Water Supply

IWA PUBLISHING

Water Supply Vol 24 No 11, 3776 doi: 10.2166/ws.2024.243

Evaluation of pilot-scale drinking water treatment trains using a panel of bioassays

Renske Hoondert[a,*], Milou M. L. Dingemans[a,b], Tineke Slootweg[c], Cheryl Bertelkamp[a,d], Franca Kramer[e], Wim Oorthuizen[e], Bram Martijn[f], Petra Scholte[d], Rinnert Schurer[g] and Corine Houtman[c,h]

[a] KWR Water Research Institute, Chemical Water Quality and Health, P.O. Box 1072, 3430 BB Nieuwegein, The Netherlands
[b] Utrecht University, Institute for Risk Assessment Sciences, Yalelaan 104–106, 3584 CM Utrecht, The Netherlands
[c] The Water Laboratory, P.O. Box 734, 2003 RS Haarlem, The Netherlands
[d] Waternet, Drinking Water Treatment, Korte Ouderkerkerdijk 7, 1096 AC Amsterdam, The Netherlands
[e] Dunea, P.O. Box 756, 2700 AT Zoetermeer, The Netherlands
[f] PWN Technologies R&D, P.O. Box 66, 1619 ZH Andijk, The Netherlands
[g] Evides Water Company, P.O. Box 4472, 3006 AL Rotterdam, The Netherlands
[h] Amsterdam Institute for Life and Environment (A-LIFE), VU University Amsterdam, Van der Boechorstraat 3, 1081 BT Amsterdam, The Netherlands
*Corresponding author. E-mail: renske.hoondert@kwrwater.nl

RH, 0000-0001-6990-9851; MMLD, 0000-0002-3531-7927

ABSTRACT

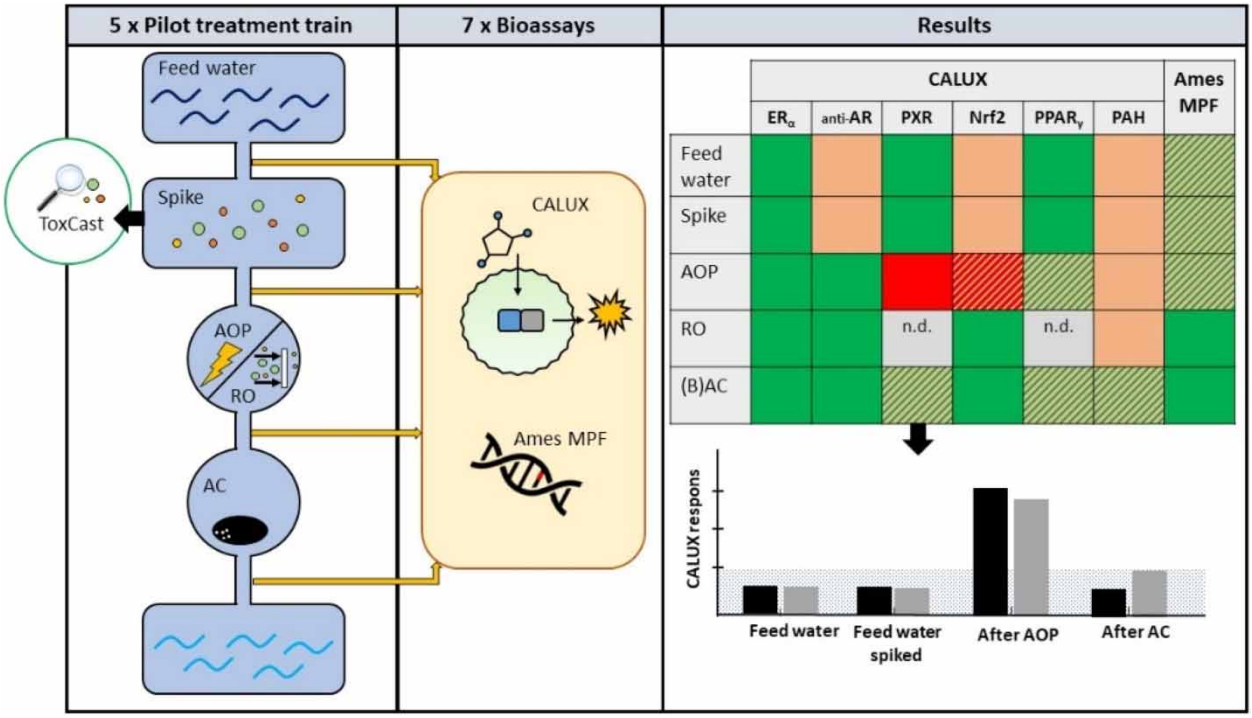
Chemical water quality monitoring is increasingly complemented with bioassays. The performance of drinking water treatment technologies was evaluated with bioassays. CALUX bioassays and the Ames fluctuation bioassay for mutagenicity were used to analyze the responses of raw feed water with and without a spiking mixture of organic contaminants in different pilot-scale water treatment trains applying advanced oxidation processes, reverse osmosis and filtration with activated carbon for the preparation of drinking water. In general, CALUX responses in the spiked feed water were in the same range as in the unspiked feed water, indicating that spiking did not significantly elevate the activity measured in the feed water. This observation was in line with a calculation of the combined additive behavior of the spiking mixture using bioactivities of the spiked chemicals in analog bioassays from ToxCast. This calculation gave no indication of a measurable response in the bioassays to be expected by the spiking, except for oxidative stress. Responses in the Ames fluctuation assay in most feed water were in some cases increased by advanced oxidation processes (AOPs) and mostly removed by active concentrations. The bioassay responses demonstrate the removal of emerging chemicals with different mechanisms of action by different water treatment technologies.

Key words: effect-based methods, pilot studies, removal, ToxCast, treatment efficiency, water quality

HIGHLIGHTS

- Bioassays were applied to evaluate pilot-scale water treatments fed with spiked influent water.
- Advanced oxidation induced the formation of compounds with different bioassay responses that are removed by activated carbon.
- Different treatment technologies remove emerging chemicals with different mechanisms of action.
- Concentration addition was applied to *in vitro* data to predict the activity of the spiking mixture.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Organic micropollutants (OMPs) end up in the environment as a result of the everyday use of substances in the modern society (Tosun *et al.* 2020). The number of new chemicals synthesized and marketed increases exponentially (Binetti *et al.* 2008; Bernhardt *et al.* 2017). Many chemicals reach surface waters via diffuse or local emissions as a result of domestic use, industrial discharge, hospital effluent, and insufficiently effective wastewater treatment (Tosun *et al.* 2020). At the same time, millions of European inhabitants depend on surface waters for their need of drinking water. Drinking water companies are as such increasingly confronted with mixtures of new and different OMPs present at low – but detectable – concentrations in their water sources (Fischer *et al.* 2017). Therefore, the companies use a multitude of screening methods to get insight into the presence of OMPs, and they often combine complementary treatment processes ('multibarrier approaches') to remove as much of the micropollutants as possible. Since these OMPs cover a wide range of physicochemical properties, a combination of treatment processes (targeting different removal mechanisms such as size exclusion, oxidation, and adsorption) is required to comply with drinking water standards that have become more stringent throughout the years (e.g., Post 2021).

The research presented in this study was conducted within a joint program set up by four Dutch surface water companies to safeguard drinking water quality. They proactively investigate the robustness of their water treatment processes for the removal of micropollutants ('Robustness of treatments'). Within this program, every 5 years the four companies assess the removal capacity for micropollutants of processes that are operated at the time or aimed to be implemented in the near future. To this aim, each time a selection was made of OMPs that were considered relevant for drinking water production. These chemicals have been frequently detected or have been detected in significant concentrations in Dutch surface waters and were selected based on (lack of) knowledge of their potential toxicity, challenges regarding effective treatment, and/or their physicochemical properties and analytical detectability. The chemicals were spiked in pilot-scale experiments and their fate during water treatment trains was assessed.

Five water treatment processes were investigated, including advanced oxidation processes, reverse osmosis (RO) in combination with ultrafiltration and (biological) activated carbon filtration steps. As targeted chemical analyses cannot completely capture the effects exerted by all chemicals in a mixture, a panel of bioassays was used to quantify the chemical burden and assess their utility for the assessment of water quality changes, alongside non-target screening (NTS) (Brunner

et al. 2020), and quantitative target chemical analysis of 34 spiked OMPs. Bioassays are biological test systems that directly detect integrated biological responses of chemicals that have a similar associated interaction with the biological targets (such as receptor proteins). As OMPs have different interactions with biological systems (mechanisms of action), appropriate bioassays need to be applied to detect their specific effects (endpoints). As mixtures of OMPs typically exert diverse types of biological effects, a panel of bioassays for effect-based monitoring needs to include several bioassays detecting different mechanisms of action. Several studies have included panels of bioassays to monitor water quality improvement by drinking water treatment processes, although the treatment processes and/or bioassay endpoints included differ (Neale *et al.* 2020; Oskarsson *et al.* 2021; Kienle *et al.* 2022). For specific purposes, such as the evaluation of the performance of water treatment technologies, a relevant and efficient set of bioassays can be selected on a case-by-case basis, considering, e.g., the specific chemical composition of local sources and the likelihood of the formation of transformation products that might exert specific biological activities. To assess whether a response may be related to potential risks, the effect level at which potential risks cannot be excluded is defined using effect-based trigger (EBT) values (van der Oost *et al.* 2017; Neale *et al.* 2022). Different methods are published, and EBT values may differ for potential risks to human health, on the one hand, and the environment, on the other (Brand *et al.*, 2013; Escher *et al.*, 2015; van der Oost *et al.* 2017; Been *et al.* 2021). EBT values were used for interpretation of the data and may not always reflect realistic risks posed to human health.

A set of CALUX reporter gene assays and the Ames fluctuation test for mutagenicity were used to study the outcome of biological activities of water samples collected before and after treatment in pilot-scale water treatment trains. As a wide variety of chemicals were spiked in the experiments in the pilot treatment plants and owing to the use of advanced oxidation processes, bioassays for mechanisms related to reactive toxicity were included: the Ames fluctuation test for mutagenicity (Heringa *et al.* 2011), CALUX tests for oxidative stress response via the nuclear factor erythroid 2-related factor 2 (Nrf2) pathways activation (Nrf2 CALUX) (van der Linden *et al.* 2014), and metabolism via activation of the aryl hydrocarbon receptor (AhR, in the polycyclic aromatic hydrocarbon (PAH) CALUX) (Pieterse *et al.* 2013). In addition, two CALUX assays for hormone-related mechanisms (estrogenic activity measured in the estrogen receptor CALUX (ERα) and anti-androgenic activity measured in the anti-androgen receptor CALUX (anti-AR)) were included for comparability with earlier water quality studies. The ERα CALUX measuring estrogenic activity (activation of the female sex steroid hormone receptor estrogen receptor α) was chosen because low levels of estrogenic activity are commonly found in surface waters (Houtman *et al.* 2020). This also holds for anti-androgenic activity, i.e., antagonistic binding on the male sex hormone receptor (androgen receptor), as measured in the anti-AR CALUX assay (Houtman *et al.* 2021).

In the present study, we aimed to look into the impact of OMPs on bioassay responses as a measure of water quality and the effects of five water treatment processes. A targeted set of bioassays was used to evaluate changes in chemical water quality in pilot-scale drinking water treatment trains fed with a mixture of micropollutants. EBTs values and ToxCast bioactivity data of analog endpoints as in the applied bioassays were used for further interpretation and evaluation of the bioassay battery. The ToxCast bioactivity data were used to predict the combined effect of the spiking mixture, using the concept of concentration addition (Loewe & Muischnek 1926).

2. METHODS

2.1. Pilot treatment trains and process conditions

Four Dutch drinking water companies participated in the study, accounting for five different pilot treatment plants treating surface water or infiltrated surface water. General descriptions of the treatment processes are given below. A schematic overview is given in Table 1. More specific treatment process conditions are provided in the Supplementary Material (Table S1).

Pilot treatment plant A consisted of advanced oxidation by O_3/H_2O_2 followed by UV/H_2O_2. Feed water was obtained from a full-scale installation. In this full-scale installation, surface water is abstracted from a side branch of the River Meuse. After dosing with $FeSO_4$ to reduce phosphate concentration and sedimentation, the water was treated by microsieve filtration and rapid sand filtration. At this point, the feed water for the pilot treatment plant was abstracted. For drinking water production, this water was infiltrated in the dunes with a residence time of approximately 2 months, followed by softening, powdered activated carbon treatment, aeration, and rapid and slow sand filtration, subsequently. Samples within the spiking experiments were taken on 5 October and 12 October 2017, several days after spiking the feed water. The average feed flow rate in this pilot treatment plant was 5 m^3/h. More information on pilot treatment plant A is outlined in Table S1 in the Supplementary Material or in the study by Brunner *et al.* (2020).

Table 1 | Collected samples (treatment steps, from top to bottom, are sequential)

Pilot treatment plant A	Pilot treatment plant B	Pilot treatment plant C	Pilot treatment plant D	Pilot treatment plant E
feed water	feed water	feed water	feed water	feed water
↓	↓	↓	↓	↓
feed water spiked	feed water spiked	feed water spiked	feed water spiked	feed water spiked
↓	↓	↓	↓	↓ ↓ ↓
after O_3/H_2O_2	after RO	after UV/H_2O_2	after O_3	After freshly reactivated BACF (20 min)
↓		↓	↓	
after UV/H_2O_2		after storage	after BACF1	After saturated BACF1 (10 min)
		↓	↓	after saturated BACF2 (20 min)
		after GAC	after BACF2	

Note: BAC, biologically active carbon; BACF, biologically active carbon filtration; RO: reverse osmosis.

Pilot treatment plant B consisted of ultrafiltration and RO. Feed water for pilot treatment plant B was abstracted from a full-scale installation in which surface water from Lake IJssel was pretreated by aeration, drum sieves, coagulation/floc formation, lamella separators, rapid sand filtration, and granular activated carbon (GAC). At this point in the treatment train, the feed water for the pilot treatment plant was abstracted. For drinking water production, this pretreated water was further treated in the full-scale plant, either by (1) ultrafiltration followed by RO (as in the pilot treatment plant) or by (2) UV in combination with H_2O_2, followed by GAC filtration and dune infiltration. After abstraction from the dunes, this water was further treated by softening, aeration, rapid sand filtration, and dosing of chlorine dioxide. Water from the two lines was mixed before distribution. The spike solution was added after the ultrafiltration step to assess the impact of the RO process. Samples within the spiking experiments were taken on 21 September and 22 September 2017, 72 and 96 h after spiking the feed water, respectively. The average feed flow rate in this pilot treatment plant was 9.7 m^3/h. More information on pilot treatment plant B is outlined in Table S1 in the Supplementary Material or in the study by Brunner *et al.* (2020).

Pilot treatment plant C consisted of UV in combination with H_2O_2 followed by GAC treatment, as described in line 2 of the full-scale plant. The feed water for pilot treatment plant C was the same as that for pilot treatment plant B. Samples within the spiking experiments were taken on 3 –5 October 2017, several days after spiking the feed water. The average feed flow rate in this pilot treatment plant was 18.3 m^3/h. More information on pilot treatment plant C is outlined in Table S1 in the Supplementary Material or in the studies by Ebrahimzadeh *et al.* (2021) and Brunner *et al.* (2020).

Pilot treatment plant D consisted of ozone treatment, followed by a two-step biologically active carbon filtration (BACF). For the BAC treatment step, carbon batches from the full-scale installation with two different operating (service) times were used, with the filter in the last step containing the batch with the shortest operating time (least preloading). Feed water was abstracted from a full-scale installation in which surface water from the Lekkanaal (a channel fed with water mainly originating from the river Rhine) at Nieuwegein was treated by dosing with $FeCl_3$, sedimentation, rapid sand filtration, dune infiltration and abstraction, and subsequent rapid sand filtration. At this point, the feed water for the pilot treatment plant

was obtained. Samples within the spiking experiments were taken on 6 November 2018 (duplicates), several hours after spiking the feed water. The average feed flow rate in this pilot treatment plant was 5 m^3/h.

Pilot treatment plant E consisted of three different treatments studied in parallel using columns filled with saturated BACF, from which samples were taken after 10 (sampling point 3) and 20 min (sampling points 2 and 4) of empty bed contact time, respectively. The feed water of the pilot trains consisted of water abstracted from the river Meuse, followed by reservoir storage that was pretreated with coagulation, sedimentation, dual media filtration, and UV disinfection. Samples within the spiking experiments were taken (duplicate) on 18 October and 19 October 2018, several hours after starting spiking the feed water. The average feed flow rate in this pilot treatment plant was 2.8 m^3/h for each BACF filter unit.

2.2. Spiking and sampling

Spike chemicals were selected based on their priority indicated in databases and reports of the Association of River Water Works (RIWA) (Slootweg & Houtman 2014; Stroomberg *et al.* 2018; Pronk *et al.* 2020; Slootweg *et al.* 2022), the Water Framework Directive (WFD) guideline (https://environment.ec.europa.eu/topics/water/water-framework-directive_en), and company-specific research (data not shown). Chemicals were selected in case measured concentrations (monitored between 2011 and 2015) repeatedly exceeded drinking water standards or target values (0.01 μg/L for genotoxic compounds, 0.10 μg/L for other biologically active compounds, and 1.00 μg/L for other anthropogenic compounds without known specific biological activity; Mons *et al.* 2013) or if often detected at concentrations above 50% of these values, and based on (lack of) toxicological information, challenges regarding effective treatment and/or their physicochemical properties, analytical detectability, and commercial availability. The selected chemicals (Table 2) were prepared in a spike solution to be added to the feed water of the pilot treatment plants, in which their concentrations were defined based on limits of quantification (LOQs) of the target analytical methods, the flowrates of the pilot treatment plants, dosing time, and knowledge on removal efficiencies leading to nominal concentrations as given in Table 2. A more detailed paragraph on the preparation of the spiking solution can be found in section I in the Supplementary Material.

Duplicate experiments took place on two different days (fall 2017 (plants A–C) and fall 2018 (plants D and E)). As the experiments took place on different days, the time between the preparation of the dosing solution and their use in the spiking experiments differed between the plants and between duplicate experiments. This could, in principle, cause the actual concentrations on the days of the experiments to deviate slightly between the pilot treatment plants. Therefore, actual concentrations of the spike chemicals in the spiked feed water were determined chemically during each experiment, with quantitative chemical analysis methods, as described in Brunner *et al.* (2020). Samples were collected from feed water, spiked feed water, and after specific treatment steps (Table 1). Water samples were stored in glass bottles at 4 °C and processed within 1 week. Details about the procedures of the experiments can be found in Brunner *et al.* (2020).

2.3. Prediction of bioactivity of the spiking mixture

2.3.1. ToxCast data

To explore if specific *in vitro* bioactivities of the chemicals in the spiking mixture would expectedly evoke responses in (one of) the CALUX bioassays, ToxCast bioactivity data (active concentration (AC_{50}); the concentration that induces half-maximum response, reported in μM) for analog assay endpoints in the ToxCast database were extracted from the CompTox dashboard, together with ToxCast AC_{50} data for the reference compounds and experimentally obtained AC_{50} values (Tables S2 and S3 in the Supplementary Material) (accessed 11 January 2022, via https://comptox.epa.gov/dashboard/). This analysis assumed that analog ToxCast *in vitro* assay responses were predictive of the response in CALUX bioassays if their respective endpoints can be considered mechanistically analog (Been *et al.* 2021). Secondary read-outs, such as cell viability or control conditions, were excluded. Only bioassay-compound combinations with an active hit call were included in the analysis. The concentration response series only got an active hit call (and high-quality rating) when they met three criteria: (i) the Hill model should emerge as the best model, (ii) the top of the modeled curve must be above the efficiency threshold (efficacy cutoff), and (iii) for at least one concentration, the median response must be above the efficacy threshold). AC_{50} concentrations (in μM) were converted to AC_{50} concentrations in μg/L, based on the molecular weight of the compound.

2.3.2. Mixture toxicity

If the potencies of all individual spike compounds in the applied bioassays were known, it would be possible to predict the combined effect of the spiking mixture, by using the concept of concentration addition (Loewe & Muischnek 1926).

Table 2 | Spiked chemicals

Chemical name	Use category	Nominal spiking concentration (μg/L)	Spiked feed water A (μg/L)		Spiked feed water B (μg/L)		Spiked feed water C (μg/L)		Spiked feed water D (μg/L)		Spiked feed water E (μg/L)	
			1	2	1	2	1	2	1	2	1	2
Acesulfame-K	Sweetener	0.1	13	14	11	7.3	11	10	8	8.4	9	8.8
AMPA	Glyphosate metabolite	0.02	NA	NA	NA	NA	NA	NA	NA	NA	2	2
Aniline	Industrial chemical	0.01	1.1	0.98	0.91	0.4	0.81	0.8	0.5	0.51	0.55	0.52
Barbital	Pharmaceutical	1	11	12	8.8	6	8.3	9.3	8.3	9.2	8.5	8.2
1H-benzotriazole	Industrial chemical	0.01	1.9	1.9	1.3	0.98	1.4	1.3	1	1.1	1.3	1.3
Carbamazepine	Pharmaceutical	0.01	1.2	1.2	1.1	0.68	0.95	0.96	0.93	0.95	0.97	0.96
Carbendazim	Fungicide	0.01	0.41	0.45	0.38	0.24	0.29	0.32	0.65	0.66	0.75	0.73
Diatrizoic acid	Contrast agent	0.03	5.1	4.9	4.1	2.7	4	3.9	2.4	2.5	2.5	2.5
Diclofenac	pharmaceutical	0.01	0.11	0.12	0.12	0.07	0.09	0.09	0.17	0.18	0.18	0.17
Dimethenamid	Herbicide	0.01	1.2	1.2	1	0.64	0.93	0.94	0.66	0.67	0.71	0.72
Dimethomorph	Fungicide	0.01	1.1	1.1	1	0.6	0.79	0.78	0.79	0.78	0.81	0.79
EDTA	Industrial chemical	0.05	NA	NA	NA	NA	NA	NA	NA	NA	13	12
Phenobarbital	Pharmaceutical	0.5	11	11	9.1	6.2	7.5	8.6	7.3	7.6	7.3	7.2
Furosemide	Pharmaceutical	0.03	1.7	1.6	1.8	1	0.92	0.97	1.5	1.5	1.5	1.6
Gabapentin	Pharmaceutical	0.01	1.6	1.6	1.3	0.68	1.2	1.2	0.85	0.87	0.97	0.95
Glyphosate	Herbicide	0.05	NA	NA	NA	NA	NA	NA	NA	NA	4.3	4.3
HFPO-DA (Gen-X)	Industrial chemical	0.2	13	13	12	6.9	9.6	10	8.6	9	9.5	9.7
HMMM	Industrial chemical	0.03	0.27	0.26	0.22	0.24	0.23	0.22	<LOQ	<LOQ	0.14	0.13
Hydrochlorothiazide	Pharmaceutical	0.05	5.2	5.8	4.6	2.5	4	3.9	4.3	4.4	4.5	4.7
Melamine	Industrial chemical	0.05	9.1	9.3	6	3.8	5.4	5.4	5	5	6.9	6.6
Metformin	Pharmaceutical	0.05	5.9	5.5	4.9	2.9	4.8	4.6	3.7	3.9	4.2	3.8
4-methyL-1H-benzotriazolea	Industrial chemical	0.01	1.2	1.3	0.94	0.64	0.86	0.83	0.78	0.8	1	0.98
5-methyl-1H-benzotriazolea	Industrial chemical	0.01	1.7	1.7	1.4	0.84	1.2	1.3	1	1.1	1.2	1.1
Propranolol	Pharmaceutical	0.01	2	2	2.4	1.3	1.4	1.5	1.7	1.8	1.8	1.8
Pyrazole	Industrial chemical	0.5	11	13	11	6.8	8.9	9.2	8.9	8.7	7.6	7.3
Sucralose	Sweetener	1	13	13	9.3	6.2	8.2	8.4	10	11	11	12
Terbuthylazine	Herbicide	0.01	0.57	0.58	0.51	0.33	0.46	0.44	0.3	0.31	0.38	0.38
Tetraglyme	Industrial chemical	0.03	4	4	3.7	2.3	3.3	3.2	2.7	2.8	3	3
TFA	Industrial chemical	0.05	6.5	6.3	5.2	7.3	6	5.9	5.8	5.9	5.4	5.7
Tiamulin	Pharmaceutical	0.01	2.4	1.5	1.9	0.9	0.68	0.65	1.6	1.8	1.6	1.6
Triphenylphosphine oxide (TPPO)	Industrial chemical	0.01	1.3	1.3	1.2	0.76	1	0.99	0.92	0.94	0.93	0.93
Tramadol	Pharmaceutical	0.01	1.3	1.2	1.1	0.66	0.93	0.89	1.1	1.2	1.2	1.2
Urotropin	Industrial chemical	0.05	2.9	3	2.8	0.69	2.8	2.7	1.5	1.4	1.1	1.1
M/P-Xylene	Industrial chemical	0.05	NA	NA	NA	NA	NA	NA	5.1	5.4	NA	NA

[a]Mixture of 35% 4-methyl-1H-benzotriazole and 65% 5-methyl-1H-benzotriazole.

Source: PUBCHEM and analytical methods included in Brunner *et al.* (2020), and the analyzed concentrations in spiked feed water introduced in the different pilot treatment plants (two replicates)

However, since individual potencies were not available for most of the 34 spiked compounds in the applied bioassays, ToxCast bioactivity data (AC_{50} data) in bioassays for analog endpoints to the CALUX bioassays in our study were collected and used to predict a possible mixture effect. As no analog endpoints were available for responses in the Ames fluctuation test for mutagenicity (yet), this analysis was not performed for the Ames tests. Expected mixture effects for each pilot treatment plant were calculated based on toxic units for each individual chemical i (TU is the ratio between the actual concentration measured at the location and the concentration evoking an effect in the assay (AC_{50})).

$$TU_i = \frac{c_i}{AC_{50,i}} \tag{1}$$

Individual potencies of compounds were expressed relative to those of the reference compounds in the CALUX assays (Table S3; anti-AR CALUX – flutamide; ER CALUX – 17β-estradiol; Nrf2 CALUX – curcumin; PAH CALUX – benzo(a)pyrene; PXR CALUX – nicardipine; and PPARγ CALUX – rosiglitazone) by calculating their relative potency factors (RPFs). RPFs for each of the 14 chemicals were calculated based on the chemical-specific AC_{50} values for the analog ToxCast assays (Table S2) and the (measured) AC_{50} values for the reference compounds for the specific CALUX assay endpoints.

$$RPF = \frac{AC_{50-ref}}{AC_{50-x}} \tag{2}$$

If no experimental data were available (i.e., in the case of Nrf2 CALUX), the AC_{50} of the reference compound was based on the arithmetic mean of the AC_{50} values of the analogous ToxCast assays (Table S4).

The measured concentrations in the pilot treatment plants could then be expressed in equivalents of the reference compound by multiplying the respective concentrations by the corresponding RPF (Equation (3)). However, prior to calculating the equivalents, AC_{50} values in μM were needed to be converted to μg/L to match the concentrations of the feed water in the pilot treatment plants.

$$EQ_i = c_i \cdot RPF_i \tag{3}$$

The total expected mixture toxicity, expressed in equivalents, was determined according to the concentration addition principle.

$$\sum_{i=1}^{n} EQ_i = EQ_{mix} \tag{4}$$

If AC_{50} values for multiple analog ToxCast assays were available per endpoint, the most conservative value was chosen. As much data were lacking on the effects of individual chemicals, we consider the mixture toxicity as described above as a measure for the minimal expected activity of the mixture.

2.4. Bioassays

2.4.1. Performance of CALUX bioassays

CALUX® cells were supplied by BioDetection Systems B.V. (Amsterdam, the Netherlands). CALUX bioassays for anti-AR, ERα, PAHs, oxidative stress response (tested in the Nrf2 CALUX), activation of the PXR receptor and activation of the PPARγ receptor were performed according to the supplier's protocols. The PXR CALUX and PPARγ CALUX were only performed at pilot treatment plant D, based on observations in earlier company research (data not shown). Water samples (1 L) were taken from the feed water and treated water of the pilot treatment plants in two independent experiments (acting as duplicates), stored in glass bottles at 4 °C and processed within 1 week. Water samples were concentrated by solid-phase extraction and evaporation of the solvents (Houtman *et al.* 2018; Alygizakis *et al.* 2019).

For anti-AR, ERα, PAH, and Nrf2 CALUX, compounds in the samples were extracted on Oasis hydrophilic-lipophilic balanc solid phase extraction (HLB SPE) cartridges, as described in Houtman *et al.* (2018), and extracts were reconstituted in 50 μL dimethylsulfoxide (DMSO). From these extracts, 3-, 10-, 30-, and 100-fold dilutions were prepared in DMSO. Extracts and dilutions were tested in the ERα CALUX and antagonistic mode of the AR CALUX (anti-AR CALUX) at 0.1% extract concentration, as described previously (Van der Linden *et al.* 2008; Houtman *et al.* 2018) with reference compounds 17β-

estradiol (E2) and flutamide, respectively. The anti-androgenic activity was tested in the presence of DHT at the EC_{50} level. Induction of the AhR was tested in the same extracts and dilutions at an exposure concentration of 0.8% in the PAH CALUX, as described by Pieterse *et al.* (2013), with benzo(a)pyrene (B(a)P) as a reference compound. Oxidative stress was tested at an exposure concentration of 1.0% in the Nrf2 CALUX as described by van der Linden *et al.* (2014) with curcumin as a reference compound. For the PXR and PPARγ CALUX, separate extracts were prepared, using similar SPE cartridges and volumes, but using methanol and acetonitrile as solvents for the elution. PXR and PPARγ CALUX were exposed to 0.1% extract concentrations, according to Gijsbers *et al.* (2011) and Alygizakis *et al.* (2019). Nicardipine and rosiglitazone were used as their respective reference compounds.

2.4.2. Performance of the Ames fluctuation test

The Ames fluctuation test uses genetically modified *Salmonella typhimurium* bacteria to investigate the potential induction of DNA mutations by exposure to chemicals or environmental samples to determine a potential for genotoxicity (Reifferscheid *et al.* 2011). The Ames fluctuation test was performed, as reported previously (Heringa *et al.* 2011) with minor modifications (regarding the used *Salmonella* strains, culture media, statistics, and interpretation). Strains TA98 and TA100 were used to detect frame-shift mutations and base-pair substitution, respectively. Water samples were concentrated by SPE using Oasis® HLB cartridges (Waters Corporation, Milford, CT, USA), evaporating the solvents and dissolving the 10,000 times concentrated water sample extract in DMSO, as reported previously (Kolkman *et al.* 2013). The extracts were stored at −18 °C prior to bioassay testing. Extracts and procedure controls were tested in triplicate with and without the S9 enzyme mix in two independent experiments. Solvent (DMSO) control and positive controls (in DMSO) were also tested in triplicate [TA98 − S9: 20 μg/mL 4-nitroquinoline N-oxide (4-NQO) and 500 μg/mL4-nitro-o-phenylenediamine (4-NOPD); TA98 + S9: 5 μg/L 2-aminoanthracene (2-AA); TA100 − S9: 12.5 μg/mL nitrofurantoin (NF); TA100 + S9: 20 μg/mL 2-AA]. Histidine, nutrient broth no. 2 oxoid, 2-AA, $MgCl_2{\cdot}6H_2O$, $NaH_2PO_4{\cdot}H_2O$, and $Na_2HPO_4{\cdot}2H_2O$ were obtained in analytical grade from Boom (Meppel, the Netherlands). NaCl and KCl were purchased from Avantor Performance Materials B.V. (Deventer, the Netherlands). 4-NOPD, 4-NQO, NF, D-glucose-6-phosphate, nicotinamide adenine dinucleotide phosphate, and ampicillin were purchased from Sigma-Aldrich (Zwijndrecht, the Netherlands). The 24- and 96-well plates were supplied by Greiner Bio-One (Alphen A/d Rijn, the Netherlands) and the Corning 384-well plates from Sigma-Aldrich. Cytotoxicity (reduced optical density at 595 nm) may result in false-negative responses, while increased cell growth may result in higher sensitivity of the assay. Therefore, cell growth and survival were checked using a custom cytotoxicity test with subsamples of the exposure cultures in a medium with histidine.

2.5. Data analysis and interpretation

2.5.1. CALUX bioassays data

Responses of dilutions of the water extracts giving response between the LOQ and the EC_{50} were interpolated in the standard curve and responses were expressed as equivalents of the reference compounds per liter of water (Houtman *et al.* 2004). Responses were benchmarked against EBTs for environmental toxicity (van der Oost *et al.* 2017) and human health (Been *et al.* 2021). As spiking concentrations were not based on biological activities in this study, a risk evaluation of the spiked water as such based on the bioassay results was less relevant. EBTs were used here to evaluate unspiked feed water for environmental quality and drinking water for human health and more generally to benchmark bioassay responses for which numerical results can vary widely. For the Nrf2 CALUX and the PPARγ CALUX, the EBTs were below the respective LOQs in these tests. By definition, every exceedance of the LOQ also exceeds the EBT, and the exceedance of the EBT cannot be ruled out when positive responses are not measured.

2.5.2. Ames fluctuation test data

Ames results were expressed as the number of cell culture wells (total number of wells: 48) in which a color change of the pH indicator in the culture medium is observed. Maximum (25) and minimum (10) average numbers of wells in which a color change is observed are considered for the solvent control and positive control conditions, respectively. Test conditions were compared to solvent and procedure controls (as a control for potential false-positive results caused by contamination during sample preparation). A chi-square test (for data with binomial distribution) was used to determine statistically significant differences ($p < 0.05$) between the control and the test conditions. As no dilutions were included in the Ames fluctuation test (unlike in the CALUX assays), a lot of uncertainty may be introduced when testing only a single sample. In addition, the Ames fluctuation test only results in a yes/no answer and no quantitative result (effect concentration), and concentrations of compounds in surface

waters and influents fluctuate considerably in space and time. Therefore, (at least) two independent Ames fluctuation tests were performed in this study. A tested sample was considered mutagenic if a statistically significant response was replicated in two independent experiments in at least one of the test conditions (TA98 − S9, TA98 + S9, TA100 − S9, and TA100 + S9), based on an earlier work by Beal *et al.* (2023). This was different from the assessment method for the Ames fluctuation test based on a baseline. However, in practice, results may be similar, unless there is a lot of variability in concentration data.

3. RESULTS AND DISCUSSION

3.1. Prediction of possible bioactivities of the spiking mixture

3.1.1. Prediction of possible activities of the spiking mixture in the anti-AR, ERα, PAH, Nrf2, PXR, and PPARγ CALUX bioassays

Information on the activity on analog ToxCast assay endpoints was available for a subset of 14 of the spiked chemicals (Supplementary Table S3). For each of the CALUX tests applied here, several of the spiked chemicals elicit responses in analog ToxCast assay endpoints, albeit at higher concentrations than observed in the pilot treatment plant experiments. For none of the individual spiked chemicals, the concentration measured in the spiked feed water would be expected to elicit a response in an analog ToxCast endpoint for any of the CALUX tests applied here. Individual spiked chemicals thus do not explain (part of) the observed responses. For the other chemicals included in the spiked mixture of chemicals fed into the pilot treatment plants, only negative (inactive) bioactivity data were available in the CompTox chemical dashboard (Table 3). For two chemicals, the majority of the reported bioactivity data had a negative (inactive) hit call: phenobarbital (245/255) and melamine (212/219). The total predicted mixture activities, expressed in equivalents of the reference compound, are shown in Table 4. The predicted mixture activities slightly deviate between pilot treatment plants because the actual concentrations of the spiking compounds were slightly different between the pilot experiments (see Table 2). The predicted mixture activities in the anti-AR, ERα, PAH, and PPARγ CALUX were undetectably low. Only in the Nrf2 and PXR CALUX, some measurable mixture activity (a maximum of 2.2-fold the LOQ) was predicted. In the ERα and PPARγ CALUX, indeed no activity was measured in the pilot plant samples (Figure 1; detailed discussion of pilot plant results will follow in the next section). While we did not expect mixture toxicity, activities were measured in the feed waters in the anti-AR CALUX (all plants) and PAH CALUX (with the exception of treatment installation C), implying that other compounds already present in the feed water contributed to CALUX responses and/or that possibly one or more compounds lacking AC_{50} data for all analog assays in ToxCast contributed as well. In addition, for Nrf2 (except in the feed water of plants A and E) and PXR CALUX, no activities were measured, while (small) mixture activities were expected based on mixture toxicity calculations. This might be due to the fact that the AC_{50} values might not 100% be equal or representative between the CALUX assays used in this study and the analog assays for similar endpoints in ToxCast. This might specifically hold for urotropin for which a high potency (low AC_{50}) for Nrf2 was found in ToxCast, leading to high predicted mixture activity, while no activity in the Nrf2 CALUX was found in the spiked water from the pilot plants B, C, and D.

3.1.2. Likelihood of possible activity of the spiking mixture in bioassays for other toxicological endpoints

The bioactivity of spike compounds reported in the ToxCast database for all other ToxCast assay endpoints was considered to give an indication if bioassays for other endpoints would be valuable additions to the panel of bioassays in future studies evaluating water treatments. Only a subset of 14 micropollutants in the spiked feed water activity could reasonably be expected if these assays were applied. In 28 chemical endpoint combinations, bioactivity was elicited in one of the 156 ToxCast assay endpoints included in this study. These endpoints included mainly assays based on human embryonic kidney cell lines and human liver cancer cell lines. These assay endpoints were not considered analogous to the CALUX assays in the present study and may be used in a follow-up study to monitor the removal of compounds eliciting bioactivity. Table S2 provides chemicals that elicited responses for at least one out of 156 different ToxCast assay endpoints. However, the AC_{50} values were in most cases higher than the concentrations in the pilot treatment plant experiment. Table S3 provides chemicals for which AC_{50} data from ToxCast bioassays were available, but none of the concentrations in the spiked feed water exceeded these AC_{50} data.

3.2. CALUX bioassays results of pilot installation samples

The CALUX bioassay results of the duplicate samples from the pilot installations are shown in Figure 1.

Table 3 | RPFs for all compounds included in the analysis

		Bioactivity of spiked chemicals micropollutants													
		RPFs													
	Analog toxcast assay endpoints	1H-benzotriazole	5-methyl-1H-benzotriazole	Carbamazepine	Carbendazim	Diatrizoic acid	Dimethenamid	Dimethomorph	Phenobarbital	Furosemide	HMMM	Hydrochlorothiazide	Terbuthylazine	Triphenylphosphine oxide (TPPO)	Urotropin
Anti-AR CALUX	TOX21_AR_BLA _antagonist_ratio						1.8E-02	4E-03					8.3E-03		
	TOX21_AR_LUC _MDAKB2_Antagonist _0.5 nM_R1881						7.6E-02	1.3E-01							
	TOX21_AR_LUC_MDAKB2 _Antagonist_10 nM_R1881						4E-03	3.5E-02	8.9E-02						
ERαCALUX	ATG_Era_TRANS_up					4E-08					7E-08		4E-06		
	ATG_ERE_CIS_up	5E-08		1E-07		3E-07					2E-07	2E-06	8E-07		
	TOX21_Era_BLA _Agonist_ratio												1E-07		
	TOX21_Era_LUC _VM7_agonist				2E-07	3E-08							2E-06		
Nrf2CALUX	ATG_NRF2_ARE_CIS_up				8.3E-1		3.5E0	8.7E-01		1.4E-01				2.8E-01	2.1E01
	TOX21_ARE_BLA _agonist_ratio						2.2E0								
PAH CALUX	ATG_Ahr_CIS_up	2E-04	5E-03		5E-03			1E-04							
	TOX21_AhR_LUC_agonist				2E-03										
PXR CALUX	ATG_PXR_TRANS_up			3.2E-03									1.4E-01	6.4E-01	
	ATG_PXRE_CIS_up			1.8E-02			2.7E-01	5.2E-02			5.8E-03	9.5E-02	7E-02	6.7E-02	
PPARγ CALUX	ATG_PPARg_TRANS_up						1.9E-02			1.7E-03					
	ATG_PPRE_CIS_up							3.2E-02				2.4E-02			
	TOX21_PPARg_BLA _agonist_ratio														

Table 4 | Total calculated mixture activity (in equivalents of the corresponding reference compound) for each pilot treatment plant and CALUX assay type

		ANTI-AR CALUX (MG-EQ/L)	ER CALUX (PG EQ/L)	NRF2 CALUX (MG CURC-EQ/L)	PAH CALUX (NG-EQ/L)	PXR CALUX (MG-EQ/L)	PPARG CALUX (MG-EQ/L)
SPIKED FEED WATER A (EQ)	1	0.27	13	66	1.5		
	2	0.27	14	68	1.5		
SPIKED FEED WATER B (EQ)	1	0.24	11	63	1.2		
	2	0.15	7	18	0.8		
SPIKED FEED WATER C (EQ)	1	0.19	10	62	1.1	1.3	122
	2	0.20	10	60	1.1	1.3	125
SPIKED FEED WATER D (EQ)	1	0.18	9	35	1.1		
	2	0.19	10	33	1.2		
SPIKED FEED WATER E (EQ)	1	0.19	10	27	1.3		
	2	0.18	11	27	1.2		
LOQ		*1.4*	*34*	*30*	*10*	*1.1*	*150*

Notes: The orange cells marked in red have detectable activities (>LOQ). No bioassays have been performed for samples indicated by gray cells.

3.2.1. Anti-androgenic activity in the anti-AR CALUX

Unspiked feed water from all pilot treatment plants, except the duplicate of plant A, induced responses in the anti-AR CALUX assay (Figure 1, top panel). Many micropollutants in water are known to exhibit anti-androgenic activity. Therefore, it is likely that anti-androgenic activity was already present in the unspiked feed water due to the presence of unknown anti-androgens in the source waters (Aït-Aïssa *et al.* 2010; Christen *et al.* 2014; Liscio *et al.* 2014). The presence of anti-androgenic activity was a common observation in the unspiked feed water of these plants (for plants A, B, C, and E; Been *et al.* 2021), which may possibly be due to the presence of unknown anti-androgens in the source waters (i.e., raw feed waters). The composition of surface waters can fluctuate considerably over a short time. This can explain the differences in the concentration of the anti-androgenic activities between the duplicates of plant A, as this plant uses water from the river Meuse as its source.

The spiked feed water showed anti-androgenic activity in all plants and duplicates in a similar range as the unspiked feed water (except the duplicate of plant A). No detectable activity from the spiking mixture was expected. The measured activity was therefore most likely due to substances already present in the (undosed) feed water (see paragraph above). The highest anti-androgenic activity was measured in the unspiked and spiked feed waters of plant D. Sometimes, the water at this location contains elevated levels of anti-androgenic activity, as was also the case in fall 2018 (Houtman *et al.* 2021). Apparently, the pretreatment performed before the water was used as feed for the pilot fails to completely remove the substances that cause the anti-androgenic activity. This underlines the importance of a multibarrier approach in drinking water treatment as endorsed and applied by the companies participating in this study. The anti-androgenic activity was not detected after the treatments, indicating efficient removal by both advanced oxidation (O_3/H_2O_2 at plant A, UV-H_2O_2 at plant C, and O_3 at plant D), RO (at plant B), and activated carbon (BACF at plant E). At pilot treatment plant E, freshly reactivated carbon removed the compounds with anti-androgenic activity completely, whereas some activity at levels below EBT remained after using the saturated BACFs.

The unspiked and spiked feed water of plant D exceeded the EBT for environmental quality (anti-androgenic activity). The identification of responsible compounds was a topic of earlier research. Pesticides were identified by effect-directed analysis in water from the Meuse and in waste water treatment plant effluent to contribute to this activity (Houtman *et al.* 2020, 2021). A study investigating the nature of anti-androgenic activity at other drinking water sources, including the Lekkanaal, is

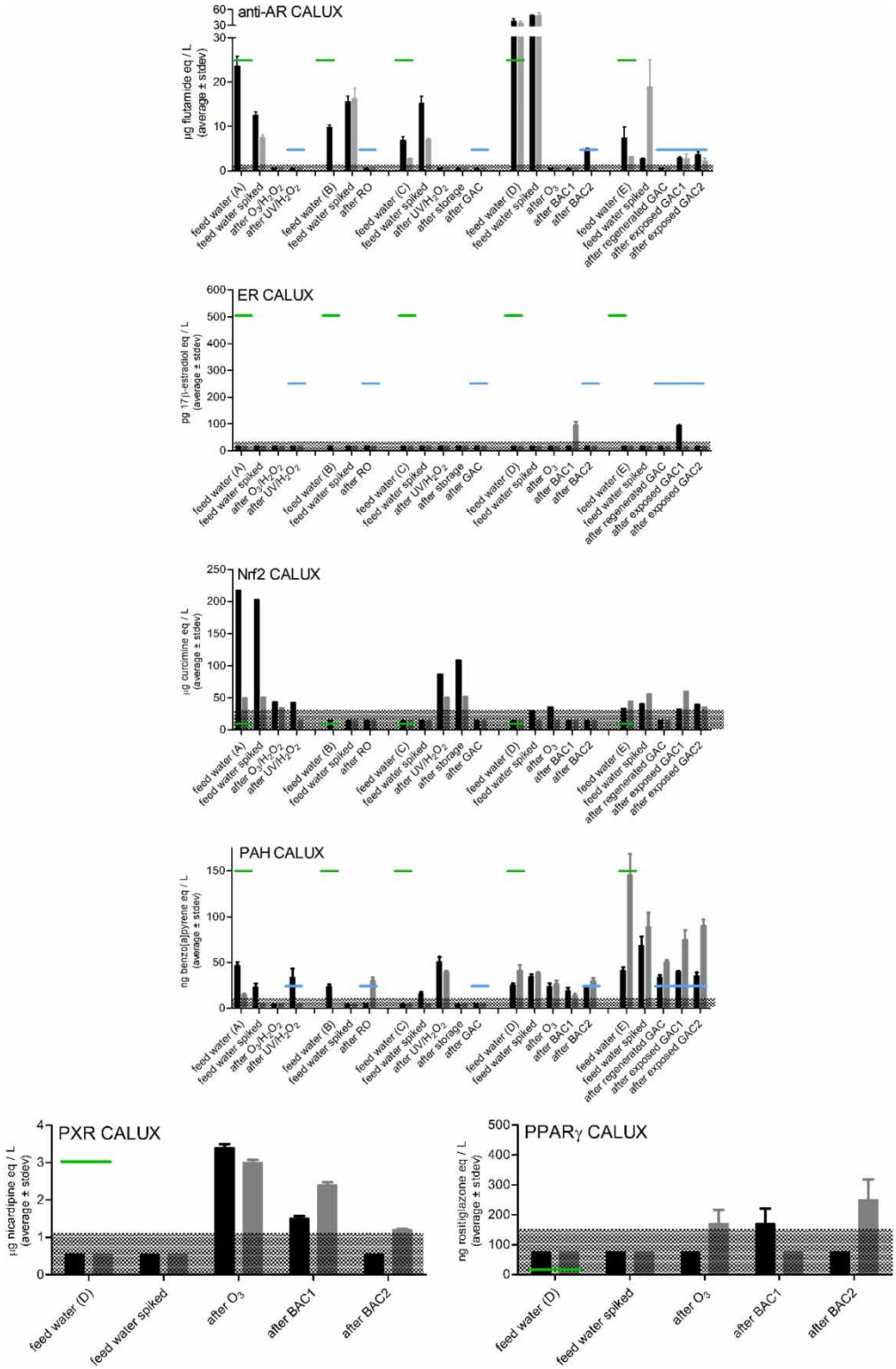


Figure 1 | Anti-AR, ERα, Nrf2, and PAH CALUX responses induced by water samples collected from pilot treatment plants (a)–(e). Treatment steps were sequential unless indicated in Table 1. The dotted area indicates the LOQ. The blue and green lines represent human EBTs (Been *et al.* 2021) and environmental EBTs (van der Oost *et al.* 2017), respectively. PXR and PPARγ CALUX responses induced by water samples collected from pilot treatment plant D. Treatment steps were sequential unless indicated in Table 1. The dotted gray area indicates the LOQ. The blue and green areas represent human EBT (Been *et al.* 2021) and environmental EBT (van Der Oost *et al.* 2017), respectively.

ongoing. All unspiked and spiked feed water samples exceeded the human EBT for anti-androgenicity (Been *et al.* 2021). However, all drinking water samples were below this value. This indicates that the applied treatment steps – even after spiking of feed water with micropollutants – removed anti-AR CALUX responses to a level at which no risk was expected for drinking water consumers from anti-androgenic activity.

3.2.2. Estrogenic activity in the ER*α* CALUX

No estrogenic responses in the ER*α* CALUX were observed in any of the pilot treatment plants in the unspiked and spiked feed water samples. This indicates that the spiking mixture indeed did not show measurable estrogenic activity, in line with the calculated mixture toxicity. For the treated samples from plants A, B, and C, no activity was observed. The capability to remove compounds with estrogenic activity of the investigated treatment processes at these plants therefore cannot be judged from these experiments. Anyhow, it can be concluded that the estrogenic activity was not generated during the treatment.

At pilot treatment plants D and E, a low activity was observed for one of the duplicates after activated carbon filtration. Possibilities are that desorption of OMPs accumulated at the (B)AC adsorptive sites occurs, where OMPs had accumulated in the GAC bed by adsorption onto the GAC medium during the earlier passage of pretreated water. This would be in line with the fact that no activity was seen for the reactivated carbon at pilot treatment plant E. Estrogenic activities in all samples were below EBT for both ecological as well as human health, indicating that no risk is expected from estrogenic activity for both the aquatic environment as for drinking water consumers.

3.2.3. Oxidative stress response in the Nrf2 CALUX

Unspiked and spiked feed waters of plants abstracting water from the Meuse (A and E) showed oxidative stress in the Nrf2 CALUX, while unspiked and spiked feed waters from plants B, C, and D showed no oxidative stress (Figure 1). Unknown compounds present in the raw Meuse water could explain this activity, and temporal fluctuations in the chemical water quality of this river could explain the high variability in the measured concentrations. Oxidative stress response was observed neither in unspiked feed water of the other plants nor after spiking in these plants, indicating no measurable contribution of the spiking mixture. This is in contrast with the calculated mixture toxicity of the spiking mixture (see Table 4), indicating a measurable activity only for the endpoint of Nrf2 induction. The calculated oxidative stress response of the spiking mixture was almost only (>90%) attributed to a single compound, urotropin. However, these results were based on only a single AC_{50} for urotropin for a single analogous bioassay from ToxCast. It is possible that the data are not representative of the activity of urotropin in the Nrf2 CALUX and the mixture effect is overestimated.

The activity in the spiked and unspiked feed water of plant A decreased considerably after oxidation with O_3/H_2O_2. On the contrary, a slightly elevated oxidative stress response was observed after UV/H_2O_2 treatment at plant C, probably due to the transformation of compounds. Compounds exerting this activity were removed by sequential activated carbon treatment. At plants B and D, (almost) no activity in the Nrf2 CALUX was observed. In plant E, reactivated GAC successfully eliminated the oxidative stress response, whereas exposed carbon did not. In plant D, saturated carbon (BAC1 and BAC2) eliminated the low activity produced after O_3/H_2O.

Both the spiked and unspiked feed waters of plants A and E exceeded the environmental EBT. A possible exceedance of the environmental EBT in both feed waters of the other plants cannot be judged as the EBT is lower than the LOQ of the assay. Since the EBT was based on background levels instead of toxicity data, it should be considered an indicator for overall chemical stress and not directly as a measure of low micropollutant risk.

3.2.4. AhR activation in the PAH CALUX

Unspiked feed water to the pilot treatment plants A, B, D, and E induced responses in the PAH CALUX. The presence of PAH-like activity in surface water samples was not surprising, as the AhR can be activated by a variety of chemicals, including many PAHs, polychlorinated biphenyls (PCBs), dioxins, and many naturally occurring chemicals such as flavonoids (Barouki *et al.* 2012). The spike mixture apparently did not induce additional response in the PAH CALUX as in most samples (8 out of 10) the activity in the spiked feed water was not higher than in the unspiked water, which was in agreement with the calculated mixture toxicity. In addition, the spiking mixture did not contain PAHs.

In water samples collected from pilot treatment plant C, moderate responses were induced after treatment with UV/H_2O_2. After additional treatment with BACF, the compounds inducing these responses were completely removed, while this was not the case in the other plants. The PAH-like activity was reduced when treating the water with O_3 at plants A (O_3/H_2O_2) and D (O_3). An increase in activity was seen for one of the duplicates of RO (plant B).

After activated carbon treatments (BAC1 at plant D) and reactivated BACF (plant E), a reduction of activity was observed. However, slight increases in activity were seen after activated carbon treatment (BAC2 at plant D) and saturated BACF1 and 2 (plant E). One of the duplicates of the feed water at plant E was close to the environmental EBT, as derived by van der Oost *et al.* (2017), but none of the samples exceeded this value. Half of the treated waters exceeded the human health EBT.

3.2.5. Xenobiotic response in PXR CALUX and PPARγ CALUX

The PXR CALUX and PPARγ CALUX were only applied in pilot treatment plant D. In both tests, no responses were induced by unspiked feed water and spiked feed water. The calculation of mixture toxicity indicated an activity only just above (18%) the LOQ. Therefore, it was no surprise that no activity was found in practice. This may, e.g., be explained by a recovery of the contributing spiking during sample preparation slightly below 100%, which is quite common for sample preparation methods that have to cover a very broad range of chemicals, as in this study. Alternatively, ToxCast analog may not be representative of the CALUX assays included in this study. After treatment with ozone, high responses in the PXR CALUX were induced, indicating the formation of transformation products that activate the PXR. These bioactive compounds were removed after two additional treatment steps with BAC. In the PPARγ CALUX, small and unreplicated responses were observed after the additional treatment steps. After treatment with O_3 or activated carbon, very low activity in the PPARγ CALUX was observed in one of the independent duplicates. Because these activities were only just above the LOQ and not consistent in the duplicate experiments, they might be just a consequence of experimental error. The possible exceedance of the environmental EBT by the feed waters could not be evaluated, as the EBT was lower than the LOQ.

3.3. Ames fluctuation test results (mutagenic response)

TA98 and TA100 strains were used in the Ames fluctuation test, in the absence or presence of an enzymatic S9 mixture to model bioactivation. An overview of these results is presented in Figure 2, the underlying data in Table S7. Overall, most responses were observed in the TA98 strain without metabolic activation. In none of the samples, metabolic activation induced additional responses in the TA98 strain (in comparison to the response in the absence of S9). In the TA100 strain, responses after metabolic activation were increased only in a few cases (in comparison with those observed in the absence of S9).

Unspiked and spiked feed water evoked replicated positive responses in one or more of these test conditions for water in pilot treatment plants A, B, C, and E. The presence of mutagenic activity in surface waters has been observed before and might partly originate from natural compounds (Kool & Van Kreyl 1988). The outcome of mutagenic activity during drinking water treatment is considered a relevant endpoint for monitoring by drinking water companies. This holds especially if oxidative processes are applied due to the previous finding that under certain process conditions UV/H_2O_2 treatment led to the formation of byproducts with mutagenic activity that, importantly, were removed by subsequent filtration over activated carbon (Hofman-Caris *et al.* 2015). In our study, this finding was confirmed: an increase in mutagenic activity was observed by UV/H_2O_2 treatment (plant C; TA98 with and without S9) and by O_3 (plant D; TA98 without S9), although the latter may be a false-positive response related to induced cell growth as observed in the experiment. Removal of mutagenic activity by activated carbon filtration was seen at all pilot treatment plants with activated carbon (plant C (TA 98 without S9), plant D (TA98 without S9), and plant E (all four test conditions)).

A mutagenic response in the Ames test does not necessarily indicate a human or ecological health risk. This would depend on the responsible compounds and if they also would be mutagenic for target organisms in the aquatic environment after exposure to current concentrations. For drinking water safety, it is important that the activated carbon filtration following oxidative processes shows complete removal of mutagenicity. As the participating drinking water companies all apply subsequent BACF filtration in their full-scale plants (including plant A, although this company did not sample an 'after BACF' sample in its pilot experiment), mutagenic activity is not expected in their final drinking water and as such is not expected to pose a health risk.

To estimate the contribution of individual spiked chemicals to the responses observed in the Ames test, a high-level inventory of data on the mutagenic, genotoxic, and carcinogenic potential of the spiked chemicals was collected from GENE-TOX (PubChem) and ToxNet and safety data sheets. For about a third of the spiked chemicals, indications for mutagenic activity were found in any or more of these data sources (data not included). However, as it is only reported whether the substance has mutagenic potential, and without any information on potency (differences), it was not possible to estimate the contribution of the spiked chemicals.

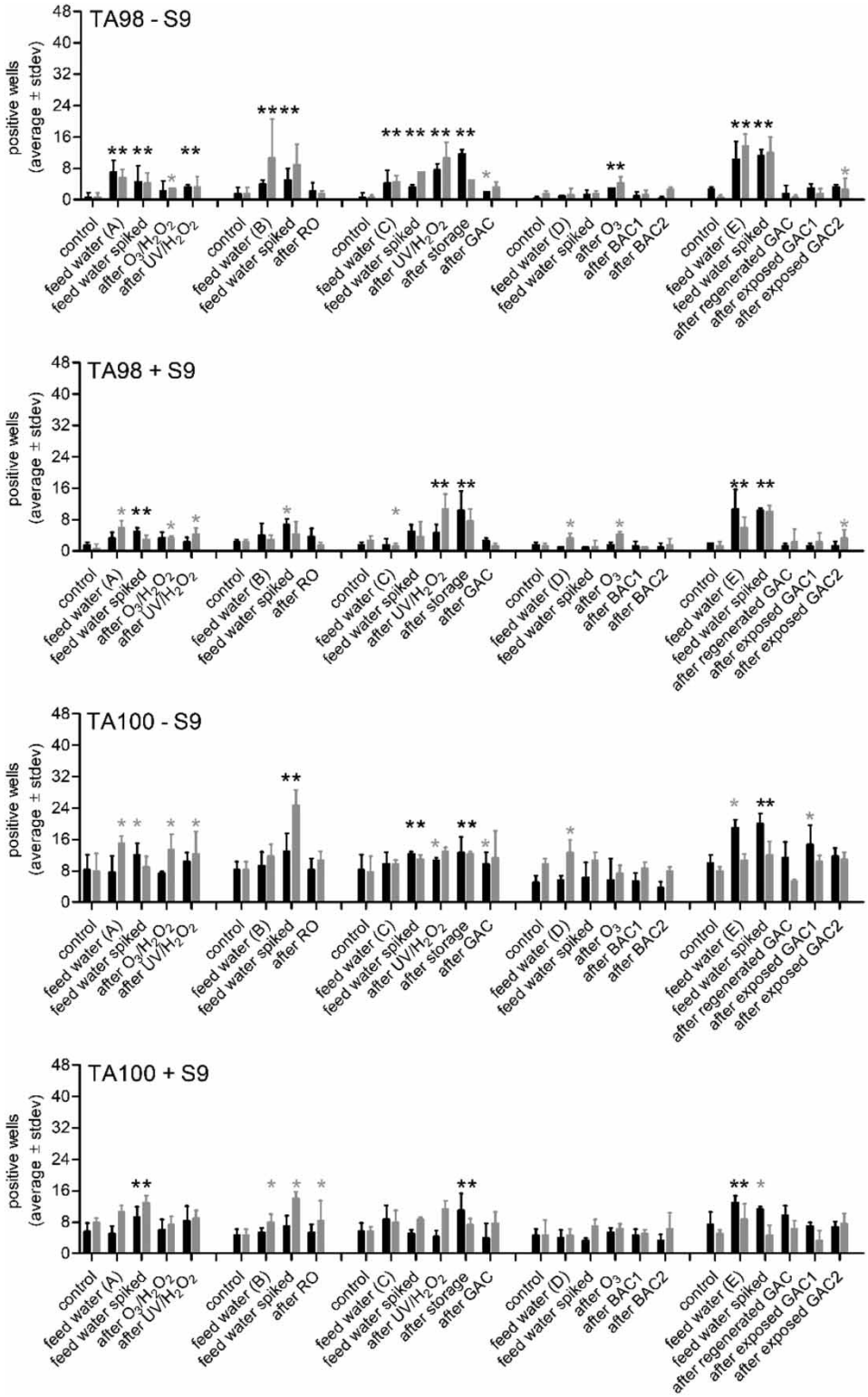


Figure 2 | Ames fluctuation test results from pilot treatment plants (a)–(e) in TA98 ± S9 and TA100 ± S9. Bars represent the average of the triplo assay, results of duplicate samples were side-by-side. Gray asterisks indicate an unreplicated statistically significant effect (deviation from Evian control). Double black asterisks indicate a replicated response in both independent duplicate samples. For the observed positive effects in pilot treatment plant D, it cannot be excluded that this response is associated to increased cell growth (false positive).

3.4. General discussion

3.4.1. Biological activity of the spiked mixture

The performance of various pilot-scale water treatment trains using spiked feed water was evaluated using a panel of bioassays. The concentrations of the spike chemicals that emerge in Dutch surface water were chosen to ascertain the detectability of the chemicals and quantification of their removal percentage up to 2 log units, and not to ensure measurable responses in the applied bioassays. Within the planning of the project, it was not feasible to determine the biological activity of the spiking mixture in the bioassays. Thanks to the public availability of the potency data in ToxCast, it was possible to generate a first, albeit incomplete, idea of any bioassay activity that should be expected from the spiking. The actual influence of spiking on the bioassay activity during the pilot experiments was evaluated by comparing measured activities in unspiked and spiked feed water. No activity of the spiking mixture (i.e., higher activity in spiked than in unspiked feed water) was observed in the bioassays. Nevertheless, the outcome of activity in the anti-AR, PAH, and Nrf2 CALUX during the various treatments could be evaluated as these activities were already detectable in (most) unspiked feed waters. For the remaining assays (ERα, PPARγ, and PXR), showing activity neither in the unspiked nor in the spiked feed water, it was not possible to determine the capacity of the tested treatments to remove the substances that cause these activities. The results were still valuable, as they could serve to monitor the possible introduction of activity during treatment steps as was seen for estrogenic activity by saturated activated carbon.

Bioassays specifically designed to detect specific mechanisms of action are inherently sensitive. Therefore, not every positive response is by definition associated with potential risks to human health or the environment. To assess whether a response may be related to potential risks, the effect level at which potential risks cannot be excluded was defined using EBT values (van der Oost *et al.* 2017; Neale *et al.* 2022). Different methods are published and EBT values may differ for potential risks to human health, on the one hand, and the environment, on the other. In the present study, potential risks were determined based on (spiked and unspiked) feed water in pilot treatment plants, rather than full-scale installations used for (drinking) water production. Therefore, EBT values were only used for interpretation of the data and may not reflect realistic risks posed to human health due to actual exposure to contaminants in the unspiked water. The absence of activity for some endpoints in the feed waters underlines that trying to mutually adjust the spiking mixture and the selection of bioassays can help optimize the experimental setup to obtain the maximum information from the pilot experiments, as far as this is possible within the practical restrictions set by the scale of the pilot treatment plants and chemical techniques. This can be done by including in the spike mixture at least some active compounds in active concentrations for each of the selected endpoints and as such ascertain the presence of activity in the spiked feed water and the possibility to evaluate its removal. Another possibility is the other way around: to select bioassays according to the endpoints for which the chosen spiking compounds were active. To this aim, the evaluation of all endpoints in ToxCast as performed here is a good approach. In the next edition of the 'Robustness of treatments' program for which preparations have started recently, the spiking mixture is tested in candidate bioassays after which the definitive selection of bioassays is made. In multiple CALUX bioassay experiments, biological activity was detected in the unspiked samples, which is likely due to unknown compounds. To perform a proper risk assessment, identification of the chemicals causing the activity is needed. To this aim, a dedicated analysis of possible correlations between the outcome of activity and that of specific NTS features (a set of elements of substances) can serve as a first step.

3.4.2. Capability of treatments to remove bioactive compounds

The *advanced oxidative processes* O_3/H_2O_2–UV/H_2O_2, O_3, and UV/H_2O_2 showed good removal of anti-androgenic activity. This was in line with the high removal of OMPs by O_3/H_2O_2–UV/H_2O_2 published by Timmers *et al.* (2022) (spiked compounds with this approach (90%)). Oxidative processes generally do not completely mineralize chemicals but break them down to transformation products. These transformation products depend on the chemicals in the feed water and on process conditions (e.g., lamp intensity, oxidant concentrations) and may have their own biological activities. This may explain that PAH-like activity, unlike anti-androgenic activity, increased by both O_3/H_2O_2–UV/H_2O_2 and UV/H_2O_2, as well as mutagenic activity by O_3/H_2O_2 treatment. For oxidative stress, varying results were obtained, which contrasts with the finding of another study by Neale *et al.* (2020), who reported a decrease in oxidative stress response after the implementation of UV treatment. Treated waters in the study by Neale *et al.* (2020) had low effects compared to EBT values for activation of estrogenic activity and oxidative stress. While equal results were obtained for estrogenic activity in the present study, this was not the case for oxidative stress, as the obtained responses sometimes exceeded the EBTs. However, in the study by Neale *et al.* (2020), it

remains unclear which of the treatment processes contributed most to the elimination of the oxidative stress response. Byproducts may have been formed during oxidative processes, including UV/H_2O_2 and O_3/H_2O_2, which may be active in the Nrf2 CALUX, the PAH CALUX, and the PXR CALUX (in plant D). This was also suggested in the studies by Oskarsson *et al.* (2021) and Kienle *et al.* (2022). Indeed, NTS confirmed the formation of unknown transformation products during the oxidative processes in the spiking experiments in the pilot installations described here (Brunner *et al.* 2020).

RO removed anti-androgenic activity successfully. Other activities were (almost) not present in the (un)spiked feed water, and as such their removal could not be judged. RO (based on filtration) is a powerful technique for the removal of micropollutants. This was confirmed by the parallel observed high removal of spiked compounds and >95% of NTS features decreasing in intensity with this technique (Brunner *et al.* 2020).

Activated carbon (BAC and GAC) showed good removal of anti-androgenic and PAH-like activity and oxidative stress. This was in line with the good removal rates of the spike compounds found for this treatment (Brunner *et al.* 2020). Saturated carbon was still capable of removing the anti-androgenic activity. PAH-like activity and oxidative stress, on the contrary, were not removed. These results are in accordance with the findings of Oskarsson *et al.* (2021) and Kienle *et al.* (2022), which report the effective removal of AhR-inducing (PAH-like) and anti-androgenic chemicals in treatment plants, including those using (unsaturated) GAC. During use, activated carbon is loaded with dissolved organic carbon as well as with OMPs that occupy the active sites at the GAC surface, turning the carbon into saturated carbon over time. The thermal reactivation of saturated GAC is a costly process and is environmentally unfriendly due to its high CO_2 footprint (Sierka 2011). This demonstrates that the reactivation interval of carbon must be balanced carefully between performance (water quality improvement with respect to the removal of OMPs triggering a response in the bioassays) and costs and emissions.

Quantitative target analysis based on partially the same pilot setups published earlier showed a very good removal of spiked chemicals in the drinking water treatment trains (Brunner *et al.* 2020). Removal rates were quantified to be (averaged over all spiked chemicals) 95–98%, depending on the pilot treatment plant. The NTS provided, although only in a semiquantitative manner, together with the effect-based monitoring using bioassays, a more complete assessment of treatment performance. This previous research demonstrated that, although feature intensities decreased due to treatments (2–5-fold), this was much less than the percentages found by the target analysis of the spike compounds. In addition, also an increase in feature intensities and the introduction of new features were observed (Brunner *et al.* 2020), suggesting that (part of the) compounds already present in the unspiked feed water were more persistent than spiked chemicals and/or transformed into transformation products.

The applied combination of techniques created a more integrated picture of the chemical water quality during drinking water treatment than any individual technique could achieve. The bioassays proved their added value in measuring the removal (or sometimes introduction) of bioactive chemicals by the water treatments. They reflected the fate of the biological activities of the total mixtures in the effluent. In addition, the results confirmed that in some cases, bioactive substances (byproducts and transformation products) may be formed during treatment processes (e.g., ozonation), which may lead to increased responses in PAH CALUX and/or Nrf2 CALUX. Although broadening the range of micropollutants, the applied approach still does not provide a complete view of all OMPs present in water. First, due to the need for extraction for the bioassays, most very polar micropollutants were probably lost during sample preparation. Current research is therefore ongoing to model the purification of polar compounds. Second, the panel of bioassays applied now does not cover all endpoints that might be relevant for ecosystems or human health. This was shown by the fact that data mined from ToxCast suggested that some of the spiked micropollutants were active in other bioassays. Extension of the panel with bioassays for other endpoints may therefore also help obtain a more complete picture of the chemical loads and water quality in terms of human toxicology during water treatments.

4. CONCLUSIONS

In the present study, we demonstrated that OMPs impact bioassay responses and, therefore, water quality. Furthermore, the study showed clear differences between different water treatments in their ability to remove specific biologically active substances, sometimes introducing bioactive compounds. Nevertheless, most pilot treatment plants did not show any detectable bioassay activities in their final treated waters, thanks to the combination of complementary treatment processes ('multibarrier approaches') to remove as many of the micropollutants as possible. The activities most often observed in the final treated waters from the pilot plants were oxidative stress (plants A and E) and PAH-like activity (plants A, B, D, and E). In these

plants, bioactivities were already present in a similar or higher range in the unspiked and spiked feed water. However, in plant C, treatment with UV/H_2O_2 increased oxidative stress and PAH-like activity, indicating the formation of bioactive compounds during advanced oxidation. In plant D, treatment with ozone resulted in high responses in the PXR CALUX. In both plants C and D, the formed bioactive compounds were removed after the additional treatment steps with GAC or BAC, respectively. Other activities were generally removed by ozonation and/or sequential activated carbon treatment.

Bioassays in this study proved their added value in measuring the occurrence of bioactive chemicals. The applied combination of techniques created a more integrated picture of the chemical water quality during drinking water treatment than any individual technique could achieve. It appeared that most of the measured bioactivity was not due to the spiked chemicals. It is therefore recommended to investigate the nature of and conditions that lead to the presence of activity-inducing compounds in drinking water sources and specific treatment steps. This requires advanced analytical approaches for the identification of active compounds as many compounds can cause oxidative stress and AhR activation, and only a small fraction of the response in a water sample can be typically explained by known chemicals.

FUNDING

This research was funded by the DPWE Joint Research Program (project: 402124) of the Dutch water companies Dunea, PWN, Waternet, and Evides.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

ACKNOWLEDGEMENTS

Merijn Schriks, Kirsten Baken, Wolter Siegers, and Bas Wols are kindly acknowledged for their contributions to the experimental setup. Astrid Reus, Roberta Hofman-Caris, and Sanah Shaikh are kindly acknowledged for their support.

REFERENCES

Aït-Aïssa, S., Laskowski, S., Laville, N., Porcher, J.-M. & Brion, F. (2010) Anti-androgenic activities of environmental pesticides in the MDA-kb2 reporter cell line, *Toxicology In Vitro*, **24** (7), 1979–1985.

Alygizakis, N. A., Besselink, H., Paulus, G. K., Oswald, P., Hornstra, L. M., Oswaldova, M., Medema, G., Thomaidis, N. S., Behnisch, P. A. & Slobodnik, J. (2019) Characterization of wastewater effluents in the Danube river basin with chemical screening, in vitro bioassays and antibiotic resistant genes analysis, *Environment International*, **127**, 420–429.

Barouki, R., Aggerbeck, M., Aggerbeck, L. & Coumoul, X. (2012) The aryl hydrocarbon receptor system. *Drug metabolism and drug interactions*, **27** (1), 3–8. https://doi.org/10.1515/dmdi-2011-0035.

Beal, M. A., Audebert, M., Barton-Maclaren, T., Battaion, H., Bemis, J. C., Cao, X., Chen, C., Dertinger, S. D., Froetschl, R. & Guo, X. (2023) Quantitative in vitro to in vivo extrapolation of genotoxicity data provides protective estimates of in vivo dose, *Environmental and Molecular Mutagenesis*, **64** (2), 105–122.

Been, F., Pronk, T., Louisse, J., Houtman, C., van der Velden-Slootweg, T., van der Oost, R. & Dingemans, M. M. (2021) Development of a Framework to Derive Effect-Based Trigger Values to Interpret CALUX Data for Drinking Water Quality, *Water Research*, **193**, 116859.

Bernhardt, E. S., Rosi, E. J. & Gessner, M. O. (2017) Synthetic chemicals as agents of global change, *Frontiers in Ecology and the Environment*, **15** (2), 84–90.

Binetti, R., Costamagna, F. M. & Marcello, I. (2008) Exponential growth of new chemicals and evolution of information relevant to risk control, *Annali-Istituto Superiore di Sanità*, **44** (1), 13.

Brand, W., de Jongh, C. M., van der Linden, S. C., Mennes, W., Puijker, L. M., van Leeuwen, C. J., van Wezel, A. P., Schriks, M. & Heringa, M. B. (2013) Trigger values for investigation of hormonal activity in drinking water and its sources using CALUX bioassays, *Environ. Int.*, **55**, 109–118.

Brunner, A. M., Bertelkamp, C., Dingemans, M. M., Kolkman, A., Wols, B., Harmsen, D., Siegers, W., Martijn, B. J., Oorthuizen, W. A. & Ter Laak, T. L. (2020) Integration of target analyses, non-target screening and effect-based monitoring to assess OMP related water quality changes in drinking water treatment, *Science of The Total Environment*, **705**, 135779.

Christen, V., Crettaz, P. & Fent, K. (2014) Additive and synergistic antiandrogenic activities of mixtures of azol fungicides and vinclozolin, *Toxicology and Applied Pharmacology*, **279** (3), 455–466.

Ebrahimzadeh, S., Wols, B., Azzellino, A., Martijn, B. J. & van der Hoek, J. P. (2021) Quantification and modelling of organic micropollutant removal by reverse osmosis (RO) drinking water treatment, *Journal of Water Process Engineering*, **42**, 102164.

Escher, B. I., Neale, P. A. & Leusch, F. D. (2015) Effect-based trigger values for in vitro bioassays: Reading across from existing water quality guideline values, *Water Res.*, **81**, 137–148.

Fischer, A., Ter Laak, T., Bronders, J., Desmet, N., Christoffels, E., van Wezel, A. & van der Hoek, J. P. (2017) Decision support for water quality management of contaminants of emerging concern, *Journal of Environmental Management*, **193**, 360–372.

Gijsbers, L., Man, H.-Y., Kloet, S. K., de Haan, L. H., Keijer, J., Rietjens, I. M., van der Burg, B. & Aarts, J. M. (2011) Stable reporter cell lines for peroxisome proliferator-activated receptor γ (PPARγ)-mediated modulation of gene expression, *Analytical Biochemistry*, **414** (1), 77–83.

Heringa, M. B., Harmsen, D. J., Beerendonk, E. F., Reus, A. A., Krul, C. A., Metz, D. H. & Ijpelaar, G. F. (2011) Formation and removal of genotoxic activity during UV/H(2)O(2)-GAC treatment of drinking water, *Water Research*, **45** (1), 366–374.

Hofman-Caris, R. C., Harmsen, D. J., Puijker, L., Baken, K. A., Wols, B. A., Beerendonk, E. & Keltjens, L. L. (2015) Influence of process conditions and water quality on the formation of mutagenic byproducts in UV/H_2O_2 processes, *Water Research*, **74**, 191–202.

Houtman, C. J., Van Oostveen, A. M., Brouwer, A., Lamoree, M. H. & Legler, J. (2004) Identification of estrogenic compounds in fish bile using bioassay-directed fractionation, *Environmental Science & Technology*, **38** (23), 6415–6423.

Houtman, C. J., ten Broek, R. & Brouwer, A. (2018) Steroid hormonal bioactivities, culprit natural and synthetic hormones and other emerging contaminants in waste water measured using bioassays and UPLC-TQ-MS, *Science of The Total Environment*, **630**, 1492–1501.

Houtman, C. J., ten Broek, R., van Oorschot, Y., Kloes, D., van der Oost, R., Rosielle, M. & Lamoree, M. H. (2020) High resolution effect-directed analysis of steroid hormone (ant)agonists in surface and wastewater quality monitoring, *Environmental Toxicology and Pharmacology*, **80**, 103460.

Houtman, C. J., Brewster, K., ten Broek, R., Duijve, B., van Oorschot, Y., Rosielle, M., Lamoree, M. H. & Steen, R. J. C. A. (2021) Characterisation of (anti-)progestogenic and (anti-)androgenic activities in surface and wastewater using high resolution effect-directed analysis, *Environment International*, **153**, 106536.

Kienle, C., Werner, I., Fischer, S., Lüthi, C., Schifferli, A., Besselink, H., Langer, M., McArdell, C. S. & Vermeirssen, E. L. M. (2022) Evaluation of a full-scale wastewater treatment plant with ozonation and different post-treatments using a broad range of in vitro and in vivo bioassays, *Water Research*, **212**, 118084.

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. (2013) Sample preparation for combined chemical analysis and in vitro bioassay application in water quality assessment, *Environmental Toxicology and Pharmacology*, **36** (3), 1291–1303.

Kool, H. & Van Kreyl, C. (1988) Mutagenic activity in drinking water prepared from groundwater: A survey of ten cities in the Netherlands, *Science of The Total Environment*, **77** (1), 51–60.

Liscio, C., Abdul-Sada, A., Al-Salhi, R., Ramsey, M. H. & Hill, E. M. (2014) Methodology for profiling anti-androgen mixtures in river water using multiple passive samplers and bioassay-directed analyses, *Water Research*, **57**, 258–269.

Loewe, S. & Muischnek, H. (1926) Über Kombinationswirkungen. 1. Mitteilung: Hilfsmittel der Fragestellung. Naunyn-Schmiedelbergs, *Archiv fur Experimentelle Pathologie und Pharmakologie*, **114**, 313–326.

Mons, M. N., Heringa, M. B., van Genderen, J., Puijker, L. M., Brand, W., van Leeuwen, C. J., Stoks, P., van der Hoek, J. P. & van der Kooij, D. (2013) Use of the threshold of toxicological concern (TTC) approach for deriving target values for drinking water contaminants, *Water Research*, **47** (4), 1666–1678.

Neale, P. A., Feliers, C., Glauch, L., König, M., Lecarpentier, C., Schlichting, R., Thibert, S. & Escher, B. I. (2020) Application of in vitro bioassays for water quality monitoring in three drinking water treatment plants using different treatment processes including biological treatment, nanofiltration and ozonation coupled with disinfection, *Environmental Science: Water Research & Technology*, **6** (9), 2444–2453.

Neale, P. A., Escher, B. I., de Baat, M. L., Enault, J. & Leusch, F. D. L. (2023) Effect-Based Trigger Values Are Essential for the Uptake of Effect-Based Methods in Water Safety Planning, Environmental toxicology and chemistry, 42 (3), 714–726. https://doi.org/10.1002/etc.5544.

Oskarsson, A., Rosenmai, A. K., Mandava, G., Johannisson, A., Holmes, A., Tröger, R. & Lundqvist, J. (2021) Assessment of source and treated water quality in seven drinking water treatment plants by in vitro bioassays – Oxidative stress and antiandrogenic effects after artificial infiltration, *Science of The Total Environment*, **758**, 144001.

Pieterse, B., Felzel, E., Winter, R., Van Der Burg, B. & Brouwer, A. (2013) *PAH-CALUX, An Optimized Bioassay for AhR-Mediated Hazard Identification of Polycyclic Aromatic Hydrocarbons (PAHs) as Individual Compounds and in Complex Mixtures*. American Chemical Society, pp. 11651–11659.

Post, G. B. (2021) Recent US state and federal drinking water guidelines for per-and polyfluoroalkyl substances, *Environmental Toxicology and Chemistry*, **40** (3), 550–563.

Pronk, T. E., Vries, D., Kools, S. A. E., Hofman-Caris, R. & Stroomberg, G. J. (2020) *Removal Requirement and Purification Treatment Effort for Dutch Rhine Water From 2000–2018 RIWA – Association of River Waterworks Rhine*. RIWA-Rijn report, Nieuwegein, The Netherlands.

Reifferscheid, G., Buchinger, S., Cao, Z. & Claus, E. (2011) Identification of mutagens in freshwater sediments by the Ames – Fluctuation assay using nitroreductase and acetyltransferase overproducing test strains, *Environmental and Molecular Mutagenesis*, **52** (5), 397–408.

Sierka, R. A. (2011) An Energy Saving Process for the Reactivation of Activated Carbon Saturated with Organic Contaminants. In: Atimtay, A., Sikdar, S. (eds) *Security of Industrial Water Supply and Management*. NATO Science for Peace and Security Series C: Environmental Security. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-1805-0_13.

Slootweg, T. & Houtman, C. J. (2014) *Toxicologische Evaluatie van Organische Microverontreinigingen in de Rijn bij Lobith Aangetroffen in 2010 en 2011*. The Netherlands: Amsterdam, pp. 1–52.

Slootweg, T., Van Genderen-de Kloe, G., Rousseau, B., Oomen, T. & Bannink, A. (2022) *Drinking Water Relevant Substances in the Meuse, RIWA – Vereniging van Rivierwaterbedrijven Meuse*. RIWA – Maas report.

Stroomberg, G. J., Neefjes, R. E. M., De Jonge, J. A., Bannink, A. D., Van de Haar, G. & Zwamborn, C. C. (2018) *Jaarrapport 2017 De Rijn, RIWA-Rijn, Vereniging van Rivierwaterbedrijven*. RIWA-Rijn report, Nieuwegein, The Netherlands.

Timmers, P. H. A., Slootweg, T., Knezev, A., van der Schans, M., Zandvliet, L., Reus, A., Vughs, D., Heijnen, L., Knol, T., El Majjaoui, J., van de Wielen, P., Stuyfzand, P. J. & Lekkerkerker-Teunissen, K. (2022) Improved drinking water quality after adding advanced oxidation for organic micropollutant removal to pretreatment of river water undergoing dune infiltration near The Hague, Netherlands, *Journal of Hazardous Materials*, **429**, 128346.

Tosun, J., Schaub, S. & Fleig, A. (2020) What determines regulatory preferences? Insights from micropollutants in surface waters, *Environmental Science & Policy*, **106**, 136–144.

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, *Environmental Science & Technology*, **42** (15), 5814–5820.

van der Linden, S. C., von Bergh, A. R., van Vught-Lussenburg, B. M., Jonker, L. R., Teunis, M., Krul, C. A. & van der Burg, B. (2014) Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, **760**, 23–32.

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

First received 12 April 2024; accepted in revised form 24 October 2024. Available online 5 November 2024