Environmental Toxicology and Chemistry—Volume 41, Number 10—pp. 2512–2526, 2022
Received: 10 December 2021 | Revised: 23 January 2022 | Accepted: 21 July 2022

Environmental Toxicology

Evaluation of Three ISO Estrogen Receptor Transactivation Assays Applied to 52 Domestic Effluent Samples

Eszter Simon,[a,b,1] Carolin Riegraf,[a,*,1] Andrea Schifferli,[a] Daniel Olbrich,[a] Thomas Bucher,[a,c] and Etiënne L. M. Vermeirssen[a]

[a]Swiss Centre for Applied Ecotoxicology, Dübendorf, Switzerland
[b]Air Pollution Control and Chemicals Division, Industrial Chemicals Section, Federal Office for the Environment, Bern, Switzerland
[c]Current affiliation: ETH Zurich, Lehrlabor Biologie, Zürich, Switzerland

Abstract: Estrogens are released to the aquatic environment by wastewater treatment plant (WWTP) effluents and can affect wildlife. In the last three decades, many in vitro assay platforms have been developed to detect and quantify estrogenicity in water. In 2018, the International Organization for Standardization (ISO) standardized protocols became available for three types of in vitro estrogen receptor transactivation assays (ERTAs) detecting estrogenicity in 96-well plates (ISO19040 1-3). Two ERTAs—lyticase Yeast Estrogen Screen (L-YES) and *Arxula* YES (A-YES)—use genetically modified yeast strains, whereas the third utilizes stably transfected human cells. One human cell based assay is ERα-CALUX, which is based on a genetically modified human bone osteosarcoma cell line. In the present study, we characterized the performance, comparability, and effectiveness of these three ERTAs, including an evaluation involving proposed water quality thresholds (effect-based trigger values [EBTs]). For a robust evaluation, we collected 52 effluent samples over three sampling campaigns at 15 different WWTPs in Switzerland. Estrogen receptor transactivation assay results were correlated and compared with results from chemical analysis targeting known estrogens. The three ERTAs showed comparable data over all campaigns. However, the selection of EBTs plays a significant role in the interpretation and comparison of bioassay results to distinguish between acceptable and unacceptable water quality. Applying a fixed cross-assay EBT for effluent of 4 ng L^{-1} resulted in varying numbers of threshold exceedances ranging between zero and four samples depending on the ERTA used. Using assay-specific EBTs showed exceedances in eight samples (ERα-CALUX) and in one sample (A-YES), respectively. Thus, proposed EBTs do not produce similar risk profiles across samples and further refinement of assay-specific EBTs is needed to account for assay-specific differences and to enable the application of ERTAs as effect-based methods in environmental monitoring. *Environ Toxicol Chem* 2022;41:2512–2526. © 2022 The Authors. *Environmental Toxicology and Chemistry* published by Wiley Periodicals LLC on behalf of SETAC.

Keywords: Estrogenic compounds; Water quality; Effect-based monitoring; Municipal effluents; In vitro toxicology; Estrogenicity; Reporter gene assay

INTRODUCTION

Synthetic and natural estrogens possess high biological activity and their release into the natural environment poses a threat to freshwater systems worldwide (Arlos et al., 2018; Brion et al., 2019). In high-income countries, wastewater treatment plant (WWTP) effluents have been identified as a major contributor to the release of estrogens into the environment (Adeel et al., 2017; Luo et al., 2014). Consequently, strategies have been devised to monitor the quantities of estrogens released from WWTPs. To this end, effect-based methods (EBMs) are gaining popularity as a means of water quality screening and have been successfully implemented to assess the estrogenic potential in freshwater bodies impacted by WWTP effluents (Brack et al., 2019). Effect-based method are designed to capture the effects of all compounds with a common mode of action. Analyses results are expressed as biological equivalence concentrations (BEQ), often in terms of the reference compound used in the bioassay, thus diverging from the classical determination of the individual compound concentration of analytical-chemical methods (e.g., European

This article includes online-only Supporting Information.

[1]These authors contributed equally to this work.
* Address correspondence to carolin.riegraf@oekotoxzentrum.ch
Published online 25 July 2022 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/etc.5445

Water Framework Directive [WFD]; European Union, 2000). A prominent example is the 17β-estradiol equivalence concentration (EEQ) for EBMs that target estrogens.

Agonistic effects such as estrogenicity may already affect aquatic organisms in concentrations in the low ng L^{-1} range, therefore highly sensitive bioassays are needed (Clouzot et al., 2008). A number of in vitro assay platforms have been developed to detect and quantify estrogenicity in recent years and have been broadly applied in the assessment of environmental aquatic samples (Archer et al., 2020; Dias et al., 2015; Simon et al., 2022; Yao et al., 2018). Many assay formats fall within the class of in vitro estrogen receptor transactivation assays (ERTAs; Kunz et al., 2017). For the successful implementation of ERTA-type EBMs in routine water quality monitoring, they need to be inter alia reliable and standardized. Thus, to support possible future integration within legal frameworks, three International Organization for Standardization (ISO) guidelines describing ERTAs were published in 2018 (ISO 19040 1–3). These three standardized assays are now available to determine the estrogenic potential of freshwater and wastewater within the scopes defined in their respective guideline.

The general principle of all three ERTAs is based on the binding of estrogenic compounds to the human estrogen receptor α (ERα; Kaiser et al., 2010; McDonnell, Nawaz, Densmore, et al., 1991; McDonnell, Nawaz, & O'Malley, 1991). Due to the differences between host cells and their cellular components as well as additional differing handling properties among ERTAs, the relative sensitivity to specific estrogens and mixtures of those varies from assay to assay. This leads to different effect potencies of compounds established in the different effect screens (Jarošová et al., 2014). Thus, a uniform data assessment—or BEQ/EEQ derivation—plays an essential role in the comparison of different assay methods (Escher et al., 2014; Kunz et al., 2017). To support a uniform and well-described data evaluation (Wagner et al., 2013), the ISO recently adopted and published a standard for BEQ derivation (ISO, 2022). Despite recent developments, relating BEQ results to potential implications for the aquatic environment remains challenging. The use of test-specific effect-based trigger values (EBTs) was proposed to facilitate the interpretation of bioassay results in terms of acceptable or unacceptable water quality (Brion et al., 2019; Escher et al., 2018; Kunz et al., 2017; van der Oost et al., 2017). However, different methods exist to derive EBTs and several EBTs exist for the same mode of action. These differences can lead to diverging water quality evaluation outcomes and comprehensive studies on ERTA intercomparability are currently missing.

The aim of the present study was to evaluate and intercompare the performances of three ERTAs based on their responses to a large number ($n = 52$) of WWTP effluents that discharge into the aquatic environment. The ERTAs applied in the present study are the Yeast Estrogen Screen (YES) using genetically modified *Saccharomyces cerevisiae* (lyticase YES [L-YES]; ISO, 2018a) and *Arxula adeninivorans* (*Arxula* YES [A-YES]; ISO, 2018b) as well as the ERα-CALUX, which is based on the human cell line U2OS-ERα-CALUX performed according to ISO (2018c). Effluent samples from 15 WWTPs were collected within three sampling campaigns accounting for seasonal variability in estrogenicity and generated large sample diversity to robustly test the ERTAs. High-performance liquid chromatography coupled to mass spectrometry was used to determine the concentrations of known estrogenicity eliciting compounds in the same samples. Data evaluation was performed along ISO (2022), ensuring uniform data evaluation across assay formats. Finally, different evaluation criteria were applied to the established results to compare the ERTAs with regard to potential exceedances of threshold and assay-specific EBTs.

MATERIALS AND METHODS

Chemicals and solvents

17β-Estradiol (E2; CAS 50-28-2, purity ≥98%), 17α-ethinylestradiol (EE2; CAS 57-63-6, purity ≥98%), estrone (E1; CAS 53-16-7, purity ≥99%), estriol (E3; CAS 50-27-1, ≥97%), 4-*tert*-octylphenol (CAS 140-66-9, purity 97%), and bisphenol A (BPA; CAS 80-05-7, purity 97%) as well as the stable-labeled internal standards 17β-estradiol-2,3,4-$^{13}C_3$ solution (CAS 1261254-48-1), 17α-ethinylestradiol-20,21-$^{13}C_2$ solution (CAS 2483735-63-1; CIL), estrone-2,3,4-$^{13}C_3$ solution (CAS 1241684-29-6), estriol-2,3,4-$^{13}C_3$ solution (CAS 1255639-56-5; CIL), 4-*tert*-octylphenol-$^{13}C_6$ (CAS: 1173020-24-0), and bisphenol-A-(diphenyl-$^{13}C_{12}$) solution (CAS 263261-65-0; CIL) were purchased from Merck (if not otherwise stated) in the highest grade commercially available.

Stock solutions of individual standards (1 mg ml^{-1}) were prepared in ethanol and then combined and further diluted to a standard mix of 10 μg ml^{-1} in ethanol. This mix was further diluted for the calibration series. Internal standards (apart from 4-*tert*-octylphenol) were obtained as 100 μg ml^{-1} solutions. A mix of 1 μg ml^{-1} was prepared in ethanol and diluted 1:25 for spike solutions (40 ng ml^{-1}).

Ethanol, hexane, acetone, and methanol as well as HCl were obtained from Merck and dimethyl sulfoxide (DMSO, purity ≥99.9%) was acquired from Sigma-Aldrich. For the three ERTAs, medium for cell growth was prepared according to the ISO standard procedures ISO 19040 1–3. For L-YES, medium components were purchased from Sigma-Aldrich and medium was prepared as described in ISO (2018a). The components for the *lacZ*-buffer consisting of 10.67 g L^{-1} $Na_2HPO_4 \times 2\ H_2O$, 0.75 g L^{-1} KCl, 0.25 g L^{-1} $MgSO_4 \times 7\ H_2O$, and 1g L^{-1} sodium dodecyl sulfate (CAS 151-21-3, purity >99%) were obtained from Sigma-Aldrich. Lyticase (CAS 37340-57-1, 344 U mg^{-1}), chlorophenolred-β-D-galactopyranoside (CAS 99792-79-7), L-cysteine (CAS 52-90-4, purity ≥97%), and dithiothreitol (CAS 3483-12-3, purity ≥98%) were also purchased from Sigma-Aldrich.

For ERα-CALUX, fetal calf serum (stripped) was purchased from BioDetection Systems. Futhermore, Eagle's minimal essential medium (no phenol red), penicillin–stryptomycin (5000 U ml^{-1}), minimal essential medium nonessential amino acids solution (100×), and liquid phosphate buffer solution (pH 7.2) were obtained from Gibco. Dulbecco's modified Eagle's medium (with 1000 mg L^{-1} glucose and L-glutamine, without

sodium bicarbonate and phenol red, powder) was acquired from Sigma-Aldrich. Lysis mix was prepared according to Bio-Detection Systems and the illuminate mix was purchased from BioDetection Systems.

A-YES was bought as an Eco Testkit from New Diagnostics. Besides solutions provided with the test kit, the substrate buffer and developer were prepared using the following substances purchased from Carl Roth: *p*-nitrophenylphosphate di-sodium hexahydrate (CAS 333338-18-4), tri-sodium citrate dihydrate (CAS 6132-04-3), citric acid (CAS 77-92-9), and sodium hydroxide (CAS 1310-73-2) from Sigma-Aldrich.

Experimental scheme

The comparison of the three ISO standardized bioassays was conducted with effluent samples from independent WWTPs. The experimental scheme is depicted in Figure 1. At each of three sampling campaigns, 16–18 samples were collected and subsequently 500-fold enriched by solid-phase extraction (SPE). The extracts were then subjected to the three different ERTAs as well as to chemical analysis using liquid chromatography–mass spectrometry (LC-MS/MS).

Effluent sampling conducted at Swiss WWTPs

Three sampling campaigns were conducted each over a period of 5–6 weeks within 12 months at 15 different WWTPs within Switzerland. Sampling Campaign 1 was performed in spring 2018 (Quarter II), Campaign 2 in summer 2018 (Quarter III), and Campaign 3 in winter 2019 (Quarter I). Marked temperature differences between sampling campaigns were linked to the different seasons when sampling occurred. Mean temperatures were 15.9 ± 1.2, 21.7 ± 1.0, and 12.1 ± 1.4 °C in Campaigns 1, 2, and 3, respectively.

During each sampling campaign, treated effluent samples were collected, depending on the site-specific features, after either the secondary clarifier or a final (sand) filtration step (Supporting Information, Table S1). Sampled WWTPs received an alphabetical sample ID from A to O. Wastewater treatment plant E was also equipped with an ozonation step. Furthermore, an ozonation step was implemented in WWTPs A and O between sampling Campaigns 1 and 2. Therefore, additional samples were collected at WWTPs A, E, and O after the secondary clarifier (i.e., before ozonation). Wastewater treatment plant A also implemented a filtration step along with the ozonation. Effluents collected after the ozonation step are also marked with a lower case "o" after the WWTP sample ID.

Effluents were collected as two 24-h flow proportional mixed samples (sampling day and the previous day) in aluminum vessels previously cleaned with acetone. On the sampling day, the cooled samples were mixed proportional to discharge, resulting in 48-h samples. For each SPE run (between five and seven per sampling campaign), one procedure blank and one solvent blank were included and later tested for potential background (effect) concentrations in ERTAs and LC-MS/MS. The procedure blank was nanopure water, which underwent the entire sample preparation procedure including filtration, pH-adjustment, extraction and solvent change (see *Materials and methods* section). Ethanol served as a solvent control.

Sample enrichment by SPE

Samples were filtered through a glass fiber filter (APFD09050, Ø 90 mm, pore size 2.7 µm; Merck Millipore) and then stabilized to $pH = 3.0 \pm 0.1$ with a 30% HCl solution. Enrichment of the 2-L samples was performed by an external laboratory using SPE equipped with an automated extractor (Dionex Autotrace 280; Thermo Scientific) according to the procedure described in Simon et al. (2019). Briefly, SPE cartridges (LiChrolut EN [40–120 µm], 100 mg [bottom]; LiChrolut RP-18 [40–63 µm], 200 mg [top]; 6 ml standard PP-tubes) were conditioned in several steps using hexane, acetone, methanol, and acidified water (pH 3.0). Elution was performed with methanol. Extracts were further concentrated in a Turbo Vap (Caliper Life Sciences) to a volume of 100 µl and subsequently filled up with ethanol to a final volume of 4 ml. Extracts were stored in glass Supelco vials with Teflon caps at −45 °C. At the end of each sampling campaign, extracts were transferred to the laboratory and stored at −20 °C until analysis.

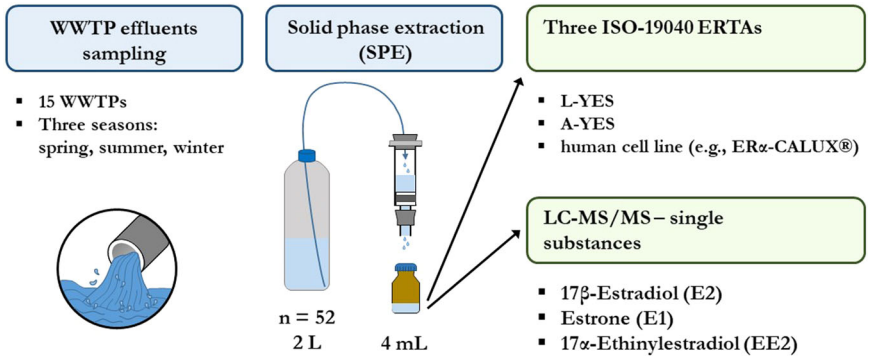

FIGURE 1: Experimental scheme for the in vitro estrogen receptor transactivation assay (ERTA) comparison. During three seasons, a total of 52 treated sewage effluent samples from 15 different wastewater treatment plants (WWTPs) were collected, enriched by solid-phase extraction (SPE), and subsequently investigated using International Organization for Standardization (ISO) standardized ERTAs (yeast estrogen screens [YES] and a human cell-line based assay) and chemical analysis (liquid chromatography–mass spectrometry [LC-MS/MS]).

Three bioassays for estrogen screening

The ISO 19040 standards suggest the testing of native water samples (ISO, 2018a, 2018b, 2018c). However, water samples are often enriched during routine analysis to increase sensitivity and to lower the quantification limits of bioassays. Testing of extracts is described in Annex D to ISO 19040 (ISO, 2018a, 2018b, 2018c). In the present study, samples were enriched by SPE as described in the *Materials and methods* section. Then, extracts were evaporated to dryness using nitrogen and re-dissolved in nanopure water (30 min under shaking), thereby enabling aqueous exposure conditions for the cells. The same aqueous extracts were used for the three ERTAs. An exception were the extracts of Campaign 1 investigated with ERα-CALUX. These extracts were re-dissolved in DMSO instead of nanopure water. Because the use of DMSO instead of nanopure water may affect results, a comparison of aqueous and DMSO exposure conditions was performed in Campaign 2. No significant differences between the two approaches could be observed (Supporting Information, Table S2). Previous method in-house evaluations also showed good comparability between aqueous and DMSO testing (data not shown). Serial dilutions of the sample extracts were prepared in glass vials to obtain optimal concentration ranges for each assay depending on previously conducted concentration range finding tests. Subsequently, the three bioassays were performed with the diluted extracts. A detailed description of the bioassays can be found in the Supporting Information. Briefly, lyticase yeast estrogen screen (L-YES) was performed according to ISO (2018a) using the *Saccharomyces cerevisiae* test strain BJ1991 (Purvis et al., 1991; Routledge & Sumpter, 1996). *Arxula* yeast estrogen screen (A-YES) was conducted according to ISO (2018b) and the protocol provided with the Eco Testkit of New Diagnostics using yeast cells of the genetically modified test strain *Blastobotrys adeninivorans* G1214 Syn: *Arxula adeninivorans* G1214 (*aleu2 aura3::ALEU2*; Simon et al., 2006). Human cell-based reporter gene assay ERα-CALUX (Sonneveld et al., 2004) was performed according to ISO (2018c). Human bone osteosarcoma cells (U2OS) were used under license from BioDetection Systems.

LC-MS/MS

In parallel to bioassays, extracts were investigated by LC-MS/MS and analyzed for the target estrogenic substances E1, E2, EE2, E3, BPA, and 4-*tert*-octylphenol. We used a method that was previously validated by comparing estrogenicity data ($n = 33$ effluent and surface water samples) with that of two other laboratories (see Könemann et al., 2018). Chemical analysis was performed in negative mode with an electrospray ionization source on an Agilent G6495A triple quadrupole mass spectrometer coupled to an ultrahigh performance liquid chromatography (UHPLC) system for chromatographic separation (Agilent 1290 Infinity II, Waters Acquity UPLC BEH Shield RP18, 130 Å, 2.1 mm × 100 mm, 1.7 µm column [p/n 186002854] with a 5-mm precolumn [p/n 186003977]). A methanol/water + 5 mM NH_3 gradient was applied as described in detail in Könemann et al. (2018) and Simon et al. (2019).

Data processing and statistical analysis

Data processing and statistical analysis were performed using Excel, GraphPad Prism 8.4.3, and R 3.6.1 (R Development Core Team, 2008) with the "ggplot" package (Wickham, 2016). Analytical data were evaluated using Agilent MassHunter Quantitative Analysis B08.00.

***Expressing bioassay results as EEQs* (ng EEQ L^{-1}).** Photometrical data were normalized to the solvent control (0%) and the highest effect of the measured reference compound E2 (100%), and subsequently presented as concentration–effect relationships (ISO, 2004). Concentration–effect relationships were fitted using a four-parameter log-logistic function.

The concentration–effect relationship of the reference compound E2 was used to interpolate the 10% effect level (PC10; ISO, 2022) of each sample and to determine E2-equivalence concentrations (ng EEQ L^{-1}) taking into account the respective sample dilution at 10% effect level (relative enrichment factor at 10% effect level [REF10]; Equation 1 and Supporting Information, Figure S1). Dilutions eliciting cytotoxicity were excluded from the quantification of estrogenicity.

$$EEQ_{Sample} = \frac{PC_{10}E2}{REF_{10}Sample} \quad (1)$$

Calculation of limit of quantification. Limits of quantification (LOQs) for the three bioassays as well as for mass spectrometry were calculated for each individual sample. The 10% effect level was specified as LOQ and expressed in ng EEQ L^{-1} using the highest REF of the respective sample. This approach is very conservative, as previous studies showed that the LOQ derived by the threefold standard deviation of the negative control corresponded with effect levels between 2% and 5% (Könemann et al., 2018). In the present study, the LOQ derived by the tenfold standard deviation of the negative control corresponded with averaged effect levels between 3.5% and 6.7% depending on the applied ERTA (data not shown). For samples showing effects >LOQ, an EEQ_{Sample} was derived (see Materials and methods section), and samples eliciting effects <LOQ were marked accordingly.

Concerning chemical analysis, limits of detection (LOD) were determined for each substance in every sample based on the signal intensity of the internal standard or the target substances using a signal-to-noise ratio (S/N) of 3. To establish the LOQ, a S/N of 10 was applied. Measured values >LOQ were quantified, values between LOQ and LOD were marked as nonrobust quantifiable values, and values <LOD were specified with 0. An overview of the determined LOQ values can be found in Supporting Information, Table S3.

Relative estrogenic potencies. Relative estrogenic potencies (REPs) were established for the estrogenic compounds E1, EE2, E3, BPA, and 4-*tert*-octylphenol in relation to E2. Analogues to extracts, working solutions of the estrogenic compounds, were evaporated to dryness using nitrogen and re-dissolved in nanopure water (30 min under shaking), thereby enabling aqueous exposure conditions for the cells. Relative

estrogenic potencies were calculated based on respective PC10 values and then averaged over the performed replicates (Table 1). The presented REPs are expressed considering the molecular weights of the respective estrogenic compounds.

Comparison of bioassay with LC-MS/MS results. The comparison of bioassays with LC-MS/MS results was performed according to previous studies (e.g., Hamers et al., 2008; Simon et al., 2010; or König et al., 2017). Briefly, analytically determined concentrations of the individual substances by LC-MS/MS can be used to calculate the expected effect in the bioassay (EEQ_{chem}). This allows a comparison with the measured effects in the bioassays (EEQ_{bio}). For the calculation of EEQ_{chem}, the measured concentration of the estrogenic substance (c_i) by LC-MS/MS was multiplied with its REP (Table 1) in the respective bioassay and subsequently summed to determine the EEQ_{chem} of the sample.

$$EEQ_{chem} = \sum c_i \cdot REP_i \quad (2)$$

Correlation and uncertainty analysis. A D'Agostino and Pearson test for normality was performed on log EEQ data (D'Agostino & Stephens, 1986). The log EEQ data followed a normal distribution and thus justified the performance of a Pearson correlation. Two exceptions were the A-YES and L-YES data in Campaign 1, for which correlation results therefore need to be considered with caution.

Comparative analysis based on threshold values. We applied threshold values for both bioassay data as well as for chemical analytical data. For bioassays, we used a trigger value of 0.4 ng EEQ L^{-1} for surface waters, which is suggested based on the predicted no-effect concentration (PNEC) of E2 (Escher et al., 2018; Kunz et al., 2015; Loos, 2012). This value was also proposed as environmental quality standard (EQS) for the former European Union-watch list compound E2 (Commission Implementing Decision, 2018). Wastewater treatment plants act as point sources by discharging their effluents into receiving waters. The exact information for outflow and river discharge was not available, thus a fixed dilution of effluent in receiving waters of 10 was assumed. This leads to a threshold for effluents of 4 ng EEQ L^{-1} (10 times the PNEC). In addition to this PNEC-based EBT, several bioassay-specific EBTs have been suggested, as described in more detail in the results and discussion sections. Second, a threshold value was used to evaluate concentrations of target substances determined by LC-MS/MS. The proposed thresholds for the individual substances was based on their 10× diluted PNEC. A 10× diluted PNEC of 36 ng L^{-1} was suggested for E1, 4 ng L^{-1} for E2, and 0.35 ng L^{-1} for EE2 (Loos et al., 2018).

TABLE 1: Overview of established relative estrogenic potencies (REPs) of selected estrogenic compounds in comparison to 17β-estradiol

		REPs of selected estrogenic compounds					
Bioassay	Replicates	E2	E1	EE2	E3	BPA	OP
A-YES	2	1.0	0.12	0.97	0.0074	1.1E−05	8.2E−06
ERα-CALUX	3	1.0	0.02	2.58	0.0879	4.2E−05	4.9E−06
L-YES	3	1.0	0.16	0.85	0.0069	4.5E−05	1.6E−05

E1 = estrone; E2 = 17β-estradiol; EE2 = 17α-ethinylestradiol; E3 = estriol; BPA = bisphenol A; A-YES = *Arxula* Yeast Estrogen Screen assay; L-YES = lyticase Yeast Estrogen Screen assay; OP = 4-*tert*-octylphenol.

RESULTS

Robust performances of three ISO standardized ERTAs

The stability of the individual test systems was investigated based on changes observed in the 50% effect concentration (EC50) of the reference compound E2 (Figure 2). All three ERTAs showed comparable stability over the test period. The EC50 is an important quality criterion of in vitro test systems to assess stability and sensitivity. Lower EC50 values indicate a higher sensitivity of the assay. An EC50 value of E2 in ERα-CALUX of $5.6 \times 10^{-12} \pm 2.4 \times 10^{-12}$M was established over the test period. In L-YES and A-YES, EC50-values of $8.0 \times 10^{-11} \pm 1.2 \times 10^{-11}$M and $4.31 \times 10^{-11} \pm 0.92 \times 10^{-11}$ M, respectively, were determined. These are all within the acceptable range set by the respective ISO guideline. Furthermore, three randomly selected samples per sampling campaign were independently measured a second time in L-YES and in the A-YES to determine the repeatability of the individual ERTAs (Supporting Information, Table S4). As seen in Supporting Information, Table S4, duplicate results showed good agreement, but results are only descriptive, because the number of samples tested with (only) double analysis is too small for a full repeatability analysis.

Comparable estrogenicity detected with A-YES, L-YES, and ERα-CALUX

A comparative assessment of estrogenicity detected with A-YES, L-YES and ERα-CALUX was conducted based on effluent samples collected within three sampling campaigns from 15 different WWTPs. The estrogenicity of effluent samples was expressed as E2 equivalence concentration (ng EEQ L^{-1} effluent). Negative controls and solvent controls showed no estrogenicity. The EEQ values in WWTP effluents determined with the three ERTAs ranged between 0.03 and 6.1 ng L^{-1}. The highest EEQ values were found in samples from Campaign 3.

Figure 3 shows a correlation between results from the three ERTAs. For a perfect agreement of the three ERTAs, all data would be located on the diagonal line (1 to 1 line) with random scatter. Overall, L-YES and A-YES results were highly correlated (Pearson correlation coefficient on log transformed data [r_P] = 0.84, $n = 37$, $p \leq 0.0001$; Supporting Information, Table S5). In particular, the results of Campaign 3 showed a high correlation ($r_P = 0.96$, $n = 15$, $p \leq 0.0001$). Comparing the correlation of A-YES and L-YES results with those obtained using ERα-CALUX, overall correlations were lower ($r_P = 0.69$, $n = 41$, $p \leq 0.0001$ and $r_P = 0.64$, $n = 38$, $p \leq 0.0001$, respectively). Regarding A-YES and L-YES, results agreed better in the higher than in the lower EEQ range (Figure 3A). A-YES

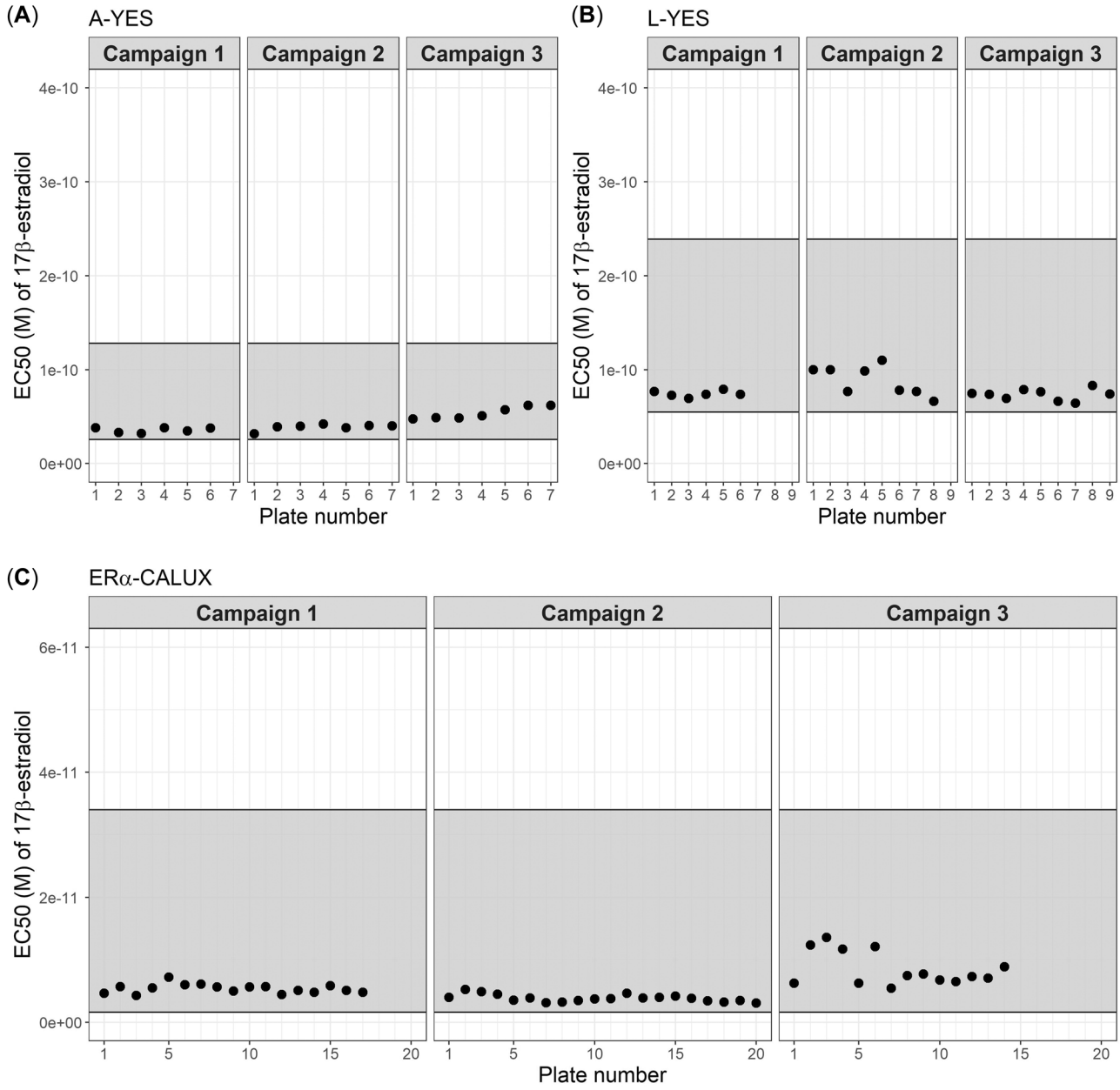

FIGURE 2: Stability of test performances depicted as variability in the 50% effect concentration (EC50) of the reference compound 17β-estradiol for the three sampling campaigns shown for (**A**) *Arxula* Yeast Estrogen Screen (A-YES), (**B**) lyticase Yeast Estrogen Screen (L-YES), and (**C**) ERα CALUX. The acceptable range according to the respective International Organization for Standardization Guideline (19040:1-3) or recent internal control data (L-YES) is shown in grey.

showed higher EEQ levels than L-YES in Campaigns 1 and 2, indicated by the location of most data points above the 1 to 1 line. Differences in Campaign 3 did not follow a clear pattern toward one particular test system, showing a good agreement of the two yeast-based ERTAs. Apart from Campaign 3, ERα-CALUX indicated higher EEQs of effluent samples compared to the yeast-based assays, L-YES (Figure 3B) and A-YES (Figure 3C). Campaign 3 displayed a different pattern in the higher EEQ range, indicating higher EEQs by the respective yeast-based assays compared to ERα-CALUX. However, in the lower EEQ range, the results of Campaign 3 data obtained by ERα-CALUX and A-YES corresponded well.

The modeled linear regression among the different bioassay comparisons were similar (Supporting Information, Figure S5). However, deviations from the 1 to 1 line were most apparent in the comparison of ERα-CALUX with L-YES (Supporting Information, Figure S5).

Extraction efficiency of target compounds eliciting estrogenic effects

The extraction efficiency of the six target substances (E1, E2, EE2, E3, BPA, and 4-*tert*-octylphenol) was determined as recovery rates of the target substances after SPE. Therefore,

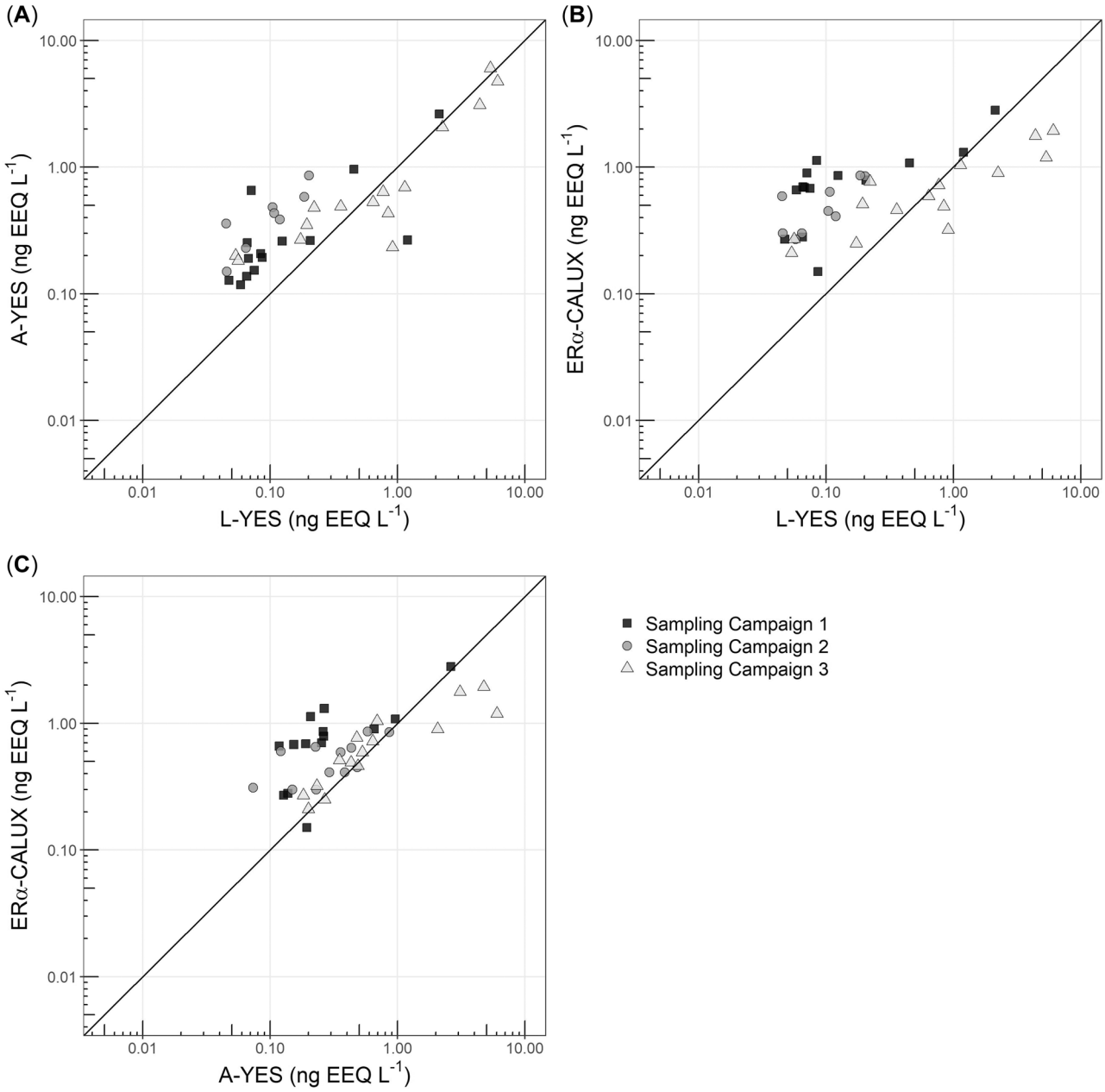

FIGURE 3: Comparison of the three in vitro estrogen receptor transactivation assay standards from the International Organization for Standardization: (**A**) *Arxula* Yeast Estrogen Screen (A-YES), (**B**) lyticase Yeast Estrogen Screen (L-YES), and (**C**) ERα-CALUX by correlation of effect concentrations detected in wastewater treatment plant (WWTP) effluents. Determined effect-concentrations in the WWTPs collected within three sampling campaigns are depicted as reference compound (17β-estradiol) equivalence concentrations per liter effluent sample (ng EEQ L^{-1}). The black solid line represents a perfect agreement of the methods.

concentrations in nanopure water spiked with defined concentrations of the target substances were compared before and after SPE. The spiked nanopure water was freshly prepared before each sampling campaign and split into several fractions to be included in each SPE run and as positive control without SPE. The extraction efficiency improved slightly along the three sampling campaigns. Regarding the biologically relevant target substances E1, E2, and EE2, recovery rates were between 67% and 77% in Campaign 1. In Campaign 2, recovery rates increased to values between 81% and 84%, and in Campaign 3, recovery rates of 72%–93% were determined. A detailed presentation of recovery rates per target substance for each individual SPE run can be found in Supporting Information, Table S6. Presented data obtained with the ERTAs or by LC-MS/MS were not corrected for extraction efficiency.

Comparison of results obtained by bioassays and LC-MS/MS

The comparison of results obtained in the bioassays and by LC-MS/MS was based on EEQs. First, relative potencies of the individual target substances were established. Then, measured

concentrations of E1, E2, and EE2 were adjusted for their estrogenic potency in the respective bioassay and summed up to obtain EEQs for LC-MS/MS results (EEQ_{chem}). Equivalence concentrations (EEQ_{chem}) were subsequently compared with the measured estrogenicity in the three bioassays (EEQ_{bio}).

The presence of the six individual substances E1, E2, EE2, E3, BPA, and 4-*tert*-octylphenol was determined in all samples. Estriol, BPA, and 4-*tert*-octylphenol showed a lower biological potency compared to E1, E2, and EE2 (Table 1). Estriol was rarely detected in the samples, but BPA and 4-*tert*-octylphenol occurred in concentrations up to 210 and 260 ng L^{-1}, respectively, showing high variations depending on WWTP and sampling campaign. However, considering the estrogenic potencies, even these elevated concentrations of BPA and 4-*tert*-octylphenol corresponded to <0.01 ng EEQ L^{-1}. The steroids E2 and EE2 were detected above their LOD (>LOQ) in only nine (six) and six (zero), respectively, out of 52 WWTP effluent samples (Supporting Information, Figure S7 and Table S8). In most samples, only the less potent E1 (Table 1) was detected. Estrone was found in 45 out of 52 samples in concentrations ranging from 0.1 to 17.4 ng L^{-1}. Evaluating LC-MS/MS data, no exceedances of the 10× diluted PNEC for individual substances were observed (10× diluted PNEC of E1 36 ng L^{-1}, E2 4 ng L^{-1}, and EE2 0.35 ng L^{-1} (Loos et al., 2018).

At low EEQ values, ERα-CALUX and A-YES data deviated significantly from LC-MS/MS data (Figure 4). With increasing EEQ values, EEQ_{bio}, and EEQ_{chem} values obtained for ERα-CALUX and A-YES showed decreased deviation. The ERα-CALUX data were subjected to higher uncertainties in the low concentrations but showed slightly higher EEQ_{bio} values compared to A-YES (Figure 4). Estrogenicity expressed as EEQ_{chem} and EEQ_{bio} was quantified above $LOQ_{A\text{-}YES} = 0.02$ ng L^{-1}, $LOQ_{L\text{-}YES} = 0.04$ ng L^{-1}, and $LOQ_{ER\alpha\text{-}CALUX} = 0.02$ ng L^{-1}. The number of samples showing no quantifiable estrogenicity determined by both LC-MS/MS and bioassay was 18 and 11 in A-YES, 20 and 14 in L-YES, and 37 and 2 in ERα-CALUX, respectively. The overall correlation of EEQ_{chem} and EEQ_{bio} was $r_P = 0.60$ ($n = 75$, $p \leq 0.0001$) with higher correlations observed for EEQ_{chem} with A-YES EEQ_{bio} ($r_P = 0.85$, $n = 32$, $p \leq 0.0001$) compared to L-YES EEQ_{bio} ($r_P = 0.71$, $n = 28$, $p \leq 0.0001$) and ERα-CALUX EEQ_{bio} ($r_P = 0.72$, $n = 15$, $p \leq 0.01$; Supporting Information, Table S5 and Figure S6).

The E2-PNEC threshold value of 4 ng EEQ L^{-1} was exceeded on four incidences (Figure 5). One incidence occurred in WWTP K during Campaign 1, indicated only by the EEQ_{chem} result but not by EEQ_{bio}. No threshold exceedance was observed in Campaign 2. In Campaign 3, samples from WWTP I and K showed values above 4 ng EEQ L^{-1} observed with both A-YES and L-YES. The L-YES data evaluation also revealed an exceedance of WWTP O. No threshold exceedance was observed with ERα-CALUX.

Besides an E2-PNEC threshold value, several EBTs have been suggested for various assays. As an example, an EBT ERα-CALUX of 0.1 ng EEQ L^{-1}, an EBT A-YES of 0.56 ng EEQ L^{-1}, and an EBT L-YES of 0.97 ng EEQ L^{-1} were proposed by Escher et al. (2018).

Furthermore, the determined EEQ_{bio} values can be set in relation to their respective EBT, taking into account the tenfold dilution on entering surface waters, to calculate an effect-based risk quotient (RQ) according to Kase et al. (2018). Applying the 10× EBTs suggested by Escher et al. (2018) for ERα-CALUX, an unacceptable risk for aquatic organisms, indicated by RQ > 1, was found in eight WWTP effluents across all samplings (Figure 6). Regarding the A-YES data, only WWTP K showed RQ > 1 (Supporting Information, Figure S8) and no exceedances were noted with L-YES (Supporting Information, Figure S9). An RQ greater than 1 was further observed in WWTP K in Campaign 1 using ERα-CALUX when applying a 10× EBT of 2 ng EEQ L^{-1} established by Jarošová et al. (2014; Figure 6).

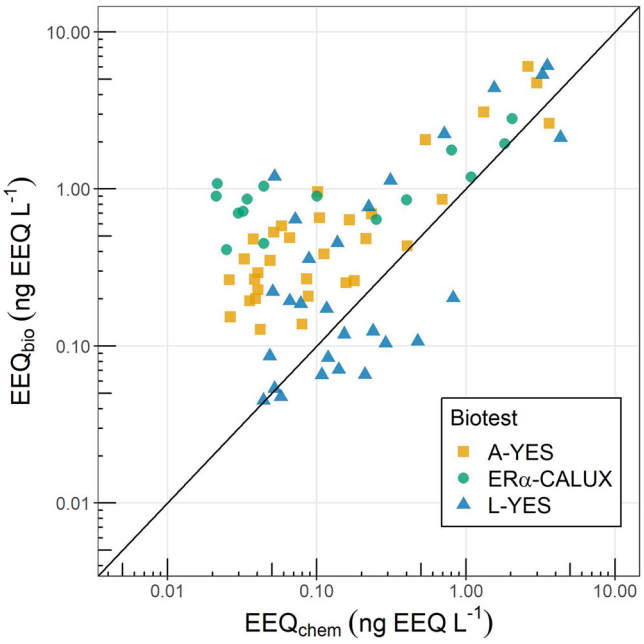

FIGURE 4: Comparison of bioassays and liquid chromatography–mass spectrometry (LC-MS/MS) by correlation of determined equivalence concentrations (EEQs). EEQ_{bio} established using three International Organization for Standardization-standardized in vitro estrogen receptor transactivation assays *Arxula* Yeast Estrogen Screen (A-YES), ERα-CALUX, and lyticase Yeast Estrogen Screen (L-YES) versus EEQ_{chem} determined by LC-MS/MS in wastewater treatment effluent samples collected during three campaigns are shown. EEQ_{chem} were determined by the sum of measured single substance concentrations multiplied with their respective relative potencies in the respective bioassays. The black line represents a perfect agreement of the methods.

DISCUSSION

Stable, sensitive, and repeatable assay performance for three ISO ERTAs using 17β-estradiol

All three ERTAs showed comparable stability over the test period with established EC50 values for E2 located in the lower part of the respective acceptable range (Figure 2). The repeatability of derived results was confirmed by independently re-analyzing several samples. The EC50 value of E2 was lowest in ERα-CALUX with 5.6 pM, indicating a higher E2 sensitivity of ERα-CALUX compared to the yeast-based assays. This finding corresponds with observations by

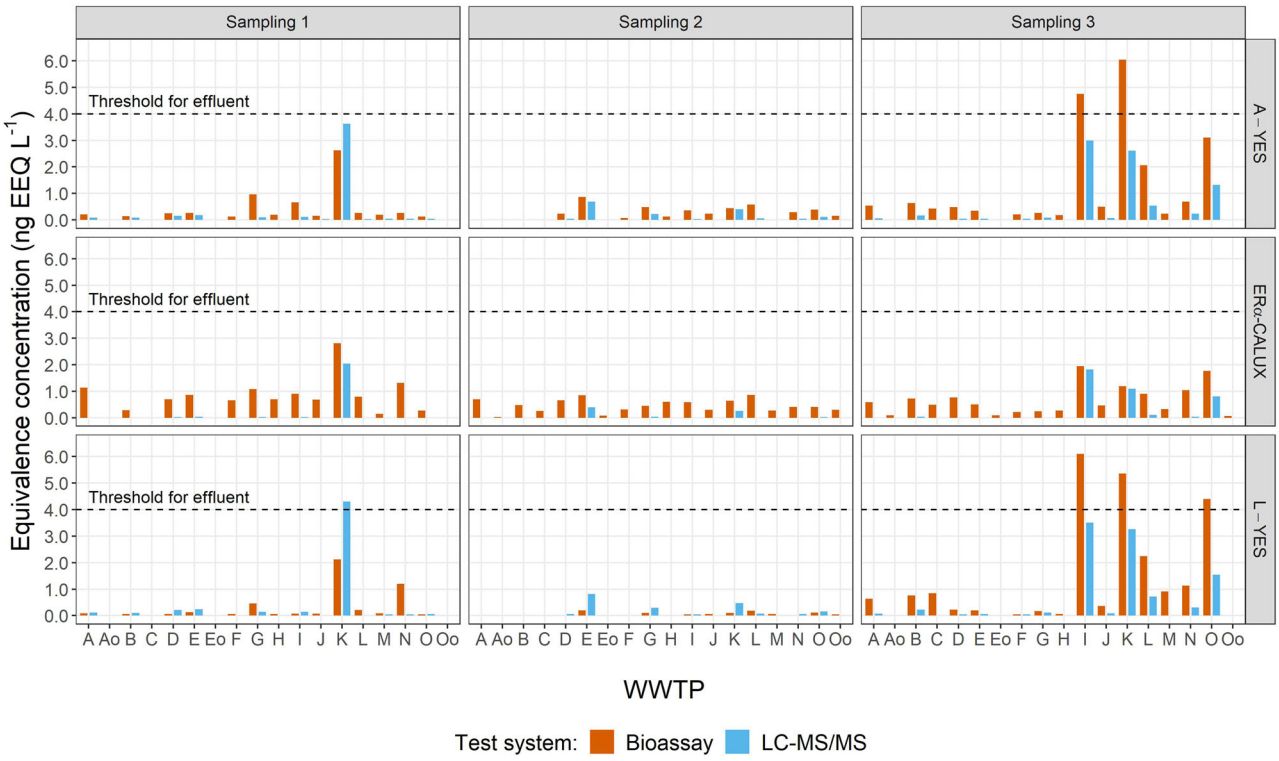

FIGURE 5: Comparison between the three International Organization for Standardization-standardized in vitro estrogen receptor transactivation assays *Arxula* Yeast Estrogen Screen (A-YES), ERα-CALUX, and lyticase Yeast Estrogen Screen (L-YES) with results obtained by liquid chromatography–mass spectrometry (LC-MS/MS). The comparison is based on measured and calculated equivalence concentrations (EEQ) in wastewater treatment effluents (A–O) during three sampling campaigns. Equivalence concentrations by LC-MS/MS were determined by the sum of measured single substance concentrations multiplied with their respective relative potencies in the respective bioassay. Black dashed horizontal lines represent the 17β-estradiol-predicted no observable effect concentration threshold value of 4 ng EEQ L^{-1} for wastewater treatment plant (WWTP) effluents.

Leusch et al. (2010), who found that a yeast-based assay was less sensitive and thus showed more nondetects in comparison to ERα-CALUX. Studies by Legler et al. (2002) and Gehrmann et al. (2018) also showed a higher E2 sensitivity of ERα-CALUX compared with yeast-based assays. In these studies, EC50 values for E2 of 6 and 5.9 pM in ERα-CALUX as well as 100 and 110 pM in L-YES were established (Gehrmann et al., 2018; Legler et al., 2002). Gehrmann et al. (2018) also reported a higher E2 sensitivity of ERα-CALUX in comparison to A-YES (EC50 for E2 = 62.4 pM). A possible reason for the increased sensitivity of ERα-CALUX toward E2 compared to yeast-based assays is the missing cell wall functioning as a potential barrier in yeast cells (Legler et al., 2002).

The repeatability of the assays differed between the ERTAs, showing relative standard deviations of 21%, 43%, and 15% for EC50 values determined in A-YES, ERα-CALUX, and L-YES, respectively. By considering the different number of replicates, relative standard errors of 4.8%, 6.0%, and 3.2% were obtained in A-YES, ERα-CALUX, and L-YES, respectively. The comparably high standard deviation in ERα-CALUX might be attributed to Campaign 3 (Figure 2).

Established REPs in the present study (Table 1) were in the same range as the REPs described in literature. Van den Belt et al. (2004) determined REPs of 0.4 for E1 and 0.9 for EE2 in another yeast-based assay, which correspond well with the determined 0.16 for E1 and 0.85 for EE2 in the present study. In addition, the REPs derived by Murk et al. (2002) of 1.2 for EE2, 1.0×10^{-5} for BPA, and 1.0×10^{-5} for 4-*tert*-octylphenol in another yeast-based assay are similar to the determined 0.85 for EE2, 4.5×10^{-5} for BPA, and 1.6×10^{-5} for 4-*tert*-octylphenol in the present study. In Murk et al. (2002), REPs in ERα-CALUX of 1.2 for EE2, 7.8×10^{-6} for BPA, and 1.4×10^{-6} for 4-*tert*-octylphenol were established. These differ slightly from the REPs in the present study with values of 2.6 for EE2, 4.2×10^{-5} for BPA and 4.9×10^{-6} for 4-*tert*-octylphenol. Brion et al. (2019) reported REPs using ERα-CALUX of 0.01 and 1.30 for E1 and EE2, respectively. For A-YES, Hettwer et al. (2018) reported REPs of 1.2 and 1.68×10^{-5} for EE2 and BPA, respectively. The values established in the present study for A-YES were 0.97 and 1.1×10^{-5} for EE2 and BPA, respectively.

The time required to perform an assay was similar for all ERTAs. The ERTA, L-YES is easier to use and has the lowest material costs followed by A-YES, and then ERα-CALUX. In the present study, we tested samples enriched by SPE and subsequently redissolved in nanopure water in glass vials as described in Annex D to ISO 19040 (ISO, 2018a, 2018b, 2018c). This enables aqueous exposure conditions for the cells, a uniform sample application procedure across all assays at the same time, and the achievement of low detection levels. Besides using glass vials, the redissolving and dilution of the sample extracts for L-YES could be performed directly in the wells, because yeast cells do not require a prior seeding and incubation in 96-well

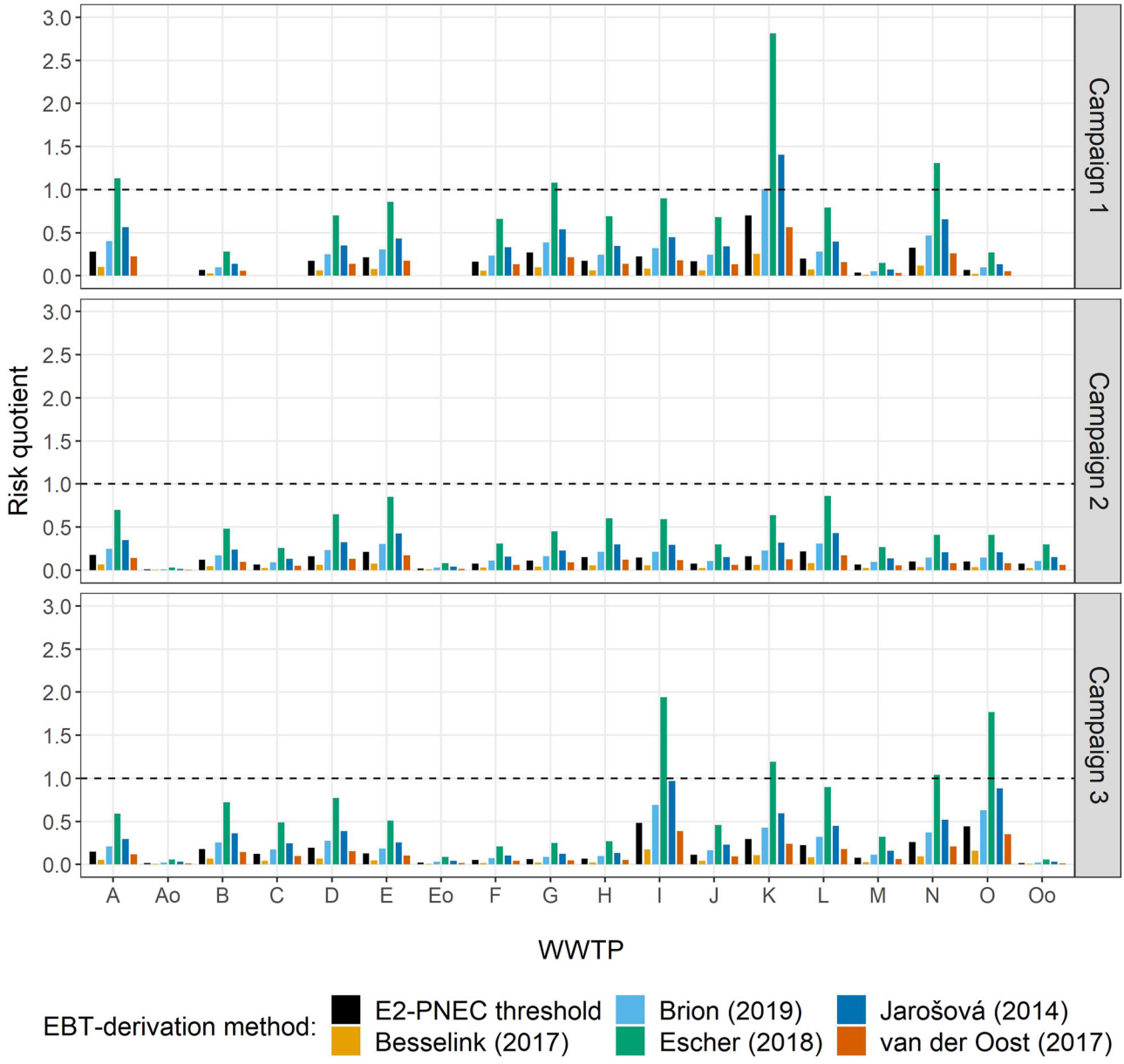

FIGURE 6: Risk quotients (RQs) of wastewater treatment plant effluents (WWTP, A-O) separately shown per sampling campaign. Risk quotients were determined by dividing ERα-CALUX derived equivalence concentrations (EEQ_{bio}) by the respective bioassay-specific effect-based trigger value (EBT). Taking into account a tenfold dilution, calculated RQs using the 17β-estradiol-predicted no observable effect concentration threshold value (E2-PNEC threshold) as well as EBTs suggested by Besselink et al. (2017), Brion et al. (2019), Escher et al. (2018), Jarošová et al. (2014) and van der Oost et al. (2017) are depicted.

plates in contrast to ERα-CALUX (ISO, 2018a). The sample preparation on 96-well plates for A-YES is more challenging to handle compared to L-YES due to its deep well plate format. In comparison to the L-YES plate, which is transparent, the deep well plate is approximately threefold deeper and opaque. This complicates the handling of small amounts of fluids in A-YES. Furthermore, special centrifuges and shakers are needed to centrifuge and shake the deep well plates. However, the yeast cells for A-YES do not have to be cultivated but only reactivated for 1 h prior to use (ISO, 2018b). Regarding ERα-CALUX, the incubation step needs to be performed in a specific CO_2 atmosphere to enable perfect growing of the cells (ISO, 2018c). The ERTA A-YES is purchased in kit format including all solutions ready to use. The ERTA ERα-CALUX meets the requirements of the generic human cell line standardized by ISO and can be run under license or by a service laboratory. For both commercial assays—A-YES and ERα-CALUX—the assay providers offer technical support.

Thus, L-YES is beneficial for cost-efficient testing of samples with highly reproducible results. If A-YES is purchased as a kit, the test is more expensive, but no time for the preparation of buffer and medium is needed. Finally, for the detection of estrogenicity occurring at low levels, for example when no sample enrichment is used, ERα-CALUX is more suitable.

The three in vitro assays show mostly similar estrogenicity patterns for WWTP effluent samples

Overall, the three ISO standardized ERTAs provided comparable results over the three sampling campaigns. Samples with low and high estrogenicity were similarly identified by the three bioassays. The ERTA, ERα-CALUX was generally more sensitive compared with the yeast-based assays and was able to quantify estrogen activity in most samples (Figures 3 and 5). Among the yeast-based assays, the number of nondetects in samples was comparable, but the A-YES indicated quantifiable estrogenicity in slightly more samples compared to the L-YES (Figure 5).

The L-YES and A-YES data show a better correlation than the yeast-based data with ERα-CALUX (Supporting Information, Table S5). Furthermore, higher uncertainty was observed for lower concentrations of estrogenic sample components, reflected by broader data scattering when comparing the different ERTAs (Figure 3). In addition, the mixture of compounds in the sample can have an effect on the test result. The ERα-CALUX is less sensitive to E1 compared to the yeast-based ERTAs (Kunz et al., 2017). Thus, higher yeast-based derived EEQs are expected in case of high E1 concentrations dominating the mixture of compounds in the sample. Particularly high E1 concentrations were found in WWTP I and K. In addition, the WWTPs L, N, O, and B showed high E1 concentrations among the investigated WWTPs (Supporting Information, Figure S7 and Table S8). This finding corresponds well with the results of the yeast-based bioassays showing elevated estrogenicity, especially in WWTPs I, K, L, and O (Figure 5). The detected estrogenicity in this sample using ERα-CALUX is comparatively low, matching its lower sensitivity toward E1.

The EEQ values of the three ERTAs ranging between 0.03 and 6.1 ng L^{-1} are comparable with findings of recent studies. Houtman et al. (2018) reported <1 ng L^{-1} in Dutch WWTP effluents and Bain et al. (2014) detected maximal values of approximately 6 ng L^{-1} in Australian WWTP effluents, both studies using ERα-CALUX. The EEQ values in treated effluents between 0.03 and 23 ng L^{-1} using ERα-CALUX were detected by Könemann et al. (2018). And EEQ values up to 0.08 ng L^{-1} were found in final WWTP effluents in Germany using A-YES (Itzel et al., 2017). Furthermore, Archer et al. (2020) reported EEQ values between 0.2 and 6.9 ng L^{-1} in South African wastewaters using another yeast-based assay.

Threshold exceedances indicated by the three ERTAs depend on the applied thresholds

Beside a direct comparison of EEQs from different ERTAs (as discussed above, *The three in vitro assays show mostly similar estrogenicity patterns for WWTP effluent samples*), assays can also be evaluated based on EEQ threshold exceedances. These threshold values or EBTs can help to distinguish between acceptable and unacceptable water quality regarding estrogenicity. However, depending on the selection of the threshold, this adds complexity to the evaluation and can produce diverse outcomes, particularly because several EBTs have been suggested for various assays. Each EBT-derivation method has its merits, but there is no consensus yet on which EBT values to use and which values are most robust to protect effects in wildlife (e.g., fish). As a starting point, we used a generic bioassay-independent threshold value of 4 ng EEQ L^{-1} for WWTP effluents across the three assays. This threshold value is based on the suggested trigger value of 0.4 ng EEQ L^{-1} for surface waters, in turn derived from the PNEC of E2. Assuming a fixed tenfold dilution of the effluents, this leads us to a tenfold PNEC-based value of 4 ng EEQ L^{-1}. However, a proportion of WWTP effluent in a receiving water that exceeds the suggested 10% would warrant a lower threshold. Applying this threshold value, no threshold exceedances (mostly <1 ng EEQ L^{-1}) were found in samples collected in spring (Campaign 1) and summer (Campaign 2). However, EEQs were somewhat higher in winter (up to 6 ng EEQ L^{-1}). Up to three samples collected in sampling Campaign 3 reached or slightly exceeded the selected threshold value of 4 ng EEQ L^{-1} for WWTP effluents in both yeast-based assays but not in the human cell line (Figure 5).

However, if bioassay-dependent EBTs are applied, the outcome changes depending on the EBT, and ERTA used. The EBTs in a range spanning from 0.1 to 3.8 ng EEQ L^{-1} are discussed for different estrogenicity assays (Been et al., 2021; Brand et al., 2013; Brion et al., 2019; Escher et al., 2018; Jarošová et al., 2014; Kase et al., 2018; Kunz et al., 2015; Loos, 2012; van der Oost et al., 2017). Different approaches for the determination of EBTs exist and are described and summarized in detail in Escher et al. (2018) and Escher and Neale (2021). Briefly, EBTs can be derived using a human health-based approach relating safe in vivo to in vitro detectable concentrations taking into account inter alia acceptable or tolerable daily intake values and estimates of bioavailability (Brand et al., 2013). The suggested ERα-CALUX EBT using this approach is 3.8 ng EEQ L^{-1} (Brand et al., 2013). Been et al. (2021) proposed an ERα-CALUX EBT of 0.25 ng EEQ L^{-1} based on relative potencies and provisional health-based guidance values in drinking water. Following different approaches, other EBTs for surface water comparing estrogenic in vivo with in vitro effects were derived. Brion et al. (2019) suggested an EBT of 0.28 ng EEQ L^{-1} for ERα-CALUX based on in vivo effects observed in zebrafish embryos. In addition, Arlos et al. (2018) established a potential link between modelled YES-EEQ concentrations and intersex in fish. Unfortunately, no EBT value is provided, but this kind of analysis predicted approximately 0.2 ng EEQ L^{-1} under a low flow scenario. This concentration would correspond to low intersex incidence (20% with 95% CI 16–24). Another approach to establish EBTs is based on EQS values considering specific uncertainty factors and estimates of species sensitivity distributions (SSD). This approach can be pursued by directly relating E2 EQS (Kunz et al., 2015) or also taking into account potency factors, in vivo derived PNECs and the distribution in WWTP effluents of different steroids to establish EBTs (Jarošová et al., 2014). In addition, Escher et al. (2015) and Escher et al. (2018) established EBTs by a combination of the derivation methods using the EQS presented

above and the focus on potential mixture effects. Bioassay-specific EBTs suggested by Escher et al. (2018) are 0.56, 0.10, and 0.97 ng EEQ L^{-1} for A-YES, ERα-CALUX and L-YES, respectively, as displayed in Supporting Information, Table S9. Recently, a mixture toxicity EBT was proposed by Escher and Neale (2021) with exemplary derived EBTs for cytotoxicity, oxidative stress, and aryl hydrocarbon receptor-activation. A smart integrated monitoring (SIMONI) strategy for the derivation of environmental EBTs based on laboratory data and field observations was proposed by van der Oost et al. (2017). The EBTs for in vivo bioassays expressed as toxic units are derived from acute effect concentrations using different safety and translation factors to cover potential chronic effects. The EBT derivation for in vitro bioassays is based on a combination of lowest observed chronic effect equivalents, SSD-derived hazardous concentration for 5% of water organisms (HC5 BEQ), and a background BEQ representing the bioassay response at eight selected sites showing good ecological status (van der Oost et al., 2017). An EBT for ERα-CALUX of 0.5 ng EEQ L^{-1} was proposed using the SIMONI strategy (van der Oost et al., 2017; Supporting Information, Table S9).

Depending on the applied EBT, the exceedance of the threshold value of a sample changes. Thus, the EEQ results obtained with the three ERTAs differ only slightly, but the outcome is rather influenced by differences in the evaluation criterion or EBT of choice. The difference between acceptable and nonacceptable water quality for an ecotoxicological status assessment can also be determined by comparing RQs of individual samples depicting the normalized data. Samples showing RQ > 1, indicating an unacceptable risk for aquatic organisms, varied between one and two for A-YES (Supporting Information, Figure S8), zero and eight for ERα-CALUX (Figure 6), and zero and three for L-YES (Supporting Information, Figure S9). Regarding A-YES, the assay-independent threshold value of 4 ng L^{-1} would indicate an exceedance by 4% of the investigated samples (two out of 52) compared to 2% applying the bioassay-specific EBT value suggested by Escher et al. (2018). A similar trend can be observed for L-YES, where 6% of the samples (three out of 52) exceeded 4 ng L^{-1} but no exceedance using the bioassay-specific EBT value proposed by Escher et al. (2018) was found. In contrast, ERα-CALUX showed zero exceedance of the samples when applying the threshold value of 4 ng L^{-1} but 15% of the samples indicated a risk when looking at the bioassay-specific EBT value (1 ng L^{-1}, taking into account a tenfold dilution of effluent in receiving water) from Escher et al. (2018). Nevertheless, the bioassay-specific EBT exceedance shown with A-YES overlapped with the exceedance indicated with ERα-CALUX. Besides the EBTs suggested in literature, we also used the EC50 values of individual compounds determined in the present study to calculate EBTs for A-YES and L-YES based on the approach suggested by Escher et al. (2018; Supporting Information, Table S9). This approach indicated RQ > 1 in the same samples as the assay-independent threshold value of 4ng L^{-1} (Supporting Information, Figures S8 and S9).

These observations highlight the need for appropriate, robust and common EBTs to assess risks across different studies. In particular, studies elucidating and correlating in vitro with in vivo results, as proposed by Brion et al. (2019) or Arlos et al. (2018), could help to improve the data basis to establish robust EBTs.

Bioassays capture more estrogenicity compared with chemical analysis

The robustness and specificity of bioassays can be assessed by comparing ERTA and chemical analytical data, provided the full mixture can be analyzed and relative potencies are known. Target compound concentrations are often low and below LOQ, and the full data range cannot be sufficiently explored. In comparison to LC-MS/MS analysis of targeted chemicals, the bioassays showed higher and overall quantifiable estrogenicity (Figure 5). An exception was WWTP K, where EEQ_{chem} was a factor of 2.3 higher compared to EEQ_{bio} (L-YES). This discrepancy could be due to the composition of the sample, because the bioassay may be less sensitive to certain components. However, this is relatively unlikely because the sample mix is similar to that of WWTP K in Campaign 3, where the discrepancy was not observed. Another possible explanation is the presence of anti-estrogens in the sample. However, this is not a good hypothesis either, because ERα-CALUX and A-YES do not indicate this. Finally, cytotoxicity could have attenuated signal intensity in L-YES, but cell density data does not support this explanation.

A comparison of EEQ_{chem} and EEQ_{bio} revealed that, especially in ERα-CALUX, the number of samples showing no quantifiable estrogenicity deviated (EEQ_{chem} 37 samples and EEQ_{bio} two samples were below LOQ). However, WWTP effluents showing high estrogenicity determined with the bioassays indicated elevated estrogenicity also in LC-MS/MS results (Figure 5). In general, L-YES data matched best with LC-MS/MS data, that is the linear regression line runs close to the 1 to 1 line (Supporting Information, Figure S6).

The general lower EEQ values for environmental samples obtained with LC-MS/MS in comparison to bioassays correspond with findings of previous studies (Kienle et al., 2019; Könemann et al., 2018; Simon et al., 2022). Possible reasons are challenges emerging with low environmental concentrations of targets. Signals of target substances could be too weak or masked by the sample matrix. Furthermore, unknown mixture components beyond the six investigated estrogens might be present, resulting in an underestimation of estrogenicity by chemical analysis. Thus, LC-MS/MS might not have been sensitive enough to detect low concentrations as the LOQ of LC-MS/MS in the present study was 0.1–0.2 ng L^{-1} for E1, 0.2–1.5 ng L^{-1} for E2, and between 0.3 and 1.5 ng L^{-1} for EE2 compared to E2 equivalent concentration-based LOQs of 0.02–0.04 ng L^{-1} in A-YES, 0.01–0.05 ng L^{-1} in ERα-CALUX, and 0.04–0.06 ng L^{-1}in L-YES (individual results for the three sampling campaigns are shown in Supporting Information, Table S3).

The main steroidal estrogen found was E1 (Supporting Information, Figure S7 and Table S8). 17β-Estradiol and EE2 were only detected in a few samples. Their chemical analysis currently remains a challenge; often neither the very low required

quality criteria, nor the detection limits (especially for EE2) are reached (Arlos et al., 2018; Loos, 2012). Low concentrations of E1 could indicate that E2 and EE2 are present below their LOQ (Könemann et al., 2018). Thus, Könemann et al. (2018) suggested accounting for a potential contribution of nondetected estrogens to the overall EEQ_{chem} by replacing data <LOQ of individual substances with LOQ/2. This would shift the data points in Figure 4, particularly for L-YES, below the 1 to 1 line, indicating an overestimation of the estrogenicity (Supporting Information, Figure S10). Nevertheless, this approach could work if LOQs are fairly low. In conclusion, nondetected steroidal estrogens were possibly present in the samples, but likely in concentrations below LOQ/2.

Seasonal fluctuations in wastewater treatment performances were observed

The purpose of the present study was to evaluate bioassay effectiveness based on a large set of appropriate environmental samples and not to monitor WWTP effluents per se. However, some aspects are noteworthy. We noticed that ozonation significantly reduced the estrogencity, leading to the lowest effects observed in WWTPs with ozonation among the investigated WWTP samples. In Campaign 3, estrogenicity was reduced from 4.4 ng EEQ L^{-1} to concentrations below LOQ. In Campaign 2, the elimination efficiency by ozonation was approximately 60%. A possible explanation is that ozonation was not yet running at full power and estrogenicity was already negligible in these WWTP effluents prior to ozonation.

In general, estrogenicity in the investigated samples was low but present in all samples with some sites showing higher values up to 6.1 ng L^{-1}. This indicates that occasionally and in this set of WWTPs, estrogens can be problematic given current EBT values. Should future EBT values become lower, this might exacerbate the problem. A comparison of the different ISO standardized ERTA performances using river water would be more difficult because EEQs might be below the LOD of both ERTAs and chemical analysis due to dilution, sorption, and degradation processes.

The highest EEQ values were found in winter (Campaign 3) and the lowest values were observed in summer (Campaign 2). This could indicate that the season might play a role in the elimination of estrogenicity during wastewater treatment. However, further research would be needed to substantiate the seasonal dependence of removal efficiency.

In general, the proposed bioassay-independent threshold of 4 ng L^{-1} was only exceeded in three samples collected during Campaign 3 (winter) regarding the bioassay results. However, if a dilution of 10× cannot be met or assay-specific EBTs are applied, increased risks to the aquatic environment are conceivable. Thus, site-specific dilution factors would need to be considered in a study focusing on monitoring.

No correlation was observed between bioassay results and WWTP performance parameters such as wastewater quantity, temperature or nutrients. However, longer hydraulic retention times seemed to increase the elimination of estrogenicity (Supporting Information, Figure S11). This corresponds well with previous findings on links between estrogenicity and hydraulic retention time (Johnson et al., 2005; Vermeirssen et al., 2008).

CONCLUSION

The present study aimed to evaluate the effectiveness of three ISO standardized ERTAs using WWTP effluent samples. Investigating numerous effluents from different sources and covering possible seasonal fluctuations allowed the characterization of the three bioassays over different concentrations and mixtures of estrogenic compounds. Investigating the performance of the ERTAs at low environmentally relevant concentrations is important because potential problems in the detection of low EEQ are considered. It was shown that estrogenicity can be assessed with all three ERTAs leading to comparable EEQs in the WWTP effluent samples, although the assays are not identical in the context of practicability. Thus, all three assays are suitable for the monitoring of estrogenicity in this matrix. The ERα-CALUX data showed the lowest EC50 value for 17β-estradiol among the three ERTAs. However, in the case of high E1 concentrations dominating the mixture, yeast-based bioassays yield higher results than ERα-CALUX. With decreasing concentrations of estrogenicity in the samples, the differences between the individual ERTA results and thus the uncertainty increased. Comparing chemical analysis targeting major estrogenicity drivers with bioassay results showed that bioassays generally captured more estrogenicity. In addition, such comparisons were only robust for higher concentration ranges, because current chemical analytical methods are not sufficiently sensitive. The interpretation of the results regarding an assessment of water quality into acceptable and unacceptable classifications depended on the selected and applied threshold value. Differences in the number of samples exceeding a threshold were observed depending on the ERTA and whether a general cross-assay threshold or bioassay-specific EBTs were applied. Bioassay-specific EBTs are a promising tool to account for assay-specific differences that would otherwise not be reflected in the results. Several EBTs for ERα-CALUX were proposed, but further research is needed to establish robust assay-specific EBTs also for the other two ERTAs.

Supporting Information—The Supporting Information is available on the Wiley Online Library at https://doi.org/10.1002/etc.5445.

Acknowledgment—We thank the support of participating WWTP personnel for providing effluent samples. Open access funding provided by ETH-Bereich Forschungsanstalten.

Conflict of Interest—The authors declare no conflict of interest.

Author Contributions Statement—**Eszter Simon**: Conceptualization; Investigation; Project administration; Writing—review & editing. **Carolin Riegraf**: Formal analysis; Data curation;

Visualization; Writing—original draft. **Andrea Schifferli**: Investigation; Data curation; Writing—review & editing. **Daniel Olbrich**: Investigation; Writing—review & editing. **Thomas Bucher**: Investigation. **Etiënne L. M. Vermeirssen**: Conceptualization; Data curation; Supervision; Writing—review & editing.

Data Availability Statement—Additional data can be found in the Supporting Information. Data, associated metadata, and calculation tools are available from the corresponding author (carolin.riegraf@oekotoxzentrum.ch).

REFERENCES

Adeel, M., Song, X., Wang, Y., Francis, D., & Yang, Y. (2017). Environmental impact of estrogens on human, animal and plant life: A critical review. *Environment International*, *99*, 107–119. https://doi.org/10.1016/j.envint.2016.12.010

Archer, E., Wolfaardt, G. M., van Wyk, J. H., & van Blerk, N. (2020). Investigating (anti)estrogenic activities within South African wastewater and receiving surface waters: Implication for reliable monitoring. *Environmental Pollution*, *263*, 114424. https://doi.org/10.1016/j.envpol.2020.114424

Arlos, M. J., Parker, W. J., Bicudo, J. R., Law, P., Hicks, K. A., Fuzzen, M. L. M., & Servos, M. R. (2018). Modeling the exposure of wild fish to endocrine active chemicals: Potential linkages of total estrogenicity to field-observed intersex. *Water Research*, *139*, 187–197. https://doi.org/10.1016/j.watres.2018.04.005

Bain, P. A., Williams, M., & Kumar, A. (2014). Assessment of multiple hormonal activities in wastewater at different stages of treatment. *Environmental Toxicology and Chemistry*, *33*(10), 2297–2307. https://doi.org/10.1002/etc.2676

Been, F., Pronk, T., Louisse, J., Houtman, C., van der Velden-Slootweg, T., van der Oost, R., & Dingemans, M. M. L. (2021). Development of a framework to derive effect-based trigger values to interpret CALUX data for drinking water quality. *Water Research*, *193*, 116859. https://doi.org/10.1016/j.watres.2021.116859

Besselink, H., Brouwer, B., & van der Burg, B. (2017). Validation and regulatory acceptance of bio-based approaches to assure feedstock, water & product quality in a bio-based economy. *Industrial Crops and Products*, *106*, 138–145. https://doi.org/10.1016/j.indcrop.2016.11.026

Brack, W., Aït-Aïssa, S., Backhaus, T., Dulio, V., Escher, B. I., Faust, M., Hilscherova, K., Hollender, J., Hollert, H., Müller, C., Munthe, J., Posthuma, L., Seiler, T.-B., Slobodnik, J., Teodorovic, I., Tindall, A. J., de Aragão Umbuzeiro, G., Zhang, X., & Altenburger, R. (2019). Effect-based methods are key. The European Collaborative Project SOLUTIONS recommends integrating effect-based methods for diagnosis and monitoring of water quality. *Environmental Sciences Europe 31*(1):10. https://doi.org/10.1186/s12302-019-0192-2

Brand, W., de Jongh, C. M., van der Linden, S. C., Mennes, W., Puijker, L. M., van Leeuwen, C. J., & Heringa, M. B. (2013). Trigger values for investigation of hormonal activity in drinking water and its sources using CALUX bioassays. *Environment International*, *55*, 109–118. https://doi.org/10.1016/j.envint.2013.02.003

Brion, F., De Gussem, V., Buchinger, S., Hollert, H., Carere, M., Porcher, J.-M., & Aït-Aïssa, S. (2019). Monitoring estrogenic activities of waste and surface waters using a novel in vivo zebrafish embryonic (EASZY) assay: Comparison with in vitro cell-based assays and determination of effect-based trigger values. *Environment International*, *130*, 104896. https://doi.org/10.1016/j.envint.2019.06.006

Clouzot, L., Marrot, B., Doumenq, P., & Roche, N. (2008). 17α-Ethinylestradiol: An endocrine disrupter of great concern. Analytical methods and removal processes applied to water purification. A review. *Environmental Progress*, *27*(3), 383–396. https://doi.org/10.1002/ep.10291

Commission Implementing Decision (EU). (2018). 2018/840 of 5 June 2018 establishing a watch list of substances for Union-wide monitoring in the field of water policy pursuant to Directive 2008/105/EC of the European Parliament and of the Council and repealing Commission Implementing Decision (EU) 2015/495 (notified under document C(2018) 3362).

D'Agostino, R. B., & Stephens, M. A. (1986). *Goodness-of-fit, techniques*. Marcel Dekker, Routledge.

Dias, A. C. V., Gomes, F. W., Bila, D. M., Sant'Anna, G. L., & Dezotti, M. (2015). Analysis of estrogenic activity in environmental waters in Rio de Janeiro state (Brazil) using the yeast estrogen screen. *Ecotoxicology and Environmental Safety*, *120*, 41–47. https://doi.org/10.1016/j.ecoenv.2015.05.013

Escher, B. I., Allinson, M., Altenburger, R., Bain, P. A., Balaguer, P., Busch, W., & Leusch, F. D. L. (2014). Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays. *Environmental Science and Technology*, *48*(3), 1940–1956. https://doi.org/10.1021/es403899t

Escher, B. I., Aït-Aïssa, S., Behnisch, P. A., Brack, W., Brion, F., Brouwer, A., Buchinger, S., Crawford, S. E., Du Pasquier, D., Hamers, T., Hettwer, K., Hilscherová, K., Hollert, H., Kase, R., Kienle, C., Tindall, A. J., Tuerk, J., van der Oost, R., Vermeirssen, E., & Neale, P. A. (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. *Science of the Total Environment 628–629*:748–765. https://doi.org/10.1016/j.scitotenv.2018.01.340

Escher, B. I., & Neale, P. A. (2021). Effect-based trigger values for mixtures of chemicals in surface water detected with in vitro bioassays. *Environmental Toxicology and Chemistry*, *40*(2), 487–499. https://doi.org/10.1002/etc.4944

Escher, B. I., Neale, P. A., & Leusch, F. D. L. (2015). Effect-based trigger values for in vitro bioassays: Reading across from existing water quality guideline values. *Water Research*, *81*, 137–148. https://doi.org/10.1016/j.watres.2015.05.049

European Union. (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy. http://data.europa.eu/eli/dir/2000/60/oj

Gehrmann, L., Bielak, H., Behr, M., Itzel, F., Lyko, S., Simon, A., & Tuerk, J. (2018). (Anti-)estrogenic and (anti-)androgenic effects in wastewater during advanced treatment: Comparison of three in vitro bioassays. *Environmental Science and Pollution Research*, *25*(5), 4094–4104. https://doi.org/10.1007/s11356-016-7165-4

Hamers, T., Kamstra, J. H., Sonneveld, E., Murk, A. J., Visser, T. J., Van Velzen, M. J. M., & Bergman, Å. (2008). Biotransformation of brominated flame retardants into potentially endocrine-disrupting metabolites, with special attention to 2,2′,4,4′-tetrabromodiphenyl ether (BDE-47). *Molecular Nutrition & Food Research*, *52*(2), 284–298. https://doi.org/10.1002/mnfr.200700104

Hettwer, K., Jähne, M., Frost, K., Giersberg, M., Kunze, G., Trimborn, M., Reif, M., Tuerk, J., Gehrmann, L., Dardenne, F., De Crook, F., Abraham, M., Schoop, A., Waniek, J., Bucher, T., Simon, E., Vermeirssen, E., Werner, A., Hellauer, K., … Uhlig, S. (2018). Validation of Arxula Yeast Estrogen Screen assay for detection of estrogenic activity in water samples: Results of an international interlaboratory study. *Science of the Total Environment*, *621*, 612–625. https://doi.org/10.1016/j.scitotenv.2017.11.211

Houtman, C. J., ten Broek, R., & Brouwer, A. (2018). Steroid hormonal bioactivities, culprit natural and synthetic hormones and other emerging contaminants in waste water measured using bioassays and UPLC-tQ-MS. *Science of the Total Environment*, *630*, 1492–1501. https://doi.org/10.1016/j.scitotenv.2018.02.273

International Organization for Standardization. (2004). *ISO 6107-6 Water quality—Vocabulary—Part 6*.

International Organization for Standardization. (2018a). *ISO 19040-1 Water quality—Determination of the estrogenic potential of water and waste water—Part 1: Yeast estrogen screen (Saccharomyces cerevisiae)* (p. 55).

International Organization for Standardization. (2018b). *ISO 19040-2 Water quality—Determination of the estrogenic potential of water and waste water—Part 2: Yeast estrogen screen (A-YES, Arxula adeninivorans)* (p. 59).

International Organization for Standardization. (2018c). *ISO 19040-3 Water quality—Determination of the estrogenic potential of water and waste water—Part 3: In vitro human cell-based reporter gene assay* (p. 46).

International Organization for Standardization. (2022). *ISO 23196 Water quality—Calculation of biological equivalence (BEQ) concentrations*.

Itzel, F., Gehrmann, L., Bielak, H., Ebersbach, P., Boergers, A., Herbst, H., & Tuerk, J. (2017). Investigation of full-scale ozonation at a municipal wastewater treatment plant using a toxicity-based evaluation concept. *Journal of Toxicology and Environmental Health, Part A*, *80*(23–24), 1242–1258. https://doi.org/10.1080/15287394.2017.1369663

Jarošová, B., Bláha, L., Giesy, J. P., & Hilscherová, K. (2014). What level of estrogenic activity determined by in vitro assays in municipal waste waters can be considered as safe. *Environment International*, *64*, 98–109. https://doi.org/10.1016/j.envint.2013.12.009

Johnson, A. C., Aerni, H. R., Gerritsen, A., Gibert, M., Giger, W., Hylland, K., & Wettstein, F. E. (2005). Comparing steroid estrogen, and nonylphenol content across a range of European sewage plants with different treatment and management practices. *Water Research*, *39*(1), 47–58. https://doi.org/10.1016/j.watres.2004.07.025

Kaiser, C., Uhlig, S., Gerlach, T., Körner, M., Simon, K., Kunath, K., & Kunze, G. (2010). Evaluation and validation of a novel *Arxula adeninivorans* estrogen screen (nAES) assay and its application in analysis of wastewater, seawater, brackish water and urine. *Science of the Total Environment*, *408*(23), 6017–6026. https://doi.org/10.1016/j.scitotenv.2010.08.050

Kase, R., Javurkova, B., Simon, E., Swart, K., Buchinger, S., Könemann, S., Escher, B. I., Carere, M., Dulio, V., Ait-Aissa, S., Hollerr, H., Valsecchi, S., Polesello, S., Behnisch, P., di Paolo, C., Olbrich, D., Sychrova, E., Gundlach, M., Schlichting, R., … Werner, I. (2018). Screening and risk management solutions for steroidal estrogens in surface and wastewater. *Trends in Analytical Chemistry*, *102*, 343–358. https://doi.org/10.1016/j.trac.2018.02.013

Kienle, C., Vermeirssen, E. L. M., Schifferli, A., Singer, H., Stamm, C., & Werner, I. (2019). Effects of treated wastewater on the ecotoxicity of small streams—Unravelling the contribution of chemicals causing effects. *PLoS One*, *14*(12), e0226278. https://doi.org/10.1371/journal.pone.0226278

Könemann, S., Kase, R., Simon, E., Swart, K., Buchinger, S., Schlüsener, M., Hollert, H., Escher, B. I., Werner, I., Aït-Aïssa, S., Vermeirssen, E., Dulio, V., Valsecchi, S., Polesello, S., Behnisch, P., Javurkova, B., Perceval, O., Di Paolo, C., Olbrich, D., … Carere, M. (2018). Effect-based and chemical analytical methods to monitor estrogens under the European Water Framework Directive. *Trends in Analytical Chemistry*, *102*, 225–235. https://doi.org/10.1016/j.trac.2018.02.008

König, M., Escher, B. I., Neale, P. A., Krauss, M., Hilscherová, K., Novák, J., & Brack, W. (2017). Impact of untreated wastewater on a major European river evaluated with a combination of in vitro bioassays and chemical analysis. *Environmental Pollution*, *220*, 1220–1230. https://doi.org/10.1016/j.envpol.2016.11.011

Kunz, P. Y., Kienle, C., Carere, M., Homazava, N., & Kase, R. (2015). In vitro bioassays to screen for endocrine active pharmaceuticals in surface and waste waters. *Journal of Pharmaceutical and Biomedical Analysis*, *106*, 107–115. https://doi.org/10.1016/j.jpba.2014.11.018

Kunz, P. Y., Simon, E., Creusot, N., Jayasinghe, B. S., Kienle, C., Maletz, S., & Vermeirssen, E. L. M. (2017). Effect-based tools for monitoring estrogenic mixtures: Evaluation of five in vitro bioassays. *Water Research*, *110*, 378–388. https://doi.org/10.1016/j.watres.2016.10.062

Legler, J., Dennekamp, M., Vethaak, A. D., Brouwer, A., Koeman, J. H., van der Burg, B., & Murk, A. J. (2002). Detection of estrogenic activity in sediment-associated compounds using in vitro reporter gene assays. *Science of the Total Environment*, *293*(1), 69–83. https://doi.org/10.1016/S0048-9697(01)01146-9

Leusch, F. D. L., de Jager, C., Levi, Y., Lim, R., Puijker, L., Sacher, F., & Chapman, H. F. (2010). Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters. *Environmental Science and Technology*, *44*(10), 3853–3860. https://doi.org/10.1021/es903899d

Loos, R. (2012). *Analytical methods relevant to the European Commission's 2012 proposal on priority substances under the Water Framework Directive*. JRC Scientific and Policy Report.

Loos, R., Marinov, D., Sanseverino, I., Napierska, D., & Lettieri, T. (2018). *Review of the 1st Watch List under the Water Framework Directive and recommendations for the 2nd Watch List*. Publications Office of the European Union.

Luo, Y., Guo, W., Ngo, H. H., Nghiem, L. D., Hai, F. I., Zhang, J., & Wang, X. C. (2014). A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Science of the Total Environment*, *473–474*, 619–641. https://doi.org/10.1016/j.scitotenv.2013.12.065

McDonnell, D. P., Nawaz, Z., Densmore, C., Weigel, N. L., Pham, T. A., Clark, J. H., & O'Malley, B. W. (1991). High level expression of biologically active estrogen receptor in *Saccharomyces cerevisiae*. *Journal of Steroid Biochemistry and Molecular Biology*, *39*(3), 291–297. https://doi.org/10.1016/0960-0760(91)90038-7

McDonnell, D. P., Nawaz, Z., & O'Malley, B. W. (1991). In situ distinction between steroid receptor binding and transactivation at a target gene. *Molecular and Cellular Biology*, *11*(9), 4350–4355. https://doi.org/10.1128/mcb.11.9.4350

Murk, A. J., Legler, J., van Lipzig, M. M. H., Meerman, J. H. N., Belfroid, A. C., Spenkelink, A., & Vethaak, D. (2002). Detection of estrogenic potency in wastewater and surface water with three in vitro bioassays. *Environmental Toxicology and Chemistry*, *21*(1), 16–23. https://doi.org/10.1002/etc.5620210103

Purvis, I. J., Chotai, D., Dykes, C. W., Lubahn, D. B., French, F. S., Wilson, E. M., & Hobden, A. N. (1991). An androgen-inducible expression system for *Saccharomyces cerevisiae*. *Gene*, *106*(1), 35–42. https://doi.org/10.1016/0378-1119(91)90563-Q

R Development Core Team. (2008). *R: A language and environment for statistical computing*.

Routledge, E. J., & Sumpter, J. P. (1996). Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. *Environmental Toxicology and Chemistry*, *15*(3), 241–248. https://doi.org/10.1002/etc.5620150303

Simon, E., Duffek, A., Stahl, C., Frey, M., Scheurer, M., Tuerk, J., & Vermeirssen, E. L. M. (2022). Biological effect and chemical monitoring of Watch List substances in European surface waters: Steroidal estrogens and diclofenac–Effect-based methods for monitoring frameworks. *Environment International*, *159*, 107033. https://doi.org/10.1016/j.envint.2021.107033

Simon, E., Lamoree, M. H., Hamers, T., Weiss, J. M., Balaam, J., de Boer, J., & Leonards, P. E. G. (2010). Testing endocrine disruption in biota samples: A method to remove interfering lipids and natural hormones. *Environmental Science and Technology*, *44*(21), 8322–8329. https://doi.org/10.1021/es101912z

Simon, E., Schifferli, A., Bucher, T. B., Olbrich, D., Werner, I., & Vermeirssen, E. L. M. (2019). Solid-phase extraction of estrogens and herbicides from environmental waters for bioassay analysis—Effects of sample volume on recoveries. *Analytical and Bioanalytical Chemistry*, *411*(10), 2057–2069. https://doi.org/10.1007/s00216-019-01628-1

Simon, K., Kunze, G., Tag, K., Steinborn, G., Körner, M., & Hanke, G. (2006). *Biosensor* (EP 08760977). Patent application 2006.

Sonneveld, E., Jansen, H. J., Riteco, J. A. C., Brouwer, A., & van der Burg, B. (2004). Development of androgen- and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid-responsive bioassays. *Toxicological Sciences*, *83*(1), 136–148. https://doi.org/10.1093/toxsci/kfi005

Van den Belt, K., Berckmans, P., Vangenechten, C., Verheyen, R., & Witters, H. (2004). Comparative study on the in vitro/in vivo estrogenic potencies of 17β-estradiol, estrone, 17α-ethynylestradiol and nonylphenol. *Aquatic Toxicology*, *66*(2), 183–195. https://doi.org/10.1016/j.aquatox.2003.09.004

van der Oost, R., Sileno, G., Suárez-Muñoz, M., Nguyen, M. T., Besselink, H., & Brouwer, A. (2017). SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: Part I–model design and effect-based trigger values. *Environmental Toxicology and Chemistry*, *36*(9), 2385–2399. https://doi.org/10.1002/etc.3836

Vermeirssen, E. L. M., Eggen, R. I. L., Escher, B. I., & Suter, M. J. F. (2008). Estrogens in Swiss Rivers and Effluents—Sampling Matters. *CHIMIA International Journal for Chemistry*, *62*(5), 389–394. https://doi.org/10.2533/chimia.2008.389

Wagner, M., Vermeirssen, E. L. M., Buchinger, S., Behr, M., Magdeburg, A., & Oehlmann, J. (2013). Deriving bio-equivalents from in vitro bioassays: Assessment of existing uncertainties and strategies to improve accuracy and reporting. *Environmental Toxicology and Chemistry*, *32*(8), 1906–1917. https://doi.org/10.1002/etc.2256

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Yao, B., Li, R., Yan, S., Chan, S.-A., & Song, W. (2018). Occurrence and estrogenic activity of steroid hormones in Chinese streams: A nationwide study based on a combination of chemical and biological tools. *Environment International*, *118*, 1–8. https://doi.org/10.1016/j.envint.2018.05.026

