DRAFT

Report of the Phase-3 Validation of the Amphibian Metamorphosis Assay

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 The OECD Program for the Amphibian Metamorphosis Assay

This report summarizes the results from an OECD study conducted in 2007 to assess the utility of the protocol for the Amphibian Metamorphosis Assay to detect weakly active thyroid system-disrupting substances and to distinguish thyroid system-related activities from activities resulting from mechanisms not directly related to thyroid system function. These studies comprise the Phase-3 of validation of the study protocol and were preceded by two inter-laboratory exercises addressing the optimization of an initial study protocol (validation Phase-1) and the inter-laboratory reproducibility of an enhanced study protocol (validation Phase-2). The Phase-3 study was designed and directed by the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA).

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; Fort et al., 2007; 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).

To date, a validated assay specifically addressing 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.

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 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, histopathological evaluation of thyroid tissue 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).

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.

Based on the results of the Phase-1 studies, an improved study protocol was established for use in the subsequent Phase-2 studies. The exposure was initiated with stage 51 *X. laevis* tadpoles. 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 included 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.

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 competitive 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. 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 conclusions were drawn when combining the results of Phase-1 and Phase-2 studies. Assessment of morphological endpoints such as developmental stage and HLL was 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 detect 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 supported the inclusion of the proposed endpoints as useful and relevant to allow for the detection of a wide array of modes of action.

The Phase-2 results indicated that some further work is necessary to identify appropriate approaches for the analysis of data from several endpoint measurements. Specific problems

included 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 suggested that additional statistical consultation and discussion within the Expert Group to resolve some of these issues has merit. In addition, further optimization and standardization of the study protocol was required to reduce the existing inter-laboratory differences in tadpole growth and development. A major aspect in this regard was a further standardization of the feeding regime.

1.2 Proposal for Phase-3 Studies

A draft proposal explaining specific experimental activities to be included in a Phase-3 validation exercise was presented at the fifth meeting of the Validation Management Group for Ecotoxicity Testing (VMG-eco) in Madrid (January 23-24, 2007). At this meeting, it was stressed that because Phase-2 already addressed the inter-laboratory reproducibility of the assay, there should not be an obligation in Phase-3 to have inter-laboratory testing of the test substances, and it was proposed that single laboratory assessment of the test substances should be sufficient. Comments raised by VMG-eco about the Phase-3 study plan were addressed in a teleconference of Expert Group members in February 2007. Specific issues discussed by the Expert Group included the choice of appropriate test substances and the need for a further standardization of the feeding protocol.

1.3 Identification of Test Substances

The Expert Group on Amphibian Testing agreed that, for the specific purposes of Phase-3, the following test substances are of equal interest in principle either as weakly active or negative control: benzophenone-2 (BP-2), 17β-estradiol (E2), potassium iodide (KI) and p,p'-DDE.

The rationale for considering the UV-filter BP-2 (CAS: 131-55-5) and p,p'-DDE (1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene; CAS: 72-55-9) as test substances was that weak anti-thyroidal activity has been shown for both compounds in a number of rodent studies (Brucker-Davis, 1998; Jarry et al., 2004; Seidlova-Wuttke et al., 2005).

The rationale for using E2 (CAS: 50-28-2) as a test substance in Phase-3 studies is that it is a potent endocrine active compound that acts via signalling pathways not directly related to the thyroid system. Given that many endocrine-disrupting chemicals possess weak or moderate

estrogenic activity, the estrogen E2 was considered as a representative for estrogenic agents. Therefore, testing of the potent estrogen E2 over a wider concentration range should provide information about the ability of the amphibian assay to distinguish thyroid system-related and unrelated mechanisms of endocrine disruption (e.g., estrogenic action).

Iodide is an important element for TH synthesis and differences in iodide supply are known to modulate the activity of the thyroid gland from fish to human. A well-known problem in assessing the thyroid system-disrupting properties of many iodine-rich compounds (e.g., IOP) is to distinguish between effects of the parent compound and effects caused by iodide which is deliberated during metabolic processes. Another aspect that qualified KI (7681 -11-0) as a possible test substance for the Phase-3 studies was the need to assess whether increased availability of iodide alone could result in developmental acceleration, an effect that is currently proposed for thyroid hormone agonists.

2 Methods

Validation Phase-3 of the Amphibian Metamorphosis Assay consisted of single laboratory assessments of specific test compounds. A total of three experimental studies were performed in two laboratories.

2.1 Experimental Protocols Used in Phase-3

Table 2 provides an overview of the Phase-3 protocol. The testing protocol used during Phase-3 was identical to the test protocol used in the Phase-2 studies with the exception of a more standardized feeding scheme being used in Phase-3 (Table 1). A detailed description of the study protocol is available 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) of the validation Phase-2 report. Below, some major aspects of the study protocol are briefly summarized.

2.2 General Study Design

The general exposure protocol for Phase-3 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.

2.3 Test Conditions

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. Aeration of treatment tanks

was required when dissolved oxygen concentrations fall below 40% of the air saturation value.

2.4 Standardized Feeding Scheme

Based on the general animal performance observed in the experiments during validation Phase-2, a more standardized feeding scheme has been developed for Phase-3 studies to minimize inter-individual differences in tadpole growth and development (Table 1). The daily food ration is given per animal, so that total amounts of food applied to the test tanks can be adjusted in cases of mortality. Another standardization was that the daily food ration must be divided into at least two smaller portions fed during the day.

Table 1. Feeding schedule with Sera Micron during Phase-3 studies

Study day	Daily food ration (mg Sera Micron per animal)
0-4	30
5-7	40
7-10	50
11-14	70
15-21	80

Bulk quantities of Sera micron were available at lab 1 and portions of it were shipped to lab 2 so that all studies during Phase-3 used the same lot of sera micron.

2.5 Sampling Schedule

The Phase-3 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.

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.

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.

2.6 Endpoint Measurements

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.

Determinations of the developmental parameters, stage and HLL, were performed 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.

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.

2.7 Test Substances

The two test substances used in Phase-3 experimental work were BP-2 and E2. Bulk quantities of BP-2 (2,2',4,4'-tetrahydroxybenzophenone; CAS: 131-55-5) were obtained from Molekula (Wimborne Dorset, UK) at a purity of 99.45%. The same lot of BP-2 (Batch: 52546) was used in both BP-2 studies. E2 (CAS: 50-28-2) was obtained from Sigma (Deisenhofen, Germany).

Table 2. Summary of main testing conditions and endpoint measurements used in the study protocol for Phase-3 validation of Amphibian Metamorphosis Assay.

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	**E2**	0.08, 0.4, 2.0, 10.0 µg/l	
	BP-2	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 histology (day 21)	Number	5 individuals per replicate tank (randomly selected)	
	Region	whole body	
	Fixation fluid	Davidson's Fixative	
Feeding	Food	Sera micron	
	Frequency/amount	see Table 1	
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	

Abbrevations: WBL – whole body length; SVL – snout-vent length

2.8 Statistical Analysis

The statistical protocol used for analysis of Phase-3 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). The statistical decision procedure is taken from *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application*, OECD Series on Testing and Assessment, No. 54, ENV/JM/MONO(2006)18. This document contains much more detail on all methods used and can be downloaded from www.oecd.org/LongAbstract/0,2546,en_2649_201185_37719579_119669_1_1_1,00.html. Some parts of this flow chart do not apply to the current datasets, for example, there was no solvent control. The complete report of the statistical analyses performed for Phase-3 study results is provided in Annex A of this document.

Some details from this document are discussed here to make clear the specific procedures used under the more general headings of the schematic. The tests used for determining the LOEC and NOEC were (1) the Dunnett (or Tamhane-Dunnett) test, possibly on transformed data to achieve normality and variance stability, (2) the Jonckheere-Terpstra test applied in step-down fashion, or (3) Dunn's test if no normalizing transform was found or, as in the case of developmental stage, the response was considered inherently non-normal.

If measurements of a response attribute (or some transformation of it) were found to be normally distributed with equal variances, then Dunnett's test was used. Where a response (or some transform) was found to be normally distributed but with unequal variances, then the Tamhane-Dunnett test (a robust form of the Dunnett test) was used. Where no normalizing transformation could be found for a response, Dunn's test was used for pairwise comparisons of treatment group medians to the control median.

Prior to performing Jonckheere trend tests, an assessment of dose-response monotonicity was performed, and serious deviation from monotonicity precluded this test. Assessment for monotonicity can be done visually from a plot, numerically from computed means, or formally. Formal tests for monotonicity are described in the previously cited reference.

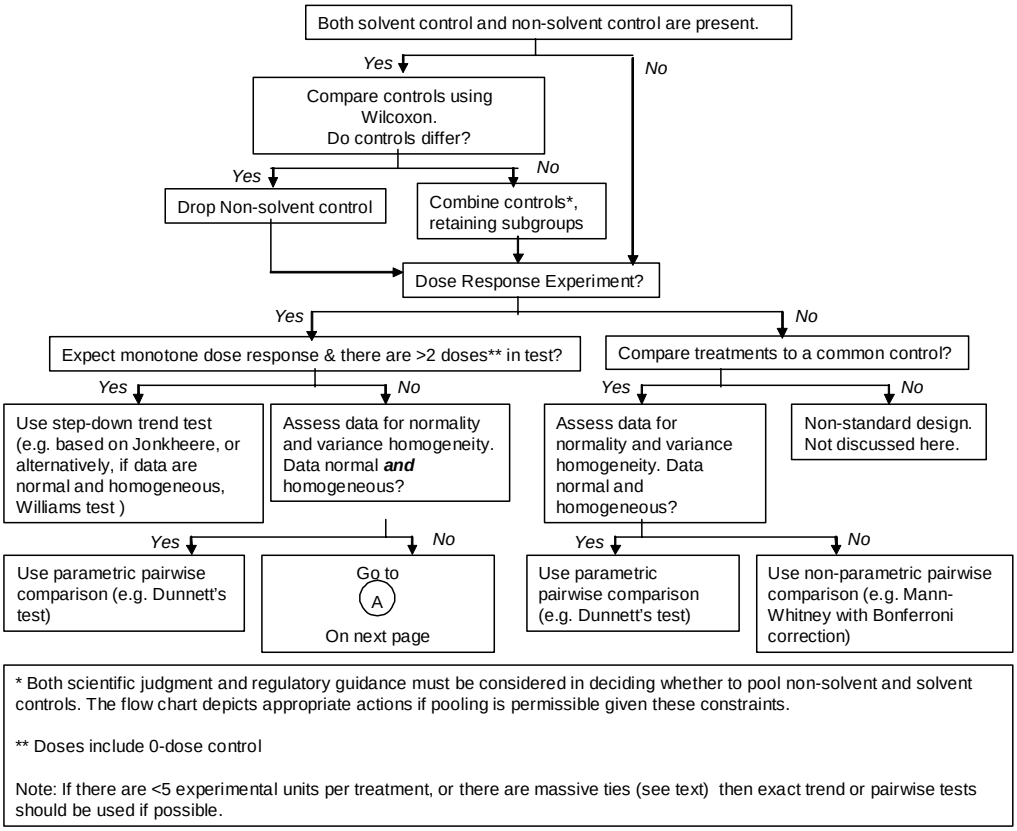
All tests were done in 2-sided fashion, since it was not known in advance what direction, if any, the chemical might affect a given response. This decision decreases the power of all the

tests relative to a 1-sided test done at the same 0.05 significance level, but no rationale justifying a 1-sided test was evident (i.e. for the primary study endpoints it was not clear that an increase would be of concern, whereas a decrease would be of no interest, or *visa versa*).

Since it has been shown in previous studies that the variable values to be analyzed are more correlated within a replicate tank than those in different tanks, the replicate tank is the unit of analysis. Thus, statistical analysis was done on the mean (or, in the case of developmental stage, the median) value for the replicate. In addition to the median for developmental stage, the replicate minimum, maximum, and interquartile range (IQR) were also analyzed.

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.

Regarding the statistical analysis of data from the severity grade scoring of thyroid tissue, specific statistical approaches have not yet been evaluated. For the analysis of severity scores determined in the Phase-3 studies, pairwise comparisons of treatment group means to the control mean were performed using Dunnett's/Tamhane-Dunnett's test.



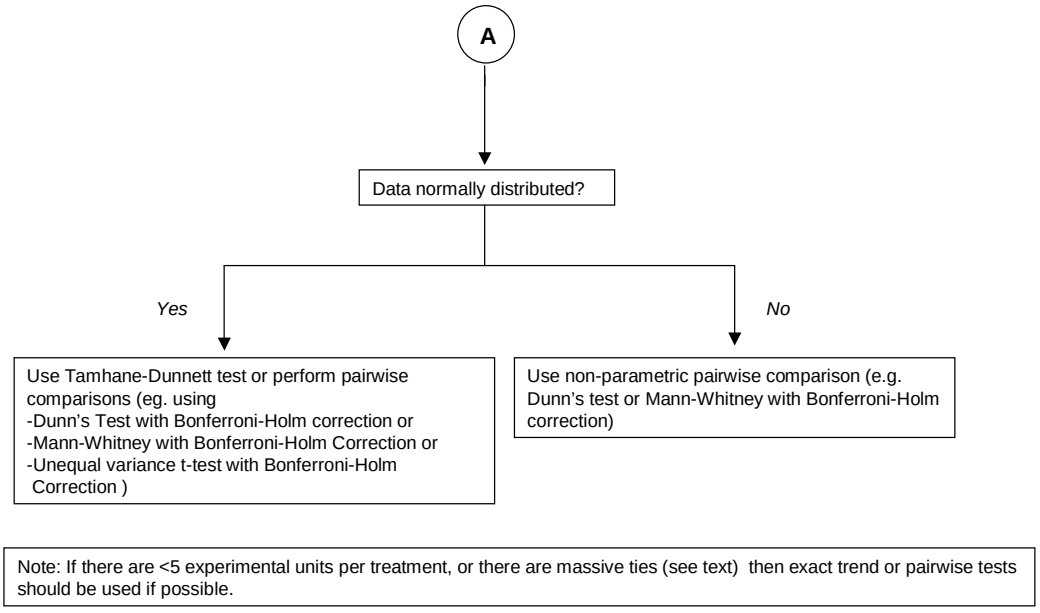


Figure 1. Statistical flow chart describing the recommended statistical protocol for Phase-2 studies.

3 Results

3.1 Control Animal Performance

In the context of the Amphibian Metamorphosis Assay, homogenous rates of tadpole growth and development in the control group are an important component of optimal test performance. 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.

Experiments conducted during the validation Phase-1 and Phase-2 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) yielded important information regarding the optimization and standardization of test parameters in order to control for an acceptable inter-individual variability of tadpole growth and development. It was particularly the analysis of results obtained during the validation Phase-2 of the Amphibian Metamorphosis Assay that highlighted the important role of standardized feeding rates for minimizing the variability of individual growth and developmental rates within the population of test animals. Consultation among participants of validation Phase-2 and discussion of Phase-2 results at the fifth meeting of the Validation Management Group for Ecotoxicity Testing (VMG-eco) in Madrid (January 23-24, 2007) suggested that further standardization of feeding rates could improve the inter-laboratory reproducibility of the study protocol.

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-3 towards their compliance with an array of recommended validity criteria.

3.1.1 Development in the Control Group

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. Table 3 summarizes the age of the tadpoles on study day 0 in the three experiments of validation Phase-3. The data suggest normal development of tadpoles in the pre-exposure phase of all three experiments. 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. Table 4 summarizes the stages observed across all four replicate tanks of the control group on study day 7 in three tests performed during validation Phase-3 experimental work. In all three experiments during Phase-3, control tadpoles showed development to early prometamorphic stages 54-55 by study day 7. Developmental rates in the control group were very similar for the E2 experiment in lab 1 and the BP-2 experiment in lab 2. Slightly slower development was observed for the control group of the BP-2 experiment in lab 1.

Table 3. Age of test animals on study day 0 (exposure initiation).

	age at test initiation (post fertilization day)	
study	**lab 1**	**lab 2**
E2	12	n.s.
BP-2	13	12

Note: n.s. no study conducted

Table 4. 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	E2	**20**			4	16	55	2
1	BP-2	**20**		1	19		54	2
2	BP-2	**20**			2	18	55	2

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.

A third criterion of normal test animal performance is the developmental stage reached by the control animal population on exposure day 21. Table 5 summarizes the distribution of stages observed in the control group on study day 21 for all three tests performed during validation Phase-3. In agreement with observations made on study day 7, developmental rates in the control group were very similar for the E2 experiment in lab 1 and the BP-2 experiment in lab 2. Slightly slower development was observed for the control group of the BP-2 experiment in lab 1. For all three experiments, however, the distribution of developmental stages was very similar across all four replicate tanks. Furthermore, in each of the three exposure studies, more than 90% of control tadpoles were distributed across 3 to 4 different developmental stages by study day 21 in each of the three exposure studies. These results demonstrate a low inter-individual variability in tadpole development in the control group in all experiments.

Table 5. Results from developmental stage determination in the control groups on study day 21.

				developmental stage (Nieuwkoop and Faber, 1994)									**stage**	**stage**
lab	**test**	**Rep.[a]**	***n*[b]**	**56**	**57**	**58**	**59**	**60**	**61**	**62**	**63**	**64**	**(median)**	**(range)**
1	E2	A	15			2	3	3	3	4			60	5
		B	15				3	5	2	4	1		60	5
		C	15			1	2	5	2	4	1		60	6
		D	15				2	4	4	5			61	5
		Pool	**60**			**3**	**10**	**17**	**11**	**17**	**2**		**60.5**	**6**
1	BP-2	A	15		3	11	1						58	3
		B	15		7	6	2						58	3
		C	15		7	4	4						58	3
		D	15		6	5	3	1					58	4
		Pool	**60**		**23**	**26**	**10**	**1**					**58**	**4**
2	BP-2	A	15				1	6	3	4	1		61	5
		B	15				1	8	3	2	1		60	5
		C	15					6	4	3	2		61	4
		D	15				1	8	5	1			60	4
		Pool	**60**				**3**	**28**	**15**	**10**	**4**		**60**	**5**

Note: [a] Stage data are shown for individual replicate tanks (A, B, C, and D) and after pooling of stage data from all four replicate tanks (pool). The number of tadpoles examined (n) and the number of tadpoles at each specific developmental stage is shown.

3.1.2 Growth in the Control Group

In addition to the monitoring of developmental stages, the study protocol 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. Mean values of WBL, SVL and wet weight measurements in individual replicate tanks of the control group on day 7 and day 21 are shown in Figure 2 and Figure 3. Control tadpoles used in the two independent experiments in lab 1 showed similar growth rates indicating low intra-laboratory variability. In comparison to lab 1, control tadpoles used in the experiment performed in lab 2 were greater in size, as judged from WBL and SVL measurements on day 7, and had increased body weights.

To obtain a first gross estimation of variability of the various growth-related parameters in the control animal population, CV values (coefficient of variation) were calculated for each replicate tank of the control group on study day 7 (Figure 2) and 21 (Figure 3) using individual animals as replicates. CV values calculated for replicate tanks indicated that the inter-individual variability of tadpole growth was low (<15% for WBL, <10% for SVL, and <25% for wet weight) on day 7 but was slightly increased by day 21 (<20% for WBL, <20% for SVL, and <35% for wet weight). When combining the information from the developmental stage determination on day 21 (Table 5) with the results from the CV calculation (Figure 3), it becomes apparent that the development of control tadpoles to more advanced stages in the E2 experiment in lab 1 and the BP-2 experiment in lab 2 resulted in an increase in the inter-individual variability compared to the BP-2 experiment in lab 1. This is a common observation in experiments with tadpoles as animals entering climax stages readily experience marked reductions in body size and wet weight due to tissue resorption. On the other hand, the variability among tank means within the control group was very low in all three experiments (CV values of <5.0% for WBL, <5.0% for SVL, and <10% for wet weight). Taken together, while slight differences in growth and development of control group tadpoles were observed across independent experiments, the currently used testing protocol efficiently controls for an acceptable inter-individual variability and a low variability between replicate tanks.

1

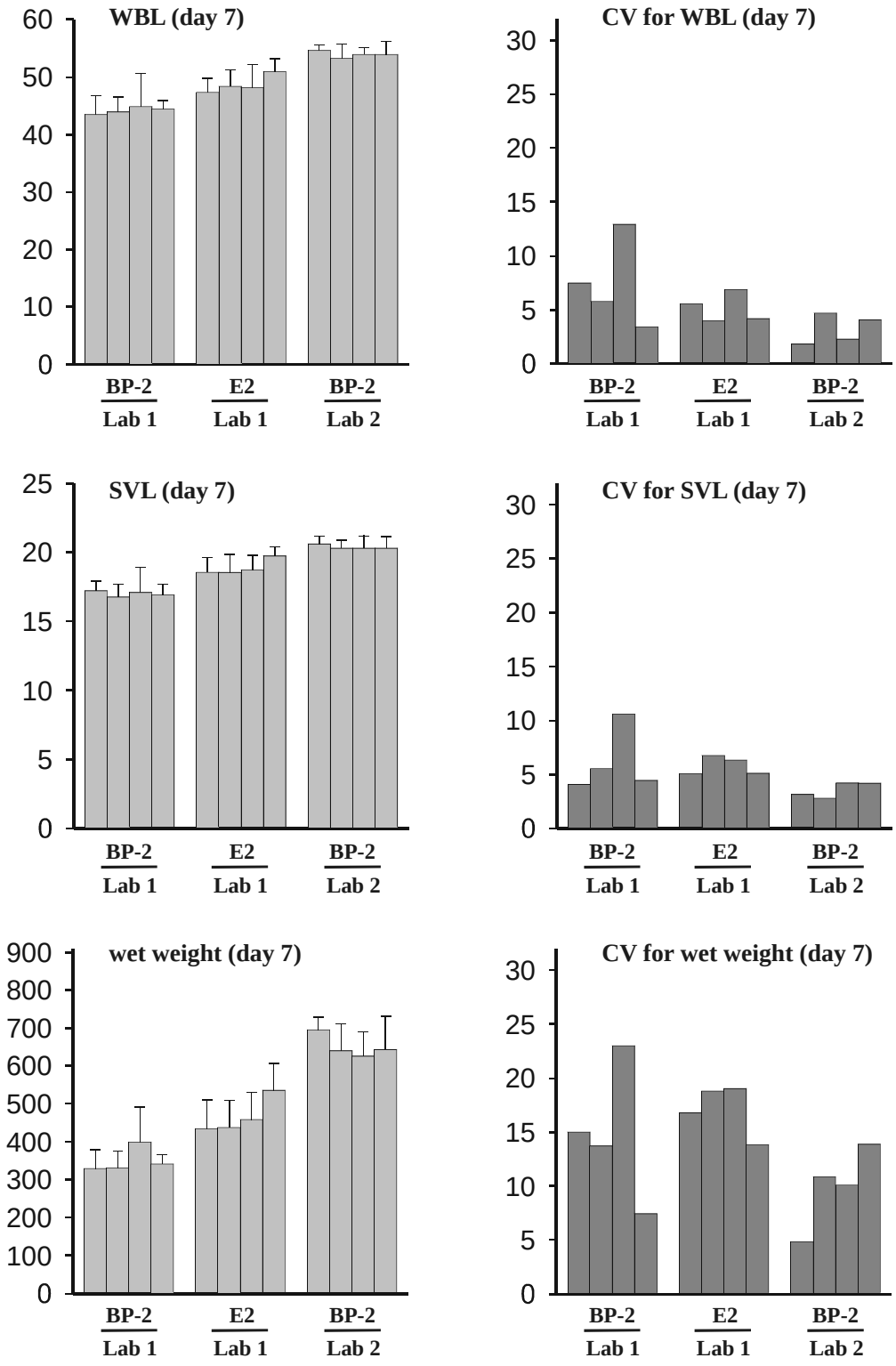


Figure 2. Growth-related parameters determined on day 7 in the control treatment groups of experiments performed with E2 and BP-2 as test substances. The left panel shows mean values of WBL (mm), SVL (mm), and wet weight (mg) for four individual replicate tanks of the control group. The right panel shows the corresponding CV values (%). Individual tadpoles (n=5) served as replicates in the determination of tank means and standard deviations.

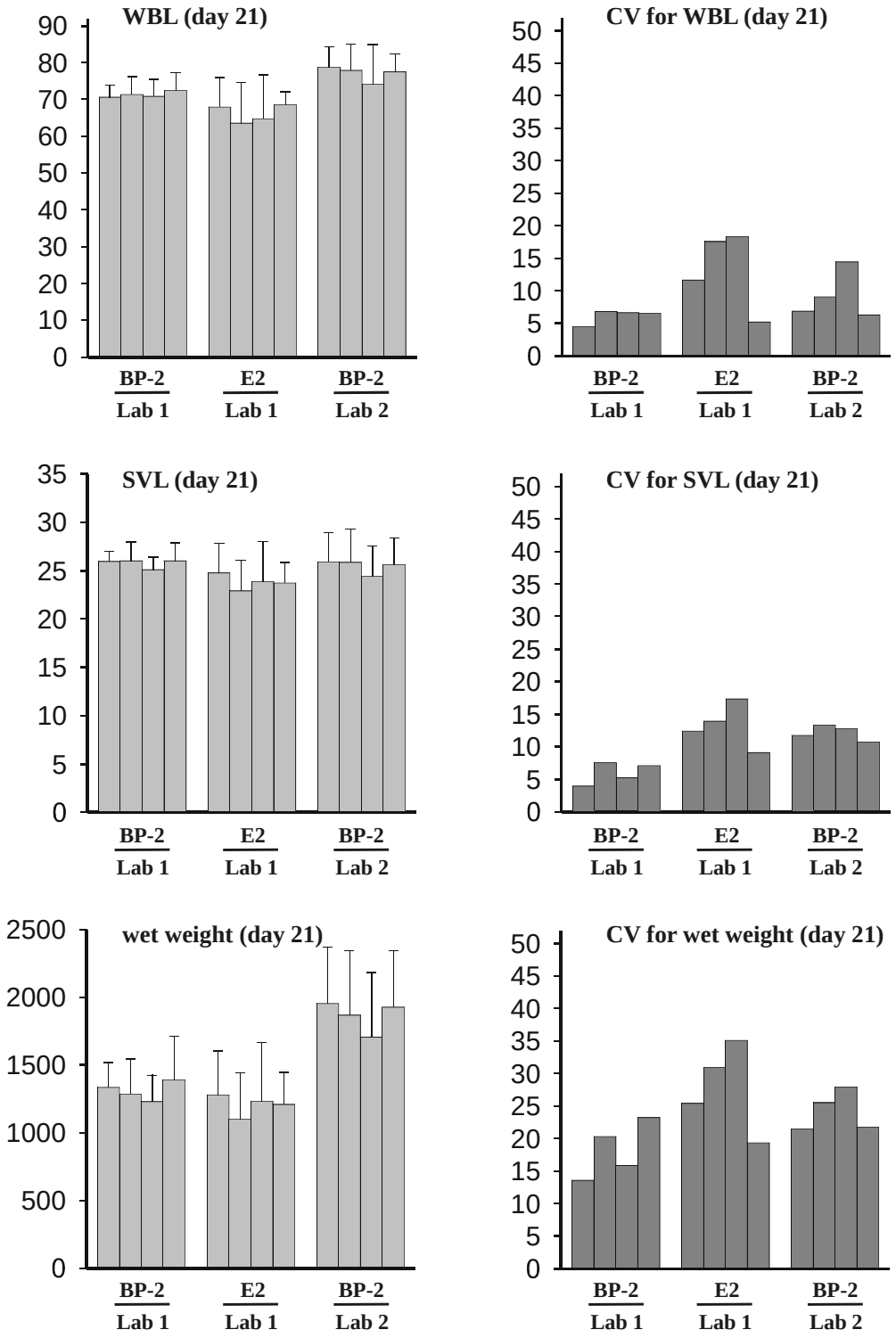


Figure 3. Growth-related parameters determined on day 21 in the control treatment groups of experiments performed with E2 and BP-2 as test substances. The left panel shows mean values of WBL (mm), SVL (mm), and wet weight (mg) for four individual replicate tanks of the control group. The right panel shows the corresponding CV values (%). Individual tadpoles (n=5) served as replicates in the determination of tank means and standard deviations.

3.2 Test Results – 17β-Estradiol (E2)

3.2.1 Overview

During the validation Phase-3 study of the Amphibian Metamorphosis Assay, one exposure experiment with 17β-estradiol (E2) as test substance was performed. The nominal aqueous concentrations for E2 were 0, 0.08, 0.4, 2.0, and 10 µg/l. Mortality was not observed in this experiment. On study day 7, analysis of growth-related parameters (e.g., WBL, SVL, wet weight) and developmental parameters (e.g., HLL and developmental stage) did not show any difference between control and E2 treatment groups. On study day 21, size and weight of E2-treated tadpoles was increased at high E2 concentrations. Results from stage determination on day 21 did not reveal significant effects of E2 on development but HLL measurements indicated a slight reduction in hind limb growth at high E2 concentrations. Histopathological assessment of thyroid tissue did not reveal remarkable effects of E2 on the thyroid gland. Gross morphological assessment of gonads showed a concentration-dependent feminization of tadpoles. Abnormal testicular development was observed at 0.4 µg/l E2 while higher E2 concentrations caused an almost complete male-to-female sex reversal as indicated by the lack of animals with testicular tissue.

3.2.2 Verification of Aqueous E2 Concentrations

The study protocol called for verification of aqueous E2 concentrations in each treatment tank at least once per week. Actual E2 concentrations in test tanks were analyzed using a sensitive ELISA technique (Abraxis LLC, Warminster, US) after appropriate dilution of test solution samples. During the course of the *in vivo* exposure, it was necessary to gradually increase (up to 1.8-fold compared to the pre-exposure phase) the amount of E2 delivered to the test tanks in order to maintain aqueous E2 concentrations close to nominal concentrations. The actual test concentrations of E2 as determined by ELISA are shown in Table 6. Mean measured E2 concentrations in replicate tanks through time ranged from 0.068 to 0.075 µg/l E2 for the 0.08 µg/l treatment group, from 0.30 to 0.37 µg/l E2 for the 0.4 µg/l treatment group, from 1.63 to 1.88 µg/l E2 for the 2.0 µg/l treatment group, and from 8.46 to 8.95 µg/l E2 for the 10 µg/l treatment group. Thus, the gradual increases of E2 delivery to the test tanks were effective in order to obtain test concentrations close to the target E2 concentrations.

Table 6. E2 concentrations (µg/l) in individual exposure tanks as determined by ELISA

nom. conc.	Rep.	E2 concentrations measured by ELISA[a]				
		day 1	day 6	day 11	day 13	day 20
0.0	A	BDL	BDL	BDL	BDL	BDL
	B	BDL	BDL	BDL	BDL	BDL
	C	BDL	BDL	BDL	BDL	BDL
	D	BDL	BDL	BDL	BDL	BDL
0.08	A	0.068 (85)	0.077 (96)	0.055 (69)	0.07 (88)	0.072 (89)
	B	0.072 (91)	0.069 (86)		0.087 (109)	0.073 (91)
	C	0.064 (79)	0.061 (76)	0.059 (73)	0.084 (105)	0.077 (96)
	D	0.071 (88)	0.058 (73)		0.080 (100)	0.080 (100)
0.4	A	0.39 (98)	0.32 (80)	0.36 (90)	0.43 (108)	0.35 (87)
	B	0.38 (95)	0.33 (83)		0.4 (99)	0.36 (91)
	C	0.38 (96)	0.34 (85)	0.35 (89)	0.32 (81)	0.42 (105)
	D	0.34 (85)	0.31 (77)		0.31 (78)	0.24 (61)
2.0	A	1.90 (95)	1.89 (95)	1.46 (73)	2.1 (105)	2.04 (102)
	B	1.55 (78)	1.25 (63)		2.08 (104)	1.86 (93)
	C	1.93 (97)	1.26 (63)	1.34 (67)	1.61 (80)	2.00 (100)
	D	1.97 (99)	1.46 (73)		1.85 (92)	1.80 (90)
10	A	8.81 (88)	7.8 (78)	9.10 (91)	8.31 (83)	9.27 (93)
	B	8.49 (85)	6.98 (69)		9.23 (92)	10.89 (109)
	C	8.22 (82)	8.88 (89)	8.35 (83)	10.12 (101)	9.21 (92)
	D	6.70 (67)	8.79 (88)		9.38 (94)	8.99 (90)

Note: [a] E2 concentrations shown are the average of duplicate measurements. The percentage of the corresponding nominal E2 concentration is presented for all measurements in parentheses. BDL below detection limit (0.05 µg/l).

3.2.3 Mortality

Mortality of tadpoles was assessed daily throughout the exposure phase. No mortality was observed in any treatment group and no gross morphological or behavioral abnormalities indicating systemic toxicity were observed throughout the exposure study.

3.2.4 Effects of E2 on Growth-related Endpoints

Possible effects of E2 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 E2 did not result in growth retardation. Results from day 21 measurements revealed that tadpoles treated with the two highest E2

concentrations had significantly increased body size and weight (Table 7). The interpretation of these results is, however, confounded by the fact that a large number of tadpoles in all treatment groups already showed development to climax stages (which is associated with weight loss and a reduction in body size). In *X. laevis* tadpoles, body size and weight are highest in tadpoles at developmental stages 59/60 and body size and weight decrease at more advanced stages. In the E2 experiment, the percentage of tadpoles at stages >60 were 50%, 53%, 40%, 30% and 18% for the control and the 0.08, 0.4, 2.0 and 10 µg/l E2 treatments, respectively. The higher mean values of WBL, SVL and wet weight at high E2 concentrations are, therefore, likely the result of the reduced number of climax stage tadpoles observed in high E2 treatments on day 21. Overall, the high number of climax stage tadpoles (stage >60) observed throughout all treatment groups on day 21 of the E2 experiment prevents a sound analysis of possible E2 effects on growth using data from day 21 growth-related measurements.

Table 7. Results from measurements of whole body length (WBL, mm), snout-vent length (SVL, mm), and wet weight (mg) on study day 7 and 21 in the exposure experiment with 17β-estradiol (E2).

	E2	**WBL**	**SVL**	**wet weight**
	µg/l	mean ± SD	mean ± SD	mean ± SD
day 7	**0.0**	49.7 ± 0.8	19.2 ± 0.3	498.1 ± 28.2
	0.08	48.7 ± 1.5	18.9 ± 0.6	466.4 ± 47.4
	0.4	49.4 ± 1.2	19.1 ± 0.4	480.9 ± 30.6
	2.0	48.2 ± 0.4	18.6 ± 0.3	460.5 ± 12.1
	10	48.6 ± 1.1	18.7 ± 0.4	467.6 ± 38.4
day 21	**0.0**	66.2 ± 2.4	23.8 ± 0.76	1205.3 ± 75.8
	0.08	66.4 ± 2.7	23.4 ± 1.06	1163.9 ± 95.2
	0.4	68.1 ± 1.9	24.1 ± 0.55	1238.4 ± 93.4
	2.0	**69.9 ± 0.7**	24.8 ± 0.69	1315.7 ± 66.4
	10	**71.9 ± 0.8**	**25.8 ± 0.92**	**1520.9 ± 90.2**

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

3.2.5 Effects of E2 on Developmental Endpoints

At the concentrations used in this study, E2 treatment caused only mild effects on tadpole development. When the standard Jonckheere-Terpstra test was used to test for significant trends of median stages (n=4 replicate tanks), no significant differences between E2 treatments and the control group were detected on day 7 or day 21. However, as noted above (section 3.2.4), tadpoles treated with the two highest E2 concentrations showed signs of a slight developmental delay as a lower number of tadpoles in the high E2 treatment groups did show development to climax stages >60. The mild developmental delay was also evident from the analyses of HLL on day 21 (but not on day 7). Compared to the control group, mean HLL was significantly lower following 21-day exposure of tadpoles to 2.0 and 10 µg/l E2 (Table 8). This statistically significant difference was observed when the analysis was based on absolute HLL values or on HLL values after normalization for WBL (HLL-W) and SVL (HLL-S). As shown in Table 8, the inhibitory effect of E2 treatment on absolute HLL was not strictly monotonic. However, when HLL data were normalized by WBL or SVL prior to statistical analyses, a concentration-response monotonicity became apparent.

Table 8. Results from measurements of hind limb length (HLL), on day 7 and 21 in the exposure experiment with 17β-estradiol (E2). Data are presented as absolute HLL (mm), and after normalization of HLL by WBL (HLL-W) or SVL (HLL-S).

	E2 µg/l	HLL mean ± SD	HLL-W mean ± SD	HLL-S mean ± SD
day 7	**0.0**	4.26 ± 0.12	0.09 ± 0.002	0.22 ± 0.006
	0.08	4.26 ± 0.22	0.09 ± 0.003	0.23 ± 0.007
	0.4	4.31 ± 0.11	0.09 ± 0.002	0.23 ± 0.007
	2.0	4.10 ± 0.10	0.08 ± 0.002	0.22 ± 0.006
	10	4.22 ± 0.09	0.09 ± 0.003	0.22 ± 0.010
day 21	**0.0**	19.4 ± 0.3	0.31 ± 0.02	0.84 ± 0.03
	0.08	19.0 ± 1.1	0.30 ± 0.03	0.83 ± 0.08
	0.4	18.6 ± 0.5	0.28 ± 0.01	0.79 ± 0.01
	2.0	**17.6 ± 0.9**	**0.26 ± 0.02**	**0.73 ± 0.03**
	10	**18.1 ± 1.2**	**0.26 ± 0.01**	**0.72 ± 0.04**

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

Table 9. Results from developmental stage determination in different treatment groups of the E2 experiment on study day 7. No significant trends of median stages (n=4 replicate tanks) were detected using the standard Jonckheere-Terpstra test.

E2 (µg/l)	Rep.[a]	*n*[b]	developmental stage				stage (median)	stage (range)
			53	54	55	56		
0	A	5		1	4		55	2
	B	5		1	4		55	2
	C	5		1	4		55	2
	D	5		1	4		55	2
	Pool	**20**		4	16		55	2
0.08	A	5			5		55	1
	B	5		2	3		55	2
	C	5		1	4		55	2
	D	5			5		55	1
	Pool	**20**		3	17		55	2
0.4	A	5		1	4		55	2
	B	5			5		55	1
	C	5		2	3		55	2
	D	5		1	4		55	2
	Pool	**20**		4	16		55	2
2.0	A	5		1	4		55	2
	B	5		1	4		55	2
	C	5		1	4		55	2
	D	5		2	3		55	2
	Pool	**20**		5	15		55	2
10	A	5		1	4		55	2
	B	5		1	4		55	2
	C	5		2	3		55	2
	D	5			5		55	1
	Pool	**20**		4	16		55	2

Note: [a] Stage data are presented for individual replicate tanks (A, B, C, and D) and after pooling of stage data from all four replicate tanks (pool). [b] The number of tadpoles examined (n) and the number of tadpoles at each specific developmental stage are shown.

Table 10. Results from developmental stage determination in different treatment groups of the E2 experiment on study day 21. No significant trends of median stages (n=4 replicate tanks) were detected using the standard Jonckheere-Terpstra test. Note the reduced number of tadpoles at stages >60 at the highest E2 concentration.

E2			developmental stage									stage	stage
(µg/l)	Rep.[a]	*n*[b]	56	57	58	59	60	61	62	63	64	(median)	(range)
0	A	15			2	3	3	3	4			60	5
	B	15				3	5	2	4	1		60	5
	C	15			1	2	5	2	4	1		60	6
	D	15				2	4	4	5			61	5
	Pool	**60**			3	10	17	11	17	2		60.5	6
0.08	A	15			2	5	2	3	3			60	5
	B	15				3	2	3	6	1		61	5
	C	15		1	1	4	2	1	6			60	6
	D	15				2	4	3	5	1		61	5
	Pool	**60**		1	3	14	10	10	20	2		61	7
0.4	A	15				7	3	2	3			60	4
	B	15		2	2	2	3	2	4			60	6
	C	15		1	2	3	1	3	5			61	6
	D	15		1	2	4	3	2	3			60	6
	Pool	**60**		4	6	16	10	9	15			60.3	6
2.0	A	15			2	5	4	2	2			60	5
	B	15		1	4	4	2		4			59	6
	C	15			2	3	5	2	2	1		60	6
	D	15			4	4	2	2	3			59	5
	Pool	**60**		1	12	16	13	6	11	1		59.5	7
10	A	15		1	2	3	7		1	1		60	7
	B	15			3	2	7	1	2			60	5
	C	15		1		6	6	1	1			60	6
	D	15				6	5	4				60	3
	Pool	**60**		2	5	17	25	6	4	1		60.0	7

Note: [a] Stage data are presented for individual replicate tanks (A, B, C, and D) and after pooling of stage data from all four replicate tanks (pool). [b] The number of tadpoles examined (n) and the number of tadpoles at each specific developmental stage are shown.

3.2.6 Effects of E2 on Thyroid Gland Histology

The study protocol called for evaluation of thyroid gland histology in five specimen 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. In the E2 experiment, the developmental stages of control tadpoles sampled for thyroid histopathology ranged between stages 59 to 62. In accordance with the developmental stages (early climax) of control animals, the light histological appearance of thyroid glands from the control group indicated an active state of the thyroid. Thyroid follicles were composed of cuboidal or columnar epithelial cells and the follicular lumen was filled with colloid. In a number of control tadpoles, mild to moderate hypertrophy of the follicular epithelium was evident. Peripheral vacuolation of the colloid was mild to moderate in most control thyroids. However, moderately severe colloid depletion was also present in hypertrophic follicles of some animals. Overall, the mild to moderate stimulation of thyroid activity in the control animals is in agreement with the peaking activity of the pituitary-thyroid axis around climax stages.

Analysis of thyroid tissue from the various E2 treatments did not reveal remarkable effects of E2 on thyroid size. The follicular architecture of the thyroids from E2-treated tadpoles did also not deviate from the control group. Minor differences between control and E2-treated animals were only noted with regard to the incidence and severity of follicular cell hypertrophy. At the two highest E2 concentrations, stimulation of follicular epithelium appeared less prominent compared to the control and the lower E2 concentrations. At 10 µg/l E2, most thyroid follicles were composed of cuboidal epithelial cells and the mild to moderate follicular cell hypertrophy observed in the majority of control samples was not as prevalent in this treatment group. A semi-quantitative analysis of the histological alterations was performed using a severity grading approach to assess the incidence and severity of three core diagnostic parameters including thyroid gland size, follicular cell hypertrophy and follicular cell hyperplasia. In order to account for the rather mild differences observed during the initial qualitative assessment, an additional grade of 0.5 was introduced into the analysis of follicular cell hypertrophy. A grade of 0.5 was defined as a small increase in epithelial cell height (cuboidal cell shape) that was distinct from the flat cell shape (grade 0) and the low columnar cell shape (grade 1). Results from these analyses are summarized in Table 11. Severity scores for follicular cell hypertrophy tended to decrease at high E2 concentrations but no statistically significant differences were detected for mean severity scores (n= 4 replicate tanks) between E2 treatments and the control group.

Table 11. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to E2. No statistically significant differences were detected for mean severity scores (n= 4 replicate tanks) between E2 treatments and the control group (p>0.05, parametric comparison to control using Dunnett's test)**.**

E2	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	3/20	0.15±0.10 (1)	1/20	0.05±0.10 (1)	15/20	0.53 ± 0.30 (2)	0/20	0.00±0.00 (0)	59-62
0.08	3/20	0.15±0.10 (1)	2/20	0.10±0.12 (1)	13/20	0.53 ± 0.34 (2)	0/20	0.00±0.00 (0)	58-62
0.4	2/20	0.10±0.20 (1)	2/20	0.10±0.12 (1)	12/20	0.48 ± 0.17 (2)	0/20	0.00±0.00 (0)	57-62
2.0	4/20	0.20±0.16 (1)	2/20	0.10±0.12 (1)	11/20	0.38 ± 0.24 (1)	0/20	0.00±0.00 (0)	57-62
10	4/20	0.20±0.16 (1)	3/20	0.15±0.19 (1)	9/19	0.27 ± 0.12 (1)	0/20	0.00±0.00 (0)	57-62

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

[b] values represent the mean ± standard deviation of severity scores (n=4 replicate tanks), the highest severity grade observed in each treatment group is given in parentheses

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

3.3 Test Results – Benzophenone-2 (BP-2)

3.3.1 Overview

During the validation Phase-3 study of the Amphibian Metamorphosis Assay, two exposure experiments with benzophenone-2 (BP-2) as test substance were performed. In both exposure studies, the nominal aqueous concentrations for BP-2 were 0, 0.75, 1.5, 3.0, and 6.0 mg/l. Tadpoles treated with the highest BP-2 concentration showed signs of weak systemic toxicity including growth retardation (lab 1) and slightly reduced survival (lab 2). Results from stage determination and HLL measurements showed BP-2 treatment to cause developmental delay in both experiments. Histopathological assessment of thyroid tissue revealed remarkable effects of the highest BP-2 concentration on the thyroid gland and provided strong evidence for disruption of the thyroid system by BP-2. Gross morphological assessment of gonads showed abnormal testicular development indicative of estrogenic activity of BP-2.

3.3.2 Verification of Aqueous BP-2 Concentrations

The study protocol called for verification of aqueous BP-2 concentrations in each treatment tank at least once per week. Actual BP-2 concentrations in test tanks were analyzed using reverse-phase HPLC and LC/MS methods during BP-2 exposure experiments. The actual test concentrations of BP-2 as determined by HPLC are shown in Table 12 (lab 1) and by LC/MS in Table 13 (lab 2). In lab 1, mean measured BP-2 concentrations in replicate tanks through time ranged from 0.72 to 0.79 mg/l BP-2 for the 0.75 mg/l treatment group, from 1.47 to 1.59 mg/l BP-2 for the 1.5 mg/l treatment group, from 2.77 to 3.09 mg/l BP-2 for the 3.0 mg/l treatment group, and from 5.53 to 5.85 mg/l BP-2 for the 6.0 mg/l treatment group. In lab 2, mean measured BP-2 concentrations in replicate tanks through time ranged from 0.61 to 0.62 mg/l BP-2 for the 0.75 mg/l treatment group, from 0.98 to 1.19 mg/l BP-2 for the 1.5 mg/l treatment group, from 2.31 to 2.41 mg/l BP-2 for the 3.0 mg/l treatment group, and from 4.54 to 5.01 mg/l BP-2 for the 6.0 mg/l treatment group.

3.3.3 Mortality

Mortality of tadpoles was assessed daily throughout the exposure phase. No mortality was observed in any treatment group in the experiment conducted in lab 1. During the BP-2 study in lab 2, one and three tadpoles died in the 1.5 and 6.0 mg/L BP-2 treatment groups, respectively.

Table 12. BP-2 concentrations (mg/l) in individual exposure tanks as determined by HPLC in lab 1.

nom. conc.	Rep.	BP-2 concentrations measured by HPLC [a]				
		day 0	day 3	day 10	day 16	mean ± SD[b]
0.0	A	n.d.	n.d.	n.d.	n.d.	n.d.
	B	n.d.	n.d.	n.d.	n.d.	n.d.
	C	n.d.	n.d.	n.d.	n.d.	n.d.
	D	n.d.	n.d.	n.d.	n.d.	n.d.
0.75	A	0.75 (100)	0.72 (96)	0.68 (91)	0.74 (99)	0.72 ± 0.03
	B	0.78 (104)	0.79 (105)	0.78 (104)	0.80 (107)	0.79 ± 0.01
	C	0.86 (115)	0.75 (100)	0.71 (95)	0.73 (97)	0.76 ± 0.07
	D	0.83 (111)	0.73 (97)	0.73 (97)	0.74 (99)	0.76 ± 0.05
1.5	A	1.70 (113)	1.40 (93)	1.55 (103)	1.46 (97)	1.53 ± 0.13
	B	1.67 (111)	1.50 (100)	1.52 (101)	1.34 (89)	1.51 ± 0.14
	C	1.63 (109)	1.48 (99)	1.45 (97)	1.31 (87)	1.47 ± 0.13
	D	1.57 (105)	2.08 (138)	1.49 (99)	1.23 (82)	1.59 ± 0.35
3.0	A	3.03 (101)	2.78 (93)	3.06 (102)	2.63 (88)	2.88 ± 0.21
	B	2.99 (100)	2.74 (91)	2.85 (95)	2.49 (83)	2.77 ± 0.21
	C	3.19 (106)	3.92 (131)	2.71 (90)	2.54 (85)	3.09 ± 0.62
	D	2.94 (98)	2.94 (98)	2.94 (98)	2.56 (85)	2.84 ± 0.19
6.0	A	6.51 (109)	5.37 (90)	5.89 (98)	4.97 (83)	5.69 ± 0.67
	B	6.19 (103)	5.49 (92)	5.77 (96)	5.96 (99)	5.85 ± 0.29
	C	5.74 (96)	5.82 (97)	5.00 (83)	5.54 (92)	5.53 ± 0.37
	D	5.69 (95)	5.79 (97)	5.45 (91)	5.23 (87)	5.54 ± 0.25

Note: [a] BP-2 concentrations shown are the average of duplicate measurements. The percentage of the corresponding nominal BP-2 concentration is presented for all measurements in parentheses. [b] Mean measured BP-2 concentrations in individual replicate tanks through time. n.d. not detected.

Table 13. BP-2 concentrations (mg/l) in individual exposure tanks as determined by LC/MS in lab 2.

nom. conc.	Rep.	BP-2 concentrations measured by HPLC[a]				
		day 1	day 9	day 15	day 21	mean ± SD[b]
0.0	A	n.d.	n.d.	n.d.	n.d.	n.d.
	B	n.d.	n.d.	n.d.	n.d.	n.d.
	C	n.d.	n.d.	n.d.	n.d.	n.d.
	D	n.d.	n.d.	n.d.	n.d.	n.d.
0.75	A	0.64 (85)	0.63 (84)	0.56 (75)	0.60 (80)	0.61 ± 0.03
	B	0.66 (88)	0.62 (83)	0.58 (77)	0.61 (81)	0.62 ± 0.03
	C	0.64 (85)	0.62 (83)	0.55 (73)	0.64 (85)	0.61 ± 0.04
	D	0.66 (88)	0.63 (84)	0.56 (73)	0.64 (85)	0.62 ± 0.04
1.5	A	1.26 (84)	1.24 (83)	1.12 (75)	1.16 (77)	1.19 ± 0.07
	B	1.25 (83)	1.14 (76)	1.15 (77)	1.20 (80)	1.18 ± 0.05
	C	1.02 (68)	0.93 (62)	1.00 (67)	0.97 (65)	0.98 ± 0.04
	D	1.04 (69)	0.94 (63)	1.01 (67)	0.95 (63)	0.98 ± 0.05
3.0	A	2.52 (84)	2.47 (82)	2.29 (76)	2.32 (77)	2.40 ± 0.11
	B	2.50 (83)	2.37 (79)	2.36 (79)	2.39 (80)	2.41 ± 0.07
	C	2.49 (83)	2.33 (78)	2.29 (76)	2.39 (80)	2.38 ± 0.09
	D	2.40 (80)	2.29 (76)	2.22 (74)	2.33 (78)	2.31 ± 0.07
6.0	A	5.10 (85)	5.06 (84)	4.98 (83)	4.90 (82)	5.01 ± 0.09
	B	5.07 (85)	4.63 (77)	5.03 (84)	4.62 (77)	4.84 ± 0.25
	C	4.73 (79)	4.40 (73)	4.55 (76)	4.46 (74)	4.54 ± 0.15
	D	4.79 (80)	4.60 (77)	4.78 (80)	4.43 (74)	4.65 ± 0.17

Note: [a] BP-2 concentrations shown are the average of duplicate measurements. The percentage of the corresponding nominal BP-2 concentration is presented for all measurements in parentheses. [b] Mean measured BP-2 concentrations in individual replicate tanks through time. n.d. not detected.

3.3.4 Effects of BP-2 on Growth-related Endpoints

Possible effects of BP-2 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. In both exposure experiments with BP-2, analysis of growth-related parameters on day 7 did not reveal effects by BP-2 treatment on tadpole growth. However, by day 21, a slight reduction in WBL, SVL and wet weight was statistically significant for the 6.0 mg/l BP-2 treatment group in the BP-2 study in lab 1 (Table 14).

Results from day 21 measurements in the BP-2 study in lab 2 showed non-monotonic responses of growth-related parameters to BP-2 treatment. Treatment of tadpoles with the intermediate concentration of 1.5 mg/l BP-2 resulted in significantly decreased mean values for WBL and wet weight ($p<0.05$; parametric comparison to control using Dunnett/Tamhane-Dunnett test) but not SVL (Table 15). No statistically significant differences were detected for the 3.0 mg/l BP-2 treatment. Treatment with the highest BP-2 concentration (6.0 mg/l), however, resulted in significant increases of mean values for SVL, WBL and wet weight ($p<0.05$; parametric comparison to control using Dunnett/Tamhane-Dunnett test).

Table 14. Results from measurements of whole body length (WBL, mm), snout-vent length (SVL, mm), and wet weight (mg) on day 7 and 21 in the exposure experiment with BP-2 in lab 1.

	BP-2 mg/l	WBL mean ± SD	SVL mean ± SD	wet weight mean ± SD
day 7	**0.0**	44.2 ± 0.6	17.0 ± 0.2	350.3 ± 33.0
	0.75	44.5 ± 1.0	16.6 ± 0.5	338.3 ± 32.0
	1.5	44.0 ± 1.2	17.0 ± 0.7	338.9 ± 21.7
	3.0	44.8 ± 0.5	17.3 ± 0.1	355.9 ± 7.8
	6.0	44.4 ± 1.0	16.8 ± 0.4	335.6 ± 22.6
day 21	**0.0**	71.3 ± 0.8	25.8 ± 0.4	1310.2 ± 68.8
	0.75	72.0 ± 0.9	26.6 ± 0.7	1391.5 ± 95.5
	1.5	71.2 ± 2.3	25.9 ± 1.0	1288.6 ± 158.8
	3.0	72.0 ± 0.3	26.1 ± 0.2	1349.6 ± 34.0
	6.0	**66.8 ± 0.9**	**23.9 ± 0.5**	**1027.8 ± 46.0**

Note: Mean values and standard deviations are shown for each treatment group (*n*=4 replicate tanks). Bold letters indicate statistically significant differences to the control group ($p<0.05$, parametric comparison to control using Dunnett/Tamhane-Dunnett test). Note that deviations from concentration-response monotonicity precluded the use of the Jonckheere-Terpstra step-down procedure.

The interpretation of the results of the study in lab 2 is, however, confounded by the fact that a large number of tadpoles in all treatment groups, with the exception of the 6.0 mg/L BP-2 group, already showed development to climax stages (which is associated with weight loss and a reduction in body size). In the BP-2 experiment in lab 2, the mean percentage of tadpoles at stages >60 were 48%, 53%, 59%, 37% and 3.5% for the control and the 0.75, 1.5, 3.0 and 6.0 mg/l BP-2 treatments, respectively. The higher mean values of WBL, SVL and wet weight at the highest BP-2 concentrations are, therefore, likely the result of the lower number of climax stage tadpoles observed in these treatment groups on day 21. Overall, the high number of climax stage tadpoles observed throughout the treatment groups on day 21 of the BP-2 experiment in lab 2 prevents a sound analysis of day 21 growth-related measurements.

Table 15. Results from measurements of whole body length (WBL, mm), snout-vent length (SVL, mm), and wet weight (mg) on day 7 and 21 in the exposure experiment with BP-2 in lab 2.

	BP-2	**WBL**	**SVL**	**wet weight**
	mg/l	mean ± SD	mean ± SD	mean ± SD
day 7	**0.0**	53.9 ± 0.6	20.4 ± 0.1	651.1 ± 30.0
	0.75	55.1 ± 1.2	**21.4 ± 0.6**	728.6 ± 54.5
	1.5	53.6 ± 1.7	20.8 ± 0.8	709.7 ± 63.0
	3.0	54.0 ± 0.9	21.2 ± 0.2	715.6 ± 19.9
	6.0	51.9 ± 1.5	20.5 ± 0.6	669.9 ± 64.5
day 21	**0.0**	77.1 ± 2.0	25.4 ± 0.7	1863.7 ± 110.9
	0.75	75.9 ± 2.6	25.1 ± 0.6	1758.1 ± 87.1
	1.5	**70.7 ± 1.4**	23.6 ± 0.7	**1643.6 ± 129.3**
	3.0	75.8 ± 3.9	25.9 ± 1.7	1773.4 ± 141.2
	6.0	**85.3 ± 1.3**	**30.4 ± 0.8**	**2125.5 ± 87.5**

Note: Mean values and standard deviations are shown for each treatment group (n=4 replicate tanks). Bold letters indicate statistically significant differences to the control group ($p<0.05$, parametric comparison to control using Dunnett/Tamhane-Dunnett test). Note that deviations from concentration-response monotonicity precluded the use of the Jonckheere-Terpstra step-down procedure.

3.3.5 Effects of BP-2 on Developmental Endpoints

In both exposure studies with BP-2, analyses of stage development on study day 7 using the standard Jonckheere-Terpstra test did not reveal significant trends of median stages (n=4 replicate tanks) between BP-2 treatments and the control group (Table 16, Table 18). Treatment with 6.0 mg/l BP-2 resulted in significant effects on tadpole development on day 21 of both BP-2 experiments. The severity of the developmental delay caused by BP-2 was, however, different in the two studies performed during validation Phase-3.

In the lab 1 study, tadpoles treated with the highest BP-2 concentrations showed a mild but statistically significant delay of development (Table 17). Concurrent analyses of treatment-related effects on hind limb growth were performed using absolute HLL values and HLL values after normalization for WBL (HLL-W) and SVL (HLL-S). When absolute HLL values were analysed for treatment-related differences, a non-monotonic concentration-response relationship was observed (Table 20). Pairwise comparison of BP-2 treatment to the control group showed significantly increased HLL values for the 0.75 and the 3.0 mg/l BP-2 treatment groups ($p<0.05$; parametric comparison to control using Dunnett/Tamhane-Dunnett test), whereas no significant differences were detected for the 1.5 and 6.0 mg/l BP-2 treatment groups. Following normalization for WBL, the increased HLL at 0.75 mg/l BP-2 was still significant ($p<0.05$; parametric comparison to control using Dunnett/Tamhane-Dunnett test), whereas no differences between the BP-2 treatments and the control group were detected after normalization of HLL by SVL (Table 20).

In the exposure study of lab 2, treatment of tadpoles with 6.0 mg/l BP-2 caused a much more marked inhibition of development. In this study, many tadpoles exposed to 6.0 mg/l BP-2 appeared arrested in early prometamorphic stages indicating a severe inhibition of development (Table 19). The developmental delay caused by BP-2 was also evident from significant reductions in hind limb growth (Table 21). Analyses of treatment-related effects on hind limb growth were performed using absolute HLL values and HLL values after normalization for WBL (HLL-W) and SVL (HLL-S). When absolute HLL values were analysed for treatment-related differences, significantly reduced HLL was detected at the two highest BP-2 concentrations (3.0 and 6.0 mg/l) on day 7 as well as on day 21 (Table 21). Following normalization of HLL by WBL, significantly reduced HLL-W was detected only at the highest BP-2 concentration on day 7 and on day 21. Following normalization of HLL by

SVL, significantly reduced HLL-S was detected at all BP-2 concentrations on day 7 but only at the highest BP-2 concentration on day 21 (Table 21).

Table 16. Results from day 7 developmental stage determination in different treatment groups of the BP-2 experiment in lab 1. No significant trends of median stages (n=4 replicate tanks) were detected using the standard Jonckheere-Terpstra test.

BP-2			**developmental stage**				**stage**	**stage**
(mg/l)	**Rep.**[a]	***n***[b]	**53**	**54**	**55**	**56**	**(median)**	**(range)**
0	A	5		5			54	1
	B	5		5			54	1
	C	5	1	4			54	2
	D	5		5			54	1
	Pool	**20**	**1**	19			54	2
0.75	A	5		5			54	1
	B	5	1	4			54	2
	C	5	1	4			54	2
	D	5		3	2		54	2
	Pool	**20**	2	16	2		54	3
1.5	A	5	1	4			54	2
	B	5		5			54	1
	C	5		5			54	1
	D	5	1	4			54	2
	Pool	**20**	2	18			54	2
3.0	A	5		5			54	1
	B	5		5			54	1
	C	5		5			54	1
	D	5	1	4			54	2
	Pool	**20**	1	19			54	2
6.0	A	5	1	4			54	2
	B	5	1	4			54	2
	C	5	1	4			54	2
	D	5		5			54	1
	Pool	**20**	3	17			54	2

Note: [a] Stage data are presented for individual replicate tanks (A, B, C, and D) and after pooling of stage data from all four replicate tanks (pool). [b] The number of tadpoles examined (n) and the number of tadpoles at each specific developmental stage are shown.

Table 17. Results from day 21 developmental stage determination in different treatment groups of the BP-2 experiment in lab 1. A significant effect on median stages was detected at 6.0 mg/l BP-2 using the standard Jonckheere-Terpstra test (n=4 replicates).

BP-2			**developmental stage**									**stage**	**stage**
(mg/l)	**Rep.[a]**	***n*[b]**	**55**	**56**	**57**	**58**	**59**	**60**	**61**	**62**	**63**	**(median)**	**(range)**
0	A	15			3	11	1					58	3
	B	15			7	6	2					58	3
	C	15			7	4	4					58	3
	D	15			6	5	3	1				58	4
	Pool	**60**			23	26	10	1				58	4
0.75	A	15			1	10	4					58	3
	B	15			5	7	2	1				58	4
	C	15			3	7	4	1				58	4
	D	15			3	8	3	1				58	4
	Pool	**60**			12	32	13	1				58	4
1.5	A	15		1	4	7	3					58	4
	B	15			5	8	2					58	3
	C	15		1	4	8	2					58	4
	D	15			7	6	1		1			58	5
	Pool	**60**		**2**	20	29	8		1			58	6
3.0	A	15			7	5	2		1			58	5
	B	15			2	8	5					58	3
	C	15			4	7	3	1				58	4
	D	15			4	5	4		2			58	5
	Pool	**60**			17	25	14	1	3			58	5
6.0	A	15			4	7	1	1	2			58	5
	B	15			8	5	1		1			57	5
	C	15			8	5	2					57	3
	D	15			8	3	3	1				57	4
	Pool	**60**			28	20	7	2	3			58	5

Note: [a] Stage data are presented for individual replicate tanks (A, B, C, and D) and after pooling of stage data from all four replicate tanks (pool). [b] The number of tadpoles examined (n) and the number of tadpoles at each specific developmental stage are shown.

Table 18. Results from day 7 developmental stage determination in different treatment groups of the BP-2 experiment in lab 2. No significant trends of median stages (n=4 replicate tanks) were detected using the standard Jonckheere-Terpstra test.

BP-2 (mg/l)	Rep.[a]	*n*[b]	developmental stage				stage (median)	stage (range)
			53	54	55	56		
0	A	5		1	4		55	2
	B	5		1	4		55	2
	C	5			5		55	1
	D	5			5		55	1
	Pool	**20**		2	18		55	2
0.75	A	5		1	4		55	2
	B	5			5		55	1
	C	5		1	4		55	2
	D	5			5		55	1
	Pool	**20**		2	18		55	2
1.5	A	5		1	4		55	2
	B	5		1	4		55	2
	C	5			5		55	1
	D	5			5		55	1
	Pool	**20**		2	18		55	2
3.0	A	5		1	4		55	2
	B	5			5		55	1
	C	5			5		55	1
	D	5			5		55	1
	Pool	**20**		1	19		55	2
6.0	A	5		1	4		55	2
	B	5		1	4		55	2
	C	5			5		55	1
	D	5		1	4		55	2
	Pool	**20**		3	17		55	2

Note: [a] Stage data are presented for individual replicate tanks (A, B, C, and D) and after pooling of data from all four replicate tanks (pool). [b] The number of tadpoles examined (n) and the number of tadpoles at each specific developmental stage are shown.

Table 19. Results from day 21 developmental stage determination in different treatment groups of the BP-2 experiment in lab 2. A significant effect on median stages was detected at 6.0 mg/l BP-2 using the standard Jonckheere-Terpstra test (n=4 replicates).

BP-2			developmental stage									stage	stage
(mg/l)	Rep.[a]	*n*[b]	55	56	57	58	59	60	61	62	63	(median)	(range)
0	A	15					1	6	3	4	1	61	5
	B	15					1	8	3	2	1	60	5
	C	15						6	4	3	2	61	4
	D	15					1	8	5	1		60	4
	Pool	**60**					**3**	**28**	**15**	**10**	**4**	**60**	**5**
0.75	A	15					1	6	7	1		61	4
	B	15					2	8		4	1	60	5
	C	15					2	3	6	4		61	4
	D	15						6	6	3		61	3
	Pool	**60**					5	23	19	12	1	61	5
1.5	A	15						5	3	7		61	3
	B	15					1	5	3	5	1	61	5
	C	15					1	5	2	6	1	61	5
	D	14					1	6	3	2	2	60.5	5
	Pool	**59**					3	21	11	20	4	61	5
3.0	A	15					3	10	2			60	3
	B	15					1	5	4	4	1	61	5
	C	15					4	7	3	1		60	4
	D	15					2	6	2	4	1	60	5
	Pool	**60**					10	28	11	9	2	60	5
6.0	A	15	1	2	1	5	2	3	1			58	7
	B	14	2	1	5	5		1				57	6
	C	14	2	2	3	1	4	2				57.5	6
	D	14	1	1		3	3	5	1			59	7
	Pool	**57**	**6**	6	9	14	9	11	2			58	7

Note: [a] Stage data are presented for individual replicate tanks (A, B, C, and D) and after pooling of data from all four replicate tanks (pool). [b] The number of tadpoles examined (n) and the number of tadpoles at each specific developmental stage are shown.

Table 20. Results from measurements of hind limb length (HLL), on day 7 and 21 in the exposure experiment with BP-2 in lab 1. Data are presented as absolute HLL (mm), and after normalization of HLL by WBL (HLL-W) and SVL (HLL-S).

	BP-2	HLL	HLL-W	HLL-S
	mg/l	mean ± SD	mean ± SD	mean ± SD
day 7	**0.0**	2.08 ± 0.03	0.047 ± 0.001	0.123 ± 0.003
	0.75	2.19 ± 0.19	0.049 ± 0.004	0.132 ± 0.011
	1.5	2.04 ± 0.13	0.046 ± 0.002	0.120 ± 0.004
	3.0	2.08 ± 0.05	0.046 ± 0.001	0.120 ± 0.003
	6.0	2.07 ± 0.11	0.047 ± 0.002	0.123 ± 0.007
day 21	**0.0**	11.19 ± 0.43	0.156 ± 0.006	0.432 ± 0.023
	0.75	**12.22 ± 0.53**	**0.169 ± 0.008**	0.459 ± 0.024
	1.5	11.12 ± 0.26	0.155 ± 0.002	0.427 ± 0.010
	3.0	**12.04 ± 0.39**	0.167 ± 0.006	0.464 ± 0.014
	6.0	11.11 ± 0.53	0.165 ± 0.010	0.464 ± 0.034

Note: Mean values and standard deviations are shown for each treatment group (n=4 replicate tanks). Bold letters indicate statistically significant differences to the control group ($p<0.05$, parametric comparison to control using Dunnett/Tamhane-Dunnett test). Note that deviations from concentration-response monotonicity precluded the use of the Jonckheere-Terpstra step-down procedure.

Table 21. Results from measurements of hind limb length (HLL), on day 7 and 21 in the exposure experiment with BP-2 in lab 2. Data are presented as absolute HLL (mm), and after normalization of HLL by WBL (HLL-W) and SVL (HLL-S).

	BP-2	HLL	HLL-W	HLL-S
	mg/l	mean ± SD	mean ± SD	mean ± SD
day 7	**0.0**	3.38 ± 0.07	0.063 ± 0.001	0.166 ± 0.003
	0.75	3.38 ± 0.10	0.061 ± 0.001	**0.158 ± 0.002**
	1.5	3.28 ± 0.08	0.061 ± 0.002	**0.157 ± 0.005**
	3.0	**3.29 ± 0.12**	0.061 ± 0.002	**0.156 ± 0.005**
	6.0	**2.94 ± 0.11**	**0.057 ± 0.002**	**0.144 ± 0.005**
day 21	**0.0**	25.22 ± 1.04	0.332 ± 0.018	1.010 ± 0.050
	0.75	24.67 ± 0.90	0.329 ± 0.006	1.002 ± 0.033
	1.5	24.59 ± 0.52	0.355 ± 0.011	1.062 ± 0.028
	3.0	**22.55 ± 1.15**	0.303 ± 0.033	0.888 ± 0.105
	6.0	**15.25 ± 3.03**	**0.180 ± 0.039**	**0.514 ± 0.114**

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

3.3.6 Effects of BP-2 on Thyroid Gland Histology

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. In both experiments performed with BP-2, remarkable effects of BP-2 on thyroid histology were observed, albeit the incidence and severity of the observed changes was different between the two exposure studies.

In the BP-2 experiment in lab 1, the developmental stages of control tadpoles sampled for thyroid histopathology ranged between stages 57 to 59. In accordance with the prometamorphic stages of control animals, the light histological appearance of thyroid glands from the control group indicated a mild to moderate activity of the thyroid. Thyroid follicles were composed of cuboidal or low columnar epithelial cells and the follicular lumen was filled with dense colloid. In only a few control tadpoles, mild to moderate hypertrophy of the follicular epithelium was evident. Peripheral vacuolation of the colloid was mild in most control thyroids. Overall, the mild stimulation of thyroid activity in the control animals is in agreement with the increasing activity of the pituitary-thyroid axis at late prometamorphic stages.

Analysis of thyroid tissue from the various BP-2 treatments did not reveal remarkable effects of the tested BP-2 concentrations on thyroid size. The follicular architecture of the thyroids from BP-2-treated tadpoles did not deviate from the control group with the exception of an increased number of moderately hypertrophic follicles in the 6.0 mg/l BP-2 group. At the highest BP-2 concentration, many thyroid follicles were composed of columnar epithelial cells, and this increased stimulation of the follicular epithelium was more prominent at 6.0 mg/l BP-2 compared to the control and the lower BP-2 concentrations. A semi-quantitative analysis of the histological alterations was performed using a severity grading approach to assess the incidence and severity of three core diagnostic parameters including thyroid gland size, follicular cell hypertrophy and follicular cell hyperplasia. In order to account for the mild differences observed during the initial qualitative assessment, an additional grade 0.5 was introduced into the analysis of follicular cell hypertrophy in lab 1. A grade of 0.5 was defined as a small increase in epithelial cell height (cuboidal cell shape) that was distinct from the flat cell shape (grade 0) and the low columnar cell shape (grade 1). Results from these analyses are summarized in Table 22. Mean severity scores (n= 4 replicate tanks) for follicular cell

hypertrophy were significantly increased at the highest BP-2 concentration (6.0 mg/l). Statistically significant differences of mean severity scores were not detected between lower BP-2 treatments and the control group.

In the BP-2 experiment in lab 2, the developmental stages of control tadpoles sampled for thyroid histopathology ranged between stages 59 to 63. In accordance with the advanced developmental stages (early climax) of control animals, the light histological appearance of thyroid glands from the control group indicated an active state of the thyroid. Thyroid follicles were composed of cuboidal or low columnar epithelial cells and the follicular lumen was filled with dense colloid. In only a few control tadpoles, mild to moderate hypertrophy of the follicular epithelium was evident.

Analysis of thyroid tissue from the various BP-2 treatments in the lab 2 study did not reveal remarkable effects of BP-2 concentrations <6.0 mg/l on thyroid size. At the highest BP-2 concentration, however, a marked increase in thyroid gland size was evident in a large number of tadpoles. Thyroid tissue from animals treated with 6.0 mg/l BP-2 was also characterized by a severe hypertrophy of the follicular epithelium and the presence of moderately severe follicular cell hyperplasia. A semi-quantitative analysis of the histological alterations was performed in lab 2 using a severity grading approach to assess the incidence and severity of three core diagnostic parameters including thyroid gland size, follicular cell hypertrophy and follicular cell hyperplasia. Results from these analyses are summarized in Table 23. Mean severity scores (n= 4 replicate tanks) for follicular cell hypertrophy, follicular cell hyperplasia and thyroid gland enlargement were significantly increased at the highest BP-2 concentration (6.0 mg/l). Statistically significant differences of mean severity scores were not detected between lower BP-2 treatments and the control group. Notably, the very severe histopathological effects observed at 6.0 mg/l BP-2 were not preceded by gradually weaker changes at the next lower BP-2 concentration (3.0 mg/l) indicating a very steep concentration-response relationship.

Table 22. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to BP-2 (lab 1 study).

BP-2	glandular hypertrophy		glandular atrophy		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(mg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	3/20	0.15±0.19 (1)	1/20	0.05±0.10 (1)	6/20	0,20 ± 0,12 (1)	0/20	0.00±0.00 (0)	57 – 59
0.75	6/20	0.35±0.34 (2)	2/20	0.10±0.12 (1)	13/20	0,40 ± 0,16 (1)	0/20	0.00±0.00 (0)	57 – 60
1.5	2/20	0.10±0.12 (1)	2/20	0.10±0.12 (1)	8/20	0,25 ± 0,06 (1)	0/20	0.00±0.00 (0)	57 – 61
3.0	4/20	0.20±0.16 (1)	1/20	0.05±0.10 (1)	13/20	0,40 ± 0,14 (1)	0/20	0.00±0.00 (0)	57 – 59
6.0	5/20	0.35±0.34 (2)	2/20	0.10±0.12 (1)	12/19	**0,68 ± 0,17 (2)**	0/20	0.00±0.00 (0)	57 - 61

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

Bold letters indicate statistically significant differences of mean severity scores (n=4 replicates) to the control group ($p<0.05$, parametric comparison to control using Dunnett's test).

Table 23. Incidence and severity of selected histological alterations in *X. laevis* tadpoles exposed to BP-2 (lab 2 study).

BP-2	glandular hypertrophy		glandular atrophy		follicular cell hypertrophy		follicular cell hyperplasia		stage[c]
(mg/L)	incidence[a]	severity[b]	incidence	severity	incidence	severity	incidence	severity	
0	3/20	0.15±0.19 (1)	0/20	0.00±0.00 (0)	5/20	0.25±0.19 (1)	7/20	0.35±0.25 (1)	59 – 63
0.75	3/20	0.15±0.10 (1)	2/20	0.10±0.20 (1)	7/20	0.35±0.19 (1)	9/20	0.45±0.19 (1)	59 – 63
1.5	4/20	0.20±0.16 (1)	0/20	0.00±0.00 (0)	11/20	0.55±0.25 (1)	10/20	0.50±0.26 (1)	59 – 63
3.0	5/20	0.30±0.26 (2)	0/20	0.00±0.00 (0)	12/20	0.65±0.44 (2)	15/20	0.80±0.43 (2)	59 – 62
6.0	18/20	**2.25±0.6 (3)**	0/20	0.00±0.00 (0)	20/20	**2.35±0.50 (3)**	20/20	**1.90±0.38 (3)**	55 – 61

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

Bold letters indicate statistically significant differences of mean severity scores (n=4 replicates) to the control group ($p<0.05$, parametric comparison to control using Dunnett's test).

4 Discussion

4.1 Purpose of the Assay

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, hind limb growth, and thyroid histology) considered in the current 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.

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) accompanied by morphometric analysis of hind limb morphogenesis (day 7 and day 21). 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. Experimental work during the validation Phase-1 and Phase-2 studies of the Amphibian Metamorphosis Assay have demonstrated 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.

4.2 The Biological Model

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.

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, 1996). In addition, the pituitary-thyroid feedback signalling is already functional at stage 51 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.

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).

Similar to other developmental processes, there is evidence that other hormones can interfere with the TH-dependent regulation of tissue remodelling during amphibian metamorphosis. In this regard, information is available for the influence of corticosteroids, prolactin and gonadal steroids on tadpole development or specific TH-regulated processes in developing tadpoles (reviewed in Denver et al., 2002; Kikuyama et al., 1993; Fort et al., 2007). For example, several studies have shown that high concentrations of corticosterone (CORT) or dexamethasone can provoke both inhibitory or accelerating effects on metamorphic development, depending in part on the timing of treatment. With regard to the inhibitory action of high CORT concentrations on the development of premetamorphic and prometamorphic *X. laevis* tadpoles, a suppressive effect of CORT on the pituitary-thyroid axis has been suggested based on the observation of reduced T4 and T3 plasma levels in CORT-treated tadpoles (Leloup-Hatey et al., 1990). Concerning the accelerating effects of CORT on the development of tadpoles at late stages, two molecular mechanisms have been proposed including CORT-induced increases in T3 receptor binding capacities and CORT-induced increases in peripheral deiodination of T4 to T3 (Shi, 1999).

Knowledge about the interaction of gonadal steroid hormones (estrogens, androgens) with the pituitary-thyroid axis or with TH action is still fragmentary and results from previous studies are sometimes inconsistent (reviewed in Fort et al., 2007; Hayes, 1997). In most experimental

studies investigating the effects of gonadal steroids on thyroid system function of amphibian tadpoles, inhibitory effects were reported at relatively high concentrations of steroidal hormones. However, many studies performed on steroidal interference with TH-dependent metamorphosis used co-exposure treatments of tadpoles or tadpole tissues to gonadal steroids plus TH to demonstrate any effect. In a study by Pickford et al. (2003), time to metamorphosis in *X. laevis* tadpoles was increased due to exposure to 2.7 µg/l E2 indicating an inhibitory effect of E2 on development. This study did, however, not address possible modes of E2 action and thyroid histopathology was performed. Summarizing the available information regarding possible mechanisms of steroid hormone action, Fort et al. (2007) concluded that interference of gonadal steroids with the thyroid system occurs most likely at the hypothalamic-pituitary level.

Another hormone that has received a great deal of attention with regard to modulation of TH-dependent metamorphic development is prolactin. Early studies using mammalian prolactin preparations could demonstrate antagonistic effects of prolactin on TH action in various peripheral tissues (Baker and Tata, 1992). Inhibitory effects on TR autoinduction have been suggested as a primary mechanism of prolactin action to antagonize TH action in peripheral tissues. It should be noted, however, that in a transgenic frog model of prolactin overexpression, no retardation of tadpole development was detectable, with the exception of blocked tail resorption in a limited number of transgene animals (Huang and Brown, 2000).

This brief review of possible hormonal interference with thyroid system function during amphibian metamorphosis should serve to highlight the need for experimental evaluation of the effect pattern caused by endocrine active compounds acting via signalling pathways not directly related to the thyroid system. Given that many endocrine-disrupting chemicals display weak or moderate estrogenic activity, the estrogen E2 has been considered as a suitable model compound to assess the effect pattern of an estrogenic chemical on the various endpoints used in the current protocol for the metamorphosis assay.

4.3 Control Animal Performance

The assessment of control animal performance has been a central aspect in the evaluation of the study protocol during the multichemical interlaboratory study during Phase-2 validation. Results from the Phase-2 exercise suggested that further optimization and standardization of the feeding regime could provide a means to reduce inter-individual variability of tadpole growth and development. Based on the results of the three experiments performed during validation Phase-3, it appears that the recommended feeding schedule with SeraMicron ensures homogenous development of the control animal population. More than 90% of the tadpoles from the control group were distributed across 3 to 4 different developmental stages by study day 21 in each of the three exposure studies. Despite the use of the standardized feeding protocol, comparison of the developmental stages of control animals in the individual experiments revealed that differences still existed between experiments. Specifically, control development in lab 1 was slower in the BP-2 study compared to the E2 study, although general testing conditions were the same in both experiments. The reason for this difference is not known, but clutch-to-clutch differences are most likely responsible for this observation. Notably, within each of the experiments, tadpole development was very homogenous. Thus, the currently used testing protocol efficiently controls for an acceptable inter-individual variability and a low variability between replicate tanks. To this end, the data generated during the Phase-3 validation support the view that the general testing conditions selected for the test protocol of the Amphibian Metamorphosis Assay provide for a high quality of test animals for the conduct of the assay.

4.4 Effects Pattern of E2 in Phase-3 Studies

17β-estradiol (E2) was used in the Phase-3 validation work as a model substance to assess the utility of the study protocol to distinguish thyroid system-related and unrelated mechanisms of endocrine disruption (e.g., estrogenic activity of E2). From the available literature, a sound assessment of the capacity of E2 to disrupt thyroid system function in amphibian species was not possible. In mammals, a large number of different rodent models has been used to assess interference of E2 with hypothalamus-pituitary-thyroid (HPT) axis function. The results from these studies are also often contradictory and clear-cut evidence for molecular mechanisms of E2 action leading to specific changes in thyroid system function is still missing. In various rodent studies reporting altered thyroid system function due to E2 treatment, the effects pattern indicate that E2 might specifically affect pituitary function, for example by altering the responsiveness of thyrotrophic cells to TRH or to negative feedback by circulating TH (Banu et al., 2001; Seidlova-Wuttke et al., 2005; Sosic-Jurjevic et al., 2006). Notably, a number of recent studies failed to detect effects of E2 administration on circulating TSH levels in various rodent protocols (Jarry et al., 2004; Lima et al., 2006; Klammer et al., 2007; Schmutzler et al., 2004; Seidlova-Wuttke et al., 2005).

A single exposure experiment with E2 was conducted during the validation Phase-3. The E2 concentrations used in this experiment were selected based on previous experience about the concentration-response relationship of estrogenic effects by E2 in *X. laevis* tadpoles. A detailed analysis of the estrogenic effects pattern caused by the applied concentrations of E2 in the Phase-3 study is currently in progress. Preliminary results suggest weak estrogenic effects at 0.08 µg/l E2, remarkable estrogenic effects on the liver and the gonads (e.g., disruption of testicular development) at 0.4 µg/l and a complete feminization of tadpoles (male-to-female sex reversal) at 2.0 and 10 µg/l E2. Based on these data, it is assumed that the tadpoles exposed to E2 experienced in fact a weak to strong estrogenic treatment.

In contrast to the strong effects on gonadal development, tadpole growth and TH-dependent development were barely affected at the E2 concentrations used in the Phase-3 study. Analysis of developmental stages on day 7 and day 21 did not reveal significant effects of E2 on stage development. Although the current statistical procedure for the analysis of differences in stage development did not detect significant effects, an inspection of the raw data suggested that, relative to the control, a reduced number of tadpoles had reached climax stages at the two highest E2 concentrations. The slightly reduced hind limb growth observed

by HLL measurements on day 21 for the two highest E2 concentrations is in accordance with a slight delay in development. Tadpole growth was not negatively affected by E2 treatment. The increased mean values of growth-related parameters (WBL, SVL, weight) determined for the 10 µg/l E2 treatment on day 21 are likely a result of a lower number of climax animals experiencing decreases in tadpole size and weight in this group. Taken together, the gross morphological endpoint measurements showed that E2 did not impair tadpole growth and caused only mild signs of developmental delay.

Histopathology of thyroid tissue revealed no remarkable differences between the control group and the various E2 treatments. E2 treatment had no effect on thyroid size and did not affect the overall follicular architecture of the thyroid glands. The only observation suggesting a difference to the control group was a slight reduction in follicular cell height at the highest E2 concentration. However, the currently employed statistical procedure (parametric comparison to control using Dunnett/Tamhane-Dunnett test) did not reveal significant differences in mean severity scores between E2 treatments and the control group. The relevance of the minimal difference in the hypertrophy grade of follicular cells between the control and the 10 µg/l E2 treatment is therefore not clear. Increased hypertrophy in the control group was related with development of tadpoles to climax stages and probably reflected the concurrent increased stimulation of the thyroid by TSH. As noted above, less tadpoles have entered early climax stages in the 10 µg/l E2 group and, therefore, the minimal differences in thyroid follicular cell activity could merely be the consequence of the slight differences in stage distribution of tadpoles sampled for thyroid histopathology.

Combining the results of the various endpoint measurements, the effects pattern of E2 did not indicate an effect on the thyroid system via one of the proposed direct modes of action. The presence of minor effects on developmental parameters in the absence of any stimulatory effect on the thyroid tissue indicates that E2 did not directly affect the TH-producing capacity of the thyroid or the amount of circulating TH. Results from validation Phase-1 and Phase-2 work have clearly demonstrated that such modes of action are sensitively detectable by histopathological assessment of thyroid tissue. The slight inhibitory effect of E2 on tadpole development as indicated by the reduced HLL on study day 21 could therefore be the result of an unspecific activity present at high E2 concentrations. For example, E2 might indirectly affect the function of the thyroid system by impairing the maturation and/or function of the neuroendocrine control circuits at the hypothalamus-pituitary level. As noted by several

authors (Fort et al., 2007; Hayes, 1997), some of the historical data on the interaction of gonadal steroids with tadpole development suggest a potential for steroidal compounds to affect pituitary function in metamorphosing tadpoles. To this end, the combined assessment of the endpoint responses observed in the E2 study does not classify E2 as a thyroid system-disrupting compound.

One of the main reasons to assess the effects pattern of E2 in the amphibian metamorphosis assay was to generate data that eventually should allow an estimation of the possible relationship between estrogenic activity and altered tadpole development and/or thyroid system function. As noted above, preliminary results from ongoing analyses demonstrate that the applied concentrations of E2 caused estrogenic effects in a concentration-dependent manner. Weak and moderate estrogenic activity was detected at 0.08 and 0.4 µg/l E2, respectively. The estrogenic activity associated with these E2 concentrations is considered to cover a range of estrogenic activities resulting from action of weakly and moderately active xenoestrogens. This assumption is based on results from various previous studies showing that several xenoestrogens are capable of disrupting gonadal development in *X. laevis* tadpoles but a complete male-to-female sex reversal has not been reported to date. Given this, the two lower E2 concentrations might be regarded as “xenoestrogenic reference concentrations”. Notably, at these concentrations, no indications for alterations in tadpole development, growth and thyroid system function were observed. These results provide strong confidence that the Amphibian Metamorphosis Assay will be capable of distinguishing thyroid system-related modes of action from the possible estrogenic action exerted by a large number of environmental chemicals.

4.5 Effects Pattern of BP-2 in Phase-3 Studies

Benzophenone-2 (BP-2) was used in the Phase-3 validation work as a model substance to assess the utility of the study protocol for the Amphibian Metamorphosis Assay to detect weakly active thyroid system disrupters. During the preparation of a proposal for validation Phase-3 experimental activities, a literature survey was conducted to identify possible candidate substances that could be classified as weakly active thyroid system disrupters. The analysis of available literature revealed that several UV filter substances including 3-4-methylbenzylidene camphor (MBC; CAS. 36861-47-9), octyl-methoxycinnamate (OMC; CAS. 5466-77-3), and 2,2',4,4'-tetrahydroxybenzophenone (benzophenone-2; CAS. 131-55-5) have recently been shown to cause modest changes in thyroid system function in a number of rodent studies (Jarry et al., 2004; Schlumpf et al., 2004; Seidlova-Wuttke et al., 2005, 2006; Schmutzler et al., 2004). To this end, the criteria used to characterize a candidate compound as a weakly active thyroid system disrupter included the demonstration of altered TH plasma levels (T4 or T3) in the absence of marked changes of plasma TSH levels and/or thyroid histopathological markers. Such effect patterns were considered to reflect the ability of a compound to alter the amount of circulating TH but that effects are too weak to override the compensatory activity of the hypothalamus-pituitary axis. The aforementioned effects pattern was in fact reported for several UV filter substances in rodent studies lasting from five days to three months (Jarry et al., 2004; Schlumpf et al., 2004; Seidlova-Wuttke et al., 2005, 2006; Schmutzler et al., 2004).

In preparation for validation Phase-3, range-finding experiments were conducted with *X. laevis* tadpoles to characterize the sublethal concentration range of various UV filter substances and to develop protocols for their use in aquatic exposure studies. Results from these experiments showed that stock solutions of MBC and OMC were difficult to prepare in the absence of solvents or carriers. However, for two other UV filters, benzophenone-1 and BP-2, protocols for preparation of appropriate stock solutions and chemical analytical verification of aqueous test concentrations were successfully established. Concerning thyroid system-disrupting effects of these benzophenones, the available literature suggested that BP-2 could meet the criteria of a weakly active compound. Oral application of BP-2 (250 and 1000 mg/kg body weight) to ovariectomized rats for five days resulted in reduced plasma levels of T4 and T3 but not in changes of plasma TSH levels (Jarry et al., 2004). In another 5-day study with rats, BP-2 administration (10 – 1000 mg/kg) caused significant reductions in plasma levels of T4, whereas plasma T3 and TSH levels were not significantly affected

(Schlecht et al., 2006). Moderate reductions in plasma T4 and T3 levels (in the absence of changes in plasma TSH levels) were also observed in a three month study with ovariectomized rats (Seidlova-Wuttke et al., 2005). In addition to displaying weak anti-thyroidal effects in various rodent studies, BP-2 has also been shown to act as an estrogenic compound in a variety of *in vitro* and *in vivo* systems (Schlumpf et al., 2004).

Based on these rodent data, it was decided that BP-2 could be an appropriate model compound to evaluate the utility of the current protocol of the Amphibian Metamorphosis Assay to detect weakly active thyroid system disrupters. The BP-2 concentrations proposed for the Phase-3 experiments were derived from initial range-finding experiments with *X. laevis* tadpoles. In these experiments (semi-static exposure system), BP-2 was well tolerated by tadpoles up to a concentration of 6 mg/L, but at BP-2 concentrations of 10 mg/L and higher, general toxicity was observed including increased mortality and abnormal swimming behaviour.

During the validation Phase-3 of the Amphibian Metamorphosis Assay, two exposure experiments were conducted using BP-2 as a test substance using the same nominal concentrations (0, 0.75, 1.5, 3.0 and 6.0 mg/l BP-2). In both experiments, minor effects on tadpole growth and survival were observed at the highest BP-2 concentration suggesting that some systemic toxicity was caused by 6.0 mg/l BP-2. However, the effects were not very marked so that these treatment groups were not excluded from the assessment of BP-2 effects on the thyroid system. Tadpole development was also affected by BP-2 at the highest test concentration resulting in a retardation of stage development (on day 21 in both experiments) and reduced hind limb growth (only in lab 2 study). The developmental delay caused by 6.0 mg/l BP-2 was much more severe in the lab 2 experiment as many tadpoles exposed to 6.0 mg/l BP-2 were developmentally arrested at early prometamorphic stages.

Histopathological assessment of thyroid tissue revealed treatment-related changes in tadpoles exposed to the highest BP-2 concentration in both exposure studies. In the lab 1 experiment, BP-2 treatment resulted in mild to moderate follicular cell hypertrophy but the size of the thyroid was not affected. Similar effects on the thyroid gland (mild to moderate increases in follicular cell hypertrophy) were observed in a preliminary exposure study with BP-2 in lab 1 where tadpoles were exposed to a single concentration of 5.0 mg/l BP-2 for 14 days.

In the lab 2 experiment, the histological effects observed in thyroid tissue of tadpoles exposed to 6.0 mg/l BP-2 were much more severe. In the latter experiment, severe hypertrophy of the follicular epithelium and moderately severe follicular cell hyperplasia were observed. The size of the thyroid was also markedly increased. A remarkable result of the thyroid histopathology evaluation in lab 2 was the steep concentration-response relationship. The very strong effects of BP-2 on the thyroid at a concentration of 6.0 mg/l were not preceded by less severe alterations at a two-fold lower concentration. The reason for this strong increase of BP-2 effects between 3.0 and 6.0 mg/l is unknown.

Combining the results of the various endpoint measurements, the effect patterns of BP-2 were very similar in both exposure studies suggesting that BP-2 acts as a thyroid system disrupter in *X. laevis* tadpoles. Notably, in both experiments with BP-2, the severity of the histopathological effects correlated with the severity of the developmental delay. In lab 1, the mild to moderate follicular cell hypertrophy observed in thyroid tissue at 6.0 mg/l BP-2 correlated well with the observation that developmental delay was only weak (hind limb growth was not affected by BP-2). In the lab 2 study, BP-2 (6.0 mg/l) caused severe alterations in thyroid histology and this finding was in agreement with a strong retardation of tadpole development.

From a mechanistic perspective, the results suggest that BP-2 acts as an anti-thyroidal compound in *X. laevis* tadpoles by reducing the levels of circulating TH and/or by blocking the action of TH leading to an inhibition of TH-dependent processes during metamorphic development. BP-2 activity was likely associated with reduced feedback signalling of TH on the hypothalamus-pituitary level which caused an increased TSH secretion and finally an increased stimulation of thyroid follicular cell activity and proliferation. With regard to a possible mode of BP-2 action on the thyroid system, a recent study by Schmutzler et al. (2007) could show that BP-2 inhibits TPO activity in an *in vitro* assay and causes reductions in plasma T4 and increases in plasma TSH levels *in vivo*. Notably, results from the *in vitro* experiments suggest that BP-2 is a more potent inhibitor of TPO activity than propylthiouracil and methimazole. The latter finding is, however, in contrast with the weak anti-thyroidal activity of BP-2 reported in several other *in vivo* studies. One explanation for this discrepancy could come from the observation by Schmutzler et al. that increasing iodine concentrations could efficiently prevent the inhibitory action of BP-2. The latter finding is important, not only in explaining the different outcome of the various rodent studies performed with BP-2,

but also concerning the apparent differences in the severity of BP-2 effects detected in the two amphibian studies performed during validation Phase-3. The total iodide concentration in the dilution water used in lab 1 (10 µg/l) appears to be higher than in lab 2 (less than 5.0 µg/l, detection limit of current method). A more sensitive technique is currently under development in lab 2 in order to determine the actual iodide levels. If the presence of lower iodide levels can be confirmed for the test system used in lab 2, the stronger effects of BP-2 on tadpole development and thyroid gland histology in lab 2 could be interpreted as the result of a lower protection of thyroid tissue by iodide from the apparent inhibitory effects of BP-2, as suggested by the recent data of Schmutzler et al. (2007). To this end, the results generated during the validation Phase-3 experiments reinforce the need for verification and reporting of iodide concentrations in the test systems of every lab performing the Amphibian Metamorphosis Assay for the detection of thyroid system disrupters.

4.6 Application of the Decision Logic

Based on the results obtained during the validation Phase-2 experimental work, a decision logic has been developed for the Amphibian Metamorphosis Assay in order to provide logical guidance to assist in the conduct and interpretation of the results of the bioassay (see flow chart in Figure 4). The decision logic was presented and discussed at the VMG-eco meeting in Madrid (January 22-24, 2007) and approved for use in validation Phase-3.

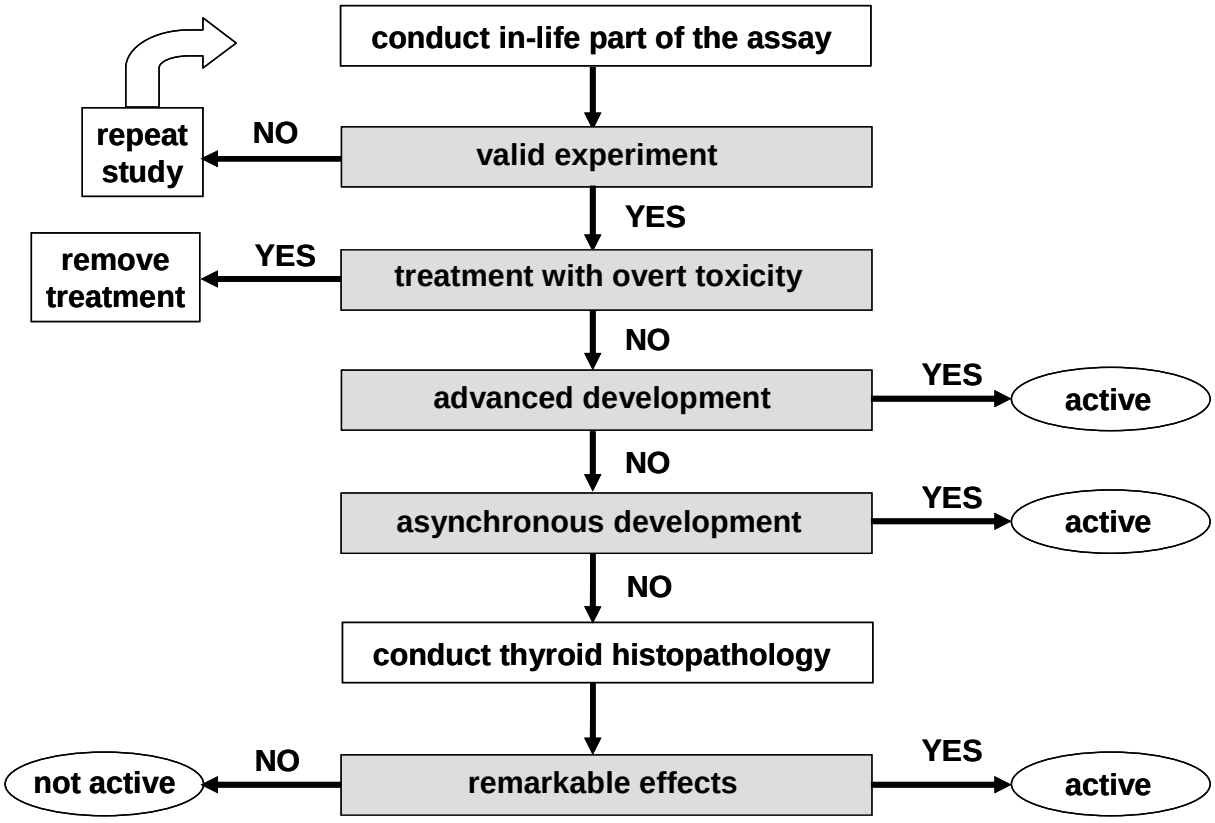


Figure 4. Flow chart describing a decision logic for the interpretation of results obtained with a test substance in the amphibian metamorphosis assay.

Applying the proposed decision logic for the interpretation of test results obtained with the test substance E2 during validation Phase-3 led to the final determination that E2 was not active on the thyroid system (Figure 5). An assessment of the general test performance indicates that E2 experiment should be regarded a valid experiment. The absence of overt toxicity implies that all test concentrations of E2 should be included in the effects assessment. Treatment of tadpoles with E2 did not cause advanced development nor were any signs of asynchronous development observed. Thyroid histopathology did not reveal remarkable effects of E2 treatment on thyroid tissue. Consequently, E2 was classified as being not active on the thyroid system given the endpoints used in the Amphibian Metamorphosis Assay.

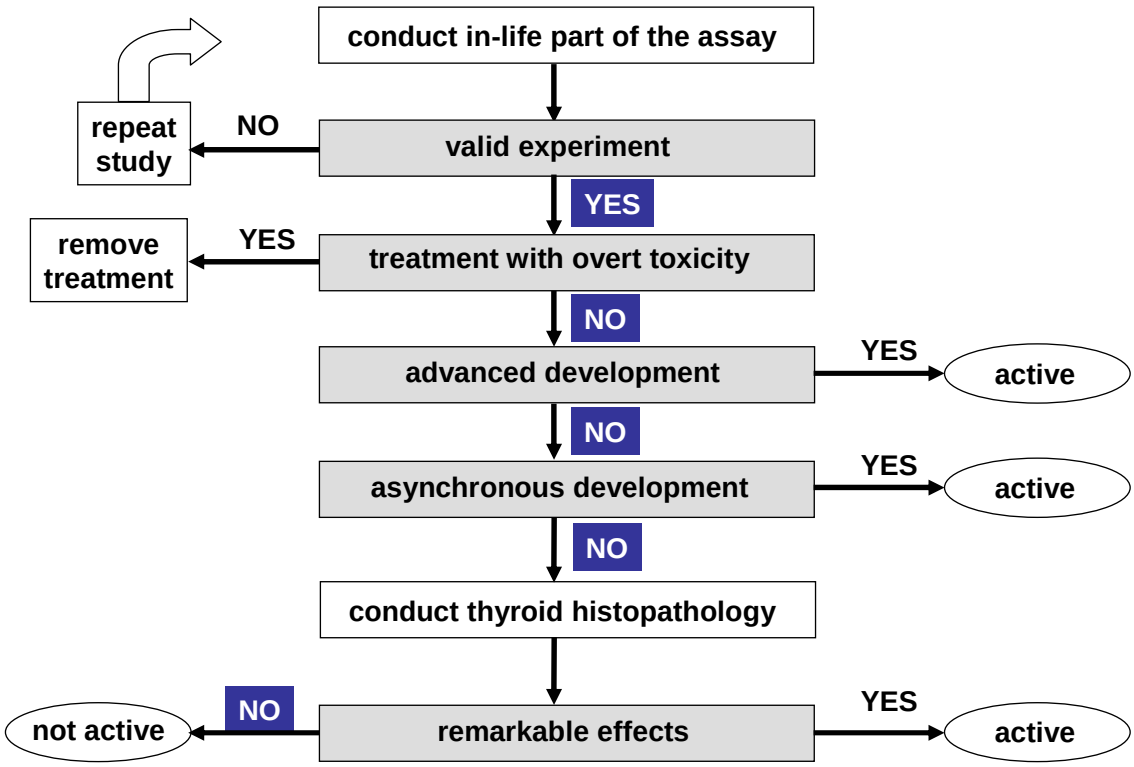


Figure 5. Flow chart describing a decision logic for the interpretation of results obtained with the test substance 17β-estradiol (E2) in the amphibian metamorphosis assay. The decisions made for E2 based on endpoint measurements are marked (dark boxes with white letters). The final determination was that E2 was not active on the thyroid system.

Applying the proposed decision logic for the interpretation of the test results obtained with the test substance BP-2 during validation Phase-3 led to the final determination that BP-2 was active on the thyroid system in both experiments (Figure 6). An assessment of the general test performance indicates that both experiments with BP-2 should be regarded valid experiments. The absence of overt toxicity implies that all test concentrations of BP-2 should be included in the assessment for biological effects. The mild growth retardation and the slightly reduced survival observed at the highest BP-2 concentration were not judged as being signs of overt toxicity, although some further guidance should be provided as to what deviations from the control will fulfill the criteria of overt toxicity. Treatment of tadpoles with BP-2 did not cause advanced development nor were any signs of asynchronous development observed. However, thyroid histopathology did reveal remarkable effects of BP-2 treatment on thyroid tissue in both exposure studies. Consequently, BP-2 was classified as being active on the thyroid system given the endpoints used in the Amphibian Metamorphosis Assay.

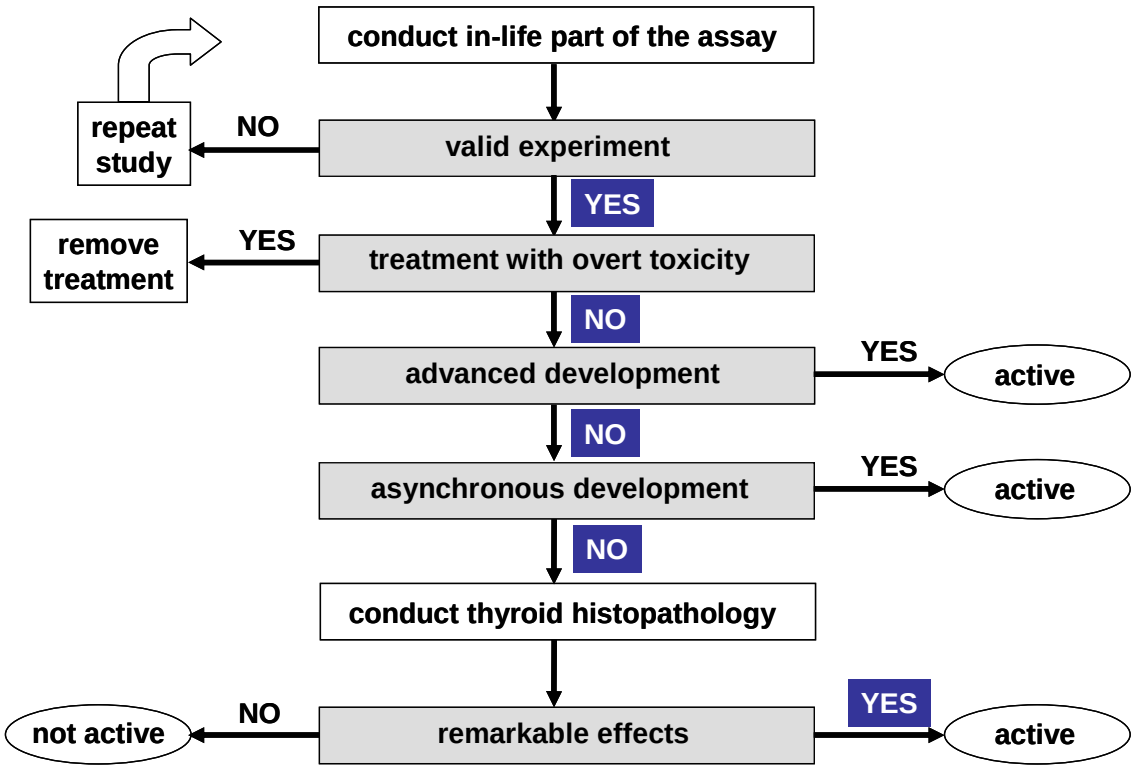


Figure 6. Flow chart describing a decision logic for the interpretation of results obtained with the test substance benzophenone-2 (BP-2) in the amphibian metamorphosis assay. The decisions made for BP-2 based on endpoint measurements are marked (dark boxes with white letters). Both experiments conducted during validation Phase-3 led to the same decisions. The final determination was that BP-2 was active on the thyroid system.

4.7 Performance of Specific Endpoints in Validation Phase-3 Studies

This section discusses the performance of the specific endpoints employed in validation Phase-3 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, some of which may not be directly 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 has been proposed to aid in the distinction of 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 endpoints.

4.7.1 Body Size and Body Wet Weight

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.

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 E2 and BP-2 on growth-related parameters by day 7 provides an important argument for the absence of overt toxicity.

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. In two experiments of the validation

Phase-3, analysis for exposure-related effects on tadpole growth by comparing treatment means for WBL, SVL and wet weight on day 21 was not possible because up to 50% of the control animals showed development to early climax stages.

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. Accordingly, the problem of a valid data interpretation towards accurate 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. One possible approach could include a stage-matched analysis of growth-related parameters.

On the other hand, the data generated from apical morphological endpoint measurements on study day 21 were robust enough to exclude the presence of overt toxicity in all three experiments of the validation phase-3. The availability of growth-related endpoint measurements is an important source of information for a sound interpretation of the study results using the proposed decision logic. Therefore, it is recommended to include measurements of WBL, SVL, and body weight in the endpoint measurements on day 21. In addition, SVL determination on day 21 is of further value as these measurements could be used to normalize HLL measurements for inter-individual differences in body size (see below).

4.7.2 Hind Limb Length

Hind limb morphogenesis is regulated by TH (Brown et al., 2005) and experimental work performed during validation Phase-1 and Phase-2 have demonstrated that HLL measurements, particularly when performed on day 7, increase the sensitivity of the assay towards detection of TH-agonistic activities. In previous experiments with T4, for example, 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. In the Phase-3 studies, neither E2 nor BP-2 showed any signs of accelerated development and hind limb growth was not enhanced on day 7 or day 21.

Determination of HLL on day 21 was a sensitive endpoint measurement to detect a mild inhibition of metamorphosis by E2 that was not evident from the analysis of median developmental stages. With regard to the BP-2 studies, detection of reduced HLL at high BP-2 concentrations on day 21 in lab 2 was consistent with the observation of a significant delay in stage development. In the other BP-2 experiment, analysis of median stages was more sensitive than HLL determination to detect the mild developmental delay caused by BP-2. To this end, it appears that it is the combined determination of HLL and developmental stages that ensures generation of relevant information to assess the effects of the test substances E2 and BP-2 on tadpole development.

However, the analysis of test substance effects on hind limb growth based on measurements of absolute HLL can be confounded by concurrent effects of a test substance on body size. This contention is supported by the close correlation between HLL and body length normally observed in stage-matched control tadpoles. Given this correlation, a need exists to identify approaches to normalize individual HLL data for differences in individual body size. Such a normalization is required to distinguish specific effects on hind limb growth from unspecific effects on overall body size. Without this distinction, the diagnostic information derived from the HLL endpoint measurement could be considerably weakened and this endpoint measurement could become prone to the generation of false positive or false negative results.

In the analysis of HLL data from the Phase-3 studies, two simple models of HLL data normalization were used including the calculation of ratios for HLL/WBL (HLL-W) or HLL/SVL (HLL-S). In the Phase-3 study, the three resulting data sets for HLL, HLL-W and HLL-S were separately analyzed using various statistical procedures and the value of data normalization was compared for each of the six HLL measurements events.

- In the E2 study, normalization of HLL by either WBL (HLL-W) or SVL (HLL-S) did not affect the outcome of the statistical analyses on study day 7 and 21. However, the HLL-S data set revealed an improved concentration-response monotonicity for E2 effects on hind limb growth on study day 21.
- In the BP-2 study in lab 1, normalization of HLL by WBL (HLL-W) and SVL (HLL-S) resulted in different outcomes of the statistical analyses on study day 21. Measurements of absolute HLL on day 21 resulted in non-monotonicity and pairwise parametric procedure detected significant increases in absolute HLL at the lowest BP-2 (0.75 mg/l) and an intermediate BP-2 concentration (3.0 mg/l) relative to the

control. HLL-W normalization removed the significant effect at the intermediate 3.0 mg/l concentration but the significant effect at the lowest BP-2 concentration was still present in this data set. In contrast, no significant differences were detectable after HLL-S normalization.

- In the BP-2 study in lab 2, normalization of HLL by WBL (HLL-W) and SVL (HLL-S) resulted in different outcomes of the statistical analyses on study day 7 and day 21. For the day 7 data sets, significant increases in absolute HLL were detected at the two highest BP-2 concentrations. HLL-W normalization removed the significant effect at the second highest BP-2 concentration. In contrast, using the data set derived from HLL-S normalization, significant decreases in HLL were detected for all BP-2 concentrations. For the day 21 data sets, significant increases in absolute HLL were detected at the two highest BP-2 concentrations and HLL-W as well as HLL-S normalization removed the significant effect at the second highest BP-2 concentration.

Based on the data sets generated during Phase-3, the comparison of the three different approaches of HLL data analysis did not demonstrate a superior value of one specific approach. However, some preliminary conclusions are possible.

- For day 21 HLL measurements, HLL-W normalization is not recommended because WBL is readily affected by development to climax stages. The concurrent tissue resorption at climax stages will result in decreasing WBL values and the positive correlation between WBL and HLL is no longer given. However, HLL-W normalization might provide a normalization approach in experiments where the test animal population does not show development to climax stages.
- SVL is much less affected by the overall reduction of body size at climax stages although SVL will also decrease due to the resorption of the anal vent at climax stages. In the three experiments during Phase-3, HLL-S normalization on study day 12 was effective in removing significant differences present at intermediate concentrations in the absolute HLL data sets of the BP-2 study in lab 1. This result suggests that the use of SVL could indeed normalize the HLL data for differences in body size as the spurious increases at intermediate concentrations are otherwise difficult to explain.

However, given the limited data sets that have been analyzed, a sound assessment of the value of different normalization approaches is not yet possible. The substantial data sets from the

Phase-2 studies should therefore be used for model calculations and subsequent statistical re-analyses of the Phase-2 study results.

4.7.3 Developmental Stage

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.

Determination of developmental stage on day 7 did not reveal any deviation from control development in the experiments with E2 and BP-2. These results are in accordance with the results of the Phase-2 studies where, with few exceptions, no statistically significant effects were observed on day 7 stage development by agonists or antagonists. Determination of developmental stage on day 21 was, however, a sensitive endpoint to detect inhibition of metamorphic development in BP-2 experiments.

However, it appears that the currently applied procedures for the statistical analyses of differences in stage development are rather weak. By merely using the median stage of replicate tanks, the information provided by stage determination data might be only partially exploited. In the E2 study during Phase-3, tanks median stages were not different between treatment groups but at high E2 concentrations, there were indications that considerably less tadpoles did show development to climax stages relative to the control and the lower E2 concentrations. It is therefore recommended to re-evaluate possibilities for alternative statistical procedures that will consider other parameters than the tank median stage.

4.7.4 Thyroid Histopathology

Histopathology of thyroid tissue was relevant and reliable for the discrimination of thyroid system-related and unspecific activities affecting tadpole development. Thyroid histopathology was clearly the most sensitive and most diagnostic endpoint to detect the effects of BP-2 on the thyroid gland. Similar observations were already made in Phase-1 and Phase-2 experiments with the TH synthesis inhibitor PTU and the inhibitor of iodide uptake sodium perchlorate.

During Phase-3, qualitatively consistent histological findings were observed in the BP-2 studies demonstrating the robustness of thyroid histopathology to identify putative anti-thyroidal compounds in the amphibian metamorphosis model (Schmutzler et al., 2007). The demonstration of mild to moderate follicular hypertrophy in the BP-2 study of lab 1 by using the graded scoring system support the value of the histopathological assessment scheme to detect even mild alterations in thyroid gland function. In addition, the scoring system also clearly revealed marked quantitative differences between the two BP-2 studies by demonstrating a much more severe effect of BP-2 on follicular cell hypertrophy/hyperplasia and glandular size. To this end, the Phase-3 studies confirmed the robustness and sensitivity of the histopathological assessment approach to detect thyroidal activity in the amphibian metamorphosis assay. During the histological assessment of thyroid tissue sampled in the E2 and BP-2 studies in lab 1, it was found that the application of an additional severity score of 0.5 improved the graded assessment of the overall mild to moderate changes in follicular cell hypertrophy. The introduction of a severity score of 0.5 (reflecting a small increase in epithelial cell height - cuboidal cell shape) to distinguish between the flat cell shape (grade 0) and the low columnar cell shape (grade 1) resulted in overall severity scores that closely reflected the results of the initial qualitative assessment.

Histopathology of thyroid tissue was also relevant for the characterization of the E2 effects pattern on tadpole development. E2 treatment caused slight retardation of tadpole development but the histopathological assessment of thyroid tissue from E2-treated tadpoles suggests that E2 did not act via one of the classical direct modes of action but presumably exerted a systemic effect that could include alteration of functions of the neuroendocrine regulatory circuits. This assumption was based on the observation that thyroid tissue did not show signs of enhanced activity but rather a tendency to a state of reduced activity.

5 Conclusion

Based on the results from Phase-1, Phase-2 and Phase-3 experimental work, 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-disrupting compounds and to distinguish activities of test substances that are not directly related to thyroid system function. 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 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 supported the inclusion of the proposed endpoints as useful and relevant for a protocol of the amphibian 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. Based on these achievements, the study protocol was challenged in Phase-3 with regard to its ability to detect the effects of a putative weakly active test substance (BP-2) and an endocrine active compounds that acts via mechanisms not directly related to the thyroid system (E2). To this end, the results of Phase-3 demonstrate the robustness and sensitivity of the test protocol to detect test substances with putative weak activity on the thyroid system. However, recent findings of a rather strong activity of BP-2 in *in vitro* assays and the marked difference in the severity of BP-2 effects on the thyroid system in two different laboratories could be interpreted that the actual potency of BP-2 to disrupt thyroid system function is strongly dependent on iodide availability. These findings further highlight the need to verify and report iodide concentrations in the test systems used for conduct of the assay. The combined assessment of responses of the currently proposed endpoint measurements also provided the basis for the evaluation of E2 as being not active on the thyroid system in the Amphibian Metamorphosis Assay. This finding is important as no evidence for thyroid system disruption was detected for a wider range of E2 concentrations encompassing treatment levels with low to strong estrogenic activity. Thus, the Phase-3 results provide further confidence that the test protocol is capable to distinguish between direct modes of action on the thyroid system and estrogenic action. The latter finding is important when considering that a fairly large number of chemicals have been shown to possess weak to moderate estrogenic activity.

6 References

Baker, B.S. and Tata, J.R. (1992): Prolactin prevents the autoinduction of thyroid hormone receptor mRNA during amphibian metamorphosis. Dev. Biol. 149, 463-467

Banu, S.K., Govindarajulu, P. and Aruldhas, M.M. (2001): Developmental profiles of TSH, sex steroids, and their receptors in the thyroid and their relevance to thyroid growth in immature rats. Steroids 67, 137-144

Brown, D.D., Cai, L., Das, B., Marsh-Armstrong, N., Schreiber, A.M. and Juste, R. (2005): Thyroid hormone controls multiple independent programs required for limb development in *Xenopus laevis* metamorphosis. Proc. Natl. Acad. Sci. USA 102, 12455-12458

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

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

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

Denver, R.J. (1996): Neuroendocrine control of amphibian metamorphosis. In: Metamorphosis. Postembryonic Reprogramming of Gene Expression in Amphibian and Insect Cells. Gilbert, L.I., Tata, J.R. and Atkinson, B.G. (eds.), Academic Press, San Diego, US, 433-464

Denver, R.J., Glennemeier, K.A. and Boorse, G.C. (2002): Endocrinology of complex life cycles. In: Amphibian Hormones, Brain and Behavior, Vol. 2, (D. Pfaff, A. Arnold, A. Etgen, S. Fahrbach, R. Moss & R. Rubin (eds), Academic Press, San Diego, CA, USA, pp 469-513.

Fort, D.J., Degitz, S., Tietge, J., Touart, L.W. (2007): The hypothalamic-pituitary-thyroid (HPT) axis in frogs and its role in frog development and reproduction. Crit. Rev. Toxicol. 37, 117-61

Hayes, T.B. (1997): Steroids as potential modulators of thyroid hormone activity in anuron metamorphosis. Am. Zool. 37, 185-194

Huang, H. and Brown, D.D. (2000): Prolactin is not a juvenile hormone in *Xenopus laevis* metamorphosis. Proc. Natl. Acad. Sci. USA 97, 195-199

Jarry, H., Christoffel, J., Rimoldi, G., Koch, L. and Wuttke, W. (2004): Multi-organic endocrine disrupting activity of the UV screen benzophenone 2 (BP2) in ovariectomized adult rats after 5 days treatment. Toxicology 205, 87-93

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

Klammer, H., Schlecht, C., Wuttke, W., Schmutzler, C., Gotthardt, I., Kohrle, J. and Jarry H. (2007): Effects of a 5-day treatment with the UV-filter octyl-methoxycinnamate (OMC) on the function of the hypothalamo-pituitary-thyroid function in rats. Toxicology 238, 192-199

Leloup-Hatey, J., Buscagalia, M., Jolivet-Jaudet, G. and Leloup, J. (1990): Interrenal function during the metamorphosis in anuran amphibia. In: Biology and Physiology of Amphibians. Fortschritte der Zoologie Vol. 38, Hanke, W. (ed.), Gustav Fischer Verlag, Stuttgart, 1990, 139-154

Lima, L.P., Barros, I.A., Lisboa, P.C., Araujo, R.L., Silva, A.C., Rosenthal, D., Ferreira, A.C., and Carvalho, D.P. (2006): Estrogen effects on thyroid iodide uptake and thyroperoxidase activity in normal and ovariectomized rats. Steroids 71, 653-659

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.

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

Pickford, D.B., Hetheridge, M.J., Caunter, J.E., Hall, A.T. and Hutchinson, T.H. (2003): Assessing chronic toxicity of bisphenol A to larvae of the African clawed frog (*Xenopus laevis*) in a flow-through exposure system. Chemosphere 53, 223-235

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

Schlumpf, M., Schmid, P., Durrer, S., Conscience M, Maerkel K, Henseler M, Gruetter M, Herzog I, Reolon S, Ceccatelli R, Faass O, Stutz E, Jarry H, Wuttke W, Lichtensteiger W. (2004): Endocrine activity and developmental toxicity of cosmetic UV filters - an update. Toxicology 205, 113-122

Schmutzler C, Hamann I, Hofmann PJ, Kovacs G, Stemmler L, Mentrup B, Schomburg L, Ambrugger P, Gruters A, Seidlova-Wuttke D, Jarry H, Wuttke W, Kohrle J. (2004): Endocrine active compounds affect thyrotropin and thyroid hormone levels in serum as

well as endpoints of thyroid hormone action in liver, heart and kidney. Toxicology 205, 95-102

Schmutzler, C., Bacinski A, Gotthardt I, Huhne K, Ambrugger P, Klammer H, Schlecht C, Hoang-Vu C, Gruters A, Wuttke W, Jarry H, Kohrle J. (2007): The ultraviolet filter benzophenone 2 interferes with the thyroid hormone axis in rats and is a potent in vitro inhibitor of human recombinant thyroid peroxidase. Endocrinology 148, 2835-2844

Seidlova-Wuttke, D., Jarry, H., Christoffel, J., Rimoldi, G., and Wuttke, W. (2005): Effects of bisphenol-A (BPA), dibutylphtalate (DBP), benzophenone-2 (BP2), procymidone (Proc), and linurone (Lin) on fat tissue, a variety of hormones and metabolic parameters: a 3 months comparison with effects of estradiol (E2) in ovariectomized (ovx) rats. Toxicology 213, 13-24

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

Sosic-Jurjevic, B., Filipovic, B., Milosevic, V., Nestorovic, N., Negic, N. and Sekulic, M. (2006): Effects of ovariectomy and chronic estradiol administration on pituitary-thyroid axis in adult rats. Life Sci. 79, 890-897

Tata, J.R. (1996): Hormonal interplay and thyroid hormone receptor expression during amphibian metamorphosis. In: Metamorphosis. Postembryonic Reprogramming of Gene Expression in Amphibian and Insect Cells. Gilbert, L.I., Tata, J.R. and Atkinson, B.G. (eds.), Academic Press, San Diego, US, 465-503

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

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