORY STUDY

Environment Directorate

Organisation for Economic Co-operation and Development

2007

Also published in the Series on Testing and Assessment:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

No. 77, *Final Report of the Validation of the Amphibian Metamorphosis Assay: Phase 2 – Multi-chemical Interlaboratory Study (2007)*

No. 78, *Final Report of the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances. Phase 2: Testing Negative Substances (2007)*

ABOUT THE OECD

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

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

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

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

FOREWORD

This document presents the report of the Phase 2 of the validation of the amphibian metamorphosis assay. The amphibian metamorphosis assay was included in the OECD conceptual framework for the testing and assessment of endocrine disrupting chemicals, for the detection of thyroid active substances in aquatic vertebrates.

Following the Phase 1 outcome described in a separate report, a second expert meeting was held in Paris (June 2004) to discuss the results of Phase 1 and make plans for further validation work. The Phase 2 was a multi-chemical inter-laboratory study for the evaluation of the relevance and reproducibility of the assay. A single protocol was used by all participating laboratories. The proposal for Phase 2 was presented to the Validation Management Group on Ecotoxicity Testing (VMG-eco) in December 2004, together with the report from Phase 1. After a few modifications to the proposal, the Phase 2 experimental work started mid-2005 and was completed in all laboratories in 2006. Extensive work was undertaken during the data analysis, in particular for power simulations on the most important endpoints of the assay. A brief summary of the power properties investigated is available in the report. The draft report of the Phase 2 was presented to the VMG-eco in January 2007. The Task Force on Endocrine Disrupters Testing and Assessment and the Working Group of National Coordinators of the Test Guidelines Programme endorsed the final draft of the report at its 19th meeting (28-30 March 2007).

The report of the Phase 1 of the validation was also presented to the EDTA Task Force and the WNT. The Phase 1 and the Phase 2 reports have to be considered together, in a logical step-wise approach towards the validation of the assay.

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

TABLE OF CONTENTS

List of Figures

List of Tables

ACKNOWLEDGEMENT

This report is the fruit of collaboration between six laboratories: Leibniz Institute of Freshwater Ecology and Inland Fisheries (Germany), the Mid-Continent Ecology Division, National Health and Environment Effects Research Laboratory (United States EPA), Fort Laboratory (United States), Towa Kagaku Co., Research and Development Institute (Japan), Brunel University (United Kingdom), and Springborn Smithers (Switzerland). The report has been prepared mainly by Dr. Robert Opitz, Leibniz Institute of Freshwater Ecology and Inland Fisheries (Germany).

SUMMARY OF PHASE-2 STUDIES

i) This report summarizes the results from an OECD inter-laboratory study in 2005-2006 to assess the reliability of the amphibian metamorphosis assay for the detection of thyroid system-disrupting substances acting through different pathways. The Phase 1 validation study, completed in 2004, consisted in the optimisation of the protocol and exposure scenario. The Phase 2 of the validation study aimed at an inter-laboratory multi-chemical testing with an harmonised protocol. The studies were designed and directed by members of the OECD Amphibian Experts Group, mainly from Germany, Japan and the United States. Other OECD member countries also participated in the validation study, namely the United Kingdom and Switzerland. This work is undertaken under the umbrella of the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA), which was established to develop and validate new methods to identify and assess substances acting through endocrine mechanisms for the protection of human health and the environment. The validation component of the work was managed by the Amphibian Experts Group of the Validation Management Group for ecotoxicity testing (VMG-eco).

ii) Endocrine disruption by environmental chemicals is an international toxicological concern. With respect to the increasing number of substances which are suspected to act as potential endocrine disruptors, the EDTA Conceptual Framework has been proposed in 2002 for the evaluation of chemicals suspected to act on endocrine systems, mainly as (anti-)etrogens, (anti-)androgens and thyroid active substances (agonists or antagonists). While well established testing models such as the uterotrophic assay or the Hershberger assay were already developed to efficiently screen substances for possible (anti-)estrogenic and (anti-)androgenic properties, other test systems were needed for e.g. aquatic species. The amphibian metamorphosis assay has been identified as a relevant test system for thyroid active substances in Vertebrate species. 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.

iii) Based on the results of the Phase-1 studies, an improved study protocol was established for use in the current Phase-2 studies. The exposure was initiated with stage 51 tadpoles of the South African clawed frog *Xenopus laevis*. A subsampling of tadpoles (n=5 animals per tank) was performed on day 7 and exposure of the remaining test animals was continued for a total of 21 days. Various apical morphological endpoints including developmental stage, hind limb length (HLL), whole body length (WBL), snout-vent length (SVL) and body weight were determined for tadpoles sampled on day 7 and all animals on day 21. In addition, five animals per tank were sampled for thyroid histopathology on day 21. Further enhancements of the Phase-2 protocol include the use of a flow-through system for aqueous exposure studies, standardization and improved guidance on endpoint measurement practices, optimization of the experimental design (sample size, replication, number of treatments) and verification of actual test substance concentrations.

iv) Three model substances representing different modes of action on the thyroid system were used in Phase-2 studies. These included sodium perchlorate, thyroxine (T4) and iopanoic acid (IOP). The perchlorate anion (PER) is a competetive inhibitor of thyroidal iodide uptake, T4 is the native thyroid hormone and IOP is an inhibitor of iodothyronine deiodinases. The Phase-2 studies involved 6 international laboratories which performed a total of 14 exposure experiments including 5 experiments with PER, 4 experiments with T4 and 5 experiments with IOP.

v) Overall, the results of the Phase-2 studies clearly demonstrated the ability of the study protocol for the amphibian metamorphosis assay to detect each of the test substances as a thyroid system-disrupting compound. The response profiles of the various endpoints were different for the individual test substances but reproducible across laboratories. The changes in specific endpoints caused by each of the test substances were in close agreement with the expected modes of action. With regard to the differential sensitivity of endpoints to detect specific modes of action the following main conclusion could be drawn when combining the results of Phase-1 and Phase-2 studies. Assessment of morphological endpoints such as developmental stage and HLL were more sensitive and reliable to detect agonists whereas thyroid histopathology was not relevant for detection of agonists. In contrast, thyroid histopathology was much more sensitive and reliable to anti-thyroidal activities resulting from either inhibition of thyroidal iodide uptake or iodide organification. The modulator of iodothyronine deiodinase activities, IOP, caused a complex effects pattern including asynchronous development of individual tadpole tissues and mild to moderate hypertrophy of the thyroid gland. The Phase-2 results support the inclusion of the proposed endpoints as useful and relevant to allow for the detection of a wide array of modes of action.

vi) The Phase-2 results indicate that some further work is necessary to identify appropriate approaches for the analysis of data from several endpoint measurements. Specific problems include the need for a normalization of HLL for differences in body size, the appropriate consideration of inter-individual differences in developmental stage during analysis and interpretation of day 21 measurements of growth-related parameters and histopathological findings. The Phase-2 results suggest that additional statistical consultation and discussion within the Expert Group for Amphibian testing to resolve some of these issues has merit. In addition, further optimization and standardization of the study protocol is required to reduce the existing inter-laboratory differences in tadpole growth and development. A major aspect in this regard is a further standardization of the feeding regime.

vii) In summary, the validation programme for the amphibian metamorphosis assay successfully achieved the goal of demonstrating the ability of the proposed study protocol to detect thyroid system disrupters. The discussion of Phase-1 and Phase-2 results culminated in a proposal for a decision logic for the conduct and interpretation of the assay.

INTRODUCTION

The OECD Program for an Amphibian Metamorphosis Assay

1. This report summarizes the results from an OECD inter-laboratory study conducted in 2005 – 2006 to assess the utility of an enhanced protocol for the Amphibian Metamorphosis Assay to reliably detect thyroid system-disrupting substances acting via different modes of action. These studies comprise the Phase-2 of validation of the study protocol and were preceded by an inter-laboratory exercise (validation Phase-1) addressing the optimization of the initial study protocol. The Phase-2 studies were designed and directed by the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA).

2. The need for the development and validation of an *in vivo* assay for detection of thyroid system-disrupting substances arises from concerns 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 regulate 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).

3. To date, a validated assay specifically adressing disruption of thyroid function has not been available. Determination of thyroid system-related endpoints has been included in the current validation of the enhanced OECD test guideline 407 but the overall assay design has not been optimized for a screening approach to detect thyroid system disrupters. In general, endpoint measurements used in the majority of recently proposed mammalian assays pay only limited attention to the possible effects of chemicals on TH signalling in peripheral tissues, despite the recent evidence for such modes and mechanisms of action for various substances.

4. The process of postembryonic development in anuran amphibians (metamorphosis) is one of the best studied biological models for TH function and, therefore, amphibian metamorphosis offers a valuable model to investigate the effects of chemicals on various aspects of thyroid system functioning. To overcome this lack of an appropriate test system, a validation program for an amphibian metamorphosis assay to detect thyroid system-disrupting compounds has been initiated in 2002. 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 (Shi, 1999). Previous work by the laboratories participating in the OECD program for validation of the Amphibian Metamorphosis Assay and by others showed that in *X. laevis* tadpoles, metamorphic development can be precociously induced and/or accelerated by agonists of the thyroid system whereas anti-thyroidal agents delay metamorphic development (Degitz et al., 2005; Opitz et al., 2005, 2006; Tietge et al., 2005). Reproducible results have also been obtained by using the close relative *X. tropicalis* species in a similar study protocol (Mitsui et al., 2006). In addition to monitoring rates of TH-dependent development, evaluation of thyroid gland histology has been proven a very sensitive and diagnostic endpoint for detection of thyroid system disruption in *X. laevis* tadpoles, particularly for agents that affect the thyroid gland's capacity to synthesize TH (Degitz et al., 2005; Opitz et al., 2006; Tietge et al., 2005).

5. During validation Phase-1 of the Amphibian test protocol, the model anti-thyroidal compound propylthiouracil (PTU) and the native prohormone thyroxine (T4) were used as test substances to compare the utility and sensitivity of two different exposure scenarios to detect modulation of thyroid system function in *X. laevis* tadpoles. The results of the Phase-1 studies were discussed at several meetings and teleconferences of the OECD *ad hoc* expert group on amphibian testing in 2004 and a summary report was presented to the Validation Management Group for Ecotoxicity Testing (VMG-eco) in 2004. While the Phase-1 studies utilized the same test substances, there were still variations in the exposure and analysis protocols used by the participating laboratories. The OECD *ad hoc* Expert Group on Amphibian Testing reviewed the results of Phase-1 and concluded that an exposure protocol comprising a 21-day treatment of initial stage 51 *X. laevis* tadpoles (Nieuwkoop and Faber, 1994) provides for enhanced sensitivity to detect both agonistic and antagonistic effects on the thyroid system.

Proposal for Phase-2 Studies

6. The Expert Group prepared a proposal for experimental activities to be included in a subsequent Phase-2 validation exercise and this proposal was reviewed in November 2004 by the VMG-eco. Comments raised by VMG-eco were addressed in a first teleconference about the Phase-2 study plan in December 2004. Additional activities to enhance the proposed study protocol included a statistical consultation on the effects of replication, treatment number and sample size on the power of several statistical tests in January 2005 and a special workshop to develop a streamlined grading system for use in the histological evaluation of *X. laevis* thyroid glands in January 2006. Together, these activities led to several enhancements and optimizations in experimental protocols to be used in Phase-2 studies and resulted in a more precise and uniform study protocol. Specific improvements in the Phase-2 protocol included 1) the utilization of flow-through exposure systems, 2) analytical verification of test chemical concentrations, 3) optimization of the statistical design of the exposure studies, and 4) further standardization of subsampling procedures and endpoint measurement techniques. The revisions of some key aspects of the initial Phase-2 protocol are briefly outlined below.

Experimental Design

7. Using data obtained during the Phase-1 experiments, extensive power analyses were conducted by John Green of Dupont and Tim Springer of Wildlife International to derive an appropriate number of test animals per replicate tank, the number of test animals required for a subsampling on study day 7 and the number and allocation of replicate tanks per treatment group. The results of these analyses were discussed among members of the Expert group and participants of Phase-2 studies at a teleconference in January 2005 in the context of practical limitations associated with conducting the experimental studies. It was decided to utilize equal allocation of treatments and controls and that the practical limit of replication is four. For this scenario, the power analyses demonstrated that the Jonckheere-Terpstra and Williams' tests, each applied in step-down manner, would provide sensitive and viable statistical procedures for analysis of test results.

8. Another important statistical issue was to decide about the sample size utilized for the subsampling on study day 7. Power analyses were conducted based on results from day 7 measurements during Phase-1 studies and evaluated the effect of different sample sizes on the power of different statistical tests using different number of replicates and allocation strategies. Assuming equal allocation of four replicates in a five treatment study (control + 4 chemical treatments), the power analyses suggested that a sample size of five animals per replicate tank does not impair the statistical power of the study protocol. Therefore, a sample size of five animals per replicate tank was considered for the day 7 subsampling of the Phase-2 study protocol.

9. Determination of developmental stages on day 21 is a core endpoint of the study protocol to detect deviations in TH-dependent metamorphic development in response to test chemical treatment. Power calculations suggested that reducing the number of organisms for developmental stage determination from a sample size of 25 to 15 animals per replicate tank does not dramaticaly reduce the statistical power of the protocol. Therefore, a sample size of 15 was used in the Phase-2 study protocol for the day 21 sampling and accordingly, the number of tadpoles per replicate tank at test initiation was set at 20 animals.

10. For the Phase 2 studies, it was thought to be appropriate to collect as much data as possible to be able afterwards to re-run power simulations to find an adequation between the minimum number of animals necessary, a practical test design, and an acceptable power of the test.

Standardization of Endpoint Measurement Techniques

11. The study protocol calls for measurements of several apical morphological endpoints including whole body length (WBL), snout-vent length (SVL), hind limb length (HLL), wet weight and developmental stage. To harmonize and standardize the methodologies for morphometric length measurements in different labs, a technical guidance document, ***Histopathology Guidance Document Part 1: Technical guidance for morphologic sampling and histological preparation*** was prepared by Christiana Grim (OSCP/EPA). This document detailed the methods for measuring body and limb lengths as well as the aquisition, fixation, and embedment of tissues for subsequent thyroid histology. This guidance document called for documentation of all test animals by digital photographs and subsequent computer-assisted image analyses of various morphological parameters.

Standardization of Histopathological Analyses of Thyroid Tissue

12. Since thyroid histology is a key diagnostic endpoint of the study protocol, efforts were undertaken to standardize the methods used to obtain, process, and analyze thyroid tissue sections in order to improve the comparability of the information. In addition to the technical guidance document part 1 (see 0), a second document, ***Histopathology Guidance Document Part 2: Approach to reading studies, diagnostic criteria, severity grading, and atlas,*** was prepared by Christiana Grim (OSCP/EPA) and drafted in Summer 2005. This document presents general guidelines for study reading practices, detailed diagnostic criteria and severity grading schemes for various observations, a reference atlas of normal microanatomy of *X. laevis* thyroid glands, an atlas of core diagnostic criteria with examples of tissues exhibiting effects with different severity grades, and a recommendation for uniform data recording and compilation.

13. In addition, a workshop "Thyroid Histopathology Consultation" (January 30 and 31, 2006) was hosted by EPA to discuss various issues associated with interpreting amphibian thyroid gland histopathology. Participants of this meeting noted that the initial grading scheme proposed in the ***Histopathology Guidance Document Part 2*** is overly complicated and proposed a more streamlined approach which places a focus on the evaluation of three different measures: increase in the number of cells (follicular cell hyperplasia), increase in the size of the cell itself (follicular cell hypertrophy), and changes in the size of the gland itself (thyroid gland atrophy or hypertrophy). As an important outcome of this workshop, a revised simplified grading system was established and the revised protocol for reading of thyroid tissue sections was circulated in March 2006 to all participating laboratories.

Identification of Test Substances

14. The Expert Group on Amphibian Testing proposed three test substance to be used in Phase-2 studies. The substances thyroxine (T4), sodium perchlorate (Na-PER), and iopanoic acid (IOP) were selected based on the fact that they represent different modes of action relevant to the thyroid system. T4, the native prohormone, was chosen to provide a compound with agonistic activity. It should be noted that T4 was also utilized in the concurrent evaluation of the enhanced protocol of the OECD Test Guideline 407 to detect thyroid system modulation in a mammalian test model. Na-PER was selected as a model inhibitor of the sodium-iodide symporter (NIS), a membrane protein mediating the active uptake of iodide by thyroid follicular cells. A decrease in iodide uptake is known to impair the capacity of the thyroid gland to produce TH and, thus, represents an anti-thyroidal mode of action that differs from previously tested compounds such as PTU (an inhibitor of iodine organification). The third test substance was IOP, a compound known to inhibit the activity of all three iodothyronine deiodinases (type I, II, and III). The developmentally regulated and very tissue-specific activities of iodothyronine deiodinases represent an important regulatory mechanism to finely adjust the levels of the biologically active TH, triiodothyronine (T3), in target tissues of TH action. The disruption of the coordinated action of various monodeiodinase enzymes is expected to cause local hyper- and hypothyrodism in a tissue-specific fashion. Among the three substances, IOP was, thus, expected to produce the most complex effects pattern.

METHODS

15. Validation phase-2 of the Amphibian Metamorphosis Assay consisted of a multi-chemical, multi-laboratory study with a total of 14 experimental studies being performed in six international laboratories. A list of participating institutions is provided in Table 1.

Table 1. List of participating institutions in the validation Phase-2 study
Throughout this report, each laboratory is referred to only by a code number, not representing the order of listing shown below.

- U.S. EPA, Duluth, Minnesota, USA (lead laboratory)
- Brunel University, UK
- Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany
- Towa Kagaku Co., LTD, Hiroshima, Japan
- Fort Environmental Laboratories, Stillwater, Oklahoma, USA
- Springborn Smithers Laboratories, Horn, Switzerland

Experimental Protocols Used in Phase-2

16. Table 2 provides an overview of the Phase-2 protocol. Below, some major aspects of the study protocol are briefly summarized, while a detailed description of the study protocol is given in Annex 1 (draft SOP for Phase-2 studies), Annex 2 (Technical guidance document – Part 1), Annex 3 (Technical guidance document – Part 2), and Annex 4 (Report from the Histopathology Workshop 2006).

General Study Design

17. The general exposure protocol for Phase-2 studies included a 21-day treatment of tadpoles of the South African clawed frog *Xenopus laevis* with four concentrations of the test substance plus a dilution water control (DWC). Each of the resulting five experimental groups consisted of four replicate tanks (equal allocation of tanks to treatments and control). Chemical treatment was accomplished by aqueous exposure of tadpoles in a flow-through system (flow-rate: 25 ml/min) and test chemical concentrations were verified by analytical chemistry. A battery of different apical morphological endpoints was analyzed after 7 and 21 days of exposure, and histological analysis of thyroid gland tissue was performed with samples obtained after 21 days of exposure.

Table 2. Summary of main testing conditions and endpoint measurements used in the study protocol for Phase-2 validation of Amphibian Metamorphosis Assay

Abbrevations: T4 – thyroxine; PER – perchlorate anion; IOP – iopanoic acid; WBL – whole body length; SVL – snout-vent length

Test animal		*Xenopus laevis*	
Exposure period		Exposure from stage 51 for 21 days * * (subsampling of 5 tadpoles per replicate tank on day 7)	
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	65, 125, 250, 500 µg/L of perchlorate	
	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 (all), 7 (subsample), and 21 (all)
		Body length (WBL and SVL)	day 7 (subsample), and 21 (all)
		Hind limb length	day 7 (subsample), and 21 (all)
		Wet weight	day 7 (subsample), and 21 (all)
		Mortality	daily observation
		Thyroid histology	day 21 (5 tadpoles per replicate tank)
Control		One dilution water control	
Larval density at test initiation		20 tadpoles per replicate tank (*Note*: 15 tadpoles per tank after day 7)	
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		4 replicates per test concentration (equal allocation of tanks to treatments)	
Acceptable mortality rate		< 5% in controls	
fixation for his-tology (day 21)	Number	5 individuals per replicate tank (randomly selected)	
	Region	whole body	
	Fixation fluid	Davidson's Fixative	
Feeding	Food	Sera micron	
	Frequency/amount	start with at least 600 mg/day/tank, then adjusted with growth 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	
Chemical analysis of test concentrations		Once a week	

Test Conditions

18. Treatment of tadpoles with the test substance was initiated at premetamorphic stage 51. Note, that throughout this report, determination of developmental stages of the test organisms was conducted according to the staging criteria of Nieuwkopp and Faber (1994). At test initiation, a total of 20 animals at stage 51 were placed in each of four replicate tanks per treatment group, so that a total of 400 animals were used for each study. All tests were conducted by using flow-through exposure systems with a test solution flow rate of 25 ml/min to each tank. The standard operating procedure called for a water temperature of 22±1°C, a pH of the test solutions between 6.5 – 8.0 and a 12:12 h light:dark photoperiod. Feeding was standardized with regard to the type of food (Sera Micron®) but less so with regarding to precise daily amount of food given to the tadpoles. A recommendation was to start with a daily food amount of at least 600 mg/tank/day and subsequent adjustment of daily food amount with age and body size of test animals. Aeration of treatment tanks was required when dissolved oxygen concentrations fall below 40% of the air saturation value.

Sampling Schedule

19. The Phase-2 study protocol included a subsampling of animals after 7 days of exposure for analysis of several apical morphological endpoints including developmental stage, hind limb length (HLL), whole body length (WBL), snout-vent length (SVL), and wet weight. Five animals were randomly removed from each treatment tank, while exposure was continued for the remaining test organisms until test termination on study day 21.

20. The exposure studies were terminated after 21 days of treatment. At test termination, all remaining test animals were analyzed for several apical morphological endpoints (stage, HLL, WBL, SVL, wet weight). In addition, five animals were randomly selected from each replicate tank for histological analysis of thyroid tissue.

21. Sampled animals were anesthetized in MS-222 and photographs were made with digital camera systems to allow for subsequent computer-assisted image analyses of various morphological parameters (HLL, WBL, SVL). Animals were then blotted dry and weighed to the nearest milligram. Tadpoles selected for thyroid histology were preserved in Davidson's solution.

Endpoint Measurements

22. Determinations of the growth-related parameters WBL, SVL and wet weight were made for all test organisms sampled on study day 7 and for all animals surviving until test termination on study day 21. The objective of these measurements was to compare their utility and sensitivity to provide information about treatment-related alterations in tadpole growth. Computer-assisted image analysis was used for quantitative analysis of WBL and SVL.

23. Determinations of the developmental parameters, stage and HLL, were peformed for all test organisms sampled on study day 7 and for all animals surviving until test termination on study day 21. The objective of these measurements was to compare the utility and sensitivity of these two endpoints to provide information about treatment-related alterations in TH-dependent development of tadpoles. Computer-assisted image analysis was used for quantitative analysis of HLL.

24. Histological assessment of thyroid gland tissue was performed for a subset of five tadpoles randomly selected from each replicate tank on study day 21. For this purpose, tadpole tissues were fixed in Davidson's solution, embedded in paraffin and sectioned at 5 µm thickness. Sections were stained with

hematoxylin and eosin and histological evaluation was performed according to a graded scoring system considering incidence and severity of selected histological alterations.

Test Substances

25. The three test substances used in Phase-2 experimental work were thyroxine (CAS-number: 6106-07-6), sodium perchlorate (CAS-number: 7601-89-0) and iopanoic acid (CAS-number: 96-83-3). All three test chemicals were supplied by the Central Chemical Repository of US Endocrine Disruptor Screening Program (Battelle Marine Sciences Laboratory, Sequim, USA). For the purpose of this study, thyroxine (T4) served as a model compound for agonistic effects, the perchlorate (PER) moiety of sodium-PER is known to act as a potent inhibitor of thyroidal iodide uptake, and IOP was regarded a model compound causing complex alterations in extrathyroidal TH metabolism due to inhibitory effects on monodeiodinases. The test concentrations of T4, PER and IOP (see Table 2) were derived from previous experimental work conducted in different laboratories (PER, IOP) and from studies performed during Phase-1 validation (T4). During the Phase-2 study, a total of five tests were performed with PER and IOP, respectively, and four tests were performed with T4 (see Table 3).

Table 3. Summary of experimental work conducted in six participating labs during validation Phase-2

Test substance	lab 1	lab 2	lab 3	lab 4	lab 5	lab 6
PER	x	x	x	x	x	-
T4	x	x	x	-	x	-
IOP	x	x	x	-	x	x

x –conduct of experimental work with this test substance, - no study performed with this substance

Statistical analysis

26. The statistical protocol used for analysis of Phase-2 data is derived from a statistical consultation with John Green of Dupont and Tim Springer of Wildlife International. A statistical flow chart is provided to indicate their expert opinion of an appropriate analysis protocol (Figure 1). This protocol is consistent with the OECD Guidance Document on the Statistical Analysis of Ecotoxicity Experiments. Some parts of this flow chart do not apply to the current datasets: There was no solvent control and there were no massive ties in the data, except for the replicate median developmental stage data. The analysis deviated from this flow chart for developmental stage by using standard Jonckheere-Tersptra and Mann-Whitney tests rather than the exact permutation versions. The reason is that in the earlier power study, there appeared to be little difference in the power properties for the exact and standard versions of these tests for experimental design used in these studies. Neither the statistical analysis nor the current power study considered exact permutation methods.

27. The recommended statistical protocol is to use the step-down Jonckheere-Terpstra test unless there is compelling evidence of a non-monotone dose-response. No such evidence was found. An alternative to the Jonckheere-Terpstra test is Williams' test, also applied in step-down fashion. Unlike The Jonckheere-Terpstra test, Williams' test requires the data be normally distributed with homogeneous variances. Fortunately, with only two exceptions, these conditions were satisfied. Where pairwise tests were desired, the Dunnett test was most powerful and appropriate, provided the data were normally distributed with homogeneous variances. Where these conditions were not satisfied, a transformation should be sought that normalizes the data and stabilizes the variances. If the data are normally distributed but heterogeneous and no variance stabilizing transform could be found, a robust version of Dunnett's test, referred to as the Tamhane-Dunnett test requiring only normality but not variance homogeneity was used.

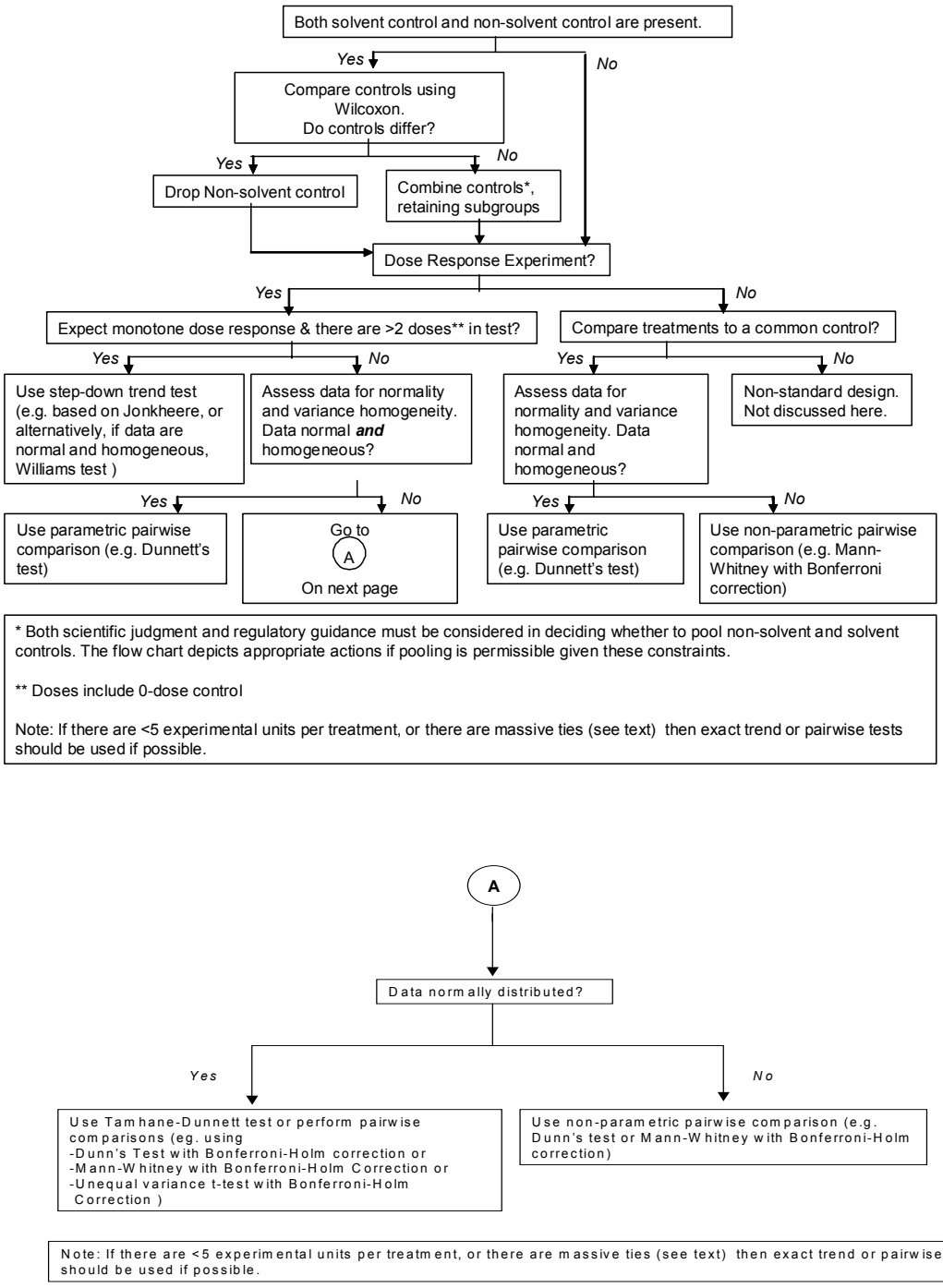

Figure 1. Statistical flow chart describing the recommended statistical protocol for Phase-2 studies

28. Where the data could not be normalized, Dunn's test was the choice among those considered in the accompanying power analysis. It is also an alternative to consider in the normal, heterogeneous case, but it has lower power than the Tamhane-Dunnett test.

29. Developmental stage is an inherently non-normal, even non-continuous response and parametric analyses such as the Williams and Dunnett tests are not applicable, nor will any transformation alter that basic fact. Thus Dunnett and Williams tests were not used in the analysis of developmental stage. The Mann-Whitney test with a Bonferroni-Holm adjustment to the *p*-values proved to have extremely low power when there are three or fewer replicates, but has more reasonable power with four or more replicates per treatment. However, no significant effects were found on any response, from any lab, on any chemical using the Mann-Whitney test. Thus, analysis was restricted to the Jonckheere-Terpstra, Dunn and Mann-Whitney tests. The issue of individual frogs versus tanks as the unit of analysis is the same for this response as for the other four. However, since developmental stage is measured on an ordinal scale, not a ratio scale, analysis is based on the replicate median rather than the replicate mean.

30. Regarding the statistical analysis of data from the severity grade scoring of thyroid tissue, a specific statistical approach is currently under development.

RESULTS

Control Animal Performance

31. In the context of the Amphibian metamorphosis assay, it is of particular importance to ensure comparable rates of tadpole growth and development across different laboratories because the developmental changes in body size, weight and morphology are important hallmarks to assess a chemical's ability to affect TH-dependent metamorphosis. In addition, the histological appearance of the thyroid gland, another major endpoint, can vary depending on the developmental stage and thus, evaluation of treatment-related changes in thyroid histology is expected to yield more consistent results when tadpole development is homogenous within a test population and comparable among different labs.

32. Experiments conducted during the validation Phase-1 of the Amphibian metamorphosis assay and additional work by various labs participating in these validation exercises (Degitz et al., 2005; Opitz et al., 2006; Tietge et al., 2005) showed that optimization and standardization of test conditions allows to obtain comparable rates of development and growth between labs leading to the generation of highly reproducible test results across different labs. Consultation among study Phase-1 participants and discussion of Phase-1 results at the OECD *ad hoc* Expert Group meeting in Paris 2004 suggested that further standardization of general test conditions and endpoint measurement techniques could improve the inter-laboratory reproducibility of the study protocol.

33. Another important issue was to enhance the sensitivity of the protocol not only by improved endpoint measurements but also by minimizing the variability of individual growth and developmental rates within the population of test animals. Given these considerations, an initial assessment of the variability of various developmental and growth-related parameters was performed to evaluate the individual experiments performed during Phase-2 towards their compliance with an array of recommended validity criteria.

Development in the Control Group

34. The study protocol uses tadpoles at developmental stage 51 for exposure initiation. A first criterion of normal test animal performance is the time required from fertilization to developmental stage 51. According to Nieuwkoop and Faber (1994) and experimental experience from several labs, *X. laevis* tadpoles should reach stage 51 within approximately 14 days postfertilization. For an interlaboratory comparison, Table 4 summarizes the age of the tadpoles at study day 0 in the different experiments of Phase-2.

Table 4. Age of test animals on study day 0 (exposure initiation)

		Age at test initiation (post fertilization day)					
study		**lab 1**	**lab 2**	**lab 3**	**lab 4**	**lab 5**	**lab 6**
IOP		13	11-13	14	n.s.	11	17
PER		12	11-13	14	-	16	n.s.
T4		13	11-13	14	n.s.	14	n.s.

Note: - data not available; n.s. no study conduced

35. A second criterion is the developmental stage reached by the control animal population within the initial 7 days of exposure. Under optimal rearing conditions, control tadpoles should reach early prometamorphic stages 54-55 until study day 7. At stages 54-55, tadpoles should have a WBL of > 35 mm and a wet weight of >300 mg. An increased number of control tadpoles that show lower developmental and growth rates is expected to indicate suboptimal testing conditions. Table 5 summarizes the stages observed across all four replicate tanks of the control group on study 7 in a total of 14 tests performed during validation Phase-2 experimental work.

36. Analysis of the stage distribution on day 7 indicated that two experiments were conspicious for sub-optimal animal performance. The observation that 75% of control animals showed development to only stage 53 within 7 days in the IOP-experiment by lab 3 suggests a relatively slow development, whilst the increased range of stages (4 stages) determined in the controls of the T4 experiment in lab 3 indicates an increased heterogeny of developmental stages already at an early time point during this experiment.

Table 5. Results from developmental stage determination in the control groups on study day 7

			developmental stage				stage	stage
lab	**test**	***n***	**52**	**53**	**54**	**55**	**(median)**	**(range)**
1	IOP	**20**			4	16	55	2
1	T4	**20**			5	15	55	2
1	PER	**20**			8	12	55	2
2	IOP	**20**			9	11	55	2
2	T4	**20**			19	1	54	2
2	PER	**20**		2	17	1	54	3
3	IOP	**20**		15	2	3	53	3
3	T4	**20**	1	2	9	8	54	4
3	PER	**20**		2	17	1	54	3
4	PER	**20**		1	6	13	55	3
5	IOP	**20**		1	10	9	54	3
5	T4	**20**		1	12	7	54	3
5	PER	**20**		9	11		54	2
6	IOP	**17**		1	14	2	54	3

Note: Stage data from all four replicate tanks of the control group were combined and the absolute number of tadpoles at each specific developmental stage are shown.

Table 6. Results from developmental stage determination in the control groups on study day 21

			developmental stage												stage	stage
lab	**test**	***n***	**55**	**56**	**57**	**58**	**59**	**60**	**61**	**62**	**63**	**64**	**65**	**66**	**(median)**	**(range)**
1	IOP	**59**			14	23	19	2	1						58	5
1	T4	**60**			15	28	12	4	1						58	5
1	PER	**60**		3	18	19	11	5	3	1					58	7
2	IOP	**60**			4	21	28	4	3						59	5
2	T4	**60**				1	19	30	7	3					60	5
2	PER	**60**		2	5	36	13	4							58	5
3	IOP	**58**				10	5	6	5	3	2	8	18	1	62	9
3	T4	**60**	1	1	5	10	22	2	8	3	3	1	1	3	59	12
3	PER	**60**		3	22	1	14	4	5	5	5		1		59	9
4	PER	**60**			5	32	17	6							58	4
5	IOP	**60**		7	19	25	6	3							58	5
5	T4	**60**		14	22	17	6	1							57	5
5	PER	**60**		10	31	9	8	1	1						57	6
6	IOP	**60**			1	1	13	18	3	11	11	2			60	8

Note: Stage data from all four replicate tanks of the control group were combined and the absolute number of tadpoles at each specific developmental stage are shown.

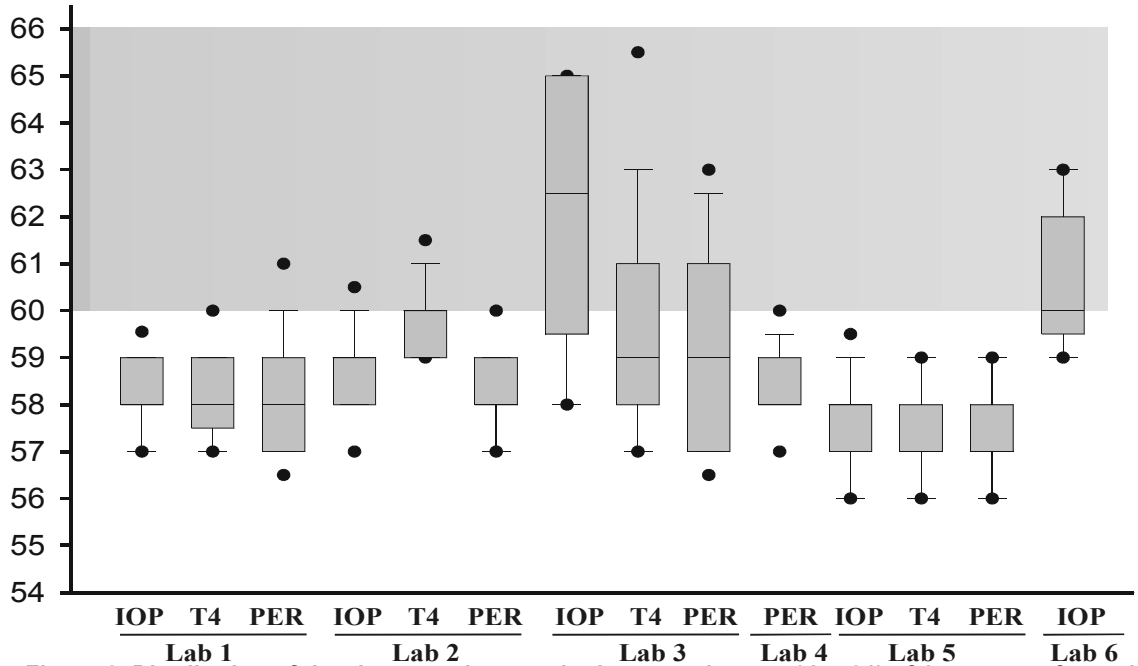

Figure 2. Distribution of developmental stages in the control group (day 21) of 14 tests performed during validation phase-2

Data within the 5% - 95% percentile interval were used for this graphical presentation. Extreme values are denoted by a black circle. The shaded area demarcates the climax stages of development.

37. A third criterion is the developmental stage reached by the control animal population on exposure day 21. Table 6 and the box plot in Figure 2 summarize the distribution of stages observed across all four replicate tanks of the control group on study 21 in a total of 14 tests performed during validation Phase-2 experimental work. From the results depicted in Figure 2, increased variability in developmental stages in the control group becomes apparent for all three experiments performed in lab 3. Of particular concern is the heterochrony in control development reported for the IOP-experiment in lab 3. Compared to all other experiments, control development in this experiment appeared very slow until day 7 (Table 5) but by day 21, ore than 50% of the control tadpoles were at late climax stages 63-66 (Table 6) and thus at much more advanced stages compared to all other experiments (Table 6). In the T4-experiment from lab 3, staging data from day 12 indicate a very high variability in individual developmental rates given that stages of control animals on day 21 are spread over a range of 12 different stages from early prometamorphic stage 55 to completion of metamorphosis at stage 66 (Table 6).

38. Feeding rates need to be considered an important factor determining not only growth but also developmental rates of tadpoles. In the study protocol, a recommendation for feeding rates was to start with a daily food ration of approximately 600 mg SeraMicron per tank, that is 30 mg food per animal, and to increase the amount of food along with tadpole growth. Table 7 provides an overview of the actual amount of food applied in the various studies during Phase-2. From these data, it is apparent that less food was provided to the animals used in lab 3. Approximately 30-50% of the food amount applied in other labs was given to the animals in lab 3. The total amount of food is also much less than recommended in the study protocol.

Table 7. Daily feeding rates (mg Sera Micron per animal) used in different exposure studies during validation Phase-2

study	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6
IOP	30 – 67	10 – 60	18 – 24	n.s.	40 – 90	23 – 37
T4	30 – 67	10 – 60	18 – 24	n.s.	40 – 90	n.s.
PER	30 – 67	10 – 60	18 – 24	-	40 – 90	n.s.

Note: - data not available; n.s. no study conducted

39. Combining the information about feeding rates (Table 7) with results from stage determination (Table 6, Figure 2) suggests that feeding rates lower than the recommended 30 mg per animal lead to a more heterogenous development in the control population. Although the box plot in Figure 2 seems to indicate faster development in control animals of experiments with low feeding rates (see lab 3 results), a closer examination of the data presented in Table 6 revealed that it is rather a subpopulation of animals undergoing accelerated development thereby greatly increasing the inter-individual variability. Overall, the control data from experiments with daily feeding rates of 30 mg per animal and higher indicate that such feeding rates control the variability of developmental rates. In addition, these feeding rates seem to ensure that only a small proportion of control animals enter climax stages.

40. Combining the information about feeding rates (Table 7) with results from measurements of growth-related parameters (WBL, SVL and wet weight),

Growth in the Control Group

41. In addition to the monitoring of developmental stages, the study protocols calls for determination of various growth-related parameters in order to allow a differential assessment of test substance effects on general growth *versus* specific developmental processes. Accordingly, one aim of this validation exercise was to compare the utility of three different parameters to characterize treatment-related effects on body size (WBL, SVL) and body weight (wet weight). Mean values for WBL, SVL and wet weight determined

for individual replicate tanks in the control groups are compared among all 14 tests performed during the Phase-2 validation study in Figure 3 for study day 7 and in Figure 5 for study day 21.

42. As shown in Figure 3 and Figure 5, general tadpole growth was highly reproducible among replicate tanks of the control group in experiments performed in lab 1, lab 2, lab 4, lab 5 and lab 6. To obtain a first gross estimation of variability of these parameters in the control animal population, CV values (coefficient of variation) were calculated for each tank using individual animals as replicates (Figure 4, Figure 6). CV values calculated for replicate tanks indicated that the variability of individual tadpole growth was low (<15% for WBL and SVL, and <30% for wet weight) in experiments performed in lab 1, lab 2, lab 4, lab 5 and lab 6. In contrast, variability of growth rates within and among replicates was markedly increased in all three experiments performed in lab 3 (see Figure 4 and Figure 6).

43. Mean values of WBL, SVL and wet weight for control treatment group and the corresponding CV values are reported in

Table 8 and Table 9. The intra-laboratory comparison of tadpole growth parameters showed highly reproducible results in lab 1, lab 2 and lab 5 and less reproducible results in lab 3. The increased CV values for all three growth-related parameters on day 21 in the experiments of lab 3 can be explained by the fact that, at the time of morphological examination, many tadpoles were at climax stages (Figure 2, Table 6). Because development to climax stages is associated with marked decreases in body size and body weight, mean values for WBL, SVL and wet weight do not provide a meaningful estimate of tadpole growth if an increased number of animals has entered climax stages 61 or higher. Inter-laboratory comparison of tadpole growth rates showed high reproducibility of tadpole growth among all labs with the exception of lab 3. A sound analysis of tadpole growth in lab 3 was not possible due to the increased variability of individual developmental rates and the large proportion of animals at climax stages.

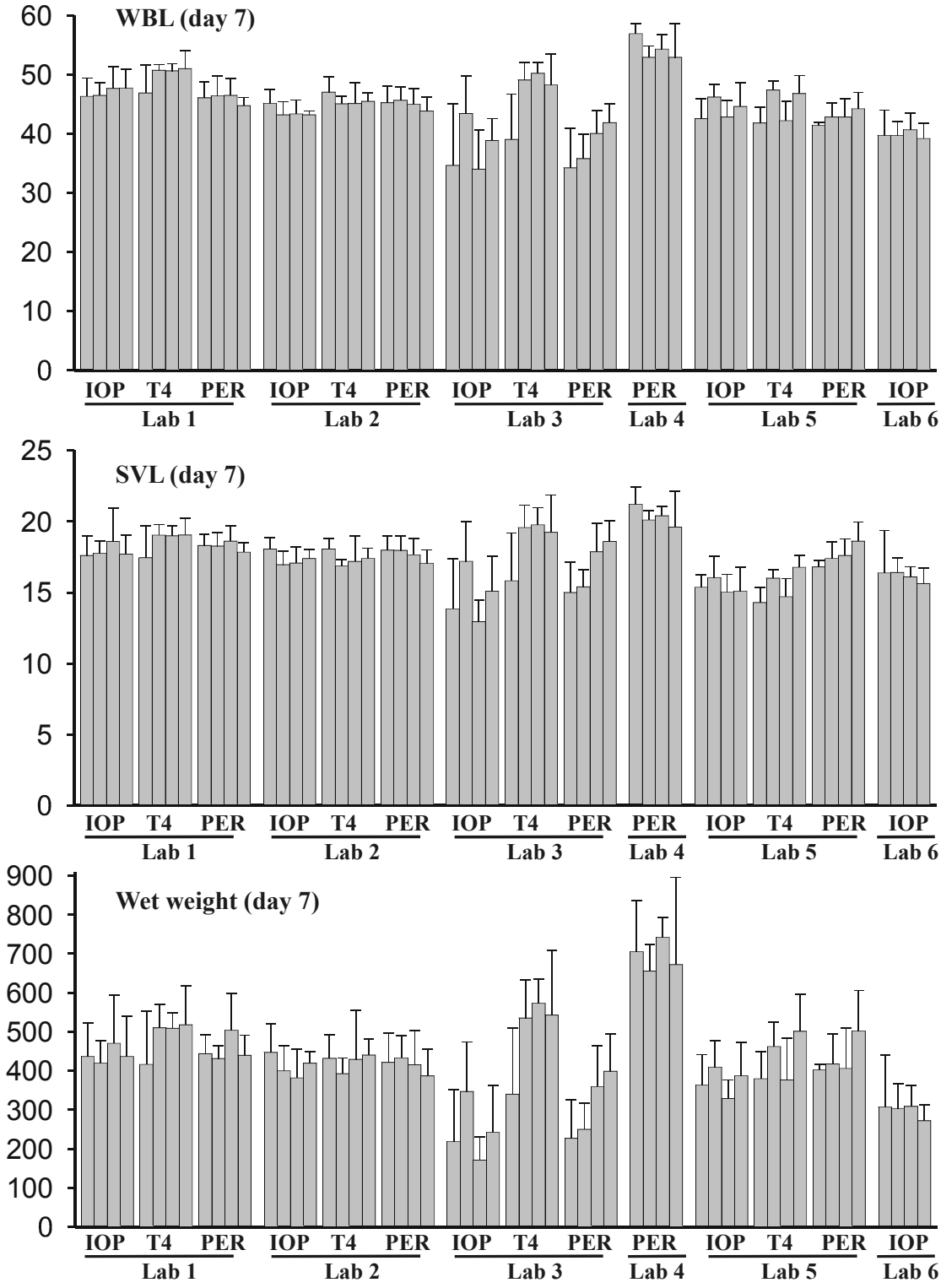

Figure 3. Mean values (± standard deviation) of WBL (mm), SVL (mm), and wet weight (mg) on day 7 in individual replicate tanks of control treatment groups

Data are shown for the 4 replicate tanks of the control treatment used in experiments with IOP, T4 and PER as test substance.

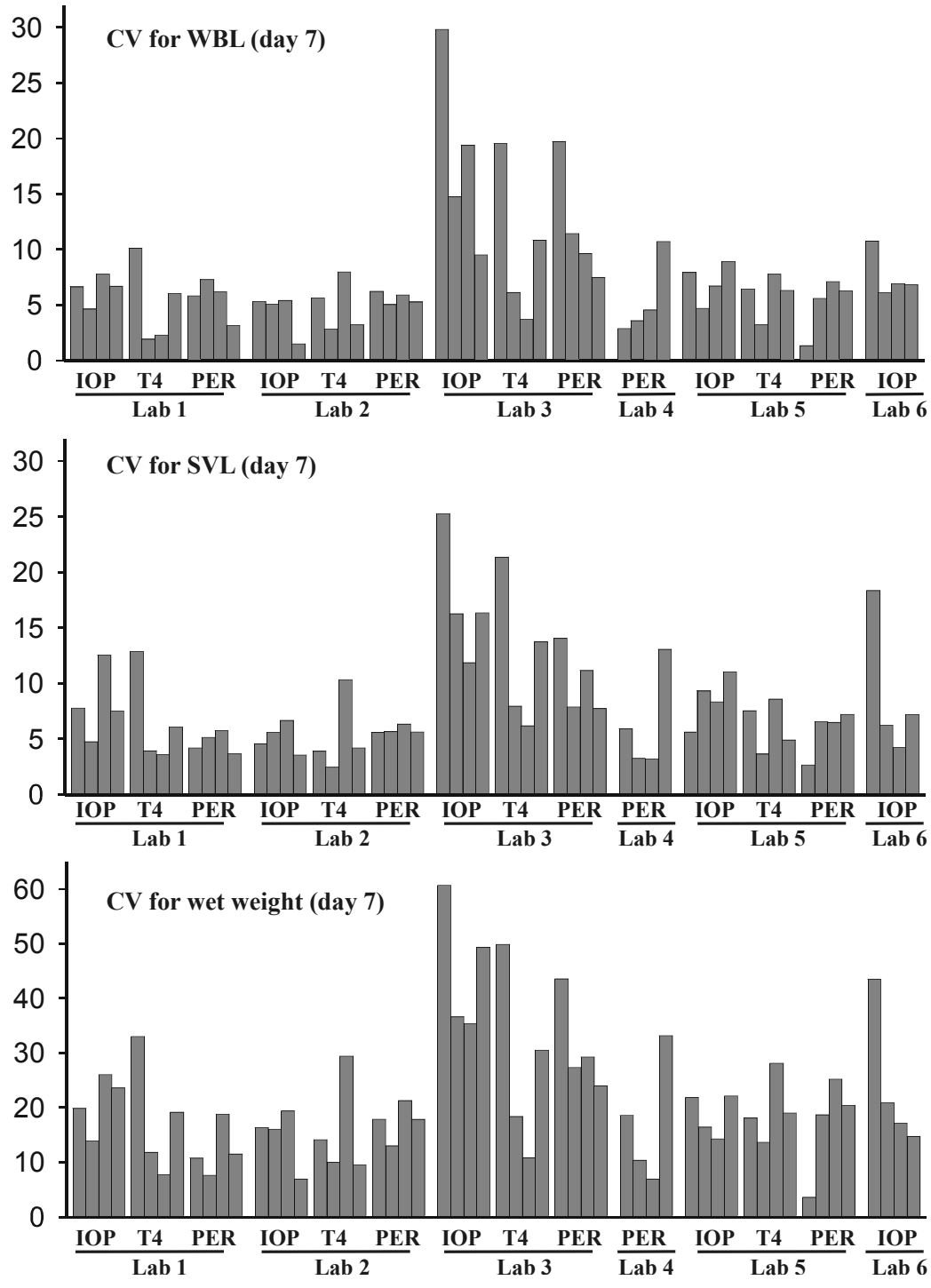

Figure 4. CV values (%) for WBL, SVL, and wet weight measurements on day 7 in replicate tanks of the control treatment group in 14 different experiments during validation Phase-2

Individual tadpoles served as replicates in the determination of tank means and standard deviations (*n*=5).

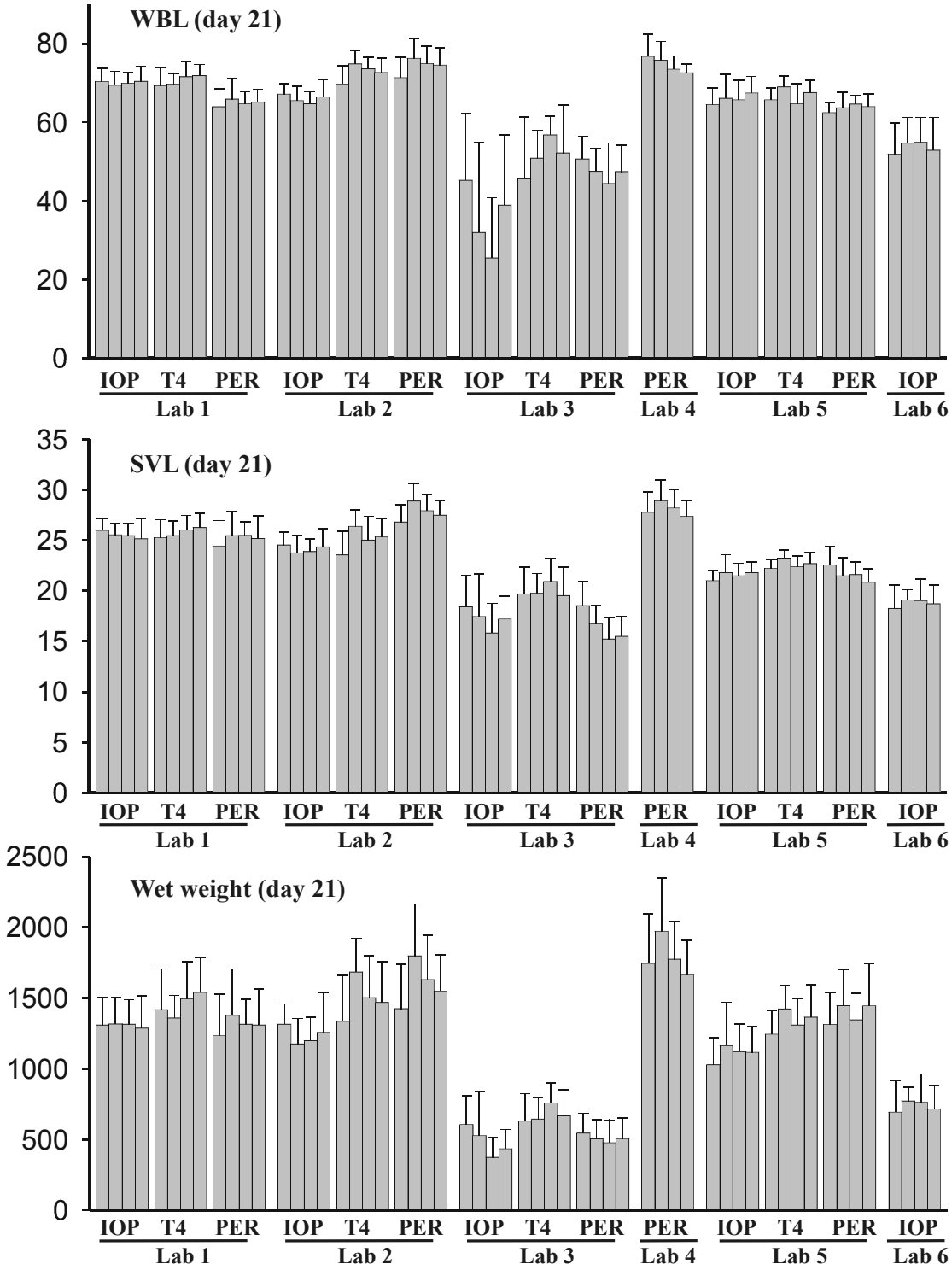

Figure 5. Mean values (± standard deviation) of WBL (mm), SVL (mm), and wet weight (mg) on day 21 in replicate tanks of the control treatment group

Data are shown for 4 replicate tanks of the control treatment used in experiments with IOP, T4 and PER as test substance.

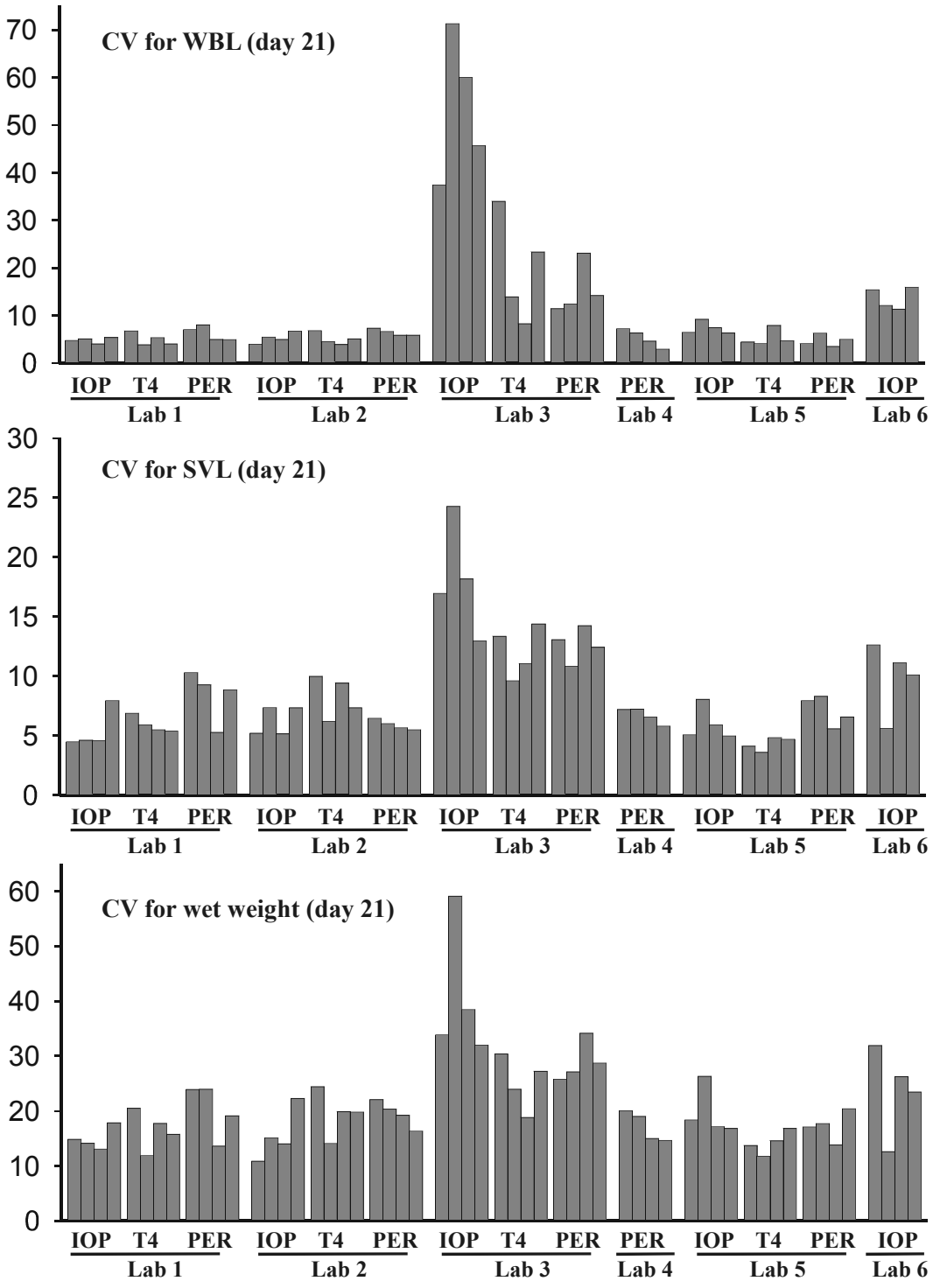

Figure 6. CV values (%) for WBL, SVL, and wet weight measurements on day 21 in replicate tanks of the control treatment group in 14 different experiments during validation phase-2

Individual tadpoles served as replicates in the determination of tank means and standard deviations (n=12-15).

Table 8. Results from WBL, SVL and wet weight measurements of control tadpoles on study day 7 in 14 experiments performed during validation Phase-2. Mean values for the control tretament group are reported together with intra-laboratory and inter-laboratory CV values.

endpoint	study	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 4 mean ±	lab 5 mean ±	lab 6 mean ±	CV[b] (all labs)
WBL (mm)	IOP	47,04 ± 0,74	43,71 ± 0,95	37,72 ± 4,35	-	44,04 ± 1,71	39,80 ± 0,60	9.8%
	PE	45,92 ± 0,81	44,92 ± 0,78	37,98 ± 3,57	54,25 ± 1,89	42,80 ± 1,14	-	
	RT4	49,79 ± 1,95	45,66 ± 0,91	46,64 ± 5,13	-	44,56 ± 2,95	-	
	CV[a]	**4.2%**	**2.2%**	**12.4%**	**-**	**2.1%**	**-**	
SVL (mm)	IOP	17,91 ± 0,45	17,37 ± 0,50	14,76 ± 1,84	-	15,39 ± 0,47	16,12 ± 0,36	8.7%
	PE	18,25 ± 0,32	17,66 ± 0,43	16,71 ± 1,78	20,33 ± 0,67	17,60 ± 0,75	-	
	RT4	18,62 ± 0,79	17,38 ± 0,50	18,60 ± 1,87	-	15,44 ± 1,15	-	
	CV[a]	**1.9%**	**0.9%**	**11.5%**	**-**	**7.8%**	**-**	
wet weight (mg)	IOP	441,1 ± 21,3	412,3 ± 28,0	245,1 ± 74,1	-	372,6 ± 34,6	298,3 ± 17,4	25.2%
	PE	455,0 ± 33,1	414,5 ± 19,6	309,2 ± 83,2	694,0 ± 38,2	432,4 ± 47,4	-	
	RT4	488,7 ± 48,2	423,8 ± 21,1	498,0 ± 106,3	-	430,3 ± 62,0	-	
	CV[a]	**5.3%**	**1.5%**	**37.5%**	**-**	**8.2%**	**-**	

Note: Mean values and standard deviations are shown (*n*=4 replicate tanks). [a] intra-laboratory CV for each endpoint measurement; [b] inter-laboratory CV for each endpoint measurement.

Table 9. Results from WBL, SVL and wet weight measurements of control tadpoles on study day 21 in 14 experiments performed during validation Phase-2. Mean values for the control tretament group are reported together with intra-laboratory and inter-laboratory CV values.

endpoint	study	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 4 mean ±	lab 5 mean ±	lab 6 mean ±	CV[b] (all labs)
WBL (mm)	**IOP**	70.04 ± 0.44	65.96 ± 1.07	35.39 ± 8.55	-	65.97 ± 1.21	53.58 ± 1.47	**8.9%**
	PEP	64.91 ± 0.82	74.30 ± 2.07	47.50 ± 2.55	74.68 ± 1.96	63.69 ± 0.91	-	
	T4	70.62 ± 1.31	72.71 ± 2.18	51.39 ± 4.52	-	66.75 ± 1.92	-	
	CV[a]	**4.6%**	**6.2%**	**18.6%**	**-**	**2.5%**	**-**	
SVL (mm)	**IOP**	25.53 ± 0.35	24.13 ± 0.37	17.23 ± 1.07	-	21.52 ± 0.38	18.78 ± 0.38	**11.6%**
	PEP	25.15 ± 0.49	27.77 ± 0.88	16.49 ± 1.51	28.06 ± 0.66	21.62 ± 0.70	-	
	T4	25.75 ± 0.47	25.08 ± 1.16	19.98 ± 0.64	-	22.64 ± 0.44	-	
	CV[a]	**1.2%**	**7.4%**	**10.3%**	**-**	**2.8%**	**-**	
wet weight (mg)	**IOP**	1307 ± 14	1236 ± 63	485 ± 102	-	1107 ± 57	736 ± 38	**20.3%**
	PEP	1309 ± 59	1601 ± 157	509 ± 28	1790 ± 131	1388 ± 68	-	
	T4	1453 ± 80	1498 ± 144	675 ± 57	-	1335 ± 77	-	
	CV[a]	**6.2%**	**13%**	**18.6%**	**-**	**11.7%**	**-**	

Note: Mean values and standard deviations are shown (*n*=4 replicate tanks). [a] intra-laboratory CV for each endpoint measurement; [b] inter-laboratory CV for each endpoint measurement (results from IOP and T4 studies in lab 3 are excluded from calculation of this CV).

Test Results – Sodium Perchlorate

Overview

44. During the validation Phase-2 study of the Amphibian Metamorphosis Assay, a total of five experiments with sodium perchlorate (Na-PER) as test substance were performed in five different laboratories (lab 1, lab 2, lab 3, lab 4, and lab 5). The nominal aqueous concentrations for the perchlorate anion (PER) were 0, 62.5, 125, 250 and 500 µg/l. Analysis of growth-related parameters (e.g., WBL, SVL, wet weight) and mortality rates did not reveal signs of systemic toxicity for any of the tested PER concentrations. Results from stage determination and HLL measurements showed PER treatment to cause developmental delay in two experiments. Histological assessment of thyroid tissue was performed in four labs. PER treatment elicited marked effects on thyroid gland histology in all four studies. Incidence and severity of histological changes occurred in a concentration-dependent manner and provided strong evidence for disruption of the thyroid system by PER.

General Test Performance

45. An initial assessment of the test performance for the five experiments conducted with PER in different labs indicated that the test performed with PER in lab 3 was characterized by increased variability in control stage development (observation of 9 different stages on day 21, ranging from stage 56 to stage 64) and lower growth rates compared to other labs. A stage-matched comparison of growth-related parameters between laboratories showed that tadpoles in the PER experiment of lab 3 had on average less than 50% of the wet weight compared to tadpoles used in other labs. The low feeding rates used in this lab are most likely responsible for this deviation from the other experiments (see section 0).

46. The study protocol called for verification of aqueous PER concentrations in each treatment tank at least once per week. Overall, the measured test concentrations were close to nominal target concentrations in all experiments as shown in

Table 10.

Mortality

47. Mortality of tadpoles was assessed daily throughout the exposure phase and the total number of dead test animals observed after 21 days of exposure is reported in Table 11 for all five experiments with PER. No mortality was observed in any treatment group in lab 3 and lab 5 and mortality was very low (< 5%) in tests performed in lab 1, lab 2 and lab 4.

Table 10. Measured concentrations (µg/L) for the perchlorate anion.

nominal conc.	replicate	lab 1	lab 2	lab 3	lab 4	lab 5
		mean ± SD (*n*)	mean ± SD (*n*)	mean ± SD (*n*)	mean ± SD (*n*)	mean ± SD (*n*)
0.0	1	n.d.	n.d.	n.d.	n.d.	n.d.
	2	n.d.	n.d.	n.d.	n.d.	n.d.
	3	n.d.	n.d.	n.d.	n.d.	n.d.
	4	n.d.	n.d.	n.d.	n.d.	n.d.
62.5	1	60.2 ± 9.7 (4)	60.9 ± 1.0 (4)	82.8 ± 12.6 (4)	77.9 ± 46.4 (4)	62.9 ± 9.9 (4)
	2	59.1 ± 2.8 (4)	60.3 ± 2.8 (4)	86.7 ± 1.8 (3)	57.9 ± 9.5 (4)	68.8 ± 8.3 (3)
	3	63.1 ± 5.2 (4)	59.4 ± 4.5 (4)	89.8 ± 6.7 (3)	67.6 ± 5.7 (4)	68.6 ± 8.4 (4)
	4	63.0 ± 3.9 (4)	58.5 ± 3.2 (4)	89.0 ± 6.4 (3)	58.5 ± 3.0 (4)	66.2 ± 13.1 (4)
125	1	116.4 ± 14.9 (4)	123.0 ± 9.6 (4)	140.1 ± 5.7 (4)	121.6 ± 7.7 (4)	151.1 ± 4.3 (5)
	2	119.1 ± 9.9 (4)	119.3 ± 6.9 (4)	150.7 ± 5.2 (3)	121.0 ± 10.5 (4)	143.8 ± 8.5 (4)
	3	116.8 ± 11.7 (4)	118.0 ± 7.7 (4)	151.4 ± 5.3 (3)	124.0 ± 5.9 (4)	138.2 ± 12.8 (5)
	4	114.3 ± 6.1 (4)	120.0 ± 3.4 (4)	153.2 ± 5.9 (3)	131.5 ± 12.6 (4)	142.1 ± 10.3 (4)
250	1	218.9 ± 17.9 (4)	240.8 ± 6.7 (4)	286.7 ± 9.6 (4)	234.6 ± 18.7 (4)	261.8 ± 35.8 (5)
	2	224.4 ± 16.8 (4)	225.8 ± 11.6 (4)	282.8 ± 8.7 (3)	233.6 ± 16.6 (4)	274.7 ± 36.0 (4)
	3	233.2 ± 16.4 (4)	230.5 ± 16.9 (4)	283.0 ± 12.5 (3)	241.5 ± 9.8 (4)	261.2 ± 36.0 (5)
	4	228.6 ± 17.4 (4)	230.3 ± 17.3 (4)	282.4 ± 6.8 (3)	241.1 ± 11.3 (4)	267.5 ± 37.1 (4)
500	1	468.1 ± 32.8 (4)	476.3 ± 8.7 (4)	573.0 ± 71.9 (4)	446.2 ± 92.1 (4)	531.2 ± 60.4 (4)
	2	491.9 ± 31.1 (4)	468.3 ± 11.9 (4)	537.8 ± 11.3 (3)	444.6 ± 87.4 (4)	524.5 ± 60.2 (4)
	3	490.5 ± 22.3 (4)	468.8 ± 9.1 (4)	535.6 ± 13.3 (3)	450.8 ± 78.1 (4)	528.8 ± 62.2 (4)
	4	491.2 ± 19.8 (4)	481.3 ± 8.7 (4)	590.8 ± 99.4 (3)	443.9 ± 68.7 (4)	513.0 ± 56.8 (6)

Note: Mean values (mean measured concentration across one replicate tank through time) and standard deviations are shown and the number of replicate measurements for each tanks are given in parentheses. n.d. not detected.

Table 11. Mortality following 21-d exposure of tadpoles to the indicated nominal concentrations of perchlorate (µg/l) as observed in different experiments

nominal conc.	lab 1	lab 2	lab 3	lab 4	lab 5
0.0	0	0	0	0	0
62.5	0	1 (1.25%)	0	2 (2.5%)	0
125	1 (1.25%)	0	0	0	0
250	1 (1.25%)	0	0	0	0
500	3 (3.75%)	0	0	0	0

Note: Data are presented as the absolute number of dead animals with the percentage of dead animals for each treatment group in parentheses. Note that two animals of the 500 µg/L PER treatment in lab 1 died by experimental error.

Effects of PER on Growth-related Endpoints

48. Possible effects of PER treatment on tadpole growth were assessed by morphometric measurements of tadpole size (WBL, SVL) and tadpole weight (wet weight). Measurements were performed for a subset of animals (n=5 animals per replicate tank) on study day 7 and for the remaining animals on study day 21. Collectively, data from day 7 and day 21 measurements indicate that treatment of tadpoles with PER did not result in growth retardation. However, results from four of five tests conducted with PER showed significant increases in tadpole size and tadpole weight (Table 12, Table 13, Table 14). Tadpole growth was not affected by PER treatment in lab 2.

49. Detection of the growth-promoting effects of PER treatment was more robust when the corresponding endpoints were analyzed on study day 21. Results from lab 4 were an exception in this regard, because significant effects at 250 and 500 µg/L PER on size and weight were only detected on study day 7 but not on study day 21. On study day 21, concentration-dependent increases in all three growth-related parameters (WBL, SVL and wet weight) were observed for PER concentrations of 125, 250 and 500 µg/L in lab 1 and for PER concentrations of 250 and 500 µg/L in lab 3. It should be noted that in lab 1, an increased number of control animals already showed development to climax stages which is associated with weight loss and a reduction in body size, thus hampering a sound assessment of growth effects at higher PER concentrations which caused a moderate inhibition of development. Significant increases in tadpole size (at 250 and 500 µg/L PER), but not wet weight, were observed in lab 5.

Table 12. Results from whole body length measurements (mm) on days 7 and 21 as determined in different experiments with perchlorate (PER) as test substance

	PER µg/l	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 4 mean ±	lab 5 mean ±
day 7	**0.0**	45.92 ± 0.81	44.92 ± 0.78	37.98 ± 3.57	54.25 ± 1.89	42.80 ± 1.14
	62.5	45.59 ± 1.75	46.30 ± 2.14	36.93 ± 2.67	53.58 ± 2.37	44.10 ± 0.89
	125	45.48 ± 0.84	44.63 ± 1.64	38.21 ± 3.32	56.27 ± 1.21[a]	44.30 ± 0.48
	250	46.19 ± 1.73	44.85 ± 3.41	39.34 ± 2.22	**56.20 ± 1.23**	45.44 ± 1.50
	500	47.18 ± 1.22	45.55 ± 0.87	37.26 ± 2.05	**56.28 ± 0.66**	43.45 ± 1.95
day 21	**0.0**	64.91 ± 0.82	74.30 ± 2.07	47.50 ± 2.55	74.68 ± 1.96	63.69 ± 0.91
	62.5	65.44 ± 0.88	77.03 ± 2.34	49.40 ± 1.81	74.61 ± 1.72	63.21 ± 0.61
	125	**67.27 ± 0.42**	77.84 ± 2.01	50.66 ± 2.84	70.96 ± 1.28	64.59 ± 1.22
	250	**68.13 ± 1.18**	77.28 ± 2.39	**52.38 ± 0.99**	74.01 ± 0.71	**66.33 ± 0.69**
	500	**68.90 ± 1.22**	76.84 ± 1.74	**53.28 ± 2.45**	74.65 ± 1.91	**66.85 ± 0.74**

Note: Mean values and standard deviations are shown (n=4; [a] n=3). Bold letters indicate statistically significant differences to the control group (p<0.05, Jonckheere-Terpstra test).

Table 13. Results from snout-vent length measurements (mm) on days 7 and 21 as determined in different experiments with perchlorate (PER) as test substance

	PER µg/l	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 4 mean ±	lab 5 mean ±
day 7	**0.0**	18.25 ± 0.32	17.66 ± 0.43	16.71 ± 1.78	20.33 ± 0.67	17.60 ± 0.75
	62.5	18.19 ± 0.54	18.01 ± 0.95	16.47 ± 0.98	20.54 ± 0.73	17.95 ± 0.53
	125	18.19 ± 0.38	17.41 ± 0.53	17.02 ± 0.97	21.23 ± 0.55[a]	18.00 ± 0.16
	250	18.75 ± 0.62	17.34 ± 1.21	16.87 ± 0.76	**21.28 ± 0.41**	18.50 ± 0.48
	500	**18.74 ± 0.50**	17.53 ± 0.50	15.71 ± 1.12	**21.25 ± 0.29**	17.15 ± 0.72
day 21	**0.0**	25.15 ± 0.49	27.77 ± 0.88	16.49 ± 1.51	28.06 ± 0.66	21.62 ± 0.70
	62.5	25.32 ± 0.29	28.33 ± 0.36	16.38 ± 0.48	28.20 ± 0.38	21.56 ± 0.68
	125	**25.74 ± 0.10**	28.61 ± 0.59	17.15 ± 0.42	28.47 ± 0.42	21.15 ± 0.56
	250	**26.35 ± 0.25**	28.19 ± 0.70	**17.53 ± 0.68**	28.62 ± 0.31	22.18 ± 0.43
	500	**26.98 ± 0.46**	28.62 ± 0.66	**17.53 ± 0.97**	28.61 ± 0.62	**23.08 ± 0.32**

Note: Mean values and standard deviations are shown (n=4; [a] n=3). Bold letters indicate statistically significant differences to the control group (p<0.05, Jonckheere-Terpstra test).

Table 14. Results from wet weight measurements (mg) on days 7 and 21 as determined in different experiments with perchlorate (PER) as test substance.

	PER µg/l	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 4 mean ±	lab 5 mean ±
day 7	**0.0**	455 ± 33	414 ± 20	309 ± 83	694 ± 38	432 ± 47
	62.5	464 ± 20	438 ± 58	283 ± 53	673 ± 133	451 ± 19
	125	475 ± 20	388 ± 39	318 ± 59	771 ± 49[a]	480 ± 29
	250	497 ± 46	397 ± 79	333 ± 47	**798 ± 86**	512 ± 22
	500	**513 ± 33**	409 ± 22	309 ± 48	**761 ± 14**	433 ± 48
day 21	**0.0**	1309 ± 59	1601 ± 157	509 ± 28	1790 ± 131	1388 ± 68
	62.5	1343 ± 42	1734 ± 70	555 ± 39	1829 ± 91.0	1300 ± 66
	125	**1457 ± 24**	1807 ± 110	**625 ± 62**	1811 ± 103	1427 ± 116
	250	**1506 ± 74**	1746 ± 155	**632 ± 49**	1843 ± 115	1419 ± 66
	500	**1584 ± 97**	1813 ± 115	**653 ± 106**	1835 ± 96	1385 ± 51

Note: Mean values and standard deviations are shown (n=4; [a] n=3). Bold letters indicate statistically significant differences to the control group ($p<0.05$, Jonckheere-Terpstra test).

50. Overall, the sensitivity of WBL and SVL measurements to detect changes in tadpole size were similar in three tests (lab 1, lab 3, and lab 4), whereas SVL was slightly less sensitive than WBL in one test (lab 5). A less consistent relationship was observed when comparing changes in tadpole size and tadpole weight. In lab 1, significant increases in wet weight were detected in all PER treatments that caused increases in WBL and SVL. In lab 3, wet weight was even increased at PER concentrations (125 µg/l) that did not led to significant effects on WBL and SVL. In contrast, significant changes in tadpole WBL or SVL were not associated with significant changes in wet weight in a test conducted in lab 5.

Effects of PER on Developmental Endpoints

51. At the concentrations used in this study, PER treatment caused developmental delay of tadpoles only in two tests (lab 3, lab 4) (Table 15 andTable 16). Detection of developmental effects was more robust when the endpoints developmental stage and HLL were analyzed on study day 21 compared to study day 7. On day 7, a weak but significant effects on median stage was only detected for the 500 µg/l PER treatment in lab 1 (Table 16). When the standard Jonckheere-Terpstra test was used to test for significant trends of day 21 median stages, significant effects of PER treatment on tadpole stage were detected for the highest PER concentration (500 µg/l) in lab 4 and for PER concentrations of 125, 250 and 500 µg/l in lab 3 (Table 16).

52. The tests conducted in lab 3 and lab 4 also revealed a reduction in hind limb growth due to PER treatment (Table 15). Compared to the control group, mean HLL was significantly lower following 21-day exposure of tadpoles to 250 and 500 µg/l PER in tests from lab 3 and lab 4. In the other three experiments, tadpole development was not significantly affected by the tested PER concentrations as judged from stage determination and HLL measurements.

Table 15. Results from hind limb length measurements (mm) on days 7 and 21 as determined in different experiments with perchlorate (PER) as test substance.

	PER	lab 1	lab 2	lab 3	lab 4	lab 5
	µg/l	mean ±	mean ±	mean ±	mean ±	mean ±
day 7	**0.0**	2.11 ± 0.12	2.49 ± 0.09	2.39 ± 0.27	2.45 ± 0.12	1.91 ± 0.10
	62.5	2.04 ± 0.05	2.52 ± 0.21	2.27 ± 0.28	2.43 ± 0.18	1.88 ± 0.08
	125	2.18 ± 0.05	2.45 ± 0.14	2.35 ± 0.22	2.62 ± 0.04[a]	1.95 ± 0.11
	250	2.02 ± 0.08	2.34 ± 0.26	2.44 ± 0.19	2.73 ± 0.13	2.01 ± 0.13
	500	2.01 ± 0.16	2.40 ± 0.12	2.32 ± 0.24	2.41 ± 0.15	1.89 ± 0.13
day 21	**0.0**	12.88 ± 0.49	14.74 ± 0.96	11.57 ± 2.17	14.21 ± 0.71	10.86 ± 0.75
	62.5	11.92 ± 0.91	14.67 ± 1.20	10.62 ± 1.76	13.22 ± 0.65	9.04 ± 1.08
	125	13.01 ± 0.31	15.65 ± 1.55	9.31 ± 1.56	13.36 ± 0.81	9.28 ± 0.60
	250	13.58 ± 0.80	15.42 ± 1.72	**7.78 ± 1.70**	**11.25 ± 0.78**	10.43 ± 1.36
	500	12.29 ± 0.21	14.49 ± 1.41	**5.18 ± 1.04**	**8.12 ± 0.57**	9.07 ± 0.40

Note: Mean values and standard deviations are shown (n=4; [a] n=3). Bold letters indicate statistically significant differences to the control group ($p<0.05$, Jonckheere-Terpstra test).

Table 16. Results from stage determination (Nieuwkoop and Faber, 1994) on days 7 and 21 in different experiments with perchlorate (PER) as test substance

	PER µg/	r	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 4 mean ±	lab 5 mean ±
day 7	0.0	1	55 (54-55)	54 (53-54)	54 (53-54)	55 (54-55)	53 (53-54)
		2	55 (54-55)	54 (54-55)	54 (53-54)	55 (54-55)	54 (53-54)
		3	55 (54-55)	54 (53-54)	54 (54-54)	55 (54-55)	54 (53-54)
		4	55 (54-	54 (54-54)	54 (54-55)	55 (53-55)	54 (53-54)
	62.5	1	55 (54-55)	54 (54-55)	54 (54-54)	54.5 (54-55)	53 (53-55)
		2	55 (53-55)	54 (53-55)	54 (54-55)	55 (54-55)	54 (53-54)
		3	55 (54-	55 (55-55)	54 (53-55)	55 (54-55)	54 (53-54)
		4		55 (54-55)	54 (53-55)	54.5 (54-55)	54 (52-54)
	125	1	55 (53-55)	55 (55-55)	54 (53-55)	55 (55-55)	54 (53-54)
		2	55 (54-55)	55 (55-55)	54 (53-55)	55 (54-55)	54 (54-54)
		3	55 (54-	55 (53-55)	54 (54-55)	55 (54-55)	53 (52-54)
		4		55 (54-55)	54 (54-54)	n.d.	54 (53-54)
	250	1	54 (53-55)	54 (53-55)	54 (54-55)	55 (54-55)	54 (54-54)
		2	55 (54-55)	54 (53-55)	54 (54-55)	55 (55-55)	54 (53-54)
		3	55 (54-	55 (55-55)	54 (53-55)	55 (55-55)	54 (53-54)
		4		55 (54-55)	54 (54-55)	55 (54-55)	54 (53-54)
	500	1	**54 (54-55)**	54 (53-55)	54 (53-55)	54 (54-55)	53 (53-54)
		2	**54 (54-55)**	54 (54-55)	54 (53-54)	54 (54-55)	54 (54-54)
		3	**54 (53-**	55 (54-55)	54 (53-54)	55 (54-55)	54 (54-54)
		4		55 (54-55)	54 (54-55)	55 (54-55)	54 (52-54)
day 21	0.0	1	58 (56-62)	58 (56-60)	57 (56-60)	59 (57-59)	57 (56-61)
		2	58 (57-60)	58 (57-60)	59 (57-65)	58 (58-60)	58 (56-59)
		3	58 (57-60)	58 (57-60)	61 (56-63)	58 (57-60)	57 (56-59)
		4	58 (56-61)	58 (56-60)	59 (57-63)	58 (57-60)	57 (56-60)
	62.5	1	58 (57-59)	58 (56-59)	58 (55-64)	58 (57-60)	57 (56-59)
		2	57 (56-62)	58 (57-60)	59 (56-62)	58 (57-59)	57 (56-60)
		3	58 (56-62)	58 (57-59)	57 (56-62)	58 (57-59)	56 (56-58)
		4	58 (56-61)	58 (58-59)	57 (56-61)	58 (57-60)	57 (56-58)
	125	1	58 (56-61)	58 (57-58)	**57 (56-59)**	58 (57-60)	57 (56-58)
		2	58 (56-60)	58 (57-60)	**57 (55-61)**	58 (57-60)	57 (56-59)
		3	58 (57-60)	58 (57-60)	**58 (55-60)**	58 (57-59)	57 (56-59)
		4	58 (56-61)	59 (58-59)	**57 (56-61)**	58 (57-60)	57 (56-60)
	250	1	58 (56-61)	58 (57-60)	**57 (55-58)**	58 (57-59)	57 (56-58)
		2	58 (57-61)	58 (57-59)	**55 (55-61)**	58 (57-59)	57 (56-60)
		3	58 (56-60)	58 (58-61)	**57 (55-59)**	58 (56-59)	57 (56-58)
		4	58 (56-60)	58 (56-60)	**57 (55-62)**	58 (57-59)	57 (56-59)
	500	1	58 (56-59)	58 (57-59)	**56 (53-57)**	**57 (54-58)**	57 (56-59)
		2	58 (57-60)	58 (56-58)	**56 (54-57)**	**57 (53-58)**	57 (56-58)
		3	58 (56-59)	58 (57-59)	**56 (55-59)**	**57 (54-59)**	57 (56-60)
		4	58 (56-60)	58 (57-59)	**55 (55-59)**	**57 (53-59)**	57 (56-59)

Note: Values shown are the median stage for each replicate tank with the total range of stages in parentheses. Bold letters indicate statistically significant differences in median stages relative to the control group ($p<0.05$, Jonckheere-Terpstra test). [a] replicate tank; n.d. no data available

Effects of PER on Thyroid Gland Histology

53. The study protocol called for evaluation of thyroid gland histology in five specimen randomly sampled from each replicate tank (*n*=20 per treatment) according to a graded scoring system which incorporates information on incidence and severity of specific histological changes. At the time of report preparation, results from histological evaluation of thyroid tissue were available for four of five experiments. In all four experiments including a histological assessment of thyroid tissue, exposure-related changes in the thyroid gland were detected (Table 17, Table 18, Table 20). The effects pattern observed following 21-day treatment with PER was very consistent among all experiments and included concentration-dependent decreases in colloid content, increases in overall thyroid gland size and hypertrophic and hyperplastic changes in the follicular epithelium. A remarkable finding was that partial colloid depletion, glandular hypertrophy and follicular cell hypertrophy occurred at all tested PER concentrations. Thus, all PER concentrations were effective in altering thyroid histology. Within the tested concentration range of PER, colloid depletion, thyroid gland hypertrophy and changes in follicular epithelium were generally mild to moderate at the lowest concentration (62.5 µg/L) but moderate to severe at the highest concentration (500 µg/L).

54. A semi-quantitative analysis of the histological alterations was performed using a 3-scale severity grading approach (5-scale scheme was used in lab 3) to assess the incidence and severity of three core diagnostic parameters including thyroid gland size, follicular cell hypertrophy and follicular cell hyperplasia. Results from these analyses are summarized in Table 17, Table 18, and Table 20 and confirmed the concentration-dependent increase in incidence as well as severity of hypertrophic and hyperplastic changes in the follicular cell epithelium.

Table 17. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to perchlorate (PER) in lab 1

PER	glandular hypertrophy		glandular atrophy		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	4/20	0.20±0.41 (1)	4/20	0.20±0.41 (1)	8/20	0.40±0.50 (1)	0/20	0.00±0.00 (0)	57-61
62.5	8/20	0.50±0.69 (2)	3/20	0.15±0.37 (1)	17/20	1.10±0.64 (2)	12/20	0.65±0.59 (2)	56-62
125	14/20	1.15±0.93 (3)	0/20	0.00±0.00 (0)	20/20	1.80±0.77 (3)	16/20	1.10±0.79 (3)	56-61
250	16/20	1.75±1.07 (3)	0/20	0.00±0.00 (0)	20/20	2.05±0.69 (3)	18/20	1.75±0.97 (3)	56-60
500	20/20	2.35±0.67 (3)	0/20	0.00±0.00 (0)	20/20	2.30±0.57 (3)	20/20	1.95±0.69 (3)	56-59

Note: [a] number of animals affected / number of animals examined
[b] Values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses
[c] Values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology

Table 18. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to perchlorate (PER) in lab 2

PER	glandular hypertrophy		glandular atrophy		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	2/20	0.10±0.31 (1)	2/20	0.10±0.31 (1)	1/20	0.05±0.22 (1)	0/20	0.00±0.00 (0)	56-60
62.5	11/20	0.65±0.67 (2)	0/20	0.00±0.00 (0)	9/20	0.45±0.51 (1)	0/20	0.00±0.00 (0)	57-59
125	15/20	0.95±0.69 (2)	0/20	0.00±0.00 (0)	11/20	0.55±0.51 (1)	10/20	0.50±0.51 (1)	57-60
250	17/20	1.50±0.76 (2)	0/20	0.00±0.00 (0)	15/20	0.85±0.59 (2)	14/20	0.70±0.47 (1)	58-60
500	18/20	1.70±0.66 (2)	0/20	0.00±0.00 (0)	20/20	1.20±0.41 (2)	17/20	1.00±0.56 (2)	57-59

Note: [a] number of animals affected / number of animals examined
[b] Values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses
[c] values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology

Table 19. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to perchlorate (PER) in lab 3

PER	glandular hypertrophy		colloid depletion		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	0/10	0.00±0.00 (0)	0/0	0.00±0.00 (0)	1/10	n.d.	10/10	1.20±0.42 (2)	59-61
62.5	9/10	1.90±1.29 (4)	10/10	2.40±0.97 (4)	10/10	n.d.	10/10	2.80±0.63 (4)	57-59
125	8/10	2.40±1.43 (4)	10/10	3.00±0.94 (4)	10/10	n.d.	10/10	2.90±0.57 (4)	57-59
250	9/10	2.90±1.73 (5)	10/10	2.90±0.74 (4)	10/10	n.d.	10/10	3.40±0.52 (4)	57-59
500	9/10	2.90±1.20 (4)	10/10	2.80±0.79 (4)	10/10	n.d.	10/10	3.40±0.70 (4)	55-59

Note: A 5-grade scoring scheme was used in this study that differed slightly from the 3-grade scheme used by the other labs.
[a] number of animals affected / number of animals examined
[b] values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses
[c] values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology
n.d. no data available

Table 20. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to perchlorate (PER) in lab 4.

PER	glandular hypertrophy		follicular cell hypertrophy		follicular cell hyperplasia		follicular lumen area		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	1/9	0.11±0.33 (1)	0/16	0.00±0.00 (0)	1/12	0.08±0.28 (1)	0/12	0.00±0.00 (0)	n.d.
62.5	3/9	0.33±0.50 (1)	8/9	1.22±0.66 (2)	4/9	0.55±0.72 (2)	9/9	1.66±0.50 (2)	n.d.
125	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
250	9/11	1.54±1.12 (3)	15/15	1.93±0.70 (3)	17/17	1.58±0.50 (2)	17/17	2.12±0.33 (3)	n.d.
500	10/15	1.73±1.38 (3)	11/11	2.63±0.67 (3)	17/18	1.44±0.70 (3)	19/19	2.94±0.23 (3)	n.d.

Note: [a] number of animals affected / number of animals examined
[b] values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses
[c] values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology
n.d. no data available

55. In addition, morphometric techniques were used by lab 1 and lab 2 to assess effects of PER treatment on epithelial cell height (Figure 7), thyroid volume (Figure 7) and thyroid cross section area (Figure 8). The morphometric data are consistent with the results from the qualitative characterization and the severity grading approach by showing significant increases in epithelial cell height (follicular cell hypertrophy), total glandular volume and maximum cross section area (thyroid gland hypertrophy).

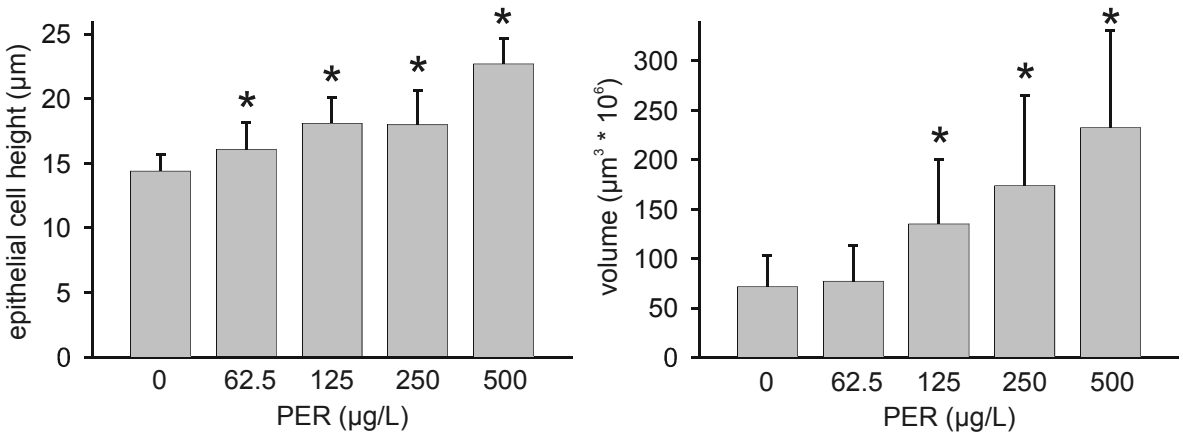

Figure 7. Morphometric analysis of thyroid tissue for changes in epithelial cell height and glandular volume following exposure to perchlorate (PER) in the lab 1 study

Columns with vertical bars represent means with standard deviations (n= 20 tadpoles per treatment). Asterisks denote significant differences compared to the control treatment group (p<0.05 Dunnett's test).

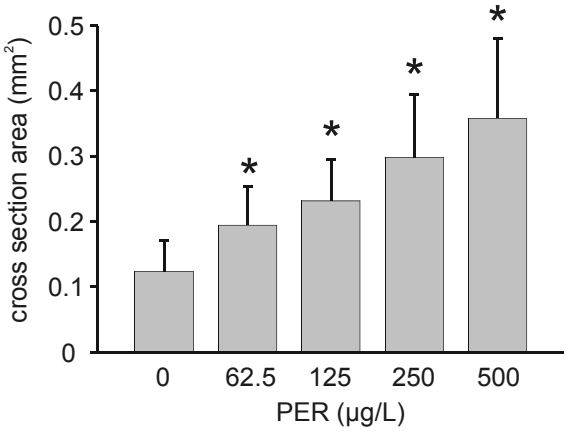

Figure 8. Morphometric analysis of thyroid tissue for changes in maximum cross section area following exposure to perchlorate (PER) in the lab 2 study

Columns with vertical bars represent means with standard deviations (n= 20 tadpoles per treatment). Asterisks denote significant differences compared to the control treatment group (p<0.05 Dunnett's test).

Test Results - Thyroxine

Overview

56. During the validation phase-2 study of the amphibian metamorphosis assay, a total of four tests with thyroxine (T4) as test substance were performed in four different laboratories (lab 1, lab 2, lab 3, and lab 5). The nominal aqueous concentrations for T4 were 0, 0.25, 0.5, 1.0, and 2.0 µg/l. Treatment of tadpoles with T4 consistently caused an acceleration of development in all tests as evident from advanced developmental stages of T4-treated tadpoles on day 21 and significantly increased mean HLL on study days 7 and 21 in the T4 treatment groups. Analysis of growth-related parameters (e.g., WBL, SVL, wet weight) showed reduced tadpole size and weight on study day 21 in those T4 treatment groups that displayed accelerated development. Overall, evaluation of growth and mortality rates did not reveal signs of systemic toxicity for any of the tested T4 concentrations. Effects of T4 treatment on thyroid gland histology were variable and histological endpoints did not display a clear concentration-response relationship. Prominent alterations in response to T4 treatment included thyroid gland atrophy at lower T4 concentrations and hypertrophy at high T4 concentrations.

General Test Performance

57. An initial assessment of validity of test performance for the four experiments conducted with T4 in different labs indicated that the T4 experiment in lab 3 was characterized by increased variability in control stage development (observation of 12 different stages on day 21, ranging from stage 55 to stage 66) and lower growth rates compared to other labs. A stage-matched comparison of growth-related parameters between different experiments showed that tadpoles in the T4 experiment of lab 3 had on average less than 50% of the wet weight compared to tadpoles used in other labs. As discussed in section 0, the low feeding rates used in this lab are most likely responsible for this deviation from the other experiments.

58. The study protocol called for verification of aqueous T4 concentrations in each treatment tank at least once per week. Actual test concentrations of T4 were close to nominal target concentrations in lab 2 as shown in

Table 21. Technical problems associated with measurements of actual T4 test concentrations were encountered in the other labs and no clear verification of test concentration was possible.

Table 21. Measured concentrations (µg/L) for thyroxine (T4)

nominal conc.	replicate	lab 1	lab 2	lab 3	lab 5
		mean ± SD (*n*)	mean ± SD (*n*)	mean ± SD (*n*)	mean ± SD (*n*)
0.0	1	n.a.	n.d.	n.a.	n.a.
	2	n.a.	n.d.	n.a.	n.a.
	3	n.a.	n.d.	n.a.	n.a.
	4	n.a.	n.d.	n.a.	n.a.
0.25	1	n.a.	0,25 ± 0,02 (4)	n.a.	n.a.
	2	n.a.	0,25 ± 0,03 (4)	n.a.	n.a.
	3	n.a.	0,25 ± 0,02 (4)	n.a.	n.a.
	4	n.a.	0,26 ± 0,02 (4)	n.a.	n.a.
0.5	1	n.a.	0,59 ± 0,01 (4)	n.a.	n.a.
	2	n.a.	0,57 ± 0,02 (4)	n.a.	n.a.
	3	n.a.	0,57 ± 0,02 (4)	n.a.	n.a.
	4	n.a.	0,56 ± 0,02 (4)	n.a.	n.a.
1.0	1	n.a.	1,10 ± 0,06 (4)	6,92 ± 11,06 (3)	n.a.
	2	n.a.	1,12 ± 0,07 (4)	0,87 ± 0,26 (2)	n.a.
	3	n.a.	1,08 ± 0,07 (4)	0,55 ± 0,03 (3)	n.a.
	4	n.a.	1,12 ± 0,04 (4)	0,49 ± 0,11 (2)	n.a.
2.0	1	n.a.	2,20 ± 0,05 (4)	2,59 ± 0,19 (2)	n.a.
	2	n.a.	2,19 ± 0,11 (4)	2,42 ± 1,67 (3)	n.a.
	3	n.a.	2,13 ± 0,02 (4)	5,25 ± 6,72 (3)	n.a.
	4	n.a.	2,18 ± 0,11 (4)	1,43 ± 0,17 (3)	n.a.

Note: Mean values (mean measured concentration across one replicate tank through time) and standard deviations are shown and the en in parentheses. n.a. not analyzed or analysis still pending; n.d. not detected number of replicate measurements for each tanks are giv

Mortality

59. Mortality of tadpoles was assessed daily throughout the exposure phase and the total number of dead test animals observed after 21 days of exposure is reported in Table 22 for all four tests with T4. No mortality was observed in any treatment group in the test performed in lab 2 and treatment-related mortality was very low (< 5%) in tests performed in lab 1 and lab 3. Lab 5 reported that all cases of mortality were due to handling errors.

Table 22. Mortality following 21-d exposure of tadpoles to the indicated nominal concentrations of thyroxine (T4) as observed in different experiments

nominal concentration µg/l T4	lab 1	lab 2	lab 3	lab 5[a]
0.0	0	0	0	1 (1.25%)
0.25	0	0	0	4 (5.0%)
0.5	0	0	0	0
1.0	0	0	0	1 (1.25%)
2.0	1 (1.25%)	0	1 (1.25%)	0

Note: Data are presented as absolute number of dead animals with the percentages of dead animals for each treatment group in parentheses. [a] Mortalities in lab 5 were reportedly all due to handling errors.

Effects of T4 on Growth-related Endpoints

60. Possible effects of T4 treatment on tadpole growth were assessed by morphometric measurements of tadpole size (WBL, SVL) and tadpole weight (wet weight). Measurements were performed for a subset of animals (n=5 animals per replicate tank) on day 7 and for the remaining animals on day 21. Day 7 measurements indicate a weak reduction in growth rates at 2.0 µg/L T4 in lab 5 and at 1.0 and 2.0 µg/L T4 in lab 3 because tadpole size (Table 23, Table 24) and tadpole weight (Table 25) were significantly lower compared to the control group. When growth-related parameters were determined on study day 21, all four tests showed a reduction in size and weight at the higher T4 concentrations. Lab 3 reported significantly reduced mean values for WBL, SVL and wet weight for all but the lowest T4 concentration. Overall, the parameters WBL, SVL and wet weight displayed a similar sensitivity to detect treatment-related effects on tadpole size and weight in all tests with T4.

Table 23. Results from whole body length measurements (mm) on days 7 and 21 as determined in different experiments with thyroxine (T4) as test substance

	T4 µg/l	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 5 mean ±
day 7	**0.0**	49.79 ± 1.95	45.66 ± 0.91	46.64 ± 5.13	44.56 ± 2.95
	0.25	49.90 ± 2.02	46.95 ± 0.49	48.01 ± 1.73	45.25 ± 1.39
	0.5	50.66 ± 0.97	46.90 ± 1.40	44.83 ± 2.17	44.95 ± 1.00
	1.0	50.90 ± 1.34	46.83 ± 1.20	**41.57 ± 3.98**	44.25 ± 1.68
	2.0	50.37 ± 1.33	46.59 ± 1.44	**41.59 ± 4.54**	**40.65 ± 0.90**
day 21	**0.0**	70.62 ± 1.31	72.71 ± 2.18	51.39 ± 4.52	66.75 ± 1.92
	0.25	70.46 ± 0.97	74.42 ± 1.09	47.40 ± 2.56	66.50 ± 1.21
	0.5	72.25 ± 1.00	74.46 ± 1.26	45.34 ± 6.50	66.53 ± 1.68
	1.0	68.58 ± 2.52	72.03 ± 1.75	**40.74 ± 2.11**	**61.20 ± 1.87**
	2.0	**66.11 ± 1.48**	**66.31 ± 1.73**	**30.02 ± 2.83**	**52.91 ± 3.32**

Note: Mean values and standard deviations are shown (*n*=4). Bold letters indicate statistically significant differences to the control group (p<0.05, Jonckheere-Terpstra test).

Table 24. Results from snout-vent length measurements (mm) on days 7 and 21 as determined in different experiments with thyroxine (T4) as test substance.

	T4 µg/l	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 5 mean ±
day 7	**0.0**	18.62 ± 0.79	17.38 ± 0.50	18.60 ± 1.87	15.44 ± 1.15
	0.25	18.74 ± 0.65	18.12 ± 0.15	19.46 ± 0.46	15.56 ± 0.34
	0.5	18.72 ± 0.48	17.84 ± 0.62	17.88 ± 0.74	15.14 ± 0.58
	1.0	18.96 ± 0.77	17.85 ± 0.90	**17.00 ± 1.89**	15.12 ± 0.40
	2.0	18.87 ± 0.23	17.57 ± 0.68	**16.12 ± 1.80**	**14.21 ± 0.31**
day 21	**0.0**	25.75 ± 0.47	25.08 ± 1.16	19.98 ± 0.64	22.64 ± 0.44
	0.25	25.86 ± 0.41	26.33 ± 0.43	19.39 ± 0.42	22.55 ± 0.15
	0.5	26.31 ± 0.28	26.46 ± 0.36	**18.34 ± 0.75**	22.22 ± 0.50
	1.0	24.70 ± 0.95	25.09 ± 0.60	**17.72 ± 0.53**	**20.79 ± 0.38**
	2.0	**23.32 ± 0.72**	**22.66 ± 0.73**	**15.62 ± 0.36**	**18.71 ± 0.83**

Note: Mean values and standard deviations are shown (*n*=4). Bold letters indicate statistically significant differences to the control group (p<0.05, Jonckheere-Terpstra test).

Table 25. Results from wet weight measurements (mg) on days 7 and 21 as determined in different experiments with thyroxine (T4) as test substance.

	T4 µg/l	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 5 mean ±
day 7	**0.0**	489 ± 48	424 ± 21	498 ± 106	430 ± 62
	0.25	485 ± 45	478 ± 8	523 ± 36	448 ± 49
	0.5	491 ± 27	473 ± 33	441 ± 44	409 ± 41
	1.0	508 ± 52	463 ± 58	**413 ± 83**	398 ± 37
	2.0	487 ± 31	441 ± 50	**389 ± 109**	**318 ± 17**
day 21	**0.0**	1453 ± 80	1498 ± 143	675 ± 57	1335 ± 77
	0.25	1447 ± 75	1652 ± 98	620 ± 27	1323 ± 47
	0.5	1548 ± 60	1666 ± 51	**559 ± 84**	1277 ± 97
	1.0	1293 ± 165	1439 ± 134	**494 ± 51**	**1012 ± 55**
	2.0	**1141 ± 68**	**1151 ± 56**	**390 ± 10**	**697 ± 83**

Note: Mean values and standard deviations are shown (*n*=4). Bold letters indicate statistically significant differences to the control group (p<0.05, Jonckheere-Terpstra test).

Effects of T4 on Developmental Endpoints

61. At the concentrations used in this study, T4 treatment caused a consistent acceleration of tadpole development in all experiments. The two main endpoint measurements, developmental stage and HLL, showed a differential sensitivity to detect developmental acceleration depending on the observation time point (see Table 26 and Table 27). By using the exact Jonckheere-Terpstra test, significant increases in HLL were detected for all four T4 concentrations (0.25 – 2.0 µg/L) on study day 7 in lab 1 and lab 2 (Table 26). Notably, data from concurrent developmental stage determinations on day 7 in these two tests did not show a significant effect of T4 on median stages (Table 27). The effects pattern observed on day 7 in the third test was somewhat different as the two highest T4 concentrations (1.0 and 2.0 µg/L) caused significant increases in HLL as well as as significant increases in median stages. No deviation from control development on day 7 was observed in lab 3.

62. When analyzed on study day 21, HLL measurements became less sensitive to detect developmental acceleration by T4. Significantly increased mean HLL values were still detected after treatment of tadpoles with 1.0 and 2.0 µg/l T4 but lower T4 concentrations did not caused significant effects on HLL on day 21. Determination of developmental stages revealed signifcant acceleration of development at 2.0 µg/l in all tests and at 1.0 µg/l in three tests (lab 1 and lab 3). Overall, HLL measurements and stage determination on day 21 displayed an equal sensitivity to detect the T4-related acceleration of morphological development.

Table 26. Results from hind limb length measurements (mm) on days 7 and 21 as determined in different experiments with thyroxine (T4) as test substance

	T4 µg/l	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 5 mean ±
day 7	**0.0**	2.20 ± 0.09	2.49 ± 0.13	3.80 ± 0.41	2.42 ± 0.12
	0.25	**2.44 ± 0.15**	**2.74 ± 0.08**	4.57 ± 0.41	2.58 ± 0.15
	0.5	**2.48 ± 0.02**	**2.73 ± 0.05**	4.10 ± 0.19	2.56 ± 0.08
	1.0	**2.70 ± 0.24**	**3.20 ± 0.22**	3.74 ± 0.63	**3.32 ± 0.24**
	2.0	**3.29 ± 0.05**	**3.73 ± 0.19**	4.39 ± 0.28	**4.40 ± 0.10**
day 21	**0.0**	13.96 ± 0.69	19.94 ± 1.04	13.71 ± 2.17	13.17 ± 1.20
	0.25	13.39 ± 0.46	20.19 ± 0.77	13.54 ± 1.56	12.97 ± 1.64
	0.5	14.97 ± 0.49	20.78 ± 1.14	13.63 ± 2.62	14.51 ± 1.09
	1.0	**15.09 ± 1.23**	22.13 ± 1.03	**14.70 ± 1.01**	**18.29 ± 0.58**
	2.0	**17.97 ± 0.61**	**24.13 ± 0.63**	**17.72 ± 1.04**	**20.39 ± 0.54**

Note: Mean values and standard deviations are shown (*n*=4). Bold letters indicate statistically significant differences to the control group (p<0.05, Jonckheere-Terpstra test).

Table 27. Results from stage determination (Nieuwkoop and Faber, 1994) on days 7 and 21 in different experiments with thyroxine (T4) as test substance

	T4		lab 1	lab 2	lab 3	lab 5
	µg/	r	mean ±	mean ±	mean ±	mean ±
day 7	0.0	1	55 (54-55)	54 (54-54)	54 (52-55)	54 (54-55)
		2	55 (54-55)	54 (54-54)	54 (54-55)	54 (54-55)
		3	55 (54-55)	54 (54-55)	55 (53-55)	54 (53-55)
		4	55 (54-55)	54 (54-54)	54 (54-55)	55 (54-55)
	0.25	1	55 (54-55)	54 (54-55)	55 (54-55)	55 (55-55)
		2	55 (55-55)	54 (54-55)	55 (54-55)	54 (54-55)
		3	55 (54-55)	54 (54-55)	55 (54-56)	54 (54-55)
		4	55 (55-55)	54 (54-55)	54 (53-56)	54 (54-55)
	0.5	1	55 (55-55)	54 (54-55)	55 (54-55)	54 (54-55)
		2	55 (54-55)	54 (54-55)	54 (53-56)	54 (54-55)
		3	55 (55-55)	54 (54-55)	53 (53-54)	55 (54-55)
		4	55 (55-55)	54 (54-55)	54 (54-56)	54 (54-55)
	1.0	1	55 (55-55)	55 (55-55)	55 (54-55)	**55 (54-55)**
		2	55 (55-55)	55 (54-55)	55 (55-55)	**55 (54-55)**
		3	55 (55-55)	55 (55-55)	55 (54-55)	**55 (54-55)**
		4	55 (55-55)	55 (54-55)	55 (54-56)	**55 (55-55)**
	2.0	1	55 (55-55)	55 (55-55)	55 (53-55)	**55 (55-56)**
		2	55 (55-55)	55 (55-55)	55 (55-56)	**55 (55-56)**
		3	55 (55-55)	55 (55-56)	56 (55-56)	**56 (54-56)**
		4	55 (55-56)	55 (55-55)	54 (54-55)	**55 (55-56)**
day 21	0.0	1	58 (57-60)	60 (59-62)	59 (55-66)	57 (56-60)
		2	58 (57-61)	60 (59-60)	59 (56-64)	57 (56-58)
		3	58 (57-60)	60 (58-62)	59 (57-62)	57 (56-59)
		4	58 (57-60)	59 (59-61)	59 (57-65)	57 (56-59)
	0.25	1	57 (57-59)	59 (59-61)	59 (57-66)	56 (56-59)
		2	58 (57-60)	60 (59-62)	59 (57-66)	58 (57-59)
		3	58 (57-61)	59 (58-60)	60 (58-66)	58 (57-59)
		4	58 (57-62)	59 (59-62)	59 (57-64)	57 (56-59)
	0.5	1	58 (57-59)	59 (59-60)	59 (58-66)	58 (57-59)
		2	58 (57-60)	60 (59-62)	59 (58-63)	58 (57-60)
		3	58 (57-60)	60 (59-62)	60 (59-66)	58 (56-60)
		4	58 (58-60)	60 (59-62)	62 (58-66)	57 (56-59)
	1.0	1	**59 (57-61)**	60 (59-62)	**60 (59-66)**	**59 (58-61)**
		2	**59 (58-59)**	60 (59-62)	**63 (59-66)**	**58 (58-60)**
		3	**58 (57-62)**	60 (59-62)	**60 (59-66)**	**59 (58-61)**
		4	**59 (58-62)**	59 (58-62)	**60 (59-66)**	**58 (58-62)**
	2.0	1	**60 (59-62)**	**61 (59-62)**	**65 (63-66)**	**62 (59-63)**
		2	**60 (59-62)**	**60 (59-62)**	**66 (61-66)**	**59 (58-62)**
		3	**60 (58-61)**	**60 (59-62)**	**63 (60-66)**	**60 (58-62)**
		4	**60 (58-62)**	**61 (59-62)**	**64 (61-66)**	**60 (58-63)**

Note: Values shown are the median stage for each replicate tank with the total range of stages in parentheses. Bold letters indicate statistically significant differences in median stages relative to the control group ($p<0.05$, Jonckheere-Terpstra test). [a] replicate tank

Effects of T4 on Thyroid Gland Histology

63. The study protocol called for evaluation of thyroid gland histology in five specimen randomly sampled from each replicate tank (*n*=20 per treatment) according to a graded scoring system which incorporates information on incidence and severity of specific histological changes. At the time of report preparation, results from histological evaluation of thyroid tissue were available for three of four experiments (see Table 28, Table 29, and Table 30). Overall, T4 treatment of tadpoles caused, if any, only mild effects on thyroid gland histology. In lab 1, variable effects of T4 treatment were observed on thyroid gland size. Mild to moderate glandular hypertrophy was noted in a number of tadpoles exposed to 0.5 µg/L and 2.0 µg/L. In contrast, mild glandular atrophy was observed in tadpoles exposed to 0.25 and 2.0 µg/L T4. Neither the incidence of glandular hypertrophy or atrophy showed a clear concentration dependency. Mild glandular atrophy was also observed in 7 of 10 tadpoles from the 1.0 µg/l T4 treatment in lab 3 but not in other treatments. In the T4 study in lab 2, no distinct changes in glandular size were observed and this result was confirmed by morphometric analyses (Figure 10).

64. Observations on effects of T4 treatment on the morphology of follicular cells were also heterogenous among different studies. In lab 1, histological alterations were mainly detected at the two highest T4 concentrations of 1.0 and 2.0 µg/l and included mild to moderate follicular cell hypertrophy, mild colloid depletion and a slight increase in the incidence of mild follicular cell hyperplasia. Morphometric analyses of epithelial cell height (Figure 9) and thyroid volume (Figure 9) in lab 1 confirmed the results from the severity grading approach, showing significant increases in epithelial cell height (follicular cell hypertrophy) at 1.0 and 2.0 µg/L T4 but no statistical significant differences in total glandular volume (thyroid gland hypertrophy) between treatments. It should be noted that the thyroid hypertrophy observed at 1.0 and 2.0 µg/l T4 in lab 1 was detected in tadpoles at more advanced stages of development compared to the control animals which provided the reference tissues. In lab 2, no similar effects on the epithelial cell layer were detected and lab 3 reported rather a decrease in the incidence of mildly hypertrophic follicular cells at the higher T4 concentrations.

Table 28. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to thyroxine (T4) in lab 1

T4	glandular hypertrophy		glandular atrophy		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	2/20	0.10±0.31 (1)	1/20	0.05±0.22 (1)	6/20	0.30±0.47 (1)	0/20	0.00±0.00 (0)	57-60
0.25	2/20	0.10±0.31 (1)	7/20	0.45±0.69 (2)	7/20	0.35±0.49 (1)	0/20	0.00±0.00 (0)	57-60
0.50	8/20	0.50±0.69 (2)	5/20	0.25±0.44 (1)	5/20	0.25±0.44 (1)	0/20	0.00±0.00 (0)	57-60
1.0	3/20	0.15±0.37 (1)	9/20	0.55±0.69 (2)	12/20	0.70±0.66 (2)	7/20	0.35±0.49 (1)	57-60
2.0	6/20	0.40±0.60 (2)	4/20	0.20±0.41 (1)	20/20	1.25±0.44 (2)	9/20	0.55±0.69 (2)	59-62

Note: [a] number of animals affected / number of animals examined
[b] values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses
[c] values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology

Table 29. Incidence and severity of selected histological alterations in X. laevis tadpoles exposed to thyroxine (T4) in lab 2

T4	glandular hypertrophy		glandular atrophy		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	0/20	0.00±0.00 (0)	0/20	0.00±0.00 (0)	0/20	0.00±0.00 (0)	0/20	0.00±0.00 (0)	59-62
0.25	0/20	0.00±0.00 (0)	0/20	0.00±0.00 (0)	0/20	0.00±0.00 (0)	0/20	0.00±0.00 (0)	58-62
0.50	0/20	0.00±0.00 (0)	0/20	0.00±0.00 (0)	1/20	0.05±0.22 (1)	0/20	0.00±0.00 (0)	59-62
1.0	0/20	0.00±0.00 (0)	3/20	0.15±0.37 (1)	0/20	0.00±0.00 (0)	0/20	0.00±0.00 (0)	59-62
2.0	1/20	0.05±0.22 (1)	7/20	0.35±0.49 (1)	1/20	0.05±0.22 (1)	0/20	0.00±0.00 (0)	59-62

Note: [a] number of animals affected / number of animals examined
[b] values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses
[c] values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology

Table 30. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to thyroxine (T4) in lab 3. Note that a 5-grade scoring scheme was used in this study that differed slightly from the 3-grade scheme used by the other labs

T4	colloid depletion		glandular atrophy		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	0/10	0.00±0.00 (0)	0/10	0.00±0.00 (0)	6/10	0.70±0.67 (2)	4/10	0.40±0.52 (1)	58-59
0.25	0/10	0.00±0.00 (0)	0/10	0.00±0.00 (0)	8/10	1.20±0.79 (2)	4/10	0.40±0.52 (1)	58-59
0.50	0/10	0.10±0.32 (1)	2/10	0.40±0.84 (2)	6/10	1.10±0.99 (2)	1/10	0.10±0.32 (1)	59-60
1.0	7/10	2.40±1.84 (5)	7/10	2.10±1.52 (4)	1/10	0.20±0.63 (2)	1/10	0.10±0.32 (1)	59-61
2.0	0/10	0.00±0.00 (0)	0/10	0.00±0.00 (0)	0/10	0.00±0.00 (0)	0/10	0.00±0.00 (0)	65-66

Note: [a] number of animals affected / number of animals examined

[b] values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses

[c] values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology

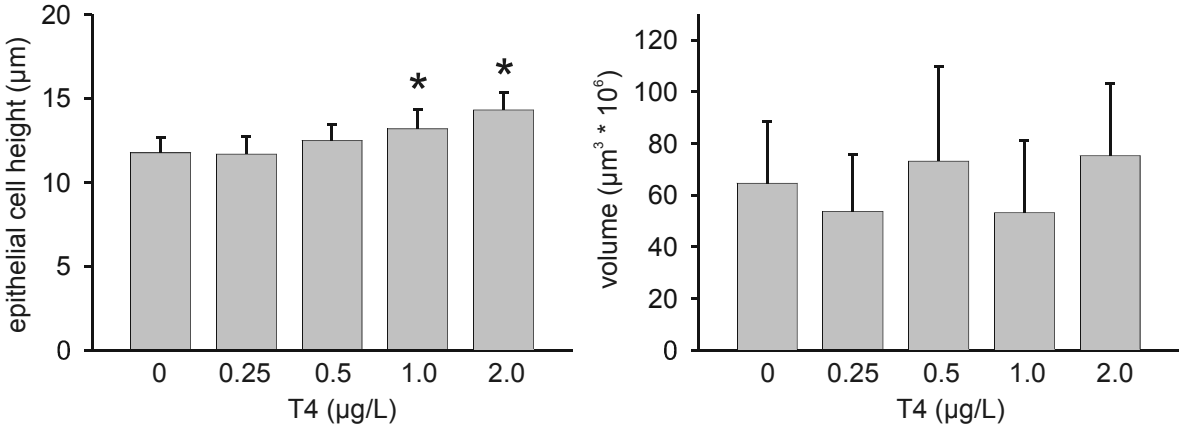

Figure 9. Morphometric analysis of thyroid tissue for changes in epithelial cell height and glandular volume following exposure to thyroxine (T4) in the lab 1 study.

Columns with vertical bars represent means with standard deviations (n= 20 tadpoles per treatment). Asterisks denote significant differences compared to the control treatment group ($p<0.05$ Dunnett's test).

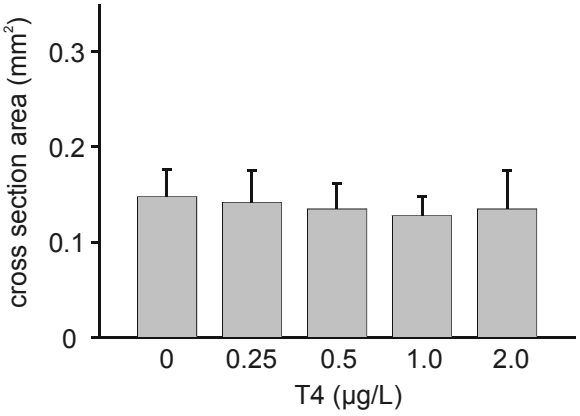

Figure 10. Morphometric analysis of thyroid tissue for changes in maximum cross section area following exposure to thyroxine (T4) in the lab 2 study.

Columns with vertical bars represent means with standard deviations (n= 20 tadpoles per treatment). No significant differences compared to the control treatment group were detected ($p<0.05$ Dunnett's test).

Test Results – Iopanoic Acid

Overview

65. During the validation phase-2 study of the amphibian metamorphosis assay, a total of five experiments with iopanoic acid (IOP) as test substance were performed in five different laboratories (lab 1, lab 2, lab 3, lab 5, and lab 6). The nominal aqueous concentrations for IOP were 0, 0.75, 1.5, 3.0, and 6.0 mg/l. The experiment in lab 6 included an additional treatment of tadpoles with 0.375 mg/l IOP. Analysis of growth-related parameters (e.g., WBL, SVL, wet weight) and mortality rates indicated mild effects of IOP treatment on tadpole growth and survival in some laboratories. In one experiment, the observation of marked growth retardation in all IOP treatment groups was regarded as the result from overt non-thyroidal toxicity by IOP treatment. With regard to metamorphic development, IOP treatment was associated with a retardation in hind limb growth and asynchronous morphological development. The latter response refers to observations made in several labs that some morphological landmarks in IOP-treated tadpoles corresponded to different developmental stages according to the criteria by Nieuwkoop and Faber (1994). It was therefore not possible to assign a specific development stage to many of the IOP-treated animals. Histological assessment of thyroid tissue after IOP exposure was performed in four labs. IOP treatment caused only mild alterations in thyroid gland histology including mild to moderate glandular hypertrophy which was associated with mild to moderate follicular cell hypertrophy.

General Test Performance

66. An initial assessment of the test performance for the five experiments conducted with IOP in different labs indicated that the test performed with IOP in lab 3 was characterized by increased variability in control stage development (observation of 9 different stages on day 21, ranging from stage 58 to stage 66) and lower growth rates compared to other labs. A stage-matched comparison of growth-related parameters between laboratories showed that tadpoles in the IOP experiment of lab 3 had on average less than 50% of the wet weight compared to tadpoles used in other labs. The low feeding rates used in this lab are most likely responsible for this deviation from the other experiments (see section 0).

67. The study protocol called for verification of aqueous IOP concentrations in each treatment tank at least once per week. Overall, the actual test concentrations for IOP in the different tests were close to nominal target concentrations as shown in

Table 31.

Table 31. Measured concentrations (mg/L) for iopanoic acid (IOP)

nominal conc.	replicate	lab 1	lab 2	lab 3	lab 5	lab 6[a]
		mean ± SD (*n*)	mean ± SD (*n*)	mean ± SD (*n*)	mean ± SD (*n*)	mean ± SD (*n*)
0.0	1	n.d.	n.d.	n.d.	n.d.	n.d.
	2	n.d.	n.d.	n.d.	n.d.	n.d.
	3	n.d.	n.d.	0.11 ± 0.22 (4)	n.d.	n.d.
	4	n.d.	n.d.	n.d.	n.d.	n.d.
0.75	1	0.79 ± 0.10 (5)	0.59 ± 0.17 (4)	0.70 ± 0.06 (4)	0.69 ± 0.15 (4)	0.75 ± 0.03 (6)
	2	0.79 ± 0.08 (4)	0.67 ± 0.12 (4)	0.68 ± 0.09 (4)	0.76 ± 0.13 (5)	0.72 ± 0.03 (6)
	3	0.80 ± 0.11 (4)	0.78 ± 0.05 (4)	0.67 ± 0.10 (4)	0.66 ± 0.14 (4)	0.80 ± 0.04 (6)
	4	0.80 ± 0.09 (5)	0.76 ± 0.04 (4)	0.68 ± 0.10 (4)	0.85 ± 0.40 (4)	0.80 ± 0.04 (6)
1.5	1	1.51 ± 0.10 (5)	1.41 ± 0.25 (4)	1.40 ± 0.11 (4)	1.66 ± 0.13 (3)	1.51 ± 0.08 (6)
	2	1.46 ± 0.09 (4)	1.41 ± 0.23 (4)	1.44 ± 0.12 (4)	1.58 ± 0.13 (4)	1.54 ± 0.03 (6)
	3	1.52 ± 0.10 (4)	1.65 ± 0.09 (4)	1.39 ± 0.14 (4)	1.61 ± 0.09 (4)	1.56 ± 0.04 (6)
	4	1.53 ± 0.11 (4)	1.67 ± 0.05 (4)	1.39 ± 0.13 (4)	1.60 ± 0.12 (5)	1.53 ± 0.05 (6)
3.0	1	2.85 ± 0.31 (5)	2.78 ± 0.47 (4)	3.01 ± 0.64 (4)	3.05 ± 0.51 (4)	3.02 ± 0.13 (6)
	2	2.83 ± 0.21 (4)	2.74 ± 0.43 (4)	3.02 ± 0.48 (4)	3.24 ± 0.44 (4)	3.05 ± 0.10 (6)
	3	2.84 ± 0.14 (4)	3.03 ± 0.17 (4)	2.82 ± 0.17 (4)	3.04 ± 0.62 (5)	3.08 ± 0.08 (6)
	4	3.05 ± 0.32 (5)	3.02 ± 0.21 (4)	2.78 ± 0.18 (4)	3.30 ± 0.49 (4)	3.09 ± 0.08 (6)
6.0	1	5.65 ± 0.33 (5)	5.55 ± 1.11 (4)	6.43 ± 0.49 (4)	5.85 ± 0.41 (4)	6.20 ± 0.14 (6)
	2	5.60 ± 0.21 (4)	5.59 ± 1.09 (4)	6.43 ± 0.52 (4)	6.04 ± 0.41 (5)	6.18 ± 0.11 (6)
	3	5.58 ± 0.22 (4)	6.14 ± 0.13 (4)	6.35 ± 0.49 (4)	5.46 ± 1.04 (4)	6.15 ± 0.10 (6)
	4	5.60 ± 0.25 (5)	6.26 ± 0.12 (4)	6.28 ± 0.50 (4)	5.61 ± 0.95 (4)	6.15 ± 0.13 (6)

Note: Mean values (mean measured concentration across one replicate tank through time) and standard deviations are shown and the number of replicate measurements for each tanks are given in parentheses. [a] Lab 6 also tested a concentration of 0.375 mg/l IOP and reported mean concentrations between 0.37 and 0.41 mg/l IOP for the replicate tanks of this treatment group. n.d. not detectable

Mortality

68. Mortality of tadpoles was assessed daily throughout the exposure phase and the total number of dead test animals observed after 21 days of exposure is reported in Table 32 for all five tests with IOP. No mortality was observed in any treatment group in the test performed in lab 5 and treatment-related mortality was very low (< 5%) in the test performed in lab 1. Increased mortality (≥ 5%) at the highest IOP test concentration was observed in lab 2 (6.25%), lab 3 (5%), and lab 6 (15%). In lab 6, an increased mortality (10%) was also noted in the 1.5 mg/L IOP treatment group.

Table 32. Mortality following 21-d exposure of tadpoles to the indicated nominal concentrations of iopanoic acid (IOP) as observed in different experiments

nominal mg/l	lab 1	lab 2	lab 3	lab 5	lab 6[a]
0.0	1 (1.25%)	0	2 (2.5%)	0	3 (3.75%)
0.75	2 (2.5%)	0	2 (2.5%)	0	1 (1.25%)
1.5	0	1 (1.25%)	4 (5.0%)	0	8 (10%)
3.0	0	1 (1.25%)	0	0	2 (2.5%)
6.0	0	5 (6.25%)	4 (5.0%)	0	12 (15%)

Note: Data are presented as the absolute number of dead animals with the percentages of dead animals for each treatment group in parentheses. [a] Lab 6 also tested a concentration of 0.375 mg/l IOP and reported 3 dead animals (3.75% mortality) in this treatment group.

Effects of IOP on Growth-related Endpoints

69. Possible effects of IOP treatment on tadpole growth were assessed by morphometric measurements of tadpole size (WBL, SVL) and tadpole weight (wet weight). Measurements were performed for a subset of animals (n=5 animals per replicate tank) on study day 7 and for the remaining animals on study day 21. Day 7 measurements indicate that treatment of tadpoles with IOP did not result in growth retardation during the initial exposure phase (Table 33, Table 34, Table 35). No significant deviation from growth in the control group was observed on day 7 for any of the tested IOP concentrations in four experiments. The only exception to this effects profile of IOP on growth-related parameters on day 7 was the test run with IOP in lab 3. In this test, all IOP concentrations caused a marked retardation in tadpole growth with significant effects being observed for all of the three endpoints WBL, SVL, and wet weight.

70. When growth-related endpoints were analyzed on study day 21, all IOP experiments showed a significant reduction in tadpole size and weight due to IOP treatment (Table 33, Table 34, Table 35). Notably, in three IOP tests (lab 1, lab 2, and lab 6), significantly reduced mean values were detected at all tested IOP concentrations for each of the growth-related parameters WBL, SVL, and wet weight. Less marked effects of IOP were noted in one lab (lab 5), where WBL and wet weight were significantly reduced only at the two highest IOP concentrations (1.0 and 2.0 µg/L) and SVL was significantly reduced only at the highest IOP concentration (2.0 µg/L). Overall, the parameters WBL, SVL and wet weight displayed a similar sensitivity to detect treatment-related changes in tadpole growth in response to IOP treatment. A marked exception to this effects profile of IOP on growth-related parameters on day 21 was the test run with IOP in lab 3. In this test, no effects of IOP treatment were detectable for the parameters

WBL and SVL and only the highest IOP concentration (6.0 mg/l) caused a significant reduction of tadpole wet weight.

Table 33. Results from whole body length measurements (mm) on days 7 and 21 as determined in different experiments with iopanoic acid (IOP) as test substance

	IOP mg/l	lab 1 mean ± SD	lab 2 mean ± SD	lab 3 mean ± SD	lab 5 mean ± SD	lab 6 mean ± SD
day 7	**0.0**	47.04 ± 0.74	43.71 ± 0.95	37.72 ± 4.35	44.04 ± 1.71	39.80 ± 0.60
	0.75	46.30 ± 1.79	43.56 ± 2.50	**26.31 ± 2.11**	43.40 ± 1.24	39.77 ± 1.56
	1.5	46.90 ± 0.57	44.79 ± 0.88	**25.83 ± 2.47**	43.15 ± 2.14	39.86 ± 1.64
	3.0	46.46 ± 1.50	43.44 ± 1.49	**26.84 ± 3.68**	44.65 ± 1.52	39.93 ± 0.66
	6.0	46.93 ± 0.80	41.77 ± 1.08	**24.15 ± 2.67**	43.89 ± 1.43	39.35 ± 1.19
day 21	**0.0**	70.04 ± 0.44	65.96 ± 1.07	35.39 ± 8.55	65.97 ± 1.21	53.58 ± 1.47
	0.75	**68.59 ± 0.59**	**60.97 ± 0.71**	39.37 ± 4.05	67.19 ± 1.25	**47.26 ± 1.74**
	1.5	**68.29 ± 0.50**	**57.55 ± 1.11**	41.04 ± 3.78	65.13 ± 0.90	**46.11 ± 3.68**
	3.0	**67.58 ± 0.58**	**55.45 ± 1.97**	38.34 ± 4.17	**63.25 ± 0.68**	**46.94 ± 1.51**
	6.0	**67.00 ± 1.13**	**54.57 ± 1.12**	41.27 ± 3.86	**62.81 ± 2.02**	**46.49 ± 1.02**

Note: Mean values and standard deviations are shown (*n*=4). Bold letters indicate statistically significant differences to the control group (p<0.05, Jonckheere-Terpstra test).

Table 34. Results from snout-vent length measurements (mm) on days 7 and 21 as determined in different experiments with iopanoic acid (IOP) as test substance

	IOP mg/l	lab 1 mean ± SD	lab 2 mean ± SD	lab 3 mean ± SD	lab 5 mean ± SD	lab 6 mean ± SD
day 7	**0.0**	17.91 ± 0.45	17.37 ± 0.50	14.76 ± 1.84	15.39 ± 0.47	16.12 ± 0.36
	0.75	17.99 ± 1.00	16.86 ± 0.91	**10.90 ± 0.75**	14.93 ± 0.51	16.10 ± 0.65
	1.5	17.95 ± 0.12	17.45 ± 0.71	**10.25 ± 0.87**	14.95 ± 0.68	15.92 ± 0.87
	3.0	17.71 ± 0.43	16.93 ± 0.53	**10.62 ± 0.98**	15.19 ± 0.20	16.30 ± 0.17
	6.0	17.70 ± 0.15	16.21 ± 0.39	**9.64 ± 1.06**	14.81 ± 0.30	15.72 ± 0.26
day 21	**0.0**	25.53 ± 0.35	24.13 ± 0.37	17.23 ± 1.07	21.52 ± 0.38	18.78 ± 0.38
	0.75	**24.70 ± 0.40**	**21.44 ± 0.78**	18.21 ± 0.44	21.86 ± 0.27	**18.00 ± 0.52**
	1.5	**24.60 ± 0.30**	**20.19 ± 0.79**	17.88 ± 1.21	21.57 ± 0.43	**18.04 ± 0.35**
	3.0	**24.10 ± 0.34**	**19.95 ± 0.63**	16.86 ± 0.58	20.48 ± 0.54	**17.75 ± 0.25**
	6.0	**23.72 ± 0.50**	**19.42 ± 0.24**	16.35 ± 1.22	**20.62 ± 0.69**	**17.58 ± 0.26**

Note: Mean values and standard deviations are shown (*n*=4). Bold letters indicate statistically significant differences to the control group (p<0.05, Jonckheere-Terpstra test).

Table 35. Results from wet weight measurements (mg) on days 7 and 21 as determined in different experiments with iopanoic acid (IOP) as test substance

	IOP mg/l	lab 1 mean ± SD	lab 2 mean ± SD	lab 3 mean ± SD	lab 5 mean ± SD	lab 6 mean ± SD
day 7	**0.0**	441 ± 21	412 ± 28	245 ± 74	373 ± 35	298 ± 17
	0.75	426 ± 67	409 ± 64	**95 ± 17**	339 ± 26	298 ± 29
	1.5	434 ± 11	447 ± 50	**80 ± 16**	341 ± 48	301 ± 47
	3.0	420 ± 23	410 ± 40	**93 ± 20**	374 ± 42	300 ± 14
	6.0	421 ± 24	369 ± 41	**74 ± 19**	368 ± 19	266 ± 9
day 21	**0.0**	1307 ± 14	1236 ± 63	485 ± 102	1107 ± 57	736 ± 38
	0.75	**1173 ± 52**	**932 ± 60**	440 ± 35	1119 ± 21	**612 ± 51**
	1.5	**1146 ± 34**	**831 ± 50**	450 ± 72	1032 ± 69	**607 ± 37**
	3.0	**1097 ± 29**	**806 ± 37**	375 ± 67	**879 ± 70**	**577 ± 17**
	6.0	**1043 ± 29**	**793 ± 60**	**327 ± 68**	**913 ± 66**	**533 ± 9**

Note: Mean values and standard deviations are shown (*n*=4). Bold letters indicate statistically significant differences to the control group (p<0.05, Jonckheere-Terpstra test).

Effects of IOP on Developmental Endpoints

71. With regard to metamorphic development, a main finding was that treatment of tadpoles with IOP caused a condition that could be described as asynchronous morphological development. In the case of the IOP-treated tadpoles, asynchronous morphological development referred to a condition where the various morphological landmarks normally used for staging showed characteristics of distinct development stages. As a consequence, it was not possible to determine the specific stage of a test animal. The study protocol did not specifically call for counting the number of tadpoles showing signs of asynchronous morphological development and no criteria were provided in the study protocol that could be used to identify such cases of maldevelopment in a specific and consistent manner. Hence, it depended very much on the observer performing the stage determination to identify and judge cases of asynchronous morphological development in the IOP experiments. Accordingly, reliable quantitative data concerning the incidence and severity of maldevelopment are not available. However, three labs provided some records on the incidence of remarkable cases of asynchronous morphological development (Table 36). The prevailing type of morphological abnormalities in IOP-treated animals could be characterized as retarded hind limb morphogenesis relative to advanced metamorphic changes of head structures. Another observation was that developmental changes in tail tissue appeared more advanced relative to hind limb development. The occurrence of such morphological abnormalities represents a confounding factor for a statistical assessment of median stages and individual labs dealed very differently with this problem. It is therefore not appropriate to statistically analyze the stage data summarized in Table 37.

Table 36. Percentage (%) of remarkable cases of asynchronous morphological development due to IOP treatment. Observations were exclusively reported on study day 21

	nominal conc.	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 5 mean ±	lab 6 mean ±
day 21	**0.0**	n.d.	n.d.	-	n.d.	-
	0.75	n.d.	53 ± 25	-	n.d.	-
	1.5	10 ± 4	51 ± 16	-	5 ± 6	-
	3.0	20 ± 9	43 ± 15	-	20 ± 24	-
	6.0	37 ± 4	24 ± 19	-	22 ± 14	-

Note: Mean values and standard deviations are shown (*n*=4). n.d. not detected; - observation of asynchronous morphological development was reported but cases were not quantified.

Table 37. Results from stage determination (Nieuwkoop and Faber, 1994) on days 7 and 21 in different experiments with iopanoic acid (IOP) as test substance

	IOP m	Rep	lab 1 mean ±	lab 2 mean ±	lab 3 mean ± SD	lab 5 mean ±	lab 6 mean ±
day 7	0.0	1	55 (54-55)	54 (54-55)	53 (53-55)	55 (54-55)	54 (54-54)
		2	55 (55-55)	55 (54-55)	54 (53-55)	54 (53-55)	54 (53-55)
		3	55 (54-55)	54 (54-55)	53 (53-53)	54 (54-55)	54 (54-55)
		4	55 (55-	55 (55-	53 (53-55)	54 (54-55)	54 (54-54)
	0.75	1	55 (55-55)	55 (54-55)	53 (52-53)	54 (54-54)	54 (53-54)
		2	55 (54-55)	54 (53-55)	52 (52-53)	54 (53-55)	54 (53-54)
		3	55 (54-55)	54 (53-55)	52 (52-52)	54 (54-54)	54 (54-54)
		4	55 (54-	55 (55-55)	52 (52-52)	54 (53-55)	54 (53-55)
	1.5	1	55 (54-55)	55 (54-55)	52 (52-53)	54 (53-55)	54 (53-55)
		2	55 (55-55)	55 (54-55)	52 (52-52)	54 (53-55)	54 (54-55)
		3	55 (55-55)	54 (54-55)	52 (52-52)	54 (54-54)	54 (54-55)
		4	55 (54-	55 (55-55)	52 (52-52)	55 (54-55)	54 (54-54)
	3.0	1	55 (54-55)	55 (54-55)	52 (52-52)	54 (54-54)	54 (54-54)
		2	55 (54-55)	54 (54-55)	52 (52-53)	54 (53-55)	54 (53-55)
		3	55 (54-55)	55 (54-55)	53 (52-53)	54 (54-55)	54 (53-55)
		4	55 (54-	54 (54-55)	53 (53-53)	54 (54-55)	54 (53-55)
	6.0	1	55 (55-55)	55 (55-55)	53 (52-53)	54 (53-55)	54 (54-55)
		2	55 (55-55)	55 (54-55)	52 (52-54)	54 (54-54)	54 (53-55)
		3	55 (55-55)	55 (54-55)	52 (52-52)	54 (54-55)	53 (53-53)
		4	55 (54-	55 (54-55)	52 (52-53)	54 (54-55)	54 (54-54)
day 21	0.0	1	58 (57-59)	59 (58-59)	60 (58-65)	57 (56-60)	62 (59-63)
		2	58 (57-59)	59 (57-61)	64 (58-65)	58 (56-59)	60 (59-64)
		3	58.5 (57-59)	58 (57-60)	65 (58-66)	58 (56-60)	60 (58-63)
		4	58 (57-61)	59 (58-61)	62 (58-65)	58 (56-60)	61 (57-64)
	0.75	1	58 (56-61)	57.5 (56-59)	58.5 (57.5-65)	57 (56-59)	61 (58-64)
		2	58 (56-61)	58 (57-58)	58.5 (58-64)	57 (56-59)	62 (57-65)
		3	57 (56-58)	58 (58-61)	58.5 (57.5-64)	58 (56-59)	61 (58-64)
		4	58 (56-61)	58 (58-58)	58.5 (57.5-64)	58 (56-62)	62 (57-65)
	1.5	1	58 (57-61)	58.5 (58-59)	59 (56-65)	58 (56-59)	61 (58-64)
		2	58 (56-60)	58 (58-59)	60 (56.5-65)	58 (56-59)	63 (57-64)
		3	58 (56-60)	58 (58-59)	56.5 (56.5-63.5)	58 (56-61)	62.5 (58-65)
		4	58 (57-61)	58 (57-59)	57 (56.5-62)	58 (57-62)	62 (59-64)
	3.0	1	58 (57-62)	58.5 (58-59)	57 (56.5-62.5)	59 (57-62)	61 (57-64)
		2	58 (57-61)	58 (58-59)	57 (56.5-65)	58 (56-62)	61 (59-64)
		3	57 (56-58)	58 (58-59)	59.5 (56-64)	59 (56-62)	62 (57-65)
		4	58 (57-61)	58 (58-62)	60 (56-62)	59 (58-62)	61 (56-64)
	6.0	1	58 (57-61)	58 (58-59)	56.8 (56.5-60.5)	58 (56-59)	61 (57-64)
		2	58 (57-60)	58 (58-59)	57.5 (56.5-61)	58 (56-59)	62 (58-64)
		3	58 (57-61)	58 (58-59)	57 (56-61)	59 (57-62)	62 (58-63)
		4	58 (56-61)	58 (58-59)	56.5 (56-61)	58 (56-63)	62 (59-65)

Note: Values shown are the median stage for each replicate tank with the total range of stages in parentheses. Note that in many cases, no definitive stage could be determined and the results summarized in this table represent therefore only the proportion of test animals for which staging was possible.

72. With the exception of one experiment (lab 3), IOP effects on tadpole development were almost exclusively detected on study day 21. In addition to the qualitative observation of asnychronous development, morphometric analyses of HLL on study day 21 revealed reductions in mean HLL in response to all IOP concentrations in three IOP experiments, whilst no significant effect on HLL was detectable in one test in lab 5 (Table 38). The reason for this drastic difference is not known. However, in general, the possible effects of IOP on hind limb growth have to be interpreted with caution. First, for the selected concentration range of IOP, data from HLL measurements on day 21 indicated the presence of a non-monotone concentration-response relationship in two labs (lab 1 and lab 2). While this prevented the use of the sensitive Jonckheere-Terpstra test, statistical analyses using either the Williams test or pairwise comparisons with the control group (Dunnett's test) still detected significantly reduced HLL in all IOP treatment groups. Secondly, a more important problem complicating the interpretation of reduced HLL in IOP treatments is the fact that concurrent WBL and SVL measurements indicated an overall reduction in tadpole size in the IOP treatments.

Table 38. Results from hind limb length measurements (mm) on days 7 and 21 as determined in different experiments with iopanoic acid (IOP) as test substance

	IOP mg/l	lab 1 mean ±	lab 2 mean ±	lab 3 mean ±	lab 5 mean ±	lab 6 mean ±
day 7	**0.0**	2.32 ± 0.10	2.13 ± 0.13	2.16 ± 0.69	2.26 ± 0.06	2.61 ± 0.11
	0.75	2.15 ± 0.12	2.05 ± 0.28	**0.84 ± 0.13**	2.14 ± 0.10	2.46 ± 0.23
	1.5	2.22 ± 0.07	2.17 ± 0.09	**0.80 ± 0.19**	2.09 ± 0.10	2.61 ± 0.19
	3.0	2.24 ± 0.12	2.02 ± 0.20	**0.91 ± 0.27**	2.21 ± 0.23	2.66 ± 0.11
	6.0	2.35 ± 0.09	2.22 ± 0.20	**0.91 ± 0.22**	2.15 ± 0.14	2.62 ± 0.19
day 21	**0.0**	13.96 ± 0.41	17.03 ± 1.02	15.40 ± 0.96	13.73 ± 0.46	17.50 ± 0.45
	0.75	**11.26 ± 0.97**	**14.97 ± 1.25**	**9.28 ± 1.56**	11.92 ± 0.90	**15.72 ± 0.62**
	1.5	**11.44 ± 0.25**	**15.89 ± 0.89**	**9.62 ± 1.79**	12.48 ± 2.12	**15.89 ± 0.72**
	3.0	**11.31 ± 1.26**	**15.54 ± 0.90**	**8.58 ± 1.60**	13.58 ± 0.79	**15.47 ± 0.58**
	6.0	**11.82 ± 0.56**	**13.27 ± 0.49**	**7.01 ± 1.49**	12.99 ± 0.90	**15.16 ± 0.58**

Note: Mean values and standard deviations are shown (n=4). Bold letters indicate statistically significant differences to the control group ($p<0.05$, Jonckheere-Terpstra test).

Effects of IOP on Thyroid Gland Histology

73. The study protocol called for evaluation of thyroid gland histology in five specimen randomly sampled from each replicate tank (n=20 per treatment) according to a graded scoring system which incorporates information on incidence and severity of specific histological changes. At the time of report preparation, results from histological evaluation of thyroid tissue were available for four of five experiments (see Table 39, Table 40, Table 41, and Table 42).

74. An initial assessment of histological data sets from these four studies idendified the data set from lab 3 as being not appropriate for an analysis of exposure-related changes in thyroid tissue. The major

problem with this data set was the marked deviation of developmental stages between the control group and the various IOP treatment groups. Specifically, all control tissue samples were from animals at developmental stages 64-65, whereas thyroid tissue was sampled from IOP-treated animals at stages 55-59. In addition to the marked mismatch of developmental stages of sampled animals in different treatment groups, it is particularly problematic that all reference tissues (control group) are derived from animals at late climax stages when the stimulation of hypothalamus-pituitary-thyroid axis is qualitatively and quantitatively very different from prometamorphic stages. The following analysis of IOP exposure related alterations in thyroid tissue will therefore exclusively focus on the results from the three studies performed in lab 1, lab 2 and lab 6.

75. Qualitatively similar changes in thyroid tissue were observed in all three experiments with IOP. Histological alterations included mild to moderate increases in thyroid gland size and mild to moderate follicular cell hypertrophy. Follicular cells were considered to be hypertrophic if they were tall columnar cells as opposed to simple cuboidal cells. Tall columnar cells occurred in thyroid glands of controls, but the number of animals affected and the severity of the change was increased in various IOP treatments. Mild hyperplasia of the follicular epithelium was seen in some tissue samples but was not as prominent as the former two alterations. The histological appearance of thyroids from IOP-treated animals was unique inasmuch as increases in glandular size and follicular cell hypertrophy were in most cases not associated with distinct changes in colloid content. Thus, the glands from IOP-treated animals did not show the entire spectrum of histological alterations that is typical for a moderate increase of glandular activity in *X. laevis* tadpoles.

76. Another finding from the histological assessment of thyroid tissue was that a clear concentration-response relationship could not be established for the tested concentration range of IOP. Notably, the highest mean severity values for follicular cell hypertrophy were observed in the 3.0 but not the 6.0 mg/l IOP treatment. Likewise, the incidence of affected animals differed not markedly among the various IOP treatments.

77. In addition, morphometric techniques were used by lab 1 and lab 2 to assess effects of IOP treatment on epithelial cell height (Figure 11), thyroid volume (Figure 11) and thyroid cross section area (Figure 12). The morphometric data are consistent with the results from the qualitative characterization and the severity grading approach by showing mild but significant increases in epithelial cell height (follicular cell hypertrophy), total glandular volume and maximum cross section area (thyroid gland hypertrophy). These quantitative data also confirmed the absence of a clear concentration-response relationship between IOP treatment levels and histological alterations.

Table 39. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to IOP in lab 1

IOP	glandular hypertrophy		glandular atrophy		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	4/20	0.25±0.55 (2)	2/20	0.10±0.31 (1)	6/20	0.30±0.47 (1)	2/20	0.10±0.31 (1)	57-60
750	9/20	0.45±0.51 (1)	2/20	0.10±0.31 (1)	6/20	0.35±0.59 (2)	3/20	0.15±0.37 (1)	57-61
1500	9/20	0.70±0.86 (2)	1/20	0.05±0.22 (1)	8/20	0.45±0.60 (2)	1/20	0.05±0.22 (1)	56-61
3000	9/20	0.50±0.61 (2)	0/20	0.00±0.00 (0)	10/20	0.55±0.60 (2)	5/20	0.30±0.57 (2)	57-62
6000	13/20	0.80±0.70 (2)	0/20	0.00±0.00 (0)	9/20	0.50±0.61 (2)	6/20	0.30±0.47 (1)	56-60

Note: [a] number of animals affected / number of animals examined
[b] values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses
[c] values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology

Table 40. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to IOP in lab 2

IOP	glandular hypertrophy		glandular atrophy		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	0/20	0.00±0.00 (0)	1/20	0.05±0.22 (1)	0/20	0.00±0.00 (0)	0/20	0.00±0.00 (0)	57-60
750	9/20	0.45±0.51 (1)	0/20	0.00±0.00 (0)	12/20	0.60±0.50 (1)	3/20	0.15±0.37 (1)	58-61
1500	8/20	0.40±0.50 (1)	0/20	0.00±0.00 (0)	14/20	0.70±0.47 (1)	7/20	0.35±0.49 (1)	57-59
3000	10/20	0.50±0.51 (1)	1/20	0.05±0.22 (1)	15/20	0.75±0.44 (1)	7/20	0.35±0.49 (1)	58-62
6000	3/20	0.15±0.37 (1)	0/20	0.00±0.00 (0)	7/20	0.35±0.49 (1)	6/20	0.30±0.47 (1)	58-59

Note: [a] number of animals affected / number of animals examined
[b] values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses
[c] values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology

Table 41. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to IOP in lab 3. Note that a 5 grade scoring scheme was used in this study that differed from the 3-grade scheme used by the other labs

IOP	colloid depletion		glandular atrophy		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	0/10	0.00±0.00 (0)	0/10	0.00±0.00 (0)	1/10	0.10±0.32 (1)	7/10	0.70±0.48 (1)	64-65
750	1/10	0.20±0.63 (2)	2/10	0.20±0.42 (1)	6/10	1.40±1.35 (3)	5/10	0.80±0.92 (2)	57-59
1500	2/10	0.20±0.42 (1)	8/10	1.80±1.40 (4)	5/10	0.80±1.03 (3)	4/10	0.40±0.52 (1)	56-57
3000	1/10	0.10±0.32 (1)	9/10	2.20±0.92 (3)	4/10	0.50±0.71 (2)	8/10	1.00±0.67 (2)	56-58
6000	4/10	0.70±0.95 (2)	9/10	3.30±1.34 (4)	0/10	0.00±0.00 (0)	4/10	0.40±0.52 (1)	55-57

Note: [a] number of animals affected / number of animals examined
[b] values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses
[c] values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology

Table 42. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to IOP in lab 6

IOP	glandular hypertrophy		colloid depletion		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(µg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	n.d.	n.d.	0/10	0.00±0.00 (0)	5/10	0.50±0.53 (1)	0/10	0.00±0.00 (0)	62-64
750	n.d.	n.d.	0/10	0.00±0.00 (0)	7/10	0.90±0.74 (2)	0/10	0.00±0.00 (0)	61-63
1500	n.d.	n.d.	2/10	0.20±0.42 (1)	8/10	1.10±0.74 (2)	0/10	0.00±0.00 (0)	61-63
3000	n.d.	n.d.	2/10	0.20±0.42 (1)	10/10	1.60±0.52 (2)	5/10	0.50±0.53 (1)	61-63
6000	n.d.	n.d.	3/10	0.30±0.48 (1)	8/10	1.40±0.97 (3)	2/10	0.30±0.67 (2)	61-63

Note: [a] number of animals affected / number of animals examined
[b] values represent the mean ± standard deviation of grade numbers for all animals examined, the highest severity grade observed is given in parentheses
[c] values represent the range of developmental stages (Nieuwkoop and Faber, 1994) of animals examined for thyroid histology
n.d. no data available

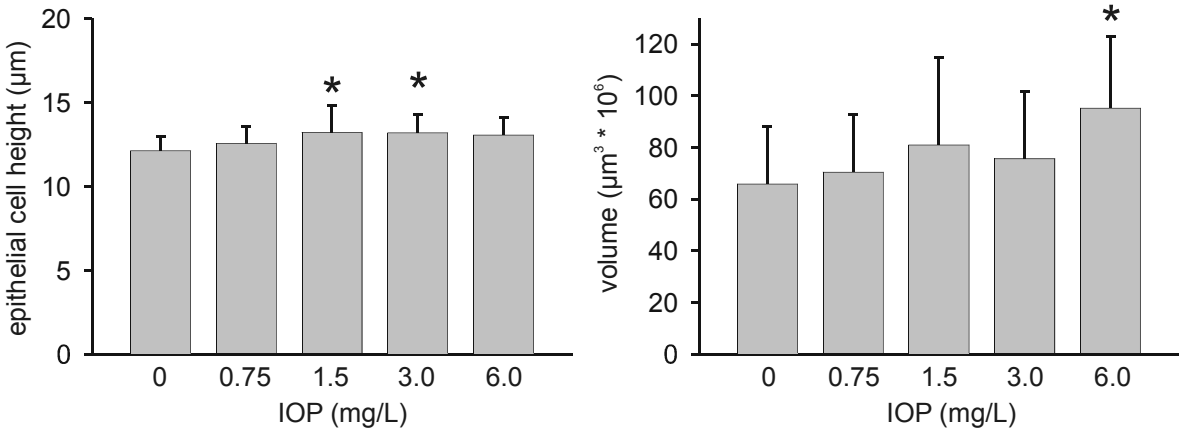

Figure 11. Morphometric analysis of thyroid tissue for changes in epithelial cell height and glandular volume following exposure to iopanoic acid (IOP) in the lab 1 study

Columns with vertical bars represent means with standard deviations (n= 20 tadpoles per treatment). Asterisks denote significant differences compared to the control treatment group ($p<0.05$ Dunnett's test).

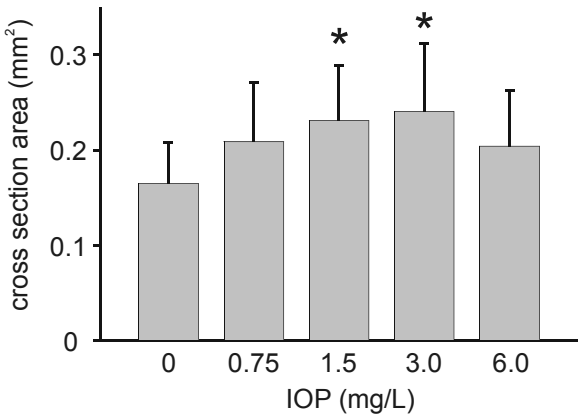

Figure 12. Morphometric analysis of thyroid tissue for changes in maximum cross section area following exposure to iopanoic acid (IOP) in the lab 2 study

Columns with vertical bars represent means with standard deviations (n= 20 tadpoles per treatment). Asterisks denote significant differences compared to the control treatment group ($p<0.05$ Dunnett's test).

DISCUSSION

Purpose of the Assay

78. The primary purpose of the Amphibian Metamorphosis Assay is to serve as a screening assay for the detection of thyroid system-disrupting activities of a test substance. It is therefore currently regarded as a potential testing tool within the tier 1 testing battery of the OECD program to detect endocrine-disrupting chemicals. The proposed experimental design and the study protocol for the conduct of the assay reflect the primary purpose as a screening assay in that the assay is not optimized towards the determination of NOEC or LOEC values or the application of regression approaches to calculate EC_x values. The diagnostic endpoints (e.g., developmental stage and thyroid histology) considered in the proposed study protocol were selected to specifically detect exposure-related changes in thyroid system function and thyroid hormone (TH) action. In order to increase the diagnostic power of the assay, additional endpoint measurements are aiming at the distinction between thyroid system-specific effects and unspecific developmental effects resulting from systemic toxicity.

79. Detection of test substance effects on thyroid system function and TH action is accomplished by the detection of alterations in the progress of TH-dependent metamorphic development and concurrent evaluation of thyroid gland histology. Monitoring for treatment-related changes in metamorphic development is accomplished through determination of the developmental stage of test animals after defined exposure periods (day 7 and day 21 in the Phase-2 protocol) accompanied by morphometric analysis of hind limb morphogenesis (day 7 and day 21 in the Phase-2 protocol). Evaluation of changes in thyroid gland histology in response to test substance treatment is performed with tissue samples obtained after the end of the 21-day exposure period. Previous experience from amphibian studies indicates that this endpoint adds not only to the diagnostic value of the assay but also provides one of the most sensitive endpoints for detection of agents that reduce the pool of TH available to peripheral tissues.

The Biological Model

80. The biological model utilized in the Amphibian metamorphosis assay is the TH-dependent metamorphic development of *X. laevis* tadpoles. Central to the understanding of this developmental process is the notion that endocrine systems, such as the thyroid system, undergo marked changes with regard to various qualitative and quantitative aspects of the regulatory pathways involved. The assay is initiated at late premetamorphic stage 51, shortly after the maturation of thyroid gland as a TH-synthesizing and TH-secreting endocrine organ. Stage 51 animals are characterized by the presence of a thyroid gland that has established follicular structures containing very little colloid which is made up by iodinated thyroglobulin, the storage form of TH in the thyroid gland. At this developmental stage, very little TH is secreted but the actual amounts are not known as no study could hitherto quantify the low amounts of TH circulating in premetamorphic tadpoles.

81. Within the first seven days of the exposure period, the control animals will develop to early prometamorphic stages 54/55. An increase in body size and weight is the most prominent change during this developmental phase while morphological changes are limited to early morphogenesis and growth of the hind limbs and initiation of fore limb morphogenesis (Nieuwkoop and Faber, 1994). Hind limb development and growth occur in response to an initial rise of circulating TH and the central role of TH in controlling early hind limb morphogenesis has been demonstrated in a great number of studies (Brown *et al.*, 2005; Buckbinder and Brown, 1992; Shi, 1999; Tata, 1998). In addition, the pituitary-thyroid feedback signalling is functional from stages 51/52 and pituitary TSH is important for thyroid gland growth and activity already during late premetamorphic stages. It has been known for long that the very

low concentrations of endogenous TH present during premetamorphic development allow for the very sensitive responsiveness of morphological and molecular endpoints to exogenous TH or TH-mimicking agents. In this regard, it is important to note that to date, no other factor than TH is able to cause precocious induction of metamorphosis-associated morphological changes in premetamorphic tadpoles.

82. During the prometamorphic phase of development, circulating concentrations of TH increase gradually due to an increased activity of the pituitary which secretes increasing amount of TSH to stimulate thyroid gland activity (Denver *et al*, 1996). Both the pituitary and the thyroid gland increase in size during prometamorphosis reflecting their increased capacity to produce TSH and TH, respectively. In response to increased TSH stimulation, follicular cells of the thyroid gland show proliferation and enhanced functional activity. The latter includes increase uptake and organification of iodide to produce TH precursors but also increased release of TH into the blood stream (Regard, 1978). Although the activity of the hypothalamus-pituitary-thyroid axis to control for increasing levels of circulating TH shows marked quantitative changes from late premetamorphosis to late prometamorphosis, the quality of this regulating circuit appears to be very much unchanged during this developmental window. In *X. laevis* tadpoles, this situation changes during the climax period of development. Results from various studies indicate that the thyrotrophic cell population of climax tadpoles starts to respond with diminished TSH synthesis to further increments in circulating TH levels so that an effective negative feedback signalling of TH on TSH-producing cells will be established at the late climax stages. Circulating TH concentrations increase dramatically within a short time period to peak values around mid-climax stages. Thereafter, TH levels decline and both the pituitary and the thyroid gland resume a state of low activity at the end of metamorphosis (Regard, 1978).

83. The developmental alterations of the different endpoints bear important implications for the design of experiments with thyroid system disrupters and the interpretation of the corresponding endpoint measurements. A problem inherent to the proposed endpoint measurements is that observations made after a fixed exposure period are potentially be confounded by the fact that organisms sampled in different treatment groups are at different developmental stages. In particular, the qualitative change in hypothalamus-pituitary-thyroid axis function during climax stages need to be considered as a confounding factor when interpreting the outcome of the assay. These aspects will be repeatedly discussed along the following assessment of test performance for the different compounds used in Phase-2 studies.

Control Animal Performance

84. In the context of the Amphibian metamorphosis assay, it is of particular importance to ensure homogenous rates of tadpole growth and development in the control treatment group. A homogenous control group is required to sensitively detect exposure-related alterations in development and growth because the developmental changes in body size, weight and morphology are the hallmarks to assess a chemical's ability to affect TH-dependent metamorphosis. Furthermore, the histological appearance of the thyroid gland can vary depending on the developmental stage and thus, evaluation of treatment-related changes in thyroid histology is expected to yield more consistent results when a homogenous set of reference tissue samples from the control treatment are available.

85. In addition, it is expected that a high proportion of climax stage animals within the control treatment group will confound not only the morphological assessments due to marked and rapid reductions in tadpole size and weight but also the interpretation of histological hallmarks of thyroid tissue sections. The latter problem is related to the marked qualitative changes of hypothalamic-pituitary-thyroid axis regulation that occur during climax including a reduction in thyroid gland stimulation by TSH towards the end of metamorphosis.

86. Given these considerations, a critical assessment of control animal performance is an important aspect in the evaluation of the study protocol used in Phase-2 studies. For 10 of the 14 experimental studies performed during the Phase-2 validation exercise, the control animal population showed the desired characteristics. In 4 experiments, the control population showed an increased variability with regard to developmental stage on day 21 as well as an increased proportion of animals in mid-climax stages. The four experiments that had a less homogenous development in the control group were performed in two labs that applied much lower feeding rates compared to the other labs. The important role of daily feeding rates for optimal development of *X. laevis* tadpoles is substantiated by two findings. First, all labs that used the recommended feeding protocol and increased the amount of food along with tadpole growth reported a relatively low inter-individual variability in developmental rates and a low proportion of control animals with accelerated development to climax stages. Secondly, a review of historical records of tadpole growth and development in lab 1, which used SeraMicron for several years, showed that feeding rates lower than the recommended amounts are associated with a greater variability in development.

87. Based on the results of the Phase-2 studies, it appears that the recommended feeding schedule with SeraMicron allows to obtain homogenous development in the controls but also minimizes the number of animals displaying accelerated development to advanced climax stages. Despite the deviations observed in some experiments, a comparison of animal performance in experiments according to the current study protocol and a large number of published studies on tadpole development in *X. laevis* showed that the various protocol modifications during the various validation exercises resulted in a superior quality of test animals for the purpose of the assay. Nevertheless, given the critical role of the feeding regime, a more strictly standardized feeding schedule should be implemented in the study protocol to further harmonize growth and developmental rates across different experiments.

Effects Pattern of PER in Phase-2 Studies

88. Sodium perchlorate was used in the Phase-2 studies as a model substance to assess the utility of the study protocol to detect inhibitory effects on thyroid iodide uptake. Since TH are iodinated thyronine derivatives, active concentration of iodide from the blood stream represents a crucial step in their biosynthesis. A remarkable capability of the thyroid gland is to concentrate iodide by a factor of 20 to 40 over blood plasma iodide concentrations (Dohan et al. 2003). This active uptake of iodide by thyroid follicular cell is an energy-dependent process that is mediated by the sodium-iodide symporter (NIS), a glycoprotein located in the basolateral membrane of thyroid follicular cells (Dohan et al. 2003). The driving force for NIS-mediated iodide transport is the inwardly directed Na+ gradient generated by Na+ K+ ATPase (Paire et al. 1998). The pivotal role of NIS in mediating thyroidal iodide uptake and thus facilitating TH synthesis is highlighted by the finding that competitive inhibition of NIS-mediated iodide uptake blocks not only thyroidal iodide uptake but impairs TH synthesis (Tonacchera et al. 2004, Wolff 1998). One of the most potent competitors of NIS-mediated iodide uptake is the PER anion with a 30-fold greater affinity towards NIS than iodide (Tonacchera et al., 2004). Similar to mammals, early studies in amphibians showed that treatment with perchlorate reduced thyroidal iodide uptake (Lynn & Dent 1960, Norris & Platt 1973). The vulnerability of the amphibian thyroid system to perchlorate has also been demonstrated in recent studies showing that long-term treatment of tadpoles with perchlorate inhibits TH-dependent development in *X. laevis* tadpoles concomitantly with the induction of classical histological lesions in the thyroid gland including colloid depletion and follicular cell hypertrophy and hyperplasia (Goleman *et al.* 2002, Tietge *et al.* 2005).

89. A total of five exposure experiments with PER were conducted during this Phase-2 exercise. Effects of PER treatment on metamorphic development were variable across the different studies and only two experiments reported statistically significant developmental retardation at the higher test concentrations. Given that iodide and PER compete at the level of the NIS protein, differences in iodide content of the dilution water were taken into consideration as a possible cofactor accounting for the

differences in the magnitude of PER effects on morphological development. Total iodide content in dilution water was determined in only three laboratories (Table 43). From the data available, it appears that even slight differences in iodide content may alter the effectiveness of PER to cause developmental delay. This contention is supported by a comparative analysis of the severity of PER effects as detected in experiments performed in lab 1 and lab 4. The test animals used for the exposure studies in both labs are all obtained from in-house breeding in lab 1, and the food applied in both studies was from the same lot. The severity of PER effects on development and on thyroid histology was increased in the experiments in lab 4 and a concurrent analysis of total iodide content of dilution water showed that iodide content was 7.5-fold lower in lab 4.

Table 43. Summary of total iodide concentrations in dilution water, developmental effects of PER and histological evaluation of thyroid tissue after PER treatment

	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6[d]
total iodide[a]	10 µg/l	n.d.	20 µg/l	1.5 µg/l	n.d.	0.75 µg/l
development[b]	>500 µg/l	>500 µg/l	125 µg/l	500 µg/l	>500 µg/l	250 µg/L
histology[c]	62.5 µg/l	62.5 µg/l	62.5 µg/l	62.5 µg/l	n.d.	16 µg/l

Note: [a] total iodide content in dilution water
[b] effective PER concentration leading to developmental delay
[c] effective PER concentration leading alterations in thyroid histology
[d] lab 6 did not perform an exposure study with PER during the validation Phase-2 and data are from a previous study in this lab on the effects of PER on *X. laevis* tadpoles
n.d. no data available

90. Despite inter-laboratory differences in PER effects on developmental parameters, all experiments could detect remarkable effects of PER treatment on thyroid histology including depletion of colloid stores, glandular hypertrophy as well as hypertrophic and hyperplastic changes in the follicular epithelium. Consistent across laboratories, effects were observed already at the lowest tested concentration of 62.5 µg/l PER and incidence and severity of these responses increased in clear concentration-dependent manner. Thus, compared to the assessment of morphological endpoints, evaluation of thyroid gland histology provided the most sensitive and consistent endpoint responses across the individual experiments.

91. In summary, the study protocol was capable of detecting the inhibitory effects of PER on the amphibian thyroid gland. Evaluation of thyroid gland histology was far more sensitive than morphological assessment of test animals in detecting the anti-thyroidal activity of PER. In experiments with PER as test substance, the study protocol yielded in highly reproducible test results.

Effects Pattern of T4 in Phase-2 Studies

92. Thyroxine (T4) was used in the Phase-2 studies as a model substance to assess the utility of the study protocol to detect agonistic effects on the thyroid system. T4 is the native prohormone synthetized by the thyroid gland of all vertebrates. The current view holds that T4 requires transformation to T3 by monodeiodinase enzymes to elicit most of the biological effects associated with TH. Acceleration of amphibian metamorphosis by T4 is a well-documented phenomenon (Degitz et al., 2005; Opitz et al., 2005). Previous studies with *X. laevis* tadpoles showed that low T4 concentrations stimulate metamorphosis without disrupting the normal sequence of morphological changes or causing overt toxicity.

93. During the Phase-2 exercise, a total of four exposure experiments were conducted with T4 as test substance. Of these four experiments, only three experiments did fulfill all validity criteria discussed in section 0. For the T4-experiment in lab 3, a major concern with regard to the validity of the experiment is the very high inter-individual variability in control animal development. The observation of a wide range

of developmental stages on day 7 (range of 4 different stages) and on day 21 (range of 12 different stages) suggest highly heterogenous rates of development in the control animal population.

94. All exposure experiments performed with T4 could clearly identify accelerating effects of T4 on metamorphic development in *X. laevis* tadpoles based on determination of developmental stages of test organisms and by morphometric analysis of hind limb length (HLL). The sensitivity of these two endpoints to reveal an acceleration of development was, however, different for day 7 and day 21 measurement. HLL measurements were more sensitive than stage determination on study day 7 whereas the two endpoints displayed a similar sensitivity to detect accelerated development on day 21. These results are in agreement with data from the Phase-1 validation study where a similar effect pattern was observed for T4.

95. 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. The reported increased sensitivity of HLL on day 7 is most likely related to the increased sensitivity of hind limb tissue to the still very low endogenous concentrations of TH during early prometamorphosis. It is anticipated that during the initial 7 days of exposure, the exogenous T4 treatment results in a relatively higher contribution to the circulating T4 levels of test animals as compared to later stage tadpoles in which endogenous T4 production is expected to play a predominant role in the provision of circulating TH.

96. Histopathological examination of thyroid gland sections revealed less marked changes following T4 treatment compared to the PER studies. Generally, mild alterations in glandular size including both atrophy and hypertrophy were observed at different treatment levels but no concentration-response relationships could be established. Similarly, mild follicular hypertrophy was observed at high T4 concentrations in lab 1 but this effect was not reproduced in experiments by other labs. The histological appearance of thyroid tissue was very variable within each of the various T4 treatment groups and there was a great overlap with the control group.

97. From a mechanistic point of view, the presence of glandular atrophy in T4-treated tadpoles is likely related to an increased negative feedback on the pituitary's activity to synthesize and release TSH leading to a reduced stimulation of the thyroid gland by TSH. On the other hand, histological assessment also indicated an increased incidence of mild glandular hypertrophy at the two highest T4 concentrations in some experiments. These mildly hypertrophic glands showed an increased number of follicles lined by columnar epithelial cell (hypertrophic cells). However, the diagnostic value of this finding is questionable because a major confounding factor in the T4 studies was that many tadpoles treated with the higher T4 concentrations were at more advanced stages compared to the control group due to accelerated development. It was not possible to discern whether specific differences in the histological appearance of thyroid tissue were due to exposure-related alterations in thyroid function or more simply a reflection of increased thyroid activity during climax stages. It has been observed in many studies on the development of the thyroid gland in anuran tadpoles that during normal tadpole development, thyroid glands become slightly hypertrophic at mid-climax stages due to a maximum stimulation of the thyroids by peaking TSH levels.

98. In summary, the study protocol was capable of detecting the stimulatory effects of T4 on metamorphic development. Assessment of tadpole morphology, particularly the morphometric analyses of hind limb growth on day 7 and developmental stage on day 21, was more sensitive and consistent with the anticipated mode of T4 action than histological evaluation of thyroid glands at test termination.

Effects Pattern of IOP in Phase-2 Studies

99. Iopanoic acid (IOP) was used in the Phase-2 studies as a model substance to assess the utility of the study protocol to detect modulators of extrathyroidal TH metabolism, specifically inhibitors of

iodothyronine deiodinases. The current view holds that T4 is the major hormone secreted by the thyroid gland and thyroidal secretion of T3 seems to be low in *X. laevis* tadpoles (Rosenkilde and Ussing, 1990). In turn, most of the T3 is generated by extrathyroidal conversion of T4 to T3 by iodothyronine deiodinases. Three types of iodothyronine deiodinases (D1, D2, D3) have been identified in vertebrates which differ in tissue distribution, substrate specificity and sensitivity to inhibiting compounds (Bianco *et al.*, 2002; Köhrle, 1996). D1 and D2 catalyze primarily the removal of one iodide from the outer tyrosine ring of T4 to produce T3, and 5'-deiodination can thus be regarded as an activating pathway of T4 metabolism. D3 catalyzes the cleaving of one iodide from the inner tyrosine rings of T4 and T3 generating inactive iodothyronine derivatives such as reverse T3 and diiodothyronines, respectively. In amphibian tadpoles, the coordinated progression of metamorphic development requires a high degree of local control of T3 production which apparently dominates over central supply of T3, at least during metamorphosis (Brown, 2005; Kühn *et al.*, 1993). Only recently, a putative *X. laevis* homologue of mammalian D1 has been identified (Kuiper *et al.*, 2006) but neither the expression profile nor the putative regulatory role of D1 during metamorphic development have been characterized so far. However, several studies have investigated the role of D2 and D3 in controlling TH action during metamorphosis of anuran tadpoles (Becker *et al.*, 1997; Cai and Brown, 2004; Galton and Hiebert, 1988; Huang *et al.*, 1999, 2001; Kawahara *et al.*, 1999). The data derived from these studies support the view that both D2 and D3 play a central role in modulating the tissue responsiveness to TH by either increasing intracellular concentrations of biologically active T3 (e.g., D2 in hind limbs) or by preventing TH action via rapid inactivation of T4 and T3 (e.g., D3 in tadpole tail). In mammals, IOP has been shown to inhibit the activity of all three iodothyronine deiodinases (Braga and Cooper, 2001) and previous studies showed that IOP has similar inhibitory activity towards amphibian deiodinases (Becker et al., 1997; Huang et al., 2001; Cai and Brown, 2004; Shintani et al., 2002).

100. During the Phase-2 exercise, a total of five exposure experiments were conducted with IOP as test substance. Of these five experiments, only four experiments did fulfill all validity criteria discussed in section 0. For the IOP-experiment in lab 3, a major concern with regard to the validity of the experiment is the very high inter-individual variability in control animal development. The observation of a wide range of developmental stages on day 21 (range of 9 different stages) suggest highly heterogenous rates of development in the control animal population. Another particular concern about this IOP experiment is related to the marked mismatch of developmental stages in the control group relative to the various IOP traetments. A high proportion of control animals were at late climax stages on day 21 while IOP-treated tadpoles showed marked retardation of growth and development.

101. Analysis of metamorphosis-associated morphological parameters showed consistently that IOP treatment caused asynchronous morphological development of tadpoles. This effect was observed in all five exposure studies. In the case of IOP, asynchronous morphological development was characterized by retarded morphogenesis of hind limbs relative to head and tail structures. It was therefore not possible to assign a specific development stage to many of the IOP-treated animals and an analysis of changes in median stages due to IOP treatment was not considered appropriate. However, the increased incidence of tadpoles showing asynchronous morphological development of tissues that are known to rely on TH action for their remodeling was considered itself as a diagnostic finding for disruption of peripheral TH action. Given that IOP is known to inhibit both 5'- and 5-deiodination of iodothyronines, the abnormal morphological development of IOP-treated animals is interpreted to reflect the presence of tissue-dependent local hypo- and hyperthyroidism. Retardation of hind limb morphogenesis by IOP treatment was very likely due to inhibition of D2 activity resulting in deficits in the conversion of T4 to T3 (Becker et al., 1997; Brown, 2005; Cai and Brown, 2004). Various studies could demonstrate the importance of local D2 activity in hind limb tissue to provide T3 for stimulation of the multiple developmental programs involved in hind limb growth and differentiation (Becker et al., 1997; Brown et al., 2005; Cai and Brown, 2004). Previous studies have shown, for example, that short-term treatment of tadpoles with IOP concentrations in the range of 6 mg/l, as used in the highest IOP treatment during Phase-2 experiments,

were capable of almost completely blocking the stimulation of tadpole tissue remodeling by exogenous T4. The results from the Phase-2 validation studies did not reveal such a severe inhibition of tissue remodeling but rather a mild to moderate retardation.

102. Still other morphological alterations observed during the Phase-2 studies suggest tissue-specific hyperthyroidism following IOP treatment. Craniofacial landmarks of metamorphic development were developmentally advanced in many IOP-treated animals compared to the controls and in some tadpoles, IOP treatment also resulted in accelerated resorption of tail fins. Local hyperthyroidism following IOP treatment was likely the result of inhibitory effects on D3 activity resulting in impaired inactivation of T3 and thus a stronger T3 stimulation of tissue remodeling. Overall, the various morphological findings of the IOP studies confirmed the proposed sensitivity of the amphibian metamorphosis model to sensitively detect alterations in peripheral TH action.

103. Despite the marked effect of IOP on morphological development, histological examination of thyroid tissue revealed only mild alterations. Exposure-related effects included mild increases in thyroid gland size and a mild to moderate hypertrophy of the follicular epithelium. Such alterations are considered to indicate a mild increase in thyroid tissue stimulation by TSH due to IOP treatment. It is currently not known how local deficits or excess of T3 may change the plasma concentrations of T3 and how much T3 is available to reduce negative feedback at the hypothalamus-pituitary level in IOP-treated tadpoles. In this regard, the histological results suggest only mild alterations of the negative feedback signalling between neuroendocrine control centers and the thyroid.

104. Furthermore, the histological appearance of the thyroid glands from IOP-treated tadpoles was somewhat unique because follicular cell hypertrophy was not closely associated with colloid depletion. In other studies using compounds that caused a mild stimulation of the thyroid gland, colloid depletion has been shown to be one of the most sensitive parameters. This deviation from the typical histology of mildly stimulated thyroid tissue may be, in part, due to the high iodide content of iopanoic acid. Given a total iodide content of 4 mg/l iodide in the highest IOP concentration (6 mg/l), abiotic and biotic metabolism of the parent compound can be expected to cause the release of large amounts of inorganic iodine. Distinct histological alterations have been reported in several rodent studies with iodine-rich substances such as IOP, ipodate or amiodarone and it has been suggested that some effects seen in thyroid histology are independent of the parent compound but are related to the massive iodine load associated with IOP treatment (Braga and Cooper, 2001; Kanno et al., 1992, 1995; Markou et al., 2001).

105. The results from histological examination of IOP-treated animals in lab 3 differed drastically from observations made in the other labs. In lab 3, thyroid tissue of tadpoles had an atrophic appearance because the glands were smaller than in the controls and contained many small follicles. However, there was a marked difference in the developmental stage of the tadpoles that were used for tissue sampling between the control group and the IOP treatments. The fact that all control animals were at very late climax stages alone already precludes a meaningful interpretation of histological results. Given that the thyroid tissue samples for the various IOP treatments were obtained from tadpoles at stages 55 to 59, a sound assessment for exposure-related alterations by histopathology was not possible.

Performance of Endpoints in Phase-2 Studies

106. This section discusses the performance of the endpoints used in Phase-2 validation studies to provide information relevant to the detection of thyroid system-disrupting activity of the test substances in the amphibian metamorphosis assay. Although being primarily regulated by TH, amphibian metamorphosis is a complex process and as such prone to perturbations originating from different modes of action, many of which may not be related to altered thyroid system function. To this point, the various endpoints included in the study protocol served to obtain information about exposure-related alterations in

(1) tadpole growth, (2) tadpole metamorphic development and (3) pituitary-thyroid axis function. While the examination of morphological landmarks associated with metamorphic development and the assessment of thyroid histopathology are important to reveal altered thyroid system function, the concurrent analysis of growth-related parameters is required to distinguish thyroid system-related activities from systemic toxicity of the test compound. Given the complexity of the biological model utilized, it has to be stressed that the interpretation of the assay results will always require a weight of evidence analysis integrating the response profile observed for all.

Body Length and Body Weight

107. Several apical morphological endpoints including whole-body length (WBL), snout-to-vent length (SVL) and body weight were used to assess possible effects of the test substances on tadpole growth. These endpoints were relevant and useful to control for the presence of possible systemic toxicity caused by the test compounds.

108. In this regard, measurements of these endpoints on day 7 were particularly relevant. During the first seven days of the exposure phase, tadpoles show substantial growth and the test animals double their body size and increase their body weight even more. Generally, the absence of significant effects of the three test compounds on growth-related parameters by day 7 provides an important argument for the absence of strong non-thyroidal toxicity.

109. The results from the PER studies confirmed previous studies showing that inhibitory effects on the thyroid system do not lead to growth retardation. In fact, anti-thyroidal substances may even cause mild increases in body size and body weight and, therefore, increases in tadpole size and weight do not indicate non-thyroidal toxicity. However, even slight reduction in tadpole growth need to be weighed against information about concurrent developmental effects. For example, the mild growth retardation observed on day 7 for the highest T4 concentrations in one T4 study, is in line with previous studies demonstrating that acceleration of development by TH agonists could be associated with slight reductions in tadpole growth rates. Generally, the stronger the acceleration of metamorphosis by TH agonists the more likely are reduced growth rates already at early developmental stages. Thus, a sound interpretation of growth-related parameters on day 7 still requires information about concurrent changes in developmental rates. On the other hand, the marked reduction in tadpole growth by day 7 in one experiment with IOP was a strong indication for non-thyroidal toxicity. Here, IOP treatment was not associated with an acceleration of development and therefore, the observation that body weight of IOP-treated tadpoles was less than 40% of the controls by day 7 should be regarded as evidence for an overt toxic effect.

110. Interpretation of day 21 measurements of growth-related endpoints is more complex than on day 7. During normal development, tadpole growth ceases at late prometamorphic/early climax stages and maximum body size and weight is normally observed in *X. laevis* tadpoles at stages 59/60. Thereafter, body size and weight decrease due to tissue resorption and a marked decrease in water content of tadpole tissues. Analysis for exposure-related effects on tadpole growth by comparing treatment means for WBL, SVL and wet weight on day 21 will thus be compromized by the development-related reduction of all these parameters in tadpoles entering climax stages.

111. Thus, robust assessment of growth effects using day 21 measurements of WBL, SVL and body weight is only possible if tadpoles have not yet reached stage 61. However, in the majority of experiments performed during Phase-2, a small proportion of test animals, even in the control group, showed development to stage 61. Accordingly, the problem of a valid data interpretation towards detection of growth effects using data from day 21 measurements requires further discussion at the Expert's Group and a statistical consultation to explore possible approaches of how to deal with these data sets.

Hind Limb Length

112. As outlined before, hind limb morphogenesis is regulated by TH (Brown et al., 2005) and results from validation Phase-1 studies indicated that HLL measurements, particularly when performed on day 7, could increase the sensitivity of the assay towards agonistic activities. This was confirmed in three of the T4 experiments where increases in HLL on day 7 provided the most sensitive endpoint response to detect the agonistic activity of T4. From a mechanistic view, the detection of increased hind limb growth on day 7 provides a diagnostic finding because so far only TH and synthetic agonists (e.g., GC-1) have been shown to cause a precocious initiation and acceleration of hind limb morphogenesis in premetamorphic tadpoles. The sensitivity of HLL measurements on day 7 to detect the effects of IOP and PER was generally low for the concentrations used in Phase-2 studies. However, experiments during Phase-1 using more portent concentrations of TH synthsis inhibitors such as PTU have shown that reduced HLL could already be detected following 7 days of exposure.

113. Determination of HLL on day 21 was useful to detect acceleration of development by T4 in all four T4 studies and to detect inhibition of metamorphosis by PER in two of five studies. In the two PER studies, detection of reduced HLL at the higher PER concentrations was consistent with the observation of retarded development by concurrent stage determination. Notably, reduced HLL was not associated with reduced tadpole size in the PER experiments. Thus, the selective retardation in morphogenesis of a TH-sensitive tissue such as the hind limbs added evidence for the presence of anti-thyroidal activities by PER.

114. Among the compounds studied in Phase-2, IOP had the most marked effect on hind limb growth. Mean HLL on day 21 was reduced at all IOP concentrations in four of five tests. This finding is consistent with the plausible inhibition of D2 activity in hind limbs by IOP and the reduced generation of local T3 (Brown et al., 2005; Cai and Brown, 2004). However, IOP treatment also caused a slight reduction in WBL at all tested concentrations raising the question whether the reduced HLL might merely reflect the reduction in overall body size. Thus, these observations in the IOP experiments point to the need to identify approaches for normalization of HLL measurements with regard to the overall body size of the test animals. Calculation of ratios for HLL/WBL or HLL/SVL could provide a promising means to achive such a normalization of HLL. The substantial data sets from the Phase-2 studies should be used for model calculations and subsequent statistical re-analyses of the Phase-2 study results.

Developmental Stage

115. Determination of the developmental stage of test animals provides an integrative measure of test substance effects on metamorphosis. It should also be stressed that stage information is an absolute prerequisite for a sound interpretation of all other endpoint responses. Given this, developmental stage determination is a highly relevant core endpoint of the assay.

116. Determination of developmental stage on day 21 was a sensitive endpoint to detect acceleration of metamorphic development in all T4 experiments. Notably, together with morphometric analyses of HLL, stage determination was a more sensitive and diagnostic endpoint to reveal the agonistic activity of T4 than histopathological examination of the thyroid. These results confirm the proposed utility of the amphibian model to sensitively detect agonistic activities by means of monitoring developmental rates of tadpoles. Morphological examination of T4-treated animals further showed that T4 treatment did not disrupt the sequential process of metamorphic development.

117. Developmental stage was a less sensitive endpoint to detect the antithyroidal activity of PER in the tested concentration range. Only two of five experiments revealed significant delay in metamorphic development due to treatment with the highest PER concentrations. In the light of the marked histological changes of thyroid tissue at all tested PER concentrations, these results support the view that alterations in

development in response to antithyroidal agents might only occur when the capacity of the hypothalamus-pituitary-thyroid axis to compensate deficits in thyroid gland function becomes exhausted.

118. Stage determination in the IOP experiments revealed an interesting aspect of this endpoint measurement, namely the detection of disruption of the sequential remodeling of tadpole tissues eventually leading to ansynchronous morphological development. This morphological effect by IOP treatment is consistent with the expected inhibitory activity of IOP on iodothyronine deiodinases in peripheral tissues. Although the developmental abnormalities prevented a precise determination of the developmental stage of test animals, the presence of asynchronous developmental may be considered a diagnostic finding by itself. The latter contention is supported by the finding that the asynchronous development was characterized by developmental heterochrony in remodeling of distinct tissues but not by malformation. However, these results also point to the need to identify criteria and examination approaches to efficiently identify and describe such morphological changes.

Thyroid Histopathology

119. Histopathology of thyroid tissue was relevant and reliable for the detection of anti-thyroidal activities. Thyroid histopathology was clearly the most sensitive endpoint to detect the effects of PER on the thyroid gland. Similar observations were made in Phase-1 experiments with the TH synthesis inhibitor PTU. During Phase-2, consistent histological findings were observed in all PER studies demonstrating the robustness of thyroid histopathology to identify inhibitors of thyroid function in the amphibian metamorphosis model. The demonstration of mild increases in thyroid gland size and mild to moderate follicular hypertrophy following exposure to IOP support the value of thyroid histopathology to detect substances with extrathyroidal modes of action. It should be noted that the improved guidance on histopathological practices and the standardized protocols to examine thyroid tissue sections for exposure-related alterations enhanced the robustness and sensitivity to detect thyroidal activity in the amphibian metamorphosis assay.

120. Histopathology of thyroid tissue was not relevant for the detection of agonistic activities of T4. Notably, heterogenous responses including mild atrophy as well as mild hypertrophy were observed in different experiments at T4 concentrations that produced clear accelerating effects on metamorphic development. Further, changes in various histological parameters were often not concentration-dependent and specific findings were not reproducible across different experiments. In the case of IOP, however, thyroid histopathology was a relevant endpoint to confirm that the distinct morphological effects caused by IOP were related to changes in thyroid system function.

121. Given the central role of thyroid histopathology as a diagnostic and sensitive endpoint to detect exposure-related alterations in thyroid system function, it is important to consider possible confounding factors that affect the interpretation of this endpoint. The results of the Phase-2 studies indicate that differences in the developmental stage of sampled test organisms need to be carefully integrated in the interpretation of histopathological findings. From the current state of knowledge on histological changes of thyroid tissue during normal development and following exposure to thyroid system-disrupting compounds, the following conclusions can be drawn of how to relate developmental stage information to histological effects patterns.

122. The thyroid gland increases in size during normal development from premetamorphic stage 51 to mid-climax stages 62/63. Furthermore, increases in functional activity of the gland are reflected by distinct changes of various histological markers including an increase in the height of the follicular epithelium and a partial depletion of colloid stores. The latter is indicated by an increase in peripheral vacuolation of the colloid. The severity of changes in the histological appearance of thyroid glands during normal development can vary among laboratories depending on various factors including culture water properties.

Although the influence of the genetic background of test organisms on specific aspects of thyroid gland activity during metamorphosis has not been investigated, it has to be considered a possible factor. Regarding the influence of culture water characteristics, the iodide content as well as background concentrations of goitrogenic substances (e.g., perchlorate, chlorate, nitrate and thiocyanate) should be considered as potent factors affecting on thyroid gland function.

123. A review of results from experiments during Phase-1 and Phase-2 as well as from published studies on the histological characteristics of thyroid glands in *X. laevis* tadpoles revealed that the severity of distinct changes in thyroid histology during normal development can differ across experiments. In most studies, key histological markers such as glandular size, epithelial cell height and colloid content showed only modest differences between stages 57 and 60, particularly when compared to the distinct alterations induced by goitrogen treatment. Thus, developmental stage differences are less likely to distort the interpretation of thyroid histology if the control samples are derived from animals within this range of stages.

124. In contrast, the thyroid of mid-climax and late climax tadpoles often shows a histological profile of enhanced activity with increases in epithelial cell height and partial colloid depletion. The severity of histological changes in climax animals relative to late prometamorphic stages could differ markedly between studies. One extreme case was reported by Saxen et al. (1968), who observed collapsed follicles in control tadpoles at mid-climax stages. Still other studies found only modest changes at climax stages (Opitz et al., 2006).

125. Two examples of the Phase-2 experiments illustrate the problems that arise when tissue samples were taken from climax tadpoles. In the T4 study of lab 1, mild to moderate stimulation of thyroid tissue was observed by histology at the highest T4 concentrations but because T4-treated animals were at advanced stages relative to the controls, it was not possible to distinguish whether the mild hypertrophy was exposure-related or simply due to the advanced stage of the T4 tadpoles. In the IOP study of lab 3, all control samples were from tadpoles at stages 64-65, whereas in the IOP treatments, tadpoles at early to late prometamorphic stages were sampled for histology. The subsequent histopathology indicated thyroid atrophy as an IOP exposure-related effect but integration of the stage information suggests that stage differences are much more plausible to account for the differences in histological appearance of the thyroids.

126. Taken together, these observation do not argue against the utility of thyroid histopathology in the amphibian metamorphosis assay but underline the need that interpretation of thyroid histopathology should integrate information about the developmental stage of test animals.

Summary of Phase-1 and Phase-2 Validation Results

127. The amphibian metamorphosis assay is intended to provide information on the possible interference of a test substance with thyroid system function and TH action. The study protocol is not as a definitive test to assess the health hazards associated with thyroid system disruption, but is designed to uncover the general capacity of test substances to interfere with the thyroid system. It is designed as a tier I screening assay providing only information about a potential toxicity concern with regard to the thyroid system. The results of the experiments conducted during validation Phase-1 and Phase-2 strongly support the utility of the proposed study protocol for the purpose of detecting thyroid-disrupting substances acting via different modes of action.

128. The primary objective of the Phase I studies 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, exposure was 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 and a dilution water control group. 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 participating laboratories in conducting related work with *X. laevis*. Prior to the Phase-1 studies, little or no work had been conducted using identical protocols and chemicals among different laboratories. Therefore, another main objective of this work was to evaluate the methodologies of three different laboratories following exposure to identical compounds and to use the outcomes of these studies to:

1. propose an array of potential endpoint measurements

2. provide guidance for standardized performance of endpoint measurements and

3. guide the development of a standardized study protocol.

129. In summary, the Phase-1 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. Thyroid histopathology was demonstrated to enhance the sensitivity and diagnostic power of the assay. These results suggested 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. Based on a critical review of Phase-1 results, an improved study protocol for a 21-day assay to be used in a validation Phase-2 study was proposed.

130. Main enhancements to the study protocol for Phase-2 included the use of a flow-through exposure system, use of an optimized experimental design based on statistical consultation and power analyses, further standardization of general test conditions (e.g., flow-rate, temperature, food type) and use of standardized protocols for endpoint measurements (e.g., measurements of apical morphological endpoints, thyroid histopathology. The Phase-2 studies were conducted with three test substances representing model compounds acting via different modes of action on the thyroid system. These compounds included perchlorate (PER), thyroxine (T4) and iopanoic acid (IOP). The results of the Phase-2 studies demonstrated clearly that each of these compounds were detectable as thyroid system-disrupting agents using the proposed study protocol. The Phase-2 studies also showed a differential sensitivity and utility of specific endpoints to detect different modes of action.

131. Combining the results from Phase-1 and Phase-2 studies, the following conclusion can be drawn regarding the capacity of the amphibian metamorphosis assay to detect various modes of thyroid system disruption.

1. Agonists of TH action are most sensitively detected by demonstration of accelerated morphological development. Morphometric analyses of hind limb length on day 7 appear to provide the most sensitive endpoint to detect agonist action. Stage determination on day 21 provided a sensitive and robust endpoint to confirm developmental acceleration. Notably, weak agonistic activities are not reliably detected by thyroid histopathology of day 21 tissue samples.

2. Inhibitors of thyroid synthesis are sensitively detected by thyroid histopathology of day 21 tissue samples. This was true for inhibitors of iodide uptake (e.g., PER) and inhibitors of iodide organification (e.g., PTU). By using the improved guidance protocols for tissue processing and

histological assessment of tissue sections, concentration-response relationships could be clearly demonstrated for the core histological parameters thyroid gland hypertrophy and folliclular cell hypertrophy/hyperplasia.

3. Modulators of iodothyronine deiodinase activity (e.g., IOP) were detected by their unique effects on morphological development. Consistent with a tissue-specific role of different iodothyronine deiodinases, IOP caused asynchronous morphological development characterized by retarded hind limb development in the presence of advanced developmental remodeling of craniofacial and tail structures. Thyroid histopathology of day 21 tissue samples also a sensitive endpoint in the IOP studies but a definitive demonstration whether inhibition of iodothyronine deiodinases or the high iodide content of IOP were causative for the histological alterations awaits further studies.

132. From the effect patterns observed for the test compounds, it becomes clear that each of the proposed endpoint measurements provided for the overall utility of the test protocol to detect various modes of action. While apical morphological endpoints appear very sensitive to disturbance of peripheral TH action, chemically-induced reductions in circulating TH are most sensitively detected by distinct changes in the histological appearance of the thyroid gland. These conclusion are very much in line with the current concept of thyroid system control of amphibian metamorphosis supporting the validity of the use of this biological model for the purpose of developing a testing tool for thyroid system disruption.

133. Furthermore, the results obtained in the Phase-1 and Phase-2 studies allow for some generalization of the utility of the amphibian metamorphosis assay to detect altered thyroid system function. It can be assumed that, irrespective of the actual mode of action, exposure-related alterations in the thyroid gland's capacity to produce and secrete TH into circulation should become detectable in the amphibian bioassay through histopathology of thyroid tissue. On the other hand, the current knowldege of tissue remodeling during amphibian metamorphosis allows the prediction that exposure-related acceleration of metamorphic development is itself a strong indication for modulation of TH action and can be regarded a diagnostic finding.

Proposal of a Decision Logic

134. The following section is intended to provide logical guidance to assist in the conduct and interpretation of the results of the bioassay (see flow chart in Figure 13). In addition to the steps described below, further discussion at the Expert's Group is required regarding the identification of criteria that define a valid exposure experiment. Such criteria could include a tolerable variability in individual rates of growth and development in the control group, a tolerable deviation of growth rates in the chemical treatments from the control group and a tolerable deviation in developmental stages of individual tadpoles sampled for thyroid histopathology.

I1: Chemical information

This test is based upon an aqueous exposure protocol whereby test chemical is introduced into the test chambers via a flow through system. The purposes for using flow through methods includes: maintenance of desired test chemical concentrations, maintenance of adequate dissolved oxygen concentrations, removal of nitrogenous wastes, fecal material, and residual food, and the removal of potential test chemical metabolites originating from the test organisms. Flow through methods, however, introduce constraints on the types of chemicals that can be tested, as determined by the physicochemical properties of the compound. Therefore, prior to using this protocol, baseline information about the chemical should be obtained that is relevant to determining the testability. Common characteristics which indicate that the chemical may not be testable include: high octanol water partitioning coefficients (log Kow), high

volatility, susceptibility to hydrolysis, and susceptibility to photolysis under ambient laboratory lighting conditions. Other factors may also be relevant to determining testability and should be determined on a case by case basis.

D1: Testability

Information from I1 should be used to determine testability. If an efficacious test is not possible for the chemical, then the negative default is to T1.

T1: Do not test

Chemical is not suitable for testing using this protocol.

P1: Conduct test

If the outcome of D1 is yes, then the test can be conducted according to the protocol.

D2: Advanced development

Advanced development can be evaluated in one of two ways. First, the general developmental stage can be evaluated using the standardized approach detailed in NF. Second, specific morphological features may be quantified, such as hind limb length, whose length is positively associated with agonistic effects on the thyroid hormone receptor. If statistically significant advances in development occur, then the test indicates that the chemical is thyroid active (T2). If development is not advanced, then the toxicity of the compound should be evaluated (D3).

T2: Thyroid active

Advanced development is only known to occur through effects which are thyroid hormone related. These may be peripheral tissue effects such as direct interaction with the thyroid hormone receptor (such as with T4) or effects which alter the circulating thyroid hormone levels. In either case, this is considered sufficient evidence to indicate that the chemical has thyroid activity. Results from Phase-1 and Phase-2 experiments indicate that thyroid histopathology based on day 21 tissue samples does not provide a sensitive and relevant endpoint to confirm agonistic activities.

D3: Toxicity

Evaluation of non-thyroidal toxicity is an essential element of the test to reduce the probability of false positive outcomes. Excessive mortality, of course, is an obvious indication that other toxic mechanisms are occurring. Similarly, reductions in growth, as determined by wet weight and/or organism length, also indicate non-thyroidal toxicity. Apparent increases in growth are commonly observed with compounds that negatively effect normal development. Consequently, larger animals do not indicate non-thyroidal toxicity. Other endpoints may also be considered here, including: edema, hemorrhagic lesions, lethargy, reduced food consumption, erratic/altered swimming behavior, etc. If the compound elicits clear signs of overt toxicity, then the compound is considered apparently thyroid inactive (T3) even if there is an accompanying retardation in development. The rationale for this decision is that overt toxicity is likely to retard development through non-thyroidal actions, such as interference with normal metabolism and physiology, reduced feeding, and other toxicity related complications. If no overt toxicity is observed, then histopathology of the thyroid gland is indicated (P2). Further guidance will be required to distinguish between what should be considered signs of overt toxicity and what characterizes a rather modest retardation of growth that will not immediately lead to step T3.

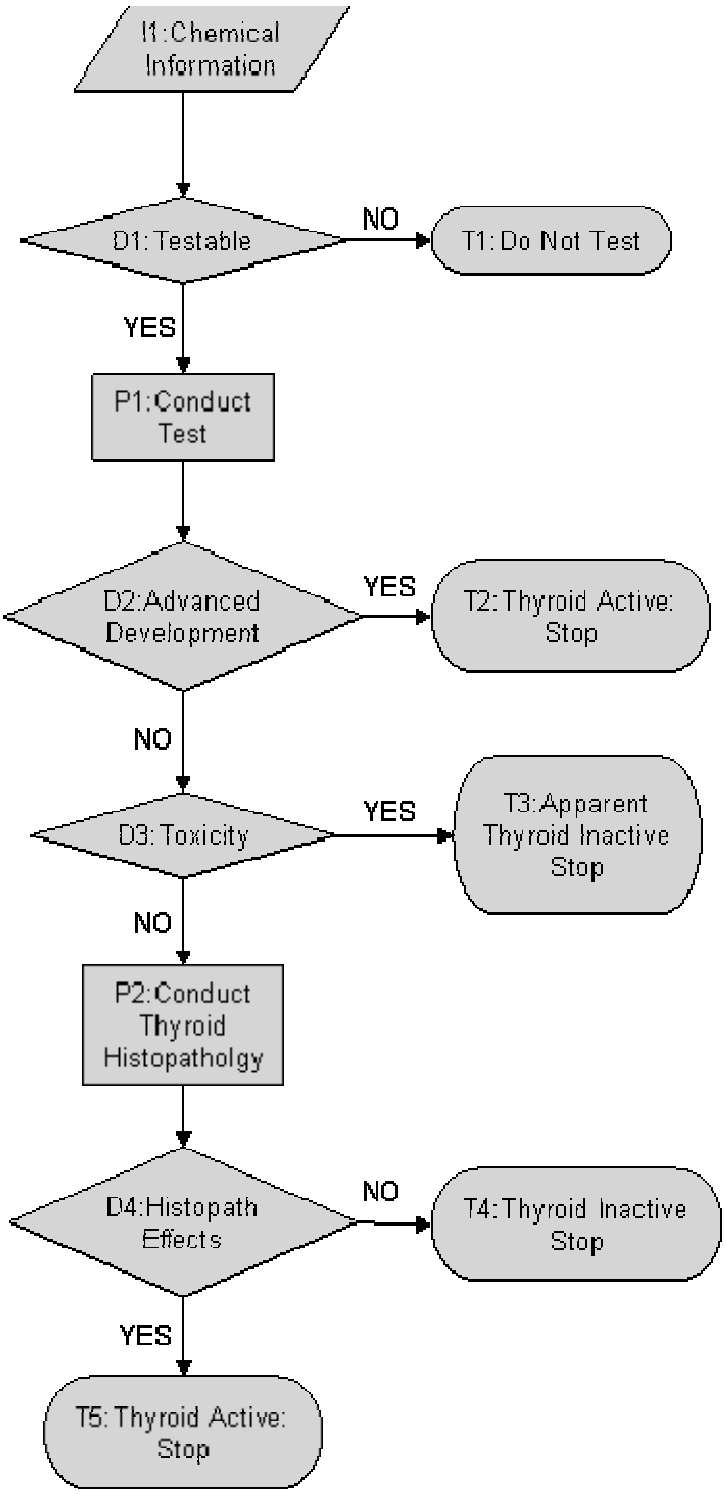

Figure 13. Flow chart describing the proposal of a decision logic for the conduct and the interpretation of the amphibian metamorphosis assay

Footnote to D3: In a multi-dose study, those dose groups without overt toxicity may be used to evaluate the thyroid and thyroid- sensitive endpoints. If an unconfounded dose group is not available, then the study should be rerun.

T3: Apparent thyroid inactive

Apparent thyroid inactivity is based on the confounding effects of overt toxicity. This does not indicate that the compound has no thyroidal effects, but is an operational definition.

P2: Conduct thyroid histopathology

If the chemical does not cause overt toxicity and does not accelerate development, then histopathology of the thyroid gland should be evaluated using the appropriate guidance document. This analysis needs to be conducted whether or not the chemical retards development or has no effect on development. Developmental retardation, in the absence of toxicity, is a strong indicator of anti-thyroid activity, but the developmental stage analysis is less sensitive and less diagnostic than the histopathological analysis of the thyroid gland. Therefore, conducting histopathological analyses of the thyroid glands is required at this point.

D4: Histopathological effects

Histopathology of the thyroid glands is the most sensitive and diagnostic endpoint in this assay. Effects on thyroid histology have been demonstrated in the absence of developmental effects, as noted earlier. If changes in thyroid histopathology of the thyroid occur, then the chemical is considered to be thyroid active (T5). If no changes are observed in thyroid histopathology, then the chemical is considered to be thyroid inactive (T4). The rationale for this decision point is that the thyroid gland is under the influence of TSH and any chemical which alters circulating thyroid hormone sufficiently to alter TSH secretion will result in histopathological changes in the thyroid gland. Various modes and mechanisms of action can alter circulating thyroid hormone. So, while this endpoint is indicative of a thyroid related effect, it is insufficient at determining which mode or mechanism of action is related to the response.

T4: Thyroid inactive

This is the final determination that the chemical on test has no detectable effects on the thyroid pathway.

T5: Thyroid active

This is the final determination that the chemical on test effects the thyroid pathway.

CONCLUSION

135. In conclusion, the validation programme for the amphibian metamorphosis assay successfully achieved the goal of demonstrating the ability of the proposed study protocol to detect thyroid system disrupters. The Phase-1 and Phase-2 studies could clearly show that the proposed combination of different apical morphological endpoint measurements with thyroid histopathology provides for a a highly sensitive and reproducible test method to detect a wide array of modes and mechanisms of action that disrupt thyroid system function. The response profiles of the various endpoints were different for the individual test substances but reproducible across laboratories. The changes in specific endpoints caused by each of the test substances were in close agreement with the expected modes of action. The Phase-2 results support the inclusion of the proposed endpoints as useful and relevant for a protocol of the amphibain metamorphosis

assay. The discussion of Phase-1 and Phase-2 results culminated in a proposal for a decision logic for the conduct and interpretation of the assay.

136. The Phase-2 results indicate that some further work is necessary to identify appropriate approaches for the analysis of data from several endpoint measurements. Specific problems include the need for a normalization of HLL for differences in body size, the appropriate consideration of inter-individual differences in developmental stage during analysis and interpretation of day 21 measurements of growth-related parameters and histopathological findings. The Phase-2 results suggest that additional statistical consultation and discussion within the Expert Group for Amphibian testing to resolve some of these issues has merit. In addition, further optimization and standardization of the study protocol is required to reduce the existing inter-laboratory differences in tadpole growth and development. A major aspect in this regard is a further standardization of the feeding regime.

INVESTIGATION OF THE POWER PROPERTIES OF IMPORTANT ENDPOINTS OF THE AMPHIBIAN METAMORPHOSIS ASSAY

137. This section provides a short overview of the power properties of the developmental stage and hind-limb length endpoints, under the conditions (or close to the conditions) of the Phase 2 experimental design. Several other power studies were run with variations of the Phase 2 experimental test design, and are not presented here to keep this section concise. When improvement of power properties is desirable, solutions were proposed (e.g. for development stage). The increase of power versus the technical feasibility of these improvements and the resources involved are not discussed in this section, but remain for discussion for drafting a Test Guideline in the future.

Development stage endpoint

138. The following tables present the power simulations for the development stage endpoint under the conditions of the assay as it was performed in Phase 2. As the percent change to be detected increases, the power also increases; thus it is important to know the size of the effect to be detected. In the case of the development stage endpoint, a change of 5% corresponds to a shift of approximately 3 stages.

139. Williams and Dunnett tests are shown but should be ignored (violate requirements of tests, because of the non-normality of development stage data). The Mann Whitney tests with a Bonferroni-Holm adjustment to the p-values proved to have extremely low power when there are three or fewer replicates, but has reasonable power with four or more replicates per treatment.

140. Generally, the Jonckheere-Terpstra test demonstrates the most acceptable power properties (shaded cells); an effect should be detected at the highest dose in 61% of the cases simulated using the Phase 2 test design.

Power to detect an effect at Dose 3 (second highest concentration: doses are 0, 1, ,2 ,3 ,4), equal allocation of reps

5 doses (4 concentrations+control); Sample size n=15; Replicates= 4 (equal allocation)

Statistical test	Percent change to be detected		
	5%	10%	20%
Dunn	12	33	82
Dunnett	35	82	100
Jonckh-Terpst.	57	92	100
Mann Whitney	2	6	100
Williams	-	-	14

Power to detect an effect at Dose 4 (highest concentration: doses are 0, 1, ,2 ,3 ,4)

5 doses (4 concentrations+control); Sample size n=15; Replicates= 4 (equal allocation)

Statistical test	Percent change to be detected		
	5%	10%	20%
Dunn	26	80	100
Dunnett	37	92	100
Jonckh-Terpst.	61	98	100
Mann Whitney	3	10	28
Williams	-	-	-

141. The power of the test for this endpoint could be improved by the following amendments to the experimental design (for example for the Jonckheere test):

- 6-dose design (Control+5 concentrations)
  - 5 reps of 5 subjects (85% and 98% power to detect 5% change at Dose 3 and at Dose 5 respectively)
  - 3 reps of 10 subjects (76% and 93% power to detect 5% change at Dose 3 and Dose 5 respectively)
- 5-dose design (Control+4 concentrations)
  - 6 reps of 15 subjects (74% and 76% power at Dose 3 and Dose 4 respectively)
  - 5 reps of 20 subjects (70% and 73% power at Dose 3 and Dose 4 respectively)

Hind-limb length endpoint

142. Hind-limb length is a continuous value, unlike development stage data, and is normally distributed. This endpoint is mainly used for the detection of agonistic effects and measured at day 7 of exposure. The size of the effect measured is about a 50% increase (i.e. from about 2.4 mm to 3.6 mm) of the hind-limb length when the tadpole is exposed to a thyroid hormone agonist.

143. No power simulation is available for the type of experiment run in Phase 2, i.e. for 5 doses. However, power simulation was made for the experiments with 4 doses and with 6 doses.

144. Generally, the Jonckheere-Terpstra test demonstrates good power properties (shaded cells).

Power to detect an effect at Dose 3 (highest concentration: doses are 0, 1, ,2 ,3)

4 doses (3 concentrations+control); Sample size n=5; Replicates= 4 (equal allocation)

Statistical test	Percent change to be detected				
	10%	20%	30%	40%	50%
Dunn	6	17	35	58	77
Dunnett	9	22	44	68	87
Jonckh-Terpst.	16	38	62	83	94
Mann Whitney	6	14	28	47	66
Williams	16	35	59	81	93

Power to detect an effect at Dose 5 (Highest concentration: doses are 0, 1, ,2 ,3 ,4, 5)

6 doses (5 concentrations+control); Sample size n=5; Replicates= 4 (equal allocation)

Statistical test	Percent change to be detected				
	10%	20%	30%	40%	50%
Dunn	4	16	35	57	71
Dunnett	7	26	54	82	95
Jonckh-Terpst.	21	52	80	94	99
Mann Whitney	0	0	0	0	0
Williams	-	-	-	-	-

REFERENCES

Brucker-Davis, F. (1998): Effects of environmental synthetic chemicals on thyroid function. Thyroid 8, 827-856

Becker, K.B., Stephens, K.C., Davey, J.C., Schneider, M.J. and Galton, V.A. (1997): The type 2 and type 3 iodothyronine deiodinases play important roles in coordinating development in *Rana catesbeiana* tadpoles. Endocrinol. 138, 2989-2997

Bianco, A.C., Salvatore, D., Gereben, B., Berry, M.J. and Larsen, P.R. (2002): Biochemistry, cellular and molecular biology, and physiological roles of the iodothyronine selenodeiodinases. Endocr. Rev. 23, 38-89

Braga, M. and Cooper, D.S. (2001): Oral cholecystographic agents and the thyroid. J. Clin. Endocrinol. Metab. 86, 1853-1860

Brown, D.D. (2005): The role of deiodinases in amphibian metamorphosis. Thyroid 15, 815-821

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

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. (2005): Progress towards development of an amphibian-based thyroid screening assay using *Xenopus laevis*. Organismal and thyroidal responses to the model compounds 6-propylthiouracil, methimazole, and thyroxine. Toxicol. Sci. 87, 353-364

Dohan, O., De la Vieja, A., Paroder, V., Riedel, C., Artani, M., Reed, M., Ginter, C.S. and Carrasco, N. (2003): The sodium/iodide symporter (NIS): characterization, regulation, and medical significance. Endocr. Rev. 24, 48-77

Galton, V.A. and Hiebert, A. (1988): The ontogeny of iodothyronine 5'-monodeiodinase activity in *Rana catesbeiana* tadpoles. Endocrinol. 122, 640-645

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

Huang, H.C., Cai, L.Q., Remo, B.F. and Brown, D.D. (2001): Timing of metamorphosis and the onset of the negative feedback loop between the thyroid gland and the pituitary is controlled by type II iodothyronine deiodinase in *Xenopus laevis*. Proc. Natl. Acad. Sci. USA 98, 7348-7353

Kanno, J., Goda, Y., Sato, K., Yoshihira, K. and Hayashi, Y. (1995): FD&C Red No. 105 (Rose Bengal B) neutralizes the thyroid tumor promoting effect of iodine-deficient diet in rats. Toxicology 99, 107-113

Kanno, J., Onodera, H., Furuta, K., Maekawa, A., Kasuga, T., and Hayashi, Y. (1992): Tumor-promoting effects of both iodine deficiency and iodine excess in the rat thyroid. Toxicol. Pathol. 20, 226-235

Kawahara, A., Gohda, Y. and Hikosaka, A. (1999): Role of type III iodothyronine 5-deiodinase gene expression in temporal regulation of *Xenopus* metamorphosis. Dev. Growth Diff. 41, 365-373

Köhrle, J. (1996): Thyroid hormone deiodinases: a selenoenzyme family acting as gate keepers to thyroid hormone action. Acta Med. Austriaca 23, 17-30

Kühn, E.R., Mol, K.A. and Darras, V.M. (1993): Control strategies of thyroid hormone monodeiodination in vertebrates. Zool. Sci. 10, 873-885

Kuiper, G.G., Klootwijk, W., Morvan, D.G., Destree, O., Darras, V.M., Van der Geyten, S., Demeneix, B., and Visser, T.J. (2006): Characterization of recombinant *Xenopus laevis* type I iodothyronine deiodinase: substitution of a proline residue in the catalytic center by serine (Pro132Ser) restores sensitivity to 6-propyl-2-thiouracil. Endocrinology 147, 3519-3529

Lynn, W.G. and Dent, J.N. (1960): The action of various goitrogens in inhibiting localization of radioiodine in the thyroid and thymus glands of larval tree frogs. Biol. Bull. 118, 430-438

Markou, K., Georgopoulos, N., Kyriazopoulou, V., and Vagenakis, A. G. (2001): Iodine-Induced hypothyroidism. Thyroid 11, 501-510

Mitsui, N., Fujii, T., Miyahara, M., Oka, T., Kashiwagi, A., Kashiwagi, K., Hanada, H., Urushitani, H., Santo, N., Tooi, O. and Iguchi, T. (2005): Development of metamorphosis assay using *Silurana tropicalis* for the detection of thyroid system-disrupting chemicals. Ecotoxicol. Environ. Saf. 64, 281-287

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

Norris, D.O. and Platt, J.E. (1973): Effects of pituitary hormones, melatonin, and thyroidal inhibitors on radioiodide uptake by the thyroid glands of larval and adult tiger salamanders, *Ambystoma tigrinum* (Amphibia: Caudata). Gen. Comp. Endocrinol. 21, 368-376

Opitz, R., Braunbeck, T., Bogi, C., Pickford, D. B., Nentwig, G., Oehlmann, J., Tooi, O., Lutz, I., and Kloas, W. (2005): Description and initial evaluation of a *Xenopus* metamorphosis assay for detection of thyroid system-disrupting activities of environmental compounds. Environ. Toxicol. Chem. 24, 653-664

Opitz, R., Hartmann, S., Blank, T., Braunbeck, T., Lutz, I., and Kloas, W. (2006): Evaluation of Histological and Molecular Endpoints for Enhanced Detection of Thyroid System Disruption in *Xenopus laevis* Tadpoles. Toxicol. Sci. 90, 337-348

Paire, A., Bernier-Valentin F, Rabilloud R, Watrin C, Selmi-Ruby S, Rousset B. (1998): Expression of alpha- and beta-subunits and activity of Na^+K^+ ATPase in pig thyroid cells in primary culture: modulation by thyrotropin and thyroid hormones. Mol. Cell. Endocrinol. 146, 93-101

Rosenkilde, P. and Ussing, A.P. (1990): Regulation of metamorphosis. In: Biology and Physiology of Amphibians. Fortschritte der Zoologie Vol. 38, Hanke, W. (ed.), Gustav Fischer Verlag, Stuttgart, 1990, 125-138

Shi, Y.-B. (1999): Amphibian Metamorphosis. From Morphology to Molecular Biology. John Wiley & Sons, New York, US, 1999, pp. 288.

Shintani, N., Nohira, T., Hikosaka, A. and Kawahara, A. (2002): Tissue-specific regulation of type III iodothyronine 5-deiodinase gene expression mediates the effects of prolactin and growth hormone in *Xenopus* metamorphosis. Dev. Growth Diff. 44, 327-335

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. (2005): Metamorphic inhibition of *Xenopus laevis* by sodium perchlorate: effects on development and thyroid histology. Environ. Toxicol. Chem. 24, 926-933

Tonacchera, M., Pinchera, A., Dimida, A., Ferrarini, E., Agretti, P., Vitti, P., Santini, F., Crump, K. and Gibbs, J. (2004): Relative potencies and additivity of perchlorate, thiocyanate, nitrate, and iodide on the inhibition of radioactive iodide uptake by the human sodium iodide symporter. Thyroid 14, 1012-1019

Wolff, J. (1998): Perchlorate and the thyroid gland. Pharmacol. Rev. 50, 89-105

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