Critical Perspectives

The Extended Amphibian Metamorphosis Assay: A Thyroid-Specific and Less Animal-Intensive Alternative to the Larval Amphibian Growth and Development Assay

Lisa S. Ortego,[a] Allen W. Olmstead,[b] Lennart Weltje,[c] James R. Wheeler,[d] Audrey J. Bone,[e] Katherine K. Coady,[e] Chris S. Banman,[f] Natalie Burden,[g] and Laurent Lagadic[h,]*

[a]Bayer U.S. LLC, Crop Science, Environmental Effects and Risk Assessment, Cary, North Carolina, USA
[b]BASF Corporation, Agricultural Solutions–Ecotoxicology, Research Triangle Park, North Carolina, USA
[c]BASF SE, Agricultural Solutions–Ecotoxicology, Limburgerhof, Germany
[d]Shell International B.V., Shell Health, The Hague, The Netherlands
[e]Bayer U.S. LLC, Crop Science, Environmental Effects and Risk Assessment, Chesterfield, Missouri, USA
[f]SynTech Research, Stilwell, Kansas, USA
[g]National Centre for the Replacement, Refinement, & Reduction of Animals in Research, London, United Kingdom
[h]Bayer AG, Research and Development, Crop Science, Environmental Safety, Monheim am Rhein, Germany

Abstract: The amphibian metamorphosis assay (AMA; US Environmental Protection Agency [USEPA] test guideline 890.1100 and Organisation for Economic Co-Operation and Development test guideline 231) has been used for more than a decade to assess the potential thyroid-mediated endocrine activity of chemicals. In 2013, in the context of the Endocrine Disruptor Screening Program of the USEPA, a Scientific Advisory Panel reviewed the results from 18 studies and recommended changes to the AMA test guideline, including a modification to a fixed-stage design rather than a fixed-time (i.e., 21-d) design. We describe an extended test design for the AMA (or EAMA) that includes thyroid histopathology and time to metamorphosis (Nieuwkoop–Faber [NF] stage 62), to address both the issues with the fixed-time design and the specific question of thyroid-mediated adversity in a shorter assay than the larval amphibian growth and development assay (LAGDA; Organisation for Economic Co-Operation and Development test guideline 241), using fewer animals and resources. A demonstration study was conducted with the EAMA (up to NF stage 58) using sodium perchlorate. Data analyses and interpretation of the fixed-stage design of the EAMA are more straightforward than the fixed-time design because the fixed-stage design avoids confounded morphometric measurements and thyroid histopathology caused by varying developmental stages at test termination. It also results in greater statistical power to detect metamorphic delays than the fixed-time design. By preferentially extending the AMA to NF stage 62, suitable data can be produced to evaluate thyroid-mediated adversity and preclude the need to perform a LAGDA for thyroid mode of action analysis. The LAGDA remains of further interest should investigations of longer term effects related to sexual development modulated though the hypothalamus–pituitary–gonadal axis be necessary. However, reproduction assessment or life cycle testing is currently not addressed in the LAGDA study design. This is better addressed by higher tier studies in fish, which should then include specific thyroid-related endpoints. *Environ Toxicol Chem* 2021;40:2135–2144. © 2021 The Authors. *Environmental Toxicology and Chemistry* published by Wiley Periodicals LLC on behalf of SETAC.

* Address correspondence to laurent.lagadic@bayer.com

INTRODUCTION

The amphibian metamorphosis assay (AMA) has been used for more than a decade to assess potential thyroid-mediated activity of chemicals, including pesticides. In 1998, the Endocrine Disruptor Screening and Testing Advisory Committee, a Federal Advisory Committee formed by the US Environmental Protection Agency (USEPA), recommended the AMA as a tool to screen for chemicals with thyroid agonist or antagonist activity. The international effort to validate the assay

resulted in the final test guideline referenced as both USEPA Office of Prevention, Pesticides, and Toxic Substances guideline 890.1100 (US Environmental Protection Agency 2015) and Organization for Economic Co-operation and Development (OECD; 2015) test guideline 231. In 2009, the AMA was required and performed for substances on list 1 of the USEPA's Endocrine Disruptor Screening Program (EDSP), most of which are pesticide active substances (US Environmental Protection Agency 2009). This was the first broad application of the AMA test guideline.

In 2013, a Scientific Advisory Panel (SAP) convened by the USEPA was tasked with evaluating the performance of the EDSP tier 1 (screening) assays, including the AMA. The SAP reviewed the available test results from 18 studies (a subset of list 1 studies) and recommended changes to the AMA test guideline that would improve the conduct and interpretation of the assay (US Environmental Protection Agency 2013a). Although the USEPA raised no specific concerns about the SAP acceptance of the fixed-stage design, the SAP recommendations have not yet been implemented in the test guideline because the USEPA has not undertaken revision of the EDSP tier 1 guidelines since their finalization. Therefore, the 2009 version of the AMA test guideline remains the current one.

The standard AMA is initiated by exposing tadpoles at Nieuwkoop and Faber (NF; 1994) stage 51 to 3 concentrations of the test substance, and then data for a number of endpoints are collected at days 7 and 21. At termination, animals will be in various stages of metamorphosis. The test guideline has a performance criterion intended to minimize the stage distribution range: in the control and the10th and 90th percentiles of the developmental stage distribution should not span more than 4 stages. However, historical control evaluations have shown that this criterion can be challenging to meet (Coady et al. 2014). The endpoints measured in the AMA, including snout–vent length (SVL), hindlimb length (HLL), and thyroid histology, are influenced by the stage of development, and issues arise when attempting to compare these endpoints among tadpoles of differing developmental stages (Pawlowski et al. 2019). Because thyroid histology will also reflect developmental stage (Grim et al. 2009), comparisons across developmental stages are problematic and can result in misinterpretation (US Environmental Protection Agency 2013b) and in difficulties identifying matching individuals across treatments. In addition, morphometric measurements between pre- and post-NF stage 60 tadpoles should not be compared due to the morphological remodeling that takes place during the metamorphic climax. Sixteen of the 18 AMA reports analyzed by the USEPA for the SAP meeting had tadpoles at more than NF stage 60, with 4 of the 18 studies having a high degree (>20%) of tadpoles exceeding NF stage 60 (US Environmental Protection Agency 2013b). Due to these considerations, the SAP recommended that the USEPA considers a modification to a fixed-stage design rather than a fixed-time (i.e., 21-d) design (US Environmental Protection Agency 2013a). This recommendation was the starting point of various attempts to improve the AMA by using a time-to-stage design (Ortego et al. 2014; Haselman et al. 2016).

More recently, European regulations have set out scientific criteria for determination of the endocrine-disrupting properties of Plant Protection Product and Biocidal Product active substances (European Commission 2017, 2018). The European evaluation of endocrine disruptors is in line with the definition of the World Health Organization, International Programme on Chemical Safety (2002) and therefore requires sufficient data on both mechanistic endocrine activity and adverse effects (e.g., an impact on reproduction), and a plausible biological link between the 2. For the thyroid modality, the suggested study for evaluating endocrine activity in nonmammalian vertebrates is the AMA, whereas for adversity, the larval amphibian growth and development assay (LAGDA; US Environmental Protection Agency 2015; Organisation for Economic Co-operation and Development test guideline 241; 2015) is recommended (European Chemicals Agency and European Food Safety Authority 2018). The LAGDA was designed to evaluate adverse apical effects that are under the control of multiple endocrine pathways and characterize potential concentration–response relationships. Thyroid histopathology and time to metamorphosis (NF stage 62) may respond to interference with the hypothalamic–pituitary–thyroid (HPT) axis in the LAGDA (Organisation for Economic Co-operation and Development 2018a) and are therefore considered "T-mediated" endpoints in European Chemicals Agency and European Food Safety Authority (2018). We propose an extended test design for the AMA (or EAMA) that includes thyroid histopathology and time to metamorphosis (NF stage 62), to address both the known issues with the fixed-time design (US Environmental Protection Agency 2013a, 2013b) and the specific question of T-mediated adversity in a shorter, better targeted assay, using fewer animals and resources compared with the LAGDA.

Purpose and validation of the LAGDA protocol

The LAGDA was initially designed as one of the proposed tier 2 assays of the USEPA EDSP (US Environmental Protection Agency 2013c). The protocol describes a long-term chronic toxicity test with the African clawed frog *Xenopus laevis* that considers growth and development from the embryo through the early juvenile period over 130 d or more. The test focuses on early development, growth, and partial sexual development that is assessed via gonad histology (Table 1). With respect to the USEPA approach, the LAGDA was not necessarily designed to elucidate mechanisms. However, according to the OECD (Organisation for Economic Co-operation and Development 2018a), it offers endpoints that provide specific information about endocrine disruption or impaired reproduction (Table 1). Most of these specific endocrine endpoints are likely to respond to interference with the hypothalamic–pituitary–gonadal axis, whereas thyroid histopathology and time-to-metamorphosis may respond to interference with the HPT axis, as may the growth endpoint, although growth responses are not necessarily thyroid specific (Organisation for Economic Co-operation and Development 2018a). The LAGDA is on level 4 of the OECD conceptual framework in that it provides information about

TABLE 1: Primary (●) and optional (○) endpoints and timing of measurements in the amphibian metamorphosis assay (AMA) and the larval amphibian growth and development assay (LAGDA) according to Organisation for Economic Co-operation and Development (OECD) test guidelines 231 (2009) and 241 (2015), respectively

Biological endpoint	AMA (OECD test guideline 231)			LAGDA (OECD test guideline 241)			
	Daily	Day 7	Day 21	Daily	Interim sampling, larval NF stage 62	Interim sampling, culled juvenile NF stage 66	Termination, juvenile
Mortality[a]	●			●			
Clinical signs of toxicity and abnormal behavior[a]	●			●			
Wet body weight[a]		●	●		●	●	●
Total body length[a]		○	○		●		
Snout–vent length[b]		●	●		●	●	●
Hindlimb length[b]		●	●				
Developmental stage[b]		●	●			●	
Time to NF stage 62 (metamorphosis)[b]					●		
Liver somatic index[a]						●	●
Phenotypic/genetic sex ratio[a]						●*	●
Plasma vitellogenin[a]							○
Thyroid histopathology			●**		●		
Gonad, reproductive duct, liver and kidney histopathology							●

[a]Nonspecific for thyroid.
[b]Nonspecific for thyroid unless accompanied by significant thyroid histopathology (Organisation for Economic Co-operation and Development 2009, 2015, 2018; Dang 2019).
*Phenotypic sex only (based on gonad morphology).
**Overt and significant changes in apical endpoints indicating developmental acceleration or asynchrony may preclude the necessity to perform histopathological analysis of the thyroid glands. However, thyroid histology may be required by some regulatory authorities regardless of the apical responses.

adverse effects on endocrine-relevant endpoints but is not comprehensive, nor does it cover a life cycle (Organisation for Economic Co-operation and Development 2018a). However, it is intended to serve as a higher tier test with an amphibian model for collecting definitive concentration–response information on adverse effects, suitable for use in ecological risk assessment (US Environmental Protection Agency 2013c; Organisation for Economic Co-operation and Development 2015).

The general experimental design of the LAGDA entails exposing NF-stage 8 to 10 *X. laevis* embryos to a minimum of 4 increasing concentrations of a test chemical and a control until 10 wk after the median time to NF stage 62 is reached in the control, with an interim subsample at NF stage 62 (Table 1). The endpoints evaluated during the course of the present study (Table 1) include those indicative of general toxicity (i.e., mortality, abnormal behavior, and growth [SVL and weight]), as well as endpoints designed to characterize specific endocrine modes of action targeting estrogen-, androgen-, steroidogenesis-, or thyroid-mediated physiological processes (i.e., thyroid histopathology, gonad and gonad duct histopathology, abnormal development, plasma vitellogenin [optional], and genotypic/phenotypic sex ratios; Organisation for Economic Co-operation and Development 2015). The thyroid-specific endpoints obtained from NF stage 62 tadpoles are time-to-stage 62, histopathology of the thyroid gland, and morphometric analysis. There are 4 replicates in each test concentration and 8 replicates in the control. Initial density is 20 embryos/replicate, resulting in a minimum number of 480 animals in a 4-concentration test design with a negative

control, or 640 animals if a solvent control were to be included. After NF stage 66 or after 70 d (whichever occurs first), the density is adjusted by randomly selecting 10 juveniles, which are euthanized, and then growth and gross necropsy with identification of phenotypic sex are recorded (US Environmental Protection Agency 2013c; Organisation for Economic Co-operation and Development 2015).

The validation of the LAGDA was conducted in 5 different laboratories using reference chemicals that are not specific for the thyroid modality (Table 2). Only benzophenone-2, tested in one laboratory, is known to act as an inhibitor of thyroperoxidase (see Schmutzler et al. 2007) but is also an agonist of the estrogen receptor (see Schlecht et al. 2004; Seidlová-Wuttke et al. 2004). Due to the limited number of chemicals and laboratories involved in the validation of the LAGDA, interlaboratory reproducibility is not well documented (Organisation for Economic Co-operation and Development 2015). In addition, a number of inconsistencies were reported during the validation exercise (US Environmental Protection Agency 2013c). These mainly concern 1) organism growth, indicating feeding and husbandry issues; 2) pathology findings that were not consistent across laboratories; 3) unsuccessful thyroxine measurements in all laboratories but one for 2 reference chemicals only (Haselman et al. 2016); and 4) highly variable vitellogenin measurements, making it difficult to interpret the induction levels (US Environmental Protection Agency 2013c). To a certain extent the issues are acknowledged in the OECD version of the LAGDA test guideline (Organisation for Economic Co-operation and Development 2015): "it is

TABLE 2: Reference substances tested in the interlaboratory validation exercise of the larval amphibian growth and development assay[a]

	Prochloraz	4-tert-octylphenol	17β-trenbolone	Benzophenone-2
Known mode of action	Aromatase inhibitor, AR antagonist	ER agonist	AR agonist	ER agonist, TPO inhibitor
Lab A	●	●	●	
Lab B	●	●		
Lab C	●	●		
Lab D	●			
Lab E				●

[a]US Environmental Protection Agency (2013).
AR = androgen receptor; ER = estrogen receptor; TPO = thyroperoxidase.

anticipated that when a sufficient number of studies is available, the TG [test guideline] will be reviewed and if necessary revised in light of experience gained."

To date, the LAGDA has been implemented only occasionally in a very limited number of contract research organizations (CROs). In the open literature, beyond publications on the validation studies, only one study (Fort et al. 2017) has been reported. In fact, most CROs have no experience with the LAGDA, so the issues highlighted during the validation exercise remain unresolved.

The AMA and its extended version (EAMA)

The AMA went through an extensive validation process (Organisation for Economic Co-operation and Development 2007a, 2007b, 2007c, 2008a) supported by an OECD detailed review paper (Organisation for Economic Co-operation and Development 2004a) and further scientific reviews in peer-reviewed journals (see Pickford 2010; Miyata and Ose 2012; Dang 2019; Glaberman et al. 2019). The test has been widely used since the first protocol validation effort (Opitz et al. 2005, 2006) and the publication of the USEPA and OECD test guidelines in 2009, so CROs and industry and university laboratories (see Grim et al. 2009; Coady et al. 2010) have acquired significant experience with the test protocol.

The AMA is a screening assay intended to empirically identify substances that may interfere with the normal function of the HPT axis (Organisation for Economic Co-operation and Development 2009). The general experimental design entails exposing NF stage 51 X. laevis tadpoles to a minimum of 3 increasing concentrations of a test chemical and a dilution water control for 21 d. There are 4 replicates for each test concentration and the control. The number of tadpoles at test initiation is 20/replicate for all groups including the dilution water control, resulting in 320 animals, or 400 when a solvent control is included.

The biological endpoints include daily observations of survival, and measurements of HLL, SVL, wet body weight, and developmental stage at days 7 and 21, whereas thyroid histology, if conducted, is done at test termination (day 21; Table 1). Although developmental acceleration, asynchronous development, and delayed development in the absence of systematic toxicity may be interpreted as sufficient evidence of potential thyroid activity, it is commonly acknowledged that the

response of these apical endpoints can be considered thyroid relevant when accompanied by significant thyroid histopathological findings such as moderate or severe follicular hypertrophy and/or hyperplasia (Organisation for Economic Co-operation and Development 2009, 2018a; Pickford 2010; Dang 2019).

The apical endpoints measured in the AMA may be relevant when one is considering individual responses that could potentially be population relevant. Therefore, the European Chemicals Agency and European Food Safety Authority (2018) considers that both the AMA and the LAGDA can be used to investigate potential endocrine-related adverse effects in amphibians. In principle, to have the T-mediated adversity sufficiently investigated, the results from all the LAGDA T-mediated parameters would be needed (European Chemicals Agency and European Food Safety Authority 2018). However, if the T-mediated parameters recommended in the AMA are negative, this suggests that the chemical is not thyroid active and would negate the need for a LAGDA to further study thyroid-related adverse effects (European Chemicals Agency and European Food Safety Authority 2018).

The relevance of the AMA as an indicator of adversity can be enhanced by extending the test duration and adding endpoint measurements at NF stage 62 as in the LAGDA (Figure 1; Table 3). The NF stage 62 encompasses metamorphic climax, and the time to reach NF stage 62 is, therefore, sensitive to modes of action that limit peak thyroid hormone levels in the developing tadpole, especially at later stages of metamorphosis (see Denver 2009; Vitt and Caldwell 2013; Olker et al. 2018; Haselman et al. 2020). Increasing the number of test concentrations to 4 or 5 (from 3) would improve the resolution of effects, allowing no-observed-effect concentration or effect concentration, percentage determinations—not foreseen when the AMA is run as a screening test (Wheeler et al. 2014; Organisation for Economic Co-operation and Development 2018a). This might be particularly important if the endpoints are used in population modeling, risk assessment, or similar exercises.

An alternative fixed-stage test design (EAMA) has been developed to account for some of the above limitations of the fixed-time design of the AMA (US Environmental Protection Agency 2013a; Ortego et al. 2014; Haselman et al. 2016; Olker et al. 2018; Haselman et al. 2020). In this alternative design, the experimental set-up, acclimation, initiation, and daily observations are similar to the current AMA
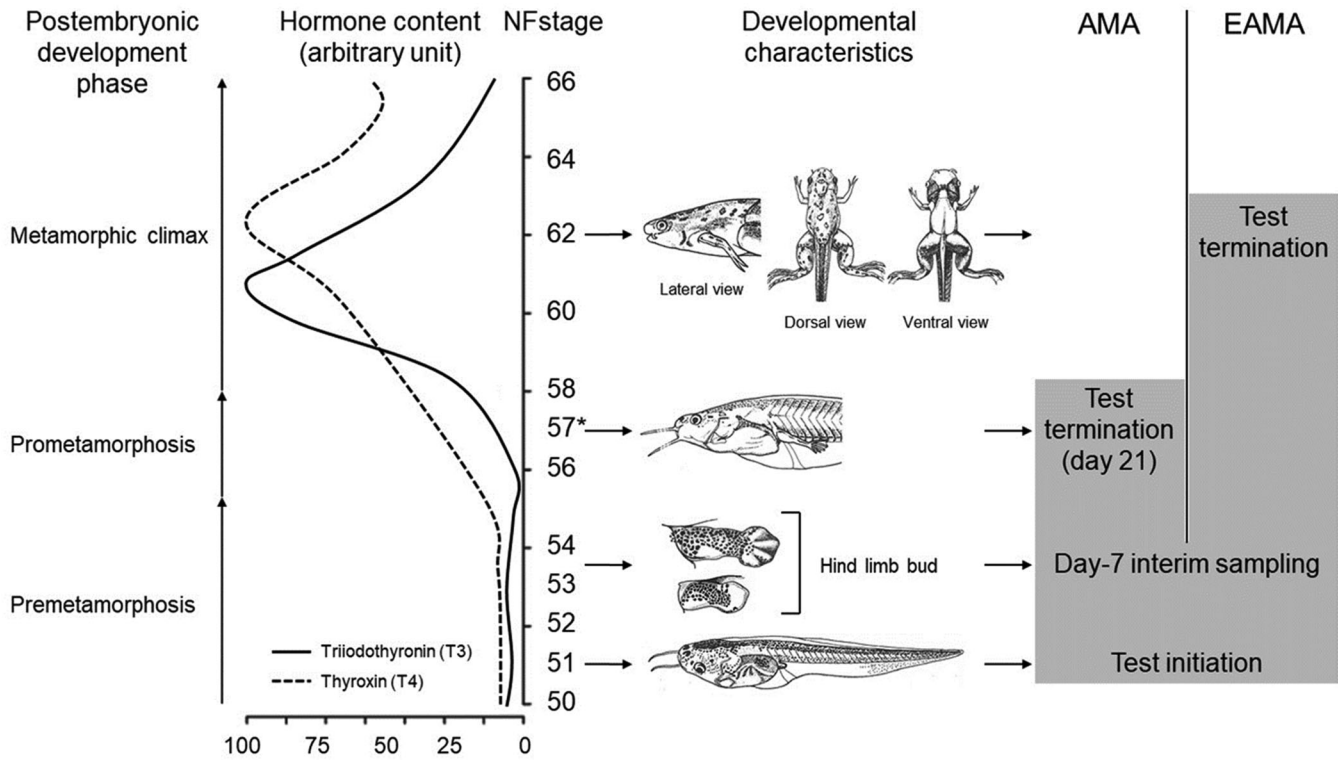
FIGURE 1: *Xenopus laevis* postembryonic development phases covered by the amphibian metamorphosis assay (AMA) and the extended assay design (EAMA). Up to the metamorphic climax, the increase in thyroid hormone levels results in morphological changes (e.g., hindlimb development) and histological changes in the thyroid gland (see Grimm et al. 2009). In *X. laevis*, thyroid hormone levels are highest between NF stages 58 and 62. Following the metamorphic climax (NF stage 62–66), decreasing levels of thyroid hormone are accompanied by the regression of thyroid follicles (not shown). Images from *X. laevis* developmental stage according to Nieuwkoop and Faber (1994) were taken from Xenbase (1994). The gray area shows that, in contrast to the AMA, the EAMA encompasses most of the postembryonic development up to the metamorphic climax. *, Minimum median developmental stage of controls at test termination according to OECD test guideline 231 (Organisation for Economic Co-operation and Development 2009).

guidelines. Day 7 endpoints (HLL, SVL, wet body weight, and developmental stage) are the same (Table 3). However, after day 7, tadpoles are observed daily and when the desired stage (i.e., NF stage 62) is reached, those animals are removed from the test tanks, the endpoints are measured (as noted on day 7), and tadpoles are processed for histological analysis (Table 3). This change in study design also reflects the design in the LAGDA (Table 1), in which time to NF stage 62 is used to monitor potential changes in developmental rate.

A demonstration study was conducted to explore the fixed-stage design using sodium perchlorate as a model thyroid-active substance. Sodium perchlorate was chosen to enable comparison of the results obtained in the EAMA with those generated in the AMA during the validation effort (Organisation for Economic Co-operation and Development 2008b). Although NF stage 58 was chosen as the terminal stage in this demonstration study based on the easily identifiable morphological marker of forelimb emergence (as in Saka and Tada 2021, using *Silurana tropicalis*, formerly *Xenopus tropicalis*), extending the assay to NF stage 62 would more inclusively cover metamorphic climax and sensitively assess modes of action responsible for limiting peak thyroid hormone levels (Olker et al. 2018; Haselman et al. 2020). Assessing stage

62 also aligns the EAMA with the assessment of T-mediated endpoints in the LAGDA. In addition, the new extended test

TABLE 3: Primary (●) and optional (○) endpoints and timing of measurements in the fixed-stage test design of the extended amphibian metamorphosis assay

Biological endpoint	Daily	Day 7	Termination (NF stage 62)
Mortality[a]	●		
Clinical signs of toxicity and abnormal behavior[a]	●		
Wet body weight[a]		●	●
Total body length[a]		○	○
Snout–vent length (SVL)[b]		●	●
Hindlimb length (HLL)[b,c]		●	●
Developmental stage[b]		●	●
Thyroid histology			●[d]

[a]Nonspecific for thyroid.
[b]Nonspecific for thyroid, unless accompanied by significant thyroid histopathology (Organisation for Economic Co-operation and Development 2009, 2015, 2018; Dang 2019).
[c]HLL can be normalized for analysis by dividing it by the SVL as required by the AMA test guideline (see also discussion in Pawlowski et al. 2019).
[d]Overt and significant changes in apical endpoints indicating developmental acceleration or asynchrony may preclude the necessity to perform histopathological analysis of the thyroid glands. However, thyroid histology may be required by some regulatory authorities regardless of the apical responses.

design allows for appropriate comparisons of endpoints across the same developmental stage, and hence for a more robust analysis and interpretation of the study results.

EAMA demonstration study

A demonstration study was conducted to evaluate the fixed-stage test design, or EAMA. The test substance chosen was sodium perchlorate ($NaClO_4$), a known thyroid inhibitor (blocking iodine uptake; Pickford 2010; Pleus and Corey 2018) that was used in the OECD interlaboratory validation of the AMA test guideline (Organisation for Economic Co-operation and Development 2007a). Nominal test concentrations were 31.25, 62.5, 125, and 250 µg/L (ppb). Test concentrations were successfully confirmed analytically for the high- and low-test levels. The test protocol is provided in the Supplemental Data (S1). In short, the USEPA guideline for the AMA (guideline 890.1100; US Environmental Protection Agency 2009) was generally followed. There were 4 replicates with 20 NF stage 51 tadpoles/replicate. The temperature was 21.8 to 22.9 °C. The tadpoles were fed Liquid Frog Brittle Slurry (60 mg/mL) twice daily according to guideline recommendations. Food ration increased throughout the study to account for tadpole growth but was not altered when tadpoles were removed from the tank as they reached NF stage 58. The tadpoles were exposed under flow-through conditions until 95% of all surviving tadpoles across all treatments reached NF stage 58 or day 30, whichever occurred first. We chose NF stage 58 because it was the closest to NF stage 60 with a conspicuous morphological trait that would be easy to recognize (i.e., forelimb emergence).

The observational endpoints were SVL, hindlimb length (further normalized to SVL, i.e., nHLL), developmental stage, wet weight, and daily observations of mortality. Thyroid histology was not included in the demonstration study. Available literature data were used to confirm that thyroid histological changes induced by sodium perchlorate do not depend on the age of the tadpoles. There are consistent reported thyroid histopathological findings in tadpoles following exposure to sodium perchlorate for 8 to 70 d (Tietge et al. 2005, 2010; Goleman and Carr 2006; Opitz et al. 2009). It appears that thyroid histopathology at later stages are not less relevant (sensitive or mechanistically informative) than earlier assessments. Because the perchlorate data are comparable across studies, it was not deemed relevant to include thyroid histology in the demonstration study, which focused on quantifying time-related development.

All AMA test guideline validity criteria were met for analytical variability, mortality, overt toxicity, dissolved oxygen, pH, and temperature. Mean-measured concentrations were 105% of nominal at the high- and low-treatment levels. Because a proportional diluter was employed, it can be reliably assumed that the intermediate exposure levels were also close to nominal. Survival in the control and all treatment groups was >98%. There were no mortalities or toxic signs related to exposure to the test substance, supporting the idea that the sodium perchlorate exposures did not exceed the maximum tolerated concentration. A few animals in the control and treatment groups exhibited bent tails; the maximum number of bent tail observations/replicate were 2, 1, 1, 1, and 3 for the control and 31.25, 62.5, 125, and 250 µg/L treatment groups, respectively. Bent tail observations in the present study were lower (generally <5%) than frequently observed in the AMA (Coady et al. 2014). For instance, the bent tail percentage often exceeded 20% in the AMA validation (US Environmental Protection Agency 2013b). No significant differences were detected in day 7 endpoints (mortality, developmental stage, normalized HLL, SVL, wet body weight). Sampling of individuals that reached NF stage 58 started on study day 17. Study termination occurred on day 27, when 95% of the tadpoles in the test reached NF stage 58. These observations corroborate the results of a comparable study implemented as an informal ring testing exercise of a *Xenopus* metamorphosis assay, in which tadpoles reached NF stage 58 within 28 d in the controls of 6 of 10 tests (Opitz et al. 2005).

Data analysis of the time to reach NF stage 58 was performed using the Cox mixed-effects model with a significant treatment difference determined from a 2-sided Dunnett's test ($\alpha = 0.05$; Green et al. 2018) using R statistical software (R Core Team 2020). This approach was also used to analyze time to NF stage 62 data in a LAGDA (Haselman et al. 2016). Body morphometric analysis was analyzed with the Jonckheere–Terpstra step-down test ($\alpha = 0.05$) using CETIS statistical software (Tidepool Scientific Software 2000–2012).

The results are provided in Table 4. There were significant effects on the following parameters: time to NF stage 58 (delayed at 250 µg/L), SVL (increased at 125 and 250 µg/L), and wet body weight (increased at 62.5, 125, and 250 µg/L). In comparison with the AMA interlaboratory validation, the results from the EAMA were similar at test termination (Table 4). Observed results were consistent with perchlorate's known thyroid antagonism. Specifically, exposure increased the time to a specific developmental stage (NF stage 58), suggesting delayed development analogous to the reduction in median stage achieved by day 21 in the AMA. Effects on body morphometrics were also observed that are consistent, although not necessarily diagnostic of T-mediated delayed development without corroborating histopathology. For detecting delayed development, the fixed-stage design revealed significant differences from the control at an exposure level (250 µg/L) that only 2 of the 5 validation laboratory studies detected. This finding suggests that the EAMA would be at least as sensitive in detecting antithyroid activity as the AMA.

Further experimental and statistical considerations

Some EAMAs have been recently implemented to address European Union regulatory requirements for the thyroid modality. These studies use NF stage 62 as the termination stage to align with the LAGDA test guideline, which includes a time to NF stage 62 endpoint in its design. This also addresses potential concerns that terminating at NF stage 58 may not be as sensitive

TABLE 4: Summary table comparing the extended amphibian metamorphosis assay (EAMA) results (present study) with the amphibian metamorphosis assay (AMA) interlaboratory validation results for sodium perchlorate[a]

Assessment time	Biological endpoint	AMA inter-laboratory validation					Extended AMA
		Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	
Daily	Mortality	–	–	–	–	–	–
Day 7	Developmental stage	–	–	–	–	–	–
	Wet body weight	↑	–	–	↑	–	–
	SVL[b]	↑	–	–	↑	–	–
	nHLL[c]	–	–	–	–	–	–
Day 21 (AMA) or	Developmental delay	–	–	Yes	Yes	–	Yes
Day 27 (EAMA)	Wet body weight	↑	–	↑	–	–	↑
	SVL[b]	↑	–	↑	–	↑	↑
	nHLL[c]	–	–	↑	↓	–	–

[a]Data from Organisation of Economic Co-operation and Development (2007a).
[b]Snout-vent length.
[c]Normalized hinlimb length.
– no effect; ↑ increase; ↓ decrease; as statistically compared with the control.
SVL = snout–vent length; nHLL = normalized hindlimb length.

for detecting all thyroid modes of action (Olker et al., 2018). Running the EAMA up to NF stage 62 would incorporate the metamorphic climax, an important developmental milestone. Whereas using NF stage 58 resulted in a test that is approximately 6 d longer than the 21-d AMA (6 d in the demonstration study; 5–7 d in Saka and Tada 2021 with *Silurana*, formerly *Xenopus*, *tropicalis*), using NF stage 62 would result in a test that is approximately 10 to 20 days longer (time appears to vary based on laboratory animal husbandry).

Fixed-stage determination has been commonly used in amphibian development tests (Ankley et al. 1998; Allran and Karasov 2000; Opitz et al. 2006; Olker et al. 2018; Haselman et al. 2020). Using ethylenethiourea as the test substance, Opitz et al. (2006) demonstrated that both the determination of *X. laevis* developmental stage after a defined exposure period (exposure day 20 in their study) and the monitoring of tadpoles for the time required until forelimb emergence, which characterizes the transition of tadpoles from NF stage 57 to 58, consistently detected developmental delay and even complete inhibition of metamorphosis for concentrations ≥25 mg/L. These developmental effects were associated with a range of severe histological alterations in the thyroid gland, including a high prevalence of follicular cell hypertrophy and hyperplasia (Opitz et al. 2006). Olker et al. (2018) exposed tadpoles to a known iodotyrosine deiodinase inhibitor (3-nitro-L-tyrosine) from NF stage 50 until NF stage 62 or 1 wk after all the control tadpoles reached NF stage 62, whichever came first. In that study, whereas the time to NF stage 58 was not affected by the 3-nitro-L-tyrosine treatment, time to NF stage 62 was significantly greater at the top 2 treatment levels, with 20 and 60% of the tadpoles in these treatment levels, respectively, not able to reach NF stage 62. Signs of asynchronous development were also noted among 3-nitro-L-tyrosine–exposed tadpoles, with the observation of some tissues progressing as expected (e.g., adult-like skin pigmentation), whereas other tissues did not develop along expected timelines (e.g., retention of gills), indicating that that EAMA study design was able to detect asynchronous development.

Delay in metamorphosis is one of the primary apical outcomes of thyroid toxicity in amphibians. It is sensitive both to effects on the thyroid and to a higher level endpoint, which is potentially relevant for ecological risk assessments. In the standard, fixed-time design AMA, this is evaluated at the NF stage after 21 d of exposure, whereas in the EAMA, the fixed-stage design results in an endpoint expressed as the time to reach a particular stage (e.g., NF stage 62). These are 2 different data types, categorical and continuous, respectively, with different statistical methods required for analysis. Given the importance of this endpoint, the sensitivity of the fixed-stage design should be considered. Monte Carlo simulations were run using R statistical software (R Core Team 2020) comparing delay of development endpoints in both study designs. Simulated fixed-stage AMA studies were randomly generated using parameters derived from previous control studies (see details in the Supplemental Data, S2). Delays in time to NF stage 62 were imposed on the 2 highest treatment levels, and stage at day 21 was estimated by assuming proportional stage durations. The results indicate that the fixed-stage design resulted in greater statistical power to detect metamorphic delays than the fixed-time design. For example, with a 1-d delay in time to NF stage 62, a significant difference was detected in the fixed-stage analysis in 41.5% of the simulations, but only in 13.5% using the fixed-time data. It should also be noted that interpreting an increase in time to a specific stage is easier than a shift in stage at a specific day, and may be more meaningful from an ecological perspective.

There are some important advantages to the EAMA design, as previously discussed. There is no confounding effect of stage on morphometric measurements at test termination. The need to censor data based on exceeding NF stage 60 is eliminated, and thus statistical power increases (more animals can be included in the calculations). In the case of histopathological analysis, this design ensures that the required stage-matched samples are available at test termination. Some additional considerations resulting from the differences in experimental

methodology between the AMA and EAMA designs are as follows. 1) Tadpoles are assessed for developmental stage daily after day 21 (up to 10–20 additional d compared with the standard AMA, in which NF stage 62 is used as the terminal stage. 2) Consequently, study duration can be 10 to 20 d longer; such flexible study durations, as opposed to a fixed test duration, also exists, for example, in OECD test guideline 210 (Organisation for Economic Co-operation and Development 2013), in which the test duration differs depending on the test species (e.g., trout vs fathead minow), or in OECD test guidelines 218 and 219 (Organisation for Economic Co-operation and Development 2004b, 2004c), in which the test duration is in the range of 20 to 28 d for *Chironomus riparius* and *C. yoshimatsui*, and 28 to 65 d for *C. tentans*, to accommodate for variation in the developmental time of the different species. 3) Study terminates over several days rather than on a single day. 4) All surviving individuals are exposed throughout metamorphosis, although the exposure duration may be longer or shorter for some individuals. 5) There is the potential for growth differences mediated by density-dependent processes among treatments. 6) Food adjustment is recommended after termination sampling begins. 7) Some tadpoles may not reach NF stage 62 if there are treatments that impede the completion of metamorphosis. In this case, termination of the EAMA can be planned for a specific time based on, for example, 1 wk after all the control tadpoles have reached NF stage 62.

Two of the performance criteria evaluated at day 21 (test termination) in the AMA understandably lack applicability to the fixed-stage design. These are: 1) NF stage 57 is the minimum median developmental stage in the control at day 21; and 2) the 10th and 90th percentiles of the developmental stage distribution should not differ by >4 stages at day 21.

These criteria would need to be revisited. For instance, because NF stage 58 tadpoles are easily identifiable due to forelimb emergence, a noninvasive determination of the number of tadpoles at NF stage 58 or higher on day 21 in the EAMA would provide a reasonably reliable evaluation to satisfy the standard AMA criteria of the minimum median NF stage of the control tadpoles being ≥57 at day 21. All other performance criteria and all test acceptability/validity criteria used in the AMA are equally applicable to the EAMA.

CONCLUSIONS

Flexible study durations are commonly implemented in OECD test guidelines (e.g., 218, 219, and 443; Organisation for Economic Co-operation and Development 2004b, 2004c, 2018b) because they allow a more comprehensive assessment of potential effects of chemicals on the development of organisms. Use of the EAMA to evaluate thyroid activity addresses some important disadvantages of the original AMA test guideline (US Environmental Protection Agency 2013a, 2013b). Specifically, data analysis and interpretation of the fixed-NF stage 62 design of the EAMA become more straightforward by avoiding the confounding morphometric measurements and thyroid histopathology caused by varying developmental stages at termination. This also negates the need to censor data based on exceeding NF stage 60, thus preserving more animals to be used in the statistical analysis. The fixed-stage design resulted in greater statistical power to detect metamorphic delays than the fixed-time design. A demonstration study terminating at NF stage 58 and published data from studies terminating at NF stage 62 (see Olker et al. 2018; Haselman et al. 2020) suggest that the extended test design has similar or even higher sensitivity to that demonstrated in the AMA validation effort. The NF stage 62 EAMA is longer in duration and more complex to implement, but provides a more scientifically robust design.

The current LAGDA design, validation, and implementation have not demonstrated appropriate strength of this assay for use in determing T-mediated adversity. In addition, the LAGDA is an animal-intensive assay, using at least 480 animals (excluding a range-finder and potential repeats; Lagadic et al. 2019). We have shown that modification of the AMA to a fixed-stage design improves the performance and interpretation of the assay for assessing thyroid-related activity. From an animal welfare perspective, the EAMA with a negative control and 3 treatment levels would use 30% fewer animals than the LAGDA (320 vs 480 in the EAMA vs the LAGDA, respectively). With 4 treatment levels, as recommended in the LAGDA, the use of animals in the EAMA is reduced by 17% (400 vs 480 in the EAMA vs the LAGDA, respectively). Another important consideration is that the EAMA is 3 to 4 times shorter than the LAGDA, thus reducing animal suffering. Also, due to the thorough validation of the AMA, the failure rate of the EAMA is lower than that of the LAGDA (Lagadic et al. 2019), necessitating fewer repeats.

By extending the AMA to NF stage 62, thereby increasing the exposure duration, the data generated on apical endpoints are sufficient to evaluate T-mediated adverse effects that are potentially relevant at the population level. A positive EAMA would therefore be sufficient to draw conclusions on T-mediated endocrine disruption when accompanied by additional mechanistic evidence of T-mediated endocrine activity from in vitro or in vivo amphibian embryo assays, and in this case, a LAGDA would not be needed. The LAGDA remains of further interest should longer term investigations of possible adversity with regard to sexual development (mediated by endocrine-active substances modalities) be necessary. However, reproduction assessment or life-cycle testing is currently not included in the LAGDA study design. This is better addressed by higher tier testing in fish, which should then include specific thyroid-related endpoints. Currently, there is a major effort funded through the European Union Horizon 2020 program, the EURION cluster of 8 research projects (European Cluster to Improve Identification of Endocrine Disruptors 2020) focusing on different aspects of screening and testing methodologies. Within this cluster, the Endocrine Guideline Optimisation (ERGO) project is investigating additional thyroid-specific endpoints that could potentially be included in existing fish test guidelines (Holbech et al. 2020). This would offer the possibility to address T-mediated effects in studies in which growth, sexual development, and reproduction can be simultaneously evaluated,

compared with amphibian test guidelines that only address growth, metamorphosis, and sexual development.

The data we present illustrate the added scientific value of the EAMA. Consequently, regulators and validating bodies such as the OECD should consider revising the existing amphibian test guidelines and endocrine disrupter assessment strategies to integrate the advantages of fixed-stage designs for the identification of T-mediated endocrine activity and adversity.

REFERENCES

Allran JW, Karasov WH. 2000. Effects of atrazine and nitrate on Northern leopard frog (*Rana pipiens*) larvae exposed in the laboratory from posthatch through metamorphosis. *Environ Toxicol Chem* 19:2850–2855.

Ankley GT, Tietge JE, DeFoe DL, Jensen KM, Holcombe GW, Durhan EJ, Diamond SA. 1998. Effects of ultraviolet light and methoprene on survival and development of *Rana pipiens*. *Environ Toxicol Chem* 17:2530–2542.

Coady K, Marino T, Thomas J, Currie R, Hancock G, Crofoot J, McNalley L, McFadden L, Geter D, Klecka G. 2010. Evaluation of the amphibian metamorphosis assay: Exposure to the goitrogen methimazole and the endogenous thyroid hormone L-thyroxine. *Environ Toxicol Chem* 29:869–880.

Coady KK, Lehman CM, Currie RJ, Marino TA. 2014. Challenges and approaches to conducting and interpreting the amphibian metamorphosis assay and the fish short-term reproduction assay. *Dev Reprod Toxicol* 101:80–89.

Dang Z. 2019. Endpoint sensitivity in Amphibian Metamorphosis Assay. *Ecotoxicol Environ Saf* 167:513–519.

Denver RJ. 2009. Endocrinology of complex life cycles: Amphibians. In Pfaff DW, Arnold AP, Etgen AM, Fahrbach SE, Rubin RT, eds, *Hormones, Brain and Behavior*, 2nd ed. Academic, Cambridge, MA, USA, pp 707–745.

European Chemicals Agency and European Food Safety Authority with the technical support of the Joint Research Centre, Andersson N, Arena M, Auteri D, Barmaz S, Grignard E, Kienzler A, Lepper P, Lostia AM, Munn S, Parra Morte JM, Pellizzato F, Tarazona J, Terron A, Van der Linden S 2018. Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009. *EFSA J* 16:5311.

European Cluster to Improve Identification of Endocrine Disruptors. 2020. New testing and screening methods to identify endocrine-disrupting chemicals (EDCS). [cited 2021 April 9]. Available from: https://eurion-cluster.eu/

European Commission. 2017. Commission Delegated Regulation (EU). 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. *Official J Eur Union* L301:1–5.

European Commission. 2018. Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) No 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting properties. *Official J Eur Union* L101:33–36.

Fort DJ, Mathis MB, Pawlowski S, Wolf JC, Peter R, Champ S. 2017. Effect of triclosan on anuran development and growth in a larval amphibian growth and development assay. *J Appl Toxicol* 37:1182–1194.

Glaberman S, Kiwiet J, Aubee CB. 2019. Evaluating the role of fish as surrogates for amphibians in pesticide ecological risk assessment. *Chemosphere* 235:952–958.

Goleman WL, Carr JA. 2006. Contribution of ammonium ions to the lethality and antimetamorphic effects of ammonium perchlorate. *Environ Toxicol Chem* 25:1060–1067.

Grim KC, Wolfe M, Braunbeck T, Iguchi T, Ohta Y, Tooi O, Wolf DC, Tietge J. 2009. Thyroid histopathology assessments for the amphibian metamorphosis assay to detect thyroid-active substances. *Toxicol Pathol* 37:415–424.

Green JW, Springer TA, Holbech H. 2018. *Statistical Analysis of Ecotoxicity Studies*. John Wiley & Sons, Hoboken, NJ, USA.

Haselman JT, Kosiana PA, Korte JJ, Olmstead AW, Iguchi T, Johnson RD, Degitz SJ. 2016. Development of the Larval Amphibian Growth and Development Assay: Effects of chronic 4-tert-octylphenol or 17β-trenbolone exposure in *Xenopus laevis* from embryo to juvenile. *J Appl Toxicol* 36:1639–1650.

Haselman JT, Olker JH, Kosian PA, Korte JJ, Swintek JA, Denny JS, Nichols JW, Tietge JE, Hornung MW, Degitz SJ. 2020. Targeted pathway-based *in vivo* testing using thyroperoxidase inhibition to evaluate plasma thyroxine as a surrogate metric of metamorphic success in model amphibian *Xenopus laevis*. *Toxicol Sci* 175:236–250.

Holbech H, Matthiessen P, Hansen M, Schüürmann G, Knapen D, Reuver M, Flamant F, Sachs L, Kloas W, Hilscherova K, Leonard M, Arning J, Strauss V, Iguchi T, Baumann L. 2020. ERGO: Breaking down the wall between human health and environmental testing of endocrine disrupters. *Int J Mol Sci* 21:2954.

Lagadic L, Bender K, Burden N, Salinas ER, Weltje L. 2019. Recommendations for reducing the use of fish and amphibians in endocrine-disruption testing of biocides and plant protection products in Europe. *Integr Environ Assess Manag* 15:659–662.

Miyata K, Ose K. 2012. Thyroid hormone-disrupting effects and the amphibian metamorphosis assay. *J Toxicol Pathol* 25:1–9.

Nieuwkoop PD, Faber J, eds. 1994. *Normal Table of* Xenopus laevis *(Daudin)*. Garland, New York, NY, USA.

Organisation for Economic Co-operation and Development OECD. 2004a. Detailed review paper on amphibian metamorphosis assay for the detection of thyroid active substances. *OECD Series on Testing and Assessment, No. 46*. Paris, France.

Organisation for Economic Co-operation and Development. 2004b. Test No. 218: Sediment-water chironomid toxicity using spiked sediment. *OECD Guidelines for the Testing of Chemicals*, Section 2. Paris, France.

Organisation for Economic Co-operation and Development. 2004c. Test No. 219: Sediment-water chironomid toxicity using spiked water. *OECD Guidelines for the Testing of Chemicals*, Section 2. Paris, France.

Organisation for Economic Co-operation and Development. 2007a. Final report of the validation of the amphibian metamorphosis assay: Phase 2—Multi-chemical interlaboratory study. *Series on Testing and Assessment No. 77*. ENV/JM/MONO(2007)24. Paris, France.

Organisation for Economic Co-operation and Development. 2007b. Final report of the validation of the amphibian metamorphosis assay for the detection of thyroid active substances: Phase 1—Optimisation of the test protocol. *Series on Testing and Assessment No. 76*. ENV/JM/MONO(2007)23. Paris, France.

Organisation for Economic Co-operation and Development. 2007c. Guidance document on amphibian thyroid histology. *Series on Testing and Assessment No. 82*. ENV/JM/MONO(2007)31. Paris, France.

Organisation for Economic Co-operation and Development. 2008a. Organisation for Economic Cooperation and Development. Report of the validation of the amphibian metamorphosis assay (phase 3). *Series*

2144

Environmental Toxicology and Chemistry, 2021;40:2135–2144—L.S. Ortego et al.

on Testing and Assessment No. 91. ENV/JM/MONO(2008)18. Paris, France.

Organisation for Economic Co-operation and Development. 2008b. Report of the validation peer review for the amphibian metamorphosis assay and agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the follow-up of this report. *OECD Series on Testing and Assessment No. 92*. ENV/JM/MONO(2008)19. Paris, France.

Organisation for Economic Co-operation and Development. 2009. Test No. 231: Amphibian metamorphosis assay. *OECD Guidelines for the Testing of Chemicals*, Section 2. Paris, France.

Organisation for Economic Co-operation and Development. 2013. Test No. 210: Fish, early-life stage toxicity test. *OECD Guidelines for the Testing of Chemicals*, Section 2. Paris, France.

Organisation for Economic Co-operation and Development. 2015. Test No. 241: The larval amphibian growth and development assay (LAGDA). *OECD Guidelines for the Testing of Chemicals*, Section 2. Paris, France.

Organisation for Economic Co-operation and Development. 2018a. Revised guidance document 150 on standardised test guidelines for evaluating chemicals for endocrine disruption. *OECD Series on Testing and Assessment*. Paris, France.

Organisation for Economic Co-operation and Development. 2018b. Test No. 443: Extended one-generation reproductive toxicity study. *OECD Guidelines for the Testing of Chemicals*, Section 4. Paris, France.

Olker JH, Haselman JT, Kosian PA, Donnay KG, Korte JJ, Blanksma C, Hornung MW, Degitz SJ. 2018. Evaluating iodide recycling inhibition as a novel molecular initiating event for thyroid axis disruption in amphibians. *Toxicol Sci* 166:318–331.

Opitz R, Braunbeck T, Bögi C, Pickford DB, Nentwig G, Oehlmann J, Tool O, Lutz I, 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, 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.

Opitz R, Schmidt F, Braunbeck T, Wuertz S, Kloas W. 2009. Perchlorate and ethylenethiourea induce different histological and molecular alterations in a non-mammalian vertebrate model of thyroid goitrogenesis. *Mol Cell Endocrinol* 298:101–114.

Ortego L, Olmstead A, Banman C, Hall T. 2014. Improving the Amphibian Metamorphosis Assay: Experience with a time-to-stage design. Proceedings, 35th SETAC North America Annual Meeting, Vancouver, BC, Canada, 9–13 November 2014.

Pawlowski S, Dammann M, Weltje L, Champ S, Mathis M, Fort DJ. 2019. Is normalized hindlimb length measurement in assessment of thyroid disruption in the amphibian metamorphosis assay relevant? *J Appl Toxicol* 39:1164–1172.

Pickford DB. 2010. Screening chemicals for thyroid-disrupting activity: A critical comparison of mammalian and amphibian models. *Crit Rev Toxicol* 40:845–892.

Pleus RC, Corey LM. 2018. Environmental exposure to perchlorate: A review of toxicology and human health. *Toxicol Appl Pharmacol* 358:102–109.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Saka M, Tada N. 2021. Acute and chronic toxicity tests of systemic insecticides, four neonicotinoids and fipronil, using the tadpoles of the western clawed frog *Silurana tropicalis*. *Chemosphere* 270:129418.

Schlecht C, Klammer H, Jarry H, Wuttke W. 2004. Effects of estradiol, benzophenone-2 and benzophenone-3 on the expression pattern of the estrogen receptors (ER) alpha and beta, the estrogen receptor-related receptor 1 (ERR1) and the aryl hydrocarbon receptor (AhR) in adult ovariectomized rats. *Toxicology* 1–2:123–130.

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.

Seidlová-Wuttke D, Jarry H, Wuttke W. 2004. Pure estrogenic effect of benzophenone-2 (BP2) but not of bisphenol A (BPA) and dibutylphtalate (DBP) in uterus, vagina and bone. *Toxicology* 1–2:103–112.

Tidepool Scientific Software. 2000–2012. CETIS Ver 1.8.7.4. McKinleyville, CA, USA.

Tietge JE, Holcombe GW, Flynn KM, Kosian PA, Korte JJ, Anderson LE, Wolf DC, Degitz SJ. 2005. Metamorphic inhibition of *Xenopus laevis* by sodium perchlorate: Effects on development and thyroid histology. *Environ Toxicol Chem* 24:926–933.

Tietge JE, Butterworth BC, Haselman JT, Holcombe GW, Hornung MW, Korte JJ, Kosian PA, Wolfe M, Degitz SJ. 2010. Early temporal effects of three thyroid hormone synthesis inhibitors in *Xenopus laevis*. *Aquat Toxicol* 98:44–50.

US Environmental Protection Agency. 2009. Endocrine Disruptor Screening Program test guideline 890.1100: Amphibian metamorphosis (frog). EPA 740-C-09-002. Office of Prevention, Pesticides, and Toxic Substances. Washington, DC.

US Environmental Protection Agency. 2013a. Scientific review of the Endocrine Disruptor Screening Program (EDSP); tier 1 screening assay and battery performance. EPA-HQ-OPP-2013-0075-0023. Washington, DC.

US Environmental Protection Agency. 2013b. Endocrine Disruptor Screening Program SAP review of EDSP tier 1 screening: Assay and battery performance. EPA-HQ-OPP-2013-0075-0003. Washington, DC.

US Environmental Protection Agency. 2013c. Validation of the larval amphibian growth and development assay: Integrated summary report. Washington, DC.

US Environmental Protection Agency. 2015. Endocrine disruptor screening program test guideline 890.2300: Larval Amphibian Growth and Development Assay (LAGDA). EPA 740-C-15-001. Office of Chemical Safety and Pollution Prevention, Washington, DC.

Vitt LJ, Caldwell JP, eds. 2013. *Herpetology—An Introductory Biology of Amphibians and Reptiles*, 4th ed. Elsevier, New York, NY, USA.

Wheeler JR, Weltje L, Green J. 2014. Mind the gap: Concerns using endpoints from endocrine screening assays in risk assessment. *Chemosphere* 69:289–295.

World Health Organization, International Programme on Chemical Safety. 2002. Global assessment of the state-of-the-science of endocrine disruptors. Geneva, Switzerland. [cited 2021 January 15]. Available from: https://www.who.int/foodsafety/publications/chemical-food/en/

Xenbase. 1994. *Xenopus laevis* stage series: Complete. [cited 2021 January 21]. Available from: https://www.xenbase.org/anatomy/alldev.do