Amended Draft validation report for the draft new Test Guideline for the Rapid Estrogen ACTivity *In Vivo* assay (REACTIV)

For second WNT-review by 11 March 2024

Validation of the Rapid Estrogen ACTivity In Vivo (REACTIV) Assay for the Detection of Estrogen Axis Active Chemicals

Draft Report

January 2024

FOREWORD

This document describes the design and results of the validation exercise for the Rapid Estrogen ACTivity In Vivo (REACTIV) Assay. This method was developed for the detection of estrogen axis active substances. It is performed in 6-well plate format and can serve as a quick screen for potential estrogen axis disrupting substances. The purpose of the validation exercise was to determine whether the standard operating procedure (SOP) could be successfully transferred across laboratories, to determine variability between laboratories and to verify the absence of false positives by testing compounds presumed to be inert.

The REACTIV assay is being validated through an international effort via the OECD. The OECD has been working with member countries on the validation and harmonization of testing methods for the detection of chemicals that interfere with the estrogen, androgen and thyroid pathways.

Table of contents

FIGURES

TABLES

ABREVIATIONS AND DEFINITION

CEFAS: Centre for Environment, Fisheries and Aquaculture Science, UK**.**
DPH: Day post hatch.
EE2: 17α-ethinylestradiol.
Eleutheroembryo: The eleutheroembryonic life stage is post-hatch, but before the embryo is capable of independently feeding on exogenous food supplies and is a stage of on-going embryonic development. In some regulatory jurisdictions, the eleutheroembryonic period is regarded as a non-protected life stage in this context (OECD, 2014). Applying this definition to *O. latipes* positions this period of development from stage 39 (hatching stage) to stage 42 (formation of structures required for prey capture including the teeth of the upper jaw, the otolith, and the shape of all fins) (Iwamatsu, 2004).
ER: Estrogen receptor.
GFP: Green fluorescent protein.
LC50: Median lethal concentration is the concentration of a test chemical that is estimated to be lethal to 50% of the test organisms within the test duration.
LPL: Laboratoires des Pyrénées et des Landes.
LOEC: The lowest observed effect concentration is the lowest tested concentration at which the test chemical is observed to have a statistically significant effect.
MS222: tricaine methanesulfonate.
NOEC: The no observed effect concentration is the tested concentration immediately below the LOEC.
SEM: Standard error to the mean.
Runs: The repeat experiments performed for each chemical. Three runs are performed for each test chemical and each run utilises different independently prepared test solutions.
***chgh-gfp*:** Transgenic medaka line harbouring a genetic construction consisting of a 2047 base pairs of the Japanese medaka *choriogenin H* promoter upstream of GFP coding sequence.
Spiked mode: Part of the REACTIV assay performed in the presence of 30 µg/L of testosterone.
Unspiked mode: Part of the REACTIV assay performed in the absence of testosterone.
UVCB: Substances of unknown or variable composition, complex reaction products or biological materials.

ACKNOWLEDGEMENTS

This work is the collaborative effort of six laboratories which generously performed the experiments described here.

The following laboratories and their staff took part in the REACTIV assay interlaboratory validation exercise:

- Health and Environmental Risk Division, National Institute for Environmental Studies, Tsukuba, Japan. Dr Takako Yasuda, Dr Takahiro Yamagishi and Dr Haruna Watanabe performed the experiments. Prof. Hiroshi Yamamoto and Prof. Taisen Iguchi (Nanobioscience, Yokohama City University) coordinated the work in Japan.
- Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Weymouth, UK: Dr Marion Sebire, Ellen Blaker, Robert McFarling and Dr Ioanna Katsiadaki performed the experiments. Dr Marion Sebire and Dr Ioanna Katsiadaki supervised the work at CEFAS.
- DHI, Hørsholm Denmark: Gert Hansen, Mette Albrektsen and Jane Bergstrøm performed the experiments. Dr Gitte Ingelise Petersen supervised the work at DHI.
- Texas Christian University, Fort Worth, Texas, USA. Zach Aldrete, Dalton Allen, Evan Burchfiel, Rachael Carlson, Bridgette Fischer, Rashidat Jimoh, Katie Solomons and Catherine Wise performed the experiments. Prof. Marlo Jeffries supervised the work at the Texas Christian University.
- University of Heidelberg, Centre for Organismal Studies, Aquatic Ecology and Toxicology, Heidelberg, Germany. Winnie Henderson, Laura Behnstedt, Maximilian Kraft performed the experiments. Dr Lisa Baumann supervised the work.
- Laboratoire WatchFrog, Evry, France: Camille Zany, Amira Chikhaoui and Coralie Barrier performed the experiments. Dr Andrew Tindall supervised the experiments at WatchFrog and designed and coordinated the validation study.

Andrew Tindall (tindall@watchfrog.fr), wrote the validation report and the draft test guideline.

1. INTRODUCTION

1.1. Objectives of the Validation Study

The overall objective of the validation exercise for the REACTIV assay was to establish the relevance of the assay to detect any potential estrogen axis activity of compounds acting at different points within the estrogen axis and via different modes of action. A second aim was to assess the transferability and reproducibility of the assay by comparing results obtained by a variety of laboratories in six different countries and three different continents.

1.2. Assay Development/Background

Earlier versions of the protocol for the REACTIV assay have been performed at the lead laboratory for over ten years. A number of scientific publications detail some of this validation work and a number of other publications demonstrate the results obtained with the assay.
The transgenic model was created under contract to the lead laboratory (WatchFrog) by Amagen, France in 2008 using a transgene provided by WatchFrog and based on a publication by Kurauchi et al. (2005).
After the lead laboratory developed the assay protocol and characterised its response to a number of estrogenic and non-estrogenic chemicals, a scientific report was published detailing the key points of the assay characterisation (Spirhanzlova *et al.*, 2016). At this point the test was carried out for 48 h using medaka eleutheroembryos heterozygous for the *chgh-gfp* transgene.
Key points of this work were:
- The selection of a transgenic line exhibiting inducible GFP signal in the liver in response to estrogen axis activity, but also non-inducible, basal fluorescence in certain cells around the mouth and in cardiac muscle fibres. The presence of this ectopic, non-inducible signal in the heart allowed easily selection of transgenic fry prior to exposure as only half of the fry from the heterozygous x wild-type cross were transgenic. Similar cases of basal, non-inducible expression of *gfp* in *chgh-gfp* lines have been described by Kurauchi et al. (2008).
- Determination of the sensitivity of the assay to a model estrogen (17α-ethinylestradiol, EE2). The lowest observable effect concentration (LOEC) was 34 ng/L EE2 following a 24 h exposure. As previously published, the sensitivity is increased to 15 ng/L EE2 with an EC50 of 71.9 ng/L following a 48 h exposure (
Figure 1).

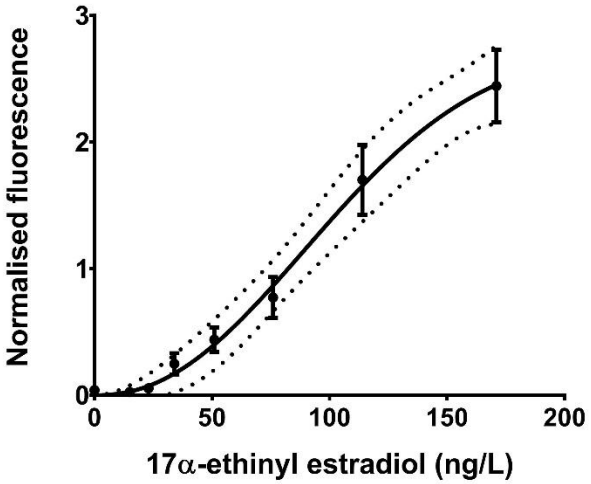


Figure 1: Concentration-response curve for ethinylestradiol.

Reprinted with permission (Spirhanzlova *et al.*, 2016).

Demonstration that eleutheroembryos can metabolise testosterone into estradiol via the enzyme P450 aromatase *in vivo* at this early developmental life stage.

The ability of the transgenic line to identify anti-estrogenic chemicals was also demonstrated using the model selective estrogen response modulator (SERM) tamoxifen.

In order to evaluate the specificity of the assay, six chemicals were tested that were expected to be inactive on the estrogen axis and steroidogenesis. None of the groups of fry exposed to these chemicals showed a statistically significant difference in fluorescence compared to the relevant control group in either unspiked or spiked mode when tested at 100 mg/L.

In addition, in order to further characterise the assay specificity, reference hormones for endocrine pathways that may give cross-talk with the estrogen axis were also tested at high concentrations (10^{-7}M). No statistically significant difference was observed for aldosterone (mineralocorticoid axis) or progesterone (progestin axis) in unspiked or spiked mode.

The most active thyroid hormone, T3, did, however, induce a slight inhibition in estrogen axis signalling in spiked mode at this very high concentration.

Due to a number of publications (Hayashi *et al.*, 2010; Kitano *et al.*, 2012; Sato *et al.*, 2005) indicating crosstalk between the estrogen and glucocorticoid axes via inhibition of aromatase expression by corticosteroids, it was expected that an inhibition of estrogen axis activity may be observed when testing corticosteroids in spiked mode. Therefore, a concentration range of dexamethasone was tested. As shown in Figure 2, dexamethasone induced a slight inhibition in estrogen axis signalling at the high concentrations of 392 µg/L and 3.92 mg/L. This effect would also be expected in other test guidelines sensitive to estrogen axis disruption.

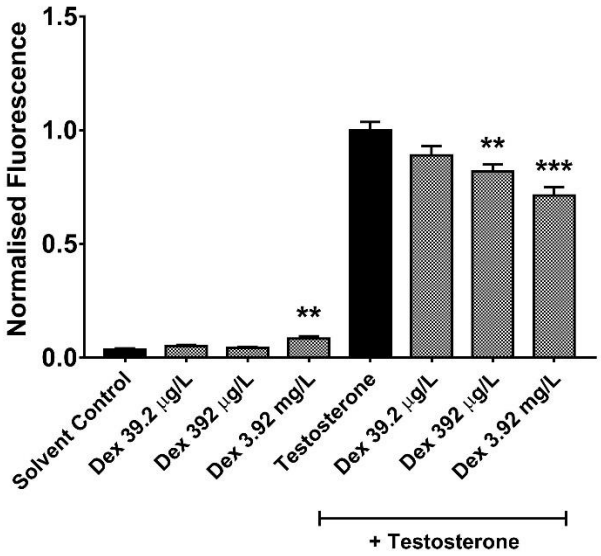


Figure 2: Concentration-response curve for dexamethasone.

Reprinted with permission (Spirhanzlova *et al.*, 2016).

Together, the results of these studies indicate that the assay is sensitive, specific and robust. They also indicate that the assay is capable of detecting estrogen receptor agonists, aromatisable androgens, aromatase activity modulators and SERMs. Unpublished internal studies have also shown that the assay is capable of detecting chemicals inducing or inhibiting aromatase at the gene expression level as either pro- or anti-estrogenic chemicals, respectively.

Currently, unpublished results from an internal study have also shown that the *chgh-gfp* line can detect changes in estrogen axis activity caused by 5α-reductase inhibiting chemicals. By inhibiting the conversion of testosterone to 5α-dihydrotestosterone (DHT) via the enzyme 5α-reductase, these chemicals increase the available pool of testosterone that can be converted in estradiol, therefore, increasing estrogen axis activity. Figure 3 below shows this exact trend when the type I, II and III 5α-reductase inhibitor dutasteride is tested in spiked mode. No fluorescence was observed for any concentration of dutasteride in unspiked mode.

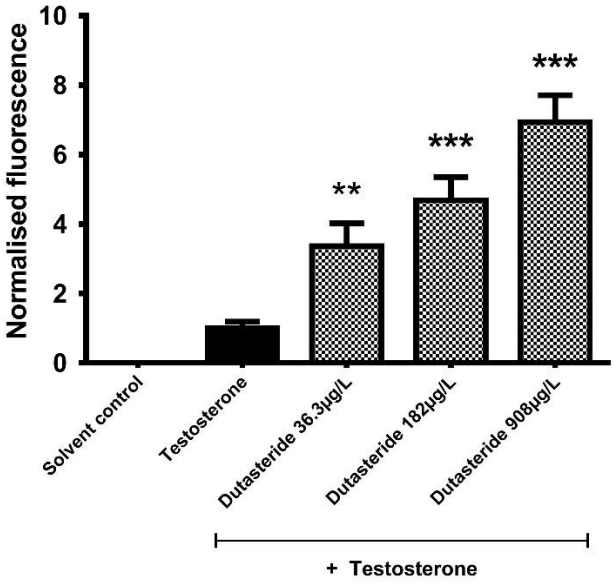


Figure 3: Concentration-response curve for the pharmaceutical 5α-reductase inhibitor dutasteride.

Recent improvements to the assay include the development of the *chgh-gfp* line into a homozygous line, meaning that eleutheroembryos do not need to be sorted to remove non-transgenic embryos prior to an exposure study. All eleutheroembryos can now be used from a homozygous x homozygous cross.

The first step of confirming transferability of the assay was performed by all laboratories taking part in the validation exercise. A series of experiments was carried out with the reference control EE2 to verify that the assay is performing correctly in all laboratories. Upon completion of this first step the second step of determining the reproducibility of the assay was initiated. This involved all partner laboratories testing a series of reference active chemicals acting via a range of mechanisms of action to determine the scope of the assay and the rate of false negatives; as well as a series of chemicals that are expected to be inert to determine the rate of false positives.
Key scientific publications involving the *chgh-gfp* line are listed at the end of the reference section below.
The inter-laboratory validation study is expected to confirm transferability of the assay, determine variability between laboratories and verify the absence of false negatives and positives by testing compounds presumed to be active and inert, respectively.

The following information supports the transferability of this protocol to the participating laboratories.

1) Medaka fish are already a widely used model organism across OECD countries. They are also widely accepted and validated as a test species in numerous OECD test guidelines including: OECD TG 203 (Fish Acute Toxicity Test; OECD,

2019a), OECD TG 210 (Fish Early Life Stage Toxicity Test; OECD, 1992), OECD TG 212 (Fish Short Term Toxicity Test on Embryo and Sac-fry Stages; OECD, 1998), OECD TG 229 (Fish Short Term Reproduction Assay; OECD, 2012), OECD TG 230 (21-day Fish Assay; OECD, 2009), OECD TG 234 (Fish Sexual Development Test; OECD, 2011) and OECD TG 240 (Medaka Extended One Generation Reproduction Test; OECD, 2015).

2) An additional advantage of medaka is that they are reared in conditions that are almost identical to those of zebrafish. Laboratories with previous zebrafish husbandry experience were able to successfully rear and reproduce medaka (Texas Christian University, USA).
3) Founders, adult homozygous medaka for breeding embryos for on-site testing, were made available to participants in the ring test.
4) Embryos were available for shipping from a breeding/production site to another testing site.

1.3. Test organism

The test species selected for the REACTIV assay is the medaka (*Oryzias latipes*). This species of fish is a well-established small model organism, having been extensively studied since the beginning of the twentieth century when Aida (Aida, 1921) linked sex to body colouration in certain strains of medaka. Medaka is an ideal model for studying the vertebrate sex steroid axes. Sexual differentiation has been extensively studied (Kondo *et al.*, 2009) and medaka estrogen receptors (ERs) and AR show conformational conservation for endocrine disrupting chemicals when compared to human receptors (Cui *et al.*, 2009). In addition, steroidogenesis pathways are highly conserved among vertebrates, with a high concordance in the identification of endocrine active chemicals between fish and rat assays carried out in the context of the U.S. Environmental Protection Agency Endocrine Disruptor Screening Program (Ankley and Gray, 2013). Medaka were also the first vertebrate species after humans in which the master sex determining gene (*dmy*) was identified (Masuyama *et al.*, 2012; Matsuda *et al.*, 2002). This fact besides its clear importance in itself also highlights the extent to which the genetic basis of sex determination has been studied in medaka and allows definitive determination of the genetic sex of medaka.

As with mammals, medaka possess a XX/XY sex determination system (Aida, 1921; Yamamoto, 1958, 1955). It is also possible to determine the phenotypic sex of medaka morphologically due to a dimorphism in their dorsal and anal fins.

Due in part to these characteristics which allow clear confirmation of sex reversal due to the action of EDCs, the effects of exogenous estrogens and anti-estrogens have been extensively studied in medaka. It has been well demonstrated that exposure to estrogens during development can cause genetically male (XY) medaka to develop a female phenotype (Dang and Kienzler, 2019; Iwamatsu *et al.*, 2005; Knörr and Braunbeck, 2002; Kobayashi and Iwamatsu, 2005; Lei *et al.*, 2013; Paul-Prasanth *et al.*, 2013; Scholz and Gutzeit, 2000; Spirhanzlova *et al.*, 2020). There is also a clear causal link between exposure to exogenous estrogens and increased expression of vitellogenins and choriogenins (Ishibashi *et al.*, 2016; Lee Pow *et al.*, 2016; Scholz *et al.*, 2005).

A number of transgenic estrogen axis reporter models have been developed in medaka utilising either a *vitellogenin* or *choriogenin* gene promoter to drive expression of *gfp* (Kurauchi *et al.*, 2005, 2008; Salam *et al.*, 2008; Spirhanzlova *et al.*, 2016; Ueno *et al.*, 2004; Zeng *et al.*, 2005). Of these models those utilising the *choriogenin H* promoter are the most sensitive and respond the most rapidly.

The welfare of the eleutheroembryos in this assay was of major concern. Medaka were raised according to established husbandry protocols (Kinoshita *et al.*, 2009). The REACTIV assay we describe is performed entirely using life stages of medaka which do not fall under the scope of the European Union Directive 2010/63/EU on the protection of animals used for scientific purposes. Currently no screening assays have been validated for estrogen axis disruption using eleutheroembryonic life stages. Details on the advantages of the use of this *in vivo* eleutheroembryonic assay compared to *in vitro* assays are given in section 2.1. It is expected that the validation and subsequent use of the REACTIV assay as a frontline *in vivo* screening tool will reduce the number of tests performed using regulated life stages of fish. The reduction in tests might concern tests capable of capturing estrogen axis disruption such as the Fish Short Term Reproduction Assay (TG229), the 21-day Fish Assay (TG230) and the Fish Sexual Development Test (TG234). In addition, the transgenic line is homozygous for the transgene, this has the advantage that all eleutheroembryos produced are capable of being used in the REACTIV assay. Therefore, no non-transgenic eleutheroembryos are produced which would could not be used and would require euthanising.

1.4. Genetic construct

The c*hgh-gfp* transgenic line used in the REACTIV assay harbours 2.047 kb of the medaka *choriogenin H* gene promoter immediately upstream of the start codon driving expression of Green Fluorescent Protein (GFP) coding sequence (Figure 4).

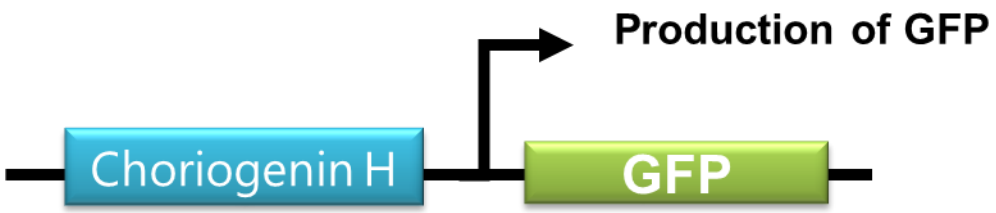


Figure 4 : The transgene present in the *chgh-gfp* medaka line used in the REACTIV assay.

The *chgh-gfp* transgene is expressed in the liver of the medaka in response to activation of estrogen axis signalling. There is also a non-inducible ectopic expression of GFP in some cells of the heart and head at eleutheroembryonic life stages. This allows visual confirmation that the developing fry are transgenic.

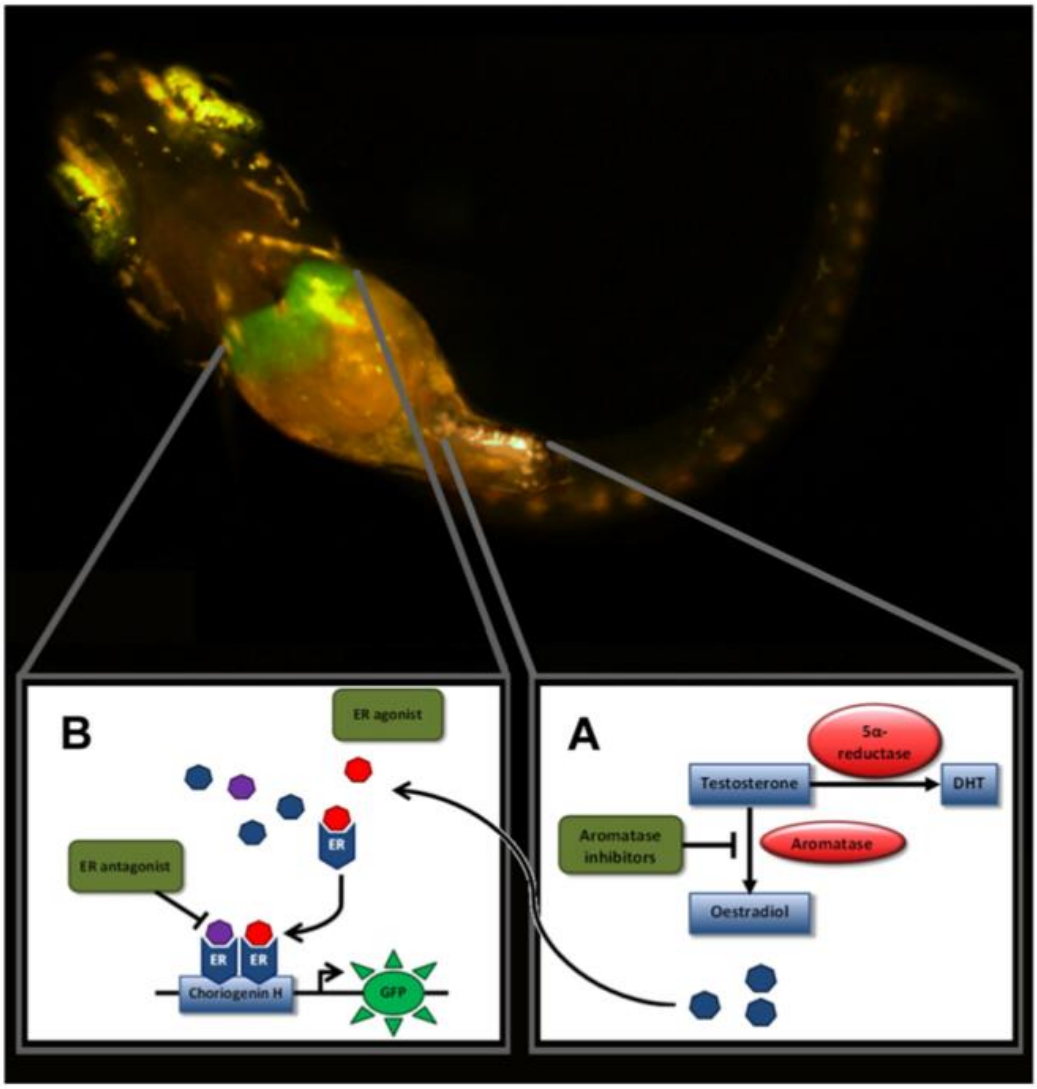


Figure 5 : Schematic diagram of the induction of GFP via aromatase in choriogenin H-GFP medaka.

A) Testosterone in the exposure medium is converted to estradiol via aromatase and DHT via 5α-reductase principally in the gonads. B) estradiol in turn binds to ERs. Ligand-receptor complexes then bind to estrogen response elements (EREs) within

the choriogenin H promoter driving expression of GFP in the liver. Reprinted with permission (Spirhanzlova *et al.*, 2016).

The promoter region present in the transgene has been shown to contains putative estrogen response elements (ERE) and the expression of the transgene has been demonstrated to be significantly modulated in the presence of ER agonists, antagonists and compounds inducing or inhibiting steroidogenic enzymes (Kurauchi *et al.*, 2005, 2008; Spirhanzlova *et al.*, 2016).

Choriogenin genes, much like vitellogenin, are required for egg production in fish. Their expression is upregulated in response to estrogen axis signalling. As a terminal step, their expression and the expression of GFP in the *chgh-gfp* medaka line represents the overall or net effects of both endogenous and exogenous factors altering estrogen axis signalling (alterations in production, transport, metabolism and excretion of hormones as well as activation and inhibition of ER; Figure 6). All laboratories taking part in the validation exercise used the *chgh-gfp* medaka line provided by the lead laboratory. This homozygous line is now beyond the F15 generation and has consistently shown a stable level of GFP expression in response to challenge with an estrogen.

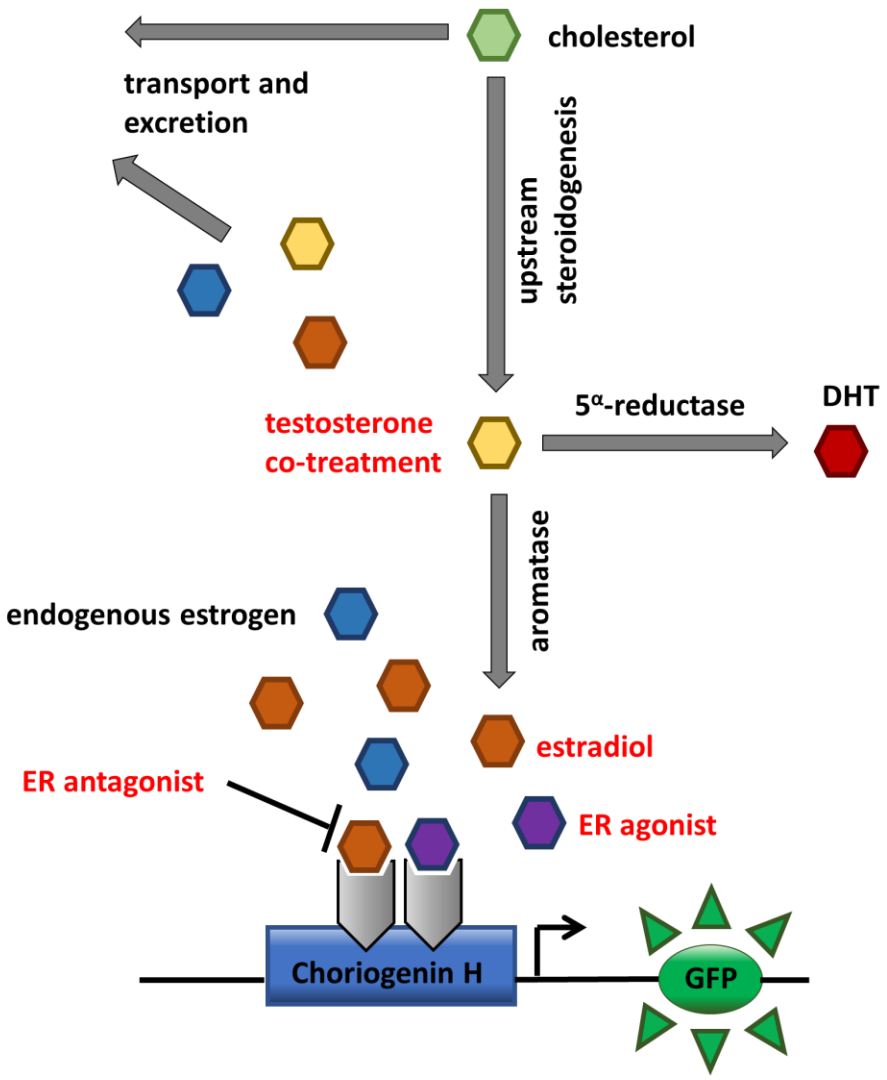


Figure 6 : Schematic representation of the factors that can influence the magnitude of the fluorescence response of *chgh-gfp* eleutheroembryos.

Grey arrows indicate processes that can be either inhibited or upregulated and alter the quantity of estrogens present to interact with estrogen receptors. Black text indicates endogenous factors and red text indicates exogenous factors.

2. PURPOSE AND OBJECTIVES

2.1. Purpose of the assay

The impact of endocrine disruptors on the health of humans and wildlife is now undeniable. REACH and equivalent international legislation require the testing of chemicals for endocrine activity. In 1998, the OECD initiated a program of work to develop new or update existing test guidelines for the screening and testing of endocrine disruptors.

At present, there is one eleutheroembryonic fish assay which has been adopted as a test guideline for the detection of estrogen axis disruption, the EASZY assay (TG 250). However, the EASZY assay is limited to detecting ER agonists which can pass the chorion and the blood-brain barrier as noted in the test guideline.

Two *in vitro* assays have been validated as test guidelines for the detection of estrogen axis disruptors (TG 493, TG 455), however, these assays can only detect interactions between the test chemical and ERs.

The exception among existing *in vitro* test guidelines is the H295R assay (TG 456). This *in vitro* assay is carried out in the H295R cell line which expresses genes coding for all key steroidogenic enzymes. It is, therefore, possible to identify disruption of steroidogenesis by quantifying testosterone and estradiol at the end of the assay and comparing to controls. However, the H295R assay does present a number of limitations as described in the test guideline, notably:

- False negatives are expected for chemicals requiring metabolic activation as the metabolic capacity of the cell line is unknown and is likely to be limited;
- Disruption of the key steroidogenic enzyme 5α-reductase is not expected to be detected by the assay as its metabolite DHT is not measured. Although, it should be noted that a proposal has been submitted to the OECD regarding the enhancement of TG 456, including DHT measurement, and accepted on the OECD workplan in 2022 (Project 4.159);
- The *in vitro* nature of this test means that chemicals disrupting the hypothalamic-pituitary-gonadal (HPG) axis will not be detected as this can only be studied in intact animals.

Taken together, this indicates that currently no test guidelines allow the detection of any of the following mechanisms of action without the use of laboratory animals covered by Directive 2010/63/EU:

- Disruption of the HPG axis;
- Disruption of the key steroidogenic enzyme 5α-reductase;
- Any form of estrogen axis disruption requiring metabolic activation.

The REACTIV assay can fill this gap at life stages not falling under the scope of Directive 2010/63/EU by detecting:

- ER agonists;
- ER antagonists;
- Chemicals altering the activity and/or expression of aromatase;
- Disruption of the key steroidogenic enzyme 5α-reductase;
- Chemicals requiring metabolic activation in order to disrupt the estrogen axis.

It should also be noted that the proposed REACTIV assay is based on the use of entire organisms. Metabolic capacity is not the only advantage of the use of whole organisms. These models also take into account other factors which are difficult or impossible to model *in vitro*, for example effects on hormonal transport and crosstalk between the different endocrine axes in a natural physiological context.
Validation of the REACTIV assay as a test guideline would, however, allow a level 3 test to be carried out in entire organisms not considered as laboratory animals covered by Directive 2010/63/EU. This would provide mechanistic information specific to the activity of the test chemical on the estrogen axis, with a positive result indicating a change in estrogen receptor binding and transactivation of target genes specific to the estrogen axis. The REACTIV assay would also be upstream of assays involving adult animals, this means that in the worst-case scenario, it could guide the selection of test concentrations for adult/juvenile animal models as well as the selection of the most adapted higher tier test to perform.

2.2. Major characteristics of the assay

The REACTIV assay involves aqueous exposures of eleutheroembryonic medaka in a multi-well format to detect modulation of estrogen axis signalling by potential estrogen axis active chemicals. The assay is transcriptional-based and uses a transgenic medaka line containing the *chgh-gfp* genetic construct (see Figure 5 above) to detect the activity of estrogenic chemicals such as ER agonists and antagonists as well as aromatisable androgens and chemicals altering the expression or enzymatic activity of key enzymes such as aromatase and 5α-reductase. The assay measures the ability of a chemical to activate or inhibit transcription of the genetic construct, whether directly through binding to ERs or by modifying the metabolism of testosterone. The endpoint measured is the fluorescence of the transgenic eleutheroembryos. When transcription of the genomic construct is activated or inhibited following chemical exposure, the eleutheroembryos express more or less GFP and, therefore, emit more or less fluorescence compared to unexposed eleutheroembryos.

The assay measures GFP fluorescence in the transgenic *chgh-gfp* eleuthero-embryos by fluorescence imaging using a fluorescence microscope. An automated image analysis macro is then used to remove fluorescence generated by endogenous pigments in the medaka eleutheroembryos (melanophores, iridiphores, xanthophores, leucophores) (Braasch *et al.*, 2009; Loire *et al.*, 2013; Wakamatsu *et al.*, 2001). The automated macro produces an Excel sheet containing a numerical value of the GFP signal in each eleutheroembryo.

Control eleutheroembryos are maintained in test medium (see section 3.2.1). If a pro-estrogenic chemical is present in the exposure media of the test groups, an increase in fluorescence signal is expected.

A second set of exposures are carried out in the presence of testosterone, with the groups exposed to media containing the test chemical spiked with testosterone being compared to a testosterone alone control group (spiked mode). The aim of these exposures is to activate the estrogen axis through conversion of testosterone to estradiol by aromatase enzyme.

595 This allows the identification of anti-estrogenic chemicals acting either through ER
596 antagonism or other mechanisms of alteration of the ability of the ER to bind estradiol such
597 as modulation of aromatase expression or activity or conversion of testosterone to non-
598 aromatisable DHT or via downregulation of ER expression. In addition, pro-estrogenic
599 chemicals can be identified acting through direct ER agonism, modulation of the ER to
600 increase its affinity for estradiol, upregulation of ER expression, modulation of aromatase
601 activity or inhibition of 5α-reductase.
602
603 Results can be evaluated in terms of the lowest observable effect concentration (LOEC).
604 They can also be evaluated as EE2 equivalence by comparing the induction or inhibition
605 of fluorescence to the EE2 standard curve included in each experiment.
606
607 A brief overview of the schedule of the REACTIV assay is given below.
608

609

Overview of schedule:

Day -1: Collect newly hatched eleutheroembryos. Eleutheroembryos hatch at day post fertilisation 10 under our conditions (26°C).

Day 0:

- Prepare exposure solutions including controls.
- Add 8 day post hatch 0 eleutheroembryos to each well.
- Remove the maximum amount of liquid without drying the eleutheroembryos (maximum remaining volume 800 µL).
- Fill each well of 6-well plates with 8 mL of each exposure solution (1 well per exposure condition) under a chemical hood.
- Incubate the plates at 26 °C in a 14:10 light:dark cycle.
    - ⇨ Do not feed the embryos during the experiment.

Day 1 (+ or- 30 min):

- Note mortality and dispose of any dead eleutheroembryos.
- Rinse the eleutheroembryos by transfering them to new annotated 6-well plates with 8 mL of dechlorinated water or mineral water per well.
- Anesthetise the eleutheroembryos with MS222 at 200 mg/L by adding 2 mL/well of MS222 1 g/L if they require positioning for reading.
- Capture a colour image of the ventral side of each eleutheroembryo including the liver in the image.
- Euthanise eleutheroembryos in MS222 at 1 g/L.

610

2.3. General experimental design

The assay is performed to determine the potential of a test chemical to modulate the estrogen axis under sublethal concentrations. For the validation process a five-concentration test design was used (8 fry per well x 1 well = 8 fry exposed per concentration). In the test guideline a minimum of five concentrations is recommended to maintain the same sensitivity. Newly hatched (day post hatch zero; DPH0) *chgh-gfp* medaka fry are used for the REACTIV assay. The test is terminated at DPH1 after 24 h of exposure. They are not fed before or during the test as the test is terminated at stage 40 (Iwamatsu, 2004). Yolk is still present until stage 41/42 and is used as the source of energy for the development of the eleutheroembryo.

The test is run in two modes "spiked" and "unspiked" i.e., with and without the addition of testosterone. In spiked mode all groups are spiked with 30 µg/L of testosterone. Eleuthero-embryonic life stages of medaka do not synthesise enough estrogens from androgens to generate a GFP signal. Therefore, spiking with testosterone is necessary in order to detect chemicals acting on estrogen distribution, metabolization, degradation and ER antagonists.

The control groups include:

a. Test medium and/or solvent control: 1 well with 8 organisms/well is exposed to test medium plus 0.2% DMSO. This control defines the basal fluorescence level in the test medium. If a solvent is used, then this group is exposed to test medium plus the solvent used at the same concentration as all other groups. In some cases, such as a solvent being used with no historical data available, both groups may be required.
b. EE2 488 ng/L: 1 well with 8 organisms/well is exposed to 488 ng/L of EE2. This control establishes a close to maximal fluorescence observable for most mechanisms of action. It is also equivalent to the lowest concentration of EE2 inducing a statistically significant reduction in fecundity in a published 21-day medaka assay (Seki *et al.*, 2002).
c. Testosterone 30 µg/L: Two wells with 8 organisms/well are exposed to 30 µg/L of testosterone. This control serves to induce estrogen axis signalling via endogenous conversion of testosterone to estradiol. Induction of estrogen signalling in "testosterone spiked mode" allows inhibition of estrogen axis signalling through ER antagonism, aromatase inhibition or repression of aromatase expression to be detected. It also allows induction of estrogen axis signalling through mechanisms such as increased aromatase expression or inhibition of 5α-reductase to be detected. Data from two wells are pooled for this control to increase confidence in the mean fluorescence value.
d. Induction control for spiked groups: 1 well with 8 organisms/well is exposed to 64 ng/L of EE2 plus 30 µg/L of testosterone. This control group confirms that an induction of fluorescence can be observed above that of the testosterone 30 µg/L control group. Under 21-day flow-through conditions (OECD, 2009) in medaka, 64 ng/L of EE2 is the lowest

concentration shown to have a physiological effect, consisting of testis-ova in one third of male fish (Seki *et al.*, 2002).

e. Inhibition control for spiked groups: 1 well with 8 organisms/well is exposed to 10 µg/L of fadrozole plus 30 µg/L of testosterone. At 10 µg/L, fadrozole induces a modification in the gonadosomatic ratio of male fish within an OECD testing protocol (OECD 229) (Ankley *et al.*, 2002).

The following additional control groups are optional, but are recommended for calibration of reading parameters in naïve laboratories as well as for quality control purposes. They were performed in all validation exercise experiments. They constitute an EE2 standard curve and as well as quality control purposes they can be used to derive a concentration-response relationship for EE2 allowing the results to be expressed in EE2 equivalents. The calculation of equivalence values is not required and is for informative purposes only as the result of the assay is that the test chemical is active or inactive only. If equivalence values are to be calculated, the optional controls below should be included in each run.

f. EE2 34 ng/L: 1 well with 8 organisms/well is exposed to 34 ng/L of EE2. This control serves as part of the EE2 standard curve, allowing an EE2 equivalence value to be read off the standard curve for any active test chemicals.

g. EE2 51 ng/L: 1 well with 8 organisms/well is exposed to 51 ng/L of EE2. This control serves as part of the EE2 standard curve, allowing an EE2 equivalence value to be read off the standard curve for any active test chemicals.

h. EE2 76 ng/L: 1 well with 8 organisms/well is exposed to 76 ng/L of EE2. This control serves as part of the EE2 standard curve, allowing an EE2 equivalence value to be read off the standard curve for any active test chemicals.

i. EE2 114 ng/L: 1 well with 8 organisms/well is exposed to 114 ng/L of EE2. This control serves as part of the EE2 standard curve, allowing an EE2 equivalence value to be read off the standard curve for any active test chemicals.

j. EE2 171 ng/L: 1 well with 8 organisms/well is exposed to 171 ng/L of EE2. This control serves as part of the EE2 standard curve, allowing an EE2 equivalence value to be read off the standard curve for any active test chemicals.

After 24 hours of exposure, the eleutheroembryos are imaged with a colour camera and GFP long pass filters. An image of the ventral region including the liver of each organism is captured. Image analysis software is then used to quantify the GFP signal to allow estrogen axis activity to be compared between different controls and exposure groups.

2.4. Replication

One test is composed of three independent and valid runs using 8 organisms/treatment group/run (Figure 7). Each run should be performed using independent solutions. The runs should be ideally conducted sequentially but could be conducted in parallel. The raw data for a given test chemical is obtained by pooling the data from the three runs to obtain n=24 fluorescence values in each treatment group.

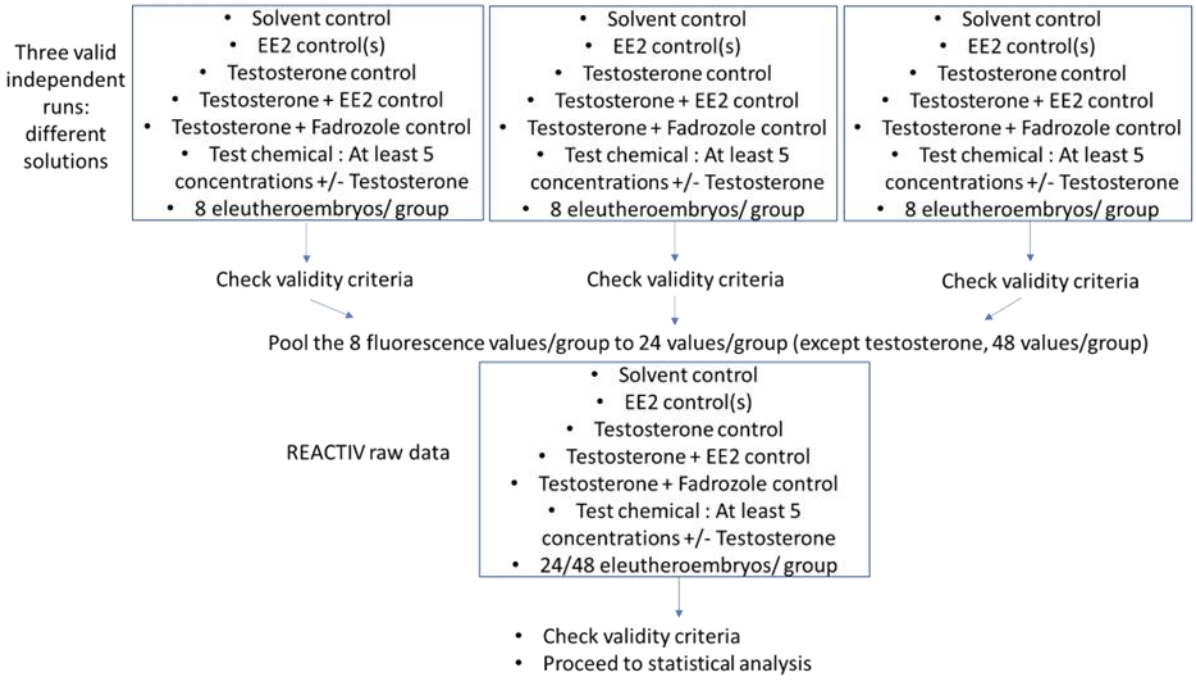


Figure 7: Overview of the REACTIV assay. "+/- Testosterone" refers to spiked and unspiked groups. A REACTIV assay is composed of three independent runs and utilises 360 eleutheroembryos in total. All non-optional controls should be performed in each run. If a solvent is being used for the first time or for the first time at a certain concentration, a test medium control should also be included.

3. Validation study

3.1. Specific goals

1) Each laboratory performed two calibration experiments. The first was with EE2 only to define the correct imaging parameters to prevent saturation of GFP signal. The second calibration experiment was with all assay controls in order to adjust image capture settings to obtain the optimal sensitivity in fluorescence readings.
2) An interlaboratory validation study was then carried out with the six participating laboratories and inter-laboratory variability was determined. A concentration range of EE2 was included in every run. Ten additional estrogen axis active chemicals were tested in between two and six laboratories. The chemicals were chosen to cover a range of modes of action on the estrogen axis. Six expected inert chemicals were also tested in two to six laboratories, to obtain sufficient information on the reliability, reproducibility, and sensitivity of the assay. An additional estrogen axis active chemical, atrazine, was tested in the lead laboratory. An EE2 concentration range served as positive controls in all experiments for the unspiked part of the test. Testosterone plus EE2 and testosterone plus fadrozole served as positive controls in all experiments for the testosterone spiked exposures. All test chemicals were tested in the presence and absence of 30 µg/L of testosterone.
3) Performance was compared between the participating laboratories. Reliability and reproducibility across laboratories, and sensitivity of the assay were determined.

3.2. Overview of Test Conditions

The lead laboratory, WatchFrog, provided two of the participating laboratories with homozygous *chgh-gfp* medaka eggs in advance of the validation study to allow them to begin breeding colonies. These two laboratories were the Japanese and American partners. All experiments carried out by these partners were performed using eleutheroembryos bred in their laboratory.

For the interlaboratory validation, all participating laboratories were asked to test seven estrogen axis active chemicals and three expected inert chemicals. Fadrozole was also included at a single concentration as a control in all experiments, although it was also tested at multiple concentrations by two partners (France and UK). The seven active chemicals were selected to cover a range of modes of action expected to increase or decrease estrogen axis activity. All of these chemicals were tested at five concentrations in the presence and absence of 30 µg/L testosterone. In addition, six concentrations of EE2 were evaluated in each experimental run. These are shown in Table 1 below.

	Test chemicals	Mode of action
Pro-estrogenic	EE2	strong ER agonist
	Testosterone	aromatisable androgen
	BPA	weak ER agonist
	Dutasteride	5α-reductase inhibitor
	Estrone	natural ER agonist
	17β-Estradiol	strong natural ER agonist
	Triphenyl phosphate	Multiple (see section 3.5.10)
Anti-estrogenic	Prochloraz	aromatase transcription inhibitor
	Anastrozole	aromatase enzyme inhibitor
	Tamoxifen	SERM
	Fadrozole	aromatase enzyme inhibitor
Presumed inert	Amantadine	antiviral, antiparkinsonian
	Arabinose	monosaccharide
	Atenolol	beta blocker
	Cromolyn	mast cell stabilizer
	Cefuroxime	cephalosporin antibiotic
	Saccharin	artificial sweetener

Table 1: Pro-estrogenic, anti-estrogenic and presumed inert chemicals tested by multiple laboratories within the interlaboratory exercise.

Blue indicates control included in all runs of the REACTIV assay.

An additional estrogen axis active chemical was tested in the lead laboratory (Table 2).

	Test chemicals	Mode of action
Pro-estrogenic	Atrazine	Aromatase expression inducer

Table 2: Pro-estrogenic chemical tested uniquely by the lead laboratory.

Preliminary experiments were performed in the lead laboratory using the proposed substances to determine appropriate testing concentrations. In concordance with validated OECD test guidelines such as the XETA assay (TG248; OECD, 2019b) the maximum test concentration was set to the lowest concentration among the solubility limit, the maximum tolerated concentration or 100 mg/L. The maximum tolerated concentration was defined as the highest concentration resulting in less than 15% combined mortality and sublethal effects such as malformation or immobility. As there were only eight eleutheroembryos per condition per run, the maximum tolerated concentration was set to less than 15%

mortality or sublethal effects rather than 10% as 10% would exclude a test group if a single eleutheroembryo suffered mortality or sublethal effects.

For the synthetic or natural hormones (17α-ethinylestradiol, EE2; 17β-estradiol; estrone and testosterone) lower concentrations were tested as following the above guidance would have led to all test concentrations inducing the maximal fluorescence observable, saturating the system. Lower concentrations were also tested for the same rationale for the pharmaceutical aromatase inhibitors (anastrozole and fadrozole).

The five concentrations tested for each chemical are shown below in Table 3. All participating laboratories assayed each test chemical from the same batch and lot number, in the presence and absence of 30 µg/L testosterone.

The testing began with a calibration experiment. The goal of the calibration steps was to ensure that all laboratories attain a similar amplitude of response and sensitivity to the reference chemical EE2 despite differences in imaging equipment used to read the experiment. The calibration required two steps:

1) Determining the optimal imaging settings to allow a satisfactory amplitude of GFP induction to be obtained with a concentration of 488 ng/L of EE2.
2) Applying these settings for the quantitation of three runs of a concentration-response experiment with six concentrations of EE2 as well as the other assay controls (testosterone, testosterone + EE2 and testosterone + fadrozole) to check the amplitude of induction and sensitivity with increasing concentrations of testosterone and to ensure that the other assay controls elicit a detectable GFP response.

Once a laboratory demonstrated its ability to run the calibration experiments with an expected dynamic concentration response for EE2 and the spike mode controls, it then obtained the agreement of the lead laboratory to begin to test the chemicals of interest.

The conditions for the interlaboratory validation exercise are summarized in Table 3 below:

810
811

Test species		Japanese medaka (*Oryzias latipes*) eleutheroembryos
Exposure period		Day post hatch (DPH) 0 for 24 h
Criteria for selecting test individuals		Primary criterion was developmental stage and health of animal (alive and no-malformations)
Solvent control		Test medium 0.2% DMSO
Ethinylestradiol (EE2) standard curve		34, 51, 76, 114, 171, 488 ng/L
Testosterone		30 µg/L
Testosterone + EE2		30 µg/L + 64 ng/L, respectively
Testosterone + Fadrozole		30 µg/L +10 µg/L, respectively
Concentration of test chemicals	Anastrozole	2.9, 1.45, 0.73, 0.36, 0.18 mg/L 20, 4, 0.8, 0.16, 0.032 µg/L
	Atrazine	10, 5, 2.5, 1.25, 0.625 mg/L
	BPA	5, 4, 3, 2, 1 mg/L
	Dutasteride	1, 0.5, 0.25, 0.1, 0.01 mg/L
	17β-Estradiol	540, 270, 135, 68, 34 ng/L
	Estrone	25, 5, 1, 0.2, 0.04 µg/L
	Prochloraz	1, 0.5, 0.25, 0.13, 0.063 mg/L
	Fadrozole	87, 44, 22, 11, 5.4 µg/L 10, 2, 0.4, 0.08, 0.016 µg/L
	Tamoxifen	483, 242, 121, 60.4, 30.2 µg/L
	Testosterone	300, 100, 33, 11, 3.7 µg/L
	Triphenyl phosphate	0.89, 0.59, 0.4, 0.26, 0.18 mg/L
	Cromolyn	1000, 100, 10, 1, 0.1 µg/L
	Cefuroxime	10, 1, 0.1, 0.01, 0.001 mg/L
	Saccharin	100, 10, 1, 0.1, 0.01 mg/L
	Amantadine	10, 1, 0.1, 0.01, 0.001 mg/L
	Atenolol	100, 10, 1, 0.1, 0.01 mg/L
	Arabinose	100, 10, 1, 0.1, 0.01 mg/L
Exposure regime		24 hours. No renewals. No feeding.
Endpoints		Total fluorescence of the liver of eleutheroembryos imaged ventrally
Eleutheroembryos per test condition		Eight eleutheroembryos per well (six-well plate) with one well per test condition
Volume of test medium		8 mL per well
Test medium		See section 3.2.1
Replication		1 well per test condition
Incubation conditions during exposure		26°C +/- 1°C, 14:10 light:dark cycle
Measurement time		24 h
Number of Experimental Runs		Experiments were run 3 times for each chemical. Each experiment was

	performed with different preparations of the test chemical.

812

813 **Table 3: Conditions of the REACTIV assay. Two concentration ranges are given for**
814 **anastrozole and fadrozole as the initial concentration range caused a maximal**
815 **response at all concentrations and was lowered for subsequent laboratories.**

816
817 A sample assay design included the following chemicals and test concentrations as outlined
818 in Table 4 below. *Note: No more than two chemicals have been run per assay per week.*
819
820

Test Group	Exposure Medium	Number of wells (8 eleuthero-embryos/well)	Number of eleuthero-embryos
Solvent control	Test medium + solvent	1	8
Positive control-activation – concentration 1	EE2 34 ng/L + solvent	1	8
Positive control-activation – concentration 2	EE2 51 ng/L + solvent	1	8
Positive control-activation – concentration 3	EE2 76 ng/L + solvent	1	8
Positive control-activation – concentration 4	EE2 114 ng/L + solvent	1	8
Positive control-activation – concentration 5	EE2 171 ng/L + solvent	1	8
Positive control-activation – concentration 6	EE2 488 ng/L + solvent	1	8
Test chemical	Test chemical + solvent (5 concentrations)	1 per concentration (5 per test chemical)	40
Spiked mode reference control	Testosterone 30 µg/L + solvent	2	16
Spiked mode activation control	Testosterone 30 µg/L + EE2 64 ng/L + solvent	1	8
Spiked mode inhibition control	Testosterone 30 µg/L + fadrozole 10 µg/L + solvent	1	8
Testosterone + test chemical	Test chemical + testosterone 30 µg/L + solvent (5 concentrations)	1 per concentration (5 per chemical)	40
	TOTAL	21	168
	TOTAL - three experimental runs	63	504

Table 4: Assay Design with one test chemical

Test medium

Any test medium was allowed which permitted normal growth and development of medaka including glass bottled still mineral water, spring water, well water and charcoal-filtered tap water.

3.2.1. Because local water quality can differ substantially from one area to another, analysis of water quality should be undertaken to screen for potential contaminants (including heavy metals) and chemicals likely to interfere with the assay, particularly if historical data on the appropriateness of the water for raising medaka are not available. Special attention should be given to copper, chlorine and chloramine, all of which are toxic to medaka eleutheroembryos.

Some chemical characteristics of an acceptable test medium suitable for medaka can be found in below. However, any medium that supports the normal growth and development of medaka and allows the test validity criteria to be met such as glass bottled Evian™ water is suitable as a test medium.

Characteristic	Recommended range	Tolerance
Dechlorinated	-	Essential
Particle filtered	25 µm	Recommended
Activated charcoal filtered	-	Recommended
Conductivity	230-290 micro Siemens	Recommended
Temperature	26°C	26-30°C
pH	7.2-8.2	Essential

Table 5: Characteristics of water suitable for performing the REACTIV assay.

Alternatively, if a synthetic solution is to be used, one option is Medaka Medium. A stock solution of 10x Medaka Medium has the following composition:

- NaCl 5 g/L
- $CaCl_2$ 0.151 g/L
- $MgSO_4$ 0.098 g/L
- KCl 0.15 g/L
- NaOH 1N 1.25 mL/L

3.2.2. This solution was diluted ten-fold with reverse osmosis water to obtain the 1x working solution. The pH was then adjusted between 7.2-8.0 with a solution of 1N NaOH.

Test and control solutions

Test solutions of the chosen concentrations were prepared by dilution of a stock solution prepared in DMSO. The final concentration of DMSO was 0.2% in all test and control solutions.

Test and control solutions were prepared each day that they were required from aliquots of stock solutions in DMSO stored at -20°C. Any remaining thawed stock solution was discarded and was not refrozen.

Test validity

3.2.3. REACTIV experiments were judged valid during the validation exercise if the following criteria were met.

- The combined mortality and/or malformations and invalid data due to poorly positioned eleutheroembryos did not exceed one eleutheroembryo in each control group and in at least five treatment groups in the presence and absence of testosterone. Groups not meeting these criteria were considered compromised.

For the test to be valid, the following criteria should be met for the pool of the three runs, and if they are not, all three runs are considered invalid:

- A statistically significant fluorescence induction for the EE2 488 ng/L and testosterone controls compared to the solvent control.
- A statistically significant fluorescence induction for the testosterone plus EE2 control compared to the testosterone control.
- A statistically significant fluorescence inhibition for the testosterone and fadrozole control compared to the testosterone control.
- For the pool of the three runs, a test should have at least five uncompromised test concentrations. A treatment group (ideally 24 individuals) is considered uncompromised if in each of the three runs (ideally 8 individuals per run) it passes validity criteria (combined mortality, and/or malformations and invalid data due to poorly positioned eleutheroembryos should not exceed one eleutheroembryo).

These validity criteria were applied after image quality control was performed. If a 3.2.4. minor deviation from the validity criteria was observed, the consequences were considered in relation to the reliability of the test data.

Training

Personnel from the participating laboratories were not trained in person and performed the assay based on a written protocol. Videos of a key step (eleutheroembryo positioning for imaging) were provided to one of the participating laboratories.

Equipment

The following fluorescence imaging systems were used by the participating laboratories (Table 6).

3.2.5.

Partner laboratory	Microscope	Objective	Fluorescent filters	Fluorescence source	Camera	Software
Denmark	Nikon Ts2R	Plan apo 2x/0.1	ex470/40, em500LP, dicroic 495 nm	Built-in LED	DP74 - 0.55x Nikon camera adapter	Olympus Cellsens Dimension
Germany	Nikon SMZ 1000	Plan apo 1.0x; 8x zoom	GFP-LP	X-cite 120Q	Baumer TXD14C	Micromanager
Japan	Olympus MVX10	MVPLAPO 1X	Olympus U-MGFPHQ/XL	Olympus U-HGLGPS	Olympus DP80	Olympus Cellsens Standard
UK	Olympus IX-83	Plan achromat 2x	Olympus Cube U-F19002 GFP AT LP (ex475/40 em515LP dichroic 505)	CoolLED PE-300 LED illuminator	Olympus DP74 Camera, with a 0.63X C-mount adapter	Olympus Cellsens Dimension
USA	Nikon SMZ800N	Plan apo 1.0x; 8x zoom	P-EFLC GFP LP AT FILTER SET	SOLA SE II 365 Light Engine	Nikon DS-FI3	Nikon NIS Elements BR
France	Olympus IX-73	Plan apo PLN2X/0.06	Olympus Cube U-F19002 GFP AT LP (ex475/40 em515LP dichroic 505)	Prior L200/D 200W	Olympus DP74	Olympus Cellsens Dimension

Table 6: Imaging systems used for the interlaboratory validation study.

3.3.1.

3.3. Results of the interlaboratory validation study

Statistical methodology

The lead laboratory proposed a statistical method, this statistical decision tree was discussed with an independent statistical expert (Zhenglei Gao, Bayer) who has given her agreement that this statistical workflow is valid for the analysis of data from the similar RADAR assay.

The applied statistical workflow began with an image quality check to remove any images of badly positioned eleutheroembryos or other images that are not expected to provide an accurate measurement of fluorescence in the liver of the eleutheroembryos.

Data were then analysed following the directives of the OECD for the analysis of ecotoxicology experiments (OECD, 2006). Sample data was examined and variance was found to be homogenous as determined by Levene's test. Each experimental group was then analysed to determine whether there was a normal distribution of values. If the values followed a normal distribution, an analysis of variance (ANOVA) was conducted, followed by a parametric post-hoc test (Dunnetts post-test). If the values of one or more experimental groups were not normally distributed, a variance test (Kruskal-Wallis) was conducted, followed by a non-parametric post-hoc test (Dunns post-hoc test) to compare the groups with each other. Statistical significance was shown as: * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$; ns : not significant $p > 0.05$.

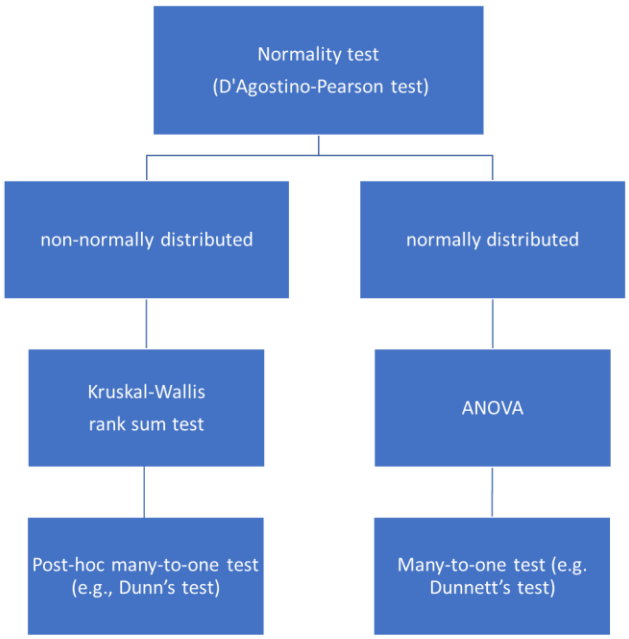


Figure 8: Flow chart for the statistical analysis of measured fluorescence

3.3.2.

False positive rate

3.3.3.

The statistical tests used which are considered positive for P<0.01 (**) control the false positive results at the 1% level.

Establishing a decision logic

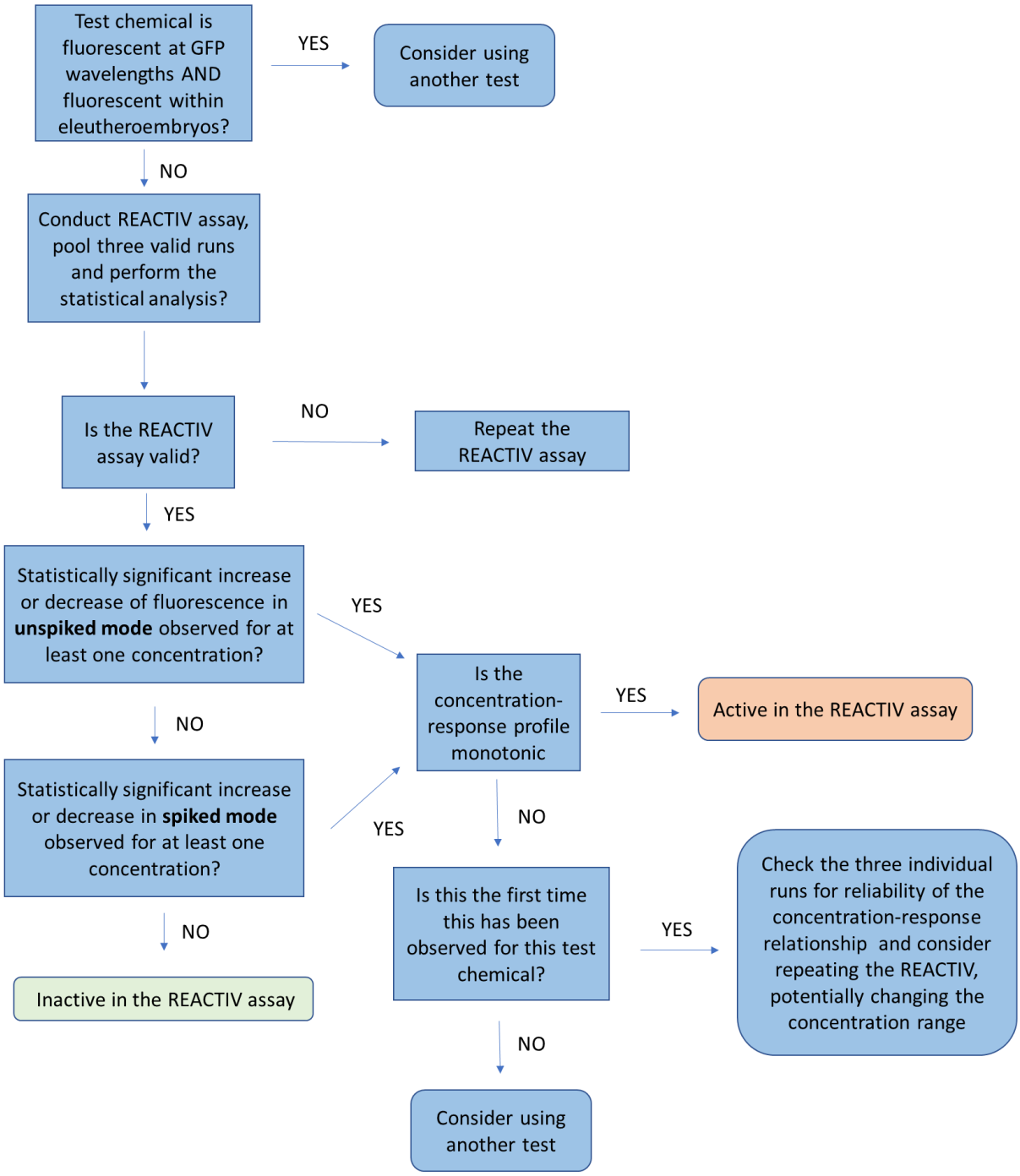
The REACTIV assay is intended to be used as a screening assay. The result obtained with will, therefore, likely influence decisions regarding further testing with additional assays.

A decision logic was developed for the REACTIV assay to provide logical assistance in the conduct and interpretation of the result of the bioassay (Figure 9). This decision logic is based on three valid runs pooled for statistical analysis (see Figure 7). A test chemical is considered to give a positive result in the REACTIV assay if at least one concentration tested is active in either unspiked or testosterone spiked mode and a concentration-response relationship is observed.

-In unspiked mode, an active concentration is defined as a concentration giving a statistically significant fluorescence increase or decrease compared to the test medium control.

-In testosterone spiked mode, an active concentration is defined as a concentration giving a statistically significant fluorescence increase or decrease compared to the 30 µg/L testosterone control.

Fluorescence decreases in unspiked mode are rare as the eleutheroembryos do not synthesise high levels of estrogens at this development stage. If a statistically significant fluorescence decrease is observed in unspiked mode, it could indicate that the REACTIV assay is not appropriate for the test chemical, or a potential problem with the organisms or the test conditions which may require further investigations. Individual runs should be considered to determine if the statistically significant fluorescence decrease is present in the three runs and best professional judgement should then be used to decide between repeating: none of the runs, only one run using a new batch of organisms; a complete REACTIV, possibly using a lower concentration range; or performing a different estrogen axis activity test.



3.3.4.

Figure 9: Decision logic for the conduct of the REACTIV assay

Establishing NOEC and LOEC

The result of the REACTIV assay is intended to be a classification of the test chemicals into potentially "estrogen axis active" or "estrogen axis inactive". The results of the REACTIV assay are expressed here in terms of LOEC and NOEC to allow the comparison

of the results between the participating laboratories as a decisional aid for the possible validation of the assay.
The LOEC is defined as the lowest concentration found to be active either in unspiked or spiked mode. The NOEC is defined as the concentration tested immediately below the LOEC.

3.4. Results of Analyses

The results presented here are the results obtained with the statistical approach and decision logic described above. Due to the fact that staff involved in this interlaboratory validation study were not experienced with the REACTIV assay or handling medaka eleuthero-embryos, a higher tolerance for mortality and exclusion of inadequate images was employed. This criterion has now been refined in the draft test guideline to no less than 87.5% of the expected number of values per group (no more than one eleutheroembryo per group), for controls and interpretable test chemical groups, following removal of data due to dead, malformed or immobile eleutheroembryos as well as images of poorly positioned eleutheroembryos.
Had the 12.5% limit for excluded data been applied to this data set, a small number of individual runs would have been excluded due to violation of the 12.5% limit by one or more control groups.

3.4.1. *Calibration*

Selecting image capture settings

Each laboratory performed an initial experiment to determine the optimal imaging settings to allow a satisfactory amplitude of GFP induction to be obtained. This involved exposing 40 eleutheroembryos to 488 ng/L of EE2 and adjusting parameters relating to image capture (white balance, gain, exposure time). These parameters were then fixed for all future experiments. This step had already been performed a number of years previously at the lead laboratory and, therefore, was not repeated.

Determining linearity and sensitivity to EE2 and performance of the controls

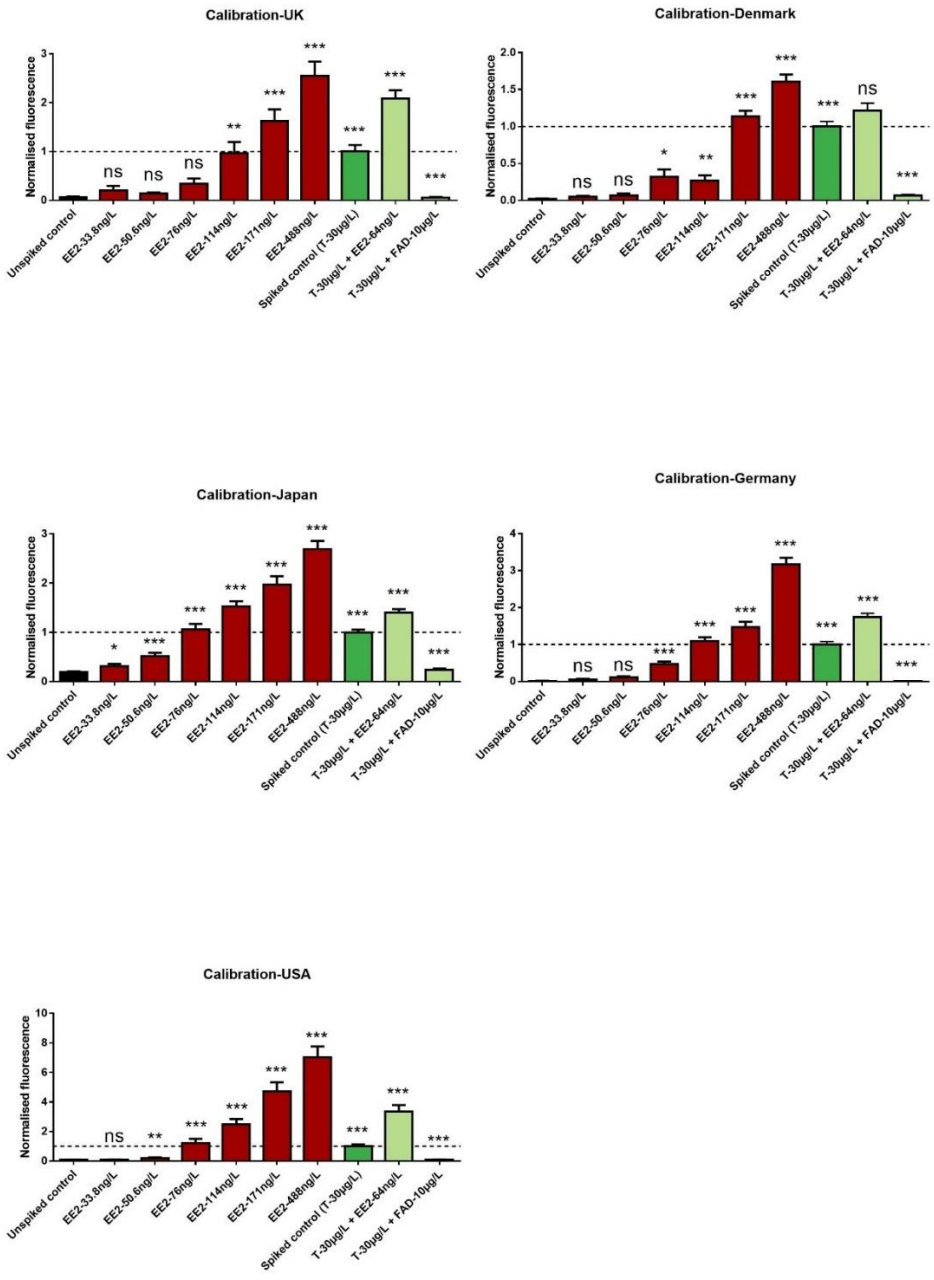
Applying the image capture parameters that were determined in the previous calibration experiment, three runs of the REACTIV assay were carried out by each participating laboratory. No test chemical was included in these three runs, which were limited to control groups only.

The first aim of this experiment was to generate data to allow an image analysis workflow to be selected which allowed background (non-GFP) fluorescence signal to be minimised prior to quantification of the images produced by each laboratory. The second aim was to verify that the fluorescence values obtained for a set of control solutions passed validity

criteria when using the image capture parameters that had previously been determined. The lead laboratory did not perform this step as they had previously optimised and tested the image capture parameters for their imaging system.

Figure 10 shows the results of the second calibration experiment. All five naïve laboratories obtained results showing increasing fluorescence with increasing concentrations of EE2 and statistically significant differences in fluorescence values for the assay controls when compared to their relevant control group.

It should be noted that the data shown for the UK is based on two experimental runs and not three due to an error when capturing the images for run two. This explains the reduced sensitivity for EE2 of 114 ng/L. The other four participating laboratories obtained EE2 sensitivities of 51 ng/L (Japan and USA) and 76 ng/L (Germany) and 114 ng/L (Denmark) as shown in Table 7.



1033

1034 **Figure 10: Mean and SEM of fluorescence for assay controls employed within the**
1035 **REACTIV assay obtained during calibration.**

1036 Fluorescence values were normalised to the mean of the testosterone group, the
1037 value of this group is indicated with a dashed line.

1038

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
UK	ns	ns	ns	**	***	***	***	***	***
Denmark	ns	ns	*	**	***	***	***	***	***
Japan	*	***	***	***	***	***	***	***	***
Germany	ns	ns	***	***	***	***	***	***	***
USA	ns	**	***	***	***	***	***	***	***

Table 7: Summarised statistical results for the second step of the calibration experiment.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green.

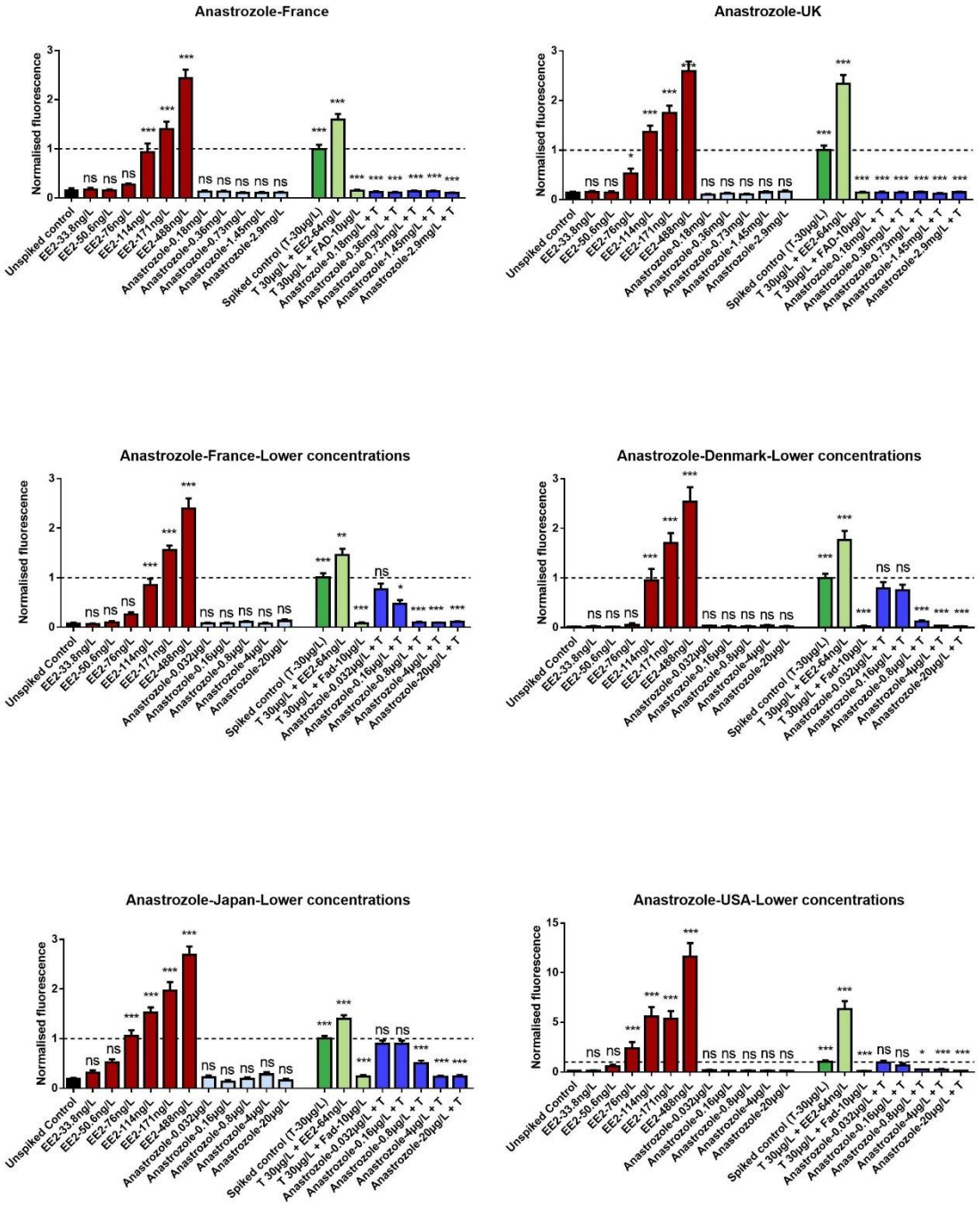
3.5. Results for estrogen axis active chemicals

Anastrozole Results

3.5.1.

Anastrozole is a pharmaceutical aromatase enzyme inhibitor. As such it was expected that by blocking conversion of androgens to estrogens, it would decrease estrogen axis activity and, therefore, fluorescence in the presence of testosterone by blocking its conversion to estradiol. In the absence of appreciable levels of testosterone (unspiked mode) it was expected that anastrozole would have no effect on fluorescence levels.

1053 Figure 11



1054
1055 shows the results obtained when testing anastrozole. A summary of the statistical analysis
1056 is provided in Table 8. It can be noted that all four laboratories that tested anastrozole
1057 obtained an LOEC of 76-114 ng/L for the EE2 controls. All laboratories also observed
1058 statistically significant differences for the spiked controls. No statistically significant
1059 variation in fluorescence was observed in unspiked mode for anastrozole in any of the
1060 laboratories. The first two laboratories (France and the UK) to test anastrozole observed an

extremely strong and statistically significant inhibition of fluorescence in spiked mode for all concentrations tested (0.18-2.9 mg/L). The lead laboratory (France) then tested a lower concentration range (0.032-20 µg/L) which was set as the concentration range for all subsequent laboratories.

A decrease in fluorescence that was proportional to the concentration of anastrozole was observed in the presence of testosterone in all four laboratories that tested the lower concentration range. All four laboratories detected a statistically significant decrease in fluorescence with 4 µg/L of anastrozole, with three of the laboratories detecting a decrease at 0.8 µg/L (France, Denmark and Japan).

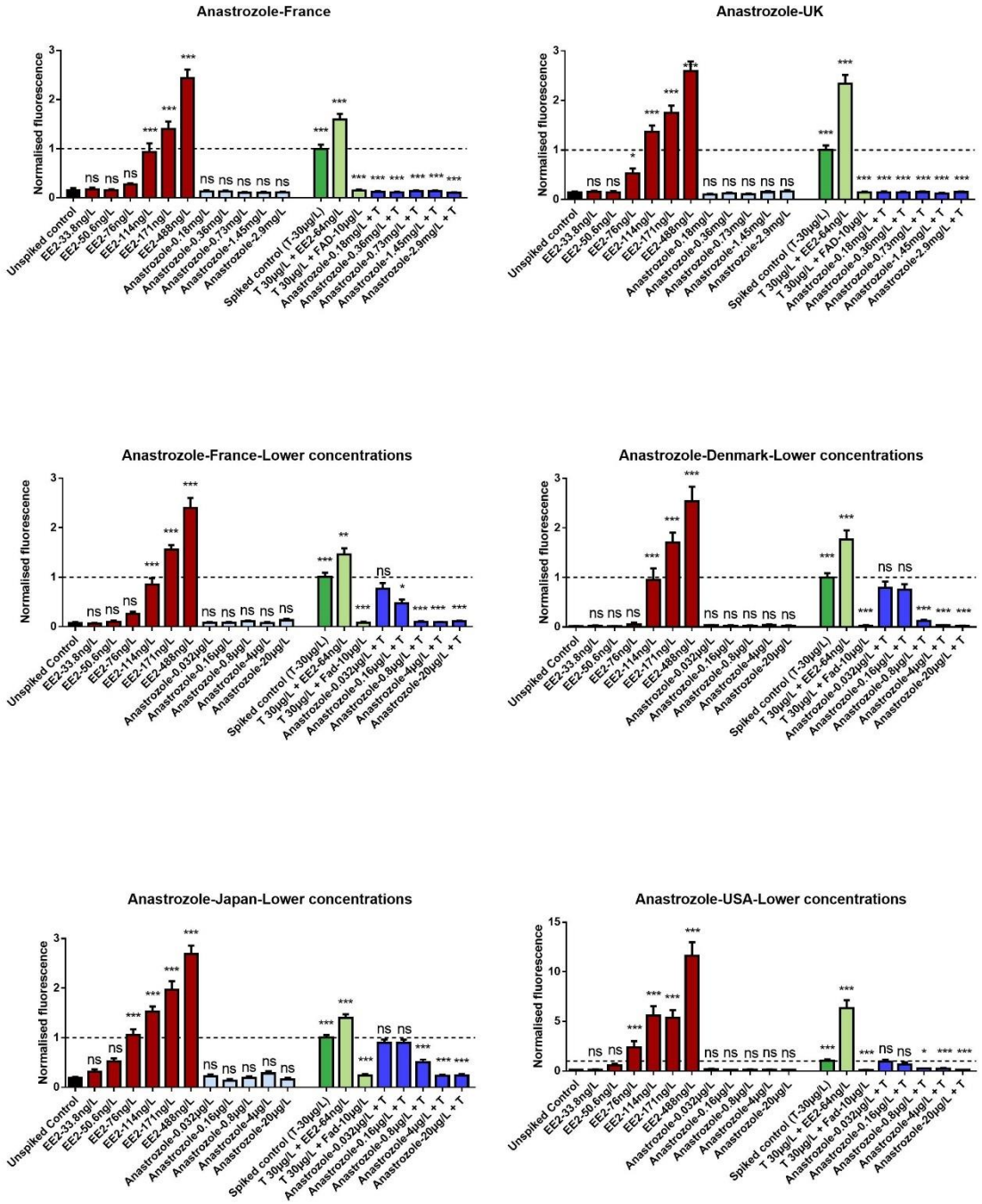


Figure 11: Mean and SEM of measured fluorescence for anastrozole.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
France	ns	ns	ns	***	***	***	***	***	***
UK	ns	ns	*	***	***	***	***	***	***
France	ns	ns	ns	***	***	***	***	**	***
Denmark	ns	ns	ns	***	***	***	***	***	***
Japan	ns	ns	***	***	***	***	***	***	***
USA	ns	ns	***	***	***	***	***	***	***

Laboratory	Anastrozole (µg/L)					Anastrozole + Testosterone (µg/L)				
	0.032	0.16	0.8	4	20	0.032	0.16	0.8	4	20
France	ns	ns	ns	ns	ns	ns	*	***	***	***
Denmark	ns	ns	ns	ns	ns	ns	ns	***	***	***
Japan	ns	ns	ns	ns	ns	ns	ns	***	***	***
USA	ns	ns	ns	ns	ns	ns	ns	*	***	***

Table 8: Summarised statistical results for the anastrozole experiments.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green. Laboratory names marked in italics indicate assays performed with the lower concentration range of anastrozole.

3.5.2. *Bisphenol A Results*

Bisphenol A (BPA) is a plasticiser which has been widely published as having pro-estrogenic activity (Ni *et al.*, 2022).

Figure 12 below shows the mean and SEM for each concentration of BPA in each laboratory. An increase in fluorescence was obtained in each laboratory with increasing concentrations of BPA. Only the Danish laboratory did not observe a statistically significant increase in fluorescence in unspiked mode. However, all six laboratories observed a significant increase in fluorescence in testosterone-spiked mode when compared to the testosterone control.

The LOEC for this increase in spiked mode was 2 mg/L (UK and Germany), 3 mg/L (Japan, France and USA) and 4 mg/L (Denmark). As indicated in Table 9. All spiked mode controls showed the expected statistically significant changes in fluorescence in all six laboratories and all laboratories observed a similar concentration-response to the EE2 controls.

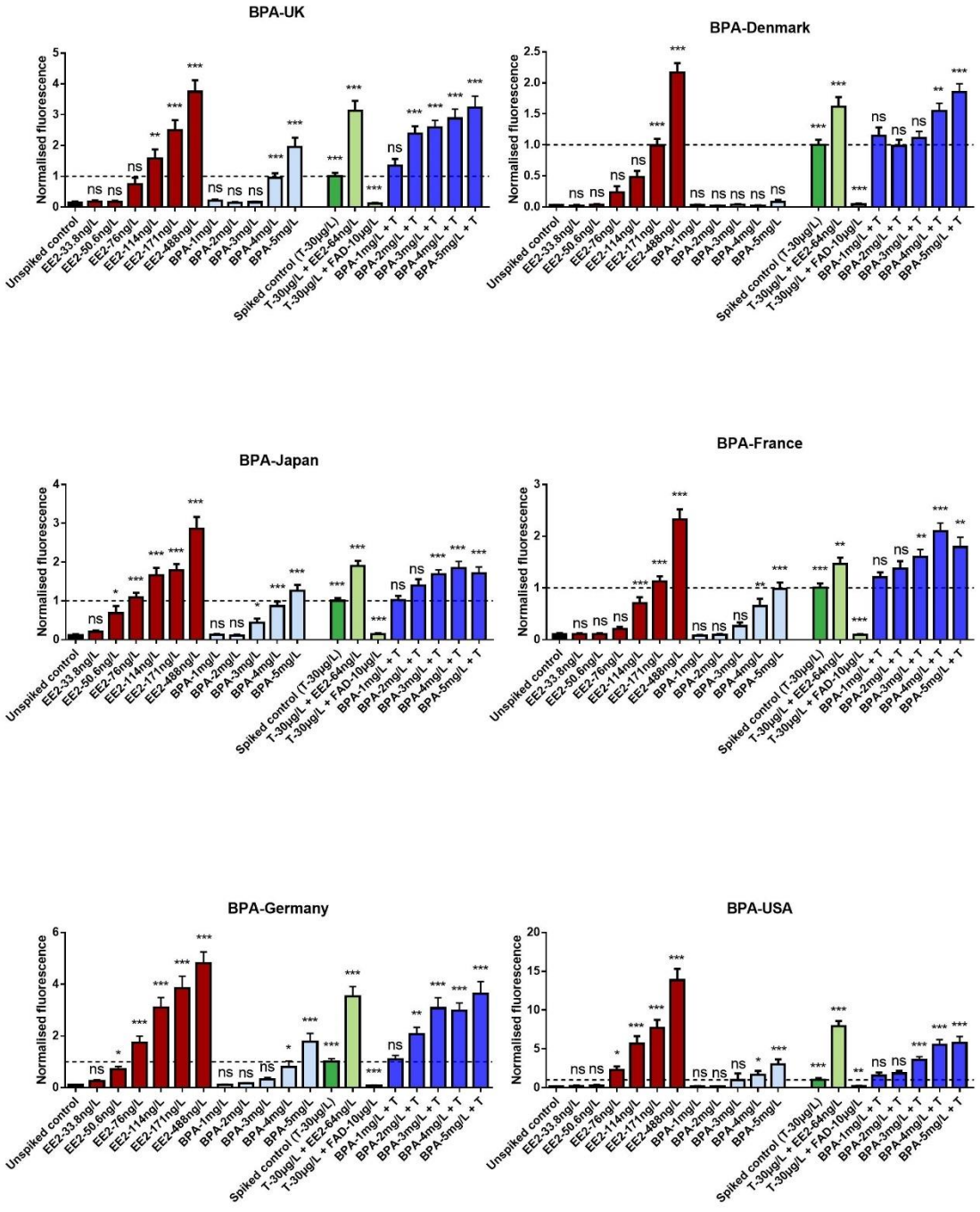


Figure 12: Mean and SEM of measured fluorescence for bisphenol A.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
UK	ns	ns	ns	**	***	***	***	***	***
Denmark	ns	ns	ns	ns	***	***	***	***	***
Japan	ns	*	***	***	***	***	***	***	***
France	ns	ns	ns	***	***	***	***	**	***
Germany	ns	*	***	***	***	***	***	***	***
USA	ns	ns	*	***	***	***	***	***	**

Laboratory	BPA (mg/L)					BPA + Testosterone (mg/L)				
	1	2	3	4	5	1	2	3	4	5
UK	ns	ns	ns	***	***	ns	***	***	***	***
Denmark	ns	ns	ns	ns	ns	ns	ns	ns	**	***
Japan	ns	ns	*	***	***	ns	ns	***	***	***
France	ns	ns	ns	**	***	ns	ns	**	***	**
Germany	ns	ns	ns	*	***	ns	**	***	***	***
USA	ns	ns	ns	*	***	ns	ns	***	***	***

Table 9: Summarised statistical results for the BPA experiments.

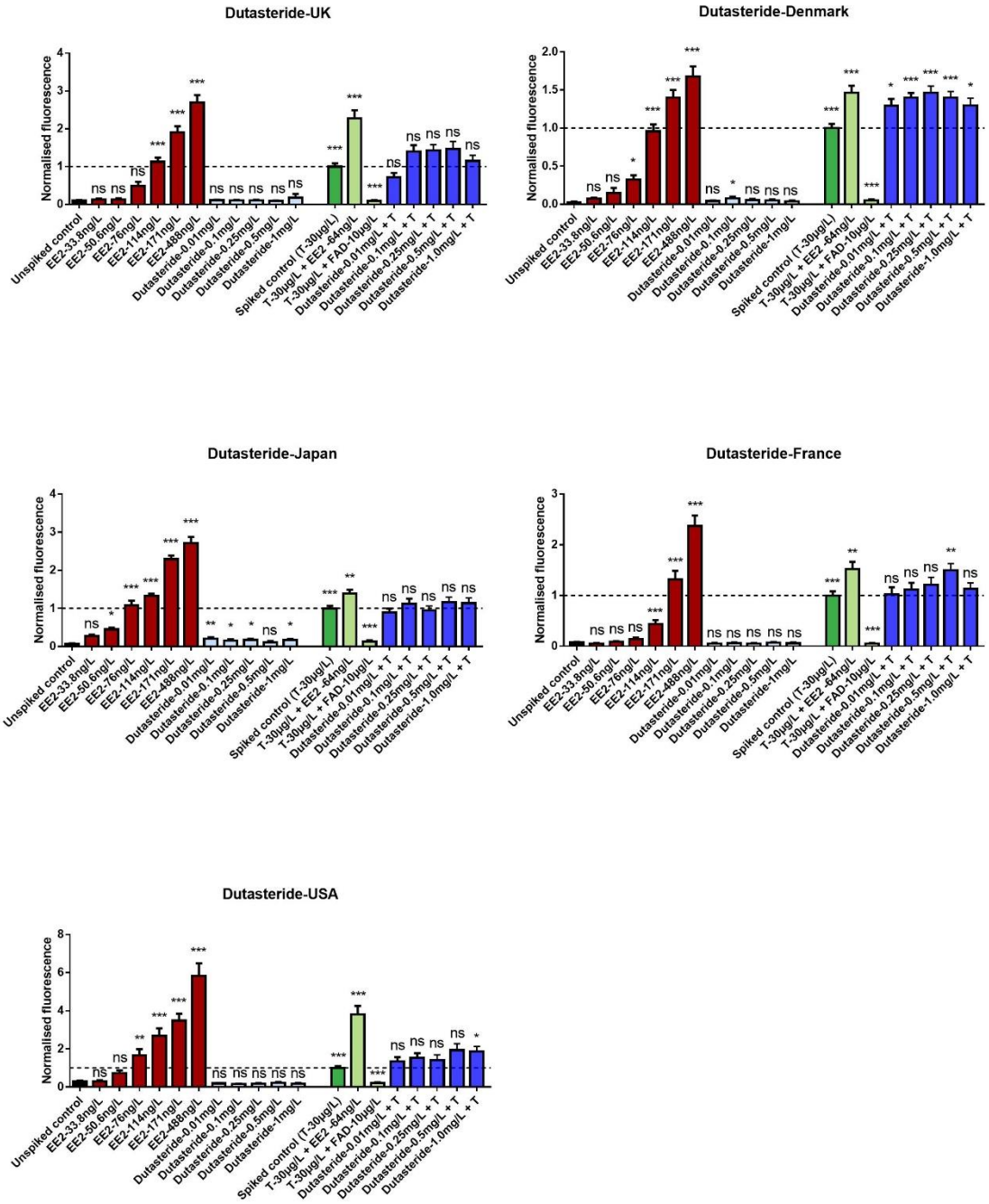
Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green.

Dutasteride Results

Dutasteride is a pharmacological inhibitor of type I and II 5α-reductase which has also been shown to act as an AR antagonist in certain cell lines (Chhipa *et al.*, 2013). The inhibitory action of this pharmaceutical on 5α-reductase activity blocks conversion of testosterone to the non-aromatisable androgen dihydrotestosterone (DHT), increasing the pool of testosterone available for conversion to estradiol. Therefore, the REACTIV assay would be expected to give an increase in fluorescence in testosterone-spiked mode and no effect in unspiked mode.

Figure 13 below shows the mean and SEM for the assay controls and tested concentrations of dutasteride in each laboratory. All spiked mode controls were statistically significant for the experiments performed by each partner laboratory (Table 10). A concentration-dependent increase in fluorescence was observed for the EE2 controls in all laboratories with a LOEC of 76 or 114 ng/L. As expected no statistically significant variation in fluorescence $P<0.01$ (indicated by two stars) was observed in unspiked mode. Except in the Japanese laboratory where the lowest tested concentration showed a statistically significant increase in fluorescence.

In spiked mode, only the Danish and French laboratories obtained an increase in fluorescence $P<0.01$. Interestingly, the highest concentration tested (1 mg/L) had a lower mean fluorescence value than the previous three concentrations in both laboratories. This may indicate sub-lethal toxicity. A similar, but statistically insignificant profile was observed in the tests performed by the UK, and the USA.



1140

1141 **Figure 13: Mean and SEM of measured fluorescence for dutasteride.**

1142 Fluorescence values were normalised to the mean of the testosterone 30 µg/L
1143 group, the value of this group is indicated with a dashed line.

1144
1145

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
UK	ns	ns	ns	***	***	***	***	***	***
Denmark	ns	ns	*	***	***	***	***	***	***
Japan	ns	*	***	***	***	***	***	**	***
France	ns	ns	ns	***	***	***	***	**	***
USA	ns	ns	**	***	***	***	***	***	***

Laboratory	Dutasteride (mg/L)					Dutasteride + Testosterone (mg/L)				
	1	2	3	4	5	1	2	3	4	5
UK	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Denmark	ns	*	ns	ns	ns	*	***	***	***	*
Japan	**	*	*	ns	*	ns	ns	ns	ns	ns
France	ns	ns	ns	ns	ns	ns	ns	ns	**	ns
USA	ns	ns	ns	ns	ns	ns	ns	ns	ns	*

Table 10: Summarised statistical results for the dutasteride experiments.

Results corresponding to a statistically significant variation of fluorescence (P<0.01) are highlighted in green.

3.5.4. *17β-Estradiol Results*

The results obtained when testing the natural estrogen receptor agonist, 17β-estradiol, are shown below in Figure 14. A summary of the statistical analysis is given in Table 11 and shows that all six laboratories identified a statistically significant (P<0.01) difference in mean normalised fluorescence for the spiked controls. All participating laboratories also obtained a concentration-dependent increase in fluorescence for the EE2 controls.

Likewise, a concentration-dependent increase in fluorescence was observed in all six laboratories for 17β-estradiol in both spiked and unspiked mode. The LOECs in unspiked mode were 135 ng/L (UK, France, Japan and USA) and 270 ng/L (Denmark) and in spiked mode were 68 ng/L (UK, France, Japan and USA) and 135 ng/L (Denmark). Higher LOECs of 540 ng/L (unspiked mode) and 270 ng/L (spiked mode) were obtained for the German laboratory, despite clear inductions in fluorescence at lower concentrations. Upon examination of the three individual runs, a clear concentration-dependent increase in fluorescence was observed for runs one and three, with no induction whatsoever for 17β-estradiol in either spiked or unspiked mode for run two. All controls performed normally in run two, suggesting that the test chemical was omitted from the run. Unfortunately, no analytical samples were taken for this run by this laboratory, preventing confirmation of this hypothesis.

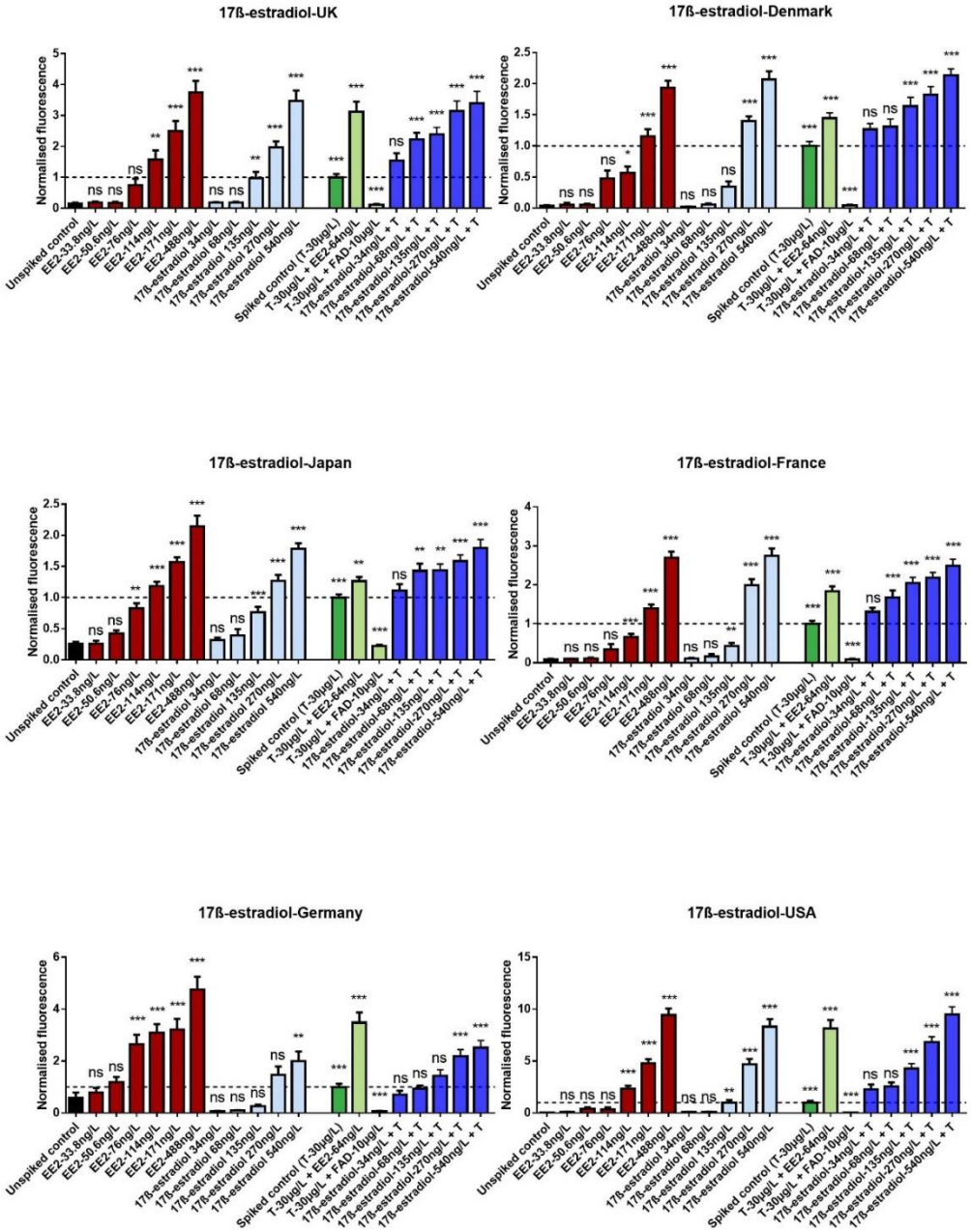


Figure 14: Mean and SEM of measured fluorescence for 17β-estradiol.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
UK	ns	ns	ns	**	***	***	***	***	***
Denmark	ns	ns	ns	*	***	***	***	***	***
Japan	ns	ns	**	***	***	***	***	**	***
France	ns	ns	ns	***	***	***	***	***	***
Germany	ns	ns	***	***	***	***	***	***	***
USA	ns	ns	ns	***	***	***	***	***	***

Laboratory	17β-estradiol (ng/L)					17β-estradiol + testosterone (ng/L)				
	34	68	135	270	540	34	68	135	270	540
UK	ns	ns	**	***	***	ns	***	***	***	***
Denmark	ns	ns	ns	***	***	ns	ns	***	***	***
Japan	ns	ns	***	***	***	ns	**	**	***	***
France	ns	ns	**	***	***	ns	***	***	***	**
Germany	ns	ns	ns	ns	**	ns	ns	ns	***	***
USA	ns	ns	**	***	***	ns	***	***	***	***

Table 11: Summarised statistical results for the 17β-estradiol experiments.

Results corresponding to a statistically significant variation of fluorescence (P<0.01) are highlighted in green.

3.5.5. *Estrone Results*

An additional natural estrogen receptor agonist, estrone, was also tested in two laboratories, the results are shown below in Figure 15. A summary of the statistical analysis is given in Table 12 and shows that both laboratories identified a statistically significant (P<0.01) difference in mean normalised fluorescence for the spiked controls. Both laboratories also obtained a concentration-dependent increase in fluorescence for the EE2 controls.
Likewise, a concentration-dependent increase in fluorescence was observed in both laboratories for estrone in both spiked and unspiked mode. The LOECs in unspiked mode were 5 µg/L in both the French and German laboratories and were 1 µg/L in spiked mode for both laboratories.

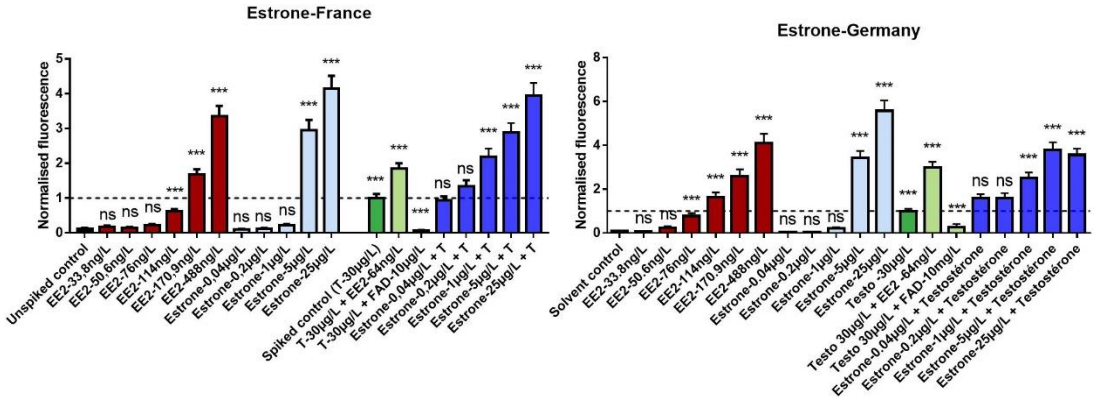


Figure 15: Mean and SEM of measured fluorescence for estrone.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
France	ns	ns	ns	***	***	***	***	***	***
Germany	ns	ns	***	***	***	***	***	***	***

Laboratory	estrone (µg/L)					estrone + testosterone (µg/L)				
	0.004	0.2	1	5	25	0.004	0.2	1	5	25
France	ns	ns	ns	***	***	ns	ns	***	***	***
Germany	ns	ns	ns	***	***	ns	ns	***	***	***

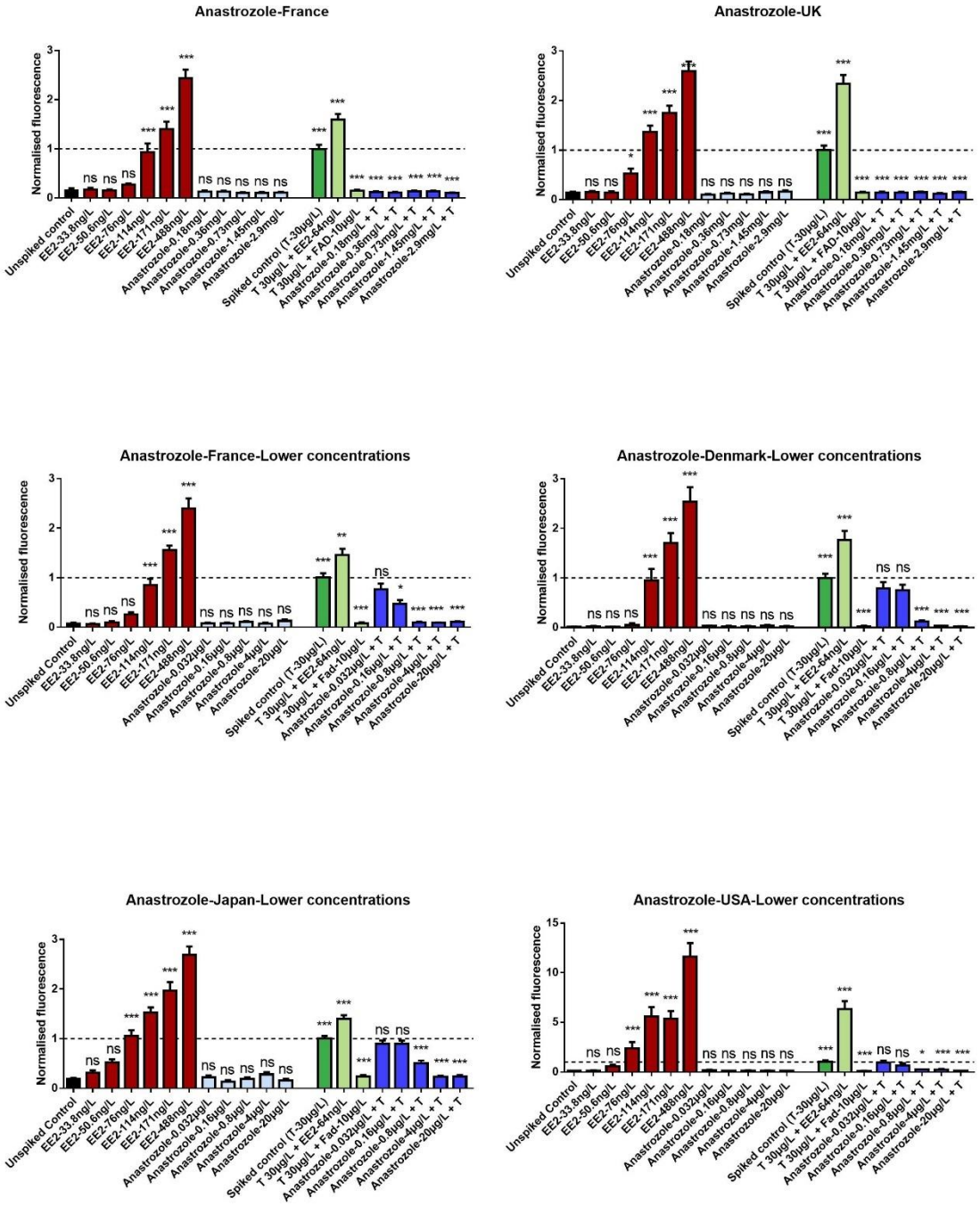
Table 12: Summarised statistical results for the estrone experiments.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green.

Fadrozole Results

Like anastrozole, fadrozole is a pharmaceutical aromatase enzyme inhibitor. It is, therefore, also expected to block conversion of androgens to estrogens and decrease estrogen axis activity and, therefore, fluorescence in the presence of testosterone by blocking its conversion to estradiol in testosterone-spiked mode. In unspiked mode, it was expected that fadrozole would have no effect on fluorescence levels due to the low level of testosterone present at eleutheroembryonic life stages.

1223 Figure 16



1224
1225 shows the results obtained for fadrozole. A summary of the statistical analysis is provided
1226 in Table 13. It can be noted that all four laboratories that tested fadrozole obtained an LOEC
1227 of 76-114 ng/L for the EE2 controls. All laboratories also observed statistically significant
1228 differences for the spiked controls. No statistically significant variation in fluorescence was
1229 observed in unspiked mode for fadrozole in any of the laboratories. The first two
1230 laboratories (France and the UK) to test fadrozole observed an extremely strong and

statistically significant inhibition of fluorescence in spiked mode for all concentrations tested (5.4-87 µg/L). The lead laboratory (France) then tested a lower concentration range (0.016-10 µg/L) which was set as the concentration range for all subsequent laboratories. A decrease in fluorescence that was proportional to the concentration of fadrozole was observed in the presence of testosterone in all three laboratories that tested the lower concentration range (France, Germany and USA). The LOEC for fadrozole in these laboratories ranged from 0.4-10 µg/L.

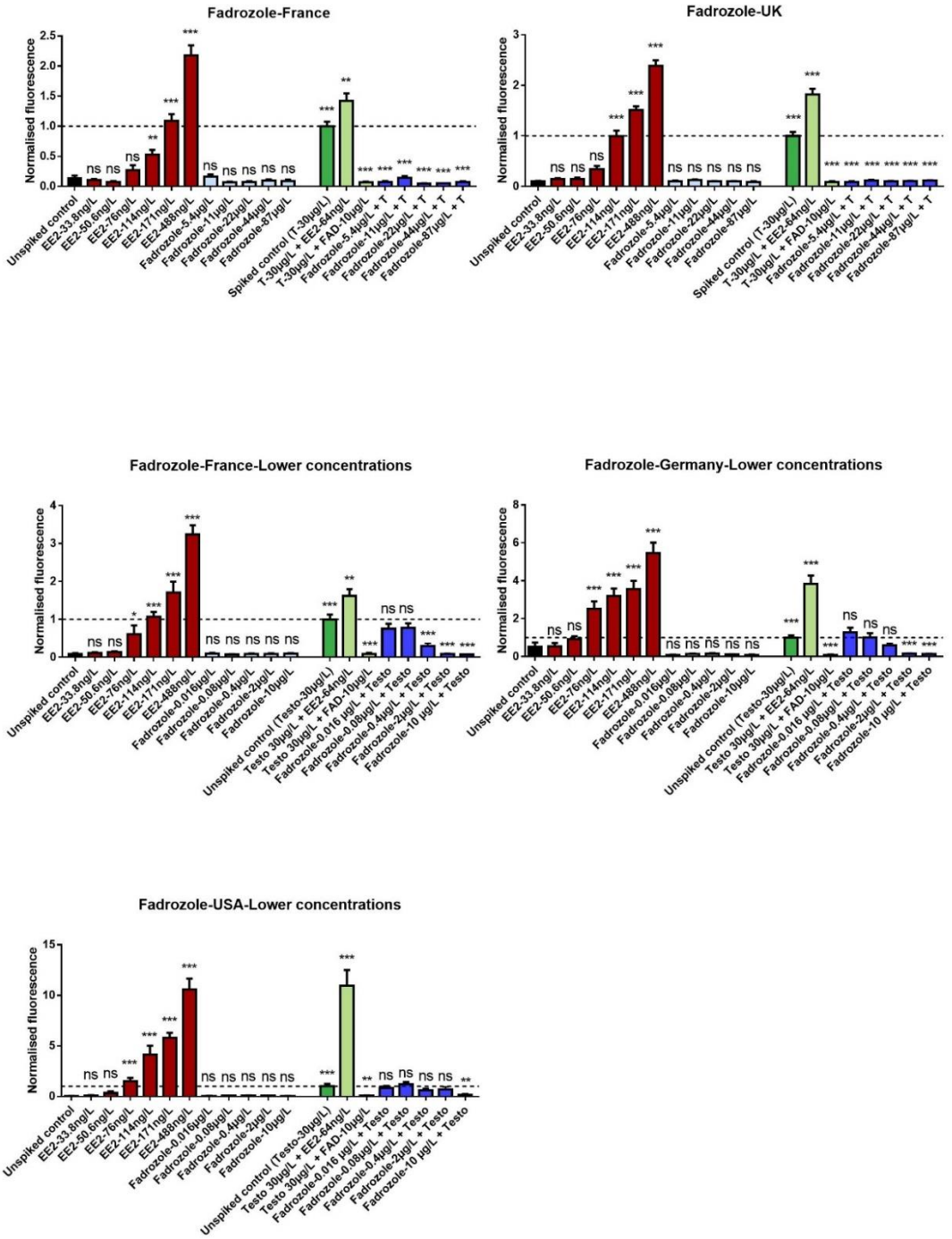


Figure 16: Mean and SEM of measured fluorescence for fadrozole.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
France	ns	ns	ns	**	***	***	***	**	***
UK	ns	ns	ns	***	***	***	***	***	***
France	ns	ns	*	***	***	***	***	**	***
Germany	ns	ns	***	***	***	***	***	***	***
USA	ns	ns	***	***	***	***	***	***	**

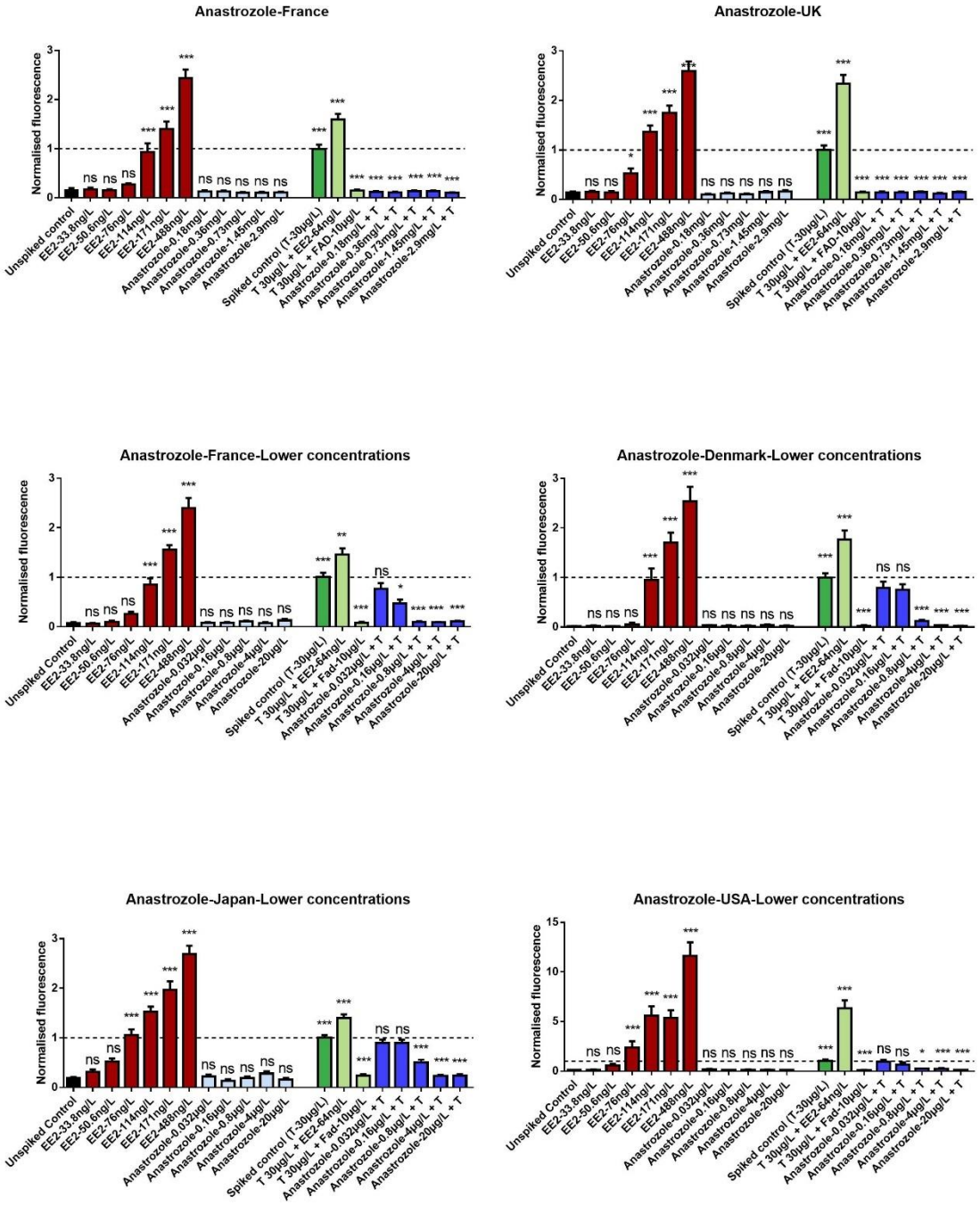
Laboratory	Fadrozole (µg/L)					Fadrozole + Testosterone (µg/L)				
	0.016	0.08	0.4	2	10	0.016	0.08	0.4	2	10
France	ns	ns	ns	ns	ns	ns	ns	***	***	***
Germany	ns	ns	ns	ns	ns	ns	ns	ns	***	***
USA	ns	ns	ns	ns	ns	ns	ns	ns	ns	**

Table 13: Summarised statistical results for the fadrozole experiments.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green. Laboratory names marked in italics indicate assays performed with the lower concentration range of fadrozole.

3.5.7. *Prochloraz Results*

The imidazole fungicide prochloraz has been shown to inhibit the expression of aromatase enzyme (Higley *et al.*, 2010). It would, therefore, be expected to decrease estrogen axis activity by inhibiting the conversion of androgens to estrogens. Figure 17



shows the results obtained for prochloraz. A summary of the statistical analysis is provided in Table 14.

It can be noted that all six laboratories that tested prochloraz obtained an LOEC of 51-114 ng/L for the EE2 controls. All laboratories also observed statistically significant differences for the spiked controls.

None of the six laboratories testing prochloraz observed a statistically significant variation in fluorescence in unspiked mode.

As expected from its mode of action, a decrease in fluorescence that was proportional to the concentration of prochloraz was observed in the presence of testosterone in five of the six laboratories, with only the UK laboratory failing to detect a statistically significant effect. The LOEC for this inhibition of estrogen signalling in spiked mode, for the five laboratories detecting it, was 0.063-0.5 mg/L.

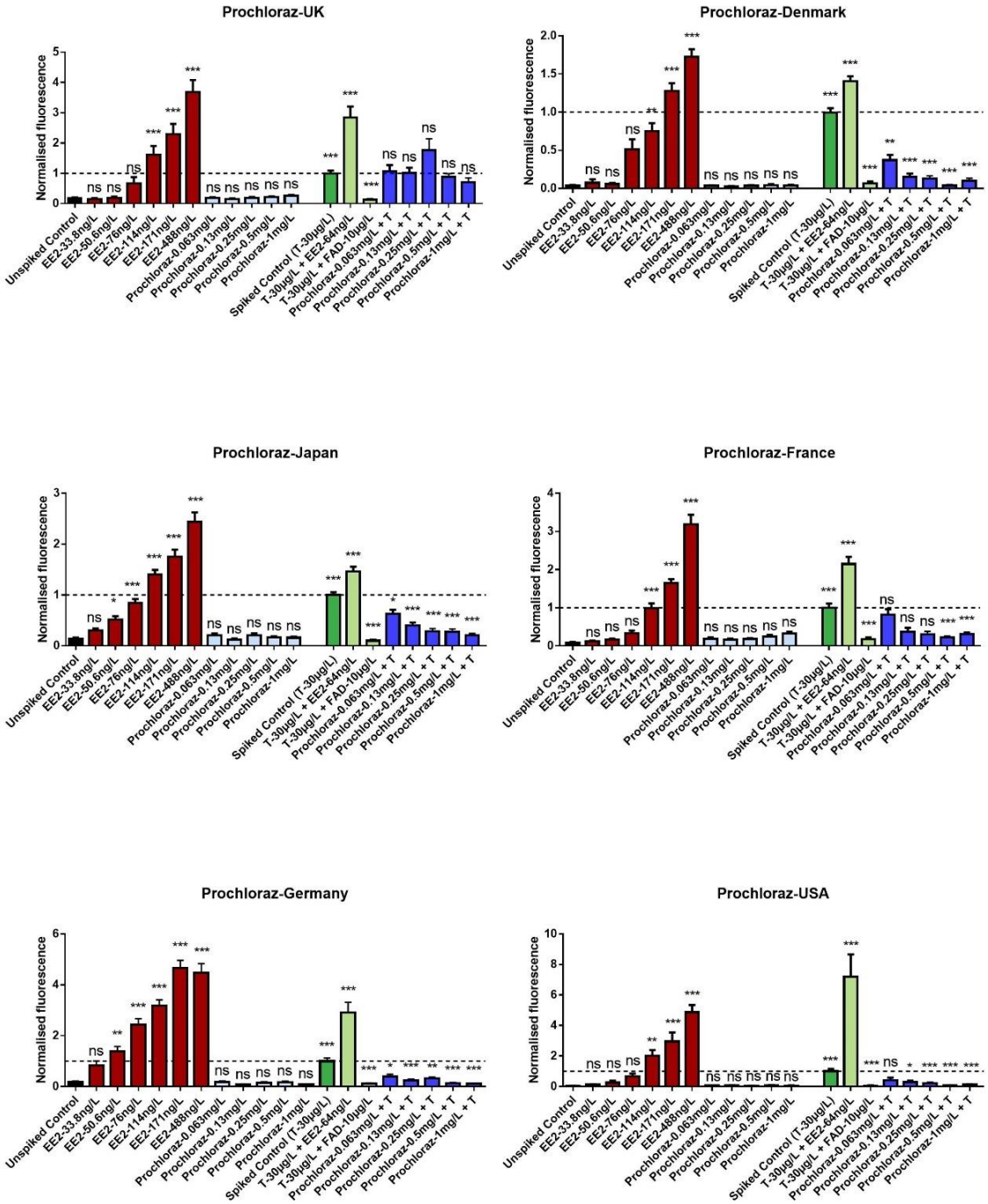


Figure 17: Mean and SEM of measured fluorescence for prochloraz.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
UK	ns	ns	ns	***	***	***	***	***	***
Denmark	ns	ns	ns	**	***	***	***	***	***
Japan	ns	*	***	***	***	***	***	***	***
France	ns	ns	ns	***	***	***	***	***	***
Germany	ns	**	***	***	***	***	***	***	***
USA	ns	ns	ns	**	***	***	***	***	***

Laboratory	Prochloraz (mg/L)					Prochloraz + Testosterone (mg/L)				
	0.063	0.13	0.25	0.5	1	0.063	0.13	0.25	0.5	1
UK	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Denmark	ns	ns	ns	ns	ns	**	***	***	***	***
Japan	ns	ns	ns	ns	ns	*	***	***	***	***
France	ns	ns	ns	ns	ns	ns	ns	ns	***	***
Germany	ns	ns	ns	ns	ns	*	***	**	***	***
USA	ns	ns	ns	ns	ns	ns	*	***	***	**

Table 14: Summarised statistical results for the prochloraz experiments.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green.

3.5.8.

Tamoxifen Results

Tamoxifen is a pharmaceutical selective estrogen response modulator (SERM). As such, it was expected to give an induction of fluorescence in the absence of estrogenic signalling (unspiked mode). Decreases in fluorescence in the presence of estrogenic signalling (spiked mode) have also been observed with longer exposures, but weren't expected here (Spirhanzlova *et al.*, 2016).

A statistically significant concentration-dependent increase in fluorescence was observed for the EE2 controls in all five laboratories with a LOEC of 76-114 ng/L. The spiked controls gave the expected statistically significant differences in all five laboratories. Four of the five laboratories detected a statistically significant induction in estrogen axis signalling in unspiked mode (UK, Japan, France and USA), with a LOEC of 242-483 µg/L. The Danish laboratory did not identify a statistically significant increase in fluorescence despite a visible concentration-response in the mean fluorescence values ($P=0.08$ for 483 µg/L).

In spiked mode, no statistically significant differences in fluorescence were observed except for the American laboratory which observed an increase in fluorescence from 121-483 µg/L.

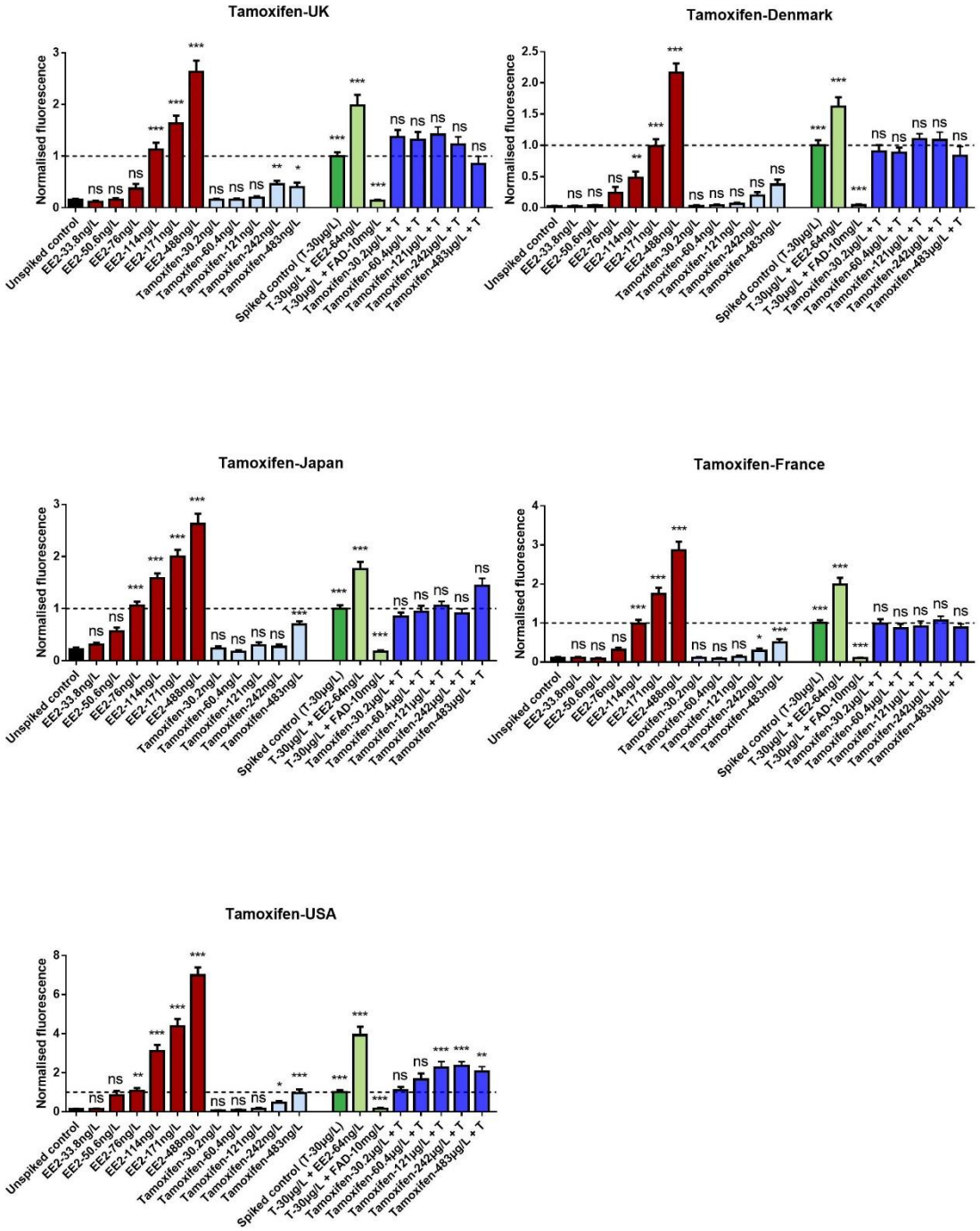


Figure 18: Mean and SEM of measured fluorescence for tamoxifen.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
UK	ns	ns	ns	***	***	***	***	***	***
Denmark	ns	ns	ns	**	***	***	***	***	***
Japan	ns	ns	***	***	***	***	***	***	***
France	ns	ns	ns	***	***	***	***	***	***
USA	ns	ns	**	***	***	***	***	***	***

Laboratory	Tamoxifen (µg/L)					Tamoxifen + Testosterone (µg/L)				
	30	60	121	242	483	30	60	121	242	483
UK	ns	ns	ns	**	*	ns	ns	ns	ns	ns
Denmark	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Japan	ns	ns	ns	ns	***	ns	ns	ns	ns	ns
France	ns	ns	ns	*	***	ns	ns	ns	ns	ns
USA	ns	ns	ns	*	***	ns	ns	***	***	**

Table 15: Summarised statistical results for the tamoxifen experiments.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green.

3.5.9. *Testosterone Results*

Testosterone is an aromatisable androgen. As such, it is metabolised into estradiol by the *chgh-gfp* eleutheroembryos. It is, therefore, expected to induce estrogen axis signalling and fluorescence in the eleutheroembryos.

Figure 19 shows the results obtained in the five different laboratories that tested testosterone. All laboratories obtained a concentration-dependent increase for the EE2 controls (Table 16) with a LOEC of 76 ng/L in all laboratories except the UK (114 ng/L). The spiked mode controls showed the expected statistically significant differences in fluorescence ($P<0.01$) in all laboratories except the testosterone + EE2 control in the Danish laboratory ($P= 0.016$).

Testosterone induced a concentration-dependent increase in fluorescence in unspiked mode in all laboratories with a LOEC of 11-33 µg/L. In testosterone-spiked mode the capacity of aromatase to further metabolise testosterone appeared saturated with only the

laboratories for Denmark and the USA detecting statistically significant ($P<0.01$) increases in fluorescence.

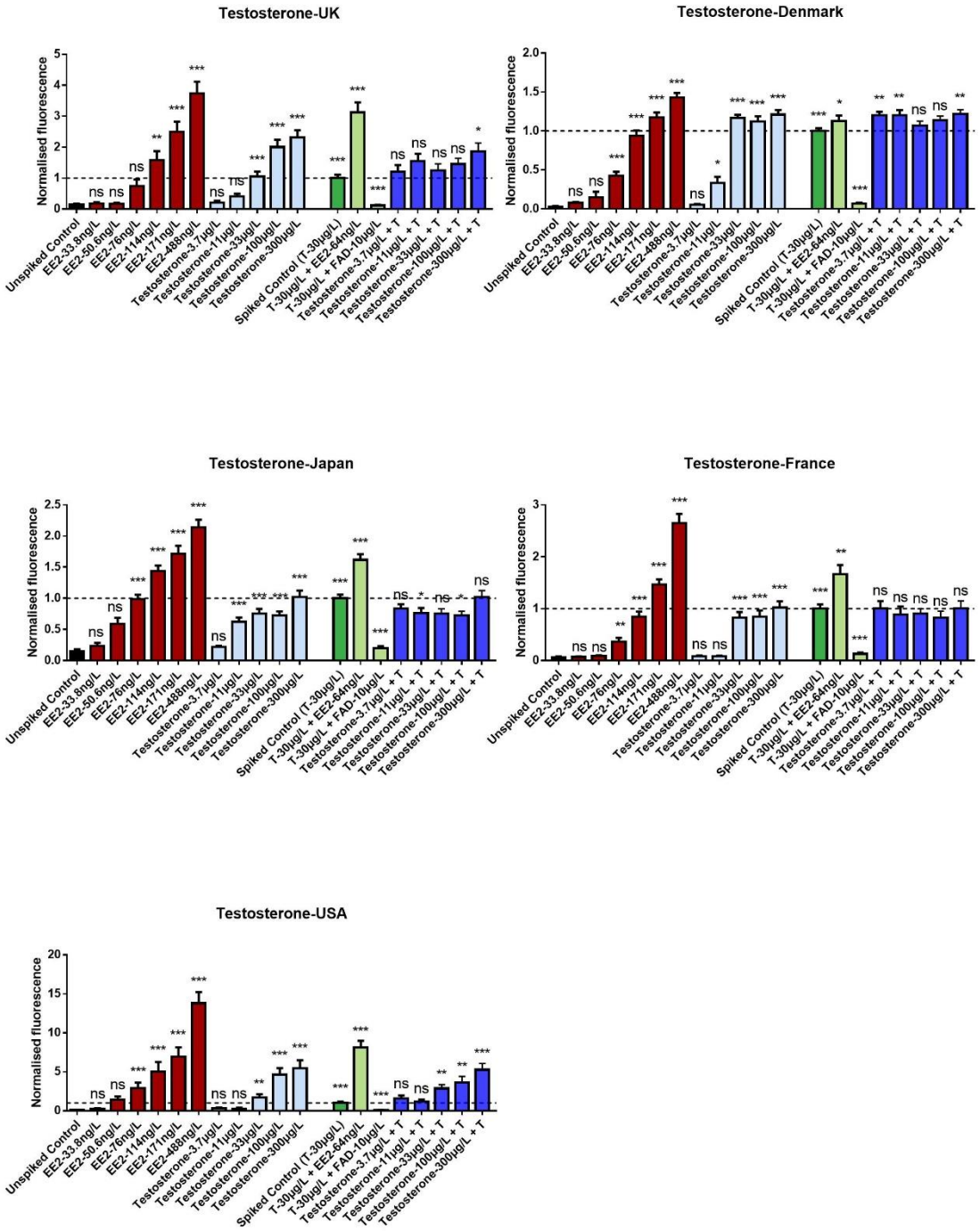


Figure 19: Mean and SEM of measured fluorescence for testosterone.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L control group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
UK	ns	ns	ns	**	***	***	***	***	***
Denmark	ns	ns	***	***	***	***	***	*	***
Japan	ns	ns	***	***	***	***	***	***	***
France	ns	ns	**	***	***	***	***	**	***
USA	ns	ns	***	***	***	***	***	***	***

Laboratory	Testosterone (µg/L)					Testosterone + Testosterone (µg/L)				
	3.7	11	33	100	300	3.7	11	33	100	300
UK	ns	ns	***	***	***	ns	ns	ns	ns	*
Denmark	ns	*	***	***	***	**	**	ns	ns	**
Japan	ns	***	***	***	***	ns	*	ns	*	ns
France	ns	ns	***	***	***	ns	ns	ns	ns	ns
USA	ns	ns	**	***	***	ns	ns	**	**	***

Table 16: Summarised statistical results for the testosterone experiments.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green.

3.5.10.

Triphenyl Phosphate Results

Triphenyl phosphate has been shown to act as an ERα agonist, recruit steroid co-activators 1 and 3 (SRC-1 and SRC-3), activate G-protein coupled ERs and to increase estradiol synthesis and estradiol/testosterone ratio (Ji *et al.*, 2022). As such, it was expected to induce estrogen axis signalling and, therefore, fluorescence in the eleutheroembryos.

Figure 20 shows the results obtained for both laboratories that tested triphenyl phosphate. Both laboratories obtained a concentration-dependent increase for the EE2 controls (Table 17) with a LOEC of 76 ng/L (Germany) and 114 ng/L (France). The spiked mode controls showed the expected statistically significant differences in fluorescence ($P<0.01$) in both laboratories.

Triphenyl phosphate induced a concentration-dependent increase in fluorescence in unspiked mode in both laboratories with a LOEC of 0.59 mg/L. A concentration-dependent increase in fluorescence was also observed in testosterone-spiked mode with the French laboratory obtaining a LOEC of 0.26 mg/L and the German laboratory detecting a

statistically significant (P<0.01) increase in fluorescence at all tested concentrations from 0.18-0.89 mg/L.

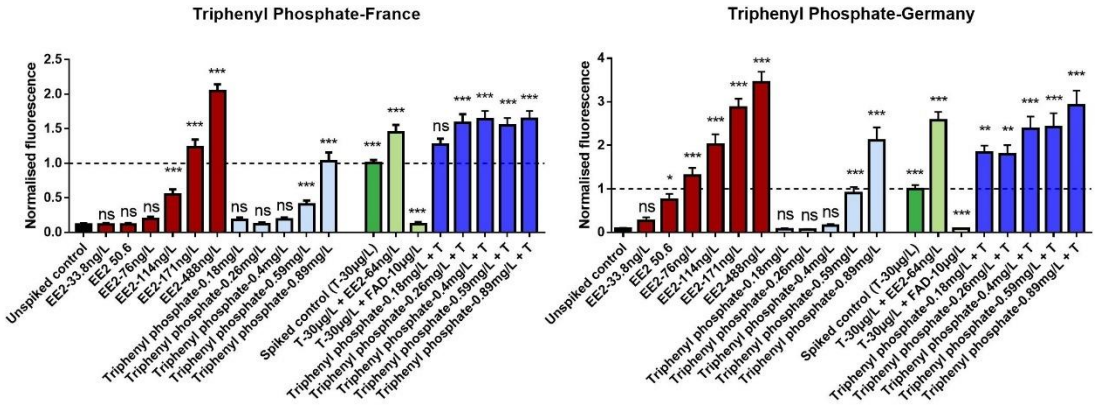


Figure 20: Mean and SEM of measured fluorescence for triphenyl phosphate.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L control group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
France	ns	ns	ns	***	***	***	***	***	***
Germany	ns	*	***	***	***	***	***	***	***

Laboratory	Triphenyl phosphate (mg/L)					Triphenyl phosphate + Testosterone (mg/L)				
	0.18	0.26	0.4	0.59	0.89	0.18	0.26	0.4	0.59	0.89
France	ns	ns	ns	***	***	ns	***	***	***	***
Germany	ns	ns	ns	***	***	**	**	***	***	***

Table 17: Summarised statistical results for the triphenyl phosphate experiments.

Results corresponding to a statistically significant variation of fluorescence (P<0.01) are highlighted in green.

3.6. Results for expected inert chemicals

Amantadine Results

Amantadine was previously used as an antiviral medication to treat influenza caused by type A influenza virus. It is still commonly used to treat Parkinson's disease. It was expected to be inert with respect to estrogen axis signalling.

3.6.1.

The results obtained during the interlaboratory validation exercise are shown in Figure 21. Table 18 shows that both laboratories testing amantadine obtained a concentration-dependent increase for the EE2 controls with a LOEC of 76 ng/L (Germany) and 114 ng/L (France). The spiked mode controls showed the expected statistically significant differences in fluorescence ($P<0.01$) in both laboratories.

No statistically significant deviation in fluorescence was recorded by either of the laboratories when eleutheroembryos were exposed to amantadine in the presence or absence of testosterone.

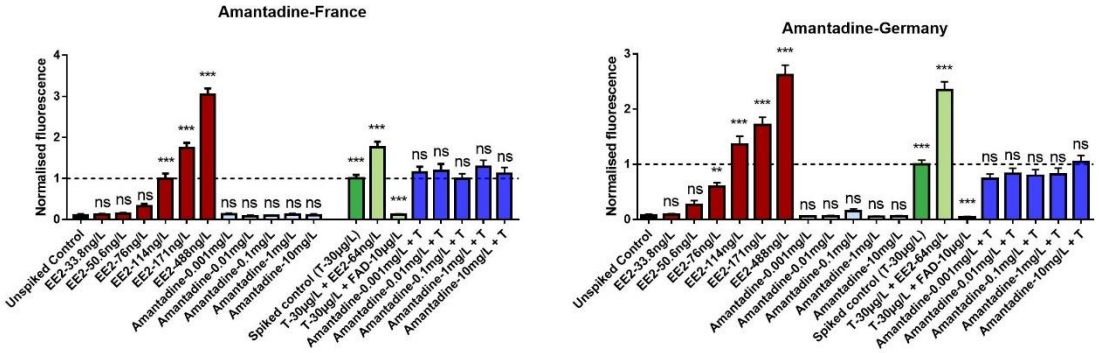


Figure 21: Mean and SEM of measured fluorescence for amantadine.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
France	ns	ns	ns	***	***	***	***	***	***
Germany	ns	ns	**	***	***	***	***	***	***

Laboratory	Amantadine (mg/L)					Amantadine + Testosterone (mg/L)				
	0.001	0.01	0.1	1	10	0.001	0.01	0.1	1	10
France	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Germany	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns

Table 18: Summarised statistical results for the amantadine experiments.

Results corresponding to a statistically significant variation of fluorescence (P<0.01) are highlighted in green.

3.6.2. *Arabinose Results*

Arabinose is a monosaccharide. As such it was expected to be inert with respect to estrogen axis signalling.

The results obtained during the interlaboratory validation exercise are shown in Figure 22. Table 19 shows that both laboratories testing arabinose obtained a concentration-dependent increase for the EE2 controls with a LOEC of 76 ng/L (USA) and 114 ng/L (France). The spiked mode controls showed the expected statistically significant differences in fluorescence (P<0.01) in both laboratories.

No statistically significant deviation (P<0.01) in fluorescence was recorded by either of the laboratories when eleutheroembryos were exposed to arabinose in the presence or absence of testosterone.

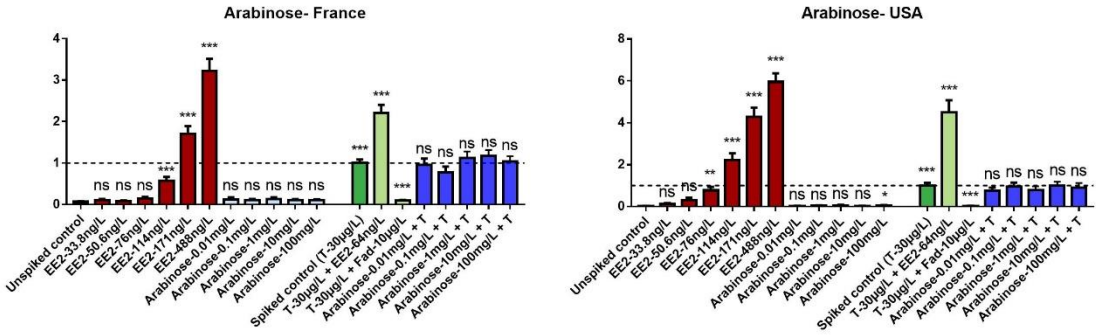


Figure 22: Mean and SEM of measured fluorescence for arabinose.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
France	ns	ns	ns	***	***	***	***	***	***
USA	ns	ns	**	***	***	***	***	***	***

Laboratory	Arabinose (mg/L)					Arabinose + Testosterone (mg/L)				
	0.01	0.1	1	10	100	0.01	0.1	1	10	100
France	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
USA	ns	ns	ns	ns	*	ns	ns	ns	ns	ns

Table 19: Summarised statistical results for the arabinose experiments.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green.

3.6.3. *Atenolol Results*

Atenolol is a pharmaceutical beta blocker regularly used to treat high blood pressure and associated chest pain. As such it was expected to be inert with respect to estrogen axis signalling.

The results obtained during the interlaboratory validation exercise are shown in Figure 23. Table 20 shows that the three laboratories testing arabinose obtained a concentration-dependent increase for the EE2 controls with a LOEC of 76 ng/L (Germany and USA) and

114 ng/L (France). The spiked mode controls showed the expected statistically significant differences in fluorescence (P<0.01) in both laboratories.
No statistically significant deviation (P<0.01) in fluorescence was recorded by either of the laboratories when eleutheroembryos were exposed to atenolol in the presence or absence of testosterone.

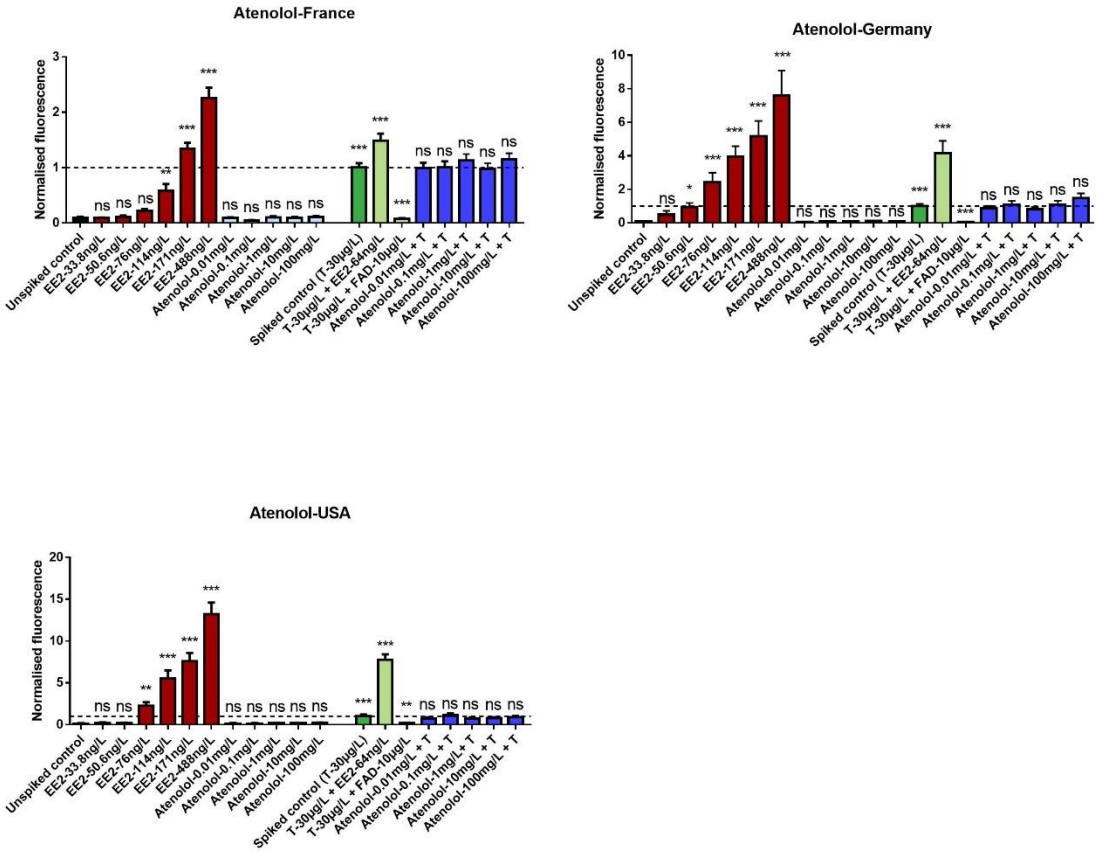


Figure 23: Mean and SEM of measured fluorescence for atenolol.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
France	ns	ns	ns	**	***	***	***	***	***
Germany	ns	*	***	***	***	***	***	***	***
USA	ns	ns	**	***	***	***	***	***	**

Laboratory	Atenolol (mg/L)					Atenolol + Testosterone (mg/L)				
	0.01	0.1	1	10	100	0.01	0.1	1	10	100
France	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Germany	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
USA	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns

Table 20: Summarised statistical results for the atenolol experiments.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green.

3.6.4. *Cefuroxime Results*

Cefuroxime is a cephalosporin antibiotic and was expected to be inert with regards to estrogen axis activity.

Figure 24 shows that a concentration dependent response was observed for EE2 controls in all laboratories. The LOEC for EE2 was between 76 and 114 ng/L (Table 21). All spiked mode controls gave the expected statistically significant differences ($P<0.01$).

No statistically significant deviation in fluorescence ($P<0.01$) was recorded by any of the five laboratories when eleutheroembryos were exposed to cefuroxime in the presence of testosterone. In the absence of testosterone, no statistically significant deviation in fluorescence ($P<0.01$) was observed except in the Japanese laboratory. The results obtained for the Japanese laboratory were highly variable between runs despite the controls performing normally. As the result for this laboratory is a non-monotonic response according to the REACTIV assay decision logic, a repeat experiment is required to determine whether this response is correct.

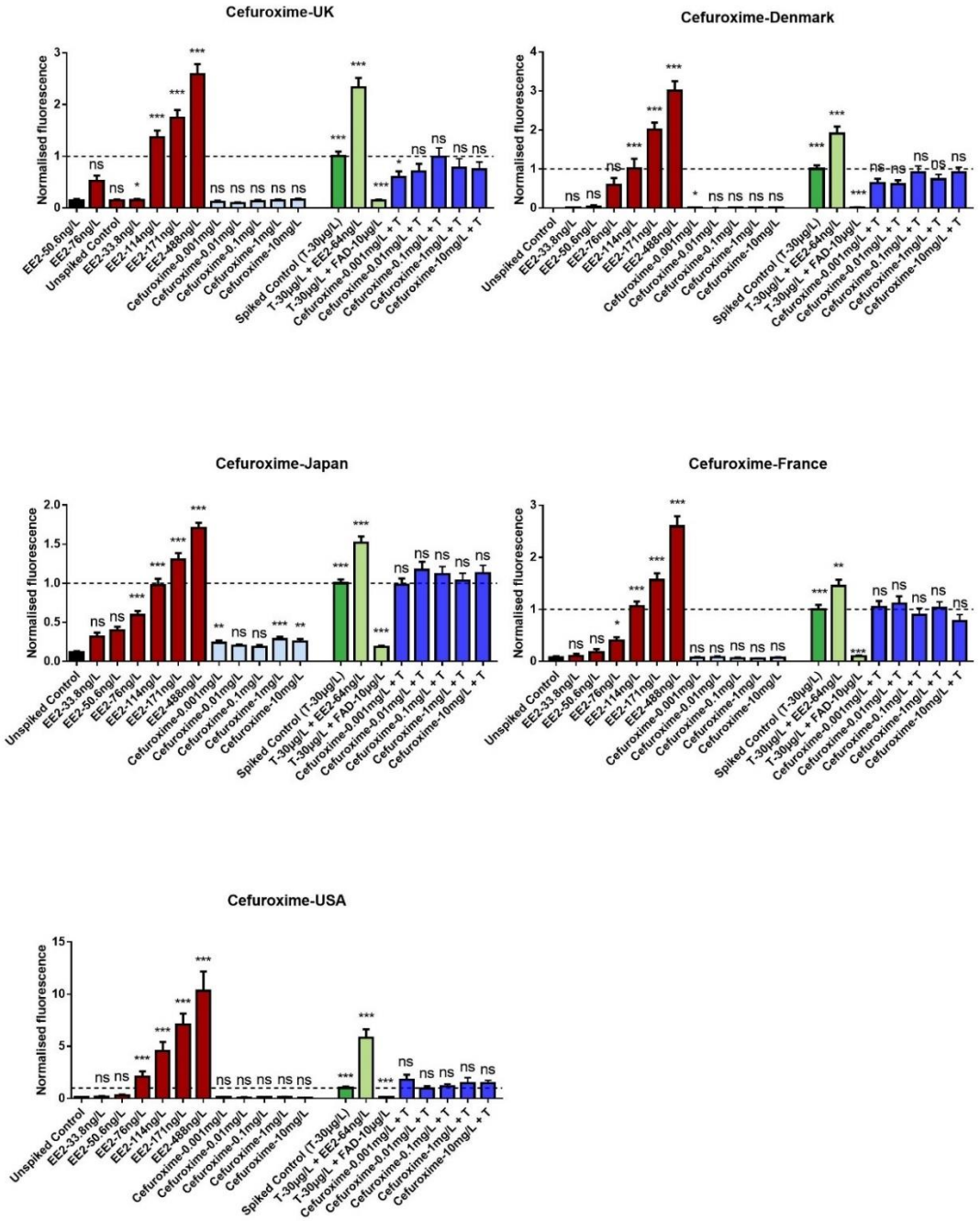


Figure 24: Mean and SEM of measured fluorescence for cefuroxime.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
UK	ns	ns	*	***	***	***	***	***	***
Denmark	ns	ns	ns	***	***	***	***	***	***
Japan	ns	ns	***	***	***	***	***	***	***
France	ns	ns	*	***	***	***	***	**	***
USA	ns	ns	***	***	***	***	***	***	***

Laboratory	Cefuroxime (mg/L)					Cefuroxime + Testosterone (mg/L)				
	0.001	0.01	0.1	1	10	0.001	0.01	0.1	1	10
UK	ns	ns	ns	ns	ns	*	ns	ns	ns	ns
Denmark	*	ns	ns	ns	ns	ns	ns	ns	ns	ns
Japan	**	ns	ns	***	**	ns	ns	ns	ns	ns
France	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
USA	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns

Table 21: Summarised statistical results for the cefuroxime experiments.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green.

3.6.5. *Cromolyn Results*

Cromolyn is a pharmaceutical mast cell stabilizer and was expected to be inert with regards to estrogen axis activity.

Figure 25 shows that a concentration dependent response was observed for EE2 in all four laboratories. The LOEC for the EE2 controls was between 76 and 114 ng/L in the different participating laboratories (Table 22).

As expected, no statistically significant deviation in fluorescence was recorded by any of the four laboratories when eleutheroembryos were exposed to cromolyn in either the presence or absence of testosterone.

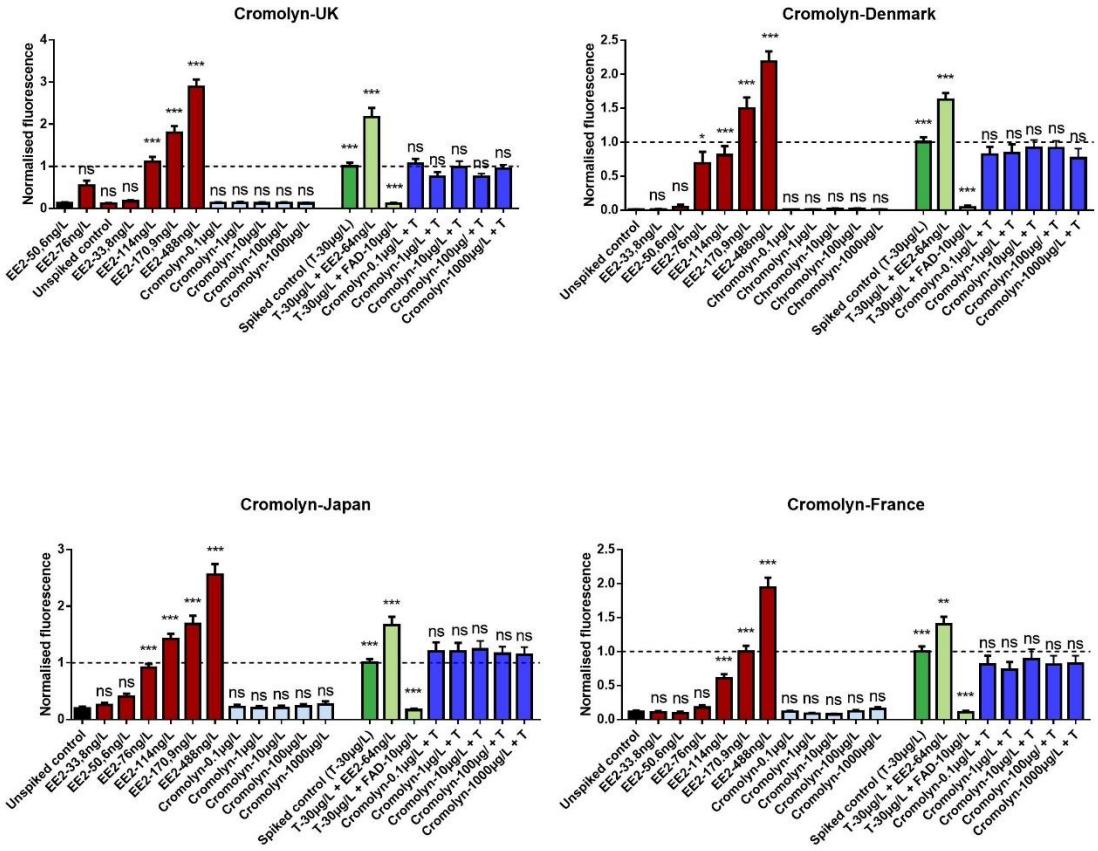


Figure 25: Mean and SEM of measured fluorescence for cromolyn.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
UK	ns	ns	ns	***	***	***	***	***	***
Denmark	ns	ns	*	***	***	***	***	***	***
Japan	ns	ns	***	***	***	***	***	***	***
France	ns	ns	ns	***	***	***	***	**	***

Laboratory	Cromolyn (µg/L)					Cromolyn + Testosterone (µg/L)				
	0.1	1	10	100	1000	0.1	1	10	100	1000
UK	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Denmark	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Japan	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
France	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns

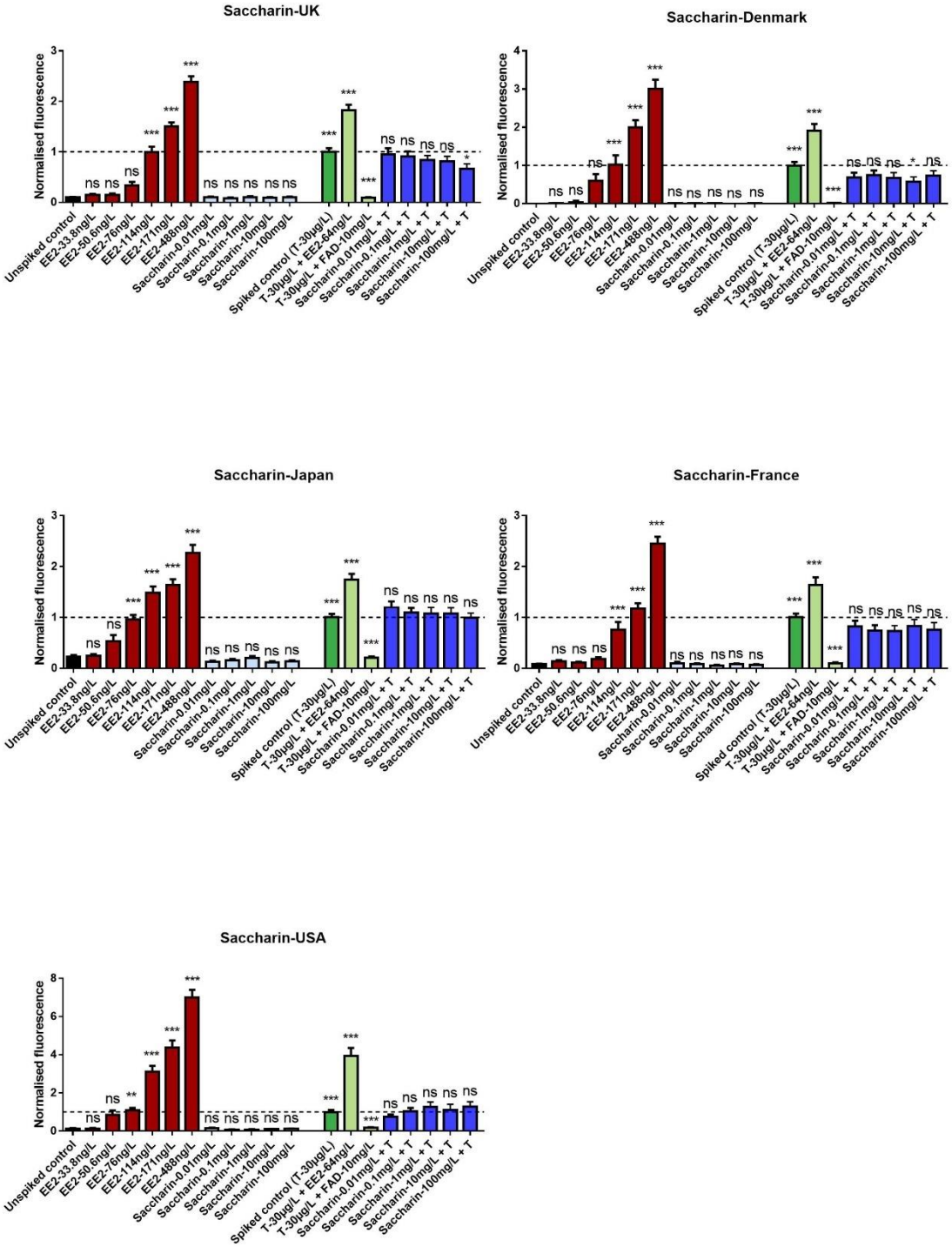
Table 22: Summarised statistical results for the cromolyn experiments.

Results corresponding to a statistically significant variation of fluorescence ($P<0.01$) are highlighted in green.

3.6.6. *Saccharin Results*

Saccharin is an artificial sweetener used in the foods, drinks and medications and was expected to be inert with regards to estrogen axis activity.
Figure 26 shows that a concentration dependent response was observed for EE2 in all five laboratories. The LOEC for the EE2 controls was between 76 and 114 ng/L in the different participating laboratories (Table 23).

As expected no statistically significant deviation in fluorescence ($P<0.01$) was recorded by any of the four laboratories when eleutheroembryos were exposed to saccharin in either the presence or absence of testosterone up to 100 mg/L.



1535

1536 **Figure 26: Mean and SEM of measured fluorescence for saccharin.**

1537 Fluorescence values were normalised to the mean of the testosterone 30 µg/L
1538 group, the value of this group is indicated with a dashed line.

Laboratory	EE2 (ng/L)						Spiked controls		
	34	51	76	114	171	488	T	T+EE2	T+FAD
UK	ns	ns	ns	***	***	***	***	***	***
Denmark	ns	ns	ns	***	***	***	***	***	***
Japan	ns	ns	***	***	***	***	***	***	***
France	ns	ns	ns	***	***	***	***	***	***
USA	ns	ns	**	***	***	***	***	***	***

Laboratory	Saccharin (mg/L)					Saccharin + Testosterone (mg/L)				
	0.01	0.1	1	10	100	0.01	0.1	1	10	100
UK	ns	ns	ns	ns	ns	ns	ns	ns	ns	*
Denmark	ns	ns	ns	ns	ns	ns	ns	ns	*	ns
Japan	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
France	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
USA	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns

Table 23: Summarised statistical results for the saccharin experiments.

Results corresponding to a statistically significant variation of fluorescence (P<0.01) are highlighted in green.

3.7. Chemical Analyses

Selected samples were retained and frozen until quantitative chemical analysis. This was performed in order to ensure that the laboratories were able to accurately prepare the chemicals used in the assay at the correct concentrations and also to measure the loss of the chemicals between initial contact with the eleutheroembryos and the end of the experiment after 24 h. It should be noted that the UK experiments took place end of 2020, and the analysis were carried out during the summer 2022, therefore some degradation of the test chemicals may be expected for some chemicals.

Due to the high number of participating laboratories and test chemicals, in order to reduce costs, analytical verification was performed for test solutions from four of the six laboratories taking part in the OECD validation including the lead laboratory (WatchFrog). The four laboratories selected were CEFAS, UK; WatchFrog, France; DHI, Denmark and Heidelberg University, Germany. This selection includes a government institute (CEFAS), two private companies (DHI and WatchFrog) and a university (Heidelberg University). The choice of laboratories for the analytical chemistry verification was not only based on the range of different types of structure that they represented, but also on the difficulties in recovering chemical samples from the USA and Japan. Several attempts were made to recover the Japanese samples with them being returned to the Japanese laboratory each time.

Analysis of the highest and lowest test concentrations for each of the selected test chemicals was performed to inform on the actual concentrations for the entire concentration range. In order to reduce costs, intermediate concentrations were not tested, as a dilution series was performed and it can be reasonably expected that the intermediate concentrations were in line with the highest and lowest concentrations.
Three runs of the assay were performed for each test chemical. In order to reduce costs, analytical verification was performed for the first run only for chemicals expected to be active in the assay (see section 3.5). As these chemicals were expected to elicit a biological response, the biological response from run one could be compared with runs two and three to inform on the presence of the test chemical. The test solutions at T0 (exposure solution prior to contact with the eleutheroembryos) and at T24 (the same solution after contact with the eleutheroembryos for 24 h) were analysed. The T24 sample informs on loss of the test chemical between the start and end of the experiment.
For the chemicals expected to be inert (see section 3.6) the analytical verification was carried out for the T0 samples for all three runs to ensure that the test chemical was present in the system. This was of particular importance for the inert chemicals as no biological response is expected in the assay. However, in order to reduce costs, the highest and lowest test concentrations were only verified for run one. For runs two and three, only the highest test concentration was verified. In addition, as for the chemicals expected to be active, analytical verification of T24 samples was performed for run one only to inform on loss of the test chemical between daily renewals of the test solutions.

All solutions were stored at -20° and sent to the contract analytical chemistry laboratories of Laboratoires des Pyrénées et des Landes (LPL) in France by the participating laboratories. The detection and quantification limits obtained by LPL are given in Table 24.

Chemical	Detection limit (µg/L)	Quantification limit (µg/ml)
Anastrozole	7.5	25
Bisphenol A	90	300
Dutasteride	0.3	4
Prochloraz	12	40
Cefuroxime	6	20
Cromolyn	15	50
Saccharin	1	3

Table 24: Detection and quantification limits for the chemical analysis.

3.7.1. ***Anastrozole***

The measured concentrations of anastrozole were very close to the nominal values prior to exposure for both the lowest tested concentration of 180 µg/L (105% nominal) and also the highest tested concentration of 2900 µg/L (103% nominal) from the original concentration

range as shown in Table 25. The same table also shows the analytical results for two of the laboratories that tested the lower concentration range. Unfortunately, the lowest concentration tested in the lower concentration range (0.032 µg/L) was below the detection limit (7.5 µg/L). However, the measured concentrations for the highest concentration tested in the lower concentration range (20 µg/L) were close to nominal (80-95%).

Following 24 h of exposure the measured concentrations ranged from 103-105% of nominal for the lowest and highest tested concentrations in the initial concentration range (Table 26). For the lower concentration range, the measured concentrations range from 95-115% of nominal for the highest tested concentration. It should be noted that these concentrations are higher than the initial concentrations, but also that this concentration (20 µg/L) is just below the quantification limit for this chemical (25 µg/L) and, therefore, some error in the measurements can be expected. It should be noted that as with the samples taken prior to exposure, the lowest test concentration in this lower range was below the analytical detection limit. The lead laboratory (France) by error sent two samples of the highest test concentration rather than the highest and lowest, both of these samples gave the same concentration (23 µg/L measured for 20 µg/L nominal).

Overall, these results indicate that there was no appreciable loss of test chemical during the REACTIV assay.

Laboratory	Run	Nominal Anastrozole Concentration (µg/L)	Measured Anastrozole Concentration (µg/L)	Measured/Nominal (%)
UK	R1	2900	3050	105
	R1	180	185	103
Denmark	R1	20	16	80
	R1	0.032	<7.5	-
France	R1	20	19	95
	R1	0.032	<7.5	-

Table 25 : Nominal and measured concentrations for anastrozole before exposure.

Laboratory	Run	Nominal Anastrozole Concentration (µg/L)	Measured Anastrozole Concentration (µg/L)	Measured/Nominal (%)
UK	R1	2900	3000	103
	R1	180	185	103
Denmark	R1	20	18	90
	R1	0.032	<7.5	-
France	R1	20	23/23	115/115
	R1	0.032	-	-

Table 26 : Nominal and measured concentrations for anastrozole after 24 h of exposure.

Bisphenol A

3.7.2

The measured concentrations of bisphenol A were very close to the nominal values both before and after exposure (91-105%) as shown in Table 27 with the exception of lowest concentration in the Danish laboratory (30% of nominal) and the highest concentration in the German laboratory (73% of nominal). Despite this the German laboratory obtained a LOEC of 2 mg/L for bisphenol A, which was only obtained by one other laboratory. However, the analytical result for the Danish laboratory, which suggests loss of test chemical during preparation of the dilution series, could explain the fact that this laboratory obtained the highest LOEC value for bisphenol A (4 mg/L).

Following 24 h of exposure the measured concentrations ranged from 73-91% of nominal (Table 28). The exception was the Danish laboratory which obtained 26% of nominal for the lowest test concentration, although as previously discussed, the analytical verification indicated that this solution was prepared at 30% of nominal prior to exposure. Taken together, these results indicate that there was no major loss of test chemical between during the exposure period.

Laboratory	Run	Nominal Bisphenol A Concentration (mg/L)	Measured Bisphenol A Concentration (mg/L)	Measured/Nominal (%)
UK	R1	5	4.62	92
	R1	1	0.91	91
Denmark	R1	5	4.72	94
	R1	1	0.30	30
France	R1	5	4.96	99
	R1	1	1.05	105
Germany	R1	5	3.65	73
	R1	1	0.98	98

Table 27 : Nominal and measured concentrations for bisphenol A before exposure.

Laboratory	Run	Nominal Bisphenol A Concentration (mg/L)	Measured Bisphenol A Concentration (mg/L)	Measured/Nominal (%)
UK	R1	5	3.65	73
	R1	1	0.74	74
Denmark	R1	5	4.36	87
	R1	1	0.26	26
France	R1	5	4.48	90
	R1	1	0.91	91
Germany	R1	5	4.19	84
	R1	1	0.84	84

Table 28 : Nominal and measured concentrations for bisphenol A after 24 h of exposure.

3.7.3. *Dutasteride*

The measured concentrations of dutasteride were very low compared to nominal values prior to exposure (8-36% of nominal) as shown in Table 29. These particularly low values, across all three laboratories for which an analytical check was performed, could be due to a problem with the solubility of the test chemical. It is less likely that they are due to loss of test chemical during the preparation of the dilution series as the recovery in terms of percentage nominal was similar for the UK and Denmark for both the highest and lowest test concentrations (18 vs 17% and 20 vs 36%, respectively).

Table 30 shows that there appears to be further loss of dutasteride during the 24 h of exposure. The lowest concentration (10 µg/L) was reduced from 1.7, 3.6 and 0.8 µg/L in the British, Danish and French laboratories, respectively, to undetectable in all three laboratories (<0.3 µg/L). The highest test concentration (1000 µg/L) was reduced from 18, 20 and 25 to 8, 10 and 2.2%, respectively.

Laboratory	Run	Nominal Dutasteride Concentration (µg/L)	Measured Dutasteride Concentration (µg/L)	Measured/Nominal (%)
UK	R1	1000	179	18
	R1	10	1.7	17
Denmark	R1	1000	199	20
	R1	10	3.6	36
France	R1	1000	246	25
	R1	10	0.8	8

Table 29 : Nominal and measured concentrations for dutasteride before exposure.

Laboratory	Run	Nominal Dutasteride Concentration (µg/L)	Measured Dutasteride Concentration (µg/L)	Measured/Nominal (%)
UK	R1	1000	79.5	8.0
	R1	10	< 0.3	-
Denmark	R1	1000	101.4	10
	R1	10	< 0.3	-
France	R1	1000	21.5	2.2
	R1	10	< 0.3	-

Table 30 : Nominal and measured concentrations for dutasteride after 24 h of exposure.

3.7.4.

Prochloraz

The measured concentrations of prochloraz were reasonably close to the nominal values prior to exposure for both the lowest tested concentration of 63 µg/L (70-100% nominal except the UK, 33% nominal) and also the highest tested concentration of 1000 µg/L (76-99% nominal except the UK, 65% nominal) as shown in Table 31.
Following 24 h of exposure the measured concentrations range from 25-65% of nominal for the lowest tested concentration and from 48-72% of nominal for the highest tested concentration (Table 32). This indicates that there was a fairly homogenous loss of test chemical during the exposure period of 22-41% compared to the measured concentrations prior to exposure, regardless of the laboratory or concentration.

Laboratory	Run	Nominal Prochloraz Concentration (µg/L)	Measured Prochloraz Concentration (µg/L)	Measured/Nominal (%)
UK	R1	1000	651	65
	R1	63	21	33
Denmark	R1	1000	791	76
	R1	63	63	70
France	R1	1000	993	99
	R1	63	50	79
Germany	R1	1000	791	79
	R1	63	63	100

Table 31 : Nominal and measured concentrations for prochloraz before exposure.

Laboratory	Run	Nominal Prochloraz Concentration (µg/L)	Measured Prochloraz Concentration (µg/L)	Measured/Nominal (%)
UK	R1	1000	479	48
	R1	63	16	25
Denmark	R1	1000	592	59
	R1	63	26	41
France	R1	1000	724	72
	R1	63	32	51
Germany	R1	1000	562	56
	R1	63	41	65

Table 32 : Nominal and measured concentrations for prochloraz after 24 h of exposure.

3.7.5.

Cefuroxime

The measured concentrations of cefuroxime were close to the nominal values prior to exposure for the highest tested concentration of 10 mg/L (84-96% nominal; Table 33). The concentrations of cefuroxime could not be measured for the lowest concentration groups as the nominal concentration (0.001 mg/L) was below the detection limit 0.006 mg/L). Following 24 h of exposure the samples from the highest test concentration showed very little difference to their initial values. This indicates that there was no appreciable loss of test chemical between media renewals (Table 34). The only exception was for the highest concentration from the UK, which for an unknown reason dropped from 88% of nominal to 40% of nominal.

Laboratory	Run	Nominal Cefuroxime Concentration (mg/L)	Measured Cefuroxime Concentration (mg/L)	Measured/Nominal (%)
UK	R1	10	8.75	88
	R1	0.001	<0.006	-
	R2	10	8.59	86
	R3	10	8.82	88
Denmark	R1	10	9.16	92
	R1	0.001	<0.006	-
	R2	10	8.44	84
	R3	10	8.64	86
France	R1	10	9.26	93
	R1	0.001	<0.006	-
	R2	10	9.64	96
	R3	10	9.26	93

Table 33 : Nominal and measured concentrations for cefuroxime before exposure.

Laboratory	Run	Nominal Cefuroxime Concentration (mg/L)	Measured Cefuroxime Concentration (mg/L)	Measured/Nominal (%)
UK	R1	10	4.03	40
	R1	0.001	<0.006	-
Denmark	R1	10	8.83	88
	R1	0.001	<0.006	-
France	R1	10	7.89	79
	R1	0.001	<0.006	-

Table 34 : Nominal and measured concentrations for cefuroxime after 24 h of exposure.

3.7.6.

Cromolyn

The measured concentrations of cromolyn were very close to the nominal values prior to exposure for the highest tested concentration of 1 mg/L (86-113% nominal; Table 35). Unfortunately, the lowest concentration of cromolyn could not be as the nominal concentration (0.0001 mg/L) was below the detection limit 0.015 mg/L).
Following 24 h of exposure the samples from the highest test concentration showed very little difference to their initial values. This indicates that there was no appreciable loss of test chemical between media renewals (Table 36). Concentrations were measured as 83-94% of nominal.

Laboratory	Run	Nominal Cromolyn Concentration (mg/L)	Measured Cromolyn Concentration (mg/L)	Measured/Nominal (%)
UK	R1	1	1.029	103
	R1	0.0001	<0.015	-
	R2	1	1.130	113
	R3	1	1.066	107
Denmark	R1	1	0.931	93
	R1	0.0001	<0.015	-
	R2	1	1.018	102
	R3	1	0.962	96
France	R1	1	0.857	86
	R1	0.0001	<0.015	-
	R2	1	0.894	89
	R3	1	0.925	93

Table 35 : Nominal and measured concentrations for cromolyn before exposure.

Laboratory	Run	Nominal Cefuroxime Concentration (mg/L)	Measured Cefuroxime Concentration (mg/L)	Measured/Nominal (%)
UK	R1	1	0.914	91
	R1	0.0001	<0.015	-
Denmark	R1	1	0.936	94
	R1	0.0001	<0.015	-
France	R1	1	0.830	83
	R1	0.0001	<0.015	-

Table 36 : Nominal and measured concentrations for cromolyn after 24 h of exposure.

3.7.7.

Saccharin

There was some variability in the measured concentrations of saccharin. Prior to exposure the highest concentration (100 mg/L) gave measured values ranging from 42-85% of nominal (Table 37). All laboratories tested at least one run with a measured value over 80% of nominal.

Following 24 h of exposure the samples from the highest test concentration show very little difference to their initial values. This indicates that there was no appreciable loss of test chemical between media renewals (Table 38).

Laboratory	Run	Nominal Saccharin Concentration (mg/L)	Measured Saccharin Concentration (mg/L)	Measured/Nominal (%)
UK	R1	100	82.698	83
	R1	0.01	0.00796	80
	R2	100	73.990	74
	R3	100	80.393	80
Denmark	R1	100	69.678	70
	R1	0.01	0.00561	56
	R2	100	69.635	70
	R3	100	73.062	73
France	R1	100	80.505	81
	R1	0.01	0.00733	73
	R2	100	84.668	85
	R3	100	42.226	42

Table 37 : Nominal and measured concentrations for saccharin before exposure.

Laboratory	Run	Nominal Saccharin Concentration (mg/L)	Measured Saccharin Concentration (mg/L)	Measured/Nominal (%)
UK	R1	100	73.063	73
	R1	0.01	0.00758	76
Denmark	R1	100	66.665	67
	R1	0.01	0.00540	54
France	R1	100	76.446	76
	R1	0.01	0.00659	66

Table 38 : Nominal and measured concentrations for saccharin after 24 h of exposure.

3.8. Results for chemicals tested uniquely in the lead laboratory

Having demonstrated the reliability of the REACTIV assay to generate reproducible results across the participating partner laboratories, an additional chemical was

tested with the REACTIV assay in the lead laboratory only. Unfortunately, due to time and financial constraints, these additional chemicals could not be tested in all participating laboratories.

Atrazine

The herbicide atrazine has been shown to increase the expression of aromatase *in vitro* (Higley *et al.*, 2010). As this is different to other modes of action included in this study, it was, therefore, decided that it would be of interest to determine the response of the REACTIV assay to this chemical.

Due to the ability of atrazine to increase aromatase expression, it would be expected to increase estrogen axis activity by increasing the conversion of androgens to estrogens. No significant differences in fluorescence were observed for any of the tested concentrations of atrazine in unspiked mode.

As expected, an increase in fluorescence was observed in spiked mode (Figure 27: Mean and SEM of measured fluorescence for atrazine tested in the lead laboratory.Figure 27). This increase was observed for all concentrations tested, but was statistically significant (P<0.01) for the highest concentration tested (10 mg/L).

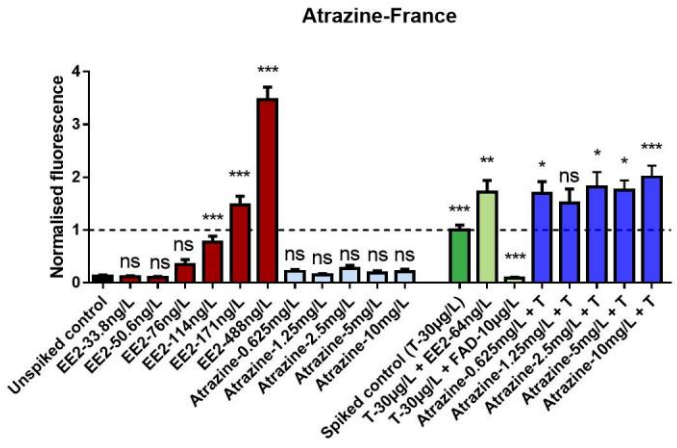


Figure 27: Mean and SEM of measured fluorescence for atrazine tested in the lead laboratory.

Fluorescence values were normalised to the mean of the testosterone 30 µg/L group, the value of this group is indicated with a dashed line. Results corresponding to a variation of fluorescence (P<0.01) are considered as significant.

3.9. Discussion

The data generated during the interlaboratory validation exercise of the REACTIV assay demonstrated that the assay is robust and produced reliable and reproducible data across six laboratories. The six laboratories used a range of different fluorescence imaging systems to read the experiments and either bred the *chgh-gfp* eleutheroembryos in house (France, Japan and the USA) or raised them from shipped embryos (Denmark, Germany and the UK). Despite this, all six laboratories were able to perform the assay with the expected sensitivity. All expected inert test chemicals were identified as inert. Likewise, for the test chemicals expected to be active, they were all identified as active with a small number of exceptions. These were prochloraz in one out of six laboratories, tamoxifen in one out of five laboratories and dutasteride in two out of four laboratories. The two laboratories failing to identify dutasteride as active (USA and UK) observed the expected increase in fluorescence in spiked mode, which did not pass the threshold for statistical significance. Results for individual chemicals are discussed in detail below.

Calibration

3.9.1.

The calibration exercise was not performed by the lead laboratory (WatchFrog, France) as they had previously calibrated their image acquisition parameters for use with the REACTIV assay. All five naïve laboratories performed the calibration exercise and obtained data showing a concentration-response relationship for EE2 with an acceptable sensitivity (51-114 ng/L). All five naïve laboratories obtained the required level of statistical significance for the spiked mode controls.

Based on these data, all five naïve laboratories were advised to proceed to testing of the expected active and inert chemicals using the image acquisition settings determined during the calibration exercise.

3.9.2.

Anastrozole

The initial concentration range selected for the pharmaceutical aromatase inhibitor anastrozole was selected as identical to that used for validation of the RADAR assay. The REACTIV assay clearly has a higher sensitivity for modulations in aromatase activity as this initial concentration range, tested in the British and French laboratories, fully inhibited the fluorescent signal at all concentrations tested. A lower concentration range was then selected, which was retested in the lead (French) laboratory and all subsequent participating laboratories.

As expected, all participating laboratories observed a lack of effect in unspiked mode and a clear concentration dependent decrease in fluorescence in spiked mode. The LOEC was also extremely reproducible, with 0.8 µg/L in three laboratories and 4 µg/L in the fourth laboratory. Analytical verification showed that the measured concentrations were very close to nominal.

3.9.3.

Bisphenol A

The weak ER agonist BPA was expected to induce fluorescence in both spiked and unspiked mode. This was the case in all six participating laboratories, except the Danish laboratory which failed to observe a statistically significant increase in unspiked mode. Despite this, all six laboratories identified BPA as active with very similar LOEC values of 4-5 mg/L in unspiked mode (except Denmark) and 2-4 mg/L in spiked mode. The lower sensitivity in the Danish laboratory may be due to dilution errors as the lowest tested concentration (1 mg/L) was found to be 30% of nominal prior to exposure.

Dutasteride

The pharmaceutical 5α-reductase inhibitor gave mixed results in the different laboratories. An increased fluorescence in spiked mode was observed in two laboratories (France and Denmark), which is in line with its mode of action which blocks conversion of testosterone to the non-aromatisable androgen dihydrotestosterone, therefore, increasing the pool of testosterone available for conversion to estradiol. Two laboratories failed to identify dutasteride as active (USA and UK). Both of these laboratories observed the expected increase in fluorescence in spiked mode, but it did not pass the threshold for statistical significance ($P<0.01$) despite a close statistical result for the American laboratory that would normally be considered as statistically significant ($P=0.02$). A non-monotonic concentration-response was observed in the Japanese laboratory and after consideration of the individual runs as required by the decision logic chart, it was decided that the three runs should be repeated. The problems encountered in detecting a statistically significant effect from dutasteride are likely to be caused by the extremely low test concentrations which were measured as 8-36% of nominal prior to exposure.

17β-Estradiol

All four laboratories that tested the natural estrogen 17β-estradiol correctly identified it as pro-estrogenic in both spiked and unspiked mode, with a similar range of LOECs in unspiked mode (135-540 ng/L) and in spiked mode (68-270 ng/L).

Estrone

As with the ER agonist 17β-estradiol, estrone was identified as pro-estrogenic in both spiked and unspiked mode in both participating laboratories (France and Germany). The LOECs in both laboratories were identical, 5 µg/L in unspiked mode and 1 µg/L in spiked mode.

Fadrozole

Fadrozole, like anastrozole is a pharmaceutical aromatase inhibitor. It is also included at a single concentration (10 µg/L) as a spiked mode control in all experiments. As with

anastrozole, the initial concentration range that was tested by France and the UK, which was identical to that used in the RADAR assay validation caused a seemingly total inhibition of the generation of GFP. A lower concentration range was then selected which was tested by France, Germany and the USA. This lower range resulted in a concentration-dependent response with a LOEC of 0.4-10 µg/L, demonstrating the higher sensitivity of the REACTIV assay for modulation in aromatase activity compared to the RADAR assay.

Prochloraz

Prochloraz has been demonstrated to inhibit transcription of aromatase (Higley *et al.*, 2010) and was tested in all six participating laboratories. As expected, in the absence of testosterone (unspiked mode), none of the laboratories detected a change in fluorescence. In the presence of testosterone (spiked mode), five of the laboratories identified an inhibition in GFP production, with only the UK failing to identify this inhibition. Interestingly, analytical verification of the exposure solutions showed that the lowest measured concentrations were in the solutions supplied by the British laboratory (65% nominal for the highest test concentration and 33% nominal for the lowest test concentration).

3.9.9. *Tamoxifen*

Tamoxifen was correctly identified as active on the estrogen axis in four of the five laboratories that tested it. Tamoxifen is a selective estrogen response modulator (SERM) and as such can exert pro- or anti-estrogenic activity. The activity observed for tamoxifen in the REACTIV assay was pro-estrogenic in unspiked mode. Only one laboratory failed to identify it as active (Denmark) and a clear concentration-dependent pro-estrogenic effect was observed in unspiked mode in the Danish laboratory, however, due to a high level of variability, this result did not reach statistical significance.

3.9.10. *Testosterone*

All five laboratories that tested testosterone found it to be pro-estrogenic via metabolic conversion by aromatase enzyme to estradiol. The dependence of the pro-estrogenic effect on enzymatic conversion by aromatase was confirmed by the testosterone + EE2 control which is included in all experiments. The LOEC values were very close between the five laboratories (11-33 µg/L). Interestingly, the ability of aromatase to transform testosterone into estradiol appeared to be saturated at around 33 µg/L in most laboratories, explaining why no clear effect in spiked mode was observed in most laboratories as these groups are exposed to 30 µg/L of testosterone from the spike alone. It is unknown why the results from the UK and USA seem to indicate saturation of aromatase enzyme activity at higher concentrations of testosterone.

Triphenyl phosphate

Triphenyl phosphate was tested in two laboratories (France and Germany). It is known act via several modes of action, as an ERα agonist, by recruiting steroid co-activators 1 and 3 (SRC-1 and SRC-3), activating G-protein coupled ERs and increasing estradiol synthesis and estradiol/testosterone ratio (Ji et al., 2022). Both participating laboratories identified triphenyl phosphate as active in both spiked and unspiked mode with a LOEC of 0.59 mg/L in both laboratories in unspiked mode and 0.18-0.26 mg/L in spiked mode.

3.9.11.

Amantadine (expected to be inert)

3.9.12.

Amantadine was tested in two laboratories (France and Germany) and showed no activity in either spiked or unspiked mode in either laboratory.

Arabinose (expected to be inert)

3.9.13.

Arabinose was tested in two laboratories (France and the USA) and showed no activity in either spiked or unspiked mode in either laboratory ($P<0.01$).

3.9.14.

Atenolol (expected to be inert)

Amantadine was tested in three laboratories (France, Germany and the USA) and showed no activity in either spiked or unspiked mode in any of the three laboratories.

3.9.15.

Cefuroxime (expected to be inert)

Cefuroxime was selected as an inert chemical for the validation of the REACTIV assay as it is biologically active, but to date, it is not known to be active on the estrogen axis. Four of the five laboratories that tested cefuroxime found it to be inert in both spiked and unspiked mode, with no changes in fluorescence level that were statistically significant ($P<0.01$). Only the Japanese laboratory identified some statistically significant differences in fluorescence in the lowest and two highest concentration groups in unspiked mode. This is believed to be due to a high level of variability between individual runs resulting in a low estimation if the negative control group. As the result for this laboratory is a non-monotonic response according to the REACTIV assay decision logic, a repeat experiment is required to determine whether this response is correct.

Analytical verification was performed for cefuroxime and the measured concentrations of cefuroxime were very close to the nominal values prior to exposure for the highest tested concentration of 10 mg/L (84-96% nominal). Unfortunately, the detection limit did not allow measurement of the lowest test concentration.

Cromolyn (expected to be inert)

3.9.16. Cromolyn was selected as the third inert chemical for the interlaboratory validation based on the same criterion as the other inert chemicals, a lack of published data showing estrogen axis activity, but also because its coloured. No active concentrations were detected by any of the four laboratories, cromolyn is, therefore, considered as estrogen axis inactive by the REACTIV assay.

Analytical verification confirmed that the measured concentrations of cromolyn were very close to nominal for the highest tested concentration (86-113% of nominal prior to exposure and 83-94% of nominal at the end of the exposure period). Unfortunately, the lowest test concentration could not be detected as it was below the analytical detection limit.

Saccharin (expected to be inert)

3.9.17. Saccharin was tested in five laboratories and showed no activity in either spiked or unspiked mode in any laboratory ($P<0.01$). Analytical verification was performed for this test substance. The highest nominal concentration (100 mg/L) was measured as 42-85% of nominal across the runs performed by the five laboratories, however, all laboratories tested at least one run with a measured value over 80% of nominal.

Chemical analysis

3.9.18. The results of the chemical analysis are discussed on a chemical-by-chemical basis above, however, a general comment should be made here. A major technical problem with the chemical analysis was the quantification and detection limits which were in some cases higher or considerably higher than the lowest test concentration. This made interpretation of the results for the lowest test concentration difficult in some cases and entirely impossible in others.

Chemical tested uniquely in the lead laboratory

Following the confirmation of reproducible classification of chemicals as pro-estrogenic, anti-estrogenic or inert across all participating laboratories, one additional chemical was tested in the lead laboratory only.

Atrazine is known to increase aromatase expression and displayed the expected increase in estrogen axis activity at 10 mg/L. Statistically significant increases in fluorescence were also observed from 0.625 mg/L but were not considered as they did not pass the P<0.01 threshold.

Test failure rate

During the interlaboratory validation exercise, a total of 246 experimental runs were performed. There was an 8.5% failure rate across all laboratories, including the five laboratories which had not previously performed the assay. Interestingly, there was also an 8.5% failure rate in the lead laboratory which had previously experience with the assay, although all experiments were performed by students with no previous experience with the assay, organism or fluorescence microscopy.

3.10. Conclusions

The REACTIV assay interlaboratory validation exercise demonstrated that the assay provides the expected results with the chemicals tested and is reproducible across laboratories (Table 39). Overall, the data generated in the six laboratories matched the expected response profiles and the test chemicals were correctly classified as estrogen axis active or inactive in each laboratory. The column labelled “Expected” indicates the expected result based on published data concerning the mode of action of the test chemical. The column labelled “ER model” shows the predicted activity of each test chemical based on computational modelling of the results of 18 *in vitro* high throughput screening assays (Browne *et al.*, 2015). When the results obtained in this study are compared with the ER model, the ER model correctly identified the inert chemicals. However, as expected, the ER model only identified the active chemicals acting as ER agonists or antagonists. Although, it should be noted that it failed to identify triphenyl phosphate as active despite one of its modes of action being agonism of ERα. As expected, the ER model failed to identify any of the test chemicals acting via alterations in aromatase or 5α-reductase activity.

		France	UK	Denmark	Japan	USA	Germany	ER model
EE2	strong ER agonist	active	active	active	active	active	active	active
Anastrozole	aromatase enzyme inhibitor	active	active	active	active	active		inert
BPA	weak ER agonist	active	active	active	active	active	active	active
Dutasteride	5α-reductase inhibitor	active	inert	active	N-M	inert		inert*
17β-estradiol	strong ER agonist	active	active	active	active	active	active	active
Estrone	strong ER agonist	active					active	active
Fadrozole	aromatase enzyme inhibitor	active	active			active	active	inert
Prochloraz	aromatase transcription inhibitor	active	inert	active	active	active	active	inert
Tamoxifen	SERM	active	active	inert	active	active		active
Testosterone	aromatisable androgen	active	active	active	active	active		
Triphenyl phosphate	multiple (see section 3.5.10)	active					active	inert
Atrazine	induced aromatase expression	active						inert
Amantadine	antiviral, antiparkinsonian	inert					inert	
Arabinose	monosaccharide	inert				inert		inert
Atenolol	beta blocker	inert				inert	inert	inert
Cefuroxime	cephalosporin antibiotic	inert	inert	inert	N-M	inert		
Cromolyn	mast cell stabilizer	inert	inert	inert	inert			
Saccharin	artificial sweetener	inert	inert	inert	inert	inert		inert

Table 39 : Summary of expected and experimental results for the REACTIV assay from the validation exercise.

* indicates that dutasteride was not evaluated with the ER model, but finasteride which has the same mode of action was evaluated and predicted to be inert. N-M indicates that a non-monotonic concentration-response profile was obtained and according to the decision logic, the test should be repeated.

The validation exercise successfully evaluated not only the REACTIV assay itself, but the optimisation and validation steps required to set up the assay in a naïve laboratory using image acquisition equipment which has not previously been employed for this assay.
Once the image acquisition and treatment steps had been optimised, the experimental protocol for the assay as well as the data treatment, statistical approach and a decision logic for classifying the test chemical as estrogen axis active or inactive were evaluated and validated. The protocol was successfully transferred to five laboratories from different OECD countries in three different continents.
In all cases, once the experimental protocol, data analysis procedure and decision logic had been applied, the expected results were obtained for the test chemicals expected to be inert. Likewise, for the test chemicals expected to be active, they were all identified as active except prochloraz in one out of six laboratories, tamoxifen in one out of five laboratories and dutasteride in two out of four laboratories. The two laboratories failing to identify dutasteride as active (USA and UK) observed the expected increase in fluorescence in spiked mode, which did not pass the threshold for statistical significance ($P<0.01$) despite a close statistical result for the American laboratory that would normally be considered as statistically significant ($P=0.02$). Chemical analysis showed that the test concentrations were very low compared to nominal prior to exposure (8-36%).
It should be noted that two experiments from the Japanese laboratory (cefuroxime and dutasteride) displayed a non-monotonic concentration-response profile and following

inspection of the individual runs as per the decision logic (see section 3.3.3), it was decided that the experiments should be repeated.

The REACTIV assay was shown to be sensitive to a range of different modes of estrogen axis activity including: ER agonism, ER antagonism, inhibition of aromatase enzyme at the protein level, inhibition or activation of aromatase expression, 5α-reductase enzyme inhibition and chemicals requiring metabolic activation.

4. REFERENCES

Aida, T. (1921) On the inheritance of color in a fresh-water fish, *Aplocheilus latipes* Temmick and Schlegel, with special reference to sex-linked inheritance. *Genetics*, **6**, 554–573.

Ankley, G.T. *et al.* (2002) Evaluation of the aromatase inhibitor fadrozole in a short-term reproduction assay with the fathead minnow (*Pimephales promelas*). *Toxicol. Sci.*, **67**, 121–130.

Ankley, G.T. and Gray, L.E. (2013) Cross-species conservation of endocrine pathways: a critical analysis of tier 1 fish and rat screening assays with 12 model chemicals. *Environ. Toxicol. Chem.*, **32**, 1084–1087.

Braasch, I. *et al.* (2009) Pigmentation pathway evolution after whole-genome duplication in fish. *Genome Biol. Evol.*, **1**, 479–493.

Browne, P. *et al.* (2015) Screening chemicals for estrogen receptor bioactivity using a computational model. *Environ. Sci. Technol.*, **49**, 8804–8814.

Chhipa, R.R. *et al.* (2013) The direct inhibitory effect of dutasteride or finasteride on androgen receptor activity is cell line specific. *Prostate*, **73**, 1483–1494.

Cui, J. *et al.* (2009) Homology-modeled ligand-binding domains of medaka estrogen receptors and androgen receptors: a model system for the study of reproduction. *Biochem. Biophys. Res. Commun.*, **380**, 115–121.

Dang, Z. and Kienzler, A. (2019) Changes in fish sex ratio as a basis for regulating endocrine disruptors. *Environ. Int.*, **130**, 104928.

Hayashi, Y. *et al.* (2010) High temperature causes masculinization of genetically female medaka by elevation of cortisol. *Mol. Reprod. Dev.*, **77**, 679–686.

Higley, E.B. *et al.* (2010) Assessment of chemical effects on aromatase activity using the H295R cell line. *Environ. Sci. Pollut. Res. Int.*, **17**, 1137–1148.

Ishibashi, H. *et al.* (2016) Gene expression analyses of vitellogenin, choriogenin and estrogen receptor subtypes in the livers of male medaka (*Oryzias latipes*) exposed to equine estrogens. *J. Appl. Toxicol.*, **36**, 1392–1400.

Iwamatsu, T. (2004) Stages of normal development in the medaka Oryzias latipes. *Mech. Dev.*, **121**, 605–618.

Iwamatsu, T. *et al.* (2005) Estradiol-17beta content in developing eggs and induced sex reversal of the medaka (*Oryzias latipes*). *J. Exp. Zoolog. Part A Comp. Exp. Biol.*, **303**, 161–167.

Ji, X. *et al.* (2022) Comparison of the mechanisms of estrogen disrupting effects between triphenyl phosphate (TPhP) and tris(1,3-dichloro-2-propyl) phosphate (TDCIPP). *Ecotoxicol. Environ. Saf.*, **229**, 113069.

Kinoshita, M. *et al.* (2009) Medaka: Biology, Management, and Experimental Protocols John Wiley & Sons.

Kitano, T. *et al.* (2012) Estrogen rescues masculinization of genetically female medaka by exposure to cortisol or high temperature. *Mol. Reprod. Dev.*, **79**, 719–726.

Knörr, S. and Braunbeck, T. (2002) Decline in reproductive success, sex reversal, and developmental alterations in Japanese medaka (*Oryzias latipes*) after continuous exposure to octylphenol. *Ecotoxicol. Environ. Saf.*, **51**, 187–196.

2085 Kobayashi, H. and Iwamatsu, T. (2005) Sex reversal in the medaka *Oryzias latipes* by brief
2086 exposure of early embryos to estradiol-17beta. *Zool. Sci.*, **22**, 1163–1167.
2087 Kondo, M. *et al.* (2009) Sex determination and sex chromosome evolution: insights from
2088 medaka. *Sex Dev.*, **3**, 88–98.
2089 Kurauchi, K. *et al.* (2005) In vivo visual reporter system for detection of estrogen-like
2090 substances by transgenic medaka. *Environ. Sci. Technol.*, **39**, 2762–2768.
2091 Kurauchi, K. *et al.* (2008) Characteristics of ChgH-GFP transgenic medaka lines, an in
2092 vivo estrogenic compound detection system. *Mar. Pollut. Bull.*, **57**, 441–444.
2093
2094 Lee Pow, C.S.D. *et al.* (2016) Sharing the roles: An assessment of Japanese medaka
2095 estrogen receptors in vitellogenin induction. *Environ. Sci. Technol.*, **50**, 8886–8895.
2096 Lei, B. *et al.* (2013) β-estradiol 17-valerate affects embryonic development and sexual
2097 differentiation in Japanese medaka (*Oryzias latipes*). *Aquat. Toxicol.*, **134–135**,
2098 128–134.
2099 Loire, N. *et al.* (2013) Optimizing fluorescent protein choice for transgenic embryonic
2100 medaka models. *Environ. Toxicol. Chem.*, **32**, 2396–2401.
2101 Masuyama, H. *et al.* (2012) Dmrt1 mutation causes a male-to-female sex reversal after the
2102 sex determination by Dmy in the medaka. *Chromosome Res.*, **20**, 163–176.
2103 Matsuda, M. *et al.* (2002) DMY is a Y-specific DM-domain gene required for male
2104 development in the medaka fish. *Nature*, **417**, 559–563.
2105 Ni, M. *et al.* (2022) Bibliometric Analysis of the Toxicity of Bisphenol A. *Int. J. Environ.*
2106 *Res. Public Health*, **19**, 7886.
2107 OECD (2006) Current Approaches in the Statistical Analysis of Ecotoxicity Data.
2108 OECD (2019a) Test No. 203: Fish, Acute Toxicity Test.
2109 OECD (1992) Test No. 210: Fish, Early-Life Stage Toxicity Test.
2110 OECD (1998) Test No. 212: Fish, Short-term Toxicity Test on Embryo and Sac-Fry Stages.
2111 OECD (2012) Test No. 229: Fish Short Term Reproduction Assay.
2112 OECD (2009) Test No. 230: 21-day Fish Assay.
2113 OECD (2011) Test No. 234: Fish Sexual Development Test.
2114 OECD (2015) Test No. 240: Medaka Extended One Generation Reproduction Test
2115 (MEOGRT).
2116 OECD (2019b) Test No. 248: *Xenopus* Eleutheroembryonic Thyroid Assay (XETA).
2117 OECD (2021) Test No. 250: EASZY assay - Detection of Endocrine Active Substances,
2118 acting through estrogen receptors, using transgenic tg(cyp19a1b:GFP) Zebrafish embrYos.
2119 Paul-Prasanth, B. *et al.* (2013) Estrogen oversees the maintenance of the female genetic
2120 program in terminally differentiated gonochorists. *Sci. Rep.*, **3**, 2862.
2121 Salam, M.A. *et al.* (2008) Detection of environmental estrogenicity using transgenic
2122 medaka hatchlings (*Oryzias latipes*) expressing the GFP-tagged choriogenin L
2123 gene. *J. Environ. Sci. Health A Tox. Hazard Subst. Environ. Eng.*, **43**, 272–277.
2124 Sato, T. *et al.* (2005) Induction of female-to-male sex reversal by high temperature
2125 treatment in Medaka, *Oryzias latipes*. *Zool. Sci.*, **22**, 985–988.
2126 Scholz, S. and Gutzeit, H. (2000) 17-alpha-ethinylestradiol affects reproduction, sexual
2127 differentiation and aromatase gene expression of the medaka (*Oryzias latipes*).
2128 *Aquat. Toxicol.*, **50**, 363–373.

Scholz, S. *et al.* (2005) Analysis of estrogenic effects by quantification of green fluorescent protein in juvenile fish of a transgenic medaka. *Environ. Toxicol. Chem.*, **24**, 2553–2561.

Seki, M. *et al.* (2002) Effect of ethinylestradiol on the reproduction and induction of vitellogenin and testis-ova in medaka (*Oryzias latipes*). *Environ. Toxicol. Chem.*, **21**, 1692–1698.

Spirhanzlova, P. *et al.* (2016) Oestrogen reporter transgenic medaka for non-invasive evaluation of aromatase activity. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.*, **179**, 64–71.

Spirhanzlova, P. *et al.* (2020) Transgenic medaka identify embryonic periods sensitive to disruption of sex determination. *Environ. Toxicol. Chem.*, **39**, 842–851.

Ueno, T. *et al.* (2004) Identification of choriogenin cis-regulatory elements and production of estrogen-inducible, liver-specific transgenic Medaka. *Mech. Dev.*, **121**, 803–815.

Wakamatsu, Y. *et al.* (2001) Fertile and diploid nuclear transplants derived from embryonic cells of a small laboratory fish, medaka (*Oryzias latipes*). *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 1071–1076.

Yamamoto, T. (1958) Artificial induction of functional sex-reversal in genotypic females of the medaka (*Oryzias latipes*). *J. Exp. Zool.*, **137**, 227–263.

Yamamoto, T.O. (1955) Progeny of artificially induced sex-reversals of male genotype (XY) in the medaka (*Oryzias latipes*) with special reference to YY-Male. *Genetics*, **40**, 406–419.

Zeng, Z. *et al.* (2005) Development of estrogen-responsive transgenic medaka for environmental monitoring of endocrine disrupters. *Environ. Sci. Technol.*, **39**, 9001–9008.

Key scientific publications involving the chgh-gfp line

Spirhanzlova P. *et al.* (2019) Composition and endocrine effects of water collected in the Kibale national park in Uganda., *Environ Pollut.* **251**, 460-468.

Escher B.I. *et al.* (2018) Effect-based trigger values for *in vitro* and *in vivo* bioassays performed on surface water extracts supporting the environmental quality standards (EQS) of the European Water Framework Directive. *Sci. Total Environ.* **628-629**, 748-765.

Altenburger R. *et al.* (2018) Mixture effects in samples of multiple contaminants - An inter-laboratory study with manifold bioassays. *Environ. Int.* **114**, 95-106.

Neale P.A. *et al.* (2017) Development of a bioanalytical test battery for water quality monitoring: Fingerprinting identified micropollutants and their contribution to effects in surface water. *Water Res.* **123**, 734-750.

Välitalo P. *et al.* (2017) Effect-based assessment of toxicity removal during wastewater treatment. *Water Res.* **126**, 153-163.

Spirhanzlova P. *et al.* (2017) Using short-term bioassays to evaluate the endocrine disrupting capacity of the pesticides linuron and fenoxycarb. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **200**, 52-58.

Tousova Z. *et al.* (2017) European demonstration program on the effect-based and chemical identification and monitoring of organic pollutants in European surface waters. *Sci. Total Environ.* **601-602**, 1849-1868.

Spirhanzlova P. *et al.* (2016) Oestrogen reporter transgenic medaka for non-invasive evaluation of aromatase activity. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **179**, 64-71.

Maroga Mboula V. *et al.* (2015) Photocatalytic degradation of estradiol under simulated solar light and assessment of estrogenic activity. *Applied Catalysis B: Environmental* **162**, 437-444.

Castillo L. *et al.* (2013) *In vivo* endocrine disruption assessment of wastewater treatment plant effluents with small organisms. *Water Sci. Technol.* **68**, 261-8.