Validation of Arxula Yeast Estrogen Screen assay for detection of estrogenic activity in water samples : results of an international interlaboratory study

Validation of *Arxula* Yeast Estrogen Screen assay for detection of estrogenic activity in water samples: Results of an international interlaboratory study

Karina Hettwer [a], Martin Jähne [a], Kirstin Frost [a], Martin Giersberg [b], Gotthard Kunze [b], Michael Trimborn [c], Martin Reif [c], Jochen Türk [d], Linda Gehrmann [d], Freddy Dardenne [e], Femke De Croock [e], Marion Abraham [f], Anne Schoop [f], Joanna J. Waniek [f], Thomas Bucher [g], Eszter Simon [g], Etienne Vermeirssen [g], Anett Werner [h], Karin Hellauer [i], Ursula Wallentits [i], Jörg E. Drewes [i], Detlef Dietzmann [j], Edwin Routledge [k], Nicola Beresford [k], Tamara Zietek [l], Margot Siebler [l], Anne Simon [m], Helena Bielak [m], Henner Hollert [n], Yvonne Müller [n], Maike Harff [n], Sabrina Schiwy [n], Kirsten Simon [o], Steffen Uhlig [a,*]

a QuoData GmbH, Prellerstr. 14, 01309 Dresden, Germany

b Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466 Seestadt OT Gatersleben, Germany

c Erftverband, Am Erftverband 6, 50126 Bergheim, Germany

d Institut für Energie- und Umwelttechnik e. V. (IUTA, Institute of Energy and Environmental Technology), Bliersheimer Str. 58-60, 47229 Duisburg, Germany

e University of Antwerp, Systemic Physiological and Ecotoxicological Research (SPHERE), Groenenborgerlaan 171/U7, 2020 Antwerp, Belgium

f Leibniz Institute for Baltic Sea Research Warnemünde, Department Marine Chemistry, Seestraße 15, 18119 Rostock, Germany

g Swiss Centre for Applied Ecotoxicology Eawag-EPFL, Überlandstraße 133, 8600 Dübendorf, Switzerland

h Technical University Dresden, Institute of Natural Science, Bioprocess Engineering, Helmholtzstraße 10, 01062 Dresden, Germany

i Technical University of Munich, Department of Civil, Geo and Environmental Engineering, Chair of Urban Water Systems Engineering, Am Coulombwall 3, 85748 Garching, Germany

j SYNLAB Umweltinstitut GmbH, Hauptstraße 105, 04416 Markkleeberg, Germany

k Brunel University London, Institute for Environment, Health and Societies, Halsbury Building, UB8 3PH Uxbridge, United Kingdom

l Technical University of Munich, Department of Nutritional Physiology, Gregor-Mendel-Straße 2, 85354 Freising, Germany

m IWW Rheinisch-Westfälisches Institut für Wasserforschung gemeinnützige GmbH, Moritzstr. 26, 45476 Mülheim an der Ruhr, Germany

n RWTH Aachen University, Institute for Environmental Research, Worringerweg 1, 52074 Aachen, Germany

o New diagnostics GmbH, Pollinger Straße 11, 81377 München, Germany

(*) Corresponding author: Steffen Uhlig, QuoData GmbH, Prellerstr. 14, 01309 Dresden, Germany, Email: uhlig@quodata.de, Phone: +49 351 40 28 867 0, Fax: +49 351 40 28 867 19.

Abstract

Endocrine-active substances can adversely impact the aquatic ecosystems. A special emphasis is laid, among others, on the effects of estrogens and estrogen mimicking compounds. Effect-based screening methods like *in vitro* bioassays are suitable tools to detect and quantify endocrine activities of known and unknown mixtures.

This study describes the validation of the *Arxula*-Yeast Estrogen Screen (A-YES®) assay, an effect-based method for the detection of the estrogenic potential of water and waste water. This reporter gene assay, provided in ready to use format, is based on the activation of the human estrogen receptor alpha. The user-friendly A-YES® enables inexperienced operators to rapidly become competent with the assay.

Fourteen laboratories from four countries with different training levels analyzed 17β-estradiol equivalent concentrations (EEQ) in spiked and unspiked waste water effluent and surface water samples, in waste water influent and spiked salt water samples and in a mixture of three bisphenols. The limit of detection (LOD) for untreated samples was 1.8 ng/L 17β-estradiol (E2). Relative repeatability and reproducibility standard deviation for samples with EEQ above the LOD (mean EEQ values between 6.3 and 20.4 ng/L) ranged from 7.5 to 21.4% and 16.6 to 28.0%, respectively. Precision results are comparable to other frequently used analytical methods for estrogens.

The A-YES® has been demonstrated to be an accurate, precise and robust bioassay. The results have been included in the ISO draft standard.

The assay was shown to be applicable for testing of typical waste water influent, effluent and saline water. Other studies have shown that the assay can be used with enriched samples, which lower the LOD to the pg/L range.

The validation of the A-YES® and the development of a corresponding international standard constitute a step further towards harmonized and reliable bioassays for the effect-based analysis of estrogens and estrogen-like compounds in water samples.

Keywords

Yeast Estrogen Screen; Effect-based monitoring tools; Estrogenicity; Interlaboratory trial; ISO; Standardization

1 Introduction

Endocrine-disrupting compounds (EDC) pose a global threat to human health and the environment (Bergman et al., 2012). Concerns about the effects of natural estrogens and estrogen mimics to the environment stem from their widespread use in personal care products (Karpuzoglu et al., 2013), industrial chemicals (Rochester et al., 2015), livestock breeding and agriculture (Gall et al., 2011) and human pharmaceutical products (Maitre 2013). Several examples include the use of estrogens in contraceptives, hormonal replacement therapy or for treating menopausal and post-menopausal symptoms. Depending on their use patterns, EDCs enter aquatic ecosystems from different human and animal sources (Adeel et al. 2017). The elimination of estrogenic micro-pollutants in conventional waste water treatment plants is incomplete so far and effluents are one of the major sources of estrogens and estrogenic transformation products into watercourses (Filby et al., 2007; Racz and Goel, 2010; Gehrmann et al., 2016). Because of the hormone-like action of estrogens and estrogen acting compounds and their ubiquity in environment, a number of efforts have been made to investigate their impacts on the ecosystems. To dates, there are numerous laboratory and field studies showing the adverse effects of natural estrogens and estrogenic EDCs in marine organisms that result in altered reproductive output in gastropods and altered sexual maturation in wild roach (Oehlmann et al., 2006; Benstead et al., 2011; Jobling et al., 2002; Jobling and Tyler, 2003; Lange A. et al., 2008).

To address the problem of EDCs and micro-pollutants in general, the European Union Water Framework Directive (WFD) was adopted in 2000 (European Commission 2000). The WFD aims to improve overall surface water quality by (i) achieving a good ecological and chemical status for surface waters and (ii) a good chemical and quantitative status for ground waters in Europe. The Article 8 of the WFD calls for the necessity of monitoring programs to establish a coherent and comprehensive overview of the water status within each river basin. Typically targeted chemical analysis strategies are applied for monitoring the chemical status. These methods are focused on the type and concentration of chemicals in a sample, and can precisely identify single known substances within environmental samples. However, they require *a priori* knowledge about the substances to be monitored. For technical and economic reasons, it is not possible to analyze, detect and quantify all substances that are present in the aquatic environment (Wernersson et al., 2015). Chemical monitoring is, therefore, performed instead only on regulated and harmful substances.

Unlike other analytical techniques, effect-based bioassays present a different approach to monitor water quality. Using bioassays offers a possible assessment of the cumulative effect of all compounds present in a sample (Silva et al., 2002). Thus, the effect of both known and unknown substances is captured (Leusch et al., 2014; Kunz et al., 2015).

Bioassays can complement a chemical analysis in various ways. They can provide additional information, or can be used as screening tools that initiate a chemical analysis in case of positive results, or can be used if routine chemical methods do not reach the required detection limits. For example, the environmentally relevant concentrations for the estrogens 17β-estradiol (E2) and 17α-ethinylestradiol (EE2) are in the low ng/L or pg/L range (Kunz et al., 2015). Here, routine analytical methods are often not sensitive enough, allowing bioassays to be used to quantify the estrogenic activity of a water sample in terms of a 17β-estradiol equivalent concentration (EEQ) (Kunz et al., 2015).

Realizing the problem of chemical mixtures in water quality monitoring, the CMEP (Chemical Monitoring and Emerging Pollutant) group, which acts in the context of Common Implementation Strategy (CIS) of the WFD, started activities on effect-based bioassays and published a technical report on aquatic effect-based monitoring tools (Wernersson et al., 2015). Beyond this report, several other recommendations for integration of bioassays in water quality monitoring are also available (Brack et al., 2016; Brack et al., 2017; Altenburger et al., 2015; Hecker and Hollert, 2011; Hamers et al., 2013; Di Paolo et al., 2016, Tousova et al. 2017).

A prerequisite for the implementation of effect-based tools in water quality monitoring is the availability of reliable and standardized test methods. On an international level, the ISO (International Organization for Standardization) is the relevant body for the development of standards. The scope of the ISO Technical Committee 147 is standardization in the field of water quality with emphasis on environmental issues. In this context, the standardization of assays for the determination of the estrogenic potential of water and waste water were included in the working program of ISO TC 147 in 2013, with *Arxula* Yeast Estrogen Screen (A-YES®) assay among them (ISO/DIS 19040-2) (ISO Standards Catalogue).

The present study describes the validation of the A-YES® within the framework of an interlaboratory trial in the context of the standardization process. This study is the first international interlaboratory validation of the A-YES®. The interlaboratory trial was organized by the working group “Genotoxicity and endocrine effects” of the DIN Standardization Committee 119-01-03-05-09. The samples comprised river water, influent and effluent of a municipal waste water treatment plant, influent of a hospital treatment plant, saline water, a sample spiked with three bisphenols (Z, S, A) and a negative control. The results of the study have been included in the draft standard (ISO/DIS 19040-2) to demonstrate the applicability of the method and to provide validation data. The ISO standard will most likely be published in spring 2018.

The validation of the A-YES® and the development of the corresponding international standard constitute a step further towards harmonized and reliable bioassays for the effect-based analysis of estrogens and estrogen-like compounds in water samples.

2 Material and methods

2.1 Coordination and participants

The study was organized by the working group “Genotoxicity and endocrine effects” of the DIN. Coordination of the interlaboratory trial for ISO/CD 19040-2 was done by QuoData GmbH with major support from the German Federal Institute of Hydrology (BfG).

Fourteen laboratories from four countries were registered for the interlaboratory trial (eleven from Germany, one each from Switzerland, Belgium and the United Kingdom). Laboratories included private companies, universities and public authorities. The laboratories had different training statuses; more than half of them performed only a few A-YES® tests prior to the start of the study. These tests consisted of the analysis of water samples according to the same testing procedure of the interlaboratory trial.

2.2 Samples and pretesting

Nine samples were analyzed: waste water, surface water, saline water, a mixture of bisphenols and a control sample (Table 1). A broad range of samples was chosen to cover the scope of the ISO/DIS 19040-2 method. As the yeast *Arxula adeninivorans* is highly salt tolerant (Gellissen 2005), a saline water sample was chosen to demonstrate the applicability of the assay for such water types. Additional information regarding the sampling sites, sample volumes, spiking and waste water treatment plants, is available in the supplementary information (SI).

A pre-test was performed to ensure that all participating laboratories received a set of good quality samples. The samples were tested in serial dilutions according to the testing scheme of the interlaboratory trial. In all tests, the samples spiked with E2, EE2 and the bisphenols showed estrogenic activity in the quantifiable range of the assay, and samples expected to show no estrogenic effect behaved accordingly.

Experiments analyzing the stability of samples of the same origin were performed in preparation of the interlaboratory trial. The effects of storage and shipping temperature and filtration were examined. Based on the results of the stability testing the samples were stored at ≤ -18°C before and after shipping. Shipping was done within 24 hours at no more than 2 °C and the temperature was recorded during shipping. Amber glass bottles (40 ml capacity) with polytetrafluoroethylene (PTFE) lined caps were used as sample containers.

Although filtration of influent samples leads to a better stability, it also causes a significant reduction in estrogenic activity. Therefore, coarse particles were removed by sedimentation at low centrifugation speeds (20 min at 3500 g at room temperature).

In addition, several tests were performed prior to the interlaboratory study to verify the inhibition potential of different kinds of samples on the reporter enzyme phytase. No significant reduction of enzymatic activity was observed for samples with high matrix load (e. g. saline and waste water samples).

Table 1: Samples analyzed with the *Arxula* Yeast Estrogen Screen assay in the interlaboratory trial. Samples were provided by the BfG (#1, #2, #3, #5, #6, #9), the Institute of Energy and Environmental Technology, IUTA (#4), the Leibniz Institute for Baltic Sea Research, IOW (#8) and QuoData GmbH (#7).

Nr.	Type	Description	Spiking
1	Waste water	Effluent, municipal-WWTP, Koblenz	--
2		Effluent, municipal-WWTP, Koblenz	10 ng/L EE2
3		Influent, municipal-WWTP, Koblenz, (centrifuged)	--
4		Influent, hospital-WWTP, Gelsenkirchen (settled)	--
5	Surface (river) water	Rhine	--
6		Rhine	15 ng/L EE2
7	Aqueous mixture of bisphenols	Mixture of BPA, BPS, BPZ	242 µg/L BPA, 48 µg/L BPS, 39 µg/L BPZ
8	Saline water	Baltic Sea	1.4 ng/L E2 and 8 ng/L EE2
9	Deionized water	Field blank (control sample)	--

2.3 Materials for participants

Each of the participants received three sets of samples. Each set contained one aliquot of the nine samples (coded). The participants were also informed about the storage requirement of the samples after arrival (≤ -18°C), as well as the testing procedure. Each participant was kindly requested to analyze at least two sample sets.

All reagents and consumables for the A-YES® were provided by new diagnostics GmbH (München, Germany) and distributed to the participants by QuoData GmbH. The material included freeze-dried *A. adeninivorans* G1214/YRC103-hERα-phyK, E2 stock solution (1 mg/L in ethanol), sterile yeast minimal medium (fivefold concentrated), sterile saline yeast minimal medium (fivefold concentrated, containing 140 g/L sodium chloride), substrate buffer (0.5 mol/L citric buffer), developer (3 mol/L NaOH), substrate tablets (5 mg p-nitrophenyl phosphate disodium salt hexahydrate, each), 0.1 mol/L HCl, 0.1 mol/L NaOH, deep well plates, microtiter plates and air permeable foil. All of the provided materials (yeast, reagents, media and buffer) for the interlaboratory trial belonged to the same batch. Participants were instructed to store the test material and reagents under cool conditions (2 – 8 °C).

2.4 Testing of samples with *Arxula* Yeast Estrogen Screen assay

The tests were performed according to the method described in ISO/DIS 19040-2, "Water quality – Determination of the estrogenic potential of water and waste water" Part 2: Yeast estrogen screen (A-YES, *Arxula adeninivorans*).

The A-YES® is a reporter gene assay for the measurement of the activation of the human estrogen receptor alpha (hERα) in the presence of a sample containing compounds which cause estrogenic effects. The assay detects the estrogenic activity of the whole sample, including possible additive, synergistic and antagonistic mixture-effects.

The *human estrogen receptor alpha* gene is constitutively expressed in the yeast cell under the control of a *TEF1* promoter. Upon entering the yeast cells, agonists of the estrogen receptor will bind to the estrogen receptor protein, forming the receptor-ligand-dimers that subsequently induce the transcription of the reporter gene *phyK*, which encodes the enzyme phytase. The activity of phytase as an indicator for the estrogenic potential of the sample is determined by using an appropriate substrate which is cleaved into a final product of reaction. This product, having a distinct color, can be detected photometrically. Quantification is carried out through calibration with E2.

The E2 stock solution (1 mg/L in ethanol) was diluted with ultrapure water according to ISO/DIS 19040-2 to obtain calibration levels with the following E2 concentrations: 80 ng/L, 40 ng/L, 20 ng/L, 8 ng/L 4 ng/L, 2 ng/L, 1 ng/L.

The conductivity and pH of the samples were measured in an aliquot of each sample with a pH/conductivity measuring device. Adjustment of the pH was done with 0.1 mol/L HCl or 0.1 mol/L NaOH. To avoid contamination of the sample by the pH/conductivity device, the determined volume of a titrated base or acid solution was added to another aliquot of the sample.

For the analysis of samples using the ISO/DIS 19040-2 method, the samples were diluted with ultrapure water or with ultrapure water adjusted with sodium chloride to the same conductivity as the sample. The analyzed dilution levels for each sample are summarized in Table 2. The dilution scheme was in accordance with the guidelines of the German waste

water ordinance (Waste Water Ordinance, 2004).

Table 2: Sample dilutions analyzed with the *Arxula* Yeast Estrogen Screen assay in the interlaboratory trial.

Dilution level without yeast suspension	Dilution level with yeast suspension	Volume (µl)			Samples tested in this dilution
		Sample	Dilution water	Yeast suspension	
1.0	1.25	400	0	100	samples 1 to 9
1.6	2.0	250	150	100	samples 1 to 9
2.4	3.0	133.3	266.7	100	samples 1 to 6 and 9
3.2	4.0	100	300	100	samples 1 to 9
4.8	6.0	66.7	333.3	100	samples 1 to 6 and 9
6.4	8.0	50	350	100	samples 1 to 9
9.6	12.0	33.3	366.7	100	samples 1 to 6 and 9

The samples were subsequently analyzed with the A-YES® according to the following procedure:

Freeze dried yeast cells were washed three times with diluted minimal medium with alternating centrifugation (3000 g) and resuspension steps. The washed cells were reactivated 1 h at 30 °C on an incubation shaker with appropriate shaking frequency. The deep well plate was filled with the negative control (ultrapure water), blanks, the dilution series of the samples and the calibration levels of the standard (E2). Replicates were as follows: two replicates for the negative control and the reference standard; four replicates for each dilution level of the sample; two replicates for the sample blank (undiluted sample); two replicates for the blank of the negative control and two replicates for the blank of the E2 concentration 80 ng/L. Based on the conductivity of the samples, fivefold minimal medium for inoculation was prepared using fivefold minimal medium with and without sodium chloride to obtain a similar salinity in all 96 wells per plate. To assess sample associated microbial growth and sample coloration blanks were inoculated with 100 µl of prepared fivefold minimal medium without yeast. The reactivated yeast was diluted with the prepared fivefold minimal medium and all wells except the blanks were inoculated with yeast medium suspension with the same cell density per well. The deep well plate was sealed with an air permeable foil and incubated 24 ± 2 h in an incubation shaker at 31 ± 1 °C with an appropriate shaker frequency depending on the laboratory. After an overnight cultivation, the yeast cells were separated from the culture by centrifugation (10 min, 700 g) and 50 µL of the clear supernatant of each well was transferred into a clear microtiter plate. Substrate solution for the reporter enzyme phytase was prepared by adding 5 mg p-nitrophenyl phosphate disodium salt hexahydrate tablets (Sigma-Aldrich) to 0.5 mol/L citric buffer to reach a concentration of 1 mg/mL. A 50-µL substrate solution was then added into each well of the microtiter plate filled with supernatant followed by a brief centrifugation step (10 s, 100 g). The plate was incubated at 37 °C for one hour and the absorption was determined at 405 nm in a microplate reader afterwards. Next, 100 µL of developing solution (3 mol/L NaOH) were pipetted to every well and the absorption of the colored product was determined at 405 nm again. The deep well plate with the remaining cell pellet was shaken rigorously to resuspend the yeast cells, and an aliquot of the yeast cell suspension of every well was transferred and diluted into a second microtiter plate with defined volume of ultrapure water. The cell density was determined by measuring

the optical density at 630 nm (alternatively at 600 or 620 nm).

2.5 Evaluation of raw data

Dose-response curve and sample EEQ

For each test performed (each plate), the raw data were analyzed as described in ISO/DIS 19040-2: raw data were analyzed for outliers, growth effects and sample coloration effects, and the corrected absorbance and induction rates were calculated.

Subsequently, the dose-response relationship for the E2 reference compound was calculated with a four-parameter logistic model (Hill 1910) (see also supplementary information).

The dilution-specific EEQ and the EEQ of the undiluted sample were calculated. To obtain the EEQ for the undiluted sample, the EC_{20} approach was followed, namely, the 20% effect level of the sample was interpolated from the 20% effect level of the E2 reference. The EEQ concentration for the undiluted sample was derived by dividing the E2 concentration needed to obtain 20% effect by the dilution level of the sample needed to obtain the 20% effect.

The caveat for this approach is that the relative growth of the tested sample dilutions has to be greater than 0.85 (compared to negative control and the E2 reference), thus this conservative method may restrict the number of usable results. Nevertheless, it ensures that the estrogenic effect of a sample is not significantly overestimated due to the growth correction.

Apparent recovery

The recovery was calculated for the samples 2, 6, 7 and 8, considering the relative potency of the A-YES® for the spiked compounds EE2, BPA, BPS and BPZ.

The relative potency is used to express the biological activity of a compound in comparison to the natural hERα ligand and calibration standard E2. It is defined by a ratio between the EC_{50} of E2 and the EC_{50} of the respective compound. The relative potencies of E2, EE2, BPA, BPS and BPZ are 1, 1.2, $1.68 * 10^{-5}$, 0 and $1.23 * 10^{-4}$, respectively. The expected EEQ of the samples 2, 6, 7 and 8 was calculated by multiplying the spiked concentration with the relative potency and summation of the result for all spiked compounds (see also supplementary information).

The recovery of a sample is defined as a ratio of the overall empirical mean and the expected mean. Subsequently the term apparent recovery is used to distinguish the recovery associated with an extraction procedure from the recovery without extraction (Burns et al., 2002).

Limit of detection (LOD), critical concentration (CC) and limit of quantification (LOQ)

The limit of detection (LOD), the critical concentration (CC) and the limit of quantification (LOQ) are performance parameters that were also determined. LOD and CC reflect the concentration at which a sample or a sample dilution is shown to mediate estrogenic activity. The limits take different sources of variation into account. Calculation of LOD is based on the plate-specific variability of the negative control and the E2 calibration levels, thus the uncertainty of the underlying four-parameter model is considered. The critical concentration is the concentration at the induction rate 1.18, which corresponds to 1.18 times the bottom curve point of the four-parameter dose response curve. This limit was determined based on long-term variability of the negative control of laboratories having a long-term experience with

the A-YES®. The LOQ is the concentration at which the calculated EEQ has an uncertainty of not more than 33.3%. The calculation is based on the concentration's 95 %-prediction interval, which is determined using the bootstrap method (Efron 1979). The calculated LOD, CC and LOQ refer to the concentration in the sample.

Validity

Prior to the evaluation of method performance, the validity of each test (each plate) was checked according to following criteria (as stated in Clause 10 of ISO/DIS 19040-2).

The following criteria relating to the negative control, the reference standard and certain characteristics of the dose-response curve were checked for: 1) EC_{50} (half maximal effective concentration) between 7 and 35 ng/L E2, 2) Critical concentration equal or below 8 ng/L E2, 3) Ratio between upper and lower level of dose-response relation ≥ 4 and ≤ 20, 4) Calibration range coverage ≥ 75%, 5) Variabilities of residuals ≤ 12% and 6) Final mean cell density after overnight incubation ≥ 1. If any of the above mentioned criteria were not met, the entire test was considered invalid.

In addition, the relative growth and variability of corrected absorbance at 405 nm for each dilution level were checked as sample-specific criteria with thresholds ranging from ≥ 0.3 to ≤ 3.5 and ≤ 15%, respectively. If one of the sample specific criteria was not met (even after outlier elimination), the respective dilution level was excluded from further statistical evaluation.

All calculations described in this section were performed using the software GAUSS (Aptech Systems, Chandler (AZ), USA), a matrix programming language for mathematics and statistics.

2.6 Evaluation of method performance

Statistical method validation was carried out according to ISO 5725-2 (ISO 5725-2 2002) using the software PROLab (QuoData GmbH, Dresden, Germany). Repeatability (s_r) and Reproducibility (s_R) standard deviations were calculated for the EEQ for each of the analyzed dilution levels and for the undiluted sample. The formulas for calculation of s_r and s_R are given as supplementary information.

Beyond the precision analysis for the sample EEQs, the following validation criteria were taken into consideration: sensitivity in terms of limit of detection (LOD), critical concentration (CC), limit of quantification (LOQ), EC_{50} (half maximal effective concentration) and apparent recovery.

The scheme of validation procedure of sample analysis with the A-YES®, the subsequent evaluation of the raw data and the final evaluation of method performance are shown in Figure 1.

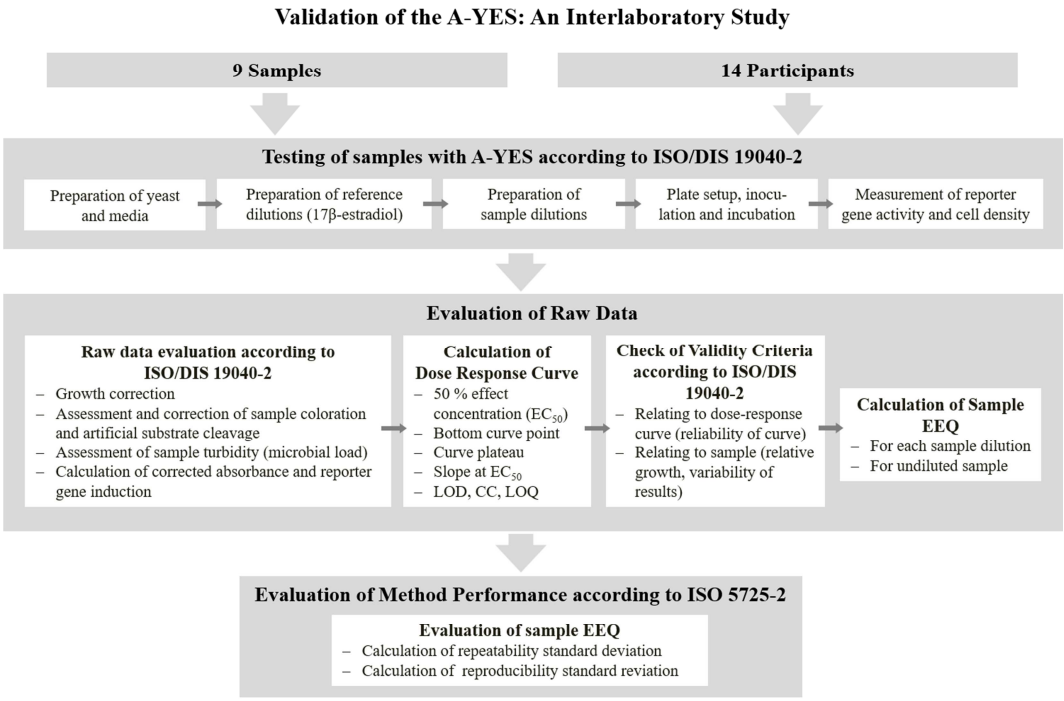

Fig. 1: Flow chart of the validation procedure.

3 Results

3.1 Useable results

Six of the 14 participating laboratories had a previous experience with the A-YES®. The remaining eight laboratories successfully established the A-YES® test within a few weeks before start of the interlaboratory trial.

One laboratory failed to follow the test procedure and results were completely excluded from further analysis. There were problems in determining the yeast cell density after incubation because the yeast pellet was not sufficiently resuspended. As a result, the cell density varied considerably, making the evaluation and correction of growth effects not possible.

For the 13 remaining laboratories, 45 out of 256 results were excluded from the statistical analysis according to ISO 5725-2. More detailed information relating to usable results is summarized in Table 3. For each sample, the number of excluded results and corresponding reasons are provided. Results were excluded because either the E2 dose-response curve was not reliable, or pipetting errors occurred, or the relative growth of a sample was too low.

Relating to the E2 dose-response curves, there were various causes that led to the exclusion of the results. On the one hand the variability of results was too high (residual standard deviation (RSD) > 12%), or the curve did not reach the upper plateau on the other hand (calibration range coverage (CRC) < 75%), or both were the case. An incomplete dose-response curve does not appear to be a reason to exclude the results completely. However, it does not allow a reliable estimation of the EC_{50}, which must be within a certain acceptable range. Causes for a variability of single results being too large and an incomplete dose-response curve may result from the following possibilities: carryover of yeast when transferring the clear supernatant from the deep well plate to a microliter plate; pipetting errors during preparation of the E2 dilution series, also by using different pipettes; air bubbles in the supernatant; inappropriate shaker frequencies and unsuitable temperature during the incubation phase. The complete 96-well plate was excluded, regardless of the results from the samples.

In line with the pipetting errors, it has occasionally occurred that yeast was mistakenly added into microtiter wells not designated for yeast. These so-called blank wells are required to evaluate and correct the inherent coloration of the sample and artificial cleavage of the enzyme substrate. Within the interlaboratory trial, samples with missing blanks were excluded from further analysis in order to ensure a uniform evaluation procedure. In the worst scenario, ignoring the blanks leads to an overestimation of the estrogenic activity of a sample.

In a few cases, the relative growth of the analyzed sample dilutions was below the threshold value of 85%. Those samples were excluded from further analysis because the inhibition of yeast cell growth by the sample was too strong (cytotoxic effects).

The total proportion of results excluded due to an unreliable E2 dose-response curve, pipetting errors and sample-related insufficient relative growth were 11.3%, 3.1% and 3.1% (in the total of 17.6%), respectively. When the training level of the participants was taken into account, the results for experienced laboratories were 5.1%, 2.7% and 1.6% (in the total of 9.4%), and for unexperienced laboratories 6.3%, 0.4% and 1.6% (in the total of 8.2%). There is a slightly higher overall exclusion rate for experienced participants caused by more

frequent pipetting errors, specifically by missing sample blanks. However, the difference is insignificant.

The largest number of invalid results was obtained for the unspiked effluent (#1) and the hospital-influent sample (#4) with 21.4% and 31.0%, respectively. In both cases, more than half of invalid results were due to non-reliable E2 dose-response curves. For the hospital-influent sample, one third of the invalid results were caused by insufficient growth of the yeast cells in the analyzed sample dilutions.

After elimination of invalid results, the method performance according to ISO 5725-2 was evaluated based on 82.4% of all results (211 out of 256 single results). Within this evaluation, outlier tests for single measurements and for laboratory mean values were performed. As a result, all values of laboratory 05 for sample #7 were excluded because the intra-laboratory variability was too high.

Table 3: Overview of number and validity of measurements for the analyzed samples. meas. = measurements; LOD = limit of detection.

			Sample								
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Laboratories			13	13	13	13	13	13	13	13	13
Measurements			28	28	28	29	29	28	29	29	28
Laboratories with valid results			13	13	12	12	13	12	12	12	13
Valid measurements			22	25	25	20	25	23	24	24	23
Invalid meas.	related to dose-response curve	RSD > 12 % and / or CRC < 75 %	4	2	2	5	3	2	4	4	3
	pipetting error	sample blank missing	1	1	1	1	1	1	0	1	1
	related to sample	relative growth < 0.85	1	0	0	3	0	2	1	0	1
	total		6	3	3	9	4	5	5	5	5
Invalid meas. (%)	related to dose-response curve	RSD > 12 % and / or CRC < 75 %	14.3	7.1	7.1	17.2	10.3	7.1	13.8	13.8	10.7
	pipetting error	sample blank missing	3.6	3.6	3.6	3.4	3.4	3.6	0.0	3.4	3.6
	related to sample	relative growth < 0.85	3.6	0.0	0.0	10.3	0.0	7.1	3.4	0.0	3.6
	total		21.4	10.7	10.7	31.0	13.8	17.9	17.2	17.2	17.9
Outliers excluded - single meas. acc. to ISO 5725-2			0	0	0	0	0	0	0	0	0
Outliers excluded - laboratory means acc. to ISO 5725-2			0	0	0	0	0	0	1	0	0
Laboratories with valid results after outlier elimination			13	13	12	12	13	12	11	12	13
Valid measurements after outlier elimination			22	25	25	20	25	23	22	24	23
Measurements <LOD			17	1	0	0	22	0	1	0	23
Measurements >LOD			5	24	25	20	3	23	23	24	0

3.2 Sensitivity

Dose-response curves for the reference standard E2 were obtained for each assay plate from each participating laboratory, in total for 99 plates. To characterize the sensitivity of the test method, the LOD, the CC, the LOQ and the EC_{50} were calculated.

The LOD was calculated for each plate considering the plate specific variability of measured values for the negative control and the reference standard. The average LOD was 1.8 ng/L E2 and ranged from 0.7 to 3.7 ng/L E2. The CC ranged from 1.5 to 3.6 ng/L E2 with an average of 2.4 ng/L E2. Both LOD and CC are a concentration above which a sample or a sample dilution is assessed to mediate estrogenic activity. But for the calculation of the limits, different sources of variation are considered (see section 2.5). The LOQ ranged from 1.1 to 6.3 ng/L E2 with an average of 2.5 ng/L E2. The EC_{50} values ranged from 8.0 to 16.5 ng/L E2 with an average of 12.2 ng/L E2.

All concentrations refer to the concentration in the sample. An LOD of 1.8 ng/L E2 means that 1.8 ng/L EEQ can be detected in a sample.

3.3 Repeatability and reproducibility for EEQs of the undiluted sample

The precision of the A-YES® relating to the EEQ of the undiluted sample was determined by means of the repeatability and reproducibility standard deviation (s_r and s_R, respectively). The results for samples with quantifiable EEQ are summarized in Table 4, together with the mean EEQ and the ratio between s_R and s_r. A graphical presentation of the results for samples with measurable EEQs is shown in Figure 2.

Results for samples showing no estrogenic effect are presented in Figure 3. For each sample the test and laboratory specific LOD is shown as upper line of the triangle. Those results are also shown to demonstrate that a negative result depends on the test specific LOD, and a negative results not always entails the estrogenic effect of 0, but rather a value between 0 and the LOD.

Waste water samples

The unspiked effluent sample (#1) was below the LOD for the majority of participants (77%). The mean EEQ (across laboratories) was 8.8 ng/L for the spiked effluent (#2), 13.9 ng/L for the municipal influent (#3) and 20.4 ng/L for the hospital influent (#4).

Repeatability standard deviations (s_r) for the municipal WWTP samples (#2, #3) were 12.9 and 17.0%, respectively. The reproducibility standard deviation (s_R) was 28% for both samples. These are the highest values compared to all other samples. Nevertheless, these precision parameters reflect a good performance.

The repeatability and reproducibility standard deviation were 8.6 and 16.7%, respectively, for the hospital-influent sample (#4). It is worth mentioning that the precision parameters almost adopt the lowest values for this sample. Only the mixture of bisphenols (#7) showed slightly better results.

These good results from the waste water effluent and influent emphasize the applicability of the A-YES® for such challenging water samples.

The ratio s_R/s_r is an additional characteristic of method performance. Usually this ratio adopts values around 2. If the ratio is near 1, the reproducibility is dominated by the repeatability

error. If it is larger than 2, reproducibility mainly depends on the systematic lab bias.

For the waste water samples, the ratios were 2.2 (#2), 1.6 (#3) and 1.9 (#4), reflecting a good balance between random repeatability error and systematic lab bias.

Surface water samples

The unspiked Rhine sample (#5) was below the LOD for almost all participants (88 %). The mean EEQ of the spiked Rhine sample (#6) was 17.4 ng/L. The repeatability and reproducibility standard deviation were 21.4 and 24.9%, respectively. Here, the ratio between s_R and s_r is 1.2, indicating that the reproducibility standard deviation was dominated by the random repeatability error.

Mixture of bisphenols

For the mixture of BPA, BPS and BPZ (#7), the mean EEQ was 6.3 ng/L with a repeatability standard deviation of 7.5% and reproducibility standard deviation of 16.6%. The precision of this sample was best compared to all other samples, possibly due to a very low complexity (no interfering substances causing strong matrix effects). The s_R/s_r value of 2.2 indicates that laboratory bias is slightly higher than a random error.

Saline water sample

The mean EEQ of the spiked Baltic Sea sample (#8) was 9.0 ng/L. The repeatability and reproducibility standard deviation were 19.6 and 24.8%, respectively. The ratio s_R/s_r was 1.3. Here again, the reproducibility standard deviation was dominated by the random error. Thus, the results are comparable to the spiked Rhine sample.

Control sample (field blank)

The field blank functioned as a control for sampling, sampling material, and transportation process. It was below LOD for all participants indicating no contamination in the samples.

Table 4: Mean EEQ with 95% confidence interval (CI), relative reproducibility (s_R) and repeatability standard deviation (s_r) as well as their ratio (s_R/s_r) for the samples with EEQ above the LOD.

Nr.	Sample	Mean (ng/L EEQ) with 95% CI	s_r %	s_R %	s_R/s_r
2	Effluent, municipal-WWTP +10 ng/L EE2	8.8 (7.5 – 10.1)	12.9	28.0	2.2
3	Influent, municipal-WWTP	13.9 (11.9 – 15.9)	17.0	28.0	1.6
4	Influent, hospital-WWTP	20.4 (18.6 – 22.2)	8.6	16.7	1.9
6	Rhine + 15 ng/L EE2	17.4 (15.4 – 19.4)	21.4	24.9	1.2
7	BPA, BPS, BPZ	6.3 (5.7 – 6.9)	7.5	16.6	2.2
8	Baltic Sea + 1.4 ng/L E2 and 8 ng/L EE2	9.0 (8.0 – 10.1)	19.6	24.8	1.3

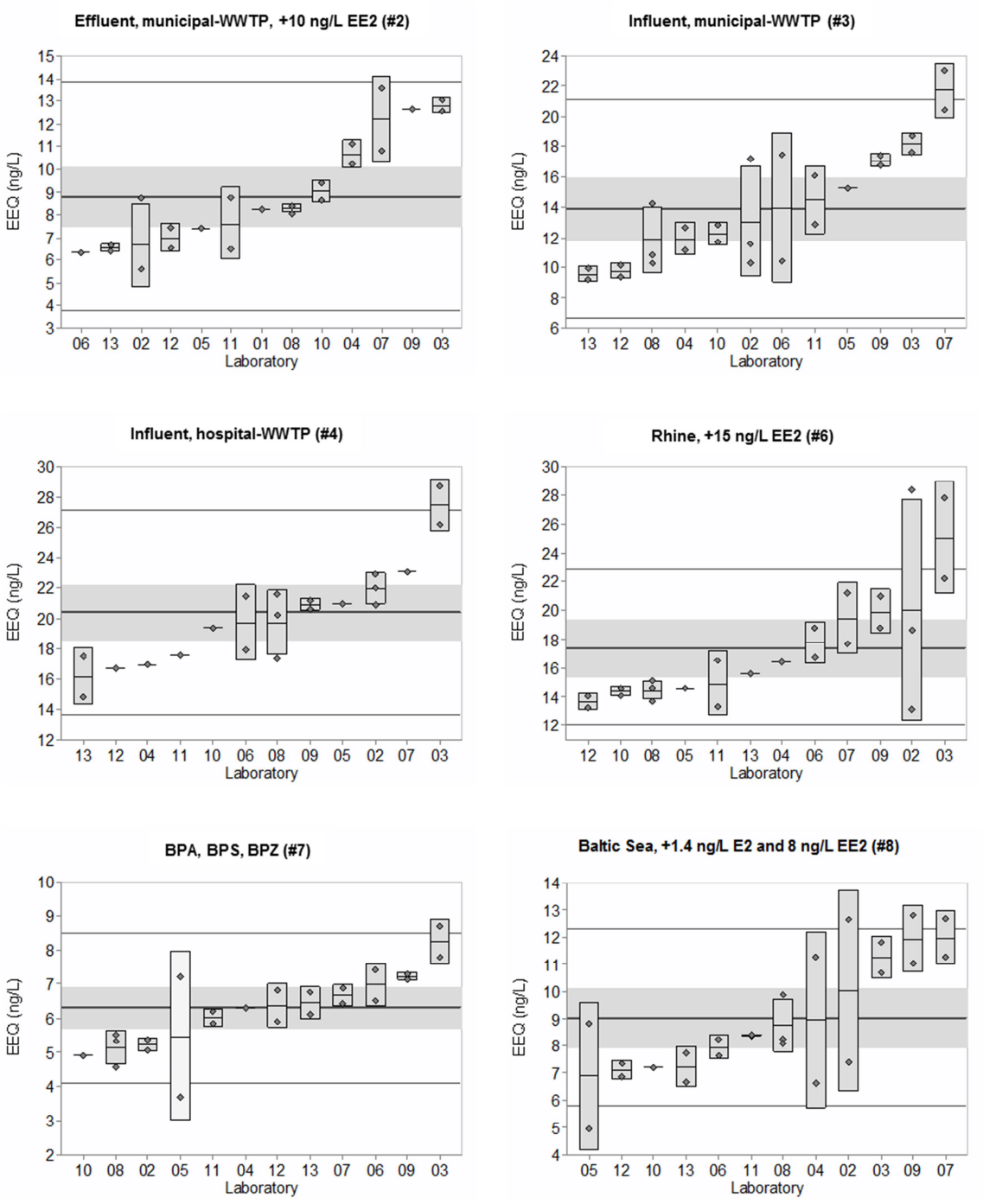

Fig. 2: Summary of results for sample #2, #3, #4, #6, #7 and #8. For each sample, the individual results of the laboratories (diamonds), the within-lab mean and variability (boxes), the overall mean (solid line) with 95% confidence interval (grey band) and the 95% prediction interval (outer lines) are shown. Results of Lab 05 for sample #7 are outliers.

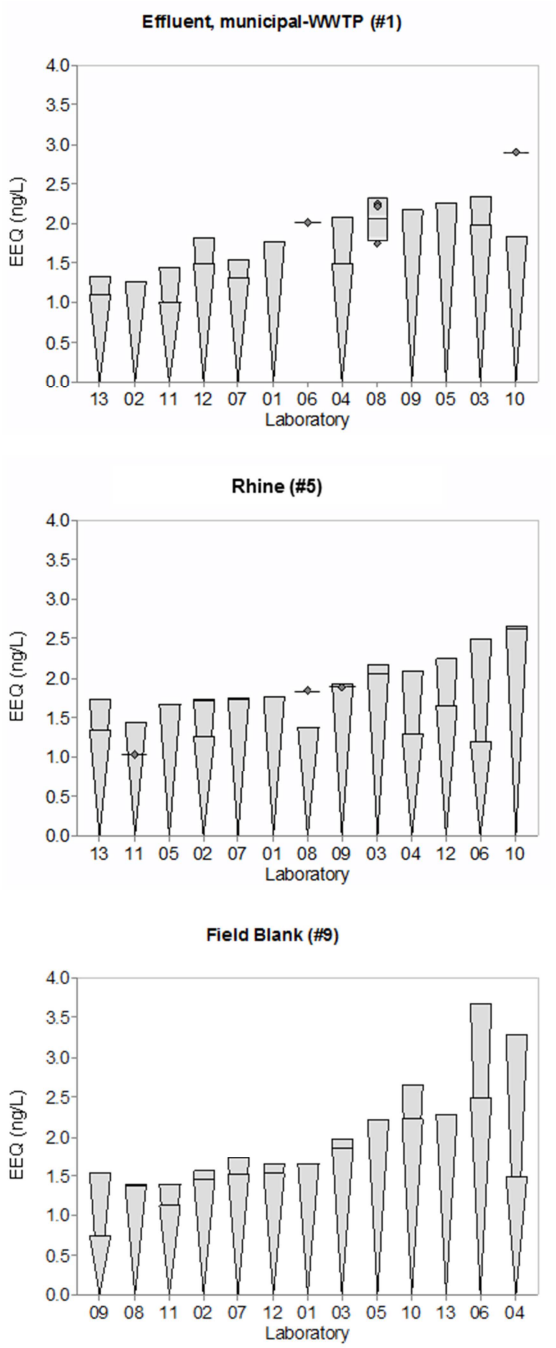

Fig. 3: Summary of results for sample #1, #5 and #9. For each sample the test and laboratory specific LOD's are shown as grey triangles (upper line) and results above the LOD are shown as diamonds. The triangles reflect that the EEQ lies between the LOD and 0.

3.4 Repeatability and reproducibility for dilution level specific EEQs

Furthermore, the repeatability and reproducibility standard deviation for the dilution level specific EEQs were determined. Table 5 summarizes these results in combination with the results from the undiluted sample.

For dilution levels with results mostly above the LOD, the relative repeatability and reproducibility standard deviation are broadly comparable to the precision parameters of the undiluted sample. For dilution levels where the majority of results were below the LOD the precision parameters are clearly larger as expected.

It is worth mentioning that for the two influent samples repeatability and reproducibility standard deviation for dilutions with EEQ around just above 2 ng/L is very good. This underpins the applicability of the A-YES® for highly complex water samples.

A graphical representation of the relative repeatability and reproducibility standard deviation for all samples and analyzed dilution levels is shown in Figure 4. Both relative standard deviations are almost exclusively below 30% in the working range of the assay, with the lower limit at approximately 2 ng/L EEQ.

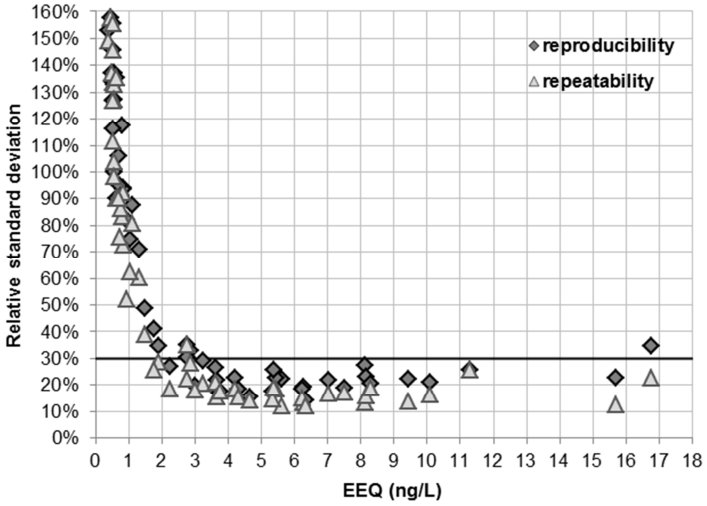

Fig. 4: Relative repeatability and reproducibility standard deviation for the dilution specific EEQs for all analyzed samples.

Another aspect to be considered is the relation between sample dilution and EEQ. Figure 5 shows the EEQ that was derived by multiplying the mean EEQ with the dilution level. Furthermore, the 95% confidence interval is indicated by the error bars. Interestingly, a systematic trend of the mean EEQs is observed from the influent samples, but not from other samples.

This is not surprising because the effluent sample, the Rhine sample, the bisphenol-mixture and the Baltic Sea sample were spiked to an extent where EE2 and E2 (or three bisphenols) are dominating the mixtures, such that the dilution level and EEQ are inversely proportional. In contrast, the municipal and the hospital influent samples are naturally contaminated.

In general, this nonlinear relation is frequently observed for bioassays (Vermeissen et al., 2006). However, it leads to the question of how samples with non-linear relation between dilution and EEQ should be analyzed and assessed.

In this study, the significance of dilution effects was determined by fitting a linear regression and checking the significance of the slope. If there is a significant dilution effect, then the slope is different from 0. If a linear regression is not appropriate, a polynomic 2nd order function may be used instead.

A significant dilution effect was observed only from the two influent samples. This effect was also present in the laboratory-specific results. The extent of the effect was different between the two influent samples, suggesting that it is more pronounced for the municipal-WWTP influent.

Because the extent of the dilution effect can vary from sample to sample, it is particularly important that this effect is analyzed and documented, and that the EEQs are calculated according to a standardized procedure. Further recommendations for the A-YES® are summarized in the discussion section.

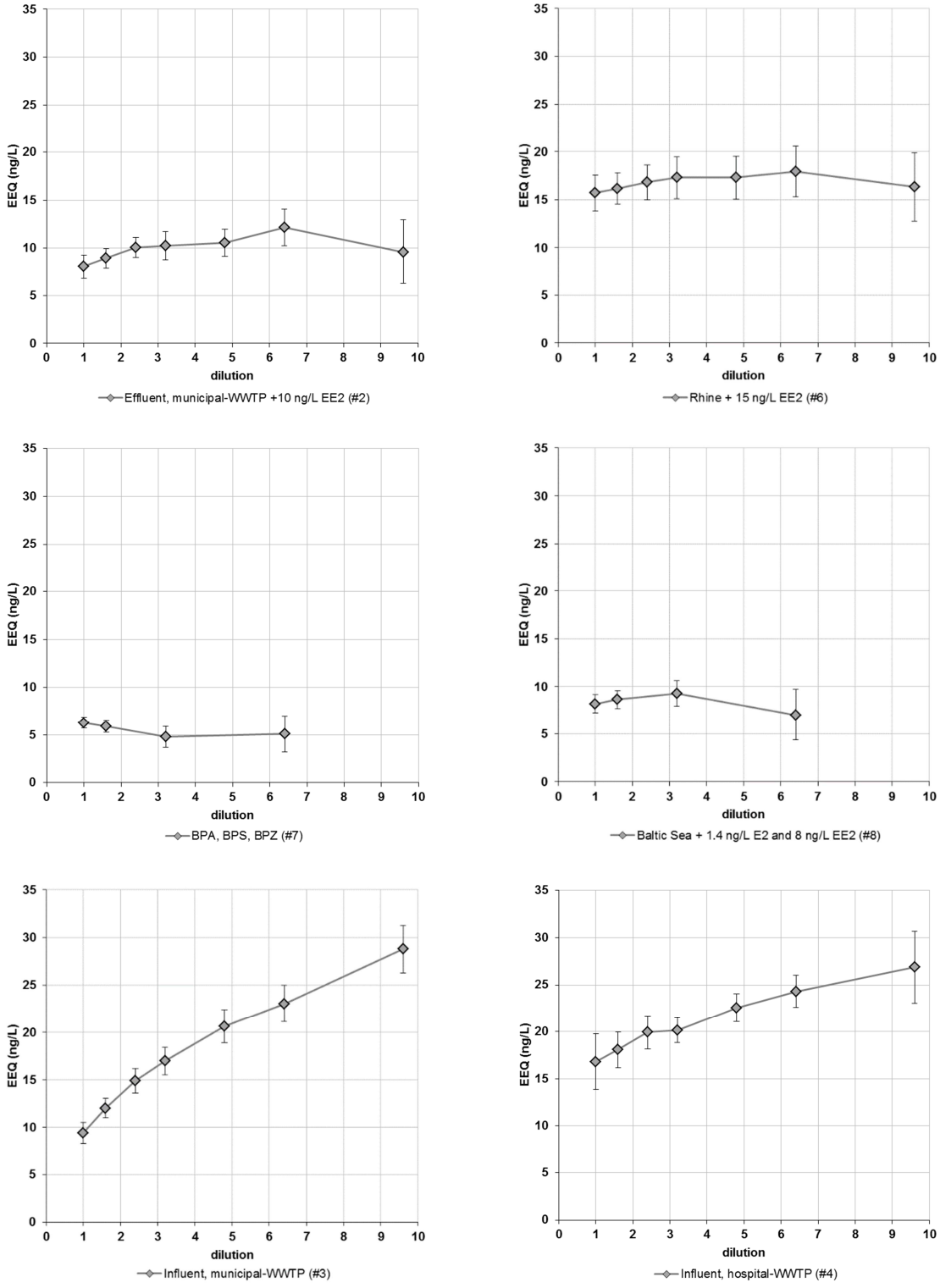

Fig. 5: Relation between sample dilution and EEQ of the undiluted sample. Error bars indicate the 95% confidence interval. EEQ was independent of dilution in samples spiked with estrogens and the mixture of bisphenols (#2, #6, #7, #8); EEQ increased with increasing dilution in two unspiked influent samples (#3, #4).

Table 5: Mean EEQ, reproducibility (s_R) and repeatability standard deviation (s_r) for the dilution specific EEQs and the undiluted sample EEQ. Meas. = measurements; LOD = limit of detection.

Sample		Dilution							Undiluted Sample
		1	1.6	2.4	3.2	4.8	6.4	9.6	
Effluent, municipal-WWTP + 10 ng/L EE2	Mean (ng/L EEQ)	8.1	5.6	4.2	3.2	2.2	1.9	1.0	8.8
	Meas. < LOD %	0.0	0.0	0.0	0.0	24.0	52.0	92.0	4.0
	s_r %	13.1	12.2	18.6	20.4	18.2	28.6	62.4	12.9
	s_R %	27.3	22.3	22.8	29.2	27.1	34.6	75.0	28.0
Influent, municipal-WWTP	Mean (ng/L EEQ)	9.4	7.5	6.2	5.3	4.3	3.6	3.0	13.9
	Meas. < LOD %	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0
	s_r %	13.5	17.0	15.2	14.5	15.4	15.4	18.1	17.0
	s_R %	22.4	19.0	18.4	17.8	18.3	18.0	19.7	28.0
Influent, hospital-WWTP	Mean (ng/L EEQ)	16.8	11.3	8.3	6.3	4.7	3.8	2.8	20.4
	Meas. < LOD %	0.0	0.0	0.0	0.0	0.0	0.0	11.5	0.0
	s_r %	22.6	25.5	18.9	12.1	14.3	17.4	35.3	8.6
	s_R %	34.5	25.5	20.4	14.2	15.4	17.4	35.3	16.7
Rhine + 15 ng/L EE2	Mean (ng/L EEQ)	15.7	10.1	7.0	5.4	3.6	2.8	1.7	17.4
	Meas. < LOD %	0.0	0.0	0.0	0.0	0.0	4.0	56.0	0.0
	s_r %	12.5	16.2	16.7	18.9	21.1	22.4	25.7	21.4
	s_R %	22.5	20.8	22.0	25.5	26.7	30.3	41.1	24.9
BPA, BPS, BPZ	Mean (ng/L EEQ)	6.3	3.7		1.5		0.8		6.3
	Meas. < LOD %	0.0	4.0		50.0		89.3		4.2
	s_r %	13.5	15.5		38.7		86.2		7.5
	s_R %	19.1	21.9		48.6		91.2		16.6
Baltic Sea + 1.4 ng/L E2 and 8 ng/L EE2	Mean (ng/L EEQ)	8.2	5.4		2.9		1.1		9.0
	Meas. < LOD %	0.0	0.0		12.0		73.1		0.0
	s_r %	15.9	18.5		28.3		80.7		19.6
	s_R %	23.1	22.8		32.7		87.5		24.8

3.5 Analysis of spiked samples

It should be noted that the recovery in terms of chemical analysis cannot be directly compared to effect-based bioassay results because such tests measure the cumulative effect of a sample including additive, synergistic, and antagonistic effects. These effects are directly influencing the recovery, especially in real samples with high matrix load. In light of this, the recovery for effect-based bioassays should not be interpreted as a validation parameter but as a feature of the sample. It is also worth mentioning that the recovery in chemical analysis usually refers to an extraction procedure. No such extraction procedure was used for the A-YES®. Therefore, the concept of apparent recovery is used herein (Burns et al., 2002).

For spiked samples #2, #6, #7 and #8 an (apparent) recovery analysis was performed (Table 6), considering the relative potency of the A-YES® for the specific compounds added to the samples. The relative potency is used to express the biological activity of a compound in comparison to the natural hERα ligand and calibration standard E2.

The apparent recovery for the spiked samples varied between 70.8 and 96.7%, with the highest values belonging to the Rhine and Baltic Sea samples. The lower recovery of the effluent sample is most likely due to the high matrix load of the sample. However, the lowest recovery was observed from the mixture of three bisphenols; although, there are no interfering matrix components present. One reason perhaps is that the expected EEQ was derived from the sum of effects of single compounds, and that the actual mixture effect of the three bisphenols may be lower (or higher) than the sum of the single effects.

Table 6: Apparent recovery for spiked samples based on the mean and expected EEQ.

Nr.	Sample	Mean (ng/L EEQ)	Expected (ng/L EEQ)	Apparent Recovery (%)
2	Effluent, municipal-WWTP +10 ng/L EE2	8.8	12.0	73.3
6	Rhine + 15 ng/L EE2	17.4	18.0	96.7
7	BPA, BPS, BPZ	6.3	8.9	70.8
8	Baltic Sea + 1.4 ng/L E2 and 8 ng/L EE2	9.0	11.0	81.8

4 Discussion

The prerequisites of measurement methods used in monitoring and regulatory purposes, e.g. in the framework of water quality, or for examination of waste water, are sensitivity, specificity, reproducibility, ease of use and cost effectiveness – regardless of a targeted chemical or an effect-based method. For the performance assessment of A-YES®, the comparisons with typically used targeted chemical methods, as well as other effect-based *in vitro* assays, are set out hereinafter. Emphasis is laid on repeatability, reproducibility, sensitivity and applicability. Furthermore, we address some aspects concerning the calculation of a sample EEQ.

Targeted chemical and effect-based methods have different approaches in measuring the estrogenic burden of a sample. Thus, a direct comparison between methods is not adequate, and the parameters' specificity and recovery are not discussed in details. Typically, the specificity of confirmatory chemical methods is very high, while effect-based tools measure

the cumulative effect of all estrogenic compounds present in a sample. This cumulative effect is more useful in answering questions relating to the effect of mixtures to a living organism. The necessity of the knowledge of individual contributions is often secondary, especially as the effect of substance mixtures may not be equal to the sum of the single effects.

Repeatability and reproducibility

In this interlaboratory study, the relative repeatability standard deviation for the A-YES® ranged from 7.5 to 21.4%, and the relative reproducibility standard deviation from 16.6 to 28.0% (EEQ for undiluted sample).

For targeted chemical analysis of steroid estrogens in aqueous samples using GC-MS or LC-MS, the reported relative reproducibility standard deviation varies from single digits to more than 50%, depending on the analyte, matrix and measured concentration (Vanderford et al., 2014). In particular, Vanderford et al. (2014) reported relative interlaboratory standard deviations between 20 and 30% for E2 and EE2 (after outlier elimination) in drinking and surface water. In a study led by Heath et al. (2010), estrogens were determined in tap water, river water and waste water influent and effluent after a solid-phase extraction. As the study reports results of a proficiency testing, each laboratory performed its own method. Either LC-MS/MS or GC-MS was used for the analyses. The relative reproducibility standard deviations for GC-MS methods ranged from 20 - 45%, and the relative repeatability standard deviations varied between 9 and 22%. When the different sample types were taken into account, the relative repeatability standard deviation ranged from 9 – 11% (tap water), 13 – 16% (surface water), 10 – 17% (influent) and 13 – 22% (effluent). Likewise, the relative reproducibility standard deviation ranged from 30 – 40% (tap water), 25 – 45% (surface water), 20 – 22% (influent) and 21 – 33% (effluent). While the LC-MS methods performed similarly for E2, the results for EE2 were unfavorably high, up to approximately 200%.

Other studies focusing on the within laboratory validation of LC-MS/MS methods for steroid estrogens report an intra- and inter-day variability for E2 and EE2 between 3 and 9% and between 3 and 11%, respectively (Fayad et al. 2013; Guo et al. 2013). Wozniak et al. (2014) conducted an in-house validation of a GC-MS method and received a relative within laboratory reproducibility ranging from 16.4 to 41.4% for E2 and EE2, respectively. Typically, these studies use standard solutions as samples, partly in combination with enrichment and extraction methods. One exception comes from the study by Fayad et al. (2013), which utilizes analyte-free effluent waste water samples in addition to standard solutions.

Comparing the results of a targeted chemical analysis to the A-YES®, we found that the relative interlaboratory reproducibility standard deviation and the relative repeatability standard deviation using the standardized A-YES® is within the ranges reported for conventional targeted chemical analyses for estrogens. This also applies particularly to different water types, as shown by the comparison of Heath et al. (2010) and Vanderford et al. (2014). These publications, however, include enrichment and extraction step, therefore this source of variability is considered as well. If only standard solutions are examined, the repeatability within one laboratory is partial, but not always smaller (see Guo et al. (2013) and Wozniak et al. (2014)). However, this depends also on the calculation method used.

Besides, the relative reproducibility standard deviation of the A-YES® is in general comparable to the conventional analytical chemistry (These et al., 2011), where a maximum relative reproducibility standard deviation of 30% is a common validation acceptance

criterion. Interestingly, the A-YES® also fulfills this requirement for challenging matrices, such as the marine water and the waste water samples, including the hospital influent.

Comparison of the relative repeatability and reproducibility standard deviation to other *in vitro* bioassays is confound by the fact that the statistical evaluation is not standardized, i.e. according to ISO 5725-2 (ISO 5725-2 2002). Following ISO 5725-2, the reproducibility standard deviation (s_R) comprises the within and between laboratory (s_r + s_L) variability. Given the published data in the literature, it often remains unclear, whether the reproducibility standard deviation (s_R) or one of its components is reported. However, it has been shown that the relative reproducibility standard deviation was between 33.6 and 72.7% for three yeast assays based on different modified *Saccharomyces cerevisiae* strains for the detection of estrogenic activity (Brix et al., 2010). These results are also comparable to the relative reproducibility standard deviation found in other interlaboratory studies, including yeast-based assays (Dhooge et al., 2006; Andersen et al., 1999). For mammalian cell-derived estrogen receptor (ER) assays, relative reproducibility standard deviations between 30 and 54% (Mehinto et al., 2015) and between 10 and 35% (Van Der Linden et al., 2008), respectively, were reported. It can be stated, that the relative reproducibility standard deviation for the A-YES® is comparable to and even better than these reported values.

Generally, bioassays possess a relatively large inherent variation (Brix et al., 2010). As such, yeast-based assays can suffer from strain variations that underlie the differences in receptor levels (Zacharewski, 1997), resulting in relatively imprecise data (Auchus, 2014). In addition, a lack of harmonization of method protocols between different laboratories using the same bioassays, along with different training levels of operators, has led to a variation in test results. The A-YES® circumvents problems caused by strain variations and variable receptor levels by using a recombinant *A. adeninivorans* strain with chromosomally located heterologous genes (*hERα*, *phyK*) (Kaiser et al., 2010). Stable mitotic *A. adeninivorans* transformants were produced after passaging on selective and non-selective media (Klabunde et al., 2003). Furthermore, the production of large batches of A-YES® cells under standardized and controlled conditions in a bioreactor, subsequent lyophilization of the cells and a continuous control of the biosensor quality, help moderate biological variations.

Sensitivity

The European Directive 2013/39/EU has included the pharmaceuticals E2, EE2 and E1 (estrone) on the watch list. Substances on this list are monitored EU-wide for the purpose of supporting future prioritization exercises in the framework of the WFD. This directive does not define environmental quality standards (EQS) for E2 and EE2. However, a former proposal for amending the priority substance list (European Commission 2011) suggested EQS for E2 and EE2. The proposed EQS for E2 of inland surface water is 0.4 ng/L, and 0.08 ng/L for other (salt) surface waters. The proposed EQS for EE2 are 0.035 and 0.007 ng/L, respectively. These are the concentrations that should not be exceeded in order to protect aquatic environment and human health.

Such low EQS values pose a major challenge for the detection using targeted chemical analysis, especially since the LOQs of the analytical methods must be at 30% of the EQS (EC 2009). In a recent report from Loos et al. (2015), analytical methods for possible WFD watch list substances such as E2 and EE2 were identified. The lowest LODs as reported by Williams et al. (2012) and Li et al. (2013) were 0.03 ng/L for E2 and 0.05 ng/L for EE2. The corresponding LOQs were 0.11 ng E2/L and 0.18 ng EE2/L (Li et al., 2013). Thus, applying

those methods, the proposed EQS for E2 of inland surface water can be monitored. Monitoring of proposed EQS for EE2 is arguably difficult, and in fact it is not achievable for other (salt) surface waters.

To accomplish LODs and LOQs in the lower pg/L range, a further development of enrichment and extraction steps is necessary. Current analytical methods apply enrichment factors that are typically around 1000, starting with a sample volume of approximately 1 L (EPA Method 539, 2010; Isobe et al., 2003). However, it must be mentioned that enrichment and extraction procedures typically increase the bias, and thus alter the method accuracy (Isobe et al., 2003) (Tomšíková et al., 2012).

Within this study, only native samples were analyzed, and no enrichment or extraction steps were applied. There are two explanations for this. On the one hand, ISO/DIS 19040-2 describes the testing of aqueous samples, regardless of any applied enrichment or extraction step. On the other hand, the variety of enrichment and extraction methods makes it impossible to validate each in combination with the A-YES®.

However, even though not demonstrated in this interlaboratory study, such enrichment and extraction steps have been used in combination with the A-YES®, as shown by Gehrmann et al. (2016). With an LOQ of 2.5 ng/L EEQ and the fact that LOQs should be at 30% of the EQS (European Commission 2009) and the A-YES® that is slightly more sensitive for EE2 than for E2 (factor of 1.2), an enrichment factor of approximately about 200 would be sufficient for detection of E2 and EE2 in inland and E2 in other (salt) surface waters. An enrichment factor of about 1000 would be effective for EE2 in (salt) surface waters. This by means holds under the assumption that enrichment and extraction steps do not alter method accuracy.

Applications: sample types

Within the framework of this interlaboratory study, it could be shown that different types of native water samples, including samples with high matrix load (municipal waste water effluent and influent, hospital waste water influent) as well as water with higher salinity (brackish and marine water), could be analyzed with the A-YES®. The robustness in terms of testable sample types is reflected by the fact that cytotoxic effects, which result in low yeast growth, occurred sporadically. It should be noted that with A-YES®, neither antibiotics are added to the yeast medium to suppress the background microflora, nor does it require sterile working during the test procedure. Owing to the robustness of the *Arxula* yeast against osmotic stress (Gellissen, 2005), the A-YES® is particularly suitable for native water samples with higher salinity. This robustness of the A-YES® offers a unique feature in comparison to the other commonly used *in vitro* assays.

In a recent study by Saranjampour et al. (2017), the effect of salinity on water solubility and partitioning potential of various pesticides and crude oil constituent has been investigated. The authors were able to show that the environmental fate estimates using these parameters indicate increased chemical sorption to the sediment, general bioavailability and toxicity in artificial seawater. They suggested that salinity should be taken into account when exposure estimates are made for marine organisms. Due to the suitability of the A-YES® for saline water samples, the assay can be a useful method to investigate the impact of salinity on the combined activity of chemical mixtures. This is also relevant in terms of the WFD, as chemicals are transported from freshwater to marine environments.

Other frequently used *in vitro* assays for the effect-based detection of estrogens and estrogen-acting compounds are reporter gene assays using the yeast *Saccharomyces cerevisiae* (yeast estrogen screen assay, YES) (Routledge and Sumpter, 1996; McDonnell et al., 1991) or are based on human cells, e. g. cell proliferation assays such as the E-Screen (Soto et al., 1995) or reporter gene assays such as MELN (Balaguer et al., 1999) or ER-CALUX® (Legler et al., 1999). The limits of quantification for the human cell-based assays are typically in the lower pg/L range (down to 100 pg/L), whereas the LOQ value for the YES is about 5 – 10 ng/L, depending on the yeast strain used (Leusch et al., 2010). Usually enriched/extracted samples are used for analysis, especially for the human cell-based assays (Leusch et al., 2010; Vanparys et al., 2010; Swart et al., 2011). This is necessary due to low concentrations of estrogens in many samples (lower pg/L range) and the requirements of the WFD for the EQS of EE2. Enrichment/extraction also reduces effects of interfering matrix components and microbial load, affecting cell metabolism and causing cell death.

Nevertheless, native water samples, especially waste water samples, are analyzed using these assays (Escher et al., 2008; Gehrmann et al., 2016). But the YES and E-Screen are less applicable for untreated samples with higher salinity (Kase et al., 2008). A recent study on methods for the analysis of waste water reported that cytotoxicity caused by matrix composition was lowest for the A-YES® in comparison to the YES and ER-CALUX® with T47D cells (Gehrmann et al., 2016). Within the interlaboratory study, it was also possible to demonstrate that the A-YES® can be used for the analysis of waste water samples because low cell toxicity was observed and the reproducibility was very good.

Sample matrix effects and calculation of EEQ

An important issue concerning effect-based *in vitro* assays is the calculation of reliable EEQs of the analyzed samples. Several approaches for the calculation are applied and described in various publications (Kunz et al. 2017, Gehrmann et al. 2016, Escher et al. 2008). If the effective concentration values of the sample dose-response curve are used (e.g. EC_{10} or EC_{50}, OECD (2016)), results from samples can deviate, showing no linear relationship between dilution and EEQ. In some cases, dose-response curves of complex samples have no parallelism to dose-response curves of the corresponding calibration standard. Thus, calculated EEQs have greater uncertainty, and estrogenicity of the samples is often underestimated (Vermeissen et al. 2006). In this interlaboratory study some of the analyzed samples revealed these characteristics.

Although one can consider such effects revealed by effect-based *in vitro* assays as a drawback, it should always be kept in mind that bioassays measure the cumulative effect of a sample, thus of complex mixtures. Therefore, it cannot be assumed that effects are linear to dilution, or relating directly to health outcomes in an intact organism. The sum of the individual effects does not necessarily correspond to the effect of the mixture. However, the matrix effects are usually alleviated by separating the sample components. An alternative to reduce effects caused by the sample composition (so called matrix effects) may be the combination of bioassays with high performance thin layer chromatography (HPTLC). After chromatographic separation of sample constituents, the bioassay test organism is directly applied on the HPTLC plate, allowing the bioassay response to be evaluated after incubation. For example, Chamas et al. (2017) describes the combination of a chromatographic separation and yeast-based reporter assays for estrogens, androgens and progestogens. Similarly, Buchinger et al. (2013) reports the coupling of thin-layer

chromatography with a *Saccharomyces cerevisiae* based assay for detection of estrogenic compounds. Nevertheless, these assays require special instrumentations and are more difficult to perform than the 96-well plate assays while often offering a poorer sensitivity.

However, for using the A-YES® assay as routine method it is important that sample matrix effects are examined, that the calculation of the EEQ is standardized, and that it is precisely documented to which dilution the calculated EEQ refers. This is particularly important for the comparability of results within and between laboratories.

Based on the results of the interlaboratory trial, it is recommended to examine samples in several dilutions and to use the EC_{20} approach for quantification for the A-YES®, if possible. More importantly, the test report should indicate whether or not the dose-response curve of the sample is parallel to the E2 reference.

The matrix effects should be considered for any cases of risk or legal assessment of environmental samples, depending on the risk of an incorrect assessment and the consequences of exceeding a threshold value.

5 Conclusions

This interlaboratory study demonstrated that the A-YES® is a suitable tool for effect-based analysis of estrogen-acting compounds and can be applied as routine method. All unexperienced participants successfully established the A-YES® within a brief period, i.e., within two weeks and the participants fulfilled the validity criteria by a high degree. It was demonstrated in this validation study that the A-YES® is suitable for analyzing different sample types, including surface, coastal, and marine water, waste water influent and effluent as well as chemicals.

The relative repeatability and reproducibility standard deviation of the A-YES® is comparable to chemical analysis methods including LC-MS-MS and GC-MS, and the assay is able to detect the estrogenic activity in the low ng/L range for untreated samples. Applying enrichment for estrogen acting substances in water samples with an appropriate enrichment factor, the A-YES® could be a feasible screening tool for monitoring the compliance of water samples with the proposed EQS for E2 and EE2 according to the WFD.

The results of the interlaboratory study have been included in the ISO draft standard. The study clearly underpins that the A-YES® is a less time-consuming and straightforward method to screen for estrogenic potentials, and can be applied in the fields of environmental monitoring in freshwater and marine habitats, waste water treatment, as well as in risk assessment of chemical compounds. The A-YES® can be included in analytical procedures incorporating bioanalytical, targeted and non-targeted approaches to achieve more comprehensive and effect-based monitoring strategies.

6 Acknowledgements

This study was part of the working program of ISO/TC 147/ SC5/WG9 and DIN Standardization Committee 119-01-03-05-09. We thank all members of the ISO and DIN working group as well as the DIN and ISO secretariat for their contributions to the ISO/DIS 19040-2. We gratefully acknowledge financial support by the German Chemical Society (GDCh).

We gratefully thank Matthias Nagel and Sarah Löwe for excellent technical assistance. The

Authors would also like to thank the staff of the municipal WWTP in Koblenz and the hospital WWTP in Gelsenkirchen (Dr. Issa Nafo).

7 Conflict of Interest

Kirsten Simon is employee of new diagnostics GmbH. All A-YES® test kits were provided by the company.

8 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. *Environ. Int.*, 99, 107-119.

Altenburger, R., Ait-Aissa, S., Antczak, P., Backhaus, T., Barceló, D., Seiler, T. B., Brion, F., Busch, W., Chipman, K., de Alda, M. L., de Aragão Umbuzeiro, G., Escher, B. I., Falciani, F., Faust, M., Focks, A., Hilscherova, K., Hollender, J., Hollert, H., Jäger, F., Jahnke, A., Kortenkamp, A., Krauss, M., Lemkine, G.F., Munthe, J., Neumann, S., Schymanski, E.L., Scrimshaw, M., Segner, H., Slobodnik, J., Smedes, F., Kughathas, S., Teodorovic, I., Tindall, A.J., Tollefsen, K.E., Walz, K.-H., Williams, T.D., Van den Brink, P.J., van Gils, J., Vrana, B., Zhang, X., Brack, W. (2015). Future water quality monitoring - Adapting tools to deal with mixtures of pollutants in water resource management. *Sci. Total Environ.*, 512–513, 540–551.

Andersen, H. R., Andersson, A. M., Arnold, S. F., Autrup, H., Barfoed, M., Beresford, N. A., Bjerregaard, P., Christiansen, L. B., Gissel, B., Hummel, R., Jørgensen, E. B., Korsgaard, B., Le Guevel, R., Leffers, H., McLachlan, J., Møller, A., Nielsen, J. B., Olea, N., Oles-Karasko, A., Pakdel, F., Pedersen, K. L., Perez, P., Skakkebœk, N. E., Sonnenschein, C., Soto, A. M., Sumpter, J. P., Thorpe, S. M., Grandjean, P. (1999). Comparison of short-term estrogenicity tests for identification of hormone-disrupting chemicals. *Environ. Health Perspect.*, 107, Suppl. 1, 89–108.

Auchus, R. J. (2014). Steroid assays and Endocrinology: Best practices for basic scientists. *Endocrinology*, *155*(6), 2049–2051.

Balaguer, P., François, F., Comunale, F., Fenet, H., Boussioux, A. M., Pons, M., Nicolas, J. C., Casellas, C. (1999). Reporter cell lines to study the estrogenic effects of xenoestrogens. *Sci. Total Environ.*, *233*(1–3), 47–56.

Bergman, Å., Heindel, J., Jobling, S., Kidd, K., Zoeller, R. T. (2012). State of the science of endocrine disrupting chemicals, 2012. *Toxicology Letters*, 211. Suppl. S3, WHO, UNEP.

Benstead, RS., Baynes, A., Casey, D., Routledge, EJ., Jobling, S. (2011). 17β-Oestradiol may prolong reproduction in seasonally breeding freshwater gastropod molluscs. *Aquatic Toxicology, 101*(2), 326–334.

Brack, W., Ait-Aissa, S., Burgess, R. M., Busch, W., Creusot, N., Di Paolo, C., Escher, B. I., Mark Hewitt, L., Hilscherova, K., Hollender, J., Hollert, H., Jonker, W., Kool, J., Lamoree, M., Muschket, M., Neumann, S., Rostkowski, P., Ruttkies, C., Schollee, J., Schymanski, E. L., Schulze, T., Seiler, T. B., Tindall, A. J., De Aragão Umbuzeiro, G., Vrana, B., Krauss, M.

(2016). Effect-directed analysis supporting monitoring of aquatic environments - An in-depth overview. *Sci. Total Environ.*, 544, 1073–1118.

Brack, W., Dulio, V., Ågerstrand, M., Allan, I., Altenburger, R., Brinkmann, M., Bunke, D., Burgess, R. M., Cousins, I., Escher, B. I., Hernández, F. J., Hewitt, L. M., Hilscherová, K., Hollender, J., Hollert, H., Kase, R., Klauer, B., Lindim, C., Herráez, D. L., Miège, C., Munthe, J., O'Toole, S., Posthuma, L., Rüdel, H., Schäfer, R. B., Sengl, M., Smedes, F., van de Meent, D., van den Brink, P. J., van Gils, J., van Wezel, A. P., Vethaak, A. D., Vermeirssen, E., von der Ohe, P. C., Vrana, B. (2017). Towards the review of the European Union Water Framework management of chemical contamination in European surface water resources. *Sci. Total Environ.*, 576, 720–737.

Brix, R., Noguerol, T. N., Piña, B., Balaam, J., Nilsen, A. J., Tollefsen, K. E., Levy, W., Schramm, K. W., Barceló, D. (2010). Evaluation of the suitability of recombinant yeast-based estrogenicity assays as a pre-screening tool in environmental samples. *Environ. Int.*, 36(4), 361–367.

Buchinger, S., Spira, D., Bröder, K., Schlüsener, M., Ternes, T., Reifferscheid, G. (2013). Direct Coupling of Thin-Layer Chromatography with a Bioassay for the Detection of Estrogenic Compounds: Applications for Effect-Directed Analysis. *Anal. Chem.*, 85, 7248−7256.

Burns, D. T., Danzer, K., Townshend A. (2002). Use of the terms “recovery” and “apparent recovery” in analytical procedures. *Pure Appl. Chem.*, 74(11), 2201-2205.

Chamas, A., Pham, H., T., M., Jähne, M., Hettwer, K., Gehrmann, L., Türk, J., Uhlig, S., Simon, K., Baronian, K., Kunze, G. (2017). Separation and identification of hormone-active compounds using a combination of chromatographic separation and yeast-based reporter assay. *Sci. Total. Environ.*, 605-606, 507–513.

Dhooge, W., Arijs, K., D’Haese, I., Stuyvaert, S., Versonnen, B., Janssen, C., Verstraete, W., Comhaire, F. (2006). Experimental parameters affecting sensitivity and specificity of a yeast assay for estrogenic compounds: Results of an interlaboratory validation exercise. *Anal. Bioanal. Chem.*, *386*(5), 1419–1428.

Di Paolo, C., Ottermanns, R., Keiter, S., Ait-Aissa, S., Bluhm, K., Brack, W., Breitholtz, M., Buchinger, S., Carere, M., Chalon, C., Cousin, X., Dulio, V., Escher, B. I., Hamers, T., Hilscherová, K., Jarque, S., Jonas, A., Maillot-Marechal, E., Marneffe, Y., Nguyen, M.T., Pandard, P., Schifferli, A., Schulze, T., Seidensticker, S., Seiler, T.-B., Tang, J., van der Oost, R., Vermeirssen, E., Zounková, R., Zwart, N., Hollert, H. (2016). Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts – Towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring. *Water Res.*, 104, 473–484.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.*, 7(1), 1-26.

EPA Method 539 (2010). United States Environmental Protection Agency, Method 539: Determination of hormones in drinking water by solid phase extraction (SPE) and liquid

chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS), EPA Document No. 815-B-10-001.

Escher, B. I., Bramaz, N., Quayle, P., Rutishauser, S., Vermeirssen, E. L. M. (2008). Monitoring of the ecotoxicological hazard potential by polar organic micropollutants in sewage treatment plants and surface waters using a mode-of-action based test battery. *J. Environ. Monit.*, *10*(5), 622.

European Commission (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. *Off. J. Eur. Union*, L 327/72 (22.12.2000),1–72.

European Commission (2009). Directive 2009/90/EC of 31 July 2009 laying down, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, technical specifications for chemical analysis and monitoring of water status. *Off. J. Eur. Union*, L 201/36 (1.8.2009),1–3.

European Commission (2011). Proposal for a Directive of the European Parliament and of the Council amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy. COM(2011) 876 Final. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2011:0876:FIN

Fayad, P. B., Prévost, M., Sauvé, S. (2013). On-line solid-phase extraction coupled to liquid chromatography tandem mass spectrometry optimized for the analysis of steroid hormones in urban wastewaters. *Talanta*, 115, 349-360.

Filby, A. L., Neuparth, T., Thorpe, K. L., Owen, R., Galloway, T. S., Tyler, C. R. (2007). Health impacts of estrogens in the environment, considering complex mixture effects. *Environ. Health Perspect.*, 115(12), 1704–1710.

Gall, H. E., Sassman, S. A., Lee, L. S., Jafvert C. T. (2011). Hormone Discharges from a Midwest Tile-Drained Agroecosystem Receiving Animal Wastes. *Environ. Sci. Technol.*, 45(20), 8755–8764.

Gehrmann, L., Bielak, H., Behr, M., Itzel, F., Lyko, S., Simon, A., Kunze, G., Dopp, E., Wagner, M., Tuerk, J. (2016). (Anti-)estrogenic and (anti-)androgenic effects in wastewater during advanced treatment: comparison of three in vitro bioassays. *Environ. Sci. Pollut. Res.*, doi:10.1007/s11356-016-7165-4.

Gellissen, G. (2005). Production of Recombinant Proteins: Novel Microbial and Eukaryotic Expression Systems. Chapter 5, *Arxula adeninivorans*. Willey-VCH 2005, 89-110.

Guo, F., Liu, Q., Qu, G. B., Song, J. T., Shi, J. B., Jiang, G. B. (2013). Simultaneous determination of five estrogens and four androgens in water samples by online solid-phase extraction coupled with high-performance liquid chromatography-tandem mass spectrometry. *J. Chromatogr. A*, 1291, 9-18.

Hamers, T., Legler, J., Blaha, L., Hylland, K., Marigomez, I., Schipper, C. A., Segner, H.,

Vethaak, A. D., Witters, H., de Zwart, D., Leonards, P. E. G. (2013). Expert opinion on toxicity profiling-report from a NORMAN expert group meeting. *Integr. Environ. Assess. Manag.*, 9(2), 185–191.

Heath, E., Kosjek, T., Andersen, H. R., Holten Lützhøft, H. C., Adolfson Erici, M., Coquery, M., Düring, R. A., Gans, O., Guignard, C., Karlsson, P., Manciot, F., Moldovan, Z., Patureau, D., Cruceru, L., Sacher, F., Ledin, A. (2010). Inter-laboratory exercise on steroid estrogens in aqueous samples. *Environ. Pollut.*, 158(3), 658–662.

Hecker, M., Hollert, H. (2011). Endocrine disruptor screening: regulatory perspectives and needs. *Environ. Sci. Eur.*, 23(1), Article 15, 1–14.

Hill, A. V. (1910). The possible effects of aggregation of the molecules of hemoglobin on its dissociation curves. *J. Physiol.*, 40, 4–7.

ISO Standards Catalogue. ISO/TC 147 – Water quality. Standards under development. (https://www.iso.org/committee/52972/x/catalogue/p/0/u/1/w/0/d/0)

ISO 5725-2. 2002. Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of standard measurement method (ISO 5725-2:1994 including Technical Corrigendum 1:2002). Berlin Germany: DIN Deutsches Institut für Normung e.V.

Isobe, T., Shiraishi, H., Yasuda, M., Shinoda, A., Suzuki, H., Morita, M. (2003). Determination of estrogens and their conjugates in water using solid-phase extraction followed by liquid chromatography-tandem mass spectrometry. *J. Chromatogr. A*, 984(2), 195–202.

Jobling, S., Beresford, N., Nolan, M., Rodgers-Gray, T., Brighty, G. C., Sumpter, J. P., Tyler, C. R. (2002). Altered sexual maturation and gamete production in wild roach (*Rutilus rutilus*) living in rivers that receive treated sewage effluents. *Biol. Reprod.*, 66(2), 272–281.

Jobling, S., Tyler, C. R. (2003). Endocrine disruption, parasites and pollutants in wild freshwater fish. *Parasitology*, 126(7), 103–108.

Kaiser, C., Uhlig, S., Gerlach, T., Körner, M., Simon, K., Kunath, K., Florschütz, K., Baronian, 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. *Sci. Total Environ.*, 408(23), 6017–6026.

Karpuzoglu, E., Holladay, S. D., Gogal Jr., R. M. (2013). Parabens: Potential impact of Low-Affinity Estrogen receptor Binding chemicals on Human health. *J. Toxicol. Environ. Health*, Part B:Critical Reviews, 16(5), 321-335.

Kase, R., Hansen, P.-D., Fischer, B., Manz, W., Heininger, P., Reifferscheid, G. (2008). Integral assessment of estrogenic potentials of sediment-associated samples. Part 1: The influence of salinity on the in vitro tests ELRA, E-Screen and YES. *Environ. Sci. Pollut. Res. Int.*, *15*(1), 75–83.

Klabunde, J., Kunze, G., Gellissen, G., Hollenberg, C. (2003). Integration of heterologous genes in several yeast species using vectors containing a -derived rDNA-targeting element. *FEMS Yeast Res.*, 4(2), 185–193.

Kunz, P. Y., Simon, E., Creusot, N., Jayasinghe, B. S., Kienle C., Maletz, S., Schifferli, A., Schönlau, C., Aït-Aïssa, S., Denslow, N. D., Hollert, H., Werner, I., Vermeissen, E. L. M. (2017). Effect-based tools for monitoring estrogenic mixtures: Evaluation of five *in vitro* bioassays. *Water research*, 110, 378–388

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. *J. Pharm. Biomed. Anal.*, 106, 107–115.

Lange, A., Katsu, Y., Ichikawa, R., Paull, G.C., Chidgey, L.L., Coe, T.S., Iguchi, T., Tyler, C.R. (2008). Altered sexual development in roach (Rutilus rutilus) exposed to environmental concentrations of the pharmaceutical 17a-ethinylestradiol and associated expression dynamics of aromatases and estrogen receptors. *Toxicol. Sci.*, *106*, 113–123.

Legler, J., Van Den Brink, C. E., Brouwer, A., Murk, A. J., Van Der Saag, P. T., Vethaak, A. D., Van Der Burg, B. (1999). Development of a stably transfected estrogen receptor-mediated luciferase reporter gene assay in the human T47D breast cancer cell line. *Toxicol. Sci.*, *48*(1), 55–66.

Leusch, F. D. L., De Jager, C., Levi, Y., Lim, R., Puijker, L., Sacher, F., Tremblay, L. A., Wilson, V. S., Chapman, H. F. (2010). Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters. *Environ. Sci. Technol.*, *44*(10), 3853–3860.

Leusch, F. D. L., Khan, S. J., Gagnon, M. M., Quayle, P., Trinh, T., Coleman, H., Rawson, C., Chapman, H. F., Blair, P., Nice, H., Reitsema, T. (2014). Assessment of wastewater and recycled water quality: A comparison of lines of evidence from in vitro, in vivo and chemical analyses. *Water Res.*, 50, 420–431.

Li, J., Fu, J., Zhang, H., Li, Z., Ma, Y., Wu, M., Liu, X. (2013). Spatial and seasonal variations of occurrences and concentrations of endocrine disrupting chemicals in unconfined and confined aquifers recharged by reclaimed water: A field study along the Chaobai River, Beijing. *Sci. Total Environ.*, 450–451, 162–168.

Loos R. (2015), Analytical methods for possible WFD 1st watch list substances. *Publ. Off. Eur. Union*, ISBN 978-92-79-44734-1, doi: 10.2788/723416.

Maitre, S. C. (2013). History of oral contraceptive drugs and their use worldwide. *Best Pract. Res. Clin. Endocrinol. Metab.*, 27(1), 3-12.

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*. *J. Steroid Biochem. Mol. Biol.*, *39*(3), 291–7.

Mehinto, A. C., Jia, A., Snyder, S. A., Jayasinghe, B. S., Denslow, N. D., Crago, J., Schlenk,

D., Menzie, C., Westerheide, S. D., Leusch, F. D. L., Maruya, K. A. (2015). Interlaboratory comparison of in vitro bioassays for screening of endocrine active chemicals in recycled water. *Water Res.*, *83*, 303–309.

OECD (2016), Test No. 455: Performance-based test guideline for stably transfected transactivation *in vitro* assays to detect estrogen receptor agonists and antagonists, OECD Publishing, Paris.

Oehlmann, J., Schulte-Oehlmann, U., Bachmann, J., Oetken, M., Lutz, I., Kloas, W., Ternes, A. T. (2006). Bisphenol A induces super-feminization in the ramshorn snail *Marisa cornuarietis* (Gastropoda: Prosobranchia) at environmentally relevant concentrations. *Environ. Health Perspect.*, 114, Suppl. 1, 127–133.

Racz, L., Goel, R. K. (2010). Fate and removal of estrogens in municipal wastewater. J. *Environ. Monit.*, 12(1), 58–70.

Rochester, J. R., Bolden, A. L. (2015). Bisphenol S and F: A Systematic Review and Comparison of the Hormonal Activity of Bisphenol A Substitutes. *Environ. Health Perspect.*, 123(7), 643–650.

Routledge, E. J., Sumpter, J. P. (1996). Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. *Environ. Toxicol. Chem.*, 15(3), 241–248.

Saranjampour, P., Vebrosky, E. N., Armbrust, K. L. (2017) Salinity impacts on water solubility and n-octanol/water partition coefficients of selected pesticides and oil constituents. *Environ. Toxicol. Chem.*, 36(9), 2274-2280.

Silva, E., Rajapakse, N., Kortenkamp, A. (2002). Something from “nothing”-eight weak estrogenic chemicals combined at concentrations below NOECs produce significant mixture effects. *Environ. Sci. Technol.*, 36, 1751-1756.

Soto, A. M., Sonnenschein, C., Chung, K. L., Fernandez, M. F., Olea, N., Olea Serrano, F. (1995). The E-SCREEN assay as a tool to identify estrogens: An update on estrogenic environmental pollutants. *Environ. Health Perspect.*, *103*, Suppl. 7, 113–122.

Swart, J. C., Pool, E. J., van Wyk, J. H. (2011). The implementation of a battery of in vivo and in vitro bioassays to assess river water for estrogenic endocrine disrupting chemicals. *Ecotoxicol. Environ. Saf.*, *74*(1), 138–143.

These, A., Klemm, C., Nausch, I., Uhlig, S. (2011). Results of a European interlaboratory method validation study for the quantitative determination of lipophilic marine biotoxins in raw and cooked shellfish based on high-performance liquid chromatography-tandem mass spectrometry. Part I: Collaborative study. *Anal. Bioanal. Chem.*, *399*(3), 1245–1256.

Tomšíková, H., Aufartová, J., Solich, P., Nováková, L., Sosa-Ferrera, Z., Santana-Rodríguez, J. J. (2012). High-sensitivity analysis of female-steroid hormones in environmental samples. *TrAC - Trends Anal. Chem.*, *34*, 35–57.

Tousova, Z., Oswalda, P., Slobodnika, J., Blahab, L., Muzc, M., Huc M., Brack, W., Kraussc, M., Di Paolod, C., Tarcaid, Z., Seiler, T-B., Hollert, H. (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.

Van Der Linden, S. C., Heringa, M. B., Man, H. Y., Sonneveld, E., Puijker, L. M., Brouwer, A., Van Der Burg, B. (2008). Detection of multiple hormonal activities in wastewater effluents and surface water, using a panel of steroid receptor CALUX bioassays. *Environ. Sci. Technol.*, *42*(15), 5814–5820.

Vanderford, B. J., Drewes, J. E., Eaton, A., Yingbo, C. G., Haghani, A., Hoppe-Jones, C., Schluesener, M. P., Snyder, S. A., Ternes, T., Wood, C. J. (2014). Results of an Interlaboratory Comparison of Analytical Methods for Contaminants of Emerging Concern in Water. Anal. Chem., 86(1), 774-782.

Vanparys, C., Depiereux, S., Nadzialek, S., Robbens, J., Blust, R., Kestemont, P., De Coen, W. (2010). Performance of the flow cytometric E-screen assay in screening estrogenicity of pure compounds and environmental samples. *Sci. Total Environ.*, *408*(20), 4451–4460.

Vermeissen, E. L.M., Suter, M. J.-F., Burkhardt-Holm, P., (2006). Estrogenicity Patterns in the swiss midland river Lützelmurg in relation to treated domestic sewage effluent discharges and hydrology. *Environ. Toxicol. Chem.*, 25(9), 2413–2422.

Waste Water Ordinance (2004). Ordinance on Requirements for the Discharge of Waste Water into Waters. Federal Minsitry for the Environment, Nature Conservation and Nuclear Saftey, Germany. 17. June 2004

Wernersson, A.-S., Carere, M., Maggi, C., Tusil, P., Soldan, P., James, A., Sanchez, W., Dulio, V., Broeg, K., Reifferscheid, G., Buchinger, S., Maas, H., Van Der Grinten, E., O'Toole, S., Ausili, A., Manfra, L., Marziali, L., Polesello, S., Lacchetti, I., Mancini, L., Lilja, K., Linderoth, M., Lundeberg, T., Fjallborg, B., Porsbring, T., Larsson, D., Bengtsson-Palme, J., Forlin, L., Kienle, C., Kunz, P., Vermeissen, E., Werner, I., Robinson, C.D., Lyons, B., Katsiadaki, I., Whalley, C., den Haan, K., Messiaen, M., Clayton, H., Lettieri, T., Carvalho, R.N., Gawlik, B.M., Hollert, H., Di Paolo, C., Brack, W., Kammann, U., Kase, R. (2015). The European technical report on aquatic effect-based monitoring tools under the water framework directive. *Environ. Sci. Eur.*, 27(1), Article 7, 1–11.

Williams, R. J., Churchley, J. H., Kanda, R., Johnson, A. C. (2012). Comparing predicted against measured steroid estrogen concentrations and the associated risk in two United Kingdom river catchments. *Environ. Toxicol. Chem.*, 31(4), 892–898.

Woźniak, B., Klopot, A., Matraszek-Żuchowska, I., Sielska, K., Żmudzki, J. (2014). Determination of natural and synthetic oestrogens in surface water using gas chromatography-mass spectrometry. *Bull. Vet. Inst. Pulawy*, 58, 603-611.

Zacharewski, T. (1997). In vitro bioassays for assessing estrogenic substances. *Environ. Sci. Technol.*, *31*(3), 613–623.

Appendix A: Supplementary data

Validation of *Arxula* Yeast Estrogen Screen assay for detection of estrogenic activity in water samples: Results of an international interlaboratory study

Karina Hettwer [a], Martin Jähne [a], Kirstin Frost [a], Martin Giersberg [b], Gotthard Kunze [b], Michael Trimborn [c], Martin Reif [c], Jochen Tuerk [d], Linda Gehrmann [d], Freddy Dardenne [e], Femke De Croock [e], Marion Abraham [f], Anne Schoop [f], Joanna J. Waniek [f],Thomas Bucher [g], Eszter Simon [g], Etienne Vermeirssen [g], Anett Werner [h], Karin Hellauer [i], Ursula Wallentits [i], Jörg E. Drewes [i], Detlef Dietzmann [j], Edwin Routledge [k], Nicola Beresford [k], Tamara Zietek [l], Margot Siebler [l], Anne Simon [m], Helena Bielak [m], Henner Hollert [n], Yvonne Müller [n], Maike Harff [n], Sabrina Schiwy [n], Kirsten Simon [o], Steffen Uhlig [a,*]

a QuoData GmbH, Prellerstr. 14, 01309 Dresden, Germany

b Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466 Seestadt OT Gatersleben, Germany

c Erftverband, Am Erftverband 6, 50126 Bergheim, Germany

d Institut für Energie- und Umwelttechnik e. V. (IUTA, Institute of Energy and Environmental Technology), Bliersheimer Str. 58-60, 47229 Duisburg, Germany

e University of Antwerp, Systemic Physiological and Ecotoxicological Research (SPHERE), Groenenborgerlaan 171/U7, 2020 Antwerp, Belgium

f Leibniz Institute for Baltic Sea Research Warnemünde, Department Marine Chemistry, Seestraße 15, 18119 Rostock, Germany

g Swiss Centre for Applied Ecotoxicology Eawag-EPFL, Überlandstraße 133, 8600 Dübendorf, Switzerland

h Technical University Dresden, Institute of Natural Science, Bioprocess Engineering, Helmholtzstraße 10, 01062 Dresden, Germany

i Technical University of Munich, Department of Civil, Geo and Environmental Engineering, Chair of Urban Water Systems Engineering, Am Coulombwall 3, 85748 Garching, Germany

j SYNLAB Umweltinstitut GmbH, Hauptstraße 105, 04416 Markkleeberg, Germany

k Brunel University London, Institute for Environment, Health and Societies, Halsbury Building, UB8 3PH Uxbridge, United Kingdom

l Technical University of Munich, Department of Nutritional Physiology, Gregor-Mendel-Straße 2, 85354 Freising, Germany

m IWW Rheinisch-Westfälisches Institut für Wasserforschung gemeinnützige GmbH, Moritzstr. 26, 45476 Mülheim an der Ruhr, Germany

n RWTH Aachen University, Institute for Environmental Research, Worringerweg 1, 52074 Aachen, Germany

o New diagnostics GmbH, Pollinger Straße 11, 81377 München, Germany

(*) Corresponding author: Steffen Uhlig, QuoData GmbH, Prellerstr. 14, 01309 Dresden, Germany, Email: uhlig@quodata.de, Phone: +49 351 40 28 867 0, Fax: +49 351 40 28 867 19.

Table of Contents

1 Additional sample information

The municipal effluent (#1) and influent (#3) samples were obtained from the waste water treatment plant in Koblenz. This plant has a mechanical and biological treatment of the waste water. The treatment plant is designed for a population equivalent of 320 000 and has a maximum annual treatment capacity of 20 million m^3. In total, approx. 12 L influent and 2 x 12 L effluent were taken. Sample #2 was obtained by separating 10 L of the effluent sample and spiking with 10 ng/L 17α-ethinylestradiol (EE2).

The influent hospital waste water sample (#4) was obtained from the waste water treatment plant at the Marienkrankenhaus in Gelsenkirchen. It was established in 2011 within the PILLS Project[1-3] and exclusively treats hospital wastewater. It is equipped with a membrane bioreactor (MBR) for the primary and secondary treatment, followed by ozonation as tertiary treatment and, finally, a sand filtration step. The hospital has approximately 580 beds, 1200 employees, and 75,000 patients per year. The WWTP is constructed for a maximum inflow of 25 m^3 wastewater per h and an average inflow of 200 m^3 per day. A total volume of 6 L was sampled and 2 L were separated for the interlaboratory trial.

The surface water sample (#5) was obtained in the city of Koblenz. Approximately 2 x 12 L were taken. Sample #6 was obtained by separating 10 L and spiking with 15 ng/L EE2.

The Baltic Sea sample (#7) was obtained in Heiligendamm. In total, 2 Liters were taken and the half was spiked with 17β-estradiol (E2) and EE2 (to achieve nominal doses of 1.4 and 8 ng/L, respectively).

In total, 1 L of the aqueous mixture of bisphenol A (BPA), bisphenol S (BPS) and bisphenol Z (BPZ) (#8) was prepared.

Samples were stored either in glass bottles (borosilicate glass) or in uncoated aluminium bottles with polytetrafluoroethylene (PTFE) lined caps.

References

[1] PILLS Project (2012): Pharmaceutical residues in the aquatic system – a challenge for the future. Insights and activities of the European cooperation project PILLS. www.pills-project.eu.

[2] Boehling, E., Adamczak, K., Nafo, I., Evenblij, H., Cornelissen, A., McArdell, C., Pahl, O., Dagot, C. (2012). Pharmaceutical input and elimination from local sources. Final report of the European cooperation project PILLS. www.pills-project.eu.

[3] Nafo, I., Lyko, S. (2012). Full-scale plant for the elimination of pharmaceuticals in hospital wastewater – Comparison of advanced treatment technologies. 16th International EWA Symposium, IFAT 2012, Munich.

2 Calculation of the dose response relationship

The dose response curve for the reference compound 17*β*-estradiol can be fitted with a four parametric logistic function (Eq. 1), which is equivalent to the Hill-Slope model[4]:

$$f(x) = y = \frac{A-D}{1+\left(\frac{X}{\mathrm{EC}_{50}}\right)^{B}} + D \qquad \text{(Eq. 1)}$$

where

y	is the corrected absorbance at concentration x
x	is the concentration which activates the test to the corrected absorbance y
A	is the mean value of y without estrogenic effects (bottom curve point)
D	is the mean value of y with the maximal activation (curve plateau)
EC_{50}	is the mean effect concentration at which the estrogenic effect attains half of its maximum
B	is proportional to the slope of the function at EC_{50}.

If more than three replicates are available an outlier test (Grubb's test)[5] can be performed for each calibration level before curve fitting. The quality of curve fitting is assessed using the relative standard deviation of residuals (RSD), with RSD ≤ 12% presenting good quality.

The inverse function can be used to calculate the concentration for a specific corrected absorbance (Eq. 2):

$$x = \left(\frac{y-A}{D-y}\right)^{\frac{1}{B}} \cdot EC_{50} \qquad \text{(Eq. 2)}$$

References

[4] Hill, A. V. (1910). The possible effects of aggregation of the molecules of hemoglobin on its dissociation curves. J. Physiol., 40, 4–7.

[5] Motulsky, H. (2010). Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking, Oxford Univ. Press, New York, 2010.

3 Calculation of the expected EEQ for spiked samples

For calculation of the EEQ for the spiked sample #2, #6, #7 and #8 the relative potencies of the added compounds were taken into account. The relative potency is used to express the biological activity of a compound in comparison to the natural hERα ligand and calibration standard E2. The relative potency (%) is calculated according to Eq. 3:

$$REP_{\text{compound}} = \frac{\text{EC}_{50}{}_{17\beta\text{-estradiol}}}{\text{EC}_{50}{}_{\text{compound}}} \times 100 \qquad \text{(Eq. 3)}$$

The relative potencies for the spiked compounds are summarized in Table 1.

Table 1: Relative potencies to 17β-estradiol (%).

Compound	Relative potency to 17β-estradiol (%)
17β-estradiol (E2)	100
17α-ethinylestradiol (EE2)	120
Bisphenol A (BPA)	0.00168
Bisphenol S (BPS)	0
Bisphenol Z (BPZ)	0.0123

The expected EEQ of compound is calculated according to Eq. 4:

$$\text{Expected EEQ} = \sum_{i=1}^{n} C_i \cdot REP_i \qquad \text{(Eq. 4)}$$

where

C_i is the concentration of compound i

REP_i is the relative potency of compound i

n is the number of compounds

The resulting expected EEQ for the samples #2, #6, #7 and #8 are shown in Table 2.

Table 2: Expected EEQ for the samples #2, #6, #7 and #8.

Nr.	Sample	Expected (ng/L EEQ)
2	Effluent, municipal-WWTP +10 ng/L EE2	10 ng/L EE2 x 1.2 = 12.0 ng/L EEQ
6	Rhine + 15 ng/L EE2	15 ng/L EE2 x 1.2 = 18.0 ng/L EEQ
7	BPA, BPS, BPZ	242 µg/L BPA x 0.0000168 + 0 µg/L BPS x 0 + 39 µg/L BPZ x 0.000123 = 0.0089 µg/L EEQ = 8.9 ng/L EEQ
8	Baltic Sea + 1.4 ng/L E2 and 8 ng/L EE2	1.4 ng/L E2 + 1.2 x 8 ng/L EE2 = 11 ng/L EEQ

4 Calculation of the repeatability and reproducibility standard deviation

Repeatability and reproducibility standard deviation were calculated according to ISO 5725-2[6]. The corresponding formulas are listed below. The respective standard deviation is calculated by the square root of the variance.

The reproducibility variance s_R^2 is calculated according to Eq. 5:

$$s_R^2 = \sqrt{s_L^2 + s_r^2} \qquad \text{(Eq. 5)}$$

with

s_L^2 is the estimate of between-laboratory variance

s_r^2 is the estimate of repeatability variance

The repeatability variance s_r^2 is calculated according to Eq. 6:

$$s_r^2 = \frac{\sum_{i=1}^{p}(n_i-1)s_i^2}{\sum_{i=1}^{p}(n_i-1)} \qquad \text{(Eq. 6)}$$

with

p is the number of laboratories (after outlier elimination)

n_i is the number of test results in one laboratory

s_i^2 is the estimate of variance within laboratory test results

The variance within laboratory test results s_i^2 is calculated according to Eq. 7:

$$s_i^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (y_{ik} - \bar{y}_i)^2 \quad \text{(Eq. 7)}$$

with

n_i is the number of test results in one laboratory

$\bar{y}_i$ is the mean value within one laboratory

y_{ik} is the single laboratory test result

The between-laboratory variance s_L^2 is calculated according to Eq. 8:

$$s_L^2 = \frac{s_d^2 - s_r^2}{\bar{\bar{n}}_j} \quad \text{(Eq. 8)}$$

with

$$s_d^2 = \frac{1}{p-1} \left[\sum_{i=1}^{p} n_i (\bar{y}_i)^2 - (\bar{\bar{y}})^2 \sum_{i=1}^{p} n_i \right]$$

$$\bar{\bar{n}}_j = \frac{1}{p-1} \left[\sum_{i=1}^{p} n_i - \frac{\sum_{i=1}^{p} n_i^2}{\sum_{i=1}^{p} n_i} \right]$$

$\bar{\bar{y}}$ is the overall mean value (across laboratories)

p is the number of laboratories (after outlier elimination)

n_i is the number of test results in one laboratory

$\bar{y}_i$ is the mean value within one laboratory

References

[6] ISO 5725-2. 2002. Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of standard measurement method (ISO 5725-2:1994 including Technical Corrigendum 1:2002). Berlin Germany: DIN Deutsches Institut für Normung e.V.

5 Flow diagram of A-YES® procedure