Approved: 25 November 2025

DOI: 10.2903/j.efsa.2025.9815

efsa JOURNAL

STATEMENT

The assessment for potential thyroid-mediated endocrine disruption in amphibians: Clarification on the use of new methods and on the interpretation of changes in thyroid histology

European Food Safety Authority (EFSA) | Lisa Baumann | Alice Baynes | Susanne Hougaard Bennekou | Kevin Crofton | Zhichao Dang | Sigmund Degitz | Jean-Baptiste Fini | Mary Gilbert | Henrik Holbech | Andrea Terron | Majorie Van Duursen | Emily McVey | Maria Arena | Aude Kienzler | Simone Rizzuto | Laura Villamar Bouza | Martin Wilks

Correspondence: Ask a Question

The declarations of interest of all scientific experts active in EFSA's work are available at https://open.efsa.europa.eu/experts.

Abstract

Amphibians (specifically *Xenopus laevis*) are used as the model species to assess potential endocrine-disrupting properties in non-mammalian species through thyroid modality. The amphibian metamorphosis assay is the most frequently available test. Attempts have been made to modify this protocol in order to make it more fit for purpose and overcome potential limitations. In light of these developments, EFSA, with the support of the Working Group on Endocrine Disruptors, under the auspices of a self-task mandate here endeavours to clarify the pros and cons of newly proposed amphibian protocols when compared with the standard guideline tests. Moreover, recommendations to facilitate the interpretation of findings in relation to changes in thyroid histopathology have been included.

KEYWORDS

amphibians, endocrine disruptors, histopathology, study protocols, thyroid

CONTENTS

Abstract 1
Summary 1
1. Introduction 3
1.1. Background and Terms of Reference as provided by the requester 3
1.2. Interpretation of the Terms of Reference 3
2. Use of a new method to conclude on the ED properties of a substance 4
2.1. Available protocol for investigating endocrine-disrupting properties in amphibians 4
2.1.1. Amphibian metamorphosis assay 5
2.1.2. The larval amphibian growth and development assay 5
2.1.3. The extended amphibian metamorphosis assay 6
2.1.4. The larval amphibian toxicity test 7
2.2. Statistical methods and related power analysis 7
2.2.1. Main outcomes 7
2.3. Conclusion on the contribution of new/amended study designs with amphibians to the testing strategy for concluding on ED properties through the T-modality (ToR1) 9
3. Interpretation of change in amphibian thyroid histology 10
3.1. Introduction 10
3.2. Method 11
3.3. Results 11
3.4. Interpretative considerations on thyroid histopathology 12
3.4.1. Illustrative examples 16
3.5. Conclusions (ToR2) 18
4. Recommendations 18
5. Future research needs 19
Abbreviations 19
Requestor 20
Question number 20
Copyright for non-EFSA content 20
References 20
Annex A 23

SUMMARY

According to Commission Regulation (EU) 2018/605 laying down the scientific criteria for the hazard identification of pesticides having endocrine-disrupting properties, scientific data generated in accordance with internationally agreed study protocols should be considered.

The ECHA/EFSA guidance on hazard identification of endocrine disruptors (ED) recommends a test according to OECD 231 (amphibian metamorphosis assay (AMA)) or, in specific circumstances, a test according to OECD test guideline (TG) 248 (*Xenopus* Eleutheroembryonic Thyroid Assay (XETA)) for the full investigation of endocrine activity through thyroid (T) modality in non-mammalian species. T-mediated endocrine adversity is fully investigated when a test according to OECD TG 241 (larval amphibian growth and development assay (LAGDA)) is available. In 2021, Ortego et al. (2021) published a protocol in which the standard AMA (TG 231) was proposed to be extended until animals reached stage Nieuwkoop and Faber (NF) 62, calling it the extended amphibian metamorphosis assay (EAMA). Several studies using this protocol have already been submitted in dossiers for the approval of pesticide active substances.

In addition, in 2024, the US EPA presented another amphibian protocol, the larval amphibian toxicity test (LATT), to the OECD Validation Management Group for ecotoxicity testing (OECD VMGeco).

Following the publication by Ortego et al. (2021) and the submission of studies in line with its protocol in pesticide dossiers, along with the ongoing work at OECD, EFSA compared the new proposed methods to the available standard protocols, i.e. AMA and LAGDA. This comparison should include a power analysis of the main diagnostic parameter(s) and growth parameters and consider the advantages and disadvantages of available standard and non-standard protocols. Moreover, EFSA provided interpretative recommendations related to changes in the thyroid histology of amphibians.

The main difference between the available amphibian testing protocols considered in this work is that, while the AMA proposes a fixed-time protocol, the LAGDA, the EAMA and the LATT propose a fixed developmental stage protocol for assessing endocrine activity and adversity through the T-modality. The use of a fixed-stage protocol overcomes some of the main limitations of the AMA, which renders non-synchronised animals for thyroid histology, thereby confounding interpretation.

For physiological reasons, in all available amphibian study designs, the larval stage must be used for an assessment of key diagnostic parameters related to the thyroid pathway. As the LAGDA is extended until completion of metamorphosis, it may be better suited for detecting population-relevant effects. However, from a biological and data analysis perspective, it remains unclear if a LAGDA is necessary to confirm or refute concerns identified by a positive AMA. The LATT, although still under development and thus not considered in the analysis of statistical power presented below, could be a valid alternative to both the AMA and LAGDA. It could provide a valuable input to the interpretation of potential thyroid-related findings, being the only amphibian study design proposing the inclusion of thyroid hormone (TH) level measurements. In addition, it overcomes some of the challenges of the LAGDA, e.g. potential issues in meeting the validity criteria due to the high mortality of early life stages and the use of many more animals.

The ECHA/EFSA guidance is very clear on the importance of using weight of evidence (WoE) considerations for the hazard identification of EDs. In line with that, this statement also recommends that changes in thyroid histology should be interpreted together with other parameters in order to understand whether a pattern of effects related to an interference with the hypothalamic–pituitary–thyroid (HPT) axis has been observed. The inclusion of TH measurements would greatly improve the overall WoE. Moreover, transcriptomic analyses of key genes of the HPT axis could provide important mechanistic information for the interpretation of developmental and thyroid histology effects.

1 | INTRODUCTION

1.1 | Background and Terms of Reference as provided by the requester

With this internal mandate EFSA is requested to develop a statement to clarify two main aspects:

1. Use of non-standard methods with amphibians to conclude on the endocrine-disrupting properties through the T-modality of a pesticide active substance by its considering advantages and disadvantages compared with available standard protocols.
2. Interpretation of changes in amphibian thyroid histology.

1.2 | Interpretation of the Terms of Reference

According to Commission Regulation (EU) 2018/605 laying down the scientific criteria for the hazard identification of pesticides having endocrine-disrupting properties through the T-modality in non-mammalian species, scientific data generated in accordance with internationally agreed study protocols should be considered, i.e. amphibian metamorphosis assay (AMA) and larval amphibian growth and development assay (LAGDA). In 2021, Ortego et al. (2021) published a protocol in which the standard AMA (test guideline (TG) 231; OECD, 2009) was extended until animals reached stage Nieuwkoop and Faber (NF) 62 (Nieuwkoop & Faber, 1994). In this protocol, therefore, a fixed developmental stage approach is proposed as an alternative to the fixed-time protocol suggested in the AMA. A standard fixed-stage-like protocol[1], however, already exists, i.e. the LAGDA (according to OECD TG 241; OECD, 2015a), which provides at interim sampling an assessment of the time to reach stage NF 62. Moreover, there is a new approach proposed by the US EPA, the larval amphibian toxicity test (LATT), which has been listed in the OECD validation workplan.

ToR1 aims to compare new proposed methods with the available standard AMA and LAGDA test protocols by highlighting the advantages and disadvantages of each method. In particular, the focus should be to provide:

- a description of the different protocols and their relevance for concluding on the potential for endocrine disruption through the T-modality and for identification of adversity relevant at the population level;
- appropriate statistical methods for the different parameters and the power of the test to detect the effects on growth and developmental parameters.

ToR2 aims to provide recommendations on the interpretation of changes in the thyroid histology of amphibians, by considering the following:

- Normal physiology of the thyroid in the control to set a baseline.
- Expected pattern of effects, e.g. continuum between follicular cell (FC) hypertrophy and hyperplasia, peak of the two findings, etc.
- Confounding factors, e.g. systemic toxicity, feeding, and environmental conditions, including water quality.

2 | USE OF A NEW METHOD TO CONCLUDE ON THE ED PROPERTIES OF A SUBSTANCE

2.1 | Available protocol for investigating endocrine-disrupting properties in amphibians

Two standard protocols are available for investigating endocrine activity and adversity in amphibians through the T-modality: the AMA (OECD 231, USEPA 890.1100; US EPA, 2009) and the LAGDA (OECD 241; OCSPP 890.2300; US EPA, 2015). In 2021, Ortego et al. (2021) proposed to extend the duration of the AMA (extended AMA, EAMA) until animals reach stage NF 62. In addition, the US EPA has also proposed a modified protocol (LATT) in which exposure was extended until the animals completed metamorphosis (stage NF 66); the overview of available test methods, considered in this analysis, life stages covered and duration of exposure are reported in Figure 1. Although other methods may be available, the analysis only considered the EAMA and the LATT as several studies following the EAMA protocol were available and the LATT was included in the OECD workplan.

An additional standard protocol, the *Xenopus* eleutheroembryonic thyroid assay (XETA; OECD 248; OECD, 2019) is also available. However, the XETA is only aimed at detecting (anti)thyroid-active compounds based on measurements of the induction of fluorescence compared with controls. Therefore, as the endpoints measured are not comparable with those of the other available protocols and only give an indication of the potential of a compound to interact with the HPT axis, it will not be considered further in this analysis.

[1]The LAGDA is referred to as a fixed-stage like protocol as the investigation of T-mediated parameters is conducted at a fixed stage, NF 62, although when considering the whole duration, it is not a fixed-stage protocol.

In the ECHA/EFSA Guidance for the hazard identification of endocrine disruptors (ED), a study in line with the AMA is recommended to fully investigate endocrine activity. If the AMA is negative, a conclusion on the lack of ED properties through the T-modality can be drawn based on the lack of endocrine activity in a sufficiently investigated data set. If this study is positive, an MoA analysis should be postulated and the need for further investigating adversity should be considered, for example, by performing a study in line with the LAGDA.

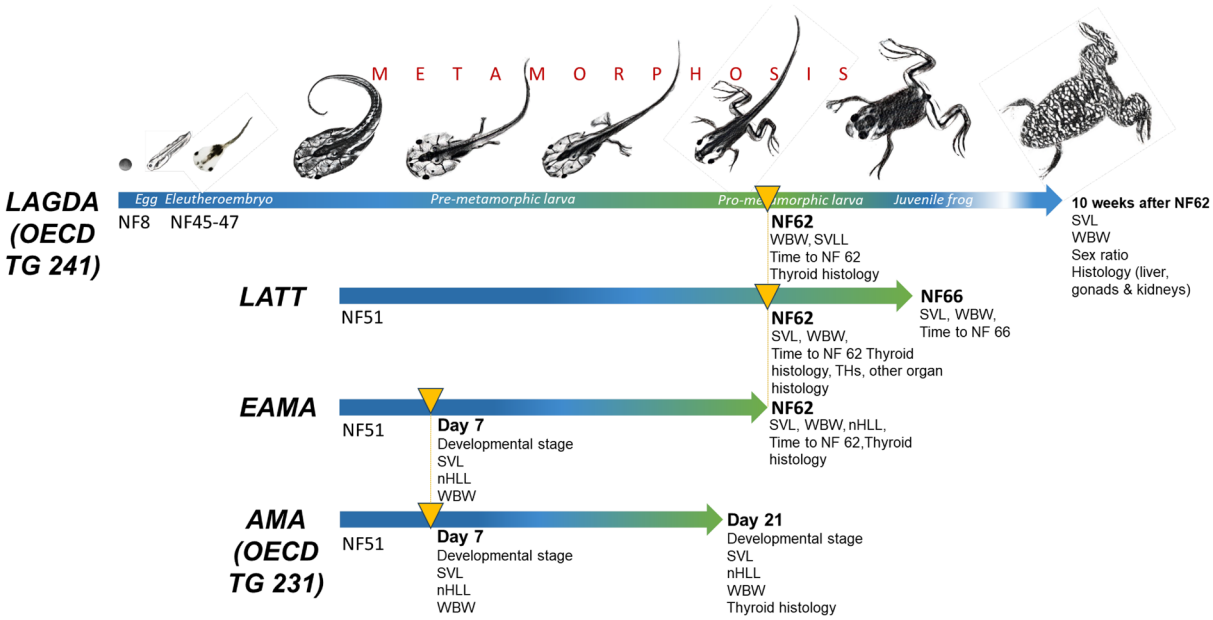

FIGURE 1 Duration of exposure and life stages covered in the available protocol with *Xenopus* sp. considered in this analysis (adapted from Couderq et al., 2020).

2.1.1 | Amphibian Metamorphosis Assay

The AMA (OECD TG 231; USEPA 890.1100) is a protocol placed at level 3 of the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters (OECD, 2018). It is considered a screening assay used to filter out substances that do not show any potential concern for disruption of the hypothalamic–pituitary–thyroid (HPT) axis.

Xenopus laevis, the African clawed frog, is tested at stage NF 51. Generally, three different concentrations of the tested substance and four replicates are used, each with 20 animals. Animals are fed Sera Micron or any other diet that has demonstrated a similar performance in such a study. If the study is performed under static conditions, the food regimen could be reduced to 50% of the daily food ration of Sera Micron. Validity and performance criteria are included in the relevant OECD/US EPA test guidelines (OECD TG 231; US EPA 890.1100). Recommended iodide content of test water should range from 0.5 to 10 µg/L. For the list of measured parameters, see Table 1.

The AMA proposes a fixed-time protocol, i.e. the exposure duration is 21 days. Two samplings are recommended: an interim sampling at Day 7 when five animals per replicate are selected and a sampling at termination after 21 days of the remaining 15 tadpoles.

TABLE 1 List of parameters and time of assessment in the AMA

Parameter	Daily	Interim sampling	
		Day 7	Day 21
Mortality	✓		
Abnormalities (e.g. abnormal behaviours)	✓		
Developmental stage		✓	✓
Snout-to-vent length (SVL)		✓	✓
Hind limb length (HLL)		✓	✓
Normalised hind limb length (nHLL)		✓	✓
Thyroid histology			✓
Wet body weight (WBW)		✓	✓

According to the ECHA/EFSA Guidance, developmental stage, normalised HLL (nHLL) and thyroid histology are diagnostic parameters for the detection of substances having endocrine-disrupting properties through the T-modality. Other parameters, such as snout-to-vent length (SVL) and wet body weight (WBW), are sensitive to but not diagnostic of T-modality and, as such, they can also provide an indication of generalised toxicity. Mortality and abnormalities (malformations) provide an indication of toxic effects other than endocrine-mediated effects.

2.1.2 | The Larval Amphibian Growth and Development Assay

The LAGDA (OECD TG 241; OCSPP 890.2300) is placed at level 4 of the OECD conceptual framework (OECD, 2018).

Xenopus laevis, the African clawed frog, at stages NF 8–10 (0 days post hatch (dph)) is tested. Generally, a minimum of four different concentrations of the tested substance, four replicates for the tested concentration, in addition to eight replicates for control(s) are used, each with 20 animals. The feeding regimen is recommended in the related OECD guideline. However, this may be subject to change provided that the animals are developing and growing appropriately. Validity criteria are also specified. Recommended iodide content ranges from 0.5 to 10 μg/L. For the list of measured parameters, see Table 2.

In the LAGDA, animals are exposed until 10 weeks after controls reach NF stage 62 (~110 dph). Two samplings are recommended, an interim sampling at the larval stage when animals reach stage NF 62 and at termination. All the diagnostic parameters for the T-modality are planned to be investigated at the interim sampling, i.e. stage NF 62. Therefore, for those parameters, the LAGDA follows a fixed-stage protocol. It is important to note that the maximum time to reach stage NF 62 for control animals is ≤ 45 days post-fertilisation. It also includes parameters that can be useful for the evaluation of the endocrine-disrupting properties through the oestrogen, androgen and steroidogenesis modality.

TABLE 2 List of parameters and time of assessment in the LAGDA

Parameter	Daily	Interim sampling stage NF 62	Termination, juvenile sampling
Mortality	✓		
Abnormalities (e.g. abnormal behaviours)	✓		
Time to reach stage NF 62		✓	
Thyroid histology		✓	
Growth and length (WBW and SVL)		✓	✓
Liver Somatic Index (LSI)			✓
Genetic/phenotypic sex ratio			✓
Organ histology (kidney, liver, gonads)			✓
Vitellogenin (optional)			✓

2.1.3 | The Extended Amphibian Metamorphosis Assay

The proposal to extend the AMA until the animals reach stage NF 62 was published by Ortego et al. (2021). The protocol is similar to the AMA with the difference that the developmental stage (fixed-time endpoint) has been replaced by time to reach stage NF 62 (fixed-stage endpoint). Therefore, although, theoretically, Ortego proposed to extend the AMA, in practice the newly proposed protocol has commonalities with both the AMA and the LAGDA. For the list of measured parameters, see Table 3.

In Ortego et al. (2021), a documentation study in which tadpoles of *X. laevis* were exposed to sodium perchlorate was presented. This study, however, did not terminate at stage NF 62 but at stage NF 58 considering its straightforward identification related to forelimb emergence. Animals were fed with liquid frog brittle slurry (LBF); it was not reported whether feeding efficiency was comparable with Sera Micron. No validity criteria were proposed.

TABLE 3 List of parameters and time of assessment in the EAMA

Parameter	Daily	Interim sampling – day 7	Termination, stage NF 62
Mortality	✓		
Abnormalities (e.g. abnormal behaviours)	✓		
Developmental stage/time to reach NF 62		✓	✓
Snout-to-vent length (SVL)		✓	✓
Hind limb length (HLL)		✓	✓
Normalised hind limb length (nHLL)		✓	✓
Thyroid histology			✓
Wet body weight (WBW)		✓	✓
Total body length (optional)		✓	✓

2.1.4 | The Larval Amphibian Toxicity Test[2]

Recently, the US EPA has proposed a different protocol, the LATT, which was included in the OECD work plan for 2025.[3] The proposal aims to extend the study until animals have completed metamorphosis, i.e. stage NF 66. The protocol proposes testing five concentrations with four replicates for treatments and control. See Table 4 for the list of measured parameters.

TABLE 4 List of parameters and time of assessment in the LATT

Parameter	Daily	Interim sampling – stage NF 62	Termination – stage NF 66
Mortality	✓		
Abnormalities (e.g. abnormal behaviours)	✓		
Time to event		✓	✓
Snout-to-vent length (SVL)		✓	✓
Hind limb length (HLL)		–	–
Normalised hind limb length (nHLL)		–	–
Thyroid histology/plasma thyroid hormones		✓	
Wet body weight (WBW)		✓	✓
Kidney and liver histology		✓	

2.2 | Statistical methods and related power analysis

From a statistical perspective, it was unclear whether the fixed stage in the EAMA is to be preferred over the fixed time of the AMA. Therefore, to address this uncertainty, an analysis of the statistical power of all investigated parameters across the two different test designs was performed. In addition, to further clarify a long-standing debate in regulatory ecotoxicology, a similar analysis was performed comparing the AMA and the LAGDA. In summary, an analysis of the statistical power of the AMA, EAMA and LAGDA was conducted on continuous variables (i.e. growth parameters, total length, wet weight, SVL, hind limb length (HLL) and normalised hind limb length (nHLL)), on the developmental stage (for the AMA) and on the time to reach stage NF 62 (for the EAMA and LAGDA). The AMA ($n = 18$) and EAMA ($n = 9$) studies were retrieved from pesticide dossiers, while the LAGDA ($n = 5$) studies were sourced from open literature, as no studies had to date been submitted in the context of the ED assessment of pesticides. The methodology and the results of such comparative analyses are reported in detail in Rizzuto et al. (2025).

2.2.1 | Main outcomes

Rizzuto et al. (2025) concluded that the statistical power of the AMA was sufficient to detect a significant effect on growth parameters. The statistical power to detect effects (both advancement and delay) on the developmental stage, however, was not constant. This parameter in the AMA was significantly affected by an increase in variance and by the effect size (i.e. the number of stages affected by an effect), whereas the number of treatment groups inducing an effect had only a negligible impact. The statistical power of the EAMA was considered in line with the AMA and LAGDA tests. The EAMA had the edge on statistical power to detect the effects on some growth parameters for both the AMA (HLL and nHLL), and the LAGDA (wet weight, also lower than in the AMA for this variable). However, it slightly underperformed compared with the LAGDA for the time to reach NF 62. For this variable, a direct comparison with the developmental stage of the AMA could not be made, considering the different nature of the two variables (i.e. quantitative vs. ordinal). However, based on a more minimalistic approach, the power of the EAMA was considered comparable with that of the AMA.

The findings from the AMA are consistent with the power analysis in OECD (2006), Annex 5, and further reinforce the idea that this test is a screening study, designed to give a more 'qualitative' rather than 'quantitative' response. As such, the AMA is not recommended to be the setting of a no observed effect concentration (NOEC) suitable for use in an environmental risk assessment. These results also highlight the fact that a significant one-stage shift may already represent a

[2]The information reported reflects the status and knowledge based on the presentation given at the OECD VMGeco in October 2025.
[3]https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/testing-of-chemicals/work-plan-test-guidelines.pdf.

concern that should be addressed with higher tier experimental designs, unless the weight of evidence (WoE) indicates that the observed pattern may not be related to a disruption in the HPT axis. In addition, based on the results, harmonisation in the 'controllable' factors (e.g. animal source, rate and type of food, iodide content) is recommended to minimise intra- and inter-replicate variability, ensuring suitable control performance and augmenting statistical power. Furthermore, to address regulatory concerns and ensure full consideration of the results of the AMA in the regulatory process, applicants could consider presenting a statistical power analysis of the results of the developmental stage (following the methods reported in Rizzuto et al., 2025).

The higher power of the LAGDA versus the EAMA to detect changes in time to reach NF 62 was expected due to the higher number of biological replicates per treatment group. The LAGDA is significantly more animal intensive with a higher number of control replicates and the inclusion of more treatment levels. This makes the LAGDA better suited to provide concentration–response data for population-relevant adverse effects and environmental risk assessment purposes. At the same time, the magnitude of the observed difference in the LAGDA compared with the other in vivo tests was unexpectedly small, questioning its contribution to an overall WoE already supporting the identification of a substance as an ED.

Overall, although the EAMA may present some added values when compared with a standard AMA (i.e. inclusion of a time-to-event parameter), it is not clear how it contributes to an MoA if a positive response has already been identified in the AMA. However, performing an EAMA, in lieu of an AMA, may be considered given its advantages documented in the Rizzuto et al. (2025) analysis.

2.3 | Conclusion on the contribution of new/amended study designs with amphibians to the testing strategy for concluding on ED properties through the T-modality (ToR1)

As described above, the four test methods analysed (AMA, LAGDA, EAMA and LATT) showed a number of similarities but also some differences. Table 5 reports the analysed protocols with a comparison of the parameters measured.

TABLE 5 Biological parameters measured in the different protocols in the scope of the current analysis

	AMA (NF 51 for Day 21)			Extended AMA (from stage NF 51 until stage NF 62)			LAGDA (NF 8–10 for 16 weeks)			LATT (from stage NF 51 until stage NF 66)		
Biological endpoints	**Daily**	**Day 7**	**Day 21**	**Daily**	**Day 7**	**NF stage 62**	**Daily**	**NF stage 62**	**10 weeks after the median time to NF stage 62**	**Daily**	**NF stage 62**	**NF stage 66**
Mortality and abnormalities	X			X			X			X		
(Wet) body weight		X	X		X	X		X	X		X	X
Hind limb length		X	X		X	X						
Snout-to-vent length		X	X		X	X		X	X		X	X
Developmental stage		X	X		X							
Time to reach NF stage 62						X		X			X	
Time to reach NF stage 66												X
Thyroid histology			X			X		X			X	
Organ histology (gonad, reproductive duct, gonads, kidney, liver)									X		X	
Liver Somatic Index									X			
Sex ratio									X			
Plasma thyroid hormones (T_3, T_4)											X	
Vitellogenin (optional)									X			

Abbreviations: T_3, triiodothyronine hormone; T_4, thyroxine hormone.

The main difference across the testing protocols is that the AMA uses a fixed-time protocol, whereas the LAGDA, the EAMA and the LATT use a fixed-stage protocol for assessing endocrine activity and adversity through the thyroid modality. Moreover, there is a difference in the life stage covered in the different protocols as the exposure window varies between the LAGDA (earlier exposure, see Table 5) and all the other protocols considered in this analysis. There is currently no evidence on the relevance of exposure during amphibian embryogenesis for the detection of an endocrine disruptor through the T-modality, and thus, its relevance is currently uncertain.

The EAMA and the LATT may overcome some of the limitations identified in the AMA and the LAGDA. The time-to-event parameter included in both the EAMA and the LATT may allow employing alternative statistical tools (Haselman, Kosian, et al., 2016; Haselman, Sakurai, et al., 2016; Ortego et al., 2021; US EPA, 2013a, 2013b), which might be increasingly informative or powerful.

In the AMA, the animals are exposed from stage NF 51 to approximately stage NF 59 (21 days), and in the EAMA from stage NF 51 to stage NF 62 (until 95% of the surviving tadpoles reached NF stage 62 or up to 45 days post-fertilisation, whichever occurs first). In both the AMA and EAMA, three concentrations are used, making the two tests not suitable for endpoint derivation and use in risk assessment.

The LAGDA and EAMA both include time to reach stage NF 62 as the main diagnostic parameter. Both protocols also include thyroid histopathology in animals reaching stage NF 62. The LAGDA, however, utilises four or more concentrations, while the EAMA includes only three concentrations. The LAGDA is also prolonged for animals remaining after subsamples reaching stage NF 62 are collected, until completion of metamorphosis stage NF 66. These three aspects of the LAGDA, i.e. time-to-event endpoint, duration of exposure and experimental design, make the test more suitable for confirming whether an effect may be relevant at the level of the population.

The LATT also includes all three positive features of the LAGDA. Exposure in the LATT starts with animals at stage NF 51, but this is advantageous as exposing early life stages is technically more challenging both in terms of concentration to be tested to reach or approach the maximum tolerated concentration (MTC) and to meet the validity criteria. The LATT is also the only protocol that proposes to measure THs. This aspect strengthens the WoE and the potential conclusion on the interference of a compound with the HPT axis. Moreover, similarly to the LAGDA, the possible inclusion of additional parameters in the LATT improves understanding and better contextualisation of the potential change in T-mediated parameters in the WoE. These include an assessment of liver and kidney histopathology, providing a better characterisation of other toxic effects and target organ toxicity. Moreover, the extension of exposure beyond stage NF 62 to completion of metamorphosis may be crucial for some T-MoA. For example, for agents that target deiodinase inhibition, the extension of observations to stage NF 66 appears necessary for the correct interpretation of the findings. Inhibition of deiodinases is characterised by an acceleration at stage NF 62 followed by complete inhibition of metamorphosis at stage NF 66 (Dang, 2022; Haselman et al., 2022).

One of the key uncertainties of the EAMA and LATT is that animals may take different times to reach stage NF 62, and this means that they may be exposed for different durations compared with the control animals. Currently, it is uncertain how to consider this in the overall WoE and the potential impact on other parameters. Nevertheless, both the EAMA and the LATT are considered an enhancement of the AMA and the fixed-stage protocol proposed in both protocols is a more robust design than the fixed time in the AMA. This may be of particular importance for the assessment of thyroid histopathology. For this specific assessment, synchronised animals are considered to be better suited, as the selection of staged-matched animals may allow the minimisation of growth-specific differences that may hamper the correct interpretation of changes observed in thyroid histology. Furthermore, the censoring of animals at stages > NF60 in the AMA for growth parameters like SVL and WBW may have an impact on the overall power of the test. Therefore, those tests should be encouraged over the AMA, although a validation has not yet been conducted.

Although it was acknowledged that the validation of the LAGDA was limited to a few chemicals (US EPA, 2013c) (only one substance known to act through multiple ED MoA including a T-MoA), it was noted that the extended AMA suggested by Ortego et al. (2021) has not been validated either. The validation of the EAMA is expected to provide information on the most appropriate validity and performance criteria to be used, although the robustness of the EAMA protocol in comparison with the AMA is not questioned. The LATT has recently been discussed at the OECD level for possible validation.

Overall, all available study designs provide an assessment of key diagnostic responses for thyroid effects at the larval stage for clear physiological reasons. From a regulatory point of view, it is still debatable whether an LAGDA, LATT or EAMA is necessary to confirm or dismiss concerns identified by a positive AMA, as well as what is the added value of a LAGDA, when an AMA is already available and positive. It is also true that the LAGDA, which extends observations until completion of metamorphosis, may be better suited for detecting population-relevant effects for some specific MoA. However, the LATT, although still under development, could be a valid alternative as it not only extends observations to the completion of metamorphosis but also includes an assessment of THs.

In summary, when an AMA is available and positive (i.e. changes in thyroid histopathology accompanied by statistically significant effects on development), considering the outcome of the power analysis in Rizzuto et al. (2025), a regulatory conclusion on the ED properties through the T-modality is possible without the need for further data. However, a fixed-stage protocol, as in the EAMA, LATT and LAGDA, possibly complemented with TH measurements, is considered a better option. Both the EAMA and the LATT are less resource-intensive, use fewer animals and overcome the issue of early life stage mortality that can plague the LAGDA. It is also acknowledged that the choice of the testing strategy is ultimately up to applicants and the conclusion on ED properties always relies on the WoE approach.

3 | INTERPRETATION OF CHANGE IN AMPHIBIAN THYROID HISTOLOGY

3.1 | Introduction

Thyroid histology is one of the key diagnostic parameters to determine whether a substance causes perturbation of the HPT axis in vertebrates in general and in amphibians in particular. Although this measurement is recommended in all the available protocols with amphibians, its current assessment and interpretation are not straightforward for a number of reasons, e.g. very specific expertise, influence of growth and reporting of data.

Methods for performing such measurement and diagnostic criteria, including severity scores, are reported in Grim et al. (2009) and OECD guidances (OECD, 2007, 2015b, 2015c). The methods are based on the subjective assessment of structural changes in thyroid morphology by a pathologist. Quantitative assessments using imaging software have not (yet) been included.

Based on these documents, two of the four core diagnostic criteria (i.e. FC hypertrophy and hyperplasia) are to be scored using absolute criteria, whereas the other two (i.e. thyroid hypertrophy and thyroid atrophy) are scored relative to the control. Although these recommendations and the severity grading scale as suggested in the OECD guidance are consistently applied, the method is still somewhat subjective, and results across studies may be reported differently leading to potential misinterpretation by assessors examining different studies and contrasting different substances.

In general, it is considered extremely useful to first examine the range of response in control animals to set the baseline for the assessment of treated animals.

In amphibians, as in all vertebrates, feedback and compensatory mechanisms of the HPT axis keep the physiological range of circulating thyroid hormone (TH) levels quite narrow and therefore do not affect the normal physiology of the thyroid gland; see Figure 2.

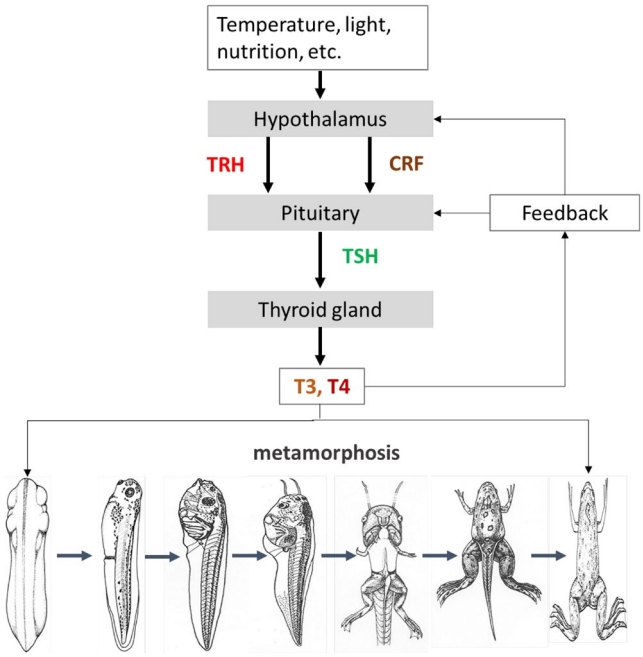

FIGURE 2 Thyroid signalling in amphibian metamorphosis (adapted from Orton and Tyler (2015) and Ikegami and Yoshimura (2017)).

As shown in Figure 2, in response to different types of stimuli (temperature, food, etc.), the hypothalamus releases thyrotropin-releasing hormone (TRH) into the portal circulation. After the TRH binds to the TRH receptor (TRHR) on the cell membrane of pituitary thyrotropes, it activates an intracellular signalling pathway and stimulates the secretion and synthesis of thyrotropin or thyroid-stimulating hormone (TSH). TSH is released into the blood and at the level of the thyroid gland, then it binds to the TSH receptor to stimulate the synthesis and release of THs. As serum TH levels rise, negative feedback on the pituitary and hypothalamus downregulates the synthesis and release of TSH. As also shown in Figure 2, in larval and adult anurans, TSH release may also be regulated by corticotropin-releasing factor (CRF) (Sternberg et al., 2011).

In larvae of anurans, metamorphosis is mediated by the HPT axis. The thyroid glands are not fully functional during premetamorphosis, and thus, development during this initial phase is generally not fully mediated by TH (stages NF 46–54). During prometamorphosis (stages NF 55–57), TH levels start to rise to initiate those morphological changes that are TH mediated. During metamorphic climax, the increase in TH synthesis leads to important changes, such as resorption of tail and gills, craniofacial and gut remodelling, and differentiation of the liver (stages NF 58–66). Although the feedback mechanisms, as described above should work similarly in vertebrates, including anurans, and therefore on the tested species *Xenopus laevis*, it is still unclear how TSH production and circulating TH concentrations rise concomitantly during

metamorphic climax in this species. The release of TSH by the pituitary gland has the consequence of increasing the occurrence of FC hypertrophy and FC hyperplasia, which represent increases in size and number, respectively, of the colloid-producing epithelial cells that line the multiple thyroid follicles. This means that, to a certain degree, FC hypertrophy and hyperplasia are expected to be observed in anurans approaching metamorphic climax and exposure to a thyroid hormone system-disrupting compound (THSDC) may exacerbate or reduce what is normally observed in control animals. This lack of mechanistic understanding underlines the need to include TH measurements in the test protocols to support the interpretation of observed effects. Transcriptomic analyses could similarly be considered.

3.2 | Method

In a manner similar to Wolf et al. (2021, 2024), EFSA analysed thyroid histology of control animals from 16 available AMAs (period 2012–2023), 7 EAMAs (period 2021–2022) and 5 LAGDA studies (2016–2018). The AMA and EAMA available in dossiers were used, whereas LAGDA data were retrieved from five published literature reports from two laboratories (Fort et al., 2017; Haselman et al., 2018; Haselman, Kosian, et al., 2016; Haselman, Sakurai, et al., 2016). To avoid any variability in the histology evaluation, only histopathology reports performed by a single laboratory and pathologist were considered for the AMA and EAMA. Overall, data from 320 control animals were analysed from AMA studies, 140 from EAMA and 125 animals from LAGDA. Analyses were limited to the occurrence of FC hypertrophy and hyperplasia, the two most commonly reported findings.

For each study, the mean percentage of occurrence of FC hypertrophy and hyperplasia was calculated by dividing the occurrence of the diagnosis by the total number of control animals used in the study. The analysis was repeated for the three different protocols.

3.3 | Results

The data collected are presented in Figures 3–5. Although the data set is limited, especially for the EAMA and LAGDA, FC hypertrophy was the most common observation, with a mean frequency of 69% in control animals from AMA studies, 77.8% in control animals of EAMA and 72.9% in LAGDA. FC hyperplasia was less frequently observed with 31.2%, 5% and 17.4% in control animals for the AMA, EAMA and LAGDA, respectively.

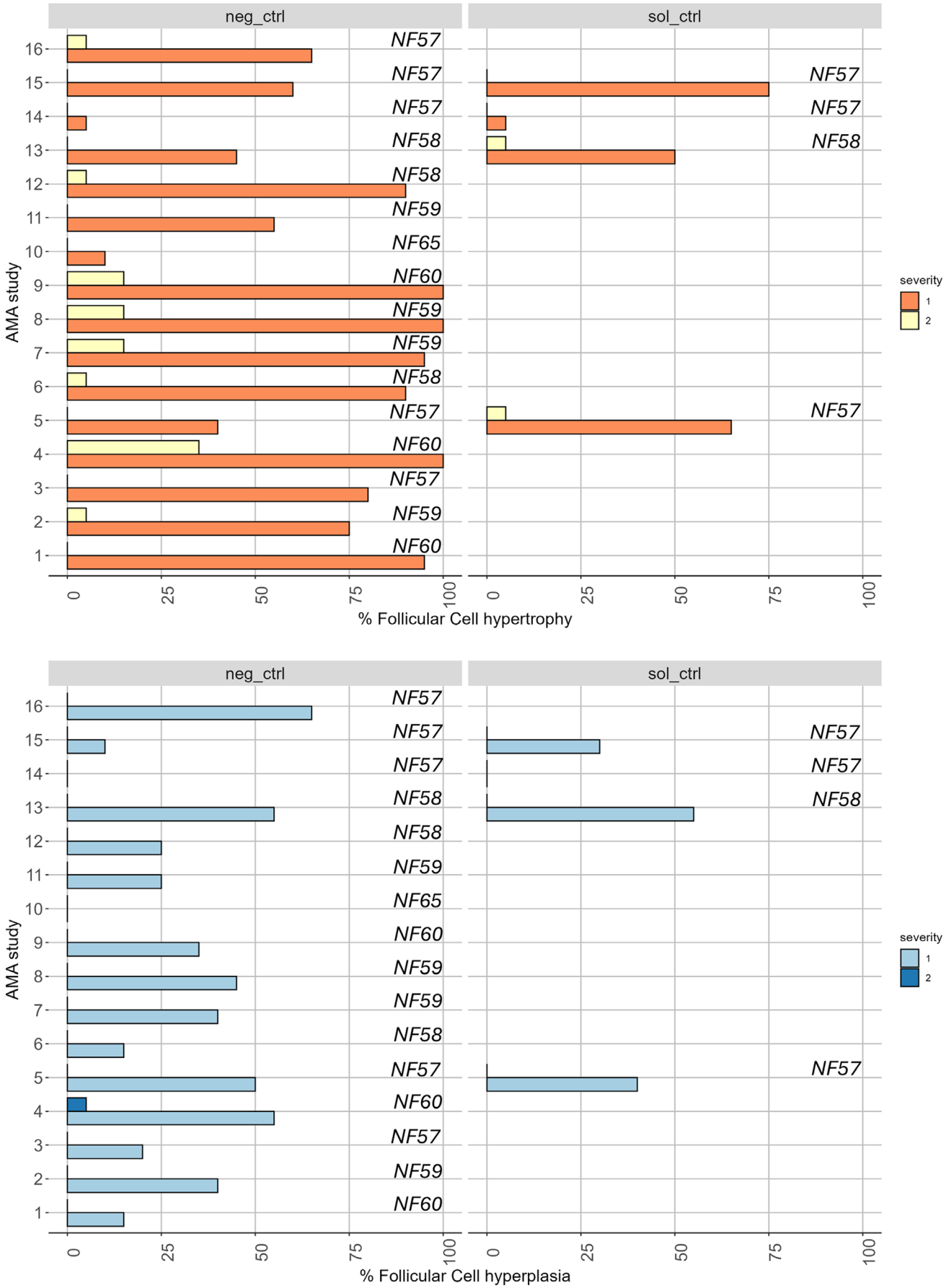

FIGURE 3 Specific occurrence of follicular cell hypertrophy (upper panel) and hyperplasia (lower panel) in both the negative (neg_ctrl) and solvent (sol_ctrl) control animals in the AMA studies. The median developmental stage is reported in italics for each study. Severity 1 = mild (~30%–50% of tissue affected), Severity 2 = moderate (~60%–80% of tissue affected).

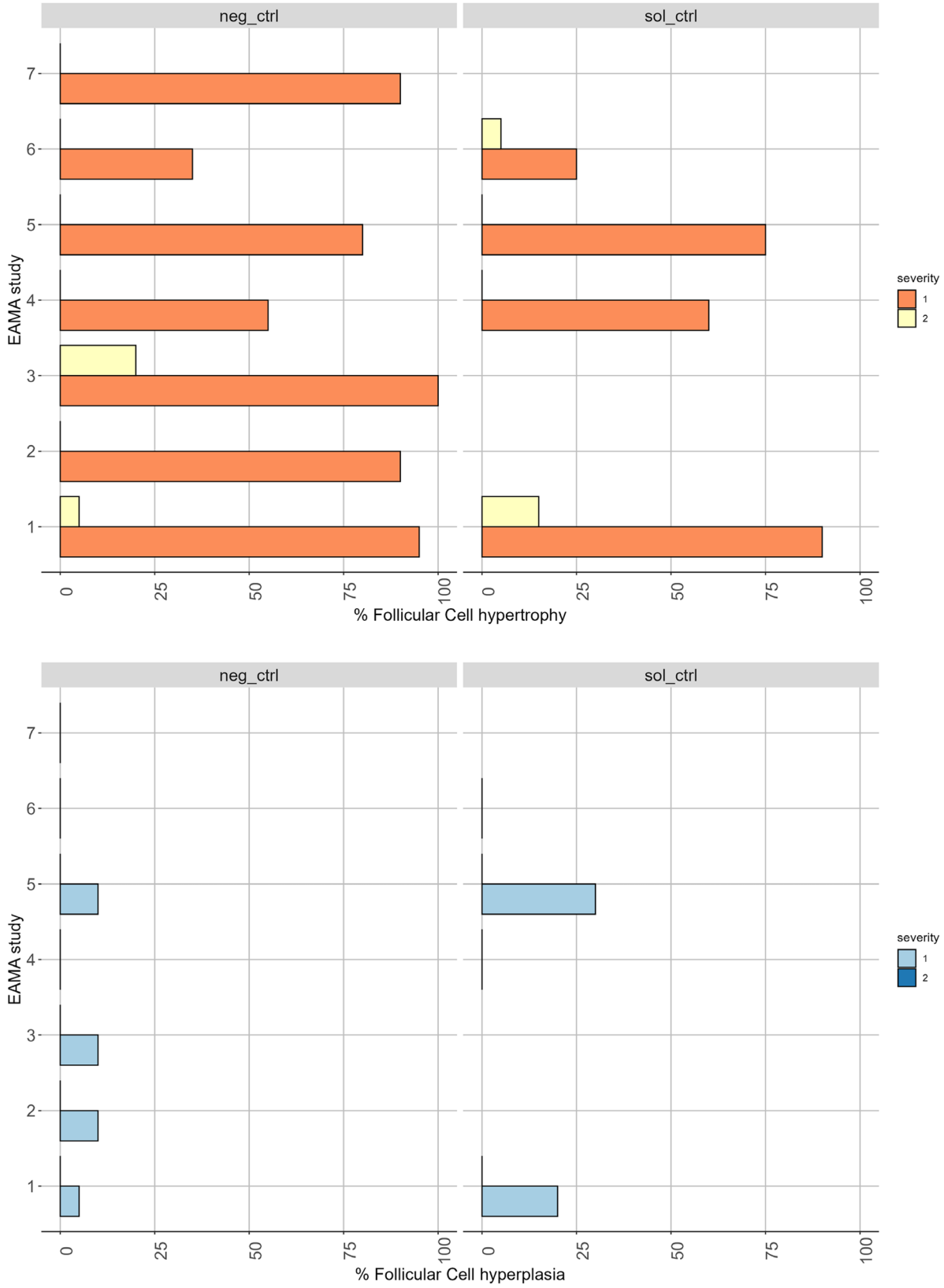

FIGURE 4 Specific occurrence of follicular cell hypertrophy (upper panel) and hyperplasia (lower panel) in both the negative (neg_ctrl) and solvent (sol_ctrl) control animals in the EAMA studies. Severity 1 = mild (~30%–50% of tissue affected), Severity 2 = moderate (~60%–80% of tissue affected).

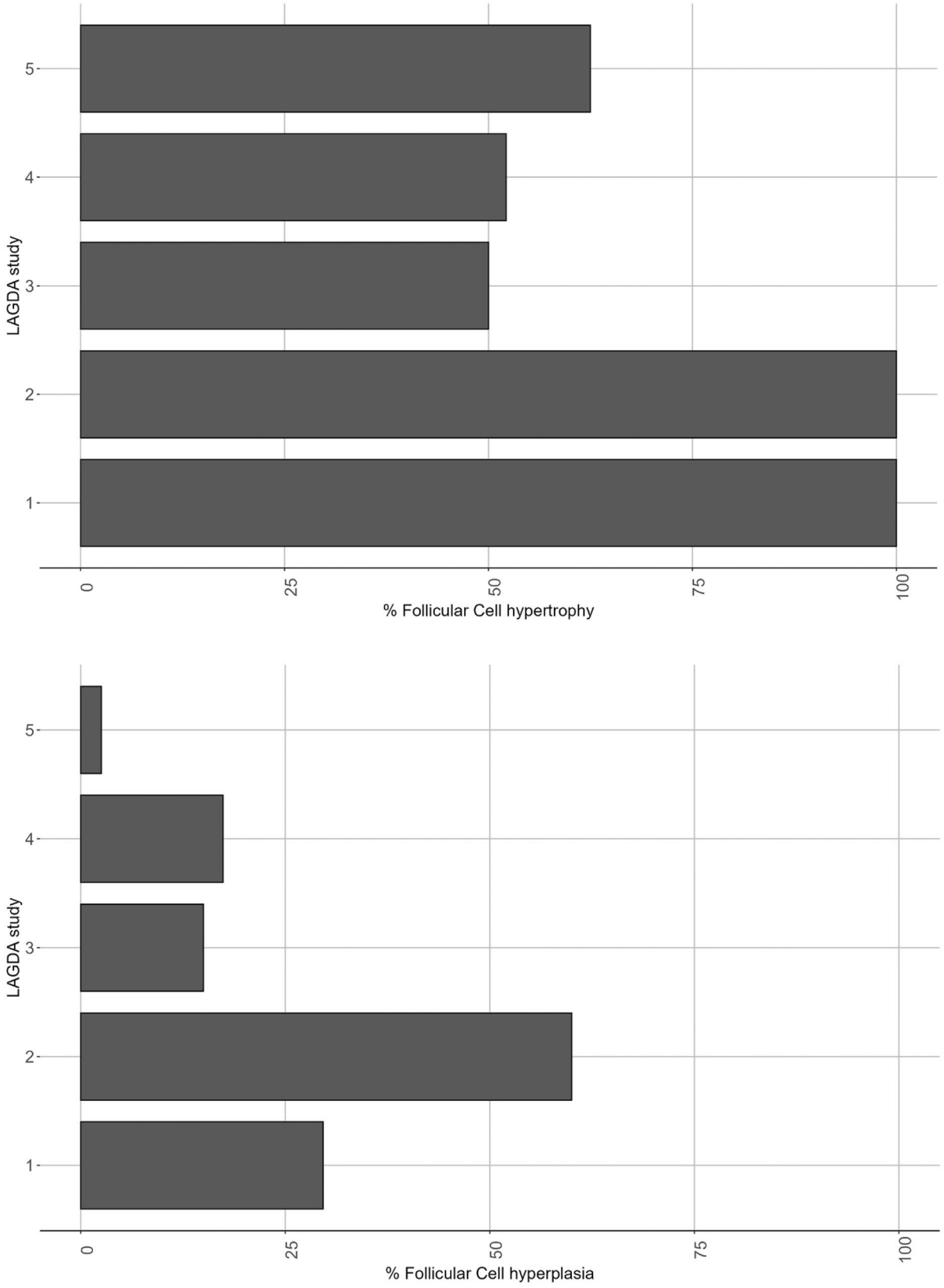

FIGURE 5 Specific occurrence of follicular cell hypertrophy (upper panel) and hyperplasia (lower panel) in control animals in the LAGDA studies.

In control animals in the AMA studies, the peak of FC hypertrophy seemed to occur when animals were at stages NF 58–60 and this seemed to confirm the trend reported by Wolf et al. (2021, 2024). Similar to Wolf et al. (2021, 2024), the occurrence of the peak of FC hyperplasia in the analysed AMA studies was less clear. In line with the principle that FC proliferation precedes major increases in cell size, it is likely that FC hyperplasia normally occurs before the peak of FC hypertrophy, and it is mainly observed at stages NF 57 and NF 58. This seems to be further confirmed by the analysis of the occurrence of FC hypertrophy and hyperplasia in the EAMA and LAGDA. In those studies, thyroid histology is assessed when animals reach stage NF 62. FC hypertrophy is still the most frequent observation, and FC hyperplasia in those studies is rarely observed,

confirming the overall conclusion from AMA studies. Moreover, as expected, FC hypertrophy also tended to decrease when the animals complete metamorphosis at stages NF 64/65, when the thyroid becomes more quiescent. Two AMA studies with animals at a median stage NF 57 showed quite a high prevalence of FC hypertrophy. This is somehow difficult to explain and can probably be considered an artefact, which however would need further confirmation with a larger dataset. It is, however, noted that Wolf et al. (2021, 2024) did not report a similar trend for FC hypertrophy for animals at stage NF 57.

3.4 | Interpretative considerations on thyroid histopathology

In line with Regulation 2018/605, a substance shall be considered to be an endocrine disruptor if:

a. *it shows an adverse effect in [an intact organism or its progeny]/[non-target organisms (NTOs)], which is a change in the morphology, physiology, growth, development, reproduction or life span of an organism, system or (sub)population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress or an increase in susceptibility to other influences;*
b. *it has an endocrine mode of action*[4]*, i.e. it alters the function(s) of the endocrine system;*
c. *the adverse effect is a consequence of the endocrine mode of action*.

In general, the identification of ED is based on the identification of a pattern of effects and should not be based on a change in a single parameter. Therefore, changes in thyroid histology should be assessed and interpreted, considering the overall pattern observed within a study and across studies with the same species and different species, including information from any *in vitro* studies. Nevertheless, it is also important to note that thyroid histopathology is known to be a sensitive diagnostic parameter to detect any compound interfering with the HPT axis in amphibians. The high sensitivity of thyroid histopathology compared with other parameters indicates the extreme capacity of the thyroid gland to compensate for a perturbation. Thus, care should be taken when changes in thyroid histopathology are observed without any effect on development. Normally two interpretative scenarios are possible, when considering the criteria as listed above:

1. Changes in the thyroid histology reflect activation of a compensatory mechanism. This means that, although changes are observed in the thyroid gland, the perturbation is overcome by inherent compensatory responses and does not result in an adverse effect on development. Therefore, in the absence of effects on the developmental stage/time to reach NF 62, if the study under assessment has enough power to detect an effect on development, it could be reasonably concluded that the endocrine activity observed does not trigger a T-mediated adversity.
2. Changes in thyroid histology do not seem to have repercussions on the development. However, if the study is not well conducted or the power of the test to detect an effect on development is insufficient, it is not possible to exclude the possibility that the substance may cause T-mediated adversity.

It is important to emphasise that the findings in either scenario need to be well contextualised in the WoE in which all available data are considered.

Furthermore, when changes in thyroid histology are observed in combination with effects on development, the direction of effects must follow a consistent pattern for both outcomes within and across studies.

The assessment and interpretation of changes in thyroid histology is facilitated in the EAMA, LATT and LAGDA as thyroid histology is assessed in animals at the same NF stage, eliminating any confounding factors such as growth and different developmental stages.

Table 6 summarises the main pattern of effects for known T-MoA in amphibians. The developmental stage or time to reach metamorphosis and thyroid histology should be given a higher weight in the overall WoE, since they are the two diagnostic parameters for detecting THSDCs.

[4]It should be highlighted that the 'endocrine mode of action' as stated in point (b) should be interpreted as 'endocrine activity' while the term 'endocrine mode of action' in point (c) covers the link between the adverse effect and the endocrine activity identified in points (a) and (b) respectively.

TABLE 6 Pattern of effects related to T-modality in amphibians for known thyroid hormone system-disrupting compounds

Substance	T-MoA	Developmental stage	Time to reach stage NF 62	Change in thyroid histology, i.e. FC hypertrophy and/or hyperplasia	nHLL[a]	SVL	WBW
PTU, ETU, methimazole	TPO inhibitors	↓	↑	↑	↓	↑[c]	↑[c]
T_4, L-thyroxine	TRH agonist	↑	↓	↓[b]	↑	↓[c]	↓[c]
Iopanoic acid	Diodinase inhibitor	Asynchronous	Asynchronous	↑	↓	↑[c]	↑[c]
Perchlorate	NIS inhibitor	↓	↑	↑	↓	↑[c]	↑[c]

Abbreviations: DIO, Diodinase; ETU, ethylene thiourea; nHLL, normalised hind limb length; PTU, propylthiouracil; SVL, snout-to-vent length; T_4, thyroxine hormone; T-MoA, T mode of action; TPO, thyroid peroxidase; TRH, thyrotropin-releasing hormone; WBW, wet body weight.

[a]nHLL is the relevant parameter since it is normalised to the SVL to consider a consistent comparison between animals by accounting for differences in overall growth.

[b]As outlined in OECD TG 231, for thyroid agonists, advanced development may be enough to conclude on the ED properties of a substance. However, if thyroid histology is also investigated, the table reports the expected pattern.

[c]The direction of the effect may be the opposite.

Table 6 details general patterns of effect but deviations are also possible. For example, changes in growth parameters may occur with or without any effect on the developmental stage or thyroid histopathology. In such cases, it is even more crucial to consider the available WoE to understand whether the observed effects are consistent with the expected pattern of endocrine effects. Other cases may present with indicators of developmental delay (e.g. reduced growth parameters and developmental stage) and thyroid gland shrinkage rather than the expected pattern of FC hypertrophy and/or hyperplasia. Caution should be exercised here as these developmental delays may be triggered by general toxicity caused by a reduction in food intake and thus negative energy balance. Concomitant changes in liver histopathology, e.g. reduced liver vacuolation could be helpful in confirming such a hypothesis (Marini et al., 2023). More often, though, sufficient supportive information is not available (e.g. reduced feeding, additional histopathological investigation) to support a reduced energy balance hypothesis. It is, therefore, recommended to include as much additional data in the study reports as possible, as this may enable identification of an effect pattern. Attention should be paid to potential inhibitors of deiodinases. In this case, it is well known that growth parameters are particularly sensitive, while it may be difficult to correctly stage the animals owing to asynchronous development. In addition, as stated in OECD TG 231, while a delay in development may be induced by other toxic MoA than ED, asynchronous or accelerated developments are considered specific to an ED MoA.

One of the most common T-MoA detected in mammals for pesticide active substances is increased TH clearance caused by the induction of certain liver metabolic enzymes, e.g. constitutive androstane receptor (CAR), Pregnane X Receptor (PXR), [UDP]-glucuronyl transferase. In contrast, this MoA has not been detected so far in amphibians (Schopfer et al., 2025). Phenobarbitals are, to date, the only known liver inducers that have shown effects compatible with a disruption of the HPT axis in amphibians, although it is also clear that those compounds may act through various MoAs. To this respect, little information is known but there may be physiological differences, e.g. different nuclear receptors (Mathäs et al., 2012; Zhao et al., 2015), leading to a dissimilar response in mammalian and non-mammalian species. Additional research would be warranted to gain a better insight into this particular MoA in non-mammalian species.

3.4.1 | Illustrative examples

It is of crucial importance to consider when and at what concentration any particular pattern of effects emerges and if these effects occur concomitantly or at higher concentrations sufficient to induce systemic toxicity or mortality. Incorporation of other scientific available data (e.g. positive *in vitro* evidence or positive evidence in mammals) should also be considered in determining concern for endocrine disruption.

The example reported in Table 7 and Figure 6 shows a delay in development both at Day 7 and at termination at the highest tested concentration. This effect was accompanied by a reduction of up to 40% in WBW in a concentration-dependent manner at both time points, Days 7 and 21. The delay in development was already evident at premetamorphosis when THs are below detectable limits (see Figure 6). Moreover, the effects in WBW were present at concentrations below those causing a developmental delay and are inconsistent with the expected pattern of higher, not lower, body weights when development is delayed. Such a pattern is, therefore, likely to be indicative of a non-ED MoA.

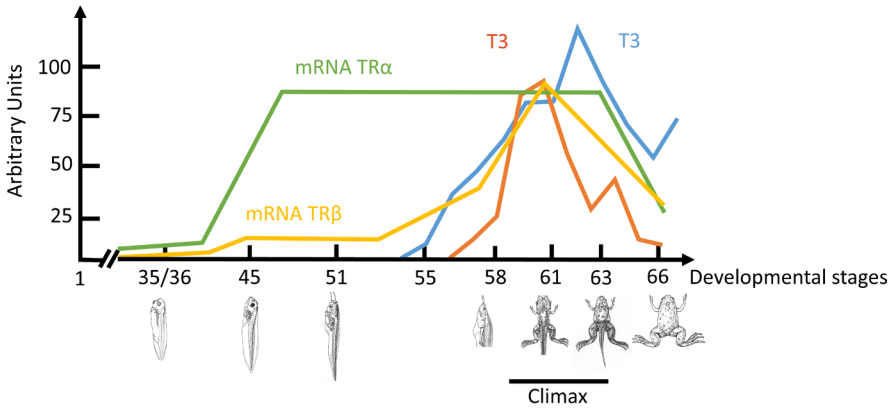

FIGURE 6 Thyroid hormone peaks in amphibians in relation to developmental stages (adapted from Holzer and Laudet, 2013).

TABLE 7 Example of a pattern prone to misinterpretation

Day 7 – Premetamorphosis			
	Median developmental stage	**WBW (mg; % difference compared with control)**	**Thyroid histology**
Negative control	54	–	
Intermediate tested concentration	54	−20%	NA
Highest tested concentration	53	−40%	NA
Termination – metamorphic climax			
Negative control	62	–	–
Intermediate tested concentration	62	−9%	–
Highest tested concentration	60	−23%	Compared with the control, one animal showed FC hyperplasia and one animal FC hypertrophy

Abbreviations: FC, follicular cell; WBW, wet body weight.

Poor performance of the control animals could also potentially lead to misinterpretation of the findings in an AMA study.

In the example reported below in Table 8, effects on nHLL, developmental stage and thyroid histopathology were observed at Day 21 in the highest treatment only.

TABLE 8 Effects observed in growth parameters, thyroid and liver histopathology pointing to different directions and thus not suggesting a pattern of T-mediated effects

Day 21						
	Developmental stage	**nHLL**	**SVL**	**WBW**	**Thyroid histology**	**Liver and pancreas histology**
Lowest treatment	–	–	–	–	–	–
Intermediate treatment	–	–	–	–	–	–
Highest treatment	↓[a]	↓	–	–	Decreased prevalence of follicular cell hypertrophy compared with the control (10 vs. 15), two tadpoles with thyroid atrophy	Decreased liver vacuolation (liver) and cell necrosis (pancreas) potentially indicating general toxicity

Abbreviations: nHLL, normalised hind limb length; SVL, snout-to-vent length; WBW, wet body weight.

[a]Significantly decreased based on the Jonckheere–Terpstra step down test on control 20th–80th percentiles.

The effects on the thyroid histopathology (i.e. decreased prevalence of FC hypertrophy) did not reflect the expected pattern that should be observed in the presence of a delay in development. However, there were some uncertainties due to the absence of a clear indication supporting the reduced energy balance hypothesis (e.g. no significant decrease in body weight) and the presence of two tadpoles showing thyroid atrophy. Effects observed on organs other than the thyroid (i.e.

liver and pancreas) could indicate potential general stress; however, the hypothesis could not be fully confirmed, considering the low number of animals tested for these organs (two tadpoles for control and highest treatment). Several ED MoAs were also postulated; however, no evidence was found to support them. Considering all the available information, the observed effects were considered probably due to other non-ED MoAs.

Tables 9 and 10 show that, for this example, an increase in all growth parameters was observed and was accompanied by a trend towards a delay in the development stage.

TABLE 9 Growth parameters showing an increasing trend compared with the control

	Developmental stage	nHLL	SVL	WBW
Negative control	–	–	–	–
Lowest treatment	↑	↑	↑	↑
Intermediate treatment	↑	↑	↑	↑
Highest treatment	↑	↑	↑	↑

Abbreviations: nHLL, normalised hind limb length; SVL, snout-to-vent length; WBW, wet body weight.

TABLE 10 Changes observed in the thyroid histopathology

	Negative control		Lowest treatment		Intermediate treatment		Highest treatment	
	Incidence	Severity	Incidence	Severity	Incidence	Severity	Incidence	Severity
Animals number	20		20		20		20	
Thyroid gland hypertrophy	8	1	11	1	11	1	13	1
Thyroid gland atrophy	3	1.3	3	1	2	1	2	1.5
FC hypertrophy	2	1	5	1	3	1	4	1
FC atrophy	4	1.3	2	1	1	1	2	1.5
Follicular lumen area, increase	2	1	1	1	0	–	4	1

Abbreviation: FC, follicular cell.

However, close examination of the raw data revealed that control animals at Day 21 exhibited a much lower body weight than expected for animals at this age (mean = 700 g). Findings in the thyroid histopathology also showed a rare pattern of changes in the control animals. While the OECD guidance on thyroid histology related only the thyroid gland hypertrophy to changes in follicular size and number, i.e. FC hypertrophy and hyperplasia, it is well known that goitres may be accompanied by different follicular cell morphologies (Opitz et al., 2009). In this study, iodine supplementation was not well reported and animals were given 20% less food than recommended for a flow-through system. Thus, it was concluded that the presence of goitres is likely to be related to non-optimal environmental conditions rather than to an ED MoA. The trend observed in growth parameters is thus considered a consequence of a poorly conducted and controlled study and the compromised health of the control animals.

3.5 | Conclusions (ToR2)

Endocrine axes are complex physiological regulatory systems. Consequently, the identification of xenobiotics such as PPPs, interfering with their correct functioning, is not a trivial process. Therefore, it is recommended that all available evidence be considered when drawing a conclusion on the potential for endocrine disruption.

In amphibians, because of the extreme capacity of the thyroid gland to compensate for a perturbation, thyroid histopathology is the most sensitive parameter to detect any compound interfering with the HPT axis. However, as described in Section 3.4, to determine if a biologically plausible pattern related to an interference with the HPT axis exists, changes in thyroid histopathology should be interpreted together with all other available data. Care should be exercised when changes in thyroid histopathology are observed without any effects on development. The inclusion of mechanistic endpoints such as TH measurements (with the potential consideration of transcriptomic analyses) represents a significant contribution to the overall WoE. However, if effects on the developmental stage or time to reach NF 62 are observed in the absence of clear changes in thyroid histology, any attribution to a T-MoA is questionable. The delay in development observed in the presence or absence of thyroid histopathology may also be induced by a number of causes, including non-ED-related toxicity of a chemical or other non-toxic mechanisms, such as food avoidance. However, to further demonstrate that the effects may not be ED related, especially when the delay in development is accompanied by a compatible change in thyroid histopathology, additional and sufficient evidence should be provided, e.g. a comparative MoA demonstrating whether the pattern of effects is more likely to be related to a non-ED MoA. Additional analyses supporting such a hypothesis should be provided, i.e. liver and kidney histopathology, as well as detailed behavioural observations.

In general, some limitations have been identified in the way thyroid histology is currently assessed and, in particular, the need for better reporting, for example, by including histology images. The need for harmonised and consistently applied criteria and a quantitative morphological analysis is a point shared among experts in the field. Such a quantitative approach has been developed for fish and is currently under OECD validation for TG 210 (FELS) in which thyroid histopathology is assessed. The protocol includes systematic and quantitative morphological measurements of thyroid follicle structure that can detect cell size changes at the low micrometre range expected, even with strong thyroid-active compounds like PTU (Gölz et al., 2023). A similar approach is recommended to harmonise and standardise thyroid histopathology in amphibians that will make the results more sensitive and reliable, and less dependent on the assessment of individual pathologists.

4 | RECOMMENDATIONS

Applicants and laboratories performing studies are given the following recommendations:

- Harmonise to the extent possible the 'controllable' factors (e.g. animal source, rate and type of food, iodide content) to help lower intra-replicate and inter-replicate variability, ensuring suitable control performances and probably allowing higher statistical power. This is particularly important for the iodide content of the water as higher iodide content, besides not being representative of iodine freshwater content, may also mask potential thyroid-mediated effects.
- Harmonise to the extent possible recommendations for the use of solvent controls and the data interpretation if both a solvent and a negative control are used in a study.
- Include a statistical power analysis of the results of the developmental stage in the AMA findings[5] (following the methods reported in Rizzuto et al., 2025), in order to address regulatory concerns and ensure full consideration of the results of the AMA in the regulatory process.
- Standardise the way thyroid histopathology is conducted and reported according to OECD (2007). Preferably, reporting should be accompanied by more representative histology images related to findings from each study, in addition to images demonstrating typical examples of the type and extent of the pathological alterations observed. Quantitative assessments should be considered.
- Conduct additional studies if an AMA is clearly positive or borderline positive to demonstrate the lack of an ED MoA. This could be done in several ways, including the testing strategy proposed in the ECHA/EFSA Guidance. In general, the first recommendation is to provide additional non-vertebrate evidence. If an additional study is warranted, a test in which additional parameters are investigated, for example, THs and histology of other organs, which would provide more useful information in the WoE assessment, would be preferable to provide, compared with one that only adds the 'time to' stage. Moreover, a test in which more than three concentrations are used would be more helpful for several reasons, for example, higher statistical power and use for endpoint setting. If the AMA is clearly positive, alternative non-ED MoAs could be postulated and a comparative MoA analysis performed.
- Validate and standardise within the OECD the newly available methods with amphibians. It would be preferable to combine all the enhancements compared with the AMA in a single protocol and to also include a method for TH measurements considering the analytical sensitivity of available techniques. Aspects to carefully consider in the course of validation would be environmental conditions as reported in bullet 1 above, use of solvents, definition of MTC, etc.

Recommendations for regulators:

- Use the AMA to draw a conclusion on the ED properties of a substance. If the AMA is the only available study, it is clearly positive and the overall WoE also supports the conclusion. Use of level 3 studies to determine adversity and endocrine activity is supported by analysis of statistical power conducted by Rizzuto et al. (2025); and reported above).

5 | FUTURE RESEARCH NEEDS

Although outside the scope of the mandate, future research needs have also been identified. In this respect, academics and test developers are recommended to:

- include gene expression in the available protocols and develop interpretative guidance;
- include neuronal endpoints, e.g. brain and eyes, and/or behavioural readouts that can help a better understanding of possible neurodevelopmental effects related to thyroid MoA in amphibians;
- further develop new approach methods (NAMs) for the detection of ED through the T-modality. This could be considered a better integration of molecular, developmental and histological endpoints in *in vivo* assays;
- validate methods for quantitative morphological analysis of thyroid histology;
- further investigate (i) the reason why amphibians may not be as sensitive to the liver enzyme induction MoA; (ii) whether

[5]Any regulatory action related to this aspect will be clarified in the technical report on the outcome of the Peer Review Experts' meeting on recurrent issues in the assessment for endocrine disruptors.

other non-mammalian species respond *in vivo* to this MoA and (iii) develop a non-mammalian test for detection of TH liver clearance, as there is evidence that this MoA may be relevant for fish (Christen & Fent, 2014; Wang et al., 2014).

ABBREVIATIONS

AMA	Amphibian Metamorphosis Assay
CRF	corticotropin-releasing hormone
dph	days post hatch
EAMA	Extended Amphibian Metamorphosis Assay
ED	endocrine disruptor
ETU	ethylene thiourea
FC	follicular cell
HLL	hind limb length
HPG	hypothalamic–pituitary–gonadal (axis)
HPT	hypothalamic–pituitary–thyroid (axis)
LAGDA	Larval Amphibian Growth and Development Assay
LATT	Larval Amphibian Toxicity Test
LBF	frog brittle slurry
MoA	mode of action
MTC	maximum tolerated concentration
NF	Nieuwkoop and Faber
nHLL	normalised hind limb length
NTOs	non-target organisms
OECD	Organisation for Economic Co-operation and Development
PPP	plant protection product
PTU	propylthiouracil
SVL	snout-to-vent length
T	thyroid
T_3	triiodothyronine hormone
T_4	thyroxine hormone
TG	test guideline
TH	thyroid hormone
THSDC	thyroid hormone system-disrupting compound
ToR	Term of Reference
TPO	thyroid peroxidase
TRH	thyrotropin-releasing hormone
TRHR	thyrotropin-releasing hormone receptor
TSH	thyroid-stimulating hormone
WBW	wet body weight
WoE	weight of evidence
XETA	Xenopus Eleutheroembryonic Thyroid Assay

REQUESTOR

EFSA

QUESTION NUMBER

EFSA-Q-2024-00116

COPYRIGHT FOR NON-EFSA CONTENT

EFSA may include images or other content for which it does not hold copyright. In such cases, EFSA indicates the copyright holder and users should seek permission to reproduce the content from the original source.

REFERENCES

Christen, V., & Fent, K. (2014). Tissue-, sex- and development-specific transcription profiles of eight UDP-glucuronosyltransferase genes in zebrafish (*Danio rerio*) and their regulation by activator of aryl hydrocarbon receptor. *Aquatic Toxicology*, *150*, 93–102. https://doi.org/10.1016/j.aquatox.2014.02.019

Couderq, S., Leemans, M., & Fini, J. B. (2020). Testing for thyroid hormone disruptors, a review of non-mammalian in vivo models. *Molecular and Cellular Endocrinology*, *508*, 110779. https://doi.org/10.1016/j.mce.2020.110779

Dang, Z. (2022). Amphibian toxicity testing for identification of thyroid disrupting chemicals. *Environmental Pollution*, *311*, 120006. https://doi.org/10.1016/j.envpol.2022.120006

Fort, D. J., Mathis, M. B., Pawlowski, S., Wolf, J. C., Peter, R., & Champ, S. (2017). Effect of triclosan on anuran development and growth in a larval amphibian growth and development assay. *Journal of Applied Toxicology*, *37*, 1182–1194. https://doi.org/10.1002/jat.3474

Gölz, L., Pannetier, P., Fagundes, T., Knörr, S., Behnstedt, L., Coordes, S., Matthiessen, P., Knapen, D., Vergauwen, L., Holbech, H., Braunbeck, T., & Baumann, L. (2023). Development of the integrated fish endocrine disruptor test (iFEDT)–Part B: Implementation of thyroid-related endpoints. *Integrated Environmental Assessment and Management*, *20*, 830–845. https://doi.org/10.1002/ieam.4828

Grim, K. C., Wolfe, M., Braunbeck, T., Iguchi, T., Ohta, Y., Tooi, O., Touart, L., Wolf, D. C., & Tietge, J. (2009). Thyroid histopathology assessments for the amphibian metamorphosis assay to detect thyroid-active substances. *Toxicologic Pathology*, *37*, 415–424. https://doi.org/10.1177/0192623309335063

Haselman, J. T., Kosian, P. A., Korte, J. J., Olmstead, A. W., & Degitz, S. J. (2018). Effects of multiple life stage exposure to the fungicide prochloraz in Xenopus *laevis*: Manifestations of antiandrogenic and other modes of toxicity. *Aquatic Toxicology*, *199*, 240–251. https://doi.org/10.1016/j.aquatox.2018.03.013

Haselman, J. T., Kosian, P. A., Korte, J. J., Olmstead, A. W., Iguchi, T., Johnson, R. D., & Degitz, S. J. (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. *Journal of Applied Toxicology*, *36*, 1639–1650. https://doi.org/10.1002/jat.3330

Haselman, J. T., Olker, J. H., Kosian, P. A., Korte, J. J., Denny, J. S., Tietge, J. E., Hornung, M. W., & Degitz, S. J. (2022). Characterization of the mechanistic linkages between iodothyronine deiodinase inhibition and impaired thyroid-mediated growth and development in *Xenopus laevis* using iopanoic acid. *Toxicological Sciences*, *187*, 139–149. https://doi.org/10.1093/toxsci/kfac014

Haselman, J. T., Sakurai, M., Watanabe, N., Goto, Y., Onishi, Y., Ito, Y., Onoda, Y., Kosian, P. A., Korte, J. J., Johnson, R. D., Iguchi, T., & Degitz, S. J. (2016). Development of the Larval Amphibian Growth and Development Assay: Effects of benzophenone-2 exposure in *Xenopus laevis* from embryo to juvenile. *Journal of Applied Toxicology*, *36*, 1651–1661. https://doi.org/10.1002/jat.3336

Holzer, G., & Laudet, V. (2013). Thyroid hormones and postembryonic development in amniotes. *Animal Metamorphosis*, *103*, 397–425. https://doi.org/10.1016/b978-0-12-385979-2.00014-9

Ikegami, K., & Yoshimura, T. (2017). The hypothalamic–pituitary–thyroid axis and biological rhythms: The discovery of TSH's unexpected role using animal models. *Best Practice & Research Clinical Endocrinology & Metabolism*, *31*, 475–485. https://doi.org/10.1016/j.beem.2017.09.002

Marini, J. P., Wolf, J. C., Mingo, V., Sayers, L. E., Jamieson, S. Y., & Wheeler, J. R. (2023). An amphibian metamorphosis assay dietary restriction study: Lessons for data interpretation. *Environmental Toxicology and Chemistry*, *42*, 1061–1074. https://doi.org/10.1002/etc.5596

Mathäs, M., Burk, O., Qiu, H., Nusshag, C., Gödtel-Armbrust, U., Baranyai, D., Römer, K., Nem, D., Windshügel, B., & Wojnowski, L. (2012). Evolutionary history and functional characterization of the amphibian xenosensor CAR. *Molecular Endocrinology*, *26*, 14–26. https://doi.org/10.1210/me.2011-1235

Nieuwkoop, P. D., & Faber, J. (1994). *Normal table of Xenopus laevis (Daudin)*. Garland Publishing Inc. ISBN 0-8153-1896-0.

OECD (Organization for Economic Cooperation and Development). (2019). TG 248: Xenopus Eleutheroembryonic Thyroid Assay (XETA), OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris. https://doi.org/10.1787/a13f80ee-en

OECD (Organization for Economic Co-operation and Development). (2006). Current Approaches in the Statistical Analysis of Ecotoxicity Data: A guidance to application (annexes to this publication exist as a separate document), OECD Series on Testing and Assessment, No. 54, OECD Publishing, Paris. https://doi.org/10.1787/9789264085275-en

OECD (Organization for Economic Co-operation and Development). (2007). *Guidance Document on Amphibian Histology*. OECD Publishing. https://doi.org/10.1787/9d831534-en

OECD (Organization for Economic Co-operation and Development). (2009). Test No. 231: Amphibian metamorphosis assay. In: OECD Guidelines for the Testing of Chemicals, Section 2 OECD Publishing, Paris, 33 pp. https://doi.org/10.1787/9789264076242-en

OECD (Organization for Economic Co-operation and Development). (2015a). Test No. 241: The Larval Amphibian Growth and Development Assay (LAGDA). In: OECD Guidelines for the Testing of Chemicals, Section 2 OECD Publishing, Paris, 38 pp. https://doi.org/10.1787/9789264242340-en

OECD (Organization for Economic Co-operation and Development). (2015b). Guidance Document on histopathology techniques and evaluation (Part I) for the larval amphibian growth and development assay (LAGDA). In: OECD series on Testing and Assessment. OECD Publishing, Paris, 28 pp. https://www.oecd.org/env/ehs/testing/37_PART1.pdf

OECD (Organization for Economic Co-operation and Development). (2015c). Guidance Document on histopathology techniques and evaluation (Part II) for the larval amphibian growth and development assay (LAGDA). In: OECD series on Testing and Assessment. OECD Publishing, Paris, 23 pp. https://www.oecd.org/env/ehs/testing/37_PART2.pdf

OECD (Organization for Economic Co-operation and Development). (2018). Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption, OECD Series on Testing and Assessment, No. 150, OECD Publishing, Paris. https://doi.org/10.1787/9789264304741-en

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. *Molecular and Cellular Endocrinology*, *298*, 101–114.

Ortego, L. S., Olmstead, A. W., Weltje, L., Wheeler, J. R., Bone, A. J., Coady, K. K., Banman, C. S., Burden, N., & Lagadic, L. (2021). The extended amphibian metamorphosis assay: A thyroid-specific and less animal-intensive alternative to the larval amphibian growth and development assay. *Environmental Toxicology and Chemistry*, *40*, 2135–2144. https://doi.org/10.1002/etc.5078

Orton, F., & Tyler, C. R. (2015). Do hormone-modulating chemicals impact on reproduction and development of wild amphibians? *Biological Reviews of the Cambridge Philosophical Society*, *90*, 1100–1117. https://doi.org/10.1111/brv.12147

Rizzuto, S., Neri, F. M., Ercolano, V., Ippolito, A., Linguadoca, A., Villamar Bouza, L., & Arena, M. (2025). Amphibian studies to investigate the endocrine-disrupting properties of chemicals through the thyroid modality: A comparison of their statistical power. *Environmental Toxicology and Chemistry*, *44*, 1435–1447. https://doi.org/10.1093/etojnl/vgaf067

Schopfer, C. R., Grözinger, F., Birk, B., Hewitt, N. J., Weltje, L., & Habekost, M. (2025). Cross-taxa extrapolation: Is there a role for thyroid hormone conjugating liver enzymes during amphibian metamorphosis? *Regulatory Toxicology and Pharmacology*, *159*, 105810. https://doi.org/10.1016/j.yrtph.2025.105810

Sternberg, R. M., Thoemke, K. R., Korte, J. J., Moen, S. M., Olson, J. M., Korte, L., Tietge, J. E., & Degitz, S. J. (2011). Control of pituitary thyroid-stimulating hormone synthesis and secretion by thyroid hormones during *Xenopus* metamorphosis. *General and Comparative Endocrinology*, *173*, 428–437. https://doi.org/10.1016/j.ygcen.2011.06.020

US EPA (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 EPA (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 EPA (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 EPA (US Environmental Protection Agency). (2013c). Validation of the larval amphibian growth and development assay: Integrated summary report.

US EPA (US Environmental Protection Agency). (2015). Endocrine Disruptor Screening Program Test Guidelines OCSPP 890.2300: Larval Amphibian Growth and Development Assay (LAGDA).

Wang, Y., Huang, H., & Wu, Q. (2014). Characterization of the zebrafish Ugt repertoire reveals a new class of drug-metabolizing UDP glucuronosyltransferases. *Molecular Pharmacology*, *86*, 62–75. https://doi.org/10.1124/mol.113.091462

Wolf, J. C., Bejarano, A. C., Fort, D. J., & Wheeler, J. R. (2021). An examination of historical control histopathology metadata from 51 Amphibian Metamorphosis Assays. *Critical Reviews in Toxicology*, *51*, 729–739. https://doi.org/10.1080/10408444.2021.1997910

Wolf, J. C., Green, J. W., Mingo, V., Marini, J. P., Schneider, S. Z., Fort, D. J., & Wheeler, J. R. (2024). Historical control histopathology data from amphibian metamorphosis assays and fathead minnow fish short term reproductive assays: A tool for data interpretation. *Aquatic Toxicology*, *267*, 106811. https://doi.org/10.1016/j.aquatox.2023.106811

Zhao, Y., Zhang, K., Giesy, J. P., & Hu, J. (2015). Families of nuclear receptors in vertebrate models: characteristic and comparative toxicological perspective. *Scientific Reports*, *5*, 8554. https://doi.org/10.1038/srep08554

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: EFSA (European Food Safety Authority). Baumann, L., Baynes, A., Bennekou, S. H., Crofton, K., Dang, Z., Degitz, S., Fini, J.-B., Gilbert, M., Holbech, H., Terron, A., Van Duursen, M., McVey, E., Arena, M., Kienzler, A., Rizzuto, S., Bouza, L. V., & Wilks, M. (2025). The assessment for potential thyroid-mediated endocrine disruption in amphibians: Clarification on the use of new methods and on the interpretation of changes in thyroid histology. *EFSA Journal*, *23*(12), e9815. https://doi.org/10.2903/j.efsa.2025.9815

ANNEX A

Outcome of the public consultation on the draft statement