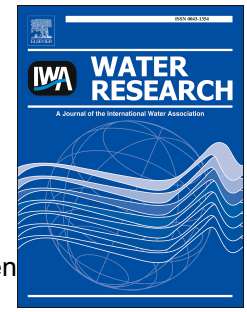
Accepted Manuscript



What you extract is what you see: Optimising the preparation of water and wastewater samples for *in vitro* bioassays

Aennes Abbas, Ilona Schneider, Anna Bollmann, Jan Funke, Jörg Oehlmann, Carsten Prasse, Ulrike Schulte-Oehlmann, Wolfram Seitz, Thomas Ternes, Marcus Weber, Henning Wesely, Martin Wagner

PII: S0043-1354(18)31067-4

DOI: https://doi.org/10.1016/j.watres.2018.12.049

Reference: WR 14345

To appear in: *Water Research*

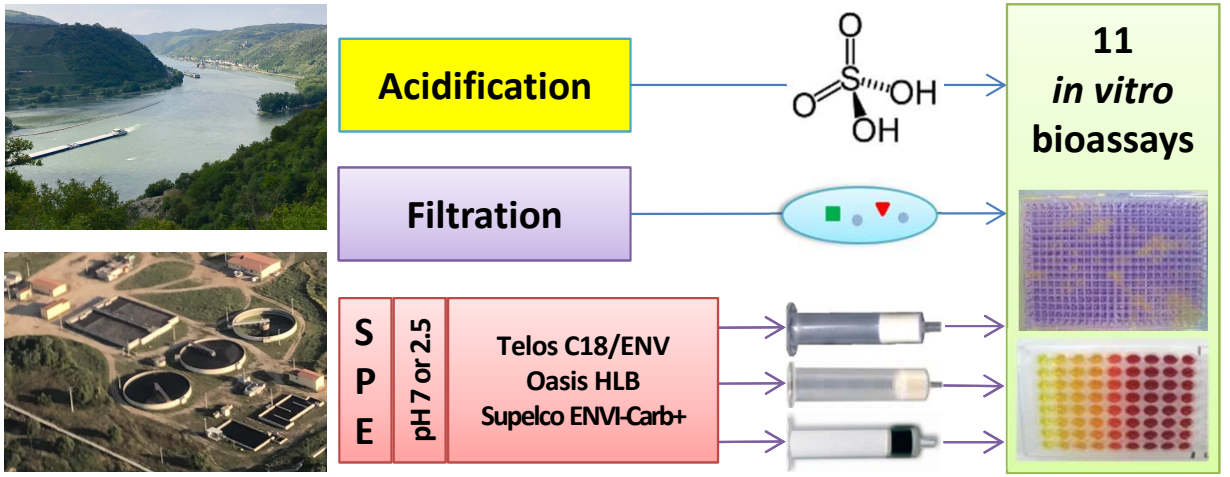
Received Date: 26 August 2018

Revised Date: 19 December 2018

Accepted Date: 21 December 2018

Please cite this article as: Abbas, A., Schneider, I., Bollmann, A., Funke, J., Oehlmann, J., Prasse, C., Schulte-Oehlmann, U., Seitz, W., Ternes, T., Weber, M., Wesely, H., Wagner, M., What you extract is what you see: Optimising the preparation of water and wastewater samples for *in vitro* bioassays, *Water Research*, https://doi.org/10.1016/j.watres.2018.12.049.

WR46063
Graphical abstract - revised version



What you extract is what you see: Optimising the preparation of water and wastewater samples for *in vitro* bioassays

Aennes Abbas[1,*] & Ilona Schneider[1,*], Anna Bollmann[2], Jan Funke[3], Jörg Oehlmann[1], Carsten Prasse[4], Ulrike Schulte-Oehlmann[1], Wolfram Seitz[2], Thomas Ternes[5], Marcus Weber[6], Henning Wesely[5], Martin Wagner[1, 7]

[1] Department Aquatic Ecotoxicology, Goethe University Frankfurt am Main, Max-von-Laue-Str. 13, D-60438 Frankfurt, Germany

[2] Zweckverband Landeswasserversorgung, Am Spitzigen Berg 1, D-89129 Langenau, Germany

[3] IWW Rheinisch-Westfälisches Institut für Wasser Beratungs- und Entwicklungsgesellschaft mbH, Moritzstraße 26, D-45476 Muelheim an der Ruhr, Germany

[4] Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore, MD, 21218, USA

[5] Federal Institute of Hydrology (BfG), Am Mainzer Tor 1, D-56068 Koblenz, Germany

[6] Department of Numerical Analysis and Modelling, Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Takustraße 7, D-14195 Berlin, Germany

[7] Department of Biology, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

[*] These authors contributed equally to the work.

Email addresses of corresponding authors: abbas@bio.uni-frankfurt.de, i.schneider@bio.uni-frankfurt.de

Abstract

The assessment of water quality is crucial for safeguarding drinking water resources and ecosystem integrity. To this end, sample preparation and extraction is critically important, especially when investigating emerging contaminants and the toxicity of water samples. As extraction methods are rarely optimised for bioassays but rather adopted from chemical analysis, this may result in a misrepresentation of the actual toxicity.

In this study, surface water, groundwater, hospital and municipal wastewater were used to characterise the impacts of common sample preparation techniques (acidification, filtration and solid phase extraction (SPE)) on the outcomes of eleven *in vitro* bioassays. The latter covered endocrine activity (reporter gene assays for estrogen, androgen, aryl-hydrocarbon, retinoic acid, retinoid X, vitamin D, thyroid receptor), mutagenicity (Ames fluctuation test), genotoxicity (umu test) and cytotoxicity. Water samples extracted using different SPE sorbents (Oasis HLB, Supelco ENVI-Carb+, Telos C18/ENV) at acidic and neutral pH were compared for their performance in recovering biological effects.

Acidification, commonly used for stabilisation, significantly altered the endocrine activity and toxicity of most (waste)water samples. Sample filtration did not affect the majority of endpoints but in certain cases affected the (anti-)estrogenic and dioxin-like activities. SPE extracts (10.4× final concentration), including WWTP effluents, induced significant endocrine effects that were not detected in aqueous samples (0.63× final concentration), such as estrogenic, (anti-)androgenic and dioxin-like activities. When ranking the SPE methods using multivariate Pareto optimisation an extraction with Telos C18/ENV at pH 7 was most effective in recovering toxicity. At the same time, these extracts were highly cytotoxic masking the endpoint under investigation. Compared to that, extraction at pH 2.5 enriched less cytotoxicity.

In summary, our study demonstrates that sample preparation and extraction critically affect the outcome of bioassays when assessing the toxicity of water samples. Depending on the

water matrix and the bioassay, these methods need to be optimised to accurately assess water quality.

Keywords

Activated carbon, advanced treatment, endocrine disrupting chemicals, micropollutants, ozonation, transformation products, tertiary treatment

Abbreviations

9-cis-RA	9-*cis* retinoic acid
4-NOPD	4-nitro-*o*-phenylenediamine
4-NQO	4-nitroquinoline N-oxide
AhR	aryl-hydrocarbon receptor
Ames	bacterial reverse mutation test
ANOVA	analysis of variance
at-RA	all-*trans* retinoic acid
CAS	Chemical Abstracts Service
CPRG	chlorophenol red-β-D-galactopyranoside
DIN	German Institute of Standardisation (Deutsches Institut für Normung)
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DOC	dissolved organic carbon
E_2	17β-estradiol
EC	European Commission
EC_{50}	Median effect concentration
EDCs	endocrine disrupting chemicals
EFF	effluent
FB	filtration basin
Flu	flutamide
GW	groundwater
hAR	human androgen receptor
hERα	human estrogen receptor α

HOS	hospital
IB	infiltration basin
INF	influent
IR	induction rate
ISO	International Standard Organisation
lacZ	bacterial gene coding β-galactosidase
LOQ	limit of quantification
MS	microsieve
n.a.	not analysed
NF	nitrofurantoin
β-NF	β-naphthoflavone
n.s.	not significant
OD	optical density
OHT	4-hydroxytamoxifen
ONPG	*o*-nitrophenyl β-D-galactopyranoside
PTFE	polytetrafluorethylene
RARα	retinoic acid receptor α
RXRα	retinoid X receptor α
SOS	inducible bacterial DNA repair system
SPE	solid phase extraction
SW	surface water
T	testosterone
T_3	3,3′,5-triiod-L-thyronine
TA100	recombinant strain of *Salmonella typhimurium*
TA98	recombinant strain of *Salmonella typhimurium*
TRα	thyroid receptor α
TSS	total suspended solids
umu	bacterial test for the determination of genotoxicity
umuC	bacterial ultra violet mutagenesis gene C
US EPA	United States Environmental Protection Agency
uvrB	gene of a bacterial DNA repair system
VDR	vitamin D receptor
WWTP	wastewater treatment plant
YAAS	yeast anti-androgen screen

YAES	yeast anti-estrogen screen
YAS	yeast androgen screen
YDS	yeast dioxin screen
YES	yeast estrogen screen

57

1 Introduction

Anthropogenic micropollutants typically occur at nanogram to microgram per litre concentrations in urban water cycles. Micropollutants may pose a risk to ecosystems as they have been associated with negative impacts on aquatic biota (Malaj et al. 2014, Prasse et al. 2015). Micropollutants are found amongst pharmaceuticals, personal care products, industrial chemicals, pesticides and biocides (Kümmerer 2011) that are emitted from different anthropogenic sources. These sources can be diffuse, such as agricultural runoffs, or point sources, such as wastewater treatment plant (WWTP) discharges. Several studies have demonstrated an incomplete removal of micropollutants and relevant toxicity after conventional wastewater treatment using activated sludge (Prasse et al. 2015). Therefore, advanced wastewater treatment technologies utilising chemical oxidation or adsorption are being developed to increase the removal of micropollutants and toxicity (Miklos et al. 2018, Rizzo 2011). *In vitro* bioassays play a crucial role for the ecotoxicological assessment of water and wastewater quality because they determine the joint toxicity caused by complex samples, often regarding a specific mode of action (Escher et al. 2014, 2018, Leusch et al. 2017). Bioassays are routinely used in monitoring campaigns and sufficiently advanced to be integrated into water and wastewater regulations (Brack et al. 2017, Escher et al. 2018).

Environmental water and wastewater samples represent complex mixtures of known and unknown chemicals (Schwarzenbach et al. 2006) and are characterised by a variable composition with respect to matrix parameters (e.g., suspended solids or dissolved organic carbon (DOC)). The toxicity of the samples is mainly determined by the type and concentration of the active, anthropogenic or natural compound(s) and their cumulative effects. However, the sample matrix can also affect the outcome of a bioassay (Janošek et al. 2007, Neale et al. 2015). In addition, samples can undergo physicochemical and biological processes that can transform or degrade the active compounds and may, therefore, modulate the biological effects under investigation.

Because of their ability to reduce matrix effects, to preserve and to concentrate dissolved organic chemicals in aqueous samples, different extraction methods, such as solid phase extraction (SPE), are used in chemical and ecotoxicological studies (Prasse et al. 2015). While sample preparation and extraction methods are commonly optimised for chemical analysis, i.e., to maximise the recovery of specific target compounds, this is rarely done in bioassay studies (Bistan et al. 2012, Neale et al. 2018, Schulze et al. 2017) because the "true" toxicity to recover remains unknown. Thus, standard extraction procedures adapted from chemical analysis are mainly used. Comparative studies have indicated that such chemical "standard" methods can be ineffective in extracting unknown, active compounds from water samples (Hendriks et al. 1994, Wagner and Oehlmann 2011). Because this can lead to an underestimation or false negative results, optimising sample preparation and extraction to recover a maximum of toxicity should be imperative for bioassay studies.

The aim of our study was to assess the impacts of common samples preparation methods on the detection of environmentally-relevant endocrine activities, genotoxicity and cytotoxicity in water and wastewater samples. These samples consisted of surface water, groundwater, hospital wastewater, raw (untreated), conventionally-treated and ozonated wastewater. These samples consisted of grab as well as composite samples with low to high contamination degrees to allow for an optimal comparison of SPE methods. The toxicity of untreated aqueous samples and samples that were acidified (24 h at pH 2.0) or filtered (1 µm pore size) was compared in eleven *in vitro* bioassays. Furthermore, the effectiveness of six SPE methods was compared by extracting samples with three SPE sorbents at acidic and neutral sample pH (2.5 and 7 right before loading). Aqueous and extracted samples were analysed using bioassays for nine human hormone receptors, the umu test and the Ames fluctuation test. The outcome of these bioassays was evaluated by a multivariate Pareto optimisation to identify the most effective sample extraction method.

2 Material and methods

2.1 Characterisation of sampling sites

Sampling locations were selected according to their relevance and representativeness regarding the water cycle in a model region in Baden-Württemberg (Germany, Table 1, samples 1−14, see Seitz and Winzenbacher 2017 for details). Samples comprised influents and effluents of three municipal WWTPs (WWTP 1−3) with activated sludge treatment, two hospital wastewaters, three rivers (surface water), influent and effluent of a filtration basin, two storm water sedimentation tanks, one storm water overflow tank (with infiltration basin), and three groundwater monitoring wells (hotspots). Additional wastewater samples were taken from a pilot WWTP (WWTP 4) in Hessen, Germany (Knopp et al. 2016), equipped with advanced treatment technologies, including a full-scale ozonation of conventionally treated effluent (activated sludge) filtered using a microsieve (MS, filtration at mesh size: 10 µm) to reduce total suspended solids (TSS, Table 1, samples 15−19). The ozonation was performed with 0.33 g O_3/g DOC.

2.2 Collection of water and wastewater samples

Wastewater samples (influent and effluent) from the municipal WWTPs in Baden-Württemberg (sampling period: April (B), July (C, D) and December (E) 2012) and the pilot WWTP in Hessen (sampling period: March (A), April (B), July 2012 (C, D) and December (E) 2012, January (F) 2013) were collected as grab (samples 1, 6, 8–14, 18) or 24 h composite samples (samples 2−5, 7, 15−17, Table 1). The results of corresponding samples (e.g., influents or effluents) were compared to each other, only, with exception of the event-driven sampling of samples 6 and 7 (FB-IN and FB-OUT, Table 1). For the collection of 24 h composite samples, wastewater was continuously pumped through polytetrafluoroethylene (PTFE) tubes into 5 L glass bottles. Bottles were kept at 4°C in darkness during sampling. Hospital effluents, surface waters, samples from storm water sedimentation and an overflow

tank (with infiltration basin) as well as groundwater hotspots were grab samples (sampling period: April (B), July (C, D) and December (E) 2012). All samples were stored at 4°C in pre-cleaned, amber glass bottles with PTFE lids and analysed (aqueous samples for acidification and filtration experiments) or further processed (comparison of SPE methods) within 48 h after sampling.

2.3 Sample preparation

2.3.1 Acidification for testing aqueous samples

One aliquot (40 mL) of the aqueous (waste)water sample was kept at the original pH, another aliquot (40 mL) was acidified with sulphuric acid (5 mol/L, purity "pro analysi") to pH 2.0 directly after sampling. After storage for 24 h at 4°C in the dark, acidified samples were neutralised with sodium hydroxide (1 mol/L, purity "pro analysi") to pH 7 prior to analysing the aqueous samples in the bioassays (in contrast to short-term acidification for SPE, 2.3.3).

2.3.2 Filtration for testing aqueous samples

One aliquot of the (waste)water sample remained unfiltered while another aliquot was filtered using glass fibre filters (Whatman GF6, pore size 1 µm) to reduce TSS. Selected filtered and unfiltered aqueous samples were tested as aqueous samples (not SPE extracts) in the *in vitro* assays (2.4). The glass fibre filters containing the retentate were suspended in ultrapure water (10 min in an ultrasonic bath) and the obtained aqueous suspensions were analysed for endocrine activity retained on the filters. A filter control was run and analysed in parallel: ultra-pure water was filtered and an empty glass fibre filter was suspended as well. Additionally, the influence of a microsieve (mesh size: 10 µm) on endocrine and genotoxic activity of conventionally treated effluent after final sedimentation at WWTP 4 was investigated by taking wastewater samples before and after the microsieve. A microsieve control was analysed as well (data not shown): fragments of the microsieve were incubated in

ultra-pure water and in methanol for 70 d and the resulting suspensions were tested in the *in vitro* bioassays.

2.3.3 Solid phase extraction

Three commonly used types of SPE sorbents were tested for the recovery of endocrine, genotoxic, and mutagenic activities: Oasis HLB (200 mg), Kinesis Telos C18/ENV (500 mg C18, 200 mg ENV) and Supelco ENVI-Carb+ (200 mg). Prior to sample loading, the cartridges were conditioned as follows: Oasis HLB and Telos C18/ENV were conditioned consecutively with 1 x 2 mL heptane, 1 x 2 mL acetone, 3 x 2 mL methanol (LC-MS Optigrade) and 4 x 2 mL ultrapure water. Supelco ENVI-Carb+ cartridges were turned (top to bottom) before they were conditioned with 1 x 2 mL acetone and 1 x 2 mL methanol. Afterwards, the columns were turned again (loading direction) and conditioned with 1 x 2 mL acetone, 3 x 2 mL methanol and 4 x 2 mL ultrapure water. For each sample, 500 mL sample was extracted at two pH values, neutral (pH 7) and acidified with sulphuric acid (3.5 mol/L) to pH 2.5.

SPE was performed within 48 h after collection and directly after acidification. The columns were dried under a stream of nitrogen and stored at -20°C. Samples extracted at neutral pH were eluted with 5 x 2 mL acidified methanol and 5 x 2 mL acetone, each containing 0.2% formic acid. Acidified samples were consecutively eluted with 5 x 2 mL methanol and 5 x 2 mL acetone at neutral pH. After adding 100 µL dimethyl sulfoxide (DMSO), the combined methanol-acetone extract was concentrated to 100 µL final volume under a gentle nitrogen stream. The extracts (5000-fold concentrated compared to the aqueous sample) were stored at -20°C until testing. A SPE blank was prepared in parallel to each sampling campaign to control for contamination by loading each column type with ultrapure water and extracting them with neutral and acidified methanol and acetone, respectively.

2.4 *In vitro* bioassays

2.4.1 Recombinant yeast screens for endocrine activities

In this study, nine recombinant yeast-based reporter-gene assays were used to detect endocrine activities: Yeast Estrogen Screen (YES, human estrogen receptor α (hERα)), Yeast Anti-Estrogen Screen (YAES), Yeast Androgen Screen (YAS, human androgen receptor (hAR)), Yeast Anti-Androgen Screen (YAAS) first described by Routledge and Sumpter (1996) and Sohoni and Sumpter (1998), Yeast Dioxin Screen (YDS, aryl-hydrocarbon receptor (AhR, Miller 1997)), as well as yeast two-hybrid assays for retinoic acid receptor α (RARα), retinoid X receptor α (RXRα), vitamin D receptor (VDR) and thyroid receptor α (TRα) introduced by Inoue et al. (2009). We used yeast-based assays rather than mammalian cell lines because they are robust in terms of cytotoxicity, because they have been validated by ISO (ISO 19040-1:2018) and to compare the results to our previous work.

All bioassays have the same principle: The activation of the respective receptor by chemicals present in the sample triggers the expression of β-galactosidase, which cleaves the chromogenic substance chlorophenol red-β-D-galactopyranoside (CPRG; CAS 99792-79-7, Sigma-Aldrich, Germany). The intensity of the colour change (yellow to red) is proportional to the agonistic activity of the sample and is measured with a photometer (Multiskan Ascent, Thermo Fisher Scientific, Braunschweig, Germany) at a wavelength of 540 nm (OD_{540}). To screen for antagonistic activities (YAES and YAAS), a known agonist is added. Thus, antagonistic compounds reduced the reporter gene activity induced by the agonist.

All bioassays were conducted in 96-well microtiter plates (f-form, VWR Darmstadt, Germany) as described previously (Völker et al. 2016, Wagner et al. 2013, Stalter et al. 2011, Wagner and Oehlmann 2009). In brief, aqueous samples were analysed in eight replicates with a dilution factor of 1.6 (i.e., 0.625-fold final sample concentration). SPE extracts were diluted 480-fold resulting in a 10.4-fold final sample concentration (0.2% v/v solvent

concentration, eight replicates). This enrichment factor was used for all SPE extracts (compare 2.2 and Table 1). After 18−22 h incubation (depending on the assay) at 30°C and 1200 rpm, cell number (absorbance at 595 nm, OD_{595}, to detect cytotoxic effects) and reporter-gene activity (OD_{540}) were determined photometrically. In each assay and experiment, concentration-response curves for the appropriate reference compound were generated (see Table S1 and Figures S1−S5 for details).

The OD_{540} was corrected for the respective cell density (OD_{595}). If > 20% cytotoxicity occurred (see 2.5) results were not used. The corrected absorbance was normalised to the negative/solvent controls (0%) and the maximum activity of the reference compound (100%) to calculate relative activities (%). For the antagonist assays, a control without agonist was used to represent 100% receptor inhibition.

The limit of quantification (LOQ) was calculated for each bioassay and experiment using the mean activity of the negative control and adding threefold it's standard deviation. As the LOQs varied between bioassays and experiments, they were not shown for the sake of clarity. However, in general only results above the LOQs were considered. In a few cases, such as estrogenic activity, lower activities were shown because of their ecotoxicological relevance (low effect threshold) and for comparing WWTP effectivities.

2.4.2 Genotoxicity assay (umu test)

Genotoxic effects were assessed using the umu test (ISO 13829) with the genetically modified *Salmonella typhimurium* strain TA1535 (pSK1002). The umu test detects primary reversible or irreversible DNA damages that induce the expression of the DNA SOS-repair system associated with the UV mutagenesis gene C (umuC gene). Genotoxic substances in the samples lead to an expression of β-galactosidase from the umuC-*lacZ* construct. The reporter-gene activity is determined by the cleavage of the chromogenic substance *o*-nitrophenyl β-D-galactopyranoside (ONPG, CAS 369-07-3, Sigma-Aldrich, Germany). The umu test was

conducted as described by Magdeburg et al. (2014). In brief, aqueous samples were analysed after sterile filtration (injection filter with PTFE membrane: pore size 0.2 µm, neoLab, Germany) with a dilution factor of 1.7 and SPE extracts in a 20-fold final sample concentration (0.4% v/v solvent) in eight replicates. Ten concentrations between 5−2000 µg/L final concentration in the well of 4-nitroquinoline N-oxide (4-NQO; CAS 56-57-5, Sigma-Aldrich, Germany) were used as genotoxic reference compound (Table S1). Cytotoxicity (OD_{595}) and genotoxicity (OD_{414}) were determined photometrically. The OD_{414} was corrected for the respective cell density (OD_{595}) if no cytotoxicity occurred (see 2.5). A linear regression line was generated using the corrected OD_{414} of the reference compound (Figure S6). The induction rate (IR) was calculated using the corrected OD_{414} of the samples. An $IR \geq 1.5$ is considered potentially genotoxic.

2.4.3 Mutagenicity assay (Ames fluctuation test)

Mutagenic effects (i.e., irreversible DNA damage) were analysed using the Ames fluctuation test (ISO/DIN 11350) with two genetically modified strains of *Salmonella typhimurium* (TA98 and TA100). The assay detects the induction of point mutations in special marker genes coding for enzymes involved in histidine biosynthesis as frameshift mutations (TA98) and base pair substitutions (TA100). To increase sensitivity, the strains TA98 and TA100 have a mutation in the *uvrB* DNA repair gene. In the absence of mutagens, the strains do not grow in histidine-free medium and a reverse mutation in the marker genes enables histidine synthesis and thus growth. This leads to a pH change in the assay medium that is determined photometrically at a wavelength of 414 nm.

The Ames test was conducted as described by Magdeburg et al. (2014). In brief, aqueous samples were tested after sterile filtration (see 2.3.2) with a dilution factor of 1.25 and SPE extracts in a 10-fold final sample concentration (0.2% v/v solvent). Mutagenic reference compounds were used as positive controls (TA98: 10 mg/L final concentration in the well 4-

nitro-*o*-phenylenediamine (4-NOPD, CAS 99-56-9, Sigma Aldrich, Germany, Table S1); TA100: 0.25 mg/L final concentration in the well nitrofurantoin (NF; CAS 67-20-9, Sigma Aldrich, Germany, Table S1). The mutagenic activity of the sample was determined photometrically with a cut-off value at a wavelength of 414 nm by counting the number of wells that shifted from purple (negative) to yellow (positive).

2.5 Data analysis

In this study, cytotoxicity was defined as a cell number in the sample of ≤ 80% compared to the negative control (solvent control) analysed in parallel in each experiment.

Statistical analyses were performed using GraphPad Prism (version 5.03, GraphPad Software Inc., San Diego, California, USA). Datasets were analysed using the D'Agostino and Pearson omnibus normality test for Gaussian distribution and the Bartlett's test for homogeneity of variances. In case of a normal distribution and equal variances significant differences between the datasets were determined using a one-way ANOVA with Dunnett's post-test. If the datasets were not normally distributed, the nonparametric Kruskal-Wallis test with Dunn's post-test was used. An unpaired t-test was used to determine significant differences between neutral and acidified samples and unfiltered and filtered samples. A p-value ≤ 0.05 was considered significant.

The mathematical part of the methodological optimisation was carried out using a Pareto strategy (Ehrgott 2000) further adapted for the multivariate optimisation, similar to the use of colour coding in *in silico* toxicology (Durmaz et al. 2015). The main optimisation criterion was to assess sample preparation methodologies that achieved the highest measured biological activity in six different parameters. Pareto thereby classified a preparation method as non-optimal, if another preparation method exists that delivers "better" values regarding *all* parameters (YES, YAS, etc.) and *all* tested samples. Non-optimal preparation methods are

289 excluded from the list leading to a ranked set of Pareto-optimal sample preparation methods.

290 The applied strategy also tackled scenarios with missing data.

3 Results and discussion

3.1 Sample acidification for testing aqueous samples

Analytical chemists use acid as a standard method to stabilise aqueous samples and prevent the biodegradation of (micro)pollutants (Prasse et al. 2015). Stabilisation is thought to occur by deactivating microorganisms (Baker and Kasprzyk-Hordern 2011, US EPA 2010) that may use target analytes as substrates. Therefore, the procedure is often adopted in ecotoxicology for conserving the toxicity of samples but often without studying its effectiveness.

The present results show that sample acidification and storage over 24 h significantly affected the endocrine activities and mutagenicity of aqueous samples compared to the samples kept at neutral pH (Figure 1, full data sets in Table S2). Focusing on a change of the endocrine activities or mutagenicity of ≥ 10%, untreated wastewater was most affected by acidification (Table S3) whereby 50% of the assays (n = 22) showed decreased activities between -13 and -94%. In case of the influent and effluent of the filtration basin 32% of the bioassays (n = 22) indicated altered activities between -13% and -37%. Groundwater (9%, n = 33), ozonated wastewater (9%, n = 11) and surface water (3%, n = 33) were least affected (Table S3).

Regarding the different bioassays, the activities in the YAES, RXR and Ames TA100 assays were most affected by acidification (Table S4). 65% of the YAES experiments showed decreased (-13 to -32%) or increased (+15 to +34%) activities (Figure 1A). The Ames TA100 was affected in 24% of the experiments with decreasing (-13 to -77%) as well as increasing mutagenicity (+17%) compared to neutral samples (Figure 1C, Table S4). Acidification caused the highest decrease of mutagenicity in the Ames TA98 with -94% followed by the RAR assay with -88% (Figure 1B). In the remaining bioassays, low endocrine or genotoxic activities were detected. Thus, no conclusion of the influence of acidification on these endpoints was possible (Figure S7, Table S2).

In summary, sample acidification led to a decrease (-13 to -94%) of activity in 81% and to an increase (+10 to +34%) of activity in 19% of the cases (n = 32). This indicates that sample

acidification significantly affects the outcomes of bioassays. Two hypotheses may explain the changes in toxicity: 1) In acidified samples, acids may interfere with active chemicals or 2) in neutral samples, microbial activity may degrade or transform the active chemicals.

Basically, the key question is whether the neutral (hypothesis 1) or the acidified sample (hypothesis 2) represent the "true" toxicity. For chemical analysis, there is consensus that acidification stabilises most compounds and prevents microbial degradation (Baker and Kasprzyk-Hordern 2011, Vanderford et al. 2011, US EPA 2010). However, our data implies that besides few exceptions the *in vitro* activity is lower at acidic compared to neutral pH (Figure 1, Table S2). Accordingly, samples at a neutral pH may better represent the actual toxicity. If this hypothesis holds true, an acidification of samples would either reduce the concentration of active chemicals by increasing adsorption to suspended matter (Baker and Kasprzyk-Hordern 2011) or by increasing hydrolysis (Prasse et al. 2015).

Alternatively, it can be assumed that the higher activity in neutral samples is an artefact caused by a change in sample composition. Here, continuous microbial activity may deconjugate compounds resulting in a higher biological activity. This occurs during biological wastewater treatment (Andersen et al. 2003, Koh et al. 2008, Wu et al. 2017). However, an on-going microbial degradation of active compounds would counteract this process (Giebner et al. 2018).

In reality, the toxicity of an aqueous sample may change at either neutral or acidic pH. As this depends on the chemical and biological composition of a sample, it is difficult to generalise which condition best represents the actual toxicity. Based on the present data, we argue that a neutral pH comes closest to reality, as the sample is minimally processed. In addition, a higher biological activity will result in a more protective water quality assessment if one accepts that the risks of false-positives outweighs the risk of false-negatives.

3.2 Sample filtration for testing aqueous samples

Sample filtration is beneficial to stabilise compounds (Baker and Kasprzyk-Hordern 2011), to avoid clogging of SPE cartridges, to remove TSS (Janex-Habibi et al. 2009) and to sterilise samples (Gehrmann et al. 2018). In the present study, unfiltered and corresponding glass fibre filtered (pore size 1 µm) aqueous samples as well as aqueous suspension of the filter retentates were compared to investigate the impacts of filtration on the toxicity. These comparisons further included a microsieve (pore size 10 µm) installed at one WWTP, which had a minimal effect on the toxicity (full data set in Table S5).

Focusing on a change of the different endocrine activities or mutagenicity of ≥ 10% again, the untreated wastewater was affected at most by filtration (Tables S5 and S6). Here, the toxicity was decreased by -20 and -54% and increased by +28 and +61% in 22% of the bioassays (n = 18, Figure 2A, 2B). For surface water, activities were altered in 14% (n = 7) of the bioassays with one affected endpoint (Figure S8). Conventionally treated wastewater and groundwater were less or not affected by filtration (Figures 2C and S8, Table S6).

Filtration had the strongest impact on the YAES (50% of the assays, n = 8; Table S7) followed by the YES and YAAS (25%, n = 8 each) and YDS (13%, n = 8). The effects observed in the other bioassays were too low to evaluate the influence of filtration on these endpoints (Figures 2 and S8, Table S5).

The aqueous suspension of the filter retentates also showed relevant changes in endocrine activities ≥ 10% in 19% (n = 36) of the yeast-based assays. The retentates were anti-estrogenic (57%, n = 7) and anti-androgenic (43%, n = 7) with activities from 21–80% (YAES) and 30–45% (YAAS, Table S5). In two samples, the endocrine activity in the filtered sample was significantly ($p \leq 0.001$) lower than in the unfiltered sample. As the retentate was also active, the activity was retained by filtration. In two cases, significantly higher ($p \leq 0.001$) activities were detected in the filtered compared to the unfiltered samples. Here, the retentate was active as well. In two YAES experiments, the endocrine activities in the filtered and unfiltered samples were on a comparable high level (84– 91%) and the retentate

was active as well (46 and 80%). One sample was not anti-androgenic as filtered and unfiltered water, but as filter retentate (45%, Figure S8, Table S5).

In summary, sample filtration led to a decrease (-18 to -54%) of activity in 33% and to an increase (+13 to +61%) of activity in 67% of the cases (n = 9) and, thus, has a significant impact on the bioassay results. The retention of particle-associated hormones and endocrine disrupting chemicals (EDCs) may explain this observation. This is supported by the detection of significant endocrine activities in the filter retentates and previous observations (Dagnino et al. 2010, Routledge 2003, Shieh et al. 2016).

Interestingly, few filtered samples had significantly higher endocrine activities than the corresponding unfiltered samples. For the WWTP effluent filtered by a microsieve we detected an approximately 2-fold increase in anti-estrogenic activity (Table S5). This may be the result of an altered ratio of agonistic and antagonistic activities (Ihara et al. 2014, Rao et al. 2014) or the leaching of "new" compounds by the filter materials (filter controls confirmed this was not the case). In the present case, dissimilar affinities towards filter materials and/or suspended solids (Ng and Cao 2015, Wangmo et al. 2018) could have resulted in a retention of antagonistic and thus increased agonistic activities in the filtrate and vice versa.

In conclusion, the application of sample filtration should be well-adjusted to the aims of a study, the characteristics of investigated (waste)water samples and bioassay specificities, as this is crucial to avoid misestimating the *in vitro* toxicity (Dagnino et al. 2010, EC 2003). In the present study, this was amongst others observed when evaluating the removal of (anti-)estrogenic and dioxin-like activities at WWTP 1 (Figure 2). Depending on whether the filtered or unfiltered samples are considered, one can conclude that the treatment in WWTP 1 either increases or decreases the toxicity.

3.3 Comparison of aqueous and extracted samples

Comparing the toxicity of aqueous samples and corresponding SPE extracts is rarely done but has a number of advantages, such as the possibility to calculate recovery rates and evaluate the environmental relevance of obtained results (Giebner et al. 2018, Muschket et al. 2017, Tousova et al. 2017, Wangmo et al. 2018).

In the present case, most aqueous samples induced minimal estrogenic, anti-androgenic and retinoic acid-like activities (Figure 3, Tables S8, S9, S10). However, anti-estrogenic activities between 21 and 91% were detected in all aqueous samples (Figure 3B). The activities were < 19% in the other bioassays (Figures 3D and S9, Table S8). In extracted samples, the estrogenic activity (≤ 8%, n = 35) was generally as low as in the corresponding aqueous samples (≤ 13%, n = 8; Figures 3 and 4, Table S9). The minor estrogenic activity detected in most samples in this study is in line with other studies on biological (Jalova et al. 2013, Keiter et al. 2006, Metcalfe et al. 2013) and advanced wastewater treatment (Ma et al. 2005, Maletz et al 2013).

The anti-estrogenic activity of the extracts was variable and, depending on the SPE method, in parts very high (13–89%, n = 35) and comparable to the corresponding aqueous samples (Figures 3B and 4). This indicated that the causative compounds were either only partially recovered or that the anti-estrogenicity of the aqueous samples is caused by the matrix (Neale et al. 2015). Interestingly, the high anti-estrogenic activities in the extracts point towards potential masking effects, whereby receptor antagonists reduce the detection of agonistic activity in water sample. This phenomenon has also been discussed by other authors (Giebner et al. 2018, Gehrmann et al. 2018, Ihara et al. 2014, Rao et al. 2014, Stalter et al. 2011). In addition, groundwater was significantly anti-estrogenic (Figure 3B, Table S8 and S9). This calls for further clarification regarding the presence of EDCs in groundwater.

In contrast, the anti-androgenic activity was low in most aqueous samples (≤ 5%, n = 7) but higher in the extracts (9–89%, n = 30, Figures 3C and 4, Table S9) indicating a successful extraction. Except for hospital wastewater, which may contain anti-androgenic

pharmaceuticals (Sohoni and Sumpter 1998, Stalter et al. 2011), the majority of aqueous samples exhibited only low androgenic and anti-androgenic activities (Figures 3C and S9, Table S8). The androgenic activities remained low in the corresponding extracts, whereas anti-androgenic activities were detected at moderate to high levels. As in case of the anti-estrogenic activity, androgen receptor antagonists may mask the androgenic activity. Such interactions were described for WWTP effluents (Leusch et al. 2017, Rao et al. 2014) and ozonated hospital wastewater (Gehrmann et al. 2018). The high removal of these activities reported for activated sludge treatment (Rao et al. 2014) and ozonation (Stalter et al. 2011) were not observed in this study.

The highest RAR activity was detected in aqueous hospital and untreated wastewater (HOS: 93%, INF-1: 23%) and corresponding extracts, depending on the SPE-method (HOS: 14–91%, INF-1: 0–54%; Figures 3E and 4, Table S9). This implies that the active compounds were only partially extracted. Only hospital and untreated wastewater induced RAR activities, which was removed in the effluent (Figure 3E, Tables S8 and S9). RXR activities were detected in extracted WWTP effluent and ozonated effluent (Figure S9, Table S8). So far, only few studies reported RAR and RXR activities in water (Inoue et al. 2009) and wastewater (Allinson et al. 2011, Inoue et al. 2011). In the experiments by Sawada et al. (2012) and Cao et al. (2009) these activities readily degraded during activated sludge treatment and lab-scale ozonation, respectively. Likewise, only a few studies exist on VDR- and TR-like activities in (waste)water samples (Escher et al. 2014, Inoue et al. 2011, Kusk et al. 2011, Leusch et al. 2017). In any case, activity levels in the present aqueous/extracted samples were negligible.

Moderate dioxin-like activities were detected in a number of extracted but none of the aqueous samples (Table S8). Highest activities were observed in raw, treated and hospital wastewater. Lowest activities were observed for ozonated wastewater and groundwater. Its removal during biological and advanced wastewater treatment has been observed in several

(Allinson et al. 2011, Loos et al. 2012, Stalter et al. 2011) but not all studies (Jia et al. 2015, Rao et al. 2014, Reungoat et al. 2010) supporting its detection in the present WWTP effluents. While none of the aqueous samples (n = 6) was active in the umu assay, 33% (n = 27) of the extracts were potentially genotoxic (Figure 3F, Tables S8 and S9). Low to moderate genotoxicity was detected in extracted hospital, raw and treated wastewater but in none of the other samples. Other studies observed genotoxicity in extracted WWTP effluents (Macova et al. 2011, Keiter et al. 2006, Escher et al. 2014). These potentials generally decreased upon ozonation (Cao et al. 2009, Misik et al. 2011).

3.4 Identifying the optimal SPE method

Similar to analytical chemistry (Baker and Kasprzyk-Hordern 2011, Maruya et al. 2016, Polo et al. 2005), SPE of (waste)water samples is advantageous for *in vitro* bioassays. Extraction prevents the microbial degradation of untreated samples and improves the detection of toxicological effects caused by low (micro)pollutant concentrations (Escher et al. 2005, Janošek et al. 2007, Macova et al. 2011, Neale et al. 2015, 2018). SPE can also minimise matrix interferences by reducing natural organic matter and excluding ions, nutrients or acids (Neale and Escher 2014, Prasse et al. 2015, Escher et al. 2018).

In contrast to chemical analysis of target compounds, the recovery of toxicity by SPE cannot be evaluated because the causative chemicals and mixture effects remain unknown. Thus, this study aimed at maximising the extraction of toxicity by comparing two mixed-mode hydrophilic/hydrophobic (Oasis HLB and Supelco ENVI-Carb+) and one composite (Telos C18/ENV) SPE sorbents. These SPE sorbents enrich a broad and heterogeneous spectrum of chemicals (Köke et al. 2018, Leusch et al. 2012, Neale et al. 2018). Extracting both neutral and acidified samples, six different SPE methods were evaluated by a semi-quantitative (3.4.1–3.4.4) approach followed by multivariate statistics (3.4.5).

3.4.1 Blanks

In parallel to the extraction of the samples, a SPE blank was prepared to control for potential contaminants in reference waters and used materials (Kolkman et al. 2013, Neale et al. 2018, Schulze et al. 2017). Each cartridge type was loaded with ultrapure water and extracted as described in 2.3.3. The extracts of the 60 SPE blanks were negative in all bioassays except in two cases (3%): Supelco ENVI-Carb+ at pH 7 and pH 2.5 in the YAAS. Here, the activities were 2% and 3% higher than the limit of quantification. In addition, a DMSO sample was included in parallel to the SPE extracts in each *in vitro* bioassay as a solvent control. These solvent controls did not induce an effect in the bioassays.

3.4.2 Cytotoxicity

Cytotoxicity is often used as indicator of the reactive toxicity of environmental samples and their overall (micro)pollutant load. It, thus, represents an important endpoint which is integrated into several water quality assessments (Escher et al. 2014, 2018, Leusch et al. 2014, Välitalo et al. 2017). However, depending on the investigated endpoint, cytotoxicity can also prevent or mask the detection of specific toxicity (see 4).

In the present study, none of the aqueous samples induced cytotoxic effects (Figure 4, Tables S8 and S9). Cytotoxicity was, however, frequently detected in SPE extracts (Figure 4). Untreated wastewater induced cytotoxicity in 50% (HOS) and 38% (INF-1) of sample extracts (n = 60, each) tested in ten *in vitro* bioassays (Table 2). For conventionally treated wastewater (EFF-1, EFF-4, EFF-4-MS, n = 54–60) cytotoxicity was observed in $\leq$ 25% of extracts (Table 2). The occurrence of cytotoxicity in extracted ozonated wastewater (sample EFF-4-MS-O_3, n = 54) and groundwater (sample GW-1, n = 60) was 35 and 2%, respectively (Table 2).

The choice of the SPE method had a substantial influence on the detection of cytotoxicity: the extracts of the Oasis HLB and the Telos C18/ENV (neutral pH) were cytotoxic in 32% and

50% of the bioassays (n = 78 each, Table 2). At acidified pH, these extracts induced similar cytotoxicity with 15 and 13%, respectively (n = 78 each, Table 2). Samples extracted with the Supelco ENVI-Carb+ at neutral pH were more cytotoxic (12%) compared to the corresponding samples that were extracted at acidified pH (not cytotoxic effects, n = 78 each, Table 2).

In general, samples extracted at neutral pH induced higher cytotoxicity than acidified samples (Figure 4) and Telos C18/ENV extracts were more cytotoxic than those of Oasis HLB and Supelco ENVI-Carb+. Thus, extraction at neutral pH with Telos C18/ENV was the method where the highest cytotoxicity was detected (Figure 4). Escher et al. (2005) found an extraction at pH 3 (using the Oasis HLB) to be more effective than pH 7 and pH 11 in a study on spiked urine samples. Stalter et al. (2011) observed this for acidified biologically-treated and ozonated wastewater. Both studies suggest that compounds with acidic moieties to be responsible for the recovered cytotoxicity. This is in contrast to the present results, which suggest that the cytotoxicity in a broad range of bioassays is extracted more effectively at neutral pH.

In a recent study by Stalter et al. (2016) the Telos ENV (without C18 sorbent) followed by the Oasis HLB recovered most cytotoxicity amongst nine other SPE sorbents from disinfected drinking water (acidified before extraction). Polar compounds adsorbed by the ENV as well as the HLB sorbent material were suspected as main causative agents. Although Stalter et al. (2016) did not compare an extraction at neutral pH the results support the effectivity of the Telos C18/ENV and Oasis HLB observed in the present study. Along the same line, a multilayer SPE based on Oasis HLB induced more cytotoxicity than a single sorbent method in a study by Neale et al. (2018).

Conventional wastewater treatment decreased the occurrence of cytotoxicity from 38% of the extracts to 7% in case of WWTP 1 (Table 2). In contrast, ozonation increased the number of cytotoxic extracts from 24 to 35% (Table 2). This observation supports earlier hypotheses on

the formation of toxic transformation products (TPs) during ozonation (Jia et al. 2015, Lundström et al. 2010, Magdeburg et al. 2014). In contrast to the WWTP samples, only 2% of groundwater extracts were cytotoxic. This is in agreement with the high water quality monitored at GW sampling sites 1−3 (Seitz and Winzenbacher 2017) as well as the rare detection of cytotoxicity in groundwater, unless influenced by landfill leachates, industrial or other contaminated sites (Baumstark-Khan et al. 2005, Baun et al. 2000).

3.4.3 Endocrine endpoints

Pooling the results according to water sample type, the highest mean estrogenic activity was found in conventionally treated wastewater (EFF-1, EFF-4, EFF-4-MS) extracted with Telos C18/ENV (pH 2.5) with 5% (n = 4) relative activity and Oasis HLB (pH 2.5) with 5% (n = 4) relative activity (Table S11, Figure S10). Samples extracted at neutral pH with the same SPE sorbents induced lower estrogenic activities (3%, n = 2; 2%, n = 3). Extracts produced with Supelco ENVI-Carb+ showed low estrogenic activity regardless of the adjusted pH.

With regard to the anti-estrogenic activity of conventionally treated (EFF) and ozonated (EFF-O_3) wastewater as well as groundwater (GW) both sorbents, Oasis HLB and Telos C18/ENV showed similar effectivity when samples were extracted at pH 2.5 (Figures 4 and S10, Tables S8 and S11). For conventionally treated wastewater (EFF) and groundwater (GW) extracted at neutral pH with the same sorbents the mean anti-estrogenic activity was higher. The highest mean anti-estrogenic activity was found in samples extracted with Supelco ENVI-Carb+ at neutral pH (62–87%, n = 1–2).

In case of the anti-androgenic activity of all sample types, acidified samples extracted with Oasis HLB and Telos C18/ENV produced similar results again (Figures 4 and S11). Because of high cytotoxicity, the activities of neutrally extracted samples could not be analysed. Treated wastewater and groundwater extracted with Supelco ENVI-Carb+ at both pH values induced lower anti-androgenic activities than the other SPE methods. As the activity in the

other bioassays was minor, no comparison of the SPE methods on these endpoints was possible (Figures S11−S14).

Based on the above results the Telos C18/ENV sorbent followed by the Oasis HLB recovered highest endocrine activities from the majority of (waste)water samples. However, the Supelco ENVI-Carb+ sorbent was more effective in recovering androgenic activities. This is in part reflected in previous studies. In a study on bottled mineral water, a C18 material recovered higher estrogenic activity compared to the Oasis HLB and Supelco ENVI-Carb+ (Wagner and Oehlmann 2011). The authors argue that non-polar chemicals are responsible for this effect. In the present study, most estrogenicity was recovered by the Telos C18/ENV (involving a similar C18 material), while Oasis HLB achieved comparable levels.

Except for estrogenicity, endocrine activities were more effectively recovered at pH 2.5. However, the more frequent detection of cytotoxicity in pH 7 extracts might have masked the respective activities. Despite the effective extraction of endocrine activities, it remained insufficient from some (waste)waters and endpoints (Figures 3 and S9, Table S8). This includes the anti-estrogenicity, which was enriched from several but not all samples. The difficulty in extracting anti-estrogenic activity has been observed and discussed in previous studies (Giebner et al. 2018).

3.4.4 Genotoxicity and mutagenicity

The highest genotoxicity (IR 4.37) was detected in the Telos C18/ENV pH 2.5 extract of untreated hospital wastewater (HOS, Tables S8 and S9, Figure S14). Seven extracts (100%) of the Oasis HLB and Telos C18/ENV sorbents at both pH 7 and 2.5 of the conventionally treated wastewater of the pilot WWTP 4 (EFF-4 and EFF-4-MS) were genotoxic with induction rates between 1.50 and 1.87. The extracts of a WWTP 1 (INF-1 and EFF-1), except one extract produced with Oasis HLB, pH 2.5, and groundwater (GW-1) did not induce

genotoxicity. All extracts produced with Supelco ENVI-Carb+ (pH 7 and pH 2.5) were not active, either.

Genotoxicity was enriched from four out of six sampling sites (Figure S14, Tables S8 and S9) but IRs remained only moderately increased compared to the corresponding aqueous samples (except for hospital wastewater). One reason for this could be that genotoxicity of (waste)water samples is generally detected at higher sample enrichment factors (e.g., 100-fold, Keiter et al. 2006, Schulze et al. 2017, Stalter et al. 2016) or at contamination hotspots (Baumstark-Khan et al. 2005, Baun et al. 2000).

In line with the efficiency of the Telos C18/ENV pH 2.5 method, Magdeburg et al. (2014) extracted genotoxicity and mutagenicity from wastewater (biological and advanced treatment) using the Oasis HLB at pH 2. Although the authors did not compare different SPE methods, their results seem in agreement with the present results. Mutagenicity and cytotoxicity were also higher in biologically-treated and ozonated wastewater extracted at pH 2 (instead of pH 7) using a C18 sorbent (Misik et al. 2011). For the other investigated *in vitro* endpoints, no SPE optimisation study was found in the literature.

3.4.5 What is the best SPE method?

Regarding the results of five types of water samples tested with five *in vitro* bioassays the most effective SPE method for the extraction of endocrine activities was Telos C18/ENV pH 7 (7x), followed by Telos C18/ENV pH 2.5 and Supelco ENVI-Carb+ pH 7 (each 5x), Oasis HLB pH 7 (4x), Oasis HLB pH 2.5 (2x) and Supelco ENVI-Carb+ pH 2.5 (1x, Table 3). To statistically distinguish between optimal (and non-optimal) SPE methods a multivariate optimisation based on Pareto was implemented (Durmaz et al. 2015, Ehrgott 2000). Pareto computed sample type and bioassay specific “Pareto optimal” methods.

The Pareto results are exemplified for conventionally treated wastewater (EFF-4) in five *in vitro* bioassays, whereby Pareto is based on the activity percentiles (Table S12) for ranking

the SPE methods (Table S13). The best extraction methods ("Pareto best") were Telos C18/ENV pH 7 followed by Oasis HLB pH 7 and Telos C18/ENV pH 2.5 (see Table S13 for detailed results). The ranking of these methods was computed as follow: Instead of looking at the "best" extraction results within a certain matrix, the "worst" results were classified as "false negative responders". The Supelco ENVI-Carb+ method at pH 2.5 was three times "Pareto-worst" as it extracted the lowest activity in a maximal number of bioassays. All other methods performed better. When an extract was cytotoxic, the result was marked with the label "cytotoxic" instead of providing a value. The Pareto algorithm is capable of evaluating data sets with a limited number of such results. In case of an excessive degree of cytotoxicity (HOS and INF-1), the corresponding SPE method was, however, not listed in the respective ranking matrix and the level of relevance decreases for this parameter. This means that the ranking for this parameter is not reaching the "worst" class anymore. This evaluation procedure was performed for all data sets referring to the different samples, SPE methods and *in vitro* bioassays to obtain the following overall ranking of "Pareto optimal" SPE methods: Regarding the five sample types, the method Telos C18/ENV at pH 7 was four times "Pareto best", followed by Oasis HLB pH 7 and pH 2.5 (each 2x, Tables 3 and S14). In terms of the five bioassays, the methods Telos C18/ENV at pH 2.5 and Supelco ENVI-Carb+ at pH 7 were two times "Pareto best", respectively (details in Table S14).

Accordingly, the method Telos C18/ENV at pH 7 was "Pareto best" regarding the effectivity in extracting different types of water and wastewater samples with respect to the highest endocrine activities (Table 3). Higher recoveries at neutral pH (over acidic and basic pH) were also observed by Tousova et al. (2017) for several endpoints also investigated in this study. The authors, however, used other sorbents for large volume SPE of surface waters. Summing up the results of the *in vitro* bioassays and Pareto optimisation, the methods Telos C18/ENV pH 7 and Oasis HLB pH 7 were optimal to enrich endocrine activities but also the highest cytotoxicity (Table 2). The corresponding methods at pH 2.5 showed good results as

627 well as lower cytotoxicity (Tables 2 and S14). The final recommendation for most effective
628 recovery of *in vitro* toxicity from diverse (waste)waters is, thus, to use the Telos C18/ENV
629 method at a sample pH of 7.

4 Challenges in optimising sample preparation for bioassays

Despite the advantages of optimising the sample preparation for bioassay analyses (Muschket et al. 2017, Neale et al. 2018, Ternes et al. 2017), a number of important challenges remain.

The first challenge is that the "true" toxicity of a sample (at a given sampling site and time) remains unknown. The reason for this is that for complex environmental samples, the causative compounds, potential mixture effects and confounding factors (e.g., matrix effects) are largely unspecified. Accordingly, each step of sampling and sample preparation and storage may change the chemical composition of a sample and its toxicity. Active compounds may be added (via contaminated materials) or removed (via adsorption to materials) during sampling, added or removed during transport and storage (via microbial activity) and added or removed during sample preparation.

Second, the differentiation between toxicity caused by anthropogenic pollutants and naturally occurring compounds, often referred to a matrix effects, remains challenging. For instance, our approach in maximising the recovery of toxicities may come at the costs of also maximising matrix effects. One such example is the co-extraction of DOC that may induce artefacts in bioassays for receptor antagonism (Neale and Escher 2014). Several confounding factors resulting in false-positive or negative result need to be considered when interpreting bioassay data (discussed in Giebner et al. 2018). However, sample preparation may not be the appropriate tool to address these. Instead, post-extraction analysis (such as effect-directed analysis) can be a way to separate the toxicity caused by anthropogenic and natural compounds.

The third challenge is the selectivity of sample extraction: While SPE methods with broad selectivity exist, an extraction of chemicals is always selective, resulting in a loss of compounds with low affinity to the sorbent (Köke et al. 2018, Neale et al. 2018, Niss et al. 2018, Stalter et al. 2016). Accordingly, the toxicity of an extract will never fully represent the toxicity of the extracted sample. Thus, the question is rather how much loss in toxicity during

extraction is acceptable. One way of addressing this is to compare the toxicity of extracts to aqueous samples (Dagnino et al. 2010, EC 2003). Another way is to optimise the recovery of toxicity. Both strategies were adopted in this study to identify the best extraction method.

The forth challenge arises from cytotoxicity masking the effect under investigation, which is often the case at high concentration factors. While cytotoxicity can be considered an important toxicological endpoint by itself outweighing the specific effect is masks, it is most commonly rather regarded an obstacle that needs to be removed. This can be achieved by diluting a sample to a non-cytotoxic concentration (Inoue et al. 2009, 2011, Leusch et al. 2017, Neale et al. 2018, Välitalo et al. 2017). However, this also dilutes the effect of interest. Alternative approaches, such as minimising the dilution of aqueous samples (Niss et al. 2018) or reducing exposure times in the bioassay as well as cleaning up the cytotoxicity (e.g., by fractionation), have so far not been widely adopted.

These challenges are connected to a range of SPE parameters. Thus, the sorbent (Chang et al. 2009, Escher et al. 2005, Stalter et al. 2016), sample volumes (Macova et al. 2011, Schulze et al. 2017), eluting solvents (Lu et al. 2010, Välitalo et al. 2017, Yang et al. 2014), fractionation steps (Leusch et al. 2017, Välitalo et al. 2017) and operating modes such as large volume or multilayer SPE (Köke et al. 2018, Schulze et al. 2017) can be optimised.

Acknowledging that it is impractical to perform an optimisation for every sample and every bioassay, a range of case studies for different matrices can be used to evaluate whether specific sample preparation methods perform generally better than others. We have taken such approach in the present study and conclude that the Telos C18/ENV method at neutral sample pH performs best in recovering multiple endocrine activities and cytotoxicity from aqueous samples.

5 Conclusions

1. Acidification of aqueous (waste)water samples significantly alters a range of *in vitro* toxicities, including anti-estrogenic, anti-androgenic and retinoic acid-like activities as well as mutagenicity. Sample filtration has a minor impact on the samples' toxicity.
2. Compared to aqueous samples, solid phase extraction enriches most *in vitro* toxicities. However, some activities (e.g., anti-estrogenicity) remain poorly extractable.
3. When comparing six SPE methods, the choice of the optimal method depends on the matrix as well as the *in vitro* endpoint.
4. In general, an extraction using Telos C18/ENV at a sample pH of 7 was most effective in recovering *in vitro* toxicity from (waste)water samples. However, these methods also co-extract a high cytotoxicity masking other endpoints. Using the same method at a sample pH of 2.5 reduced the extraction of cytotoxicity.
5. Sample preparation needs to be optimised when analysing the toxicity of water samples. While this is a resource-consuming task involving multiple methodological parameters, water quality can only be accurately assessed when the recovery of the toxicity of a sample is maximal.

6 Acknowledgements

This study was funded by the German Federal Ministry of Education and Research (BMBF) Project TransRisk "Charakterisierung, Kommunikation und Minimierung von Risiken durch neue Schadstoffe und Krankheitserreger im Wasserkreislauf" (funding number: 02WRS1275B, contract period: 11/2011 to 04/2015) within the funding measure "Risk Management of Emerging Compounds and Pathogens in the Water Cycle (RiSKWa)" which is gratefully appreciated. The authors cordially thank all TransRisk project partners for fruitful discussions and extensive collaboration that helped to compile and improve this manuscript as well as Johannes Völker and Andrea Misovic for their scientific and technical support. Furthermore, the authors would like to thank Prof. Dr. P. Cornel and Gregor Knopp and the employees at the investigated WWTP and water works for supporting this study such as by provision of samples.

References

Allinson, M., Shiraishi, F., Allinson, G., 2011. A comparison of recombinant receptor-reporter gene bioassays and a total estrogen enzyme linked immunosorbent assay for the rapid screening of estrogenic activity in natural and waste waters. Bulletin of Environmental Contamination and Toxicology, 86 (5), 461.

Andersen, H., Siegrist, H., Halling-Sørensen, B., Ternes, T. A., 2003. Fate of estrogens in a municipal sewage treatment plant. Environmental Science & Technology, 37 (18), 4021-4026.

Baker, D. R., Kasprzyk-Hordern, B., 2011. Critical evaluation of methodology commonly used in sample collection, storage and preparation for the analysis of pharmaceuticals and illicit drugs in surface water and wastewater by solid phase extraction and liquid chromatography–mass spectrometry. Journal of Chromatography A, 1218 (44), 8036-8059.

Baumstark-Khan, C., Cioara, K., Rettberg, P., Horneck, G., 2005. Determination of geno- and cytotoxicity of groundwater and sediments using the recombinant SWITCH test. Journal of Environmental Science and Health, 40 (2), 245-263.

Baun, A., Jensen, S. D., Bjerg, P. L., Christensen, T. H., Nyholm, N., 2000. Toxicity of organic chemical pollution in groundwater downgradient of a landfill (Grindsted, Denmark). Environmental Science & Technology, 34 (9), 1647-1652.

Bistan, M., Podgorelec, M., Marinšek Logar, R., Tišler, T., 2012. Yeast estrogen screen assay as a tool for detecting estrogenic activity in water bodies. Food Technology and Biotechnology, 50 (4), 427-433.

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 Directive: Recommendations for more efficient assessment and management of chemical contamination in European surface water resources. Science of The Total Environment (576), 720-737.

Cao, N., Yang, M., Zhang, Y., Hu, J., Ike, M., Hirotsuji, J., Matsui, H., Inoue, D., Sei, K., 2009. Evaluation of wastewater reclamation technologies based on *in vitro* and *in vivo* bioassays. Science of the Total Environment, 407 (5), 1588-1597.

Chang, H. S., Choo, K. H., Lee, B., Choi, S. J., 2009. The methods of identification, analysis, and removal of endocrine disrupting compounds (EDCs) in water. Journal of Hazardous Materials 172 (1), 1-12.

Dagnino, S., Gomez, E., Picot, B., Cavaillès, V., Casellas, C., Balaguer, P., Fenet, H., 2010. Estrogenic and AhR activities in dissolved phase and suspended solids from wastewater treatment plants. Science of the Total Environment, 408 (12), 2608-2615.

Durmaz, V., Weber, M., Meyer, J., Mückter, H., 2015. Computergestützte Simulationen zur Abschätzung gesundheitlicher Risiken durch anthropogene Spurenstoffe in der Wassermatrix. KA Korrespondenz Abwasser, Abfall, 3/15, 264-267.

Ehrgott, M., 2000. Approximation algorithms for combinatorial multicriteria optimization problems. International Transactions in Operational Research, 7 (1), 5-31.

Escher, B. I., Aït-Aïssa, S., Behnisch, P. A., Brack, W., Brion, F., Brouwer, A., Buchinger, S., Crawford, S. E., Du Pasquier, D., Hamers, T., Hettwer, K., Hilscherová, K., Hollert, H., Kase, R., Kienle, C., Tindall, A. J., Tuerk, J., van der Oost, R., Vermeirssen, E., Neale P. A., 2018. Effect-based trigger values for *in vitro* and *in vivo* bioassays performed on surface water extracts supporting the environmental quality standards (EQS) of the European Water Framework Directive. Science of the Total Environment, 628, 748-765.

Escher, B. I., Bramaz, N., Maurer, M., Richter, M., Sutter, D., von Känel, C., Zschokke, M., 2005. Screening test battery for pharmaceuticals in urine and wastewater. Environmental Toxicology and Chemistry, 24 (3), 750-758.

Escher, B. I., Allinson, M., Altenburger, R., Bain, P. A., Balaguer, P., Busch, W., Crago, J., Denslow, N. D., Dopp, E., Hilscherova, K., Humpage, A. R., Kumar, A., Grimaldi, M., Jayasinghe, B. S., Jarosova, B., Jia, A., Makarov, S., Maruya, K. A., Medvedev, A., Mehinto, A. C., Mendez, J. E., Poulsen, A., Prochazka, E., Richard, J., Schifferli, A., Schlenk, D., Scholz, S., Shiraishi, F., Snyder, S, Su, G., Tang, J. Y., van der Burg, B., van der Linden, S. C., Werner, I., Westerheide, S. D., Wong, C. K., Yang, M., Yeung, B. H., Zhang, X., Leusch, F. D., 2014. Benchmarking organic micropollutants in wastewater, recycled water and drinking water with *in vitro* bioassays. Environmental Science & Technology, 48 (3), 1940-1956.

Gehrmann, L., Bielak, H., Behr, M., Itzel, F., Lyko, S., Simon, A., Kunze, G., Dopp, E., Wagner, M., Tuerk, J., 2018. (Anti-) estrogenic and (anti-) androgenic effects in wastewater during advanced treatment: comparison of three *in vitro* bioassays. Environmental Science and Pollution Research, 25 (5), 4094-4104.

Giebner, S., Ostermann, S., Straskraba, S., Oetken, M., Oehlmann, J., Wagner, M., 2018. Effectivity of advanced wastewater treatment: reduction of *in vitro* endocrine activity and mutagenicity but not of *in vivo* reproductive toxicity. Environmental Science and Pollution Research, 25 (5), 3965-3976.

Hendriks, A. J., Maas-Diepeveen, J. L., Noordsij, A., Van der Gaag, M. A., 1994. Monitoring response of XAD-concentrated water in the Rhine delta: a major part of the toxic compounds remains unidentified. Water Research, 28 (3), 581-598.

Ihara, M., Ihara, M. O., Kumar, V., Narumiya, M., Hanamoto, S., Nakada, N., Yamashita N., Miyagawa S., Iguchi T., Tanaka, H., 2014. Co-occurrence of estrogenic and antiestrogenic activities in wastewater: quantitative evaluation of balance by *in vitro* ERα reporter gene assay and chemical analysis. Environmental Science & Technology, 48 (11), 6366-6373.

Inoue, D., Nakama, K., Matsui, H., Sei, K., Ike, M., 2009. Detection of agonistic activities against five human nuclear receptors in river environments of Japan using a yeast two-hybrid assay. Bulletin of Environmental Contamination and Toxicology, 82, 399-404.

Inoue, D., Nakama, K., Sawada, K., Watanabe, T., Matsui, H., Sei, K., Nakanishi T., Ike, M., 2011. Screening of agonistic activities against four nuclear receptors in wastewater treatment plants in Japan using a yeast two-hybrid assay. Journal of Environmental Sciences, 23 (1), 125-132.

International Standard Organisation (ISO)/Deutsches Institut für Normung (DIN), 2012. ISO/DIN 11350: Water quality - Determination of the genotoxicity of water and waste water - Salmonella/microsome fluctuation test (Ames fluctuation test). Geneva, Switzerland.

International Standard Organisation (ISO), 2002. ISO13829: Water quality-determination of the genotoxicity of water and waste water using the umu test. Geneva, Switzerland.

International Standard Organisation (ISO), 2018. ISO 19040-1: Water quality -- Determination of the estrogenic potential of water and waste water -- Part 1: Yeast estrogen screen (*Saccharomyces cerevisiae*). Geneva, Switzerland.

Jalova, V., Jarošová, B., Blaha, L., Giesy, J. P., Ocelka, T., Grabic, R., Jurčíková J., Vrana B., Hilscherova, K., 2013. Estrogen-, androgen-and aryl hydrocarbon receptor mediated activities in passive and composite samples from municipal waste and surface waters. Environment International, 59, 372-383.

Janex-Habibi, M. L., Huyard, A., Esperanza, M., Bruchet, A., 2009. Reduction of endocrine disruptor emissions in the environment: The benefit of wastewater treatment. Water Research, 43 (6), 1565-1576.

Janošek, J., Bittner, M., Hilscherová, K., Bláha, L., Giesy, J. P., Holoubek, I., 2007. AhR-mediated and antiestrogenic activity of humic substances. Chemosphere, 67 (6), 1096-1101.

Jia, A., Escher, B. I., Leusch, F. D., Tang, J. Y., Prochazka, E., Dong, B., Snyder, E. M., Snyder, S. A., 2015. *In vitro* bioassays to evaluate complex chemical mixtures in recycled water. Water Research, 80, 1-11.

Keiter, S., Rastall, A., Kosmehl, T., Erdinger, L., Braunbeck, T., Hollert, H., 2006. Ecotoxicological assessment of sediment, suspended matter and water samples in the upper Danube river. A pilot study in search for the causes for the decline of fish catches. Environmental Science and Pollution Research, 13 (5), 308-319.

Knopp, G., Prasse, C., Ternes, T. A., Cornel, P., 2016. Elimination of micropollutants and transformation products from a wastewater treatment plant effluent through pilot scale ozonation followed by various activated carbon and biological filters. Water Research, 100, 580-592.

Köke, N., Zahn, D., Knepper, T. P., Frömel, T., 2018. Multi-layer solid-phase extraction and evaporation − enrichment methods for polar organic chemicals from aqueous matrices. Analytical and Bioanalytical Chemistry, 410 (9), 2403-2411.

Koh, Y. K., Chiu, T. Y., Boobis, A., Cartmell, E., Scrimshaw, M. D., Lester, J. N., 2008. Treatment and removal strategies for estrogens from wastewater. Environmental Technology 29 (3), 245–267.

Kolkman, A., Schriks, M., Brand, W., Bäuerlein, P. S., van der Kooi, M. M., van Doorn, R. H., Emke, E., Reus, A. A., van der Linden, S. C., de Voogt, P., Heringa, M. B., 2013. Sample

preparation for combined chemical analysis and *in vitro* bioassay application in water quality assessment. Environmental Toxicology and Pharmacology, 36 (3), 1291-1303.

Kümmerer, K., 2011. Commentary: Emerging contaminants versus micro-pollutants. Clean – Soil, Air, Water, 39 (10), 889–890.

Kusk, K. O., Krüger, T., Long, M., Taxvig, C., Lykkesfeldt, A. E., Frederiksen, H., Emke E., Reus A. A., van der Linden S. C., de Voogt P., Bonefeld-Jørgensen, E. C., 2011. Endocrine potency of wastewater: contents of endocrine disrupting chemicals and effects measured by *in vivo* and *in vitro* assays. Environmental Toxicology and Chemistry, 30 (2), 413-426.

Leusch, F. D. L., Prochazka, E., Tan, B. L. L., Carswell, S., Neale, P., Escher, B. I., 2012. Optimising micropollutants extraction for analysis of water samples: comparison of different solid phase materials and liquid-liquid extraction. Sci. Forum Stakehold. Engagem. Build. Link. Collab. Sci. Qual., Brisbane, Queensland (pp. 191-195).

Leusch, F. D., 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 Research, 50, 420-431.

Leusch, F. D., Neale, P. A., Hebert, A., Scheurer, M., Schriks, M. C., 2017. Analysis of the sensitivity of *in vitro* bioassays for androgenic, progestagenic, glucocorticoid, thyroid and estrogenic activity: Suitability for drinking and environmental waters. Environment International, 99, 120-130.

Loos, R., Carvalho R., António D. C., Comero S., Locoro G., Tavazzi S., Paracchini B., Ghiani M., Lettieri T., Blaha L., Jarosova B., Voorspoels S., Servaes K., Haglund P., Fick J., Lindberg R. H., Schwesig D., Gawlik B. M., 2012. EU wide monitoring survey on waste water treatment plant effluents. JRC Scientific and Policy Report, JRC 76400.

Lu, G., Zhang, H., Wang, C., 2010. Assessment of estrogenic activity conducted by combining bioassay and chemical analyses of the effluent from wastewater treatment plants in Nanjing, China. Environmental Toxicology and Chemistry, 29 (6), 1279-1286.

Lundström, E., Adolfsson-Erici, M., Alsberg, T., Björlenius, B., Eklund, B., Lavén, M., Breitholtz, M., 2010. Characterization of additional sewage treatment technologies: Ecotoxicological effects and levels of selected pharmaceuticals, hormones and endocrine disruptors. Ecotoxicology and Environmental Safety, 73 (7), 1612-1619.

Ma, M., Li, J., Wang, Z., 2005. Assessing the detoxication efficiencies of wastewater treatment processes using a battery of bioassays/biomarkers. Archives of Environmental Contamination and Toxicology, 49 (4), 480-487.

Macova, M., Toze, S., Hodgers, L., Mueller, J. F., Bartkow, M., Escher, B. I., 2011. Bioanalytical tools for the evaluation of organic micropollutants during sewage treatment, water recycling and drinking water generation. Water Research, 45 (14), 4238-4247.

Magdeburg, A., Stalter, D., Schlüsener, M., Ternes, T., Oehlmann, J., 2014. Evaluating the efficiency of advanced wastewater treatment: target analysis of organic contaminants and (geno-) toxicity assessment tell a different story. Water research, 50, 35-47.

Malaj, E., von der Ohe, P. C., Grote, M., Kühne, R., Mondy, C. P., Usseglio-Polatera, P., Brack, W., Schäfer, R. B., 2014. Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. Proceedings of the National Academy of Sciences of the United States of America, 111 (26), 9549-9554.

Maletz, S., Floehr, T., Beier, S., Klümper, C., Brouwer, A., Behnisch, P., Higley E., Giesy J. P., Hecker M., Gebhardt W., Linnemann V., Pinnekamp J., Hollert H., 2013. *In vitro* characterization of the effectiveness of enhanced sewage treatment processes to eliminate endocrine activity of hospital effluents. Water Research, 47 (4), 1545-1557.

Maruya, K. A., Dodder, N. G., Mehinto, A. C., Denslow, N. D., Schlenk, D., Snyder, S. A., Weisberg, S. B., 2016. A tiered, integrated biological and chemical monitoring framework for contaminants of emerging concern in aquatic ecosystems. Integrated Environmental Assessment and Management, 12 (3), 540-547.

Metcalfe, C. D., Kleywegt, S., Letcher, R. J., Topp, E., Wagh, P., Trudeau, V. L., Moon, T. W., 2013. A multi-assay screening approach for assessment of endocrine-active contaminants in wastewater effluent samples. Science of the Total Environment, 454, 132-140.

Miklos, D. B., Remy, C., Jekel, M., Linden, K., G., Drewes, J. E., Hübner U., 2018. Evaluation of advanced oxidation processes for water and wastewater treatment – A critical review. Water Research (139), 118-131.

Miller, C. A., 1997. Expression of the human aryl hydrocarbon receptor complex in yeast – activation of transcription by indole compounds. Journal of Biological Chemistry 272 (52), 32824-32829.

Misik, M., Knasmueller, S., Ferk, F., Cichna-Markl, M., Grummt, T., Schaar, H., Kreuzinger, N., 2011. Impact of ozonation on the genotoxic activity of tertiary treated municipal wastewater. Water Research, 45 (12), 3681-3691.

Muschket, M., Di Paolo, C., Tindall, A. J., Touak, G., Phan, A., Krauss, M., Kirchner K., Seiler T. B., Hollert H., Brack, W., 2018. Identification of unknown antiandrogenic compounds in surface waters by effect-directed analysis (EDA) using a parallel fractionation approach. Environmental Science & Technology, 52 (1), 288-297.

Neale, P. A., Escher, B. I., 2014. Does co-extracted dissolved organic carbon cause artefacts in cell-based bioassays? Chemosphere, 108, 281-288.

Neale, P. A., Escher, B. I., Leusch, F. D., 2015. Understanding the implications of dissolved organic carbon when assessing antagonism *in vitro*: an example with an estrogen receptor assay. Chemosphere, 135, 341-346.

Neale, P. A., Brack, W., Aït-Aïssa, S., Busch, W., Hollender, J., Krauss, M., Maillot-Maréchal, E., Munz, N. A., Schlichting, R., Schulze, T., Vogler, B., Escher, B. I, 2018. Solid-phase extraction as sample preparation of water samples for cell-based and other *in vitro* bioassays. Environmental Science: Processes & Impacts, 20 (3), 493-504.

Niss, F., Rosenmai, A. K., Mandava, G., Örn, S., Oskarsson, A., Lundqvist, J., 2018. Toxicity bioassays with concentrated cell culture media—a methodology to overcome the chemical

loss by conventional preparation of water samples. Environmental Science and Pollution Research, 25(12), 12183-12188.

Ng, C. K., Cao, B., 2015. What exactly are you filtering out? Environmental Science and Technology, 49, 5259−5260.

Polo, M., Llompart, M., Garcia-Jares, C., Cela, R., 2005. Multivariate optimization of a solid-phase microextraction method for the analysis of phthalate esters in environmental waters. Journal of Chromatography A, 1072 (1), 63-72.

Prasse, C., Stalter, D., Schulte-Oehlmann, U., Oehlmann, J., Ternes, T., 2015. Spoilt for choice: A critical review on chemical and biological evaluation of current wastewater treatment technologies. Water Research, 87, 237−270.

Rao, K., Li, N., Ma, M., Wang, Z., 2014. *In vitro* agonistic and antagonistic endocrine disrupting effects of organic extracts from waste water of different treatment processes. Frontiers of Environmental Science & Engineering, 8 (1), 69-78.

Reungoat, J., Macova, M., Escher, B. I., Carswell, S., Mueller, J. F., Keller, J., 2010. Removal of micropollutants and reduction of biological activity in a full scale reclamation plant using ozonation and activated carbon filtration. Water Research, 44 (2), 625-637.

Rizzo, L., 2011. Bioassays as a tool for evaluating advanced oxidation processes in water and wastewater treatment. Water Research, 45 (15), 4311-4340.

Routledge, E. J., 2003. Identifying the causative agents: the use of combined chemical and biological strategies in monitoring programs. Pure and Applied Chemistry, 75 (11-12), 2461-2466.

Routledge, E. J., Sumpter, J. P., 1996. Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. Environmental Toxicology and Chemistry, 15, 241-248.

Sawada, K., Inoue, D., Wada, Y., Sei, K., Nakanishi, T., Ike, M., 2012. Detection of retinoic acid receptor agonistic activity and identification of causative compounds in municipal wastewater treatment plants in Japan. Environmental Toxicology and Chemistry, 31 (2), 307-315.

Schulze, T., Ahel, M., Ahlheim, J., Aït-Aïssa, S., Brion, F., Di Paolo, C., Froment, J., Hidasi, A. O., Hollender, J., Hollert, H., Hu, M., Kloß, A., Koprivica, S., Krauss, M., Muz, M., Oswald, P., Petre, M., Schollée, J. E., Seiler, T. B., Shao, Y., Slobodnik, J., Sonavane, M., Suter, M. J., Tollefsen, K. E., Tousova, Z., Walz, K. H., Brack, W., 2017. Assessment of a novel device for onsite integrative large-volume solid phase extraction of water samples to enable a comprehensive chemical and effect-based analysis. Science of the Total Environment, 581, 350-358.

Schwarzenbach, R. P., Escher, B. I., Fenner, K., Hofstetter, T. B., Johnson, C. A., von Gunten, U., Wehrli, B., 2006. The challenge of micropollutants in aquatic systems. Science 313(5790), 1072-1077

Seitz, W., Winzenbacher, R., 2017. A survey on trace organic chemicals in a German water protection area and the proposal of relevant indicators for anthropogenic influences. Environmental Monitoring and Assessment, 189 (6), 244.

Shieh, B. H., Louie, A., Law, F. C., 2016. Factors affecting distribution of estrogenicity in the influents, effluents, and biosolids of Canadian wastewater treatment plants. Archives of Environmental Contamination and Toxicology, 70 (4), 682-691.

Sohoni P., Sumpter J. P., 1998. Several environmental oestrogens are also antiandrogens. Journal of Endocrinology, 158, 327-339.

Stalter, D., Peters, L. I., O'Malley, E., Tang, J. Y., Revalor, M., Farré, M. J., Watson, K., von Gunten, U., Escher, B. I., 2016. Sample enrichment for bioanalytical assessment of disinfected drinking water: Concentrating the polar, the volatiles, and the unknowns. Environmental Science and Technology, 50 (12), 6495-6505.

Stalter, D., Magdeburg, A., Wagner, M., Oehlmann, J., 2011. Ozonation and activated carbon treatment of sewage effluents: Removal of endocrine activity and cytotoxicity. Water Research, 45 (3), 1015-1024.

Ternes, T. A., Prasse, C., Eversloh, C. L., Knopp, G., Cornel, P., Schulte-Oehlmann, U., Schwartz, T., Alexander, J., Seitz, W., Coors, A., Oehlmann, J., 2017. Integrated evaluation concept to assess the efficacy of advanced wastewater treatment processes for the elimination of micropollutants and pathogens. Environmental Science & Technology, 51 (1), 308-319.

Tousova, Z., Oswald, P., Slobodnik, J., Blaha, L., Muz, M., Hu, M., Brack, W., Krauss, M., Di Paolo, C., Tarcai, Z., Seiler, T. B., Hollert, H., Koprivica, S., Ahel, M., Schollée, J. E., Hollender, J., Suter, M. J., Hidasi, A. O., Schirmer, K., Sonavane, M., Ait-Aissa, S., Creusot, N., Brion, F., Froment, J., Almeida, A. C., Thomas, K., Tollefsen. K. E., Tufi, S., Ouyang, X., Leonards, P., Lamoree, M., Torrens, V. O., Kolkman, A., Schriks, M., Spirhanzlova, P., Tindall, A., Schulze, T., 2017. European demonstration program on the effect-based and chemical identification and monitoring of organic pollutants in European surface waters. Science of the Total Environment, 601, 1849-1868.

US EPA, 2010. Stability of Pharmaceuticals, Personal Care Products, Steroids, and Hormones in Aqueous Samples, POTW Effluents, and Biosolids. U.S. Environmental Protection Agency, Office of Water. p. 1-38. EPA-820-R-10-008.

US EPA, 2002. Short-term method for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. In assessments of effluents. Fourth Edition, October. Office of Water, U.S. Environmental Protection Agency Washington, DC.

Välitalo, P., Massei, R., Heiskanen, I., Behnisch, P., Brack, W., Tindall, A. J., Du Pasquier, D., Küster, E., Mikola, A., Schulze, T., Sillanpää, M., 2017. Effect-based assessment of toxicity removal during wastewater treatment. Water Research, 126, 153-163.

Vanderford, B. J., Mawhinney, D. B., Trenholm, R. A., Zeigler-Holady, J. C., Snyder, S. A., 2011. Assessment of sample preservation techniques for pharmaceuticals, personal care products, and steroids in surface and drinking water. Analytical and Bioanalytical Chemistry, 399 (6), 2227-2234.

Völker, J., Castronovo, S., Wick, A., Ternes, T. A., Joss, A., Oehlmann, J., Wagner, M., 2016. Advancing biological wastewater treatment: extended anaerobic conditions enhance the removal of endocrine and dioxin-like activities. Environmental Science and Technology, 50, 10606-10615.

Wagner, M., Vermeirssen, E. L. M., Buchinger, S., Behr, M., Magdeburg, A., Oehlmann, J., 2013. Deriving bio-equivalents from *in vitro* bioassays: assessment of existing uncertainties and strategies to improve accuracy and reporting. Environmental Toxicology and Chemistry, 32 (8), 1-12.

Wagner, M., Oehlmann, J., 2011. Endocrine disruptors in bottled mineral water: estrogenic activity in the E-Screen. Journal of Steroid Biochemistry and Molecular Biology, 127 (1), 128-135.

Wagner, M., Oehlmann, J, 2009. Endocrine disruptors in bottled mineral water: total estrogenic burden and migration from plastic bottles. Environmental Science and Pollution Research, 16, 278-286.

Wangmo, C., Jarque, S., Hilscherová, K., Bláha, L., Bittner, M., 2018. *In vitro* assessment of sex steroids and related compounds in water and sediments – a critical review. Environmental Science: Processes & Impacts, 20 (2), 270-287.

Wu, Q., Lam, J. C., Kwok, K. Y., Tsui, M. M., Lam, P. K., 2017. Occurrence and fate of endogenous steroid hormones, alkylphenol ethoxylates, bisphenol A and phthalates in municipal sewage treatment systems. Journal of Environmental Sciences, 61, 49-58.

Yang, X. L., Xia, M. Q., Chen, M., Shen, D. Q., Fu, D. F., Song, H. L., 2014. Optimization of solid-phase extraction for pretreatment of selected estrogens in sewage by response surface methodology. Polish Journal of Environmental Studies, 23 (6), 2287-2294.

Tables

Table 1: Overview of the investigated samples; WWTP: wastewater treatment plant. Details on samples 1–14 can be found in Seitz and Winzenbacher (2017).

Sample No.	Type of sample	Sample acronym	Sampling mode
1	untreated wastewater (hospital effluent)	HOS	grab
2	untreated wastewater (WWTP 1 influent)	INF-1	composite
3	conventionally treated wastewater (WWTP 1 effluent)	EFF-1	composite
4	conventionally treated wastewater (WWTP 2 effluent)	EFF-2	composite
5	conventionally treated wastewater (WWTP 3 effluent)	EFF-3	composite
6	conventionally treated wastewater (WTTP 4 influent of a filtration basin)	FB-IN	grab
7	conventionally treated wastewater (WTTP 4 effluent of a filtration basin)	FB-OUT	composite
8	surface water of an infiltration basin	IB (SW)	grab
9	surface water 1 (river)	SW-1	grab
10	surface water 2 (river)	SW-2	grab
11	surface water 3 (river)	SW-3	grab
12	groundwater 1 (hotspot)	GW-1	grab
13	groundwater 2 (hotspot)	GW-2	grab
14	groundwater 3 (hotspot)	GW-3	grab
15	conventionally treated wastewater (pilot WWTP)	EFF-4	composite
16	ozonated conventionally treated wastewater (before microsieve, pilot WWTP)	EFF-4-O_3	composite
17	conventionally treated wastewater (after microsieve, pilot WWTP)	EFF-4-MS	composite
18	ozonated microfiltered conventionally-treated wastewater (pilot WWTP)	EFF-4-MS-O_3	composite
19	tap water (pilot WWTP)	TAP	grab

1013 **Table 2:** Occurrence of cytotoxicity (%) during the analysis of all sample extracts in ten *in*
1014 *vitro* bioassays (except EFF-4-MS (F) and EFF-4-MS-O_3 (F): n = 9) pooled according to SPE
1015 method. Corresponding samples were taken on the same sampling dates in July (D) 2012 and
1016 in January (F) 2013.

sample	Oasis HLB		Telos C18/ENV		Supelco ENVI-Carb+		Sample
	pH 7	pH 2.5	pH 7	pH 2.5	pH 7	pH 2.5	mean
HOS	80	70	100	50	0	0	50 (n = 60)
INF-1	60	50	70	50	0	0	38 (n = 60)
EFF-1	0	0	30	0	10	0	7 (n = 60)
EFF-4	0	0	0	0	0	0	0 (n = 60)
EFF-4-MS (D)	0	0	50	0	0	0	8 (n = 60)
EFF-4-MS (F)	44	0	56	0	44	0	24 (n = 54)
EFF-4-MS-O_3 (F)	78	0	100	0	33	0	35 (n = 54)
GW-1	0	0	0	0	10	0	2 (n = 60)
Method mean	32 (n = 78)	15 (n = 78)	50 (n = 78)	13 (n = 78)	12 (n = 78)	0 (n = 78)	

1017

Table 3: Most effective SPE methods for the extraction of estrogenic (YES), anti-estrogenic (YAES), androgenic (YAS), anti-androgenic (YAAS) and dioxin-like (YDS) activity from water and wastewater samples (inner table, based on Table S8). In addition, "Pareto best" methods for each bioassay and sample type were computed. Double/triple listings represent equally effective methods. Hospital wastewater (HOS) and one WWTP influent (INF-1) were not analysed due to excessive cytotoxicity. Brackets: activity ≤ 10%; "-": no endocrine activity/cytotoxicity

Bioassay / Sample type	YES	YAES	YAS	YAAS	YDS	Pareto best: sample type
EFF-1	(Oasis 2.5)	Supelco 7	(Oasis 7)	Oasis 2.5	Telos 7	Oasis 2.5 Telos 7
EFF-4	(Telos 2.5)	Telos 7	(Oasis 7)	Telos 7	Telos 7	Oasis 7 Telos 7 Telos 2.5
EFF-4-MS	(Telos 2.5)	Oasis 7	(Supelco 7)	Oasis 7	Telos 7	Telos 7
EFF-4-MS-O_3	-	Supelco 7	(Supelco 2.5)	Telos 2.5	(Telos 2.5)	Supelco 7
GW-1	(Telos 7)	Telos 7	(Supelco 7)	Telos 2.5	(Supelco 7)	Oasis 7 Oasis 2.5 Telos 7
Pareto best: bioassay	Telos 2.5	Supelco 7	Supelco 7	Telos 2.5 Supelco 2.5	Telos 7	**Telos 7**

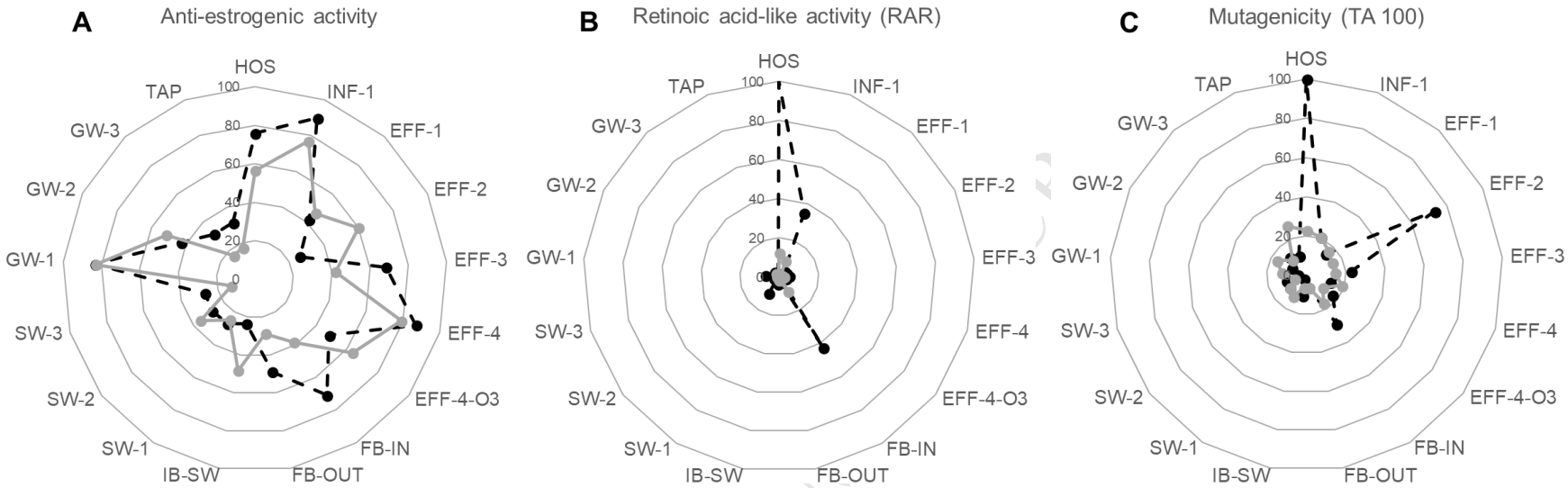


Figure 1: Impact of acidification. Anti-estrogenic activity (A), retinoic acid-like activity (RAR, B) and mutagenicity (Ames TA 100, C) of neutral (black) and acidified (grey) aqueous water and wastewater samples (mean in %). Corresponding samples (INF-1/EFF-1, EFF-4/EFF-4-O_3 and FB-IN/FB-OUT) were taken on the same sampling date in March 2012 and April 2012, respectively.

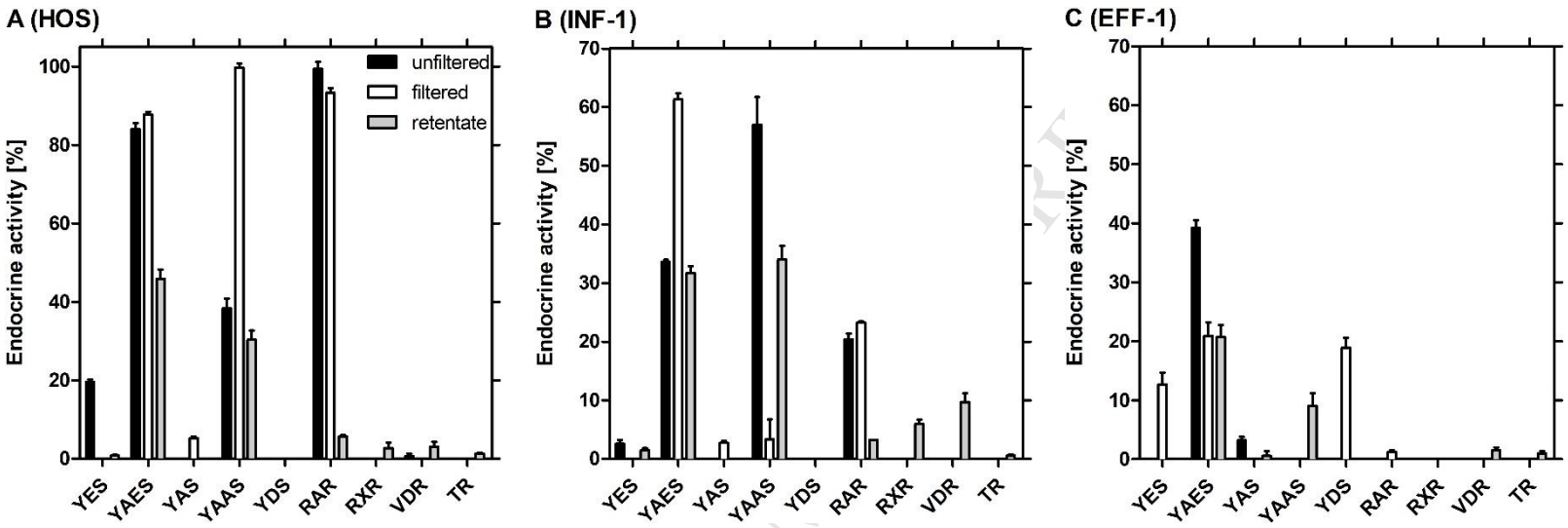


Figure 2: Impact of filtration. Endocrine activity (%, mean ± SEM) of unfiltered (black bars) and filtered (white bars) wastewater samples and the aqueous suspensions of the filter retentate (grey bars). A: untreated hospital wastewater (HOS), B: untreated municipal wastewater of WWTP 1 (INF-1), C: conventionally treated effluent of WWTP 1 (EFF-1). YES: estrogenic, YAES: anti-estrogenic, YAS: androgenic, YAAS: anti-androgenic, YDS: dioxin-like, RAR: retinoic acid-like, RXR: retinoid-X-like, VDR: vitamin D-like, TR: thyronine-like. Corresponding samples (INF-1/EFF-1) taken on the same sampling date in July 2012.

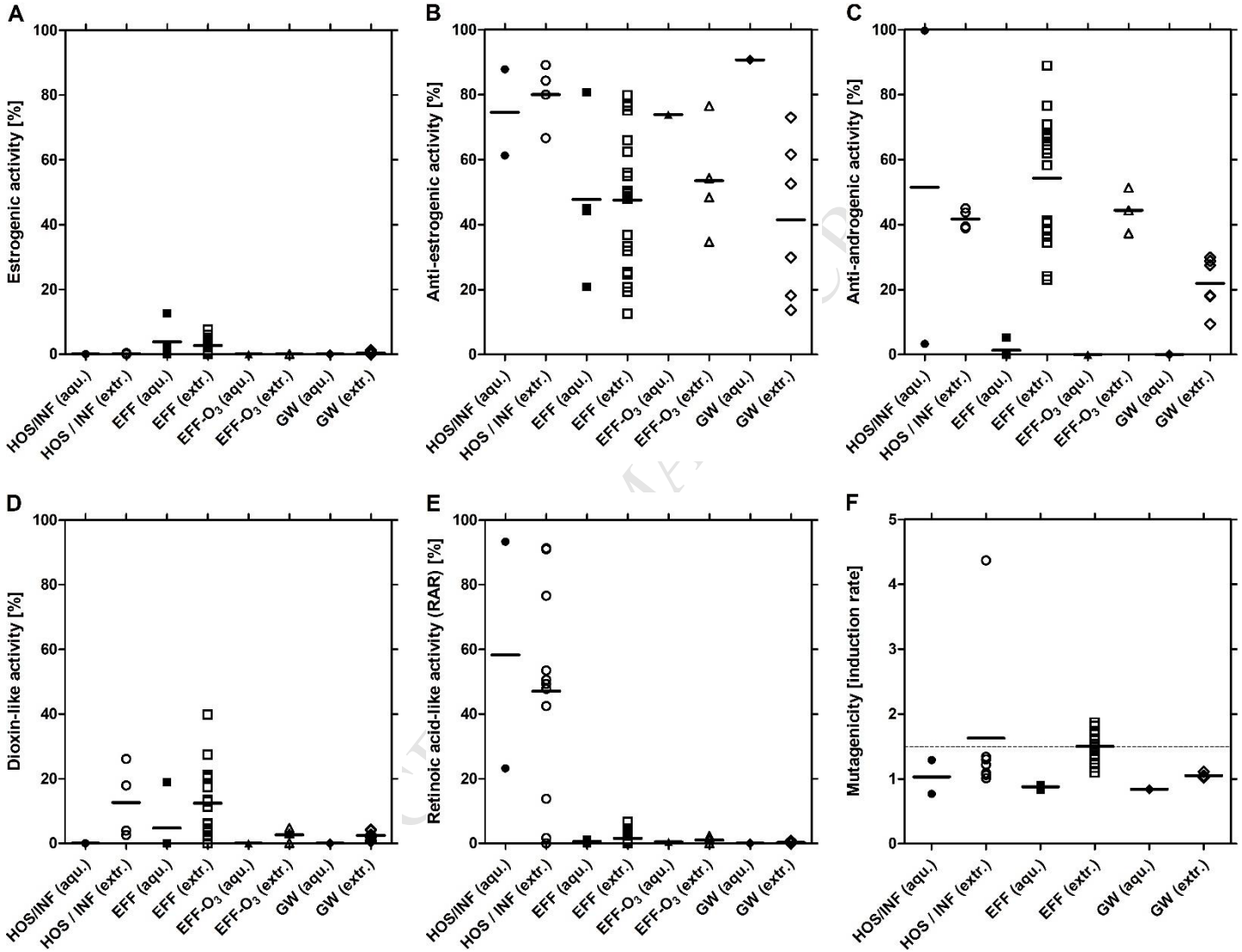

A
Estrogenic activity [%]
B
Anti-estrogenic activity [%]
C
Anti-androgenic activity [%]
D
Dioxin-like activity [%]
E
Retinoic acid-like activity (RAR) [%]
F
Mutagenicity [induction rate]
HOS/INF (aqu.)
HOS / INF (extr.)
EFF (aqu.)
EFF (extr.)
EFF-O3 (aqu.)
EFF-O3 (extr.)
GW (aqu.)
GW (extr.)

Figure 3: Comparison of aqueous and extracted samples. Estrogenic (A), anti-estrogenic (B), anti-androgenic (C), dioxin-like (D) and retinoic acid-like (RAR, E) activity in % and genotoxicity as induction rate (umu, F) of the pooled data of aqueous (aqu.) water and wastewater samples (0.63-fold final concentration) and of the corresponding 10.4-fold concentrated SPE extracts (extr.). Symbols: mean activity of the individual sample, line: mean of all samples of one sample type, filled symbol: aqueous sample, clear symbol: SPE extract, HOS: untreated hospital wastewater, INF: untreated influent, EFF: conventionally treated effluent, EFF-O_3: ozonated conventionally treated wastewater, GW: groundwater. Corresponding samples were taken within the same sampling period in July 2012 and January 2013.

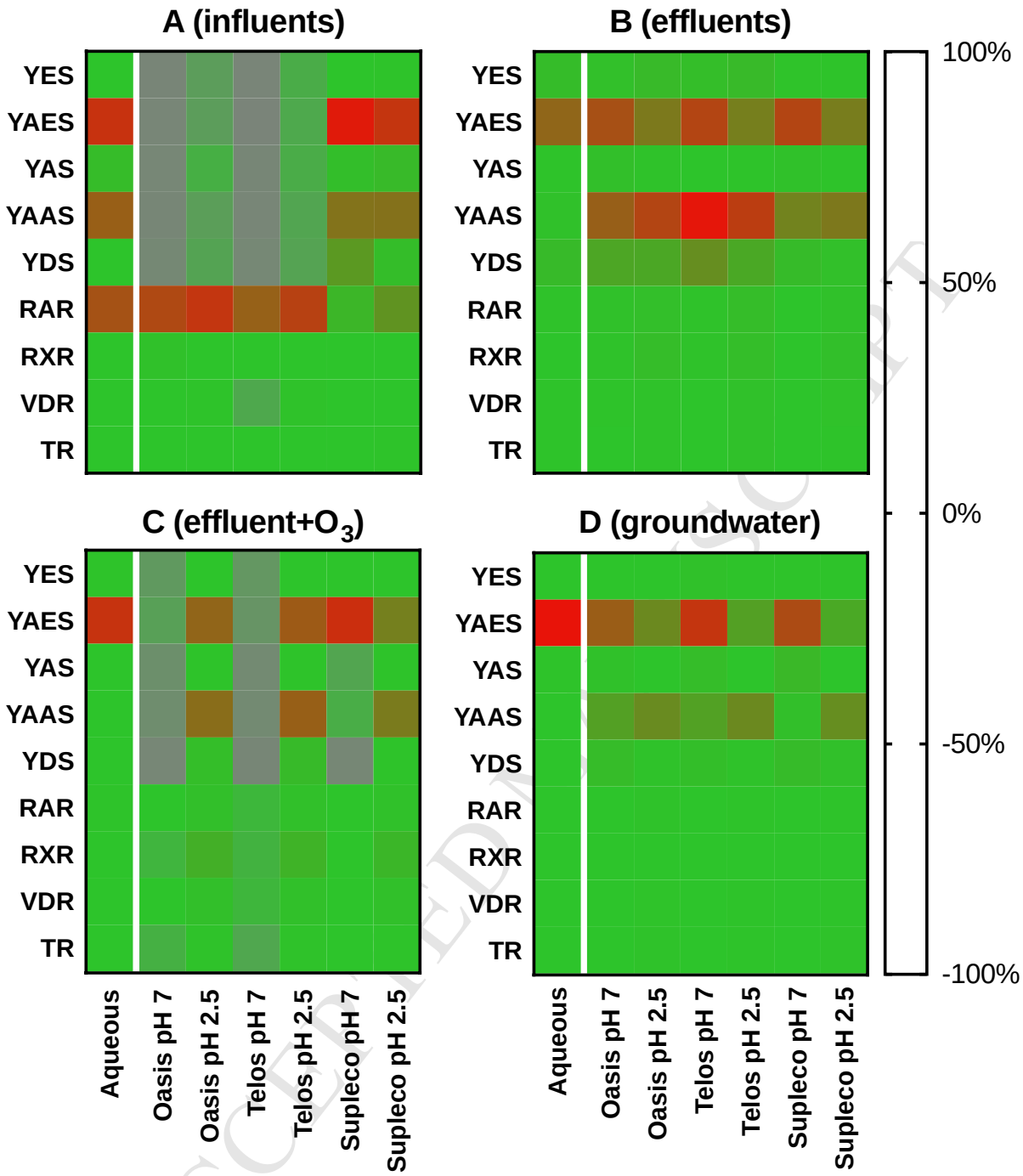


Figure 4: Comparison of the six SPE methods. Endocrine activity (0% to 100%) and cytotoxicity (0% to -100%) of aqueous samples and the corresponding SPE extracts (0.63 and 10.4-fold final concentration, respectively) of wastewater treatment plant influents (A), effluents (B), ozonated effluent (C) and groundwater (D). Six SPE methods were compared: Oasis HLB, Telos C18/ENV and Supelco ENVI-Carb+ extraction at pH 7 and pH 2.5. The results were pooled from the different samples according to water type. Green: 0.0% endocrine activity/cytotoxicity, red: 100% endocrine activity, grey: 100% cytotoxicity.

What you extract is what you see: Optimising the preparation of water and wastewater samples for *in vitro* bioassays

Aennes Abbas & Ilona Schneider, Anna Bollmann, Jan Funke, Jörg Oehlmann, Carsten Prasse, Ulrike Schulte-Oehlmann, Wolfram Seitz, Thomas Ternes, Marcus Weber, Henning Wesely, Martin Wagner

HIGHLIGHTS

- Acidification of (waste)water samples significantly affects their *in vitro* toxicity
- Filtration does not affect the toxicity of most (waste)water samples
- All six SPE methods recovered *in vitro* toxicity, depending on endpoints/matrices
- Best SPE methods were identified for each matrix and endpoint
- Multivariate optimisation identified Telos C18/ENV (pH7) as overall best SPE method