Environmental Toxicology and Chemistry—Volume 40, Number 2—pp. 487–499, 2021
Received: 8 October 2020 | Revised: 4 November 2020 | Accepted: 23 November 2020

Environmental Toxicology

Effect-Based Trigger Values for Mixtures of Chemicals in Surface Water Detected with In Vitro Bioassays

Beate I. Escher[a,b,*] and Peta A. Neale[c]

[a]Helmholtz Centre for Environmental Research–UFZ, Leipzig, Germany
[b]Center for Applied Geoscience, Eberhard Karls University of Tübingen, Tübingen, Germany
[c]Australian Rivers Institute, School of Environment and Science, Griffith University, Southport, Queensland, Australia

Abstract: Effect-based trigger (EBT) values for in vitro bioassays are important for surface water quality monitoring because they define the threshold between acceptable and poor water quality. They have been derived for highly specific bioassays, such as hormone-receptor activation in reporter gene bioassays, by reading across from existing chemical guideline values. This read-across method is not easily applicable to bioassays indicative of adaptive stress responses, which are triggered by many different chemicals, and activation of nuclear receptors for xenobiotic metabolism, to which many chemicals bind with rather low specificity. We propose an alternative approach to define the EBT from the distribution of specificity ratios of all active chemicals. The specificity ratio is the ratio between the predicted baseline toxicity of a chemical in a given bioassay and its measured specific endpoint. Unlike many previous read-across methods to derive EBTs, the proposed method accounts for mixture effects and includes all chemicals, not only high-potency chemicals. The EBTs were derived from a cytotoxicity EBT that was defined as equivalent to 1% of cytotoxicity in a native surface water sample. The cytotoxicity EBT was scaled by the median of the log-normal distribution of specificity ratios to derive the EBT for effects specific for each bioassay. We illustrate the new approach using the example of the AREc32 assay, indicative of the oxidative stress response, and 2 nuclear receptor assays targeting the peroxisome proliferator–activated receptor gamma and the arylhydrocarbon receptor. The EBTs were less conservative than previously proposed but were able to differentiate untreated and insufficiently treated wastewater from wastewater treatment plant effluent with secondary or tertiary treatment and surface water. *Environ Toxicol Chem* 2021;40:487–499. © 2020 The Authors. *Environmental Toxicology and Chemistry* published by Wiley Periodicals LLC on behalf of SETAC.

Keywords: Cytotoxicity; Specific mode of action; Environmental quality standard; Water quality; Reporter gene assay; Water pollution

INTRODUCTION

Many in vitro bioassays, particularly mammalian reporter gene assays, are increasingly sensitive and can detect effects in relatively clean waters, such as drinking water and recycled water, after enrichment with solid-phase extraction (SPE; e.g., Jia et al. 2015; Conley et al. 2017; Neale et al. 2020b). However, just because an effect is detected does not necessarily mean that the water is toxic and the chemical water quality is not acceptable. As a result, effect-based trigger (EBT) values have been proposed to help differentiate between an acceptable and an unacceptable bioassay response in a water sample. The current approach used for chemical water quality monitoring is to compare detected chemical concentrations from targeted chemical analysis to chemical guideline values. However, chemical guidelines cannot possibly capture all chemicals potentially present in water and do not account for the mixture effects that can occur between the many chemicals present. This emphasizes the need for effect-based monitoring and consequently the definition of EBTs.

A number of different approaches have been applied to derive EBTs for both drinking water and surface water. The simplest approach involves directly translating a chemical guideline value to a bioanalytical equivalent concentration (BEQ) using the bioassay's reference compound (Leusch et al. 2014; Kunz et al. 2015). This implies that the reference

This article contains online-only Supplemental Data.

* Address correspondence to beate.escher@ufz.de
Published online 30 November 2020 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/etc.4944

compound is representative of all chemicals inducing a specific effect in the assay. This is not necessarily the case, and therefore, a number of studies have determined the in vitro effect at the guideline concentration using the different potencies of the bioactive chemicals and accounted for mixture effects using a read-across approach (Escher et al. 2013, 2015, 2018b). An additional approach specific for drinking water has been proposed that converts the concentrations of reference compounds considered safe in vivo to concentrations that can be detected using in vitro assays using differences in the toxicokinetics of different compounds to correct the EBT (Brand et al. 2013). Brion et al. (2019) derived EBTs for estrogenicity by comparing the effects detected in vitro to effects detected in vivo. In addition, other studies have derived EBTs for surface water using a combination of approaches, including converting from in vivo toxicity data and field investigations, often following multiple lines of evidence in a more qualitative manner (van der Oost et al. 2017; de Baat et al. 2020), or using distributions of a large set of water samples (Besselink et al. 2017).

Previous methods to derive EBTs have relied on a very limited number of experimental effect data of single chemicals. This was successful for bioassays indicative of highly specific modes of action such as estrogenicity or inhibition of photosynthesis, where a small number of highly active chemicals dominate the mixture effects in water samples. In contrast, this method has limitations for cytotoxicity and bioassays indicative of modes of action where many chemicals are active but with lower potency, such as oxidative stress response mediated by the nuclear erythroid 2-related factor 2 (Nrf2)–antioxidant response element (ARE) pathway and activation of the peroxisome proliferator–activated receptor gamma (PPARγ) or the arylhydrocarbon receptor (AhR). Direct read-across from guideline values was not possible for low-potency chemicals, and a mixture factor had to be invoked to derive the EBT for bioassays that respond to many but low-potency chemicals (Escher et al. 2018b).

However, these biological endpoints are very important: wide screening of water samples with multiplexed assays that included more than 60 nuclear receptors and transcription factors (Escher et al. 2014; Blackwell et al. 2019) have indicated that Nrf2, AhR, the pregnane X receptor (PXR), PPARγ, and some hormone receptors are activated most prominently by many types of water samples.

We propose an alternative approach that defines safe levels for overall cytotoxicity as the point of departure. We illustrate the new method using 3 reporter gene assays often used for water quality assessment: AREc32 for the activation of oxidative stress response, PPARγ-BLA for binding to PPARγ, and AhR-CALUX for activation of AhR. From a general EBT 10% inhibitory concentration (IC10) for cytotoxicity, we derived the EBT for specific effects in reporter gene assays by analyzing the degree of specificity (specificity ratio) and accounted for mixtures by looking at the distribution of specificity ratios of large numbers of chemicals for each assay. We included effect data of 689 environmentally relevant chemicals that have been detected in European surface waters and had previously been included in water quality monitoring studies. We used experimental effect data for single chemicals from the literature but also mined the Tox21 database for developing the EBTs and for estimating what fraction of chemicals is expected to be active in a whole-water sample. Finally, we applied the new mixture EBT to several case studies on wastewater and surface water and evaluated if the EBT can differentiate between these water types and how they compare to previously proposed EBTs for these bioassays.

THEORY: DERIVATION OF MIXTURE EBTs FOR CYTOTOXICITY AND MODES OF ACTION WITH LOW SPECIFICITY

What is a safe level of cytotoxicity in cell-based bioassays?

Baseline toxicants, which act nonspecifically according to narcosis (McCarty et al. 2013), have very similar internal effect concentrations, more specifically critical membrane concentrations, with 69 mmol kg_{lip}^{-1} (95% CI 49–89 mmol kg_{lip}^{-1}) for 10% cytotoxicity derived for 7 different reporter gene cell lines that were based on 5 different cell types including MCF7, HepG2, H4IIe, HEK293H, and HEK293T (Escher et al. 2019). Because critical membrane concentrations were uniform across cell types, the quantitative structure–activity relationships (QSARs) for baseline toxicity were very similar for diverse cell lines (Escher et al. 2019). Figure 1A depicts the empirical baseline toxicity QSARs for the 3 cell lines investigated. Small differences between the empirical baseline toxicity QSARs of different cell lines stemmed from experimental variability and the setup of the assay, mainly differences in medium composition, leading to small differences in the bioavailability of the chemicals (Escher et al. 2019). Given the similarity in baseline toxicity between cell lines, we can define one EBT for cytotoxicity for all cell lines.

Similar to how safe effect levels are derived for aquatic organisms, we first have to decide on an effect level that is considered safe for cells and then translate this effect level to an EBT-IC10 for cytotoxicity. Cytotoxicity of 1% cannot be differentiated from the control and can be considered a safe level of effect. For aquatic organisms, the lower 5th percentile of species sensitivity distributions of no-observed-effect concentrations (NOECs) is typically used to derive environmental quality standards for single chemicals (Posthuma et al. 2002). Considering the variability of controls, effect level variability of NOECs, and sensitivity differences between aquatic organisms and cells, it seems reasonable to accept 1% cytotoxicity as a safe level.

How the acceptable cytotoxicity level of 1% is translated into an IC10 for cytotoxicity is illustrated in Figure 1B. At low effect levels (<30%) the concentration–response curves are typically linear (Escher et al. 2018a), so we can directly interpolate in this range. If we would accept 10% of cytotoxicity in the original water sample (green dotted line in Figure 1B), then the acceptable IC10 for cytotoxicity (concentration that causes 10% reduction of cell viability and growth) would be a relative enrichment factor (REF) of 1. The REF takes into consideration sample enrichment and dilution in the assay, with an REF of 1 indicating the unenriched sample. If we accept 1% of

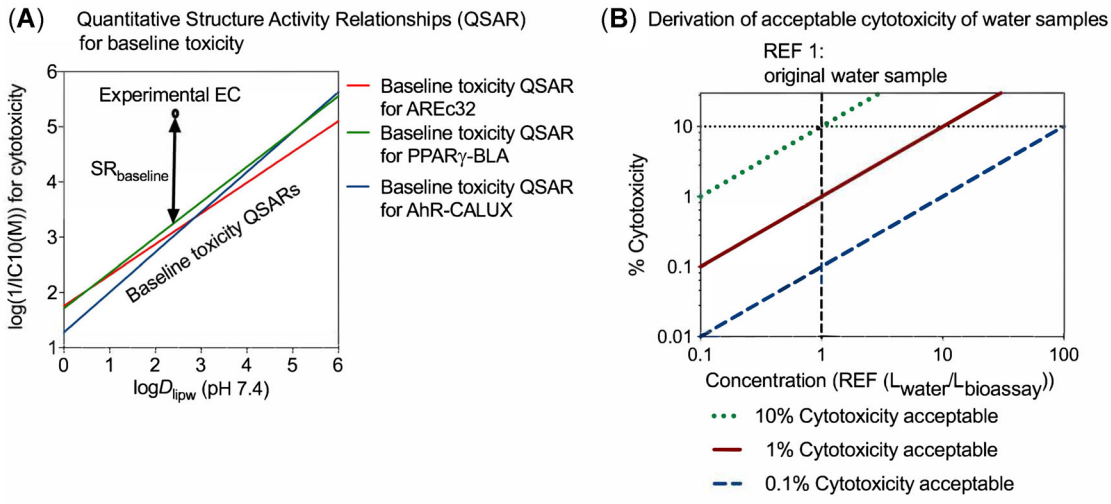

FIGURE 1: (**A**) Quantitative structure–activity relationships for baseline toxicity (Escher et al. 2019) of the 3 reporter gene assays considered: AREc32 for the activation of oxidative stress response, PPARγ-BLA for binding to the peroxisome proliferator–activated receptor gamma, and AhR-CALUX for activation of the arylhydrocarbon receptor. (**B**) Conceptual considerations for the derivation of the effect-based trigger for cytotoxicity from linear concentration–response curves (note that they are plotted on a double logarithmic scale for visualization only). AhR = arylhydrocarbon receptor; EC = effect concentration; D_{lipw} = lipid–water distribution constant; IC10 = 10% inhibition concentration; PPARγ = peroxisome proliferator–activated receptor gamma; QSAR = quantitative structure–activity relationship; REF = relative enrichment factor of a water sample in the bioassay (a REF of 1 means that the concentrations of the extracted chemicals are the same as in the original water sample); SR = specificity ratio.

cytotoxicity at an REF of 1 (red line in Figure 1B), then the acceptable IC10 for cytotoxicity would be an REF of 10. If we only accept 0.1% of cytotoxicity (blue dashed line in Figure 1B), then the acceptable IC10 for cytotoxicity would be an REF of 100.

We suggest accepting 1% of cytotoxicity in any original water sample; that is, the anchor for the accepted baseline toxicity of chemical mixtures extracted from water would correspond an EBT-IC10 of REF 10. Even if the cytotoxicity was caused by chemicals with specific modes of action, we can use the same cytotoxicity threshold because it does not matter if the effect is caused by a large molar amount of baseline toxicants or a smaller molar amount of specifically acting chemicals.

What is a safe level for chemicals with specific modes of action?

Ideally, we would need to anchor each effect of an in vitro assay applied in water quality assessment to the relevant adverse outcome pathways (AOPs). More than 200 AOPs have been developed in the last decade (Leist et al. 2017) since the concept was proposed (Ankley et al. 2010), but their applicability for risk assessment remains limited because of their chemical-agnostic feature, limited quantitative aspects, and lack of multiple networks of AOPs (Perkins et al. 2015). Furthermore, we want to be protective for any aquatic organism and protect human health in case surface water is used for occupational or recreational purposes or drinking water abstraction, but AOPs are typically specific for certain organisms.

If we use the cytotoxicity EBT-IC10 as a point of departure, then the question for a given reporter gene assay remains how many chemicals elicit specific effects in this assay and how much of this specific effect may a mixture trigger before it becomes a problem for the organism. A measure of the degree of specificity of a chemical is the so-called baseline specificity ratio ($SR_{baseline}$; Escher et al. 2020b), which describes how much more potent a chemical is in reporter gene activation (i.e., its specific mode of action) in comparison to the baseline cytotoxicity QSAR of the same cell line (Figure 2A). The $SR_{baseline}$ (Equation 1; Escher et al. 2020b) is defined as the ratio of predicted baseline toxicity ($IC10_{baseline}$) to the effect concentration for a specific endpoint (EC10; effect concentration causing 10% of maximum activation) or ECIR1.5 (effect concentration causing an induction ratio [IR] of 1.5–50% over control of IR1).

$$SR_{baseline} = \frac{IC10_{baseline}}{EC10} \text{ or } SR_{baseline} = \frac{IC10_{baseline}}{ECIR1.5} \quad (1)$$

An $SR_{baseline} \leq 1$ is not specific, $1 \leq SR_{baseline} < 10$ can be considered moderately specific (with high uncertainty), $10 \leq SR_{baseline} < 100$ is specific, and $100 \leq SR_{baseline}$ is highly specific (Escher et al. 2020b). It is also possible to define the specificity ratio against the measured IC10 in a given bioassay ($SR_{cytotoxicity}$; Escher et al. 2020b). We are not using $SR_{cytotoxicity}$ in the present study because the cytotoxicity might also be triggered by a specific mode of action. The $SR_{cytotoxicity}$ value is therefore rather an indicator of selectivity (i.e., a combination by multiplication of $SR_{baseline}$ and the toxic ratio, which is the ratio of $IC10_{baseline}$ to the experimental IC10).

Bioassays can be classified into 2 categories that are distinctly different from baseline toxicity. Baseline toxicants are expected to show a fairly tight log-normal distribution approximately specificity ratio = 1 (Figure 2B); that is, the effect concentrations are expected to be close to the IC10

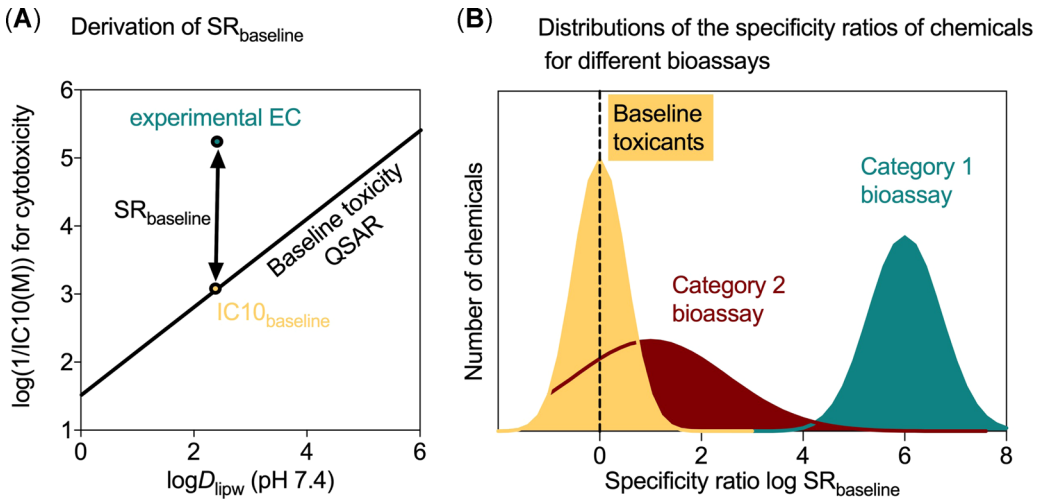

FIGURE 2: (**A**) Conceptual figure describing the derivation of the baseline specificity ratio ($SR_{baseline}$; Equation 1) from experimental effect concentrations and the inhibitory concentration causing 10% reduction in cell viability that was predicted from the quantitative structure–activity relationship for baseline toxicity. (**B**) Conceptual figure of distributions of specificity ratios log $SR_{baseline}$ for highly specific bioassays (category 1) and those that respond to many different chemicals (category 2) compared to distributions of log $SR_{baseline}$ of baseline toxicants. EC = effect concentration; D_{lipw} = lipid–water distribution constant; IC10 = 10% inhibition concentration; QSAR = quantitative structure–activity relationship.

(or no effects can be detected because cytotoxicity masks the effect).

Bioassays that are mainly triggered by highly specific chemicals have been classified as category 1 bioassays in a previous study that proposed EBTs (Escher et al. 2018b). These category 1 bioassays target highly specific, mainly receptor-mediated effects such as binding to hormone receptors. We proposed to define category 2 bioassays as those assays covering effect endpoints that are triggered by many more and more diverse chemicals that exhibit specific effects but with lower degree of specificity. Category 2 bioassays include oxidative stress response and reporter gene assays with more promiscuous nuclear receptor, such as AhR and PPARγ (Escher et al. 2018b).

In the earlier study on EBTs (Escher et al. 2018b), there were no clear criteria defined on how to differentiate category 1 from category 2 bioassays. With the definition of the specificity ratios in Escher et al. (2020b), we have a quantitative measure for the categorization. The highly specific category 1 bioassays respond to chemicals that have a relatively narrow distribution at high $SR_{baseline}$ (Figure 2B), and the responsive chemicals have been termed "defined mixtures" (Escher et al. 2018b). Their $SR_{baseline}$ values are typically ≥100, often up to 10^6 (Escher et al. 2020b). Category 2 bioassays respond to many more chemicals that have lower $SR_{baseline}$ values that typically show a wider distribution (Figure 2B). Many of the chemicals activating category 2 bioassays are presumably not known and hence are "undefined mixtures." Of course, there can be transitions between category 1 and 2 bioassays if those have bimodal or multimodal distributions of high- and low-potency ligands. Hence, we could also look at Figure 2B with different eyes—it could refer to one bioassay, in which some chemicals act highly specifically, whereas others show low potency and some are merely baseline toxicants whose measured specific effect is an artifact of the cytotoxicity burst (Judson et al. 2016; Fay et al. 2018).

If all chemicals were baseline toxicants in a given assay, we could directly apply the cytotoxicity EBT-IC10. For category 1 bioassays, existing EBTs that were read across from environmental quality standards (EQSs; Escher 2018b) remain valid because it is typically just a very small number of high-potency chemicals triggering the specific effect.

For category 2 bioassays, low-potency effect data were filtered out in the previous proposal for EBT derivation (Escher et al. 2018b). Read-across from EQS was only possible for the most potent chemicals in these category 2 bioassays, and we had to invoke a mixture factor to acknowledge that many more chemicals with lower potency were active in these bioassays. As discussed previously (Escher et al. 2018b), this approach is not very useful and is subject to major uncertainty. The resulting EBTs for category 2 bioassays were considered preliminary, although they have been applied in several case studies and slight improvements have been made by using a larger experimental single-chemical database (Neale et al. 2020a) or by harmonizing (de Baat et al. 2020) with the Smart Integrated Monitoring (SIMONI) approach (van der Oost et al. 2017).

We propose an approach for these category 2 bioassays that uses the effect data for all chemicals and their distributions of $SR_{baseline}$ to not only better differentiate between the different bioassay categories but also use these distributions to derive the EBT. First, it is important to check if the $SR_{baseline}$ values are indeed log-normally distributed. We will show in the *Results and Discussion* section how this can be practically accomplished and how to diagnose bioassays with high- and low-potency chemicals by breaking them up into several distributions.

We then suggest using the 50th percentile of the log-normal distribution, which is effectively the median, to derive

the EBT-ECIR1.5 for adaptive stress responses and the EBT-EC10 for receptor-mediated effects (Equation 2):

$$\text{EBT-ECIR1.5 or EBT-EC10} = \frac{\text{IC10}_{\text{baseline}}}{\text{SR}_{\text{baseline}}} = \frac{\text{REF 10}}{\text{SR}_{\text{baseline}}} \quad (2)$$

The median is used with the rationale that this is an integration over all $SR_{baseline}$ values provided that they follow a standard log-normal distribution, which will be checked and discussed. Also, the division by the $SR_{baseline}$ basically means that, on average, the mixture of chemicals in any sample with a very high number of chemicals responds at a lower REF than baseline toxicity. Given that all these considerations rely on linear concentration–response curves at low effect levels, one can expect that a sample that does not exceed the EBT-EC for specific effect does not exceed 1% of effect.

As for the cytotoxicity EBT-IC10, these EBT-ECIR1.5 and EBT-EC10 values are given in units of REFs of the water sample; but for comparison with the previously proposed EBT-BEQ values, they can also be related to reference compounds (Equation 3).

$$\text{EBT-BEQ} = \frac{\text{ECIR1.5}_{\text{reference compound}}}{\text{EBT-ECIR1.5}} \text{ or } \frac{\text{EC10}_{\text{reference compound}}}{\text{EBT-EC10}} \quad (3)$$

MATERIALS AND METHODS

Cell lines

The AREc32 reporter gene cell line, which was derived from MCF7 cells (Wang et al. 2006); PPARγ-BLA, derived from HEK293H cells; and AhR-CALUX (H4L7.5c2), based on H4IIe cells (Brennan et al. 2015) were frequently used in the past for water quality assessment and corresponding reporter gene assays: TOX21_ARE_BLA_Agonist_viability/TOX21_ARE_BLA_Agonist_ratio, TOX21_PPARg_BLA_Antagonist_viability/TOX21_PPARg_BLA_Agonist_ratio, and TOX21_AhR_LUC_Agonist_viability/TOX21_AhR_LUC_Agonist are included in the Tox21 database, which is accessible via the CompTox Chemicals Dashboard (US Environmental Protection Agency 2020).

Chemicals and their physicochemical properties

There were 689 environmentally relevant chemicals included in the derivation of EBTs for the 3 bioassays (Supplemental Data, Table S1). They were a subset of chemicals detected in surface water, including chemicals identified by Busch et al. (2016) as major European river pollutants (Malaj et al. 2014) and/or detected in recent case studies in surface waters across Europe (König et al. 2017; Neale et al. 2017b, 2020a). Only chemicals for which we had in-house measurements in one or more of the 3 reporter gene assays or data were available through the Tox21 database were included.

We collated physicochemical properties for the 689 chemicals (Supplemental Data, Table S1). The octanol–water partition constants ($\log K_{OW}$) of the neutral species were retrieved from the CompTox Chemicals Dashboard, with preference given to experimental data from PhysPropNCCT, followed by predictions with OPERA (Mansouri et al. 2018). The QSARs use the liposome–water partition constants (K_{lipw}) as the descriptor of hydrophobicity. For acids and bases as well as multiprotic chemicals, the speciation has to be considered. The fraction of the neutral and ionized species was calculated with the Henderson-Hasselbalch equation from the acidity constant pK_a, which was estimated with ACD/Percepta p*K*a using the GALAS algorithm (ACD/Labs, 2015 release [Build 2726]). The ionization-corrected liposome–water distribution ratios ($\log D_{lipw}$; pH 7.4) were calculated from experimental K_{lipw} values of the neutral species, K_{lipw}(neutral), and the charged species, K_{lipw}(ionized), with Equation 4.

$$\log D_{lipw} = f_{neutral} \times \log K_{lipw}(\text{neutral}) + f_{ionized} \times \log K_{lipw}(\text{ionized}) \quad (4)$$

Experimental K_{lipw} data (Betageri and Rogers 1987; Avdeef et al. 1998; Escher et al. 2000, 2017; Kwon et al. 2006; Endo et al. 2011; Bittermann et al. 2014; Henneberger et al. 2019a, 2019b; Klüver et al. 2019; Ebert et al. 2020) were given preference in Supplemental Data, Table S1. If no experimental data were available, a $\log K_{OW}$-based QSAR (Endo et al. 2011) was used to predict the $\log K_{lipw}$(neutral), and it was assumed that K_{lipw}(ionized) was 10 times lower than K_{lipw}(neutral; Bittermann et al. 2014) as detailed previously (Escher et al. 2020b).

Baseline toxicity QSARs

The $IC10_{baseline}$ values were predicted with baseline toxicity QSARs developed recently for the 3 cell lines (Escher et al. 2019). The QSARs are depicted in Figure 1B and are very similar for all 3 cell lines, in accordance with the concept of critical membrane concentrations. These baseline toxicity QSARs were initially derived for neutral species and a range of $1 < \log K_{lipw} < 5$ (Escher et al. 2019), but we extended them to ionizable chemicals by replacing the $\log K_{lipw}$ with the $\log D_{lipw}$ (pH 7.4) and applied them in a range from $0.6 < \log D_{lipw}$ (pH 7.4) < 5.4 in a previous experimental study (Escher et al. 2020b). In the present study, we extrapolated the QSARs another 0.6 log units on both ends of hydrophobicity and applied them to $0 < \log D_{lipw}$ (pH 7.4) < 6. Seventy-seven chemicals (11%) had a $\log D_{lipw}$ (pH 7.4) < 0, and 24 chemicals (3%) had a $\log D_{lipw}$ (pH 7.4) > 6 and were therefore excluded from the analysis, leaving 588 chemicals for analysis of the specificity ratio, of which 75 chemicals fell into the extrapolated range of $0 < \log D_{lipw}$ (pH 7.4) < 1 and 22 chemicals into the extrapolated range of $5 < \log D_{lipw}$ (pH 7.4) < 6. The $\log D_{lipw}$ (pH 7.4) and the predicted $IC10_{baseline}$ are listed in Supplemental Data, Table S1.

Experimental cytotoxicity and activity data of single chemicals

The IC10s and effect concentrations ECIR1.5 in AREc32 or EC10 in PPARγ-BLA or AhR-CALUX were previously measured in our laboratory (Neale et al. 2017a; Escher et al. 2019; Huchthausen et al. 2020; Neale et al. 2020a) for 121 chemicals in AREc32, 52 chemicals in PPARγ-BLA, and 89 chemicals in

AhR-CALUX (H4L7.5c2) and are listed in Supplemental Data, Table S1. In Supplemental Data, Table S1, 55 data points (20 chemicals in AREc32, 12 chemicals in PPARγ-BLA, and 23 chemicals in AhR-CALUX) were newly measured according to methods described in Neale et al. (2020a). All Tox21 entries for these assays have recently been reevaluated (Escher et al. 2020b) with a linear concentration–effect model (Escher et al. 2018a), and EC10 values for 522 chemicals in TOX21_ARE_BLA, 614 chemicals in TOX21_PPARg_BLA_Agonist_ratio, and 574 chemicals in TOX21_AhR_LUC_Agonist were also included in Supplemental Data, Table S1, marked with the source "Tox21"; but these EC10 values were previously reevaluated from publicly available concentration–response data with a linear concentration–effect model (Escher et al. 2020b). The cell lines AREc32 and AhR-CALUX are not identical to the Tox21 assays TOX21_ARE_BLA and TOX21_AhR_LUC_Agonist but are sufficiently similar to apply them as one data set (Neale et al. 2020a); PPARγ-BLA is identical to TOX21_PPARg_BLA_Agonist_ratio.

Overall 25% of all chemicals were active in ARE (30% in AREc32, 24% in TOX21_ARE_BLA_Agonist), 5.6% in PPARγ-BLA, and 20% in AhR (26% in AhR-CALUX and 18% in TOX21_AhR_LUC_Agonist).

Iceberg modeling of mixture effects in water samples

Iceberg modeling compares the mixture effects predicted from the concentrations of detected chemicals and effect data of the detected chemicals with the measured effects (Escher et al. 2020c). In the present study, we reused data from iceberg modeling of a highly diverse set of surface water quality data (tables S7–S9 in Neale et al. [2020a], summarized in Supplemental Data, Table S2) and data from wastewater-treatment plants (WWTPs; Table S9 in Neale et al. [2020c], summarized in Supplemental Data, Table S3), where more than 600 chemicals were analyzed with the same analytical method, and all bioassay data were measured with AREc32, PPARγ-BLA, and AhR-CALUX after SPE. The number of analyzed chemicals was smaller in those 2 case studies than in Supplemental Data, Table S1, and hence a smaller number of chemicals had been considered in the iceberg modeling (Neale et al. 2020a, 2020c).

For cytotoxicity, the ratio of the toxic units (TU = 1/IC10) predicted from quantified chemicals (TU_{chem}) to experimental toxic units (TU_{bio}) served as a measure of the fraction of total cytotoxicity explained (Neale et al. 2020a). For activation of the reporter genes, the ratio of BEQs (BEQ_{chem}/BEQ_{bio}), predicted from the detected chemicals' concentrations and their effect potency (BEQ_{chem}) and measured in the bioassays ($BEQ_{bio} = EC_{ref}/EC_{sample}$, when EC can be ECIR1.5 or EC10), served as a measure of the fraction of effect explained (Neale et al. 2020a).

Estimated number of cytotoxic and bioactive chemicals in water samples

We know the number of detected chemicals that were cytotoxic or active and used to predict TU_{chem} and BEQ_{chem} in the water samples of the iceberg modeling case studies (Supplemental Data, Tables S2 and S3). From "#detected cytotoxic chemicals" we can extrapolate the total number of cytotoxic chemicals with Equation 5 and the extrapolated total number of bioactive chemicals with Equation 6.

$$\text{Extrapolated total number of cytotoxic chemicals} = \#\text{detected cytotoxic chemicals}/(TU_{chem}/TU_{bio}) \quad (5)$$

$$\text{Extrapolated total number of bioactive chemicals} = \#\text{detected bioactive chemicals}/(BEQ_{chem}/BEQ_{bio}) \quad (6)$$

RESULTS AND DISCUSSION

Estimated number of cytotoxic and bioactive chemicals

In the case study on 128 highly diverse surface water samples collected during rain events (Neale et al. 2020a), cytotoxicity data were available and included for 99 water samples of AREc32, 117 samples of PPARγ, and 113 samples of AhR (Supplemental Data, Table S2). The fraction of cytotoxicity explained by the detected chemicals (TU_{chem}/TU_{bio}) was low, despite up to 64 cytotoxic chemicals being detected, and increased only slightly with increasing number of detected chemicals (Supplemental Data, Table S2; Figure 3A). The extrapolated numbers of cytotoxic chemicals (Equation 5) varied widely (Supplemental Data, Table S2) and were log-normally distributed (Figure 3B; D'Agostino & Pearson test, $p = 0.41–0.98$). The log-normal distributions were significantly different between the 3 cell lines (one-way analysis of variance, $p < 0.0001$), with AhR having a slightly lower range of mean ± standard error (SE) with 6944 ± 1316 (95% CI 4336–9552) than ARE with 25 770 ± 5456 (95% CI 14 925–36 616) and PPARγ with 13 086 ± 1615 (95% CI 9887–16 286). The mean ± SE extrapolated number of cytotoxic chemicals in all assays was 14 793 ± 1843 (95% CI 11 169–18 418), which means that we can expect roughly 15 000 cytotoxic chemicals in a water sample that act together in a concentration-additive manner (Escher et al. 2020a). Because all chemicals are cytotoxic, albeit with different potencies, we can use that number in principle as an approximation of the number of chemicals present in a complex mixture in a water sample. Of course, any complex mixture contains high and low concentrations of thousands of chemicals (Escher et al. 2020c). Furthermore, the main assumption underlying this extrapolation is that the analyzed chemicals are a representative subsample of the entire chemical universe so that extrapolation to all chemicals is possible. Another assumption is that the effect potency of the included chemicals is representative of all chemicals. This is likely not true because we have chemicals on the target list of chemical analysis, from which we know that they are problematic and known water pollutants. Nevertheless, this analysis can tell us that there must be many more chemicals active in the complex mixtures extracted from water than the ones we can quantify. Increasing the numbers of chemicals analyzed will not solve the problem; we need to complement chemical analysis with

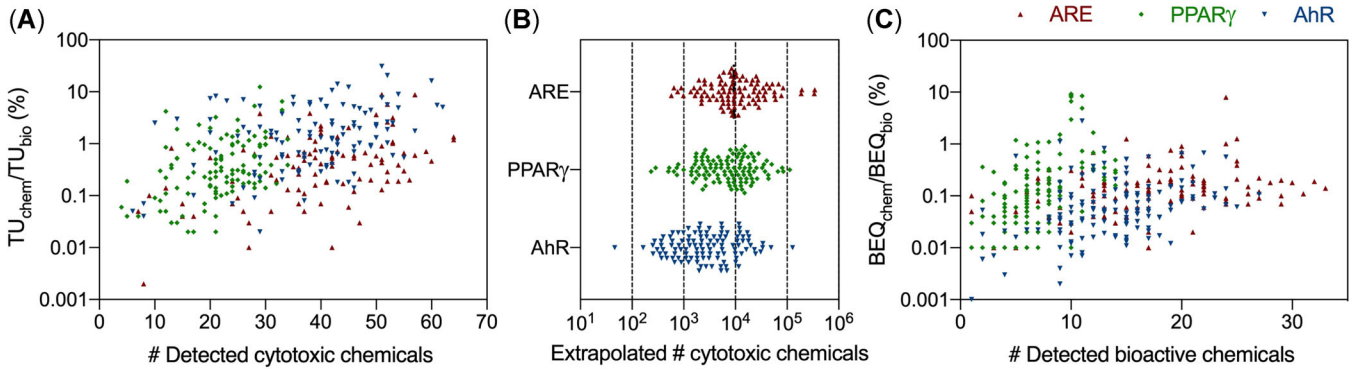

FIGURE 3: (**A**) Relationship between number of detected cytotoxic chemicals and fraction of toxic units (TUs) explained by the detected chemicals (TU_{chem}/TU_{bio}) in 128 surface water samples (Neale et al. 2020a). (**B**) Extrapolated number of cytotoxic chemicals in surface water (Equation 5). (**C**) Relationship between number of detected bioactive chemicals and fraction of bioanalytical equivalent concentrations explained by the detected chemicals in 128 surface water samples (Neale et al. 2020a) and 20 samples from wastewater-treatment plants (Neale et al. 2020c). All data are in Supplemental Data, Tables S2 and S3. AhR = arylhydrocarbon receptor; ARE = antioxidant response element; BEQ = bioanalytical equivalent concentration; PPARγ = peroxisome proliferator–activated receptor gamma.

bioassays for apical endpoints such as cytotoxicity. This underpins the need to develop EBTs that account for the mixture effects and not only rely on a few highly potent chemicals.

Activity data were available from 128 surface water samples (Neale et al. 2020a) and 20 samples from WWTPs (Neale et al. 2020c) with activity detected in 120 water samples in AREc32, 140 samples in PPARγ, and 142 samples in AhR (Supplemental Data, Tables S2 and S3). Fewer chemicals were active in these specific endpoints, and there appeared to be an even weaker correlation between the number of detected bioactive chemicals and the fraction of BEQ explained by the detected chemicals (BEQ_{chem}/BEQ_{bio}; Figure 3C), with most of BEQ_{chem}/BEQ_{bio} ranging from 0.01 to 1%. The low fraction of effect explained by the detected chemicals means that there are many more bioactive chemicals that have not been identified but need to be accounted for because they contribute to the mixture effect elicited by extracts from water samples. Because, despite efforts to detect hundreds of chemicals and assess their bioactivity individually, no substantial fraction of effect can be explained by the detected chemicals, it is imperative that we develop trigger values for mixture effects and not for groups or individual chemicals.

What is a safe level for cytotoxicity in water samples?

When applying the proposed EBT-IC10 of REF 10 to the cytotoxicity data set ($IC10 = 1/TU_{bio}$ in Supplemental Data, Table S2) of water samples collected during rainfall in small agricultural streams (Neale et al. 2020a), we can see in Figure 4 that only 2 of 72 samples in AhR, only 4 of 73 samples in PPARγ, and 7 of 62 samples in AREc32 exceeded the threshold (i.e., had IC10 <10), whereas for WWTPs (Table S2 of Neale et al. [2020c]) most influent samples were more cytotoxic than the EBT-IC10 of 10. Note that we did not enrich more than REF 4 in that study, so we cannot tell if some of the effluents were meeting the EBT-IC10. Both water types, surface and WWTP waters, were enriched by SPE, which also removed inorganic contaminants and matrix components. The good discrimination between surface water and untreated wastewater and the fact that only a small fraction of surface waters exceeded the trigger is a confirmation that the choice of the protection level of 1% was appropriate.

It is interesting to note that van der Oost et al. (2017) proposed an EC50 of REF 20 for nonspecific effects toward cells (Cytotox CALUX), bacteria, and zooplankton, which is very

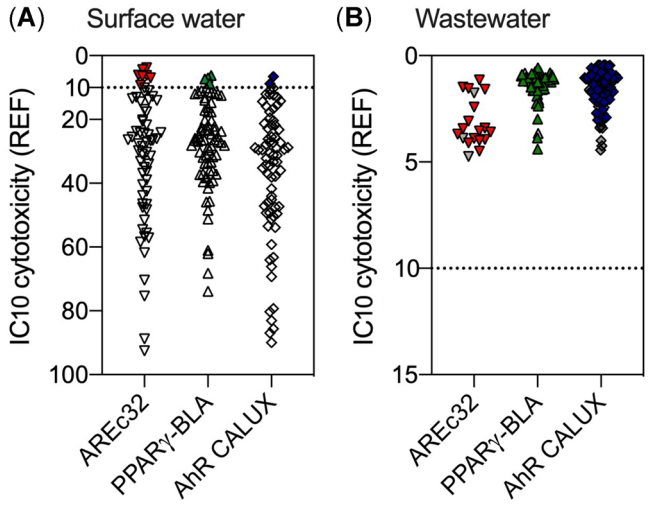

FIGURE 4: (**A**) Distribution of inhibitory concentration causing 10% reduction in cell viability (IC10) in water samples collected during rainfall in small agricultural streams; (Neale et al. 2020a) and (**B**) in wastewater-treatment plant influent (color) and effluent (gray; Neale et al. 2020c). In both figures, the dotted lines correspond to the proposed effect-based trigger (EBT) for cytotoxicity EBT-IC10 of REF 10. AhR = arylhydrocarbon receptor; PPARγ = peroxisome proliferator–activated receptor gamma; REF = relative enrichment factor.

similar to the proposed threshold for cells considering also the differences between the 10% and 50% effect levels.

Are the $SR_{baseline}$ values log-normally distributed?

As shown in Supplemental Data, Table S1, for ARE, 166 of 689 chemicals were active, 497 chemicals were inactive, and for 26 chemicals we did not have data for this pathway. For PPARγ, only 38 chemicals were active, 640 were inactive, and 11 data points were missing, whereas for AhR 133 chemicals were active, 553 were inactive, and 3 data points were missing.

The $IC10_{baseline}$ (Supplemental Data, Table S1) were predicted for all chemicals in the range of $0 < \log D_{lipw}$ (pH 7.4) < 6, leaving 152 (ARE), 34 (PPARγ), and 125 (AhR) chemicals for the $SR_{baseline}$ (Supplemental Data, Table S1) analysis.

We plotted the log $SR_{baseline}$ against the rank, expressed in probit units (Figure 5): ARE (Figure 5A), PPARγ (Figure 5B), and AhR (Figure 5C) showed remarkably good log-normal distributions with highly linear probit plots. Only very low and very high probit values deviated slightly from the linear regression. When the log $SR_{baseline}$ values of all 3 bioassays were superimposed, they showed remarkably little difference (Figure 5D).

For comparison, we also plotted the distribution of a category 1 bioassay, using the reporter gene assay for estrogenicity ERα-BLA as an illustrative example. Although the distribution of single-chemical effect data from Tox21 (Supplemental Data, Table S4; data from Escher et al. [2020b], low-potency estrogens) measured in ERα-BLA overlapped with these distributions at low log $SR_{baseline}$, there were 4 highly potent estrogenic chemicals with $SR_{baseline} > 1$ million that largely deviated from the probit regressions of the low-potency chemicals (Figure 5D). These are 17-α-ethinylestradiol, 17-α-estradiol, 17-β-estradiol, and estriol (Supplemental Data, Table S4; data from Hashmi et al. [2018]). These high-potency estrogens dominate the mixture effects in environmental waters together with the lower-potency estrone (Kase et al. 2018; Könemann et al. 2018) even if xenoestrogens such as bisphenol A and butylparaben are present at much higher concentrations because of the much lower potency of the xenoestrogens. This analysis confirms that estrogenicity bioassays are category 1 bioassays, where previously proposed EBTs remain valid (Escher et al. 2018b).

The AhR also has ligands of very high specificity, such as dioxin and dioxin-like chemicals; but these high-potency AhR ligands are very hydrophobic and do not occur in the aqueous phase but are bound to suspended particulate matter and sediments. Therefore, they are not relevant for the derivation of EBTs for surface water. If the approach was used to derive EBTs for sediments, though, one would need to consider dioxin-like chemicals.

The question may arise as to why we are not working with effect concentrations themselves but with the $SR_{baseline}$. Effect concentrations have both a toxicokinetic and a toxicodynamic component, whereas the $SR_{baseline}$ reflects toxicodynamics only. Hence, the distributions of ECIR1.5 and EC10 values of the chemicals listed in Supplemental Data, Table S1 are not log-normal (Figure 6) and cannot be used to derive EBTs.

Derivation of mixture EBTs

The log-normal distributions of $SR_{baseline}$ projected with the regression parameters from the probit analysis are plotted in Figure 7. The 50th percentiles of the distributions are highlighted with dotted lines and correspond to $SR_{baseline}$ values of 8 for ARE (Figure 7A), 33 for PPARγ (Figure 7B), and 12 for AhR (Figure 7C). With these 50th percentiles we are in the region of moderately specific ($1 \leq SR_{baseline} < 10$), but although uncertainty of this region is high for single chemicals, the distribution of large numbers of data points provides some robustness of the $SR_{baseline}$ and reflects the fact that many chemicals are active but with moderate specificity.

Although ARE, PPARγ, and AhR showed unimodal distributions, ERα-BLA had 2 distinct distributions with a maximum at $SR_{baseline}$ of 6.6 for low-potency ERα agonists and another maximum at $SR_{baseline}$ of 4×10^6 for high-potency ERα agonists (Figure 7D). This visualization confirms the analysis that ERα is a category 1 bioassay.

The EBT-ECIR1.5 for oxidative stress response can be derived from the 50th percentile of the log-normal distribution of $SR_{baseline}$ of 8, which can be converted with Equation 2 to a mixture EBT-ECIR1.5 of REF 1.2. This AREc32 EBT-ECIR1.5 is

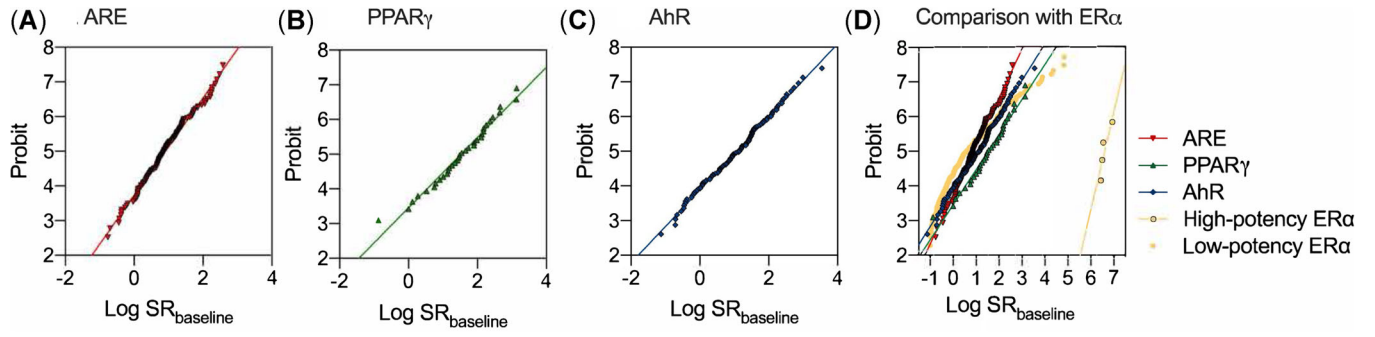

FIGURE 5: Probit plots of logarithms of specificity ratios (Supplemental Data, Table S1) for (**A**) ARE (slope = 1.4, intercept = 3.7, $r^2 = 0.9928$), (**B**) PPARγ (slope = 1.0, intercept = 3.5, $r^2 = 0.9778$), (**C**) AhR (slope = 1.1, intercept = 3.9, $r^2 = 0.9952$), and (**D**) overlay of all 3 probit plots (**A**–**C**) and comparison with ERα (Supplemental Data, Table S4). AhR = arylhydrocarbon receptor; ARE = antioxidant response element; ERα, estrogen receptor alpha; PPARγ = peroxisome proliferator–activated receptor gamma; SR = specificity ratio.

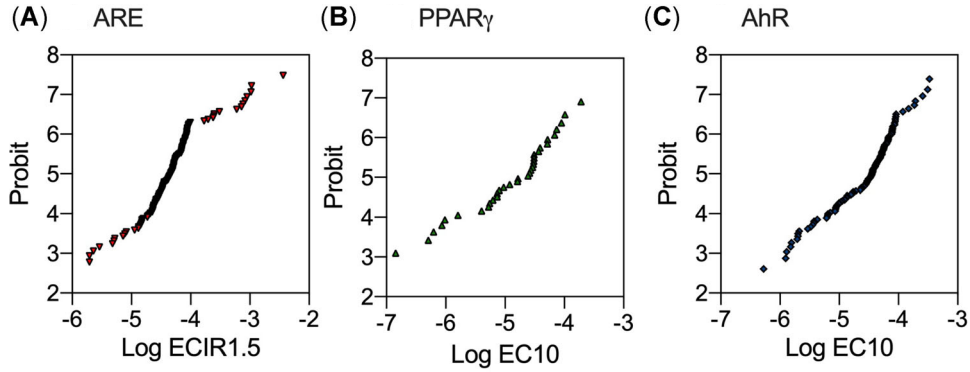

FIGURE 6: Probit plots of logarithms of effect concentrations from Supplemental Data, Table S1 for (**A**) ARE, (**B**) PPARγ, and (**C**) AhR. AhR = arylhydrocarbon receptor; ARE = antioxidant response element; EC10 = 10% effect concentration; ECIR1.5 = effect concentration causing an induction ratio IR of 1.5; PPARγ = peroxisome proliferator–activated receptor gamma.

remarkably similar to the initial EBT-ECIR1.5 of 3 derived in 2013 (Escher et al. 2013). The corresponding EBT-dichlorvos-EQ is 1.4 $mg_{dichlorvos}$/L, which is 10 times higher than the previously proposed EBT-dichlorvos-EQ of 156 $\mu g_{dichlorvos}$/L that had applied a mixture factor of 1000 (Escher et al. 2018b). In hindsight, a mixture factor of 10 000 would have been more appropriate and led to the same EBT-dichlorvos-EQ as the specificity ratio-distribution method.

The EBT-EC10 for PPARγ is now REF 0.3, which corresponds to an EBT-rosiglitazone-EQ of 1.2 $\mu g_{rosiglitazone}$/L. This is 33 times higher than the previously derived EBT-rosiglitazone-EQ of 36 $ng_{rosiglitazone}$/L (Escher et al. 2018b) and the revised value of 19 $ng_{rosiglitazone}$/L (Neale et al. 2020a), which had assumed that PPARγ is a high-potency endpoint and that no mixture factor had been considered. At the time, there was limited experimental data for the PPARγ assay available, but more recent iceberg modeling experience (Neale et al. 2020a) has shown that the fraction of effect explained by known chemicals was often <1% and that many chemicals are low-potency agonists. Hence the reclassification as a category 2 bioassay seems reasonable, although only 5.6% of the 678 chemicals with bioassay data were active (in contrast to 25% for ARE and 19% for AhR).

The mixture EBT-EC10 for AhR is REF 0.9, which corresponds to an EBT-B[a]P-EQ of 250 $ng_{B[a]P}$/L. This value is 40 times higher than the previously proposed EBT-B[a]P-EQ of 6.4 $ng_{B[a]P}$/L, which had applied a mixture factor of 100 (Escher et al. 2018b), but is similar to the SIMONI EBT-B[a]P-EQ of 150 $ng_{B[a]P}$/L for the PAH-CALUX, which was defined as twice the background B[a]P-EQ (van der Oost et al. 2017).

While the mixture factors in the previous derivation were derived only from expert knowledge and considerations of numbers of chemicals acting together, the new derivation is based on large numbers of chemicals, does not filter out any entries, and is the same approach for all category 2 bioassays.

Application of mixture EBTs in case studies

We applied the new mixture EBTs to the 2 case studies on surface water (Neale et al. 2020a) and a WWTP (Neale

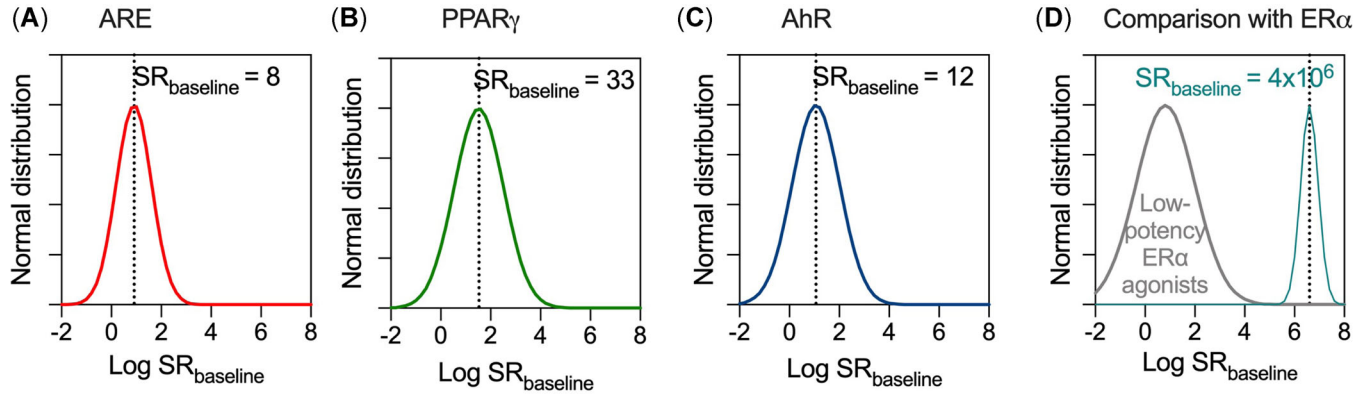

FIGURE 7: Visualization of the log-normal distributions of baseline specificity ratios ($SR_{baseline}$) and 50th percentile of the distribution for (**A**) ARE, (**B**) PPARγ, (**C**) AhR, and (**D**) ERα showed 2 distinctly different distributions of low-potency (gray) and high-potency ERα agonists. The normal distributions were constructed from the probit regression descriptors in Figure 5 (ARE, slope = 1.4, intercept = 3.7; PPARγ, slope = 1.0, intercept = 3.5; AhR, slope = 1.1, intercept = 3.9; ERα [low potency], slope = 0.4, intercept = 4.3; ERα [high potency], slope = 2.8, intercept = –13.6) using the Excel function NORMDIST((slope × log $SR_{baseline}$ + intercept), 5, 1, 0), and the medians of log $SR_{baseline}$ were calculated from the probit regression for probit = 5. AhR = arylhydrocarbon receptor; ARE = antioxidant response element; ERα, estrogen receptor alpha; PPARγ = peroxisome proliferator–activated receptor gamma.

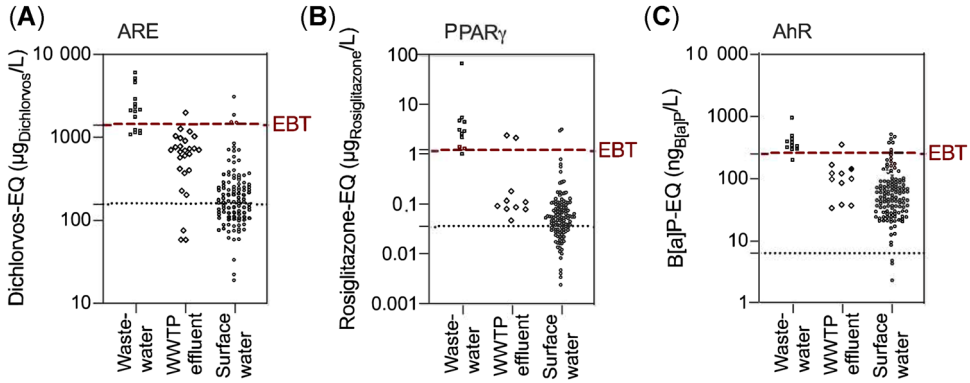

FIGURE 8: Application of the new effect-based triggers (EBTs) in field studies (circles [Neale et al. 2020a] and squares and diamonds [Neale et al. 2020c]) and previous data collections (gray symbols [Escher et al. 2018b]). Red dashed lines indicate new EBT bioanalytical equivalent concentration (BEQ), black dotted lines indicate previously proposed EBT-BEQ with the read-across method (Escher et al. 2018b). AhR = arylhydrocarbon receptor; ARE = antioxidant response element; EQ = equivalent; PPARγ = peroxisome proliferator–activated receptor gamma; WWTP = wastewater-treatment plant.

et al. 2020c) and revisited the case studies from the previous EBT paper (appendix B in Escher et al. [2018b]). With the new EBT-dichlorvos-EQ of 1.4 $mg_{dichlorvos}$/L, most of the surface water samples did not exceed the trigger for oxidative stress response (Figure 8A). In addition, the EBT-dichlorvos-EQ differentiated well between wastewater and surface water and often even differentiated well between untreated wastewater and WWTP effluent. The one exceedance of WWTP effluent was after an initial membrane filtration step that was directly followed by further treatment. The 4 surface water samples that now exceed the EBT-dichlorvos-EQ were collected from sites impacted or likely impacted by industrial and municipal WWTP effluent. Of course, we cannot use exceedance for a certain water type as a criterion for whether an EBT is valid or not, but this comparison shows us that water that we expect to be of poorer quality is exceeding the EBT-dichlorvos-EQ. The new EBT-dichlorvos-EQ is probably more realistic than the EBT of 156 $\mu g_{dichlorvos}$/L derived in 2018 using the method for high-potency chemicals and a mixture factor (Escher et al. 2018b), which yielded much lower EBT and consequently a high frequency of EBT exceedances (Neale et al. 2020a).

The same holds for PPARγ, where the EBT-rosiglitazone-EQ differentiated well between treated and untreated wastewater, with 2 wastewater samples exceeding the EBT (Figure 8B). Both were from WWTPs that applied primary treatment only (P1 and P2 in Neale et al. [2020c]). All surface water samples had acceptable water quality, with exception of 2 that were from the Danube River at Novi Sad, in areas impacted by wastewater (König et al. 2017).

For AhR (Figure 8C), one of the primary treated effluent exceeded the EBT (339 ngB[a]P/L, P1 in Neale et al. [2020c]), whereas the other effluent from primary treatment was below. Seven of the surface water samples taken during rain events exceeded the EBT, which can be rationalized by road runoff and potentially river water impacted by WWTPs (Neale et al. 2020a).

Overall the new EBTs differentiated much better between wastewater and surface water than the EBT-BEQ values derived with the read-across method (Escher et al. 2018b), which are marked as dotted lines in Figure 8.

CONCLUSION

The present study closed the last conceptual gap in our derivation strategy of EBTs (Escher et al. 2018b). It was recognized earlier that the different bioassay categories reacting to highly specific but "rare" active chemicals (category 1 bioassays) and those reacting to a much larger range of more diverse chemicals of lower potency (category 2 bioassays) could not be treated in the same way. For category 2 bioassays, a mixture factor was implemented, which was scientifically justified but still subject to assumptions.

In the present study, we proposed an independent approach that starts from scratch, that is, from acceptable cytotoxicity of 1%, not by reading across from chemical guideline values but by rooting the derivation in acceptable cell toxicity for all cells and beings. The concept of baseline toxicity also has drawbacks in case of very hydrophilic and/or reactive chemicals because they are not associated with critical membrane concentrations. Nevertheless, the concept of baseline toxicity still appears sufficiently robust to serve as a basis, on top of which specific effects can be defined. By using the specificity ratio as a measure of the degree of specificity of effect, we do not have a mixed parameter, as is the case for the effect concentrations, which are a composite of toxicokinetics and toxicodynamics. Basically, with the new approach we are disentangling toxicokinetics (indirectly described by the baseline toxicity QSAR) from toxicodynamics (represented by the $SR_{baseline}$). One weakness is that metabolism is not accounted for or is merely integrated in the toxicodynamic parameter of $SR_{baseline}$. Generally, in vitro reporter gene assays are considered not to be metabolically active; but, as has been recently demonstrated, one of the cell lines applied in the present study, AREc32, expresses Cyp1A1 after induction with B[a]p (Fischer et al. 2020).

The excellent log-normal distributions of $SR_{baseline}$ in all 3 investigated reporter gene assays serve as important underpinning of the concept. This is why we can use the 50th percentile

or median of the distribution to translate the cytotoxicity EBT to EBTs for specific modes of action. The outlined method would, of course, be applicable to any category 2 bioassay, including other reporter gene constructs targeting the same endpoints as those applied in the present study or other receptors (e.g., PXR). However, we would need a large number of single-chemical data for construction of the $SR_{baseline}$ distributions, which presently are not available. It remains to be explored if genotoxicity assays also fall into category 2 bioassays and if EBTs for genotoxicity could also be developed, which is clearly a gap at present. Only van der Oost et al. (2017) have derived genotoxicity EBTs for Ames, umuC, and p53 CALUX with EBTs derived for nonspecific endpoints reduced by an assessment factor of 10. This is actually very similar to the proposed approach with the assessment factor of 10 replaced by the specificity ratio. The distributions of $SR_{baseline}$ might also have some utility to better filter appropriate chemicals for the read-across approach of category 1 bioassays, as was evidenced by the bimodal distributions of estrogenic chemicals in the ERα-BLA assay.

Application of the new EBT in existing case studies confirmed that they differentiate unambiguously between poor and acceptable water quality. The present derivation had focused on surface water. However, with the baseline toxicity concept applicable to human cell lines, it can easily be extended to drinking water. Existing EBTs for category 1 bioassays are already mostly in a similar range for surface water and drinking water; hence, we could apply the proposed EBTs for surface water also to drinking water. Alternatively, one could also consider an additional uncertainty factor, which would effectively be equivalent to accepting only 0.1% of baseline toxicity for drinking water; but the specificity ratio factor would not change (unless there are substantial numbers of new single-chemical data) because it is bioassay-specific and not protection target-specific.

Supplemental Data—The Supplemental Data are available on the Wiley Online Library at https://doi.org/10.1002/etc.4944.

Acknowledgment—We thank the Global Water Research Coalition for supporting the present study within the framework of the project "Effect Based Monitoring in Water Safety Planning," which was jointly funded by the Public Utilities Board, Stowa-Foundation for Applied Water Research, Water Research Australia, Water Research Commission, Water Services Association of Australia, KWR–Water, SUEZ, Veolia–Research and Innovation, Griffith University, and the Helmholtz Centre for Environmental Research. We thank S. Ait-Assa, C. Arnal, F. Brion, S. Buchinger, M. Dechesne, T. Hamers, L. Henneberger, K. Hilscherova, R. Kase, F. Leusch, G. Meheut, L. Niu, E. Simon, and R. van der Oost for helpful discussions and for reviewing the manuscript. This article is dedicated to Lucy, who felt she could not wait to be born before it was submitted. Open Access funding enabled and organized by ProjektDEAL.

Disclaimer—All authors have no conflict of interest to declare. The views expressed in the present review are solely those of the authors.

Data Availability Statement—All literature data used for the modeling are compiled and referenced in Supplemental Data, Table S1.

REFERENCES

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* 29:730–741.

Avdeef A, Box KJ, Comer JEA, Hibbert C, Tam KY. 1998. pH-metric logP 10. Determination of liposomal membrane–water partition coefficients of ionizable drugs. *Pharm Res* 15:209–215.

Besselink H, Brouwer B, van der Burg B. 2017. Validation and regulatory acceptance of bio-based approaches to assure feedstock, water & product quality in a bio-based economy. *Ind Crops Prod* 106:138–145.

Betageri GV, Rogers JA. 1987. Thermodynamics of partitioning of beta-blockers in the normal-octanol-buffer and liposome systems. *Int J Pharm* 36:165–173.

Bittermann K, Spycher S, Endo S, Pohler L, Huniar U, Goss K-U, Klamt A. 2014. Prediction of phospholipid–water partition coefficients of ionic organic chemicals using the mechanistic model COSMOmic. *J Phys Chem B* 118:14833–14842.

Blackwell BR, Ankley GT, Bradley PM, Houck KA, Makarov SS, Medvedev AV, Swintek J, Villeneuve DL. 2019. Potential toxicity of complex mixtures in surface waters from a nationwide survey of United States streams: Identifying in vitro bioactivities and causative chemicals. *Environ Sci Technol* 53:973–983.

Brand W, de Jongh CM, van der Linden SC, Mennes W, Puijker LM, van Leeuwen CJ, van Wezel AP, Schriks M, Heringa MB. 2013. Trigger values for investigation of hormonal activity in drinking water and its sources using CALUX bioassays. *Environt Int* 55:109–118.

Brennan JC, He G, Tsutsumi T, Zhao J, Wirth E, Fulton MH, Denison MS. 2015. Development of species-specific Ah receptor-responsive third generation CALUX cell lines with enhanced responsiveness and improved detection limits. *Environ Sci Technol* 49:11903–11912.

Brion F, De Gussem V, Buchinger S, Hollert H, Carere M, Porcher JM, Piccini B, Feray C, Dulio V, Konemann S, Simon E, Werner I, Kase R, Ait-Aissa S. 2019. Monitoring estrogenic activities of waste and surface waters using a novel in vivo zebrafish embryonic (EASZY) assay: Comparison with in vitro cell-based assays and determination of effect-based trigger values. *Environ Int* 130:104896.

Busch W, Schmidt S, Kühne R, Schulze T, Krauss M, Altenburger R. 2016. Micro-pollutants in European rivers: A mode of action survey to support the development of effect-based tools for water monitoring. *Environ Toxicol Chem* 35:1887–1899.

Conley JM, Evans N, Mash H, Rosenblum L, Schenck K, Glassmeyer S, Furlong ET, Kolpin DW, Wilson VS. 2017. Comparison of in vitro estrogenic activity and estrogen concentrations in source and treated waters from 25 US drinking water treatment plants. *Sci Total Environ* 579:1610–1617.

de Baat ML, Van der Oost R, Van der Lee GH, Wieringa N, Hamers T, Verdonschot PFM, De Voogt P, Kraak MHS. 2020. Advancements in effect-based surface water quality assessment. *Water Res* 183:116017.

Ebert A, Allendorf F, Berger U, Goss KU, Ulrich N. 2020. Membrane/water partitioning and permeabilities of perfluoroalkyl acids and four of their alternatives and the effects on toxicokinetic behavior. *Environ Sci Technol* 54:5051–5061.

Endo S, Escher BI, Goss KU. 2011. Capacities of membrane lipids to accumulate neutral organic chemicals. *Environ Sci Technol* 45:5912–5921.

Escher B, Baumer A, Bittermann K, Henneberger L, König M, Kühnert C, Klüver N. 2017. General baseline toxicity QSAR for non-polar, polar and ionisable chemicals and their mixtures in the bioluminescence inhibition assay with *Alliivibrio fischeri*. *Environ Sci Process Impact* 19:414–428.

Escher B, Glauch L, Konig M, Mayer P, Schlichting R. 2019. Baseline toxicity and volatility cutoff in reporter gene assays used for high-throughput screening. *Chem Res Toxicol* 32:1646–1655.

Escher B, Neale PA, Villeneuve D. 2018a. The advantages of linear concentration–response curves for in vitro bioassays with environmental samples. *Environ Toxicol Chem* 37:2273–2280.

Escher BI, Ait-Aissa S, Behnisch PA, Brack W, Brion F, Brouwer A, Buchinger S, Crawford SE, Du Pasquier D, Hamers T, Hettwer K, Hilscherova K, Hollert H, Kase R, Kienle C, Tindall AJ, Tuerk J, van der Oost R, Vermeirssen E, Neale PA. 2018b. Effect-based trigger values for in vitro and in vivo bioassays performed on surface water extracts supporting the environmental quality standards (EQS) of the European Water Framework Directive. *Sci Total Environ* 628–629:748–765.

Escher BI, Allinson M, Altenburger R, Bain P, Balaguer P, Busch W, Crago J, Humpage A, Denslow ND, Dopp E, Hilscherova K, Kumar A, Grimaldi M, Jayasinghe BS, Jarosova B, Jia A, Makarov S, Maruya KA, Medvedev A, Mehinto AC, Mendez JE, Poulsen A, Prochazka E, Richard J, Schifferli A, Schlenk D, Scholz S, Shiraishi F, Snyder S, Su G, Tang J, van der Burg B, van der Linden S, Werner I, Westerheide SD, Wong CKC, Yang M, Yeung B, Zhang X, Leusch FDL. 2014. Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays. *Environ Sci Technol* 48:1940–1956.

Escher BI, Braun G, Zarfl C. 2020a. Exploring the concepts of concentration addition and independent action using a linear low-effect mixture model. *Environ Toxicol Chem* 39:2552–2559.

Escher BI, Henneberger L, Schlichting R, Fischer FC. 2020b. Cytotoxicity burst or baseline toxicity? Differentiating specific from nonspecific effects in reporter gene assays. *Environ Health Perspect* 128:077007.

Escher BI, Neale PA, Leusch FDL. 2015. Effect-based trigger values for in vitro bioassays: Reading across from existing water quality guideline values. *Water Res* 81:137–148.

Escher BI, Schwarzenbach RP, Westall JC. 2000. Evaluation of liposome–water partitioning of organic acids and bases. 1. Development of a sorption model. *Environ Sci Technol* 34:3954–3961.

Escher BI, Stapleton HM, Schymanski EL. 2020c. Tracking complex mixtures of chemicals in our changing environment. *Science* 367:388–392.

Escher BI, van Daele C, Dutt M, Tang JYM, Altenburger R. 2013. Most oxidative stress response in water samples comes from unknown chemicals: The need for effect-based water quality trigger values. *Environ Sci Technol* 47:7002–7011.

Fay KA, Villeneuve DL, Swintek J, Edwards SW, Nelms MD, Blackwell BR, Ankley GT. 2018. Differentiating pathway-specific from nonspecific effects in high-throughput toxicity data: A foundation for prioritizing adverse outcome pathway development. *Toxicol Sci* 163:500–515.

Fischer FC, Abele C, Henneberger L, Klüver N, König M, Mühlenbrink M, Schlichting R, Escher BI. 2020. Cellular metabolism in high-throughput in vitro reporter gene assays and implications for the quantitative in vitro–in vivo extrapolation. *Chem Res Toxicol* 33:1770–1779.

Hashmi MAK, Escher BI, Krauss M, Teodorovic I, Brack W. 2018. Effect-directed analysis (EDA) of Danube River water sample receiving untreated municipal wastewater from Novi Sad, Serbia. *Sci Total Environ* 624:1072–1081.

Henneberger L, Mühlenbrink M, Fischer FC, Escher BI. 2019a. C18-coated solid-phase microextraction fibers for the quantification of partitioning of organic acids to proteins, lipids, and cells. *Chem Res Toxicol* 32: 168–178.

Henneberger L, Muhlenbrink M, Konig M, Schlichting R, Fischer FC, Escher BI. 2019b. Quantification of freely dissolved effect concentrations in in vitro cell-based bioassays. *Arch Toxicol* 93:2295–2305.

Huchthausen J, Muhlenbrink M, Konig M, Escher BI, Henneberger L. 2020. Experimental exposure assessment of ionizable organic chemicals in in vitro cell-based bioassays. *Chem Res Toxicol* 33:1845–1854.

Jia A, Escher BI, Leusch FDL, Tang JYM, Prochazka E, Dong B, Snyder EM, Snyder SA. 2015. In vitro bioassays to evaluate complex chemical mixtures in recycled water. *Water Res* 80:1–11.

Judson R, Houck K, Martin M, Richard AM, Knudsen TB, Shah I, Little S, Wambaugh J, Setzer RW, Kothya P, Phuong J, Filer D, Smith D, Reif D, Rotroff D, Kleinstreuer N, Sipes N, Xia MH, Huang RL, Crofton K, Thomas RS. 2016. Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. *Toxicol Sci* 152:323–339.

Kase R, Javurkova B, Simon E, Swart K, Buchinger S, Konemann S, Escher BI, Carere M, Dulio V, Ait-Aissa S, Hollert H, Valsecchi S, Polesello S, Behnisch P, di Paolo C, Olbrich D, Sychrova E, Gundlach M, Schlichting R, Leborgne L, Clara M, Scheffknecht C, Marneffe Y, Chalon C, Tusil P, Soldan P, von Danwitz B, Schwaiger J, Palao AM, Bersani F, Perceval O, Kienle C, Vermeirssen E, Hilscherova K, Reifferscheid G, Werner I. 2018. Screening and risk management solutions for steroidal estrogens in surface and wastewater. *Trends Analyt Chem* 102:343–358.

Klüver N, Bittermann K, Escher BI. 2019. QSAR for baseline toxicity and classification of specific modes of action of ionizable chemicals in the zebrafish embryo toxicity test. *Aquat Toxicol* 207:110–119.

Könemann S, Kase R, Simon E, Swart K, Buchinger S, Schlüssener M, Hollert H, Escher BI, Werner I, Ait-Assa S, Vermeirssen E, Dulio V, Valsecchi S, Polesello S, Behnisch P, Javurkova B, Perceval O, di Paolo C, Olbrich D, Sychrova E, Schlichting R, Leborgne L, Clara M, Scheffknecht C, Marneffe Y, Chalon C, Tusil P, Soldan P, von Danwitz B, Schwaiger J, San Martin Becares MI, Bersani F, Hilscherova K, Reifferscheid G, Ternes T, Carere M. 2018. Effect-based and chemical analytical methods to monitor estrogens under the European Water Framework Directive. *Trends Analyt Chem* 102:225–235.

König M, Escher BI, Neale PA, Krauss M, Hilscherová K, Novák J, Teodorović I, Schulze T, Seidensticker S, Kamal Hashmi MA, Ahlheim J, Brack W. 2017. Impact of untreated wastewater on a major European river evaluated with a combination of in vitro bioassays and chemical analysis. *Environ Pollut* 220:1220–1230.

Kunz PY, Kienle C, Carere M, Homazava N, Kase R. 2015. In vitro bioassays to screen for endocrine active pharmaceuticals in surface and waste waters. *J Pharm Biomed Anal* 106:107–115.

Kwon JH, Liljestrand HM, Katz LE. 2006. Partitioning of moderately hydrophobic endocrine disruptors between water and synthetic membrane vesicles. *Environ Toxicol Chem* 25:1984–1992.

Leist M, Ghallab A, Graepel R, Marchan R, Hassan R, Bennekou SH, Limonciel A, Vinken M, Schildknecht S, Waldmann T, Danen E, van Ravenzwaay B, Kamp H, Gardner I, Godoy P, Bois FY, Braeuning A, Reif R, Oesch F, Drasdo D, Hohme S, Schwarz M, Hartung T, Braunbeck T, Beltman J, Vrieling H, Sanz F, Forsby A, Gadaleta D, Fisher C, Kelm J, Fluri D, Ecker G, Zdrazil B, Terron A, Jennings P, van der Burg B, Dooley S, Meijer AH, Willighagen E, Martens M, Evelo C, Mombelli E, Taboureau O, Mantovani A, Hardy B, Koch B, Escher S, van Thriel C, Cadenas C, Kroese D, van de Water B, Hengstler JG. 2017. Adverse outcome pathways: Opportunities, limitations and open questions. *Arch Toxicol* 91:3477–3505.

Leusch FDL, Khan SJ, Gagnon MM, Quayle P, Trinh T, Coleman H, Rawson C, Chapman HF, Blair P, Nice H, Reitsema T. 2014. Assessment of wastewater and recycled water quality: A comparison of lines of evidence from in vitro, in vivo and chemical analyses. *Water Res* 50:420–431.

Malaj E, von der Ohe PC, Grote M, Kuehne R, Mondy CP, Usseglio-Polatera P, Brack W, Schaefer RB. 2014. Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. *Proc Natl Acad Sci USA* 111:9549–9554.

Mansouri K, Grulke CM, Judson RS, Williams AJ. 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. *J Cheminform* 10:10.

McCarty LS, Arnot JA, Mackay D. 2013. Evaluation of critical body residue data for acute narcosis in aquatic organisms. *Environ Toxicol Chem* 32:2301–2314.

Neale PA, Altenburger R, Ait-Aissa S, Brion F, Busch W, de Aragão Umbuzeiro G, Denison MS, Du Pasquier D, Hilscherova K, Hollert H, Morales DA, Novac J, Schlichting R, Seiler T-B, Serra H, Shao Y, Tindall AJ, Tollefsen KE, Williams TD, Escher BI. 2017a. Development of a bioanalytical test battery for water quality monitoring: Fingerprinting identified micropollutants and their contribution to effects in surface water. *Water Res* 123:734–750.

Neale PA, Braun G, Brack W, Carmona E, Gunold R, König M, Krauss M, Liebmann L, Liess M, Link M, Schäfer RB, Schlichting R, Schreiner VC, Schulze T, Vormeier P, Weisner O, Escher BI. 2020a. Assessing the mixture effects in in vitro bioassays of chemicals occurring in small agricultural streams during rain events. *Environ Sci Technol* 54:8280–8290.

Neale PA, Feliers C, Glauch L, König M, Lecarpentier C, Schlichting R, Thibert S, Escher BI. 2020b. 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. *Environ Sci (Camb)* 6:2444–2453.

Neale PA, Munz NA, Ait-Aissa S, Altenburger R, Brion F, Busch W, Escher BI, Hilscherova K, Kienle C, Novak J, Seiler T-B, Shao Y, Stamm C, Hollender J. 2017b. Integrating chemical analysis and bioanalysis to evaluate the contribution of wastewater effluent on the micropollutant burden in small streams. *Sci Total Environ* 576:785–795.

Neale PA, O'Brien JW, Glauch L, König M, Krauss M, Mueller JF, Tscharke B, Escher BI. 2020c. Wastewater treatment efficacy evaluated with in vitro bioassays. *Water Res X* 9:100072.

Perkins EJ, Antczak P, Burgoon L, Falciani F, Garcia-Reyero N, Gutsell S, Hodges G, Kienzler A, Knapen D, McBride M, Willett C. 2015. Adverse

outcome pathways for regulatory applications: Examination of four case studies with different degrees of completeness and scientific confidence. *Toxicol Sci* 148:14–25.

Posthuma L, Suter GW, Traas TP. 2002. *Species Sensitivity Distributions in Ecotoxicology*. Lewis, Boca Raton, FL, USA.

US Environmental Protection Agency. 2020. CompTox Chemicals Dashboard. Washington, DC. [cited 2020 August 1]. Available from: comptox.epa.gov/dashboard

van der Oost R, Sileno G, Suarez-Munoz M, Nguyen MT, 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. *Environ Toxicol Chem* 36:2385–2399.

Wang XJ, Hayes JD, Wolf CR. 2006. Generation of a stable antioxidant response element-driven reporter gene cell line and its use to show redox-dependent activation of Nrf2 by cancer chemotherapeutic agents. *Cancer Res* 66:10983–10994.

