Published in final edited form as:
Toxicol Sci. 2020 June 01; 175(2): 236–250. doi:10.1093/toxsci/kfaa036.

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*

Jonathan T. Haselman[1,*], **Jennifer H. Olker**[1], **Patricia A. Kosian**[1], **Joseph J. Korte**[1], **Joseph A. Swintek**[2], **Jeffrey S. Denny**[1], **John W. Nichols**[1], **Joseph E. Tietge**[1], **Michael W. Hornung**[1], **Sigmund J. Degitz**[1]

[1]U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Great Lakes Toxicology and Ecology Division, Duluth, Minnesota 55804

[2]Badger Technical Services, U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Great Lakes Toxicology and Ecology Division, Duluth, Minnesota 55804

Abstract

Chemical safety evaluation is in the midst of a transition from traditional whole-animal toxicity testing to molecular pathway-based in vitro assays and in silico modeling. However, to facilitate the shift in reliance on apical effects for risk assessment to predictive surrogate metrics having characterized linkages to chemical mechanisms of action, targeted in vivo testing is necessary to establish these predictive relationships. In this study, we demonstrate a means to predict thyroid-related metamorphic success in the model amphibian *Xenopus laevis* using relevant biochemical measurements during early pro-metamorphosis. The adverse outcome pathway for thyroperoxidase inhibition leading to altered amphibian metamorphosis was used to inform a pathway-based in vivo study design that generated response-response relationships. These causal relationships were used to develop Bayesian probabilistic network models that mathematically determine conditional dependencies between biochemical nodes and support the predictive capability of the biochemical profiles. Plasma thyroxine concentrations were the most predictive of metamorphic success with improved predictivity when thyroid gland sodium-iodide symporter gene expression levels (a compensatory response) were used in conjunction with plasma thyroxine as an additional regressor. Although thyroid-mediated amphibian metamorphosis has been studied for decades, this is the first time a predictive relationship has been characterized between plasma thyroxine and metamorphic success. Linking these types of biochemical surrogate metrics to apical outcomes is vital to facilitate the transition to the new paradigm of chemical safety assessments.

[*]**Corresponding author:** Jonathan Haselman, 218-529-5244, haselman.jon@epa.gov; U.S. EPA, Great Lakes Toxicology and Ecology Division, Duluth, Minnesota 55804.

Financial Interest Declaration: The authors declare they have no actual or potential competing financial interests.

Disclaimer: The views expressed in this manuscript are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA.

Data available at: https://edg.epa.gov/metadata/catalog/main/home.page

Keywords

Thyroid; *Xenopus laevis*; Thyroperoxidase; Adverse Outcome Pathway; Bayesian network

INTRODUCTION

Chemical safety evaluation is transitioning away from reliance on whole animal testing and toward increased use of in vitro and in silico mechanistic and pathway-based approaches for hazard characterization (National Research Council, 2007; Ankley et al., 2010). This paradigm shift supports the shared international goal of reducing animal testing in toxicology and will streamline chemical screening and prioritization efforts. Evaluation of common use chemicals for thyroid disrupting properties is imperative, as thyroid hormone plays a critical role in vertebrate development including control of mammalian neurodevelopment (Axelstad et al., 2008; Gilbert et al., 2012; Gilbert and Zoeller, 2010) and amphibian metamorphosis (Brown, 2005; Brown and Cai, 2007; Fort et al., 2007; Galton, 1983; Shi et al., 1996; Tata, 2006). The U.S. EPA's Endocrine Disruptor Screening Program (EDSP) and the global Organization for Economic Cooperation and Development (OECD) have adopted standardized in vivo test guidelines to evaluate a chemical's potential to disrupt the thyroid axis in rodent and amphibian models (OECD, 2018; U.S. EPA, 2009a). *Xenopus laevis* is the model amphibian used in these chemical screening programs and has been studied extensively in the context of thyroid-mediated metamorphosis (Morvan-Dubois et al., 2008). The highly conserved nature of thyroid biology across vertebrate taxa makes *X. laevis* a useful model for characterizing mechanisms of thyroid disruption (Coady et al., 2010; Degitz et al., 2005; Hornung et al., 2015; Olker et al., 2018; Sachs and Buchholz, 2017; Tietge et al., 2005, 2010, 2013).

Recent advances in thyroid-related in vitro chemical screening assays allow large libraries of chemicals to be evaluated for their activity toward specific thyroid-related targets (Buckalew et al., 2020; Deisenroth et al., 2019; Dong et al., 2019; Hallinger et al., 2017; Hornung et al., 2018; Murk et al., 2013; Olker et al., 2019; Paul et al., 2013, 2014; Paul Friedman et al., 2016, 2019; Wang et al., 2018). To support the transition away from animal testing and toward more reliance on these in vitro approaches, however, pathway-based predictive models need to be developed to link biochemical responses to organismal outcomes relevant to risk assessment (Noyes et al., 2019). In a recent study, Hassan et al. (2020) demonstrated quantitative linkages between in vitro inhibition of thyroperoxidase (TPO) inhibition and circulating thyroid hormone (TH) in the rodent model. TPO is a membrane-bound enzyme on the apical surface of thyroid follicular cells that catalyzes the covalent binding of iodine to tyrosine residues on thyroglobulin to produce monoiodotyrosine (MIT) and diiodotyrosine (DIT). Thyroxine (T4) is produced by coupling of two DIT residues, which is the secondary mechanism of TPO catalysis (Kessler et al., 2008; Ruf and Carayon, 2006; Taurog et al., 1996). Previously, Hassan et al. (2017) developed a physiologically-based computational model that quantitatively links circulating TH with physical malformations in rat brains. Comparable models that link chemical impacts on amphibian thyroid biochemistry to relevant apical endpoints (e.g., metamorphic failure) do not presently exist.

The pharmaceuticals methimazole (MMI) and propylthiouracil (PTU) strongly inhibit TPO resulting in reduced levels of circulating thyroid hormone (TH) in rodents (Axelstad et al., 2008; Gilbert, 2011; Hassan et al., 2017, 2020; Zoeller and Crofton, 2005) and amphibians (Coady et al., 2010; Degitz et al., 2005; Tietge et al., 2010). Both chemicals were employed to validate tier 1 standardized EDSP assays and both have been categorized as reference chemicals for thyroid disruption via TPO inhibition (Wegner et al., 2016). Mercaptobenzothiazole (MBT) is a high-volume production chemical used in a variety of industrial applications such as rubber vulcanization (Ciullo and Hewitt, 1999) and inhibition of metal corrosion (Jafari et al., 2014). MBT is a potent TPO inhibitor in vitro and causes the same adverse apical outcomes in *X. laevis* larvae as MMI and PTU including thyroid gland pathologies, decreased circulating levels of TH, and arrested metamorphosis (Hornung et al., 2015; Tietge et al., 2013).

The objective of the present study was to establish a quantitative relationship between developmental thyroid biochemistry and metamorphic success/failure in *X. laevis*. To achieve this objective, a pathway-based in vivo testing approach was taken to generate data suitable for predictive modeling. The adverse outcome pathway (AOP) framework is a widely accepted approach for describing the cascading effects of toxicity, starting with a molecular initiating event (MIE), which leads to downstream key events (KEs) and results in an adverse outcome (AO) relevant to chemical risk assessment (Ankley et al., 2010). This framework supports the assembly of existing knowledge and provides a logical basis for making downstream predictions based on upstream effects. AOPs for TPO inhibition leading to mammalian neurodevelopmental outcomes (Crofton et al., 2018) and altered amphibian metamorphosis (Haselman et al., 2018; Figure 1A) have been described, providing a weight of evidence supporting the linkages between thyroid biochemistry and adverse apical outcomes across species.

In vivo exposures with *X. laevis* larvae were performed to characterize pathway-level biochemical responses to MMI, PTU, and MBT, administered at multiple exposure concentrations with temporal subsampling. This study design provided an opportunity to evaluate the concordance in effects associated with the same MIE, while the time-course information allowed for an analysis of the timing and magnitude of TH-related perturbations that may be predictive of metamorphic failure. The resulting datasets were subjected to Bayesian network analysis to determine whether metamorphic success/failure was conditionally dependent on one or more measured endpoints. The resulting networks were then used to inform the development of logistic regressions for predicting the probability of metamorphic success based on thyroid-related biochemistry.

MATERIALS AND METHODS

Study design

Three separate studies were conducted using the same study design (Supplemental Figure S.1), but each with a different model TPO inhibitor (MMI, PTU, MBT). Exposure was initiated at, or slightly before *X. laevis* pro-metamorphosis (Nieuwkoop and Faber, 1994 [NF] stages 53/54).

Each study consisted of three chemical concentrations separated by either a 0.5 (MMI) or a 0.33 (PTU and MBT) dilution factor and a Lake Superior water (LSW) negative control. Three replicate tanks were analyzed for each treatment and the control for a total of 12 test tanks per study (i.e., n = 3 per treatment and control). The highest target test concentration for each chemical was 25 mg/L (220 μM) MMI, 270 μg/L (1.6 μM) MBT and 20 mg/L (117.5 μM) PTU, which were determined based on concentrations demonstrated in previous studies performed in our laboratory to arrest metamorphosis and impact thyroid biochemistry (Degitz et al., 2005; Hornung et al., 2015; Tietge et al., 2010, 2013). Blood plasma and thyroid glands were collected from random sub-samples at 2, 4, 7 and 10 days to evaluate early temporal profiles of TH-related molecules and compensatory metrics. For this study, targeted measurements in the plasma were triiodothyronine (T3) and T4 levels, whereas targeted measurements in thyroid gland tissues were MIT, DIT, T3, T4, sodium-iodide symporter (NIS) mRNA levels (thyroid stimulating hormone [TSH]-responsive gene) and thyroid follicular cell number (FCN) as a measure of compensatory thyroid gland hyperplasia (a pathology that contributes to goiter). To evaluate apical outcomes, the remaining larvae were grown out until they either reached NF stage 62 (metamorphic climax) or until the test was terminated approximately two weeks after the last of the controls reached NF stage 62. Any larvae remaining at test termination were considered to be developmentally arrested.

Animals and study initiation

All procedures involving animal use were reviewed and approved by the Division's Animal Care and Use committee. *X. laevis* larvae were acquired from the in-house culture. Approximately 21 days post-fertilization, larvae were briefly anesthetized in 100 mg/L neutral buffered MS-222 (Argent Chemical Laboratories, Redmond, WA, USA) and sorted based on NF stage. For each of the three studies there were approximately 40% NF stage 53 and 60% NF stage 54 that were randomly mixed and distributed across the study at the start. Once the larvae were given sufficient time to recover in clean LSW, they were randomly distributed into 12 intermediate containers prior to being randomly assigned to a treatment or control tank. Within approximately 30 minutes, all larvae were transferred to their respective tanks to initiate the exposure. Larvae were fed daily using a standard blended diet consisting of *Spirulina* algae (Wardley®, Secaucus, NJ, USA), GoldFish flakes (Tetra, Blacksburg, VA, USA) and Trout Starter (Skretting North America, Tooele, UT, USA) in addition to 24-hour old hatched live brine shrimp (Bio-Marine, Inc., Hawthorne, CA, USA). The diet preparation procedure and rations are outlined in the Larval Amphibian Growth and Development Assay test guideline (OECD, 2015; U.S. EPA, 2015).

Environmental conditions

All three studies were performed in the same flow-through exposure system that contained all glass aquaria (6.5 L) receiving approximately 44 ml/min of either LSW or toxicant solution to achieve 10 tank volume additions per day. Source water was pumped directly from Lake Superior, filtered to 5 μm and UV-treated prior to being delivered directly to control tanks or being used as diluent for the various toxicant concentrations. Fluorescent lighting was on a 12-hour light: 12-hour dark cycle. Temperature was held at 21°C and measured daily in all exposure tanks; dissolved oxygen, pH and conductivity were measured

weekly in all test aquaria; alkalinity and hardness were measured twice in one replicate per treatment at different times during each study (Supplemental Table S.1). All water quality parameters were within acceptable ranges, so the data are summarized based on each study and not grouped by treatments.

Sample collection and handling

At 2, 4, 7 and 10 days of exposure, 6–8 larvae were randomly removed from each test chemical tank and control. Larvae were euthanized in buffered MS-222 (200 mg/L), weighed and developmentally staged. Blood and thyroid gland tissues were then collected according to procedures described by Tietge et al. (2010) for further biochemical analyses. Blood was collected in heparinized capillary tubes and centrifuged at 13,700 × *g* to obtain a pooled plasma sample for each replicate tank. The two thyroid glands from each larva were distributed to two of three pools per tank until all three pools had generally even amounts of glands. For example, the first larva's glands went into pools #1 and #2; the second larva's glands went into pools #3 and #1; the third larva's glands went into pools #2 and #3, and so on. All gland tubes were kept on dry ice throughout the collection period so that glands were frozen immediately when transferred to the tube. The three thyroid gland pools per tank were collected for analyses of either NIS gene expression, iodo-tyrosines/thyronines, or FCN. Blood plasma and gland samples were all stored at −80°C until further processing for analyses.

Chemical stock preparation and exposure verification

Chemical stocks were prepared from neat material (MMI: CAS# 60–56-0, purity >98%; MBT: CAS# 149–30-4, purity >98%; PTU: CAS# 51–52-5, purity >99%) acquired from Sigma-Aldrich (St. Louis, MO, USA) according to methods previously described (Degitz et al., 2005; Tietge et al., 2010, 2013). Chemical stocks were diluted to each of the exposure concentrations using an automated proportional pump-driven dilution system and delivered directly to the test tanks using a peristaltic pump. Water samples (900 µl) were collected weekly from every control and MMI, MBT and PTU exposure tank, placed into amber vials and supplemented with 100 µl acetonitrile (ACN). Samples were immediately vortex mixed and analyzed using an Agilent 1100 series high performance liquid chromatograph (HPLC) with diode array detection. Detector wavelength settings were 258 nm, 325 nm and 276 nm for MMI, MBT and PTU, respectively. HPLC mobile phases were 100% HPLC-grade water with 0.05% acetic acid (A) and 95% ACN with 5% water (B). An aliquot of sample (i.e., 25 µl for MMI, PTU and 150 µl for MBT) was injected into a LiChrosorb RP-18 column (100 mm × 4.6 mm, 5 µ particle size; Alltech, Deerfield, IL, USA) and chromatographic separation of the compounds was achieved under isocratic conditions at a flowrate of 1 ml/min. Isocratic conditions were 90%A:10%B for MMI and 60%A:40%B for MBT and PTU. Analyte concentrations were quantified using linear regression. Quality assurance samples were prepared and analyzed with each chemical sample set and consisted of LSW blanks, duplicate samples and matrix spiked samples at corresponding treatment concentrations. All quality assurance samples met quality criteria (data not shown). No test chemical was detected in any of the control tanks throughout all three studies and all measured concentrations across all three studies were within 20% of nominal target concentrations with CVs within and between tanks of <10% (Supplemental Table S.2).

Sample analyses

Iodo-tyrosines/thyronines in thyroid glands and plasma.—Thyroid gland samples were homogenized and subjected to pronase digestion to liberate iodinated compounds (MIT, DIT, T3, T4) from thyroglobulin and then processed through solid phase extraction (SPE); plasma samples were acid hydrolyzed to liberate protein-bound TH prior to being processed through SPE. Detailed methods used for pronase digestion of thyroid tissues, acid hydrolysis of plasma, SPE and subsequent analytical determination of the iodinated species by liquid chromatography tandem mass spectrometry (LC/MS/MS) have been previously described (Hassan et al., 2017; Hornung et al., 2015; Olker et al., 2018).

RNA extraction and NIS mRNA expression.—Total RNA was extracted from thyroid gland samples using RNeasy® Plus Micro kits following the manufacturer's instructions (Qiagen, Germantown, MD, USA) then quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). All samples had A260/280 ratios greater than 1.59 indicating they had sufficient quality for subsequent quantitative PCR analysis. A subset of total RNA samples representing a portion of each group of RNA extractions was run on an Agilent 2100 Bioanalyzer using an RNA 6000 pico assay (Agilent, Santa Clara, CA, USA) to confirm acceptable RNA integrity. Thyroid gland NIS and ribosomal protein L32 (RPL32) mRNA expression levels were evaluated using a one-step TaqMan real-time RT-qPCR assay (Applied Biosystems, Foster City, CA, USA) and copies of NIS mRNA transcripts were normalized to copies of RPL32. Detailed methods used for RNA extraction and to determine expression levels of NIS and RPL32 have been previously described (Olker et al., 2018; Sternberg et al., 2011; Tietge et al., 2013).

Thyroid follicular cell number.—Thyroid FCN was determined using a CyQUANT® Cell Proliferation Assay Kit (Invitrogen) as previously described by Tietge et al. (2010). Thyroid gland samples were homogenized in assay kit buffer to yield a standard volume per gland and DNA concentration was determined according to kit instructions with a BioTek Synergy 4 fluorescence microplate reader (BioTek Instruments Inc., Winooski, VT, USA). Resulting DNA concentrations were converted to cell number using 7.7 pg DNA/cell as determined by Tietge et al. (2010).

Data analyses and statistics

Data reduction and cleaning.—The focus of this study was to quantify gland MIT, DIT, T3 and T4; plasma T3 and T4; NIS gene expression in thyroid tissue; and thyroid follicular cell numbers and relate these biochemical responses to metamorphic success. For occurrences where an analyte was below the lower limit of quantification (<LLOQ) in greater than 90% of samples across all treatments and control within a time point, the data were omitted prior to endpoint response analyses. However, these data points were assigned values equal to half of the LLOQ for graphing purposes and for Bayesian network analyses. For analytes and time points that shared a combination of detectable levels and non-quantifiable levels, the non-quantifiable levels were assigned a value equal to half of the LLOQ for inclusion in the data analysis workflow. One tank in the high PTU treatment experienced high mortality and impaired growth beyond the 4 d sampling point due to unknown factors. This was identified based on the data from the 7 and 10 d sample sets.

Therefore, the data associated with this tank were removed from the study, resulting in a sample size of n = 2 for this treatment whereas all other treatments resulted in a sample size of n = 3.

Endpoint response analysis.—Mean larval weight and median NF stage were calculated using the data collected from six to eight individuals per replicate tank at the various time points. Plasma and tissues were pooled across the same 6–8 individuals for biochemical measurements, so all data from the 2, 4, 7 and 10 d time points resulted in a single value per replicate tank and a sample size of n = 3 per treatment per time point (n = 2 for the high PTU treatment). Median NF stages within each treatment at each time point are presented as the median of tank medians; no statistical analyses were performed on the temporal developmental staging data, as overall development is encompassed within the time-to-NF stage 62 analysis described below. All endpoint response analyses were implemented in R (R Core Team, 2018) with packages used for various analyses referenced accordingly. Data from a single continuous endpoint by time point within each of the three studies were independently checked for normality and homogeneity of variance using the Shapiro-Wilk and Levene's tests (car v3.0–2, Fox and Weisberg, 2011), respectively, at alpha level of 0.01. All endpoints within all time points met parametric assumptions with the original data except the 2 and 10 d time points for gland T4 in the MBT study, which met parametric assumptions following a log-transformation. Data were then analyzed using a one-way ANOVA model with Dunnett's post-hoc test (multcomp v1.4–10, Hothorn et al., 2008). Time-to-NF stage 62 data were analyzed using a Cox mixed-effects proportional hazard model (coxme v2.2–10, Therneau, 2018) with individuals failing to reach NF stage 62 by the termination of the test censored within the analysis. However, the likelihood function of the hazard model converged before its parameter estimates, so post-hoc analysis of treatment effects on time-to-NF stage 62 was implemented using a step-down likelihood ratio test (instead of a parameter-based test like Dunnett's test). A false discovery rate (Benjamini and Hotchberg, 1995) p-value adjustment was applied to all resulting statistics from the biochemical measurements to account for the numerous comparisons across analytes, time points and studies. All statistical analyses of treatment effect were performed using an alpha level of 0.05.

Bayesian network analysis.—Bayesian networks (Pearl, 1986) are graphical statistical models that use both data-driven and expert knowledge approaches to model the conditional relationships between a set of variables. Modeling the relationships between variables is accomplished using Bayes' theorem to express the probability of an event or condition, which are called nodes, given the state of another set of events or conditions (e.g., metamorphic success being conditionally dependent on plasma T4 levels). These conditional dependencies are represented by an arrow called an "edge" (Figure 1B). Thus, Bayesian networks provide the ability to logically organize mechanistic data into probabilistic models based on relationships that show conditional dependence via an edge between nodes and show conditional independence with the exclusion of edges between nodes while identifying the full chain of relationships between variables.

The Bayesian network analysis employed in this study utilized the following endpoints at each of the time points: thyroid gland iodotyrosines (MIT, DIT), gland iodothyronines (T3, T4), plasma iodothyronines (T3, T4), and time-to-NF stage 62 (i.e., metamorphic success). Metamorphic success was defined as reaching NF stage 62 by the 75th percentile of time it took the controls to reach NF stage 62 (pooled across all studies), which was calculated to be 22 days. The above endpoints represent KEs within the TPO AOP (Figure 1A) and are treated as nodes within the Bayesian network (Figure 1B). The directional connections between the nodes in the network were assigned either a forced conditional dependency (i.e., "whitelisted"; black arrow, Figure 1B), possible conditional dependency (gray arrows, Figure 1B), or suspected lack of a conditional dependency (i.e., "blacklisted"; no arrow, Figure 1B). Whitelisted relationships are based on several key events within the amphibian TPO AOP that are linked by causal relationships (Figure 1A & 1B). Additionally, nodes known to be conditionally dependent on another node, within and across KEs, by what is known about thyroid-related biochemistry are also whitelisted. For example, gland T3 levels are conditionally dependent on both TPO coupling activity of MIT and DIT residues on thyroglobulin (Deme et al., 1978), and from outer ring deiodination of free T4 in the thyroid follicular cells (Beech et al., 1993). Any relationship defined regarding treatments had to be designated as possible conditional dependencies since the chemical treatments can elicit non-specific effects and specific levels of TPO inhibition by the chemical cannot be directly defined in vivo. Relationships represented by either a black or grey arrow in Figure 1B were considered by the networks, whereas relationships between nodes not connected by an arrow were not considered by the networks.

Networks were generated separately for each time point at which data were collected (day 2, 4, 7 or 10) using the abn (Kratzer et al., 2016; Lewis, 2016; Lewis and Ward, 2013) R package (R Core Team, 2019). The abn (additive Bayesian network) package was chosen due to its ability to whitelist and blacklist relationships and because it can handle both Gaussian and Bernoulli data simultaneously. The networks themselves were fit using the Bayesian equivalent of generalized linear mixed effect models where iteratively nested Laplace approximations (INLA) via the INLA R package (Rue et al., 2009) were used to estimate mixed effects for each node. Fitting the network, also called "learning" the network, is accomplished using the exact order-based structure discovery approach (Koivisto and Sood, 2004) where internal evaluation of each network, otherwise known as the scoring method, is accomplished using marginal likelihood (MacKay, 1992). Out of network decisions (such as the maximum number of parents or nodes that can predict a given node) were made using a penalty equal to the number of parameters or edges in the network following the same process as Akaike information criterion (AIC, Akaike, 1974). In addition to AIC, overfitting was further controlled using non-parametric bootstrapping, with edges only remaining in the network if they appeared in greater than 50% of the 5000 bootstrapped networks (Kratzer et al., 2017).

Logistic regression.—The conditional dependencies between plasma T4 and metamorphic success identified by the Bayesian network analyses at days 7 and 10 informed subsequent logistic regressions to characterize the predictive relationship. Every conditional dependency (i.e., edge) influencing metamorphic success at days 7 and 10 informed the

regressions between plasma T4 and the probability of metamorphic success where plasma T4 measures were treated as continuous covariates, treatment levels as factors, and metamorphic success as a binary response. NIS gene expression in thyroid glands and thyroid follicular cell number were not included in the Bayesian networks because they are compensatory responses that influence multiple nodes in a cyclic manner that would confound the directionality of the networks. However, the data showed treatment effects on NIS gene expression before changes in plasma T4 which suggested that this response could be used as a potential predictor of metamorphic success/failure. As such, we explored the use of NIS gene expression and thyroid follicular cell number as additional predictors of metamorphic success.

RESULTS

Thyroid gland iodotyrosines and iodothyronines

Iodotyrosine and iodothyronine levels in thyroid glands consist of both free and thyroglobulin-bound compounds and represent the aggregate iodination and coupling activity of TPO. The MMI-treated larvae showed concentration-dependent decreases in MIT, DIT, T3 and T4 in thyroid glands starting at 2 d of exposure and continuing through 10 d of exposure (Figure 2A, 2D, 2G, 2J, respectively). There was a significant increase in MIT at 7 d in the low MMI treatment (55 μM) that was a transient effect appearing to be a result of active compensation (Figure 2A). By 10 d, all analytes were <LLOQ in both the high (220 μM) and medium (110 μM) MMI treatments.

MBT exposure caused a significant concentration-dependent decrease in DIT and T4 in gland tissue starting at 4 d and continuing through 10 d (Figure 2E, 2K, respectively). Although MIT and T3 levels were also lower than control levels at 10 d in the high (1.6 μM) MBT treatment, statistical analysis yielded no evidence of a chemical-induced difference of these analytes at any of the four time points (Figure 2B, 2H, respectively).

Significant concentration-dependent decreases occurred in glandular DIT and T4 starting at 4 d of PTU exposure (Figure 2F, 2L, respectively). However, DIT profiles became similar to MIT and T3 profiles at 7 and 10 d, which exhibited hormetic responses (Figure 2C, 2I, respectively). MIT and T3 were significantly increased compared to control in the medium (39.2 μM) PTU treatment at 10 d while MIT and T3 in the high (117.5 μM) PTU treatment were <LLOQ; however, the analysis did not yield any evidence of an effect from the high PTU treatment (likely due to reduced statistical power from the loss of one tank). T4 maintained the same concentration-dependent decrease from 4 d through 10 d of PTU exposure, again with n = 2, the statistical test could not provide any evidence of a treatment effect at 10 d regardless of the high (117.5 μM) PTU treatment having T4 levels <LLOQ (Figure 2L).

Plasma iodothyronines

Plasma T3 levels were all <LLOQ at 2 and 4 d across all three studies. However, there were concentration-dependent decreases in both plasma T3 and T4 starting at 7 d and continuing through 10 d in the MMI-treated larvae (Figure 3A, 3D). Plasma T3 and T4 in the high (1.6

μM) MBT treatment were decreased at 7 d but analysis only showed evidence of a treatment effect at 10 d (Figure 3B, 3E). Larvae in the PTU study had overall higher levels of plasma T3 and T4 at 10 d but measurements were more variable (Figure 3C, 3F). The higher variability together with the loss of a tank in the high (117.5 μM) PTU treatment led to a lack of evidence of an effect despite clear reductions in both T3 and T4 in the high (117.5 μM) PTU treatment at 7 and 10 d.

Growth and development

There were no treatment-related effects on body mass through 10 days of exposure in any of the three studies (Supplemental Figure S.2) and overall growth rates appeared similar across all three studies. NF stage distributions over time were beginning to show delayed development in the higher treatments beginning at 7 d, although these relationships were not analyzed statistically. However, the statistical test yielded evidence of impacts on the number of days to NF stage 62 in the high treatments of all three studies and all treatments in the MMI study caused metamorphic development to be arrested (Figure 4).

Thyroid gland compensatory responses: NIS gene expression and follicular cell number

There were concentration-dependent increases in normalized NIS gene expression starting at 4 d of exposure in all three studies and continuing through 10 d (Figure 5A–5C). The magnitude of NIS mRNA induction varied by both study and treatment but maximal responses of almost 5-fold induction in the high treatments compared to control were observed at 7 d of exposure in both the MMI and PTU studies. Concentration-dependent increases of FCN were observed starting at 7 d of exposure in the MBT and PTU studies, but the changes were subtle compared to the substantial increases in cell numbers from the high treatments of all studies by 10 d of exposure (Figure 5D–5F). The highest magnitude of response was an approximately 3.4-fold increase in cell numbers between the control and high (1.6 μM) MBT treatment at 10 d of exposure, whereas the high treatments in the MMI (220 μM) and PTU (117.5 μM) studies were both greater than 2-fold higher than control follicular cell numbers.

Bayesian network analysis

Bayesian networks were generated based on the series of known and unknown potential causal relationships informed by the amphibian TPO AOP and other known thyroid-related biochemistry, together with the corresponding empirical data generated from these three in vivo exposures. Optimized networks for each time point are represented by directed acyclic graphs (DAGs) in Figure 6A-D (2, 4, 7 and 10 d, respectively). 5000 iterations of bootstrapping removed 2 (10%), 1 (5%), 3 (23%) and 6 (30%) edges from the 2, 4, 7 and 10 d DAGs, respectively. The 2 d and 4 d DAGs do not show any conditional dependencies directly between biochemical nodes and metamorphic success but only between chemical treatment level and metamorphic success, whereas the 7 d and 10 d DAGs show conditional dependency between plasma T4 and metamorphic success.

Logistic regression to predict metamorphic success

The 7 d and 10 d Bayesian networks corroborate what is already known about thyroid-mediated metamorphosis while also supporting the conditional dependency between plasma T4 and metamorphic success. Equipped with this information, the relationships between plasma T4 levels and probabilities of reaching NF stage 62 by day 22 were demonstrated via logistic regression for days 7 and 10 (Figure 7A and 7B, respectively). The conditional dependencies identified by the Bayesian networks informed whether a treatment was used as a covariate in the regressions. For example, the day 10 DAG in Figure 6D indicates that metamorphic success was conditionally dependent on the low MMI treatment node in addition to the plasma T4 node, so the low MMI treatment was used as a factor in the logistic regression to predict metamorphic success and is included as a separate trend line (Figure 7B). Examining Figures 6 and 7 reveals that plasma T4 becomes a better predictor of metamorphic success as time of exposure increases. Plasma T4 is non-predictive of metamorphic success during the first 4 d based on the lack of conditional dependency in the 2 d and 4 d DAGs (Figures 6A and 6B), gains predictive power by 7 d (Figures 6C and 7A), and even more predictive power at 10 d (Figures 6D and 7B). This trend is reinforced by the AIC at 10 days being lower than the AIC at 7 days (Table 1).

Thyroid gland NIS gene expression and FCN have historically been robust biomarkers of thyroid axis compensation in *X. laevis* (Hornung et al., 2015; Sternberg et al., 2011; Tietge et al., 2010, 2013). These biomarkers are relatively easy to measure and actively respond to TSH stimulation with NIS gene expression being more sensitive and responsive temporally than FCN. Since these endpoints are not causal key events in the AOP and instead result from system compensation, they were not included in the Bayesian network analyses; however, their response profiles suggest they have predictive properties. What's more, these compensatory response measures provide further weight of evidence that the organism is expending energy to counteract the chemical insult via thyroid-specific processes. To evaluate their predictive capabilities at the 10 d time point, they were used as regressors in place of plasma T4 to predict probabilities of metamorphic success via logistic regression. On the basis of AIC scores, the NIS model at 10 days of exposure was more predictive than the 10 d FCN model, but neither were more predictive than 10 d plasma T4 (Table 1). To evaluate the predictive capability of the combined effects on plasma T4 and NIS expression at 10 d (the most predictive endpoints independently), plasma T4 values were adjusted based on NIS expression levels and the adjusted T4 values were used as the regressor. This model (Figure 8) demonstrated the lowest AIC score and highest predictive capability when ignoring the direct effects from treatment on metamorphic success.

The definition of metamorphic success can have an impact on the prediction of probability of success. For all predictions thus far, the definition of metamorphic success has been based on the 75^{th} percentile of controls reaching NF stage 62, which was 22 days. In order to demonstrate the impact on how metamorphic success was defined, the probability distributions were derived for the median day controls reached NF stage 62 (20 days, corresponding to the 50^{th} percentile) as opposed to 22 days. Logistic regression was performed using the plasma T4 values at 10 d of exposure with the probability distributions for reaching NF stage 62 by 20 days. The comparison of probability ranges of success

reaching NF stage 62 by either 20 days or 22 days given a range of plasma T4 levels at 10 d post NF 53/54 is presented in Table 2.

DISCUSSION

Thyroid-mediated amphibian metamorphosis is a well-characterized biological process that has been studied for decades. Characterization of how these model thyroid disrupting chemicals affect amphibian metamorphosis at the apical level has already been done using either the Amphibian Metamorphosis Assay (AMA) or a study design similar to the AMA (Coady et al., 2010; Degitz et al., 2005; Tietge et al., 2010, 2013). What has not been demonstrated before is a mathematical predictive relationship between thyroid hormone levels in developing *X. laevis* and an adverse apical outcome like metamorphic failure. The predictive linkages described here will permit inference of an adverse apical effect based on a biochemical response – a response that can be quantitatively linked to a common molecular initiating event that is currently interrogated by in vitro methods. These pathway-based predictive linkages are critical for transitioning away from animal testing and towards an increased reliance on in vitro data and in silico modeling for chemical safety evaluation.

It is worth noting that assessment of metamorphosis differs in the current study design compared to the standardized AMA guideline. Evaluating days to NF stage 62 allows a comparable continuous data point for every larva, whereas the AMA terminates at 21 d and relies on distributions of ordinal NF stage data (OECD, 2009; U.S. EPA, 2009b); this requires developmental stage-matching to appropriately compare other endpoints. When there are significant treatment-related developmental delays in the AMA, it precludes statistical analysis for those treatments not having stage-matched individuals in the controls. The other advantage to evaluating days to NF stage 62 is that this stage represents metamorphic climax, which is when THs in the blood are at peak levels (Sternberg et al., 2011). This means any perturbation upstream of blood TH levels has been successfully compensated for if the individual achieves NF stage 62. In general, we believe an individual achieving NF stage 62 should be considered as undergoing successful metamorphosis. Conversely, apical adversity would be defined as not achieving NF stage 62 at the same rate as controls and has relevance to population-level effects, especially for amphibian species that rely on ephemeral pools for larval development.

The superior benefit to using the amphibian model for thyroid toxicology research is the well-characterized linkage between thyroid function and an easily measured phenotype. Considering the objective of this study, we felt the AOP for TPO inhibition leading to altered amphibian metamorphosis was a sensible foundation from which to demonstrate development of a surrogate measure of an apical outcome. To evaluate the timing and magnitude of response relationships across KEs that could be predictive of the apical outcome, model TPO inhibitors were used to intentionally elicit expected responses, including arresting metamorphosis so the early biochemical responses could be quantitatively linked to an apical outcome. These chemicals are well-characterized and potent TPO inhibitors, but their toxicokinetics differ with regard to TPO inhibition. This is demonstrated in part by the marked difference in concentration-response profiles observed in glandular MIT, DIT and T3 between MMI and PTU (Figure 2). Regardless of these

differences in inhibitory characteristics toward TPO in the intermediate concentration treatments, all three of these model chemicals effectively blocked TH synthesis in all of the high concentration treatments allowing an evaluation of concordance of downstream effects. At the highest concentrations of all three chemicals, there was strong concordance of effects throughout the KEs measured; specifically, by 10 d of exposure, iodinated products of TPO activity in the gland and TH in the plasma were <LLOQ in most cases (Figures 2 & 3). In addition, the compensatory responses of NIS transcriptional upregulation and increased FCN (Figure 5) are indicative of increased circulating TSH in response to decreased TH feedback at the hypothalamus-pituitary. Overall, this study generated the intended quantitative data to characterize the relationships between upstream KEs and an adverse outcome and supports the concept that this specific AOP can reasonably be considered chemical agnostic when TPO is severely inhibited (e.g., near 100%).

With a mechanistic understanding of the relationships between these data, Bayesian networks were developed to integrate the three datasets into a single probabilistic model per time point. This combined data-driven and knowledge-driven analysis helped to interpret conditional dependencies mathematically and identify any potential anomalies that might challenge the intended goal of utilizing biochemical profiles to make predictions of outcome. As a result, the computationally learned graphical models generally confirmed what can be qualitatively assembled from the biochemical and time to NF stage 62 data profiles (Figures 2, 3 & 4). At 2 d and 4 d, there are no identified conditional dependencies between plasma T3/T4 profiles and metamorphic success whereas at 7 d and 10 d, there is a conditional dependency between plasma T4 and metamorphic success. This suggests plasma T4 is a better predictor of metamorphic success than plasma T3 given these data, and the predictive relationship between plasma T4 and metamorphic success becomes established beyond 4 d in the current test conditions. The increased resolution of this predictive relationship is supported by the improved fit of the logistic regression models through time (Figure 7). At 7 d of exposure, there are a series of data points associated with the low MMI and high PTU treatments that have probabilities of success equal to zero (P(NF62) = 0); however, plasma T4 levels are within the same range as plasma T4 levels of treatments that develop successfully (Figure 7A). These treatments were identified in the 7 d Bayesian network as covariates of metamorphic success alongside plasma T4, and in the essence of Bayesian statistics, are considered to be acting through one or more “hidden” nodes in the network. We propose this hidden node is related to plasma T4 levels through time (i.e., future plasma T4 levels). This phenomenon is demonstrated by the shifting of plasma T4 concentrations to the left through time in the metamorphic failures and can be observed by the high PTU treatment no longer being identified as a covariate of metamorphic success in the 10 d Bayesian network (Figure 6D). Ostensibly, only at 10 d of exposure can the high PTU treatment’s impact on metamorphic success be identified as acting through biochemical nodes. Also, comparing Figures 7A and 7B, plasma T4 concentrations are shifting further to the left (i.e., decreasing) in metamorphic failures while plasma T4 concentrations are shifting further to the right (i.e., increasing) in metamorphic successes. This supports the idea that if plasma T4 sampling were to occur at 14 d, this predictive relationship would be even stronger.

NIS gene expression and thyroid FCN are compensatory responses to thyroid axis disruption and can be diagnostic of thyroid status; inherently, they have some degree of predictivity toward thyroid-related outcomes. These endpoints were not included in the Bayesian network analyses because they are common effects of the treatments and not part of the causal linkages leading to the apical outcome. It was determined through logistic regression that these endpoints were not as predictive of metamorphic success as plasma T4 on their own. However, the combination of NIS gene expression and plasma T4 showed improved predictivity of metamorphic success as compared to plasma T4 alone. Of all the biochemical endpoints assessed in this study, NIS gene expression and plasma T4 can be the most straightforward and routine to measure making these two measurements practical to include in future targeted study designs intended to predict metamorphic failure via thyroid disruption.

According to the logistic regression models presented here that relate plasma T4 to probabilities of metamorphic success, the ranges of plasma T4 values in the 10 d models span an order of magnitude for probabilities of metamorphic success similar to the controls. Figures 7B and 8 show that any data points with less than 0.6 probability of metamorphic success are all flatlined at P(NF62) = 0.0, leaving a void of empirical data to suggest intermediate effect ranges can occur within the relationship between plasma T4 levels and metamorphic success. This is characteristic of a toxicological “tipping point” (Shah et al., 2016) where the system can no longer compensate to maintain plasma T4 levels and results in an apical outcome resembling, in this case, a binary response. Taken together, the *X. laevis* thyroid axis appears to be extremely robust in maintaining enough plasma T4 to undergo metamorphosis and it takes severe inhibition of TPO to impact circulating levels of T4 to the point of affecting metamorphic success; an order of magnitude or more decrease in circulating T4 is required to impact metamorphic success.

Although this can be demonstrated via TPO inhibition, it is unclear whether this level of impact on circulating T4 can occur through other individual, or more likely, combinations of mechanisms of thyroid disruption.

This study demonstrates that plasma T4 can be used as a predictive surrogate metric for metamorphic success when thyroid hormone synthesis is reduced due to TPO inhibition. Utilization of plasma T4 as an upstream biomarker of metamorphic success offers the possibility of more targeted, abbreviated in vivo studies to assess a chemical’s potential to cause apical effects via thyroid disruption. However, without assessments of additional upstream KEs, this measure will remain incapable of being diagnostic of the MIE leading to the plasma T4 perturbation. More work is necessary to quantitatively link the various thyroid-related MIEs currently being represented by in vitro screening assays (e.g., NIS inhibition, iodotyrosine deiodinase [IYD] inhibition, etc.) to plasma T4 levels, in addition to development of other surrogate biomarkers that can predict apical adverse outcomes occurring via MIEs that may not affect plasma T4 levels. Herein, characterization of control and perturbed *X. laevis* thyroid biology lends suitable data for computational model development that can eventually interface with in vitro screening data to predict circulating T4 concentrations, which is a metric that can now be applied to predict metamorphic success/failure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

The authors would like to thank RaeAnn Schulte, Kelby Donnay and Carsten Knutsen for their technical contributions that were key to the success of the in vivo studies and data generation. Additionally, the authors greatly appreciate the input provided by Dr. Katie O'Shaughnessy to improve this manuscript and Chad Blanksma for helping prepare the figures.

REFERENCES

Akaike H (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716–723.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK & Serrrano JA (2010). Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environmental Toxicology and Chemistry: An International Journal, 29(3), 730–741.

Axelstad M, Hansen PR, Boberg J, Bonnichsen M, Nellemann C, Lund SP, Hougaard KS & Hass U (2008). Developmental neurotoxicity of propylthiouracil (PTU) in rats: relationship between transient hypothyroxinemia during development and long-lasting behavioural and functional changes. Toxicology and applied pharmacology, 232(1), 1–13. [PubMed: 18573268]

Beech SG, Walker SW, Dorrance AM, Arthur JR, Nicol F, Lee D, & Beckett GJ (1993). The role of thyroidal type-I iodothyronine deiodinase in tri-iodothyronine production by human and sheep thyrocytes in primary culture. Journal of Endocrinology, 136(3), 361–370.

Benjamini Y, & Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1), 289–300.

Brown DD (2005). The role of deiodinases in amphibian metamorphosis. Thyroid, 15(8), 815–821. [PubMed: 16131324]

Brown DD, & Cai L (2007). Amphibian metamorphosis. Developmental biology, 306(1), 20. [PubMed: 17449026]

Buckalew AR, Wang J, Murr AS, Deisenroth C, Stewart WM, Stoker TE, & Laws SC (2020). Evaluation of potential sodium-iodide symporter (NIS) inhibitors using a secondary Fischer rat thyroid follicular cell (FRTL-5) radioactive iodide uptake (RAIU) assay. Archives of Toxicology, 1–13.

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. Environmental Toxicology and Chemistry: An International Journal, 29(4), 869–880.

Crofton K, Gilbert M, Paul Friedman K, Demeneix B, Marty M, Zoeller T (2018). Inhibition of Thyroperoxidase and Subsequent Adverse Neurodevelopmental Outcomes in Mammals. AOP-Wiki. Society for Advancement of AOPs [cited 2020 February 21]. Available from: https://aopwiki.org/aops/42

Ciullo PA, Hewitt N, Ciullo PA, & Hewitt N (1999). Compounding materials. The Rubber Formulary, 4–49.

Degitz SJ, Holcombe GW, Flynn KM, Kosian PA, Korte JJ, & Tietge JE (2005). Progress towards development of an amphibian-based thyroid screening assay using *Xenopus laevis*. Organismal and thyroidal responses to the model compounds 6-propylthiouracil, methimazole, and thyroxine. Toxicological sciences, 87(2), 353–364. [PubMed: 16002479]

Deisenroth C, Soldatow VY, Ford J, Stewart W, Brinkman C, LeCluyse EL, MacMillan DK & Thomas RS (2019). Development of an In Vitro Human Thyroid Microtissue Model for Chemical Screening. Toxicological Sciences.

Deme D, Pommier J, & Nunez J (1978). Specificity of thyroid hormone synthesis the role of thyroid peroxidase. Biochimica et Biophysica Acta (BBA)-General Subjects, 540(1), 73–82. [PubMed: 25093]

Dong H, Atlas E, & Wade MG (2019). Development of a non-radioactive screening assay to detect chemicals disrupting the human sodium iodide symporter activity. Toxicology in Vitro, 57, 39–47. [PubMed: 30738889]

Fort DJ, Degitz S, Tietge J, & Touart LW (2007). The hypothalamic-pituitary-thyroid (HPT) axis in frogs and its role in frog development and reproduction. Critical reviews in toxicology, 37(1–2), 117–161. [PubMed: 17364707]

Fox J and Weisberg S (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Galton VA (1983). Thyroid hormone action in amphibian metamorphosis. In Molecular basis of thyroid hormone action (pp. 445–483). Academic Press New York.

Gilbert ME, & Zoeller RT (2010). Thyroid hormone—impact on the developing brain: possible mechanisms of neurotoxicity. Neurotoxicology, 79–111.

Gilbert ME (2011). Impact of low-level thyroid hormone disruption induced by propylthiouracil on brain development and function. Toxicological Sciences, 124(2), 432–445. [PubMed: 21964421]

Gilbert ME, Rovet J, Chen Z, & Koibuchi N (2012). Developmental thyroid hormone disruption: prevalence, environmental contaminants and neurodevelopmental consequences. Neurotoxicology, 33(4), 842–852. [PubMed: 22138353]

Hallinger DR, Murr AS, Buckalew AR, Simmons SO, Stoker TE, & Laws SC (2017). Development of a screening approach to detect thyroid disrupting chemicals that inhibit the human sodium iodide symporter (NIS). Toxicology in Vitro, 40, 66–78. [PubMed: 27979590]

Haselman J, Hornung M, Degitz S (2018). Thyroperoxidase inhibition leading to altered amphibian metamorphosis. AOP-Wiki. Society for Advancement of AOPs [cited 2020 February 21]. Available from: https://aopwiki.org/aops/175

Hassan I, El-Masri H, Kosian PA, Ford J, Degitz SJ, & Gilbert ME (2017). Neurodevelopment and thyroid hormone synthesis inhibition in the rat: Quantitative understanding within the adverse outcome pathway framework. Toxicological Sciences, 160(1), 57–73. [PubMed: 28973696]

Hassan I, El-Masri H, Ford J, Brennan A, Handa S, Paul Friedman K, & Gilbert ME (2020). Extrapolating In Vitro Screening Assay Data for Thyroperoxidase Inhibition to Predict Serum Thyroid Hormones in the Rat. Toxicological Sciences, 173(2), 280–292. [PubMed: 31697382]

Hornung MW, Kosian PA, Haselman JT, Korte JJ, Challis K, Macherla C, Nevalainen E & Degitz SJ (2015). In vitro, ex vivo, and in vivo determination of thyroid hormone modulating activity of benzothiazoles. Toxicological Sciences, 146(2), 254–264. [PubMed: 25953703]

Hornung MW, Korte JJ, Olker JH, Denny JS, Knutsen C, Hartig PC, Cardon MC & Degitz SJ (2018). Screening the ToxCast phase 1 chemical library for inhibition of deiodinase type 1 activity. Toxicological Sciences, 162(2), 570–581. [PubMed: 29228274]

Hothorn T, Bretz F, & Westfall P (2008). Simultaneous inference in general parametric models. Biometrical Journal: Journal of Mathematical Methods in Biosciences, 50(3), 346–363.

Jafari H, Akbarzade K, & Danaee I (2014). Corrosion inhibition of carbon steel immersed in a 1 M HCl solution using benzothiazole derivatives. Arabian Journal of Chemistry.

Kessler J, Obinger C, & Eales G (2008). Factors influencing the study of peroxidase-generated iodine species and implications for thyroglobulin synthesis. Thyroid, 18(7), 769–774. [PubMed: 18631006]

Koivisto M, & Sood K (2004). Exact Bayesian structure discovery in Bayesian networks. Journal of Machine Learning Research, 5(May), 549–573.

Kratzer G, Pittavino M, Lewis F, & Furrer R (2016). abn: an R package for modelling multivariate data using additive Bayesian networks. R package version, 1.

Lewis FI (2016). ABN: modelling multivariate data with additive bayesian networks. R package version 1.0. 2.

Lewis FI, & Ward MP (2013). Improving epidemiologic data analyses through multivariate regression modelling. Emerging themes in epidemiology, 10(1), 4. [PubMed: 23683753]

MacKay DJ (1992). Bayesian interpolation. Neural computation, 4(3), 415–447.

Morvan-Dubois G, Demeneix BA, & Sachs LM (2008). *Xenopus laevis* as a model for studying thyroid hormone signalling: from development to metamorphosis. Molecular and cellular endocrinology, 293(1–2), 71–79. [PubMed: 18657589]

Murk AJ, Rijntjes E, Blaauboer BJ, Clewell R, Crofton KM, Dingemans MM, Furlow JD, Kavlock R, Köhrle J, Opitz R & Traas T (2013). Mechanism-based testing strategy using in vitro approaches for identification of thyroid hormone disrupting chemicals. Toxicology in vitro, 27(4), 1320–1346. [PubMed: 23453986]

National Research Council. (2007). Toxicity testing in the 21st century: a vision and a strategy. National Academies Press.

Nieuwkoop PD, & Faber J (1994). Normal table of. Xenopus laevis, 252.

Noyes PD, Friedman KP, Browne P, Haselman JT, Gilbert ME, Hornung MW, Barone S Jr., Crofton KM, Laws SC, Stoker TE, Simmons SO, Tietge JE, & Degitz SJ (2019). Evaluating chemicals for thyroid disruption: opportunities and challenges with in vitro testing and adverse outcome pathway approaches. Environmental health perspectives, 127(9), 095001.

OECD. (2009). Test No. 231: Amphibian Metamorphosis Assay, OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris, DOI: 10.1787/9789264076242-en.

OECD. (2015). Test No. 241: The Larval Amphibian Growth and Development Assay (LAGDA), OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris. DOI: 10.1787/9789264242340-en.

OECD. (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, 10.1787/9789264304741-en.

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. Toxicological Sciences, 166(2), 318–331. [PubMed: 30137636]

Olker JH, Korte JJ, Denny JS, Hartig PC, Cardon MC, Knutsen CN, Kent PM, Christensen JP, Degitz SJ & Hornung MW (2019). Screening the ToxCast phase 1, phase 2, and e1k chemical libraries for inhibitors of iodothyronine deiodinases. Toxicological Sciences, 168(2), 430–442. [PubMed: 30561685]

Paul KB, Hedge JM, Macherla C, Filer DL, Burgess E, Simmons SO, Crofton KM & Hornung MW (2013). Cross-species analysis of thyroperoxidase inhibition by xenobiotics demonstrates conservation of response between pig and rat. Toxicology, 312, 97–107. [PubMed: 23959146]

Paul KB, Hedge JM, Rotroff DM, Hornung MW, Crofton KM, & Simmons SO (2014). Development of a thyroperoxidase inhibition assay for high-throughput screening. Chemical research in toxicology, 27(3), 387–399. [PubMed: 24383450]

Paul Friedman K, Watt ED, Hornung MW, Hedge JM, Judson RS, Crofton KM, Houck KA & Simmons SO (2016). Tiered high-throughput screening approach to identify thyroperoxidase inhibitors within the ToxCast phase I and II chemical libraries. Toxicological Sciences, 151(1), 160–180. [PubMed: 26884060]

Paul-Friedman K, Martin M, Crofton KM, Hsu CW, Sakamuru S, Zhao J, Xia M, Huang R, Stavreva DA, Soni V & Varticovski L (2019). Limited Chemical Structural Diversity Found to Modulate Thyroid Hormone Receptor in the Tox21 Chemical Library. Environmental Health Perspectives, 127(9), 097009.

Pearl J (1986). Fusion, propagation, and structuring in belief networks. Artificial intelligence, 29(3), 241–288.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: www.R-project.org/.

Rue H, Martino S, & Chopin N (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the royal statistical society: Series b (statistical methodology), 71(2), 319–392.

Ruf J, & Carayon P (2006). Structural and functional aspects of thyroid peroxidase. Archives of biochemistry and biophysics, 445(2), 269–277. [PubMed: 16098474]

Sachs LM, & Buchholz DR (2017). Frogs model man: In vivo thyroid hormone signaling during development. genesis, 55(1–2), e23000.

Shi YB, Wong J, Puzianowska-Kuznicka M, & Stolow MA (1996). Tadpole competence and tissue-specific temporal regulation of amphibian metamorphosis: roles of thyroid hormone and its receptors. Bioessays, 18(5), 391–399. [PubMed: 8639162]

Shah I, Setzer RW, Jack J, Houck KA, Judson RS, Knudsen TB, Liu J, Martin MT, Reif DM, Richard AM & Thomas RS (2016). Using ToxCast™ data to reconstruct dynamic cell state trajectories and estimate toxicological points of departure. Environmental health perspectives, 124(7), 910–919. [PubMed: 26473631]

Sternberg RM, Thoemke KR, Korte JJ, Moen SM, Olson JM, Korte L, Tietge JE & Degitz SJ Jr (2011). Control of pituitary thyroid-stimulating hormone synthesis and secretion by thyroid hormones during *Xenopus* metamorphosis. General and comparative endocrinology, 173(3), 428–437. [PubMed: 21803044]

Tata JR (2006). Amphibian metamorphosis as a model for the developmental actions of thyroid hormone. Molecular and cellular endocrinology, 246(1–2), 10–20. [PubMed: 16413959]

Taurog A, Dorris ML, & Doerge DR (1996). Mechanism of simultaneous iodination and coupling catalyzed by thyroid peroxidase. Archives of biochemistry and biophysics, 330(1), 24–32. [PubMed: 8651700]

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. Environmental Toxicology and Chemistry: An International Journal, 24(4), 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*. Aquatic toxicology, 98(1), 44–50. [PubMed: 20153061]

Tietge JE, Degitz SJ, Haselman JT, Butterworth BC, Korte JJ, Kosian PA, Lindberg-Livingston AJ, Burgess EM, Blackshear PE & Hornung MW (2013). Inhibition of the thyroid hormone pathway in *Xenopus laevis* by 2mercaptobenzothiazole. Aquatic toxicology, 126, 128–136. [PubMed: 23178179]

Therneau TM (2018). coxme: mixed effects Cox models. R package version 2.2–10 2018.

U.S. EPA. (2009a). Series 890—Endocrine Disruptor Screening Program Test Guidelines. Washington DC: U.S. EPA, Endocrine Disruptor Screening Program www.epa.gov/test-guidelines-pesticides-and-toxic-substances/series-890-endocrine-disruptor-screening-program [accessed 21 February 2020].

U.S. EPA. (2009b). OCSPP 890.1100: Amphibian Metamorphosis Assay (AMA), Endocrine Disruptor Screening Program Test Guidelines, 890 series, www.regulations.gov, ID: EPA-HQ-OPPT-2009–0576.

U.S. EPA. (2015). OCSPP 890.2300: Larval Amphibian Growth and Development Assay (LAGDA), Endocrine Disruptor Screening Program Test Guidelines, 890 series, www.regulations.gov, ID: EPA-HQ-OPPT-2014–0766-0020.

Wang J, Hallinger DR, Murr AS, Buckalew AR, Simmons SO, Laws SC, & Stoker TE (2018). High-throughput screening and quantitative chemical ranking for sodium-iodide symporter inhibitors in ToxCast phase I chemical library. Environmental science & technology, 52(9), 5417–5426. [PubMed: 29611697]

Wegner S, Browne P, & Dix D (2016). Identifying reference chemicals for thyroid bioactivity screening. Reproductive Toxicology, 65, 402–413. [PubMed: 27589887]

Zoeller RT, & Crofton KM (2005). Mode of action: developmental thyroid hormone insufficiency—neurological abnormalities resulting from exposure to propylthiouracil. Critical reviews in toxicology, 35(8–9), 771–781. [PubMed: 16417044]

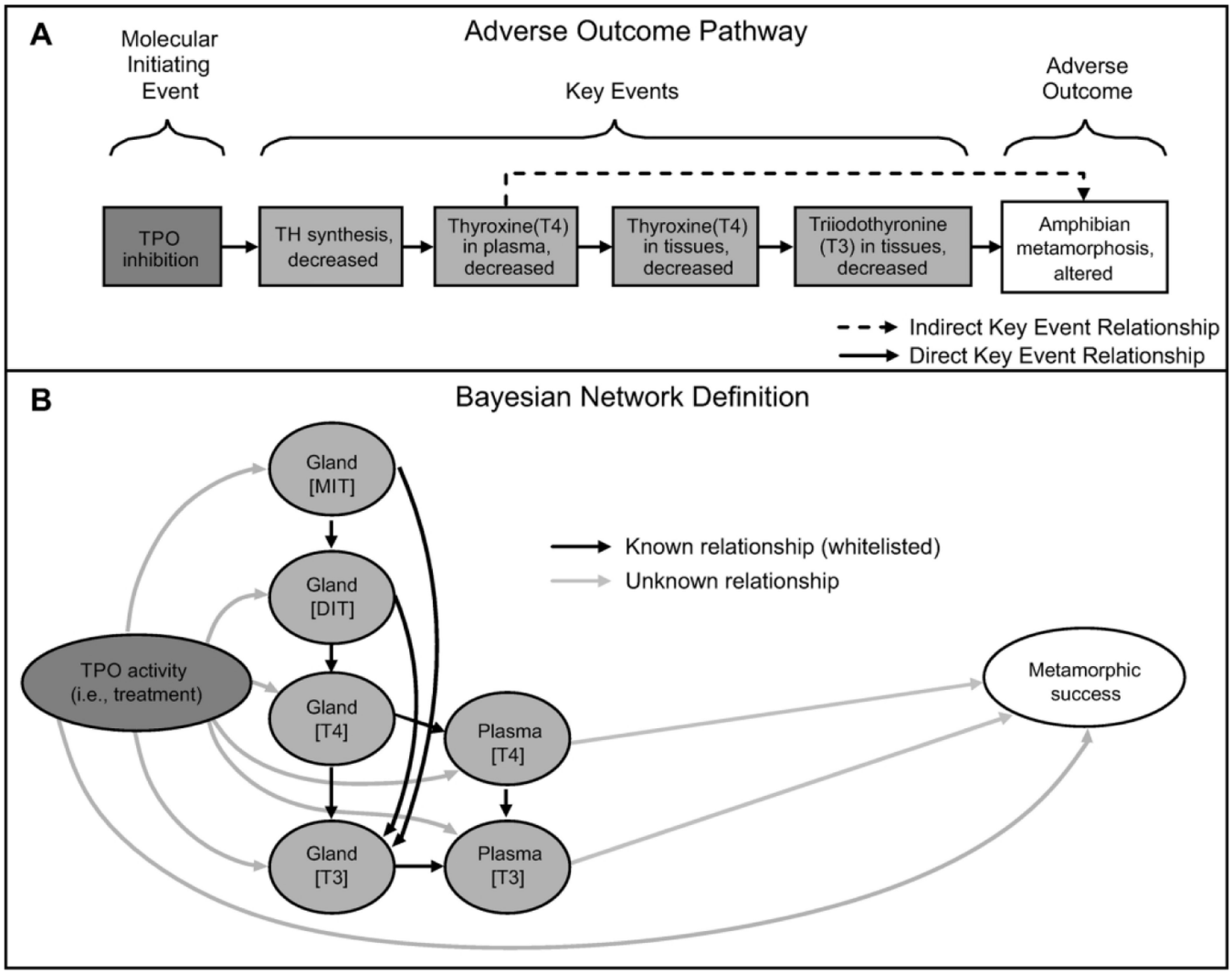

Figure 1.

The adverse outcome pathway for thyroperoxidase (TPO) inhibition leading to altered amphibian metamorphosis (A). The specified relationships defining the Bayesian network (B). TH, thyroid hormone; MIT, monoiodotyrosine; DIT, diiodotyrosine; T4, thyroxine; T3, triiodothyronine.

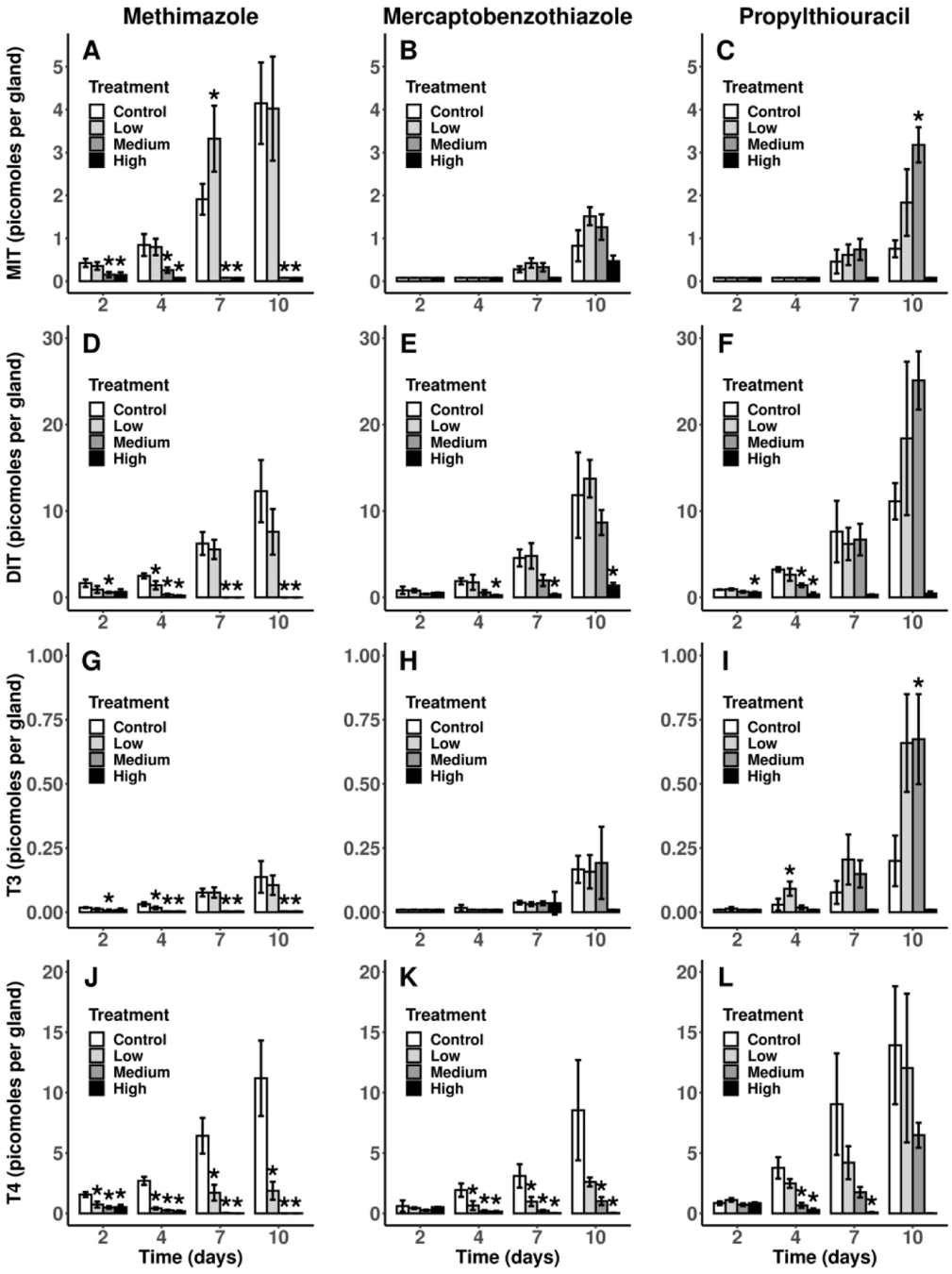

Figure 2.

Mean (± SD) levels of iodotyrosines (A-F) and iodothyronines (G-L) in thyroid glands of control *X. laevis* and larvae exposed to each of three model thyroperoxidase inhibitors during pro-metamorphosis at 2, 4, 7 and 10 d post NF 53/54. Sample sizes are n = 3 pools except for the high PTU treatment which has a sample size of n = 2 pools. Treatments with evidence of a difference with the control (FDR-adjusted $p < 0.05$) are indicated with an asterisk (*). Note - mean values represent analyte levels from a single gland equivalent. Since larvae have two glands, means should be doubled to derive the estimate representing

the total amount within the thyroid gland tissue per larvae. MIT, monoiodotyrosine; DIT, diiodotyrosine; T4, thyroxine; T3, triiodothyronine; MMI, methimazole; MBT, mercaptobenzothiazole; PTU, propylthiouracil; MMI:Low, 55 μM; MMI:Med, 110 μM; MMI:High, 220 μM; MBT:Low, 0.18 μM; MBT:Med, 0.54 μM; MBT:High, 1.6 μM; PTU:Low, 13 μM; PTU:Med, 39 μM; PTU:High, 117.5 μM.

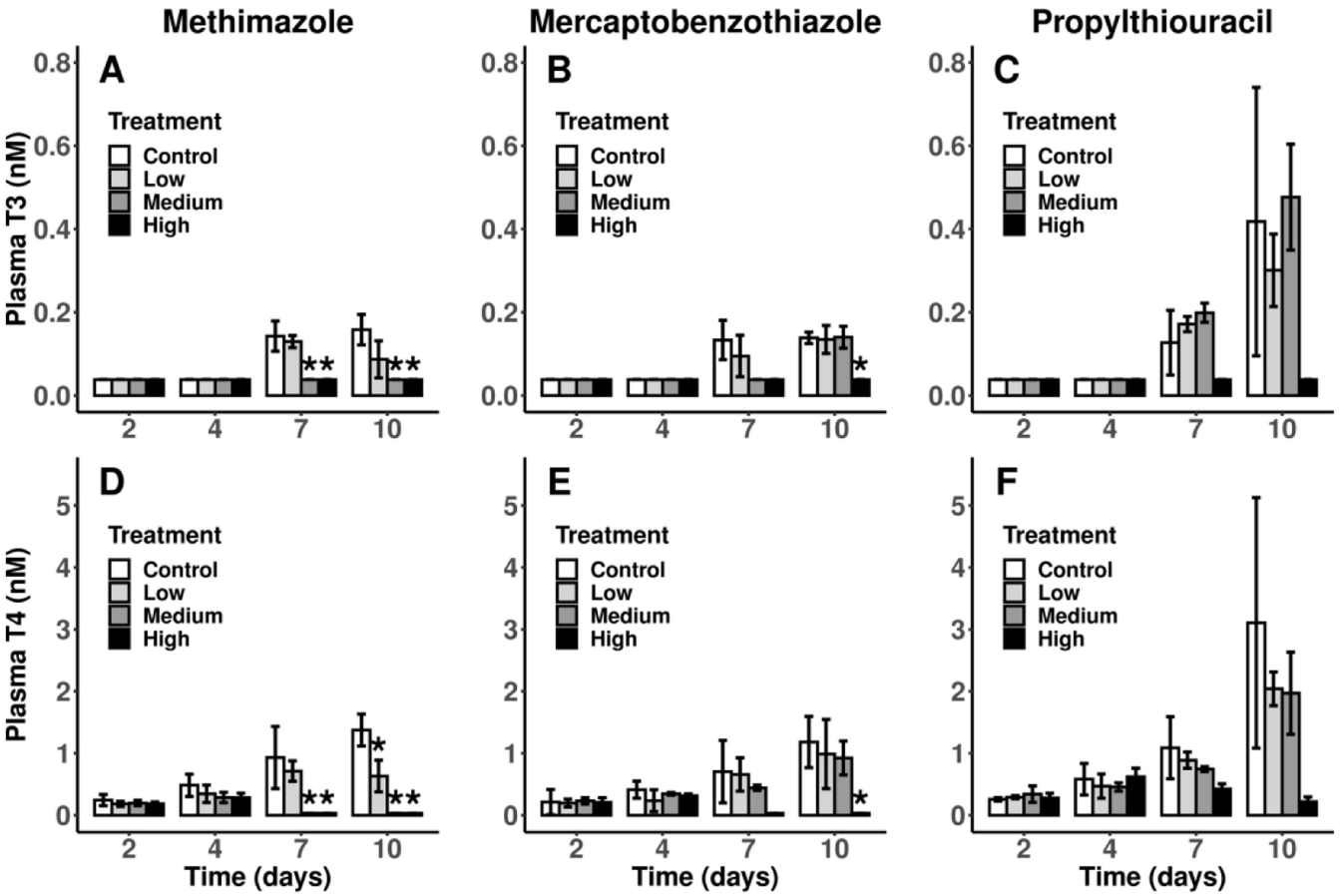

Figure 3.

Mean (± SD) levels of iodothyronines in blood plasma of control *X. laevis* and larvae exposed to each of three model thyroperoxidase inhibitors during pro-metamorphosis at 2, 4, 7 and 10 d post NF 53/54. Sample sizes are n = 3 pools except for the high PTU treatment which has a sample size of n = 2 pools. Treatments with evidence of a difference with the control (FDR-adjusted $p < 0.05$) are indicated with an asterisk (*). T4, thyroxine; T3, triiodothyronine; MMI, methimazole; MBT, mercaptobenzothiazole; PTU, propylthiouracil; MMI:Low, 55 μM; MMI:Med, 110 μM; MMI:High, 220 μM; MBT:Low, 0.18 μM; MBT:Med, 0.54 μM; MBT:High, 1.6 μM; PTU:Low, 13 μM; PTU:Med, 39 μM; PTU:High, 117.5 μM.

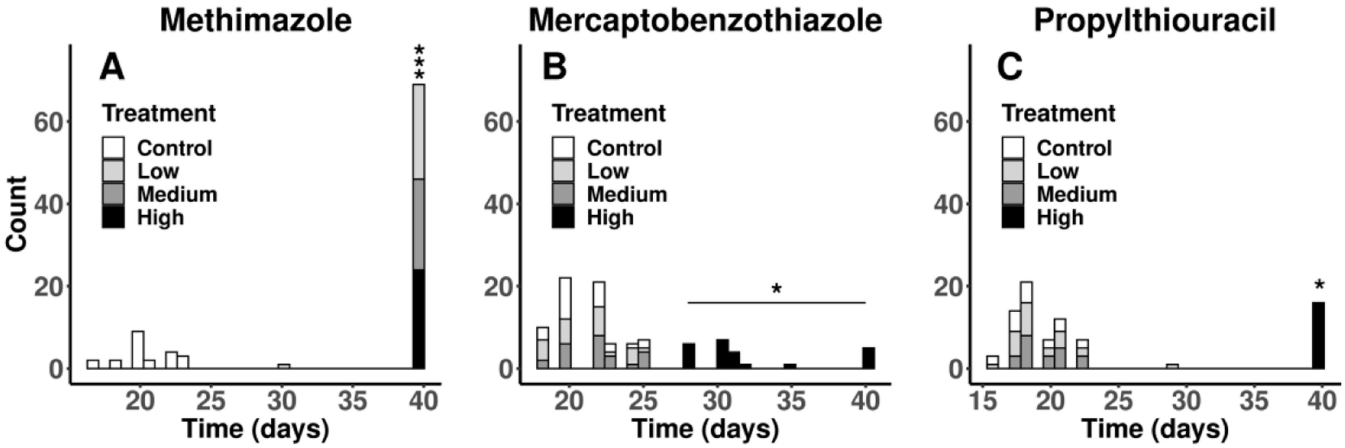

Figure 4.

Number of unexposed *X. laevis* larvae and larvae exposed to each of three model thyroperoxidase inhibitors reaching NF stage 62 by day or showing arrested metamorphic development (40 days on graphs). Treatments with evidence of a difference with the control ($p < 0.05$) are indicated with an asterisk (*). MMI, methimazole; MBT, mercaptobenzothiazole; PTU, propylthiouracil; MMI:Low, 55 μM; MMI:Med, 110 μM; MMI:High, 220 μM; MBT:Low, 0.18 μM; MBT:Med, 0.54 μM; MBT:High, 1.6 μM; PTU:Low, 13 μM; PTU:Med, 39 μM; PTU:High, 117.5 μM.

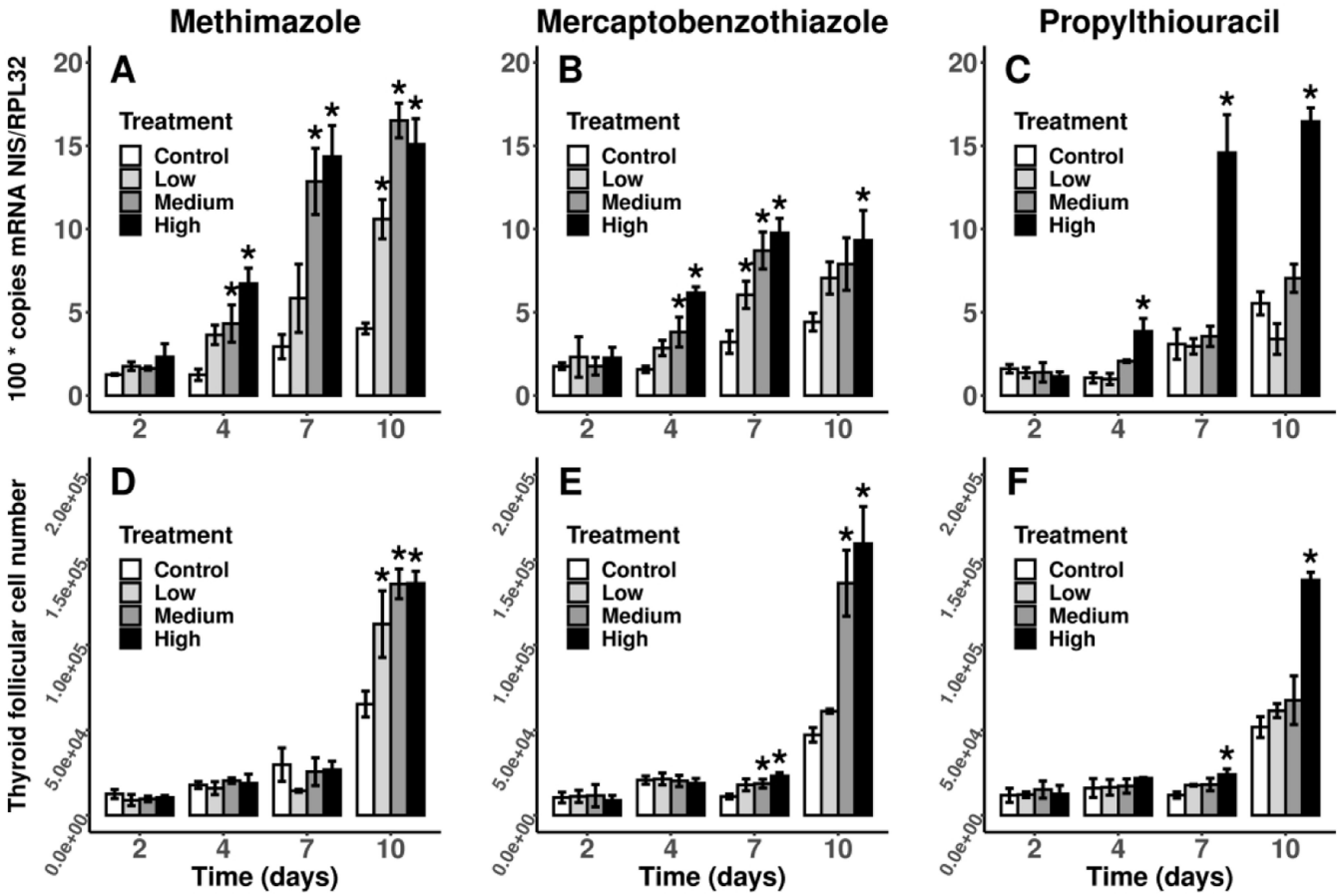

Figure 5.

Mean (± SD) levels of normalized copies of sodium-iodide symporter (NIS) mRNA (A-C) in thyroid gland tissue and numbers of thyroid follicular cell numbers (FCN) per larva (D-F) in control *X. laevis* and larvae exposed to each of three model thyroperoxidase inhibitors during pro-metamorphosis at 2, 4, 7 and 10 d post NF 53/54. Sample sizes are n = 3 pools except for the high PTU treatment which has a sample size of n = 2 pools. Treatments with evidence of a difference with the control (FDR-adjusted $p < 0.05$) are indicated with an asterisk (*). MMI, methimazole; MBT, mercaptobenzothiazole; PTU, propylthiouracil; MMI:Low, 55 μM; MMI:Med, 110 μM; MMI:High, 220 μM; MBT:Low, 0.18 μM; MBT:Med, 0.54 μM; MBT:High, 1.6 μM; PTU:Low, 13 μM; PTU:Med, 39 μM; PTU:High, 117.5 μM.

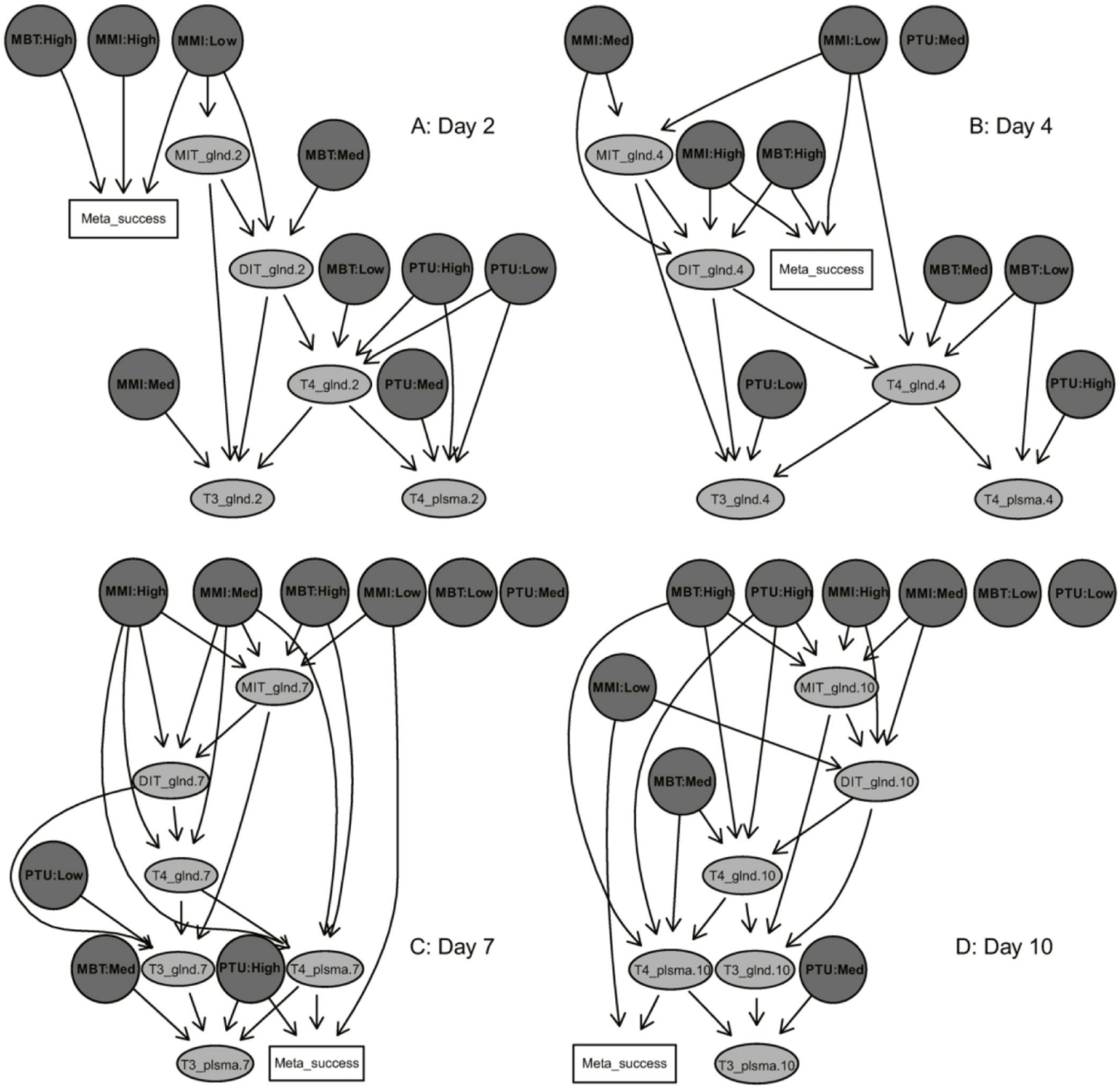

Figure 6.

Directed acyclic graphs (DAGs) of Bayesian networks at exposure days 2 (A), 4 (B), 7 (C) and 10 (D). Edges (arrows) represent conditional dependencies, or covariates, affecting downstream nodes as determined by the network analysis and subsequent bootstrapping. Chemical treatments are circles shaded dark gray with bold font, biochemical nodes are ovals shaded light gray and the apical outcome is a rectangle without shading. MMI, methimazole; MBT, mercaptobenzothiazole; PTU, propylthiouracil; MIT, monoiodotyrosine; DIT, diiodotyrosine; T4, thyroxine; T3, triiodothyronine; glnd, thyroid gland; plsma, plasma; Meta, metamorphic; MMI:Low, 55 μM; MMI:Med, 110 μM;

MMI:High, 220 μM; MBT:Low, 0.18 μM; MBT:Med, 0.54 μM; MBT:High, 1.6 μM; PTU:Low, 13 μM; PTU:Med, 39 μM; PTU:High, 117.5 μM.

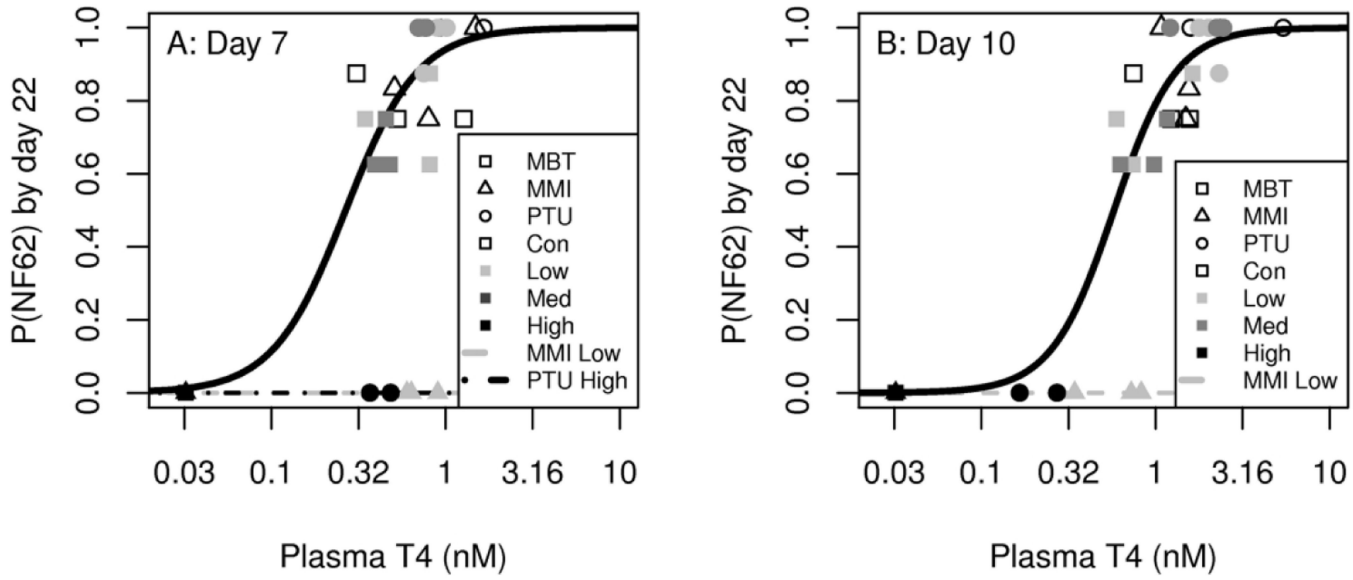

Figure 7.

Probability of reaching NF stage 62 (P(NF62)) by day 22 based on plasma thyroxine (T4) levels at 7 and 10 d of exposure (A and B, respectively). Regressions were informed by the Bayesian networks; if a treatment was identified as a covariate of metamorphic success, it was used as a factor in the resulting regression and then plotted with a separate trend line. A test for goodness of fit fails to indicate a lack of fit for both models with p-values of 0.41 (A) and 0.092 (B). Note that plasma T4 was only identified as a covariate of metamorphic success at days 7 and 10 (Figure 6C and 6D). Plasma T4 values that were <LLOQ (0.0644 nM) are overlapping in the lower left and include all three replicates of MMI:High, MMI:Med, and MBT:High. MMI, methimazole; MBT, mercaptobenzothiazole; PTU, propylthiouracil; Con, control; MMI:Low, 55 μM; MMI:Med, 110 μM; MMI:High, 220 μM; MBT:Low, 0.18 μM; MBT:Med, 0.54 μM; MBT:High, 1.6 μM; PTU:Low, 13 μM; PTU:Med, 39 μM; PTU:High, 117.5 μM.

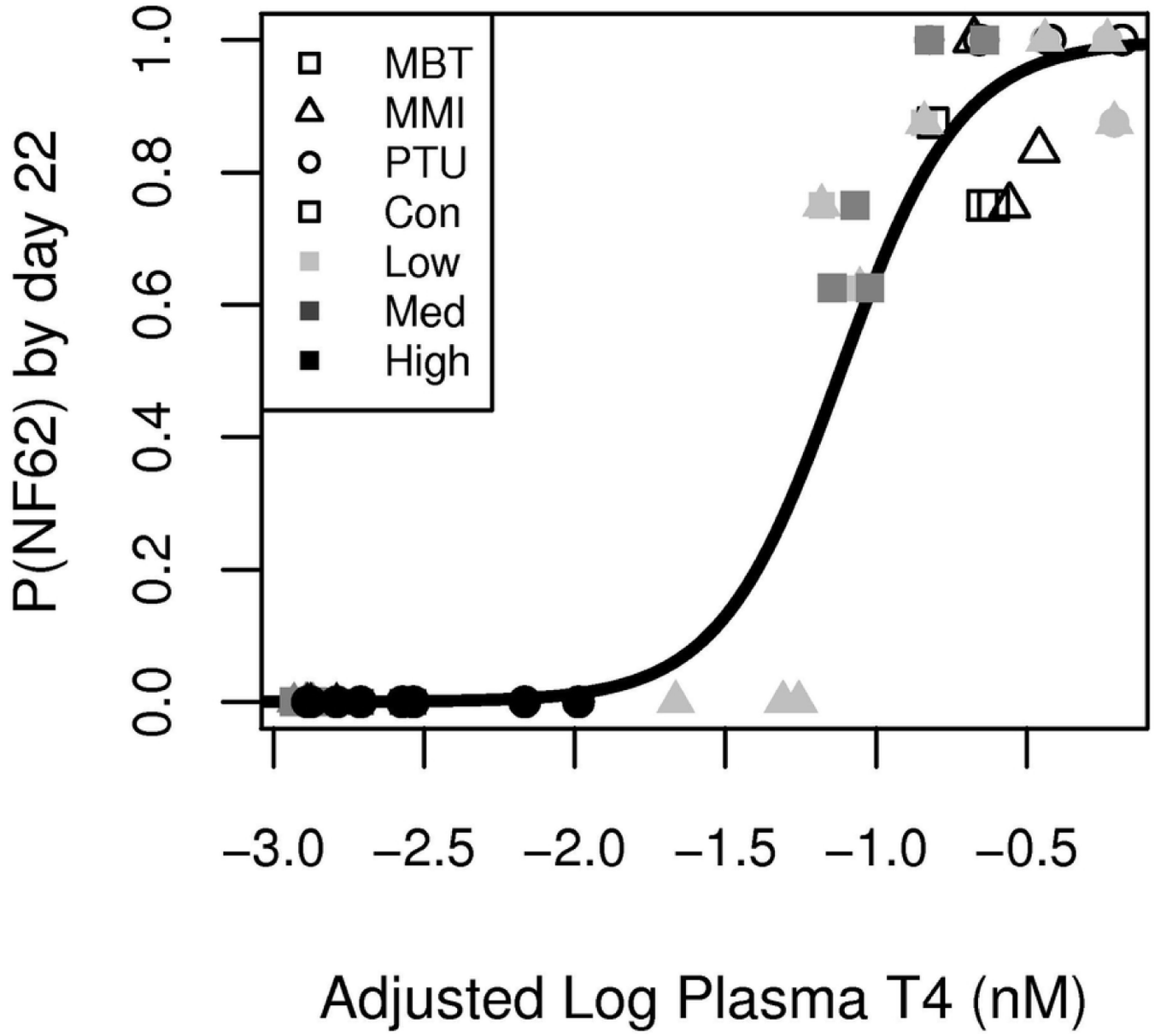

Figure 8.

Probability of reaching NF stage 62 (P(NF62)) by day 22 based on plasma thyroxine (T4) and sodium-iodide symporter (NIS) gene expression in thyroid tissue at 10 d of exposure post NF 53/54. To visually compare this model to models only using plasma T4, the predicated marginal influence of NIS gene expression in thyroid tissue on P(NF62) is converted into an equivalent amount of plasma T4 and added to plasma T4 for each sample in such a way to preserve the marginal relationship between plasma T4 and P(NF62). A test for goodness of fit fails to indicate a lack of fit for this model (p-value = 0.055). MMI, methimazole; MBT, mercaptobenzothiazole; PTU, propylthiouracil; MMI:Low, 55 μM; MMI:Med, 110 μM; MMI:High, 220 μM; MBT:Low, 0.18 μM; MBT:Med, 0.54 μM; MBT:High, 1.6 μM; PTU:Low, 13 μM; PTU:Med, 39 μM; PTU:High, 117.5 μM.

Table 1.

Akaike information criterion (AIC) scores for each logistic regression model.

Model	AIC
T4 plasma 7 d	167
T4 plasma 10 d	88
FCN 10 d	144
NIS 10 d	126
T4 plasma 10 d + NIS 10 d	74

Table 2.

Probability of reaching NF stage 62 (P(NF62)) given plasma T4 measured at 10 days.

Plasma T4 (nM)	P(NF62) by day 20	P(NF62) by day 22
T4 < 0.120	0–0.05	0–0.02
0.12 ≤ T4 < 0.170	0.05 – 0.07	0.02 – 0.05
0.17 ≤ T4 < 0.37	0.07 – 0.18	0.05 – 0.26
0.37 ≤ T4 < 0.50	0.18 – 0.25	0.26 – 0.41
0.50 ≤ T4 < 0.58	0.25 – 0.29	0.41 – 0.50
0.58 ≤ T4 < 0.91	0.29 – 0.42	0.50 – **0.75**
0.91 ≤ T4 < 1.15	0.42 – **0.50**	**0.75** – 0.84
1.15 ≤ T4 < 1.95	**0.50** – 0.66	0.84 – 0.95
1.95 ≤ T4 < 2.7	0.66 – 0.75	0.95 – 0.98
2.7 ≤ T4 < 10.9	0.75 – 0.95	0.98 – 0.999
T4 ≥ 10.9	≥ 0.95	≥ 0.999

Note: Based on the definitions used, control larvae are expected to have P(NF62) by day 20 equal to 0.5 and P(NF62) by day 22 equal to 0.75; these values are indicated by bold font in the table.