A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP6664.

Cytotoxicity Burst? Differentiating Specific from Nonspecific Effects in Tox21 *in Vitro* Reporter Gene Assays

Beate I. Escher,[1,2] Luise Henneberger,[1] Maria König,[1] Rita Schlichting,[1] and Fabian C. Fischer[1†]

[1]Department of Cell Toxicology, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany
[2]Environmental Toxicology, Center for Applied Geoscience, Eberhard Karls University Tübingen, Tübingen, Germany

Background: High-throughput screening of chemicals with *in vitro* reporter gene assays in Tox21 has produced a large database on cytotoxicity and specific modes of action. However, the validity of some of the reported activities is questionable due to the "cytotoxicity burst," which refers to the supposition that many stress responses are activated in a nonspecific way at concentrations close to cell death.

Objectives: We propose a pragmatic method to identify whether reporter gene activation is specific or cytotoxicity-triggered by comparing the measured effects with baseline toxicity.

Methods: Baseline toxicity, also termed narcosis, is the minimal toxicity any chemical causes. Quantitative structure–activity relationships (QSARs) developed for baseline toxicity in mammalian reporter gene cell lines served as anchors to define the chemical-specific threshold for the cytotoxicity burst and to evaluate the degree of specificity of the reporter gene activation. Measured 10% effect concentrations were related to measured or QSAR-predicted 10% cytotoxicity concentrations yielding specificity ratios (SR). We applied this approach to our own experimental data and to $\sim$8,000 chemicals that were tested in six of the high-throughput Tox21 reporter gene assays.

Results: Confirmed baseline toxicants activated reporter gene activity around cytotoxic concentrations triggered by the cytotoxicity burst. In six Tox21 assays, 37%–87% of the active hits were presumably caused by the cytotoxicity burst (SR < 1) and only 2%–14% were specific with SR $\geq$ 10 against experimental cytotoxicity but 75%–97% were specific against baseline toxicity. This difference was caused by a large fraction of chemicals showing excess cytotoxicity.

Conclusions: The specificity analysis for measured *in vitro* effects identified whether a cytotoxicity burst had likely occurred. The SR-analysis not only prevented false positives, but it may also serve as measure for relative effect potency and can be used for quantitative *in vitro–in vivo* extrapolation and risk assessment of chemicals. https://doi.org/10.1289/EHP6664

Introduction

The increasing abundance, number, and heterogeneity of anthropogenic chemicals in our environment call for high-throughput effect screening of chemicals while complying with the strategy to reduce, refine, and replace animal testing (Burden et al. 2015; National Research Council 2007). The application of *in vitro* cell-based bioassays has emerged in the last decade, and their implementation in miniaturized test systems increased the throughput of such bioassays, culminating in collaborative high-throughput screening (HTS) approaches like the Tox21 program (Tice et al. 2013), in which thousands of chemicals were tested in a battery of $\sim$70 *in vitro* bioassays (Attene-Ramos et al. 2013) or ToxCast™, where a smaller number of chemicals were tested in hundreds of cellular bioassays at receptor and pathway level (Hsieh et al. 2017; Judson et al. 2010) but also in cell-free assays (Sipes et al. 2013). The resulting large database is freely available on the U.S. EPA's Chemistry Dashboard (Richard et al. 2016; U.S. EPA 2019) and has been linked to *in vivo* databases (Hu et al. 2015), applied for hazard assessment (Pham et al. 2016; Reif et al. 2010) and quantitative *in vitro–in vivo* extrapolation (QIVIVE) for risk assessment (Sipes et al. 2017; Thomas et al. 2019; Wetmore 2015).

In vitro assays that apply cells stably transfected with reporter genes coupled to the biological receptor of interest are particularly promising tools enabling the classification of chemicals according to their mode of action and/or their potential to disturb biological pathways in humans (Krewski et al. 2020). Upon interaction of a chemical with the receptor, the reporter gene is expressed, triggering the synthesis of reporter proteins and enzymes, e.g., luciferase or β-lactamase, that can be detected by adding appropriate substrates to quantify the enzyme activity. ToxCast™/Tox21 includes reporter gene assays for hormone activity [e.g., estrogen receptors (Filer et al. 2014; Huang et al. 2014) or others (Kleinstreuer et al. 2015)], for activation of metabolic enzymes [e.g., aryl hydrocarbon receptor AhR (Brennan et al. 2015; Murk et al. 1996)] or adaptive stress response assays (Martin et al. 2010).

One important confounding factor of *in vitro* reporter gene activation is the cytotoxicity of the dosed chemicals. Cytotoxicity has been assessed in parallel to reporter gene activation in many of the assays in Tox21. Time-dependence of cytotoxicity was assessed for two cell lines (HepG2 and HEK293) using metabolic activity or loss of cell membrane integrity as a measure of cytotoxicity (Hsieh et al. 2017). Kinetics of cytotoxicity depended on the underlying pathway triggering cytotoxicity with activation of γ-H2AX, which is related to genotoxicity pathways being particularly fast, followed by those associated with mitochondrial disruption (Hsieh et al. 2017). However, no quantitative picture emerged. In our experience the typically used cell viability assays using fluorometric methods to detect metabolic activity or loss of cell membrane integrity are prone to artifacts and, as observed by Hsieh et al. (2017), produce inconsistent results. We are therefore routinely screening cytotoxicity during reporter gene assays measurements by assessing confluence with a cell imager directly after dosing and prior to activity measurement of the reporter protein (Escher et al. 2019). Although, for most chemicals, this imaging method provided effect concentrations similar to those of the fluorometric cell viability assay, artifacts of apparent high metabolic activity while cells have already disappeared under the microscope can be avoided.

Address correspondence to Beate I. Escher, Department Cell Toxicology, Helmholtz Centre for Environmental Research–UFZ, Permoserstraße 15, 04318 Leipzig, Germany. Telephone: 0049 341 235 1244. Fax: 0049 341 235 45 1244. Email: beate.escher@ufz.de

[†]Current address: National Institute for Environmental Studies (NIES), Center for Health and Environmental Risk Research, Onogawa 16-2, 305-8506 Tsukuba, Ibaraki, Japan.

Supplemental Material is available online (https://doi.org/10.1289/EHP6664).

The authors declare they have no actual or potential competing financial interests.

Received 11 December 2019; Revised 16 June 2020; Accepted 2 July 2020; Published 23 July 2020.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

The minimal toxicity any chemical can elicit is baseline toxicity (or narcosis) that results from disturbance of structure and functioning of the cell membrane by the presence of chemicals in the membrane (van Wezel and Opperhuizen 1995). Many functions are lost in response to baseline toxicity but mitochondria are especially affected and lose their ability for energy transduction leading to ATP depletion as membranes become permeable (Vinken and Blaauboer 2017). Lipophilic chemicals that have high sorption affinities to phospholipid membranes trigger baseline toxicity at lower dosed concentrations than hydrophilic chemicals, but the cytotoxic concentrations in the cell membranes do not differ much between different chemicals causing baseline toxicity (van Wezel and Opperhuizen 1995). We confirmed for eight reporter gene cell lines that chemicals triggered baseline toxicity when reaching a critical membrane concentration of approximately 70 mmol $\times$ L_{lip}^{-1} (Escher et al. 2019). We developed quantitative structure–activity relationships (QSARs) for these cell lines to predict the 10% inhibitory concentrations (IC_{10}) based on one chemical parameter, the liposome–water partition constant ($K_{lip/w}$) (Escher et al. 2019). Known baseline toxicants (Vaes et al. 1998) were used to derive the QSARs. The QSARs were very similar across cell lines with differences caused by the assays conditions, mainly the serum content of the medium (Escher et al. 2019).

Close to cell death, the cells activate many cellular signaling pathways hence the exposure to high concentrations of chemicals may lead to a nonspecific activation of the reporter gene, a phenomenon termed "cytotoxicity burst" (Judson et al. 2016). Judson et al. (2016) developed a statistical approach to identify cytotoxicity burst in large *in vitro* platforms that include thousands of data points. Another measure of selectivity of chemicals was the number of assays activated by a chemical and the ratio of the effect concentration of the most sensitive assay to the 10th percentile of the distribution of all assays (Thomas et al. 2013). Fay et al. (2018) recently refined the diagnosis of the cytotoxicity burst phenomenon. They proposed a diagnostic odds ratio that differentiates assays that respond in the range of cytotoxic concentrations from those that are active at much lower concentrations and hence true responders. They also related the predicted baseline toxicity concentrations from Fischer et al. (2017) with the threshold for the cytotoxicity burst and concluded that the cytotoxicity burst phenomenon is more complex than baseline toxicity.

The goal of the present study was to quantify the degree of specificity of a chemical for each reporter gene assay directly from the experimental effect data without the need to analyze them in context with other chemicals and bioassays. We hypothesized that if the reporter gene activation occurs at concentrations close to baseline toxicity, it is likely not a specific effect but that the effect resulted from the cytotoxicity burst. To test this hypothesis, we measured the cytotoxicity and reporter gene activation of seven confirmed baseline toxicants and eight additional chemicals of environmental and toxicological relevance in eight standardized and widely used *in vitro* reporter gene assays. For chemicals that are more cytotoxic than baseline toxicity, the relationship between the receptor or pathway affected and the degree of cytotoxicity enhancement over baseline toxicity was investigated. Aiming to capture the big picture in an HTS *in vitro* platform, we applied the specificity analysis to the Tox21 *in vitro* reporter gene database to evaluate which role the cytotoxicity burst may play.

Materials and Methods

Experimental Study Chemicals

We selected seven confirmed baseline toxicants (2-Phenylphenol, 3-Nitroaniline, 4-Chloro-3-methylphenol, 4-Pentylphenol, 2-Allylphenol, 2-Butoxyethanol, 2,4,5-Trichloroaniline) that had been used to set up baseline toxicity QSARs for eight reporter gene cell lines (Escher et al. 2019) and seven additional chemicals that were frequently found in environmental samples (Bisphenol A, Quinoxyfen, Fluoranthene) or in foodstuff (Genistein, Coumarin, 8-Gingerol, Zingerone). Their names and physicochemical properties are listed in Table S1. None of the chemicals tested were below the volatility cutoff of a medium–air partition constant of 10,000 (Escher et al. 2019).

The experimental liposome–water partition constants ($K_{lip/w}$) of the seven baseline toxicants were taken from Vaes et al. (1997). In the second set, the $K_{lip/w}$ stem from Kwon et al. (2006) and van der Heijden and Jonker (2009). We included ionizable chemicals that dissociate into a neutral ($f_{neutral}$) and ionized fraction ($f_{ionized}$) in the *in vitro* assay medium at pH 7.4. From the acidity constants pK_a as well as the $K_{lip/w}$ of the neutral and charged species (Table S1), ionization-corrected liposome–water distribution ratios [$D_{lip/w}$ (pH 7.4)] can be calculated (Equation 1) or directly measured (Henneberger et al. 2019, 2020).

$$\log D_{lip/w}\,(\mathrm{pH}\,7.4) = f_{neutral} \times \log K_{lip/w}(\mathrm{neutral}) + (1\text{-}f_{neutral}) \times \log K_{lip/w}(\mathrm{ion}) \quad (1)$$

Reporter Gene Assays

Eight reporter gene assays (Table 1) were performed as described in (König et al. 2017; Neale et al. 2017). The cell confluency served as surrogate for cell viability as previously described (Escher et al. 2019).

Briefly 2,500 to 5,000 cells in 30 μL medium were plated in each well of a black 384-well polystyrene microtiter plate with clear bottom (AREc32 #3764; all other cell lines BioCoat #356663, Corning), leaving the last column as control without cells for the GeneBLAzer cell lines, and incubated for 24 h at 37°C, 5% CO_2 to let the cells attach. All medium components were purchased from Gibco. Media were 90% DMEM + GlutaMAX plus 10% FBS and 100 U/mL penicillin and 100 μg/mL streptomycin for AhR-CALUX and AREc32 cells, 90% phenol red–free DMEM, 10% dialyzed FBS, 0.1 mM NEAA, 25 mM HEPES, 1 mM sodium pyruvate, 100 U/mL penicillin and 100 μg/mL streptomycin and 4 mM GlutaMAX for ARE-BLA and 98% Opti-MEM supplemented with 2% charcoal-stripped FBS 100 U/mL penicillin and 100 μg/mL streptomycin for all other GeneBLAzer cell lines. During the initial 24 h, the cell number did not increase visibly, but cells attached (Escher et al. 2019). Plated cell numbers were adjusted between 2,500 to 5,000 cells per well depending on the cell line that the confluency was ~30% to 50% prior to dosing and no more than 80% after 24 h of exposure (Escher et al. 2019). Before dosing and after additional 24 h of exposure, the cell confluency was measured with an IncuCyte S3 live cell imaging system (Essen BioScience).

The detection of activation in the reporter gene assays AREc32 and AhR-CALUX encoding for luciferase and the GeneBLAzer assays encoding for β-lactamase was performed as previously described (Escher et al. 2012; König et al. 2017; Neale et al. 2017). Briefly, after 24 h of exposure and measurement of cell confluency, AREc32 and AhR-CALUX cells were washed twice with PBS. Subsequently 10 μL of lysis buffer per well was added [25 mM Tris (AppliChem, A13790500), 1% Triton-X 100 (Geyer Chemsolute, 8059), 2 mM EDTA (AppliChem, A11040500), 2 mM DTT (Sigma-Aldrich, D0632), 10% glycerol (AppliChem, A11231000)], followed by a 15–20 min incubation at room temperature (RT) and shaking at 1,500 rpm to allow complete lysis of cells. Afterward, 40 μl of luciferase substrate buffer [20 mM Tricine (Sigma-Aldrich, T0377), 2.67 mM MgSO4 (AppliChem, 131404.1210), 33.3 mM DTT (Sigma-Aldrich, D0632), 0.1 mM EDTA (AppliChem,

Table 1. Reporter gene cell lines applied in this study and baseline toxicity QSAR (Escher et al. 2019).

Reporter gene cell line	Derived from	Slope of baseline toxicity QSAR	Intercept of baseline toxicity QSAR	Corresponding assay in Tox21 for reporter gene activation[a]	Corresponding cytotoxicity assay in Tox21[a]
AREc32	MCF7	0.56 ± 0.09	1.76 ± 0.28	(not in Tox21)	(not in Tox21)
ARE-BLA	HepG2	0.68 ± 0.08	1.19 ± 0.26	TOX21_ARE_BLA_Agonist_ratio	TOX21_ARE_BLA_Agonist_viability
AhR-CALUX (H4L7.5c2)	H4IIe	0.73 ± 0.10	1.28 ± 0.31	TOX21_AhR_LUC_Agonist[b]	TOX21_AhR_LUC_Agonist_viability[b]
PPARγ-BLA	HEK293H	0.64 ± 0.20	1.71 ± 0.69	TOX21_PPARg_BLA_Agonist_ratio	TOX21_PPARg_BLA_Antagonist_viability
AR-BLA	HEK293T	0.76 ± 0.14	1.44 ± 0.49	TOX21_AR_BLA_Agonist_ratio	TOX21_AR_BLA_Antagonist_viability
ERα-BLA	HEK293T	0.76 ± 0.10	1.54 ± 0.42	TOX21_ERa_BLA_Agonist_ratio	TOX21_ERa_BLA_Antagonist_viability
PR-BLA	HEK293T	0.70 ± 0.16	1.52 ± 0.11	(not included)	(not included)
GR-BLA	HEK293T	0.72 ± 0.13	1.67 ± 0.42	TOX21_GR_BLA_Agonist_ratio[c]	TOX21_GR_BLA_Antagonist_viability[c]

Note: AhR, aryl hydrocarbon receptor; AR, androgen receptor; ARE, antioxidant response element; BLA, GeneBLAzer reporter gene cell line; CALUX, Chemical Activated LUciferase gene eXpression; ER, estrogen receptor; GR, glucocorticoid receptor; PPARγ, peroxisome proliferator activated receptor gamma; PR, progesterone receptor; QSAR, Quantitative Structure Activity Relationship.
[a]As referenced in https://comptox.epa.gov/dashboard.
[b]In Tox21 a HepG2-based AhR-CALUX cell line was used (He et al. 2011).
[c]In Tox21 a HeLa-derived cell line was used.

A11040500), 0.261 mM coenzyme A (Sigma-Aldrich, C3144), 0.53 mM ATP (Sigma-Aldrich, A2383), 0.235 mM D-luciferin (AREc32), and 0.059 mM D-luciferin (AhR) (AAT Bioquest, ABD-12506)] was added, and luminescence was read with a Tecan Infinite® M1000 plate reader. For the detection of the activity of the β-lactamase in the GeneBLAzer® assays 8 µl of ToxBLAzer™ substrate (ThermoFisher Scientific) were added to each well of the plate, and fluorescence was measured immediately to allow elimination of autofluorescence and after 2 h incubation at RT for reporter gene activation using a Tecan Infinite® M1000 plate reader. The ToxBLAzer™ substrate allowed a ratiometric detection of the reporter gene activity by measuring the emission of blue (460 nm) and green (530 nm) light using the M1000 plate reader.

Dosing of Chemicals

Chemical stocks dissolved in DMSO or neat liquid chemicals were dosed into medium at 4 × concentrations on 384-well plates with a Tecan D300e Digital Dispenser (Tecan) and then 10 µl per well of chemicals in medium were transferred with a 96-pipette head (Hamilton Microlab Star) into the cell plates that contain cells as described above in 30 µL medium. The chemicals were dosed in final concentrations up to three times their predicted IC_{10} for baseline toxicity (Fischer et al. 2019), with the dose range depicted in the concentration–response curves (Figures S1–S15) with different symbols for each independent experiment ($n \geq 3$).

Data Evaluation

Cytotoxicity was expressed as percent inhibition of cell viability in comparison with unexposed cells (ratio of confluency of exposed to confluency of unexposed cells). The inhibitory concentration for 10% cytotoxicity, IC_{10} (Equation 2) was determined from the linear range of the concentration–cytotoxicity curves (% cytotoxicity = slope × concentration) as described previously (Escher et al. 2018).

$$IC_{10} = \frac{10\%}{\text{slope}} \tag{2}$$

The effect concentrations EC_{10} were calculated analogously from the linear concentration–effect curves (% effect = slope × concentration) with Equation 3.

$$EC_{10} = \frac{10\%}{\text{slope}} \tag{3}$$

For AREc32 and ARE-BLA, the effect concentration causing an induction ratio (IR) of 1.5, $EC_{IR1.5}$, was derived from the linear regression of the IR against the concentration (IR = 1 + slope × concentration) for IR < 4 (Escher et al. 2018).

$$EC_{IR1.5} = \frac{0.5}{\text{slope}} \tag{4}$$

Baseline Toxicity QSARs

Baseline toxicity QSARs of the form given in Equation 5 previously developed for the eight cell lines [(Escher et al. 2019), Table 1] were used to predict the IC_{10} for baseline toxicity ($IC_{10,\text{baseline}}$) of the study chemicals using the ionization-corrected log $D_{\text{lip/w}}$ (pH 7.4) calculated by Equation 1.

$$\log[1/IC_{10,\text{baseline}}\,(\text{M})] = \text{slope} \times \log D_{\text{lip/w}}(\text{pH } 7.4) + \text{intercept} \tag{5}$$

We replaced the $K_{\text{lip/w}}$ of the neutral species in the original QSAR by the ionization-corrected $D_{\text{lip/w}}$ (pH 7.4) in Equation 5 to include also charged chemicals. The expansion of QSARs for neutral chemicals to ionizable chemicals was previously described for bacteria (Escher et al. 2017) and the zebrafish embryo toxicity test (Klüver et al. 2019). A potential ion-trapping effect does not need to be accounted for, provided the pH does not deviate much from pH 7.4, because ion-trapping becomes relevant only if the intracellular pH and the extracellular pH differ by more than one pH unit (Escher et al. 2020).

Specificity Analysis

The toxic ratio (TR, Equation 6) is a measure of enhanced cytotoxicity, i.e., how much more potent a chemical is in comparison with its predicted baseline toxicity.

$$TR = \frac{IC_{10,\text{baseline}}}{IC_{10}} \tag{6}$$

For TR, it is commonly accepted that a TR ≥ 10 is associated with specific or reactive toxicity (Maeder et al. 2004), and any chemical with TR < 10 is considered a baseline toxicant.

We defined the specificity ratio as the ratio between cytotoxicity (IC_{10}) and effect concentration (EC_{10} or $EC_{IR1.5}$). The specificity ratio can relate either to the experimental IC_{10} ($SR_{\text{cytotoxicity}}$, Equation 7) or to the predicted $IC_{10,\text{baseline}}$ (SR_{baseline}, Equation 8), as conceptually illustrated in Figure 1.

$$SR_{\text{cytotoxicity}} = \frac{IC_{10}}{EC_{10}} \quad \text{or} \quad SR_{\text{cytotoxicity}} = \frac{IC_{10}}{EC_{IR1.5}} \tag{7}$$

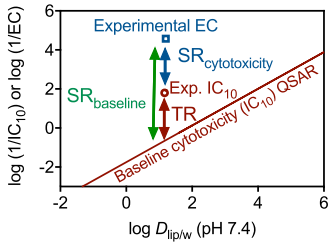

Figure 1. Conceptual illustration of the proposed specificity analysis framework. The line corresponds to the Quantitative Structure Activity Relationship (QSAR) for baseline toxicity. The experimental effect concentration [log (1/EC)] is depicted by a blue square; the experimental inhibitory concentration [log $(1/IC_{10})$] leading to 10% cytotoxicity is depicted by a red circle. The distance between the log $[1/IC_{10}$ (QSAR)] and the experimental log $(1/IC_{10})$ is the toxic ratio log TR. The distance between the experimental log $(1/IC_{10})$ and the experimental log (1/EC) is the specificity ratio log $SR_{cytotoxicity}$. The distance between the log $[1/IC_{10}$ (QSAR)] and the experimental log (1/EC) is the specificity ratio log $SR_{baseline}$. For better legibility, log(1/y) was omitted in the graph, but all measures are in form of negative logarithms. Note: EC, effect concentration; exp., experimental; QSAR, Quantitative Structure Activity Relationship; SR, specificity ratio; TR, toxic ratio.

$$SR_{baseline} = \frac{IC_{10,baseline}}{EC_{10}} \quad \text{or} \quad SR_{baseline} = \frac{IC_{10,baseline}}{EC_{IR1.5}} \tag{8}$$

In analogy with the threshold for TR, we considered $SR \leq 1$ as not specific, $1 \leq SR < 10$ as moderately specific (with high uncertainty), $10 \leq SR < 100$ as specific, and $100 \leq SR$ as highly specific.

Tox21 Data Extraction and Processing

Concentration–effect data for the 8,628 chemicals from the Tox21 10K library with available chemical identifiers (CAS number and DSSTox ID) were extracted for ARE-BLA, AhR-CALUX, PPARγ-BLA, AR-BLA, ERα-BLA, and GR-BLA (see Table 1 for Tox21 assay notation). Reporter gene activation and cell viability were downloaded from the Tox21 Concentration-Response Browser as percent of the maximum effect of the positive control (=100%) (https://sandbox.ntp.niehs.nih.gov/tox21-curve-visualization, last accessed 10 June 2019). No data for AREc32 were available in Tox21. The database includes effect data from the U.S. EPA, the National Toxicity Program, and the Federal Drug Administration laboratories. In Tox21, chemicals were partly tested multiple times in one or more laboratories, and provided that the same chemical ID was tested (equal vendor and purity), the concentration–effect data were merged and reevaluated using the same linear concentration–response analysis (Escher et al. 2018) as that used for the experimental data measured for this study, yielding single IC_{10} and EC_{10} for each chemical and assay. In the BLA-bioassays, cell viability was quantified via adenosine triphosphate (ATP) with CellTiter-Glo, and the decrease of the signal served as measure of cytotoxicity. In AhR-CALUX, the cytotoxicity was quantified with CellTiter-Fluor based on protease activity. Note that cell viability was measured by microscopy in our own experiments; thus differences in the sensitivity between the two cytotoxicity measurement methods may represent a source of error. For the Tox21 ARE-BLA, PPARγ-BLA, AR-BLA, and ERα-BLA, the reporter gene constructs, cell lines, and specific effect measurement techniques were the same as those used in our experiments. AhR agonism was measured by CALUX luciferase quantification in both Tox21 and our experiments, but Tox21 used the HG2L7.5c1 cells derived from HepG2 (He et al. 2011), whereas we applied H4L7.5c2 cells (Brennan et al. 2015). Another difference was that we used IR as an effect measure for ARE-BLA. For GR-BLA, we used the construct based on HEK293T, and Tox21 applied the HeLa-derived GR-BLA. Cytotoxicity data were available for ARE-BLA and AhR-CALUX, but for the other cell lines cytotoxicity was derived from the antagonistic assays. This difference can be justified because the potent agonists were added at very low constant concentrations, where they did not cause any cytotoxicity.

The data were fitted by linear regression as described in (Escher et al. 2018) using MATLAB R2018a with the code detailed in Text S1 and S2. EC_{10} and IC_{10} for all chemicals in the six assays were calculated by Equations 2 and 3, and standard errors were derived according to Escher et al. (2018). Concentrations triggering >30% effect and >50% cytotoxicity were excluded from the fit, because linearity was observed only for the lower portion of concentration–response curves in *in vitro* reporter gene assays (Escher et al. 2018). Chemicals with EC_{10} and $IC_{10} > 0$ and <their maximum test concentration were classified active and/or cytotoxic. EC_{10} and IC_{10} with relative standard errors >50% were classified "inconclusive" and excluded from the specificity analysis. The remaining EC_{10} and IC_{10} were analyzed for their TR, $SR_{cytotoxicity}$, and $SR_{baseline}$ by Equations 6–8.

The $IC_{10,baseline}$ of all Tox21 chemicals were predicted with the baseline toxicity QSARs (Table 1). The log $K_{lip/w}$ of the neutral species were predicted by a log K_{ow}-based QSAR [(Endo et al. 2011), Equation 9] and the log $D_{lip/w}$ (pH 7.4) with Equation 10 that was derived from Equation 1 by assuming 10 times lower affinity of the ionized species to phospholipid liposomes (Bittermann et al. 2016; Escher et al. 2020). The fraction of the neutral and ionized species was calculated with the Henderson-Hasselbalch equation from the acidity constants pK_a, which were estimated with ACD/Percepta pK_a using the GALAS algorithm (www.acdlabs.com/software/pka/).

$$\log K_{lip/w} = 1.01 \times \log K_{ow} + 0.12 \tag{9}$$

$$\log D_{lip/w} = f_{neutral} \times \log K_{lip/w} + f_{ionized} \times (\log K_{lip/w} - 1) \tag{10}$$

Only $IC_{10,baseline}$ for chemicals with $1 < \log D_{lip/w}\,(\text{pH } 7.4) < 5$ were predicted; the others were marked "outside QSAR domain." The entire workflow of data analysis of the Tox21 data is outlined in the SI, Figure S18.

Results

Cytotoxicity and Reporter Gene Activation of Baseline Toxicants

All seven evaluated baseline toxicants triggered cytotoxicity in the eight reporter gene assays and the IC_{10} were consistent with those previously reported (Escher et al. 2019). The TR of the chemicals in the assays were all within one order of magnitude around 1 ($0.1 \leq TR < 10$); hence, the measured IC_{10} were similar to the $IC_{10,baseline}$ predicted by the baseline toxicity QSARs (Figure 2A).

We previously proposed that all concentrations above the IC_{10} must be omitted before analyzing reporter gene activation (Escher et al. 2018). Here, we have purposely violated this rule to extrapolate the $EC_{IR1.5}$ and EC_{10} whenever possible, and there are a few examples, where even after the IC_{10} cutoff, valid activation would have been detected. As we measured both cytotoxicity and reporter gene activation for the baseline toxicants, $SR_{cytotoxicity}$ were calculated according to Equation 7. In AREc32, all confirmed baseline toxicants activated the oxidative stress response (Figure 2B and Figure S1), albeit with very low

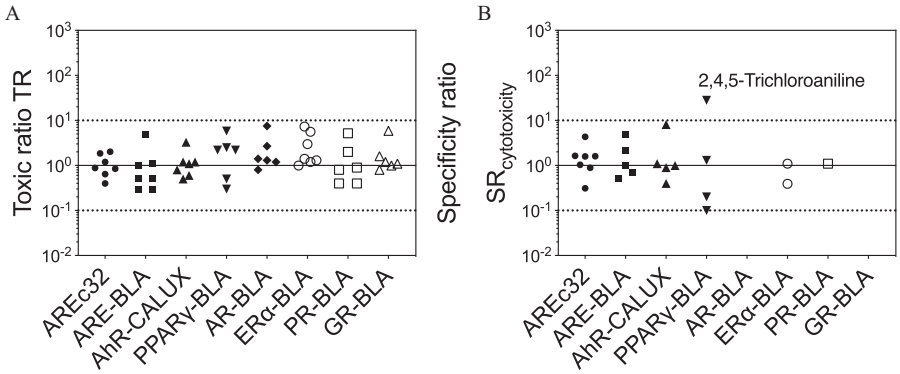

Figure 2. (A) Toxic ratios and (B) specificity ratios of the baseline toxicants (2-Phenylphenol, 3-Nitroaniline, 4-Chloro-3-methylphenol, 4-Pentylphenol, 2-Allylphenol, 2-Butoxyethanol, 2,4,5-Trichloroaniline) in eight reporter gene assays. The solid line is a $SR_{cytotoxicity}$ of 1, and the dashed lines are the thresholds of 10 and 0.1. If the effects did not exceed 10%, then no IC_{10} and/or EC_{10} could be derived, and there is no symbol in the figure. Underlying data are in Tables S2–S9.

$SR_{cytotoxicity}$ ranging from 0.3 to 4.3 (Table S2). Only five chemicals activated ARE-BLA (Figure S2) but did not exceed the SR threshold of 10 [$SR_{cytotoxicity}$ 0.5 to 5.0 (Table S3)], and five chemicals activated AhR-CALUX (Figure S3) with $SR_{cytotoxicity}$ 0.4 to 8.1 (Table S4). In PPARγ-BLA (Figure S4), 2,4,5-Trichloroaniline showed a specific effect that appeared not to be influenced by cytotoxicity with a $SR_{cytotoxicity}$ of 28 (Figure 2B and Table S5). No $SR_{cytotoxicity}$ could be deduced for the hormone receptors apart from 4-Chloro-3-methylphenol and 4-Pentylphenol in ERAα-BLA (Tables S6–S9); still low activity was recorded at cytotoxic concentrations (Figures S4–S8).

Cytotoxicity and Reporter Gene Activation of Environmental Chemicals

To further explore the cytotoxicity burst, we selected seven environmentally relevant chemicals with diverse physicochemical properties, five of which (Bisphenol A, Quinoxyfen, Fluoranthene, Genistein, Coumarin) overlapped with and were reported to be active in the Tox21 reporter gene assays corresponding to the assays performed in this study (https://comptox.epa.gov/dashboard, Tables S1–S6). The concentration–response curves in all eight reporter gene assays are depicted in Figures S9–S16, and IC_{10} and effect concentrations are listed in the corresponding Tables S10–S17. When inserting the experimental IC_{10} into the figures of all baseline QSARs (Figure S19), visual inspection indicated that the more hydrophobic chemicals were often below the baseline, which could be an artifact due to sorptive loss processes or degradation. The QSARs were developed with a test set of chemicals, whose log $K_{lip/w}$ ranged from 0.60 (2-Butoxyethanol) to 4.31 (4-Pentylphenol), but three of the environmental chemicals (Quinoxyfen, 8-Gingerol and Fluoranthene) exceeded this range (Table S1). The TR analysis of the IC_{10} for cytotoxicity (Tables S10–S17, Figure 3A) revealed that apart from one single outlier (Genistein in PR-BLA with a TR of 15), all environmental chemicals caused baseline toxicity in the cytotoxicity end point.

As expected for known estrogen agonists, Genistein ($SR_{cytotoxicity} = 118$, $SR_{baseline} = 441$) and Bisphenol A ($SR_{cytotoxicity} = 98$, $SR_{baseline} = 282$) were highly specific in the ERα-BLA (Figure 3B) and because the TR were <10, $SR_{cytotoxicity}$ and $SR_{baseline}$ were in the same range. Unexpected SR ≥10 were found for Genistein in AREc32 ($SR_{cytotoxicity} = 38$) and Quinoxyfen in PPARγ-BLA ($SR_{cytotoxicity} = 25$). All other chemicals did not show activation of the reporter gene or had $SR_{cytotoxicity} < 10$ indicative for moderate or nonspecific effects. Bisphenol A had a $SR_{cytotoxicity}$ of 1–1.5 and $SR_{baseline}$ of 1.2–4 in AREc32, ARE-BLA, and AhR-CALUX, which indicates that these activations are of low specificity, and the estrogenic effect was the only true pathway among the tested assays.

Specificity Analysis of Tox21 Effect Data

The corresponding cell lines in Tox21 are identical with exception of AhR and GR-BLA (Table 1). Table 2 summarizes the output of the toxic ratio analysis and Tables 3 and 4 the specificity analysis for the six Tox21 reporter gene assays. Individual results are in the Tables S1–S6. The number of chemicals included in the $SR_{cytotoxicity}$ analysis was limited to the availability of chemicals, for which both EC_{10} and IC_{10} were measured. Because a relatively narrow and constant concentration range was measured for all chemicals in Tox21, a large proportion of the chemicals were not tested up to baseline cytotoxic concentrations; therefore, false negative counts are possible.

The $SR_{baseline}$ and TR analyses included more chemicals than the $SR_{cytotoxicity}$ analysis because the QSAR-predicted $IC_{10,baseline}$ were used as an anchor for baseline-associated cytotoxicity. The $IC_{10,baseline}$ could be derived for approximately 60% of the chemicals (59% for ARE-BLA, 57% for AhR-CALUX, 60% for PPARγ-BLA, 57% for AR-BLA, 57% for ERα-BLA, and 57% for GR-BLA), the remainder was outside the applicability domain of the baseline toxicity QSAR of $1 < \log D_{lip/w}$ (pH 7.4) < 5.

Figure 4A and Table 2 report the TR values for all cytotoxic chemicals in the Tox21 reporter gene assays. Approximately 50% of the experimental IC_{10} values were in the range of one order of magnitude around $IC_{10,baseline}$, hence, the TR < 10 and measured cytotoxicity were baseline-associated. Both neutral and ionizable chemicals were included in this analysis. The baseline toxicity QSAR can in principle be extended to ionizable chemicals by replacing the log $K_{lip/w}$ with the $D_{lip/w}$ (pH 7.4) corrected for ionization, but this has not yet been demonstrated for mammalian cell lines. We therefore split the data set in two subsets of predominantly neutral ($f_{neutral} > 98\%$) and (partially) charged ($f_{neutral} < 98\%$) chemicals, and the resulting TR analysis is depicted in Figure S20 and Table S18. Further, 44% to 61% of chemicals were (partially) charged (Table S18). There was little difference between the TR ranges of neutral and charged chemicals (Figure S20), with slightly more chemicals with TR > 10 for the charged chemicals, which can be explained by the fact that some specific modes of action, such as uncoupling, require charged chemicals. Therefore, neutral and charged chemicals were evaluated together.

The other 50% of the cytotoxic chemicals triggered cytotoxicity at lower concentrations than predicted by the baseline

Table 2. Toxic ratio TR analysis for the six Tox21 *in vitro* reporter gene assays.

Reporter gene assay	n (chemicals)[a]	n (TR)[b]	n (TR > 10)[c]		n (1 ≤ TR ≤ 10)[c]		n (TR < 1)[c]	
ARE-BLA	7,214	694	384	(55.3%)	275	(39.6%)	35	(5.0%)
AhR-CALUX	7,968	930	524	(56.3%)	353	(38.0%)	52	(5.6%)
PPARγ-BLA	6,968	723	326	(45.1%)	349	(48.3%)	48	(6.6%)
AR-BLA	7,959	1,172	485	(41.4%)	532	(45.4%)	155	(13.2%)
ERα-BLA	7,947	610	193	(31.6%)	266	(43.6%)	151	(24.8%)
GR-BLA	7,963	599	204	(34.1%)	277	(46.2%)	118	(19.7%)

Note: The underlying data is given in Tables S1–S6.
[a]Number (n) of the evaluated chemicals [n (chemicals)].
[b]Number of chemicals n for which a TR could be derived n (TR).
[c]Binning into the TR categories (TR > 10, 1 ≤ TR ≤ 10, TR < 1). In parentheses are the percentages of categories in each bin.

QSARs. Interestingly, chemicals triggered cytotoxicity at lower concentrations than their predicted $IC_{10,baseline}$, especially in ARE-BLA and AhR-CALUX, with 55% and 56% of cytotoxicity data exceeding the TR = 10 threshold.

In addition, 5% to 25% of the chemicals with experimental IC_{10} and matching $IC_{10,baseline}$ had TR < 1 (Table 2). These numbers reflect the uncertainty and variability of the assay results because in theory a true baseline toxicant has TR = 1. Deviation to higher TR can be related to specific cytotoxicity, but those below 1 must reflect uncertainty as well as measurement artifacts.

$SR_{cytotoxicity}$ were <10 for a large proportion of the chemicals, for which both, IC_{10} and EC_{10} were reported (Figure 4B) and, apart from ARE-BLA, the majority of those were even $SR_{cytotoxicity} < 1$ (Table 3). Taking all assays together, only 9% of the active chemicals had a $SR_{cytotoxicity} > 10$ and were classified as specifically active. The highest proportion of specifically active chemicals was found for AR-BLA (14%), with another 21% in the range of $1 \leq SR_{cytotoxicity} \leq 10$, which still means that 65% of all chemicals reported as activating the androgen receptor triggered cytotoxicity at concentrations similar to that of reporter gene activation. For GR-BLA, only 2 of the 121 active chemicals exceeded the range of cytotoxicity by a factor of 10. In PPARγ-BLA, only troglitazone ($SR_{cytotoxicity} = 1{,}742$) and in GR-BLA ciclesonide ($SR_{cytotoxicity} = 3{,}390$) were highly specific, whereas the other chemicals were either of low specificity (16% and 12%, respectively, with $1 < SR_{cytotoxicity} < 10$) or did not exceed $SR_{cytotoxicity}$ of 1. Sixteen (8%) and three (1.8%) chemicals were highly specific in the AR-BLA and ERα-BLA assays, respectively, with $SR_{cytotoxicity} > 1{,}000$. All were well-known agonists of these hormone receptors. Contrarily, no chemicals exceeded a $SR_{cytotoxicity}$ of 1,000 in the ARE-BLA and AhR-CALUX assays.

The $SR_{baseline}$ analysis indicated that a larger proportion of the evaluated Tox21 chemicals were specifically active in the reporter gene assays (Table 3 and Figure 4C). Taking all chemicals together, ~60% of the active chemicals in the assays were specific with $SR_{baseline} > 10$ while 23% to 38% fell into the $1 \leq SR_{baseline} \leq 10$ range of $IC_{10,baseline}$ and were classified as moderately specific, with fewer than 25% having $SR_{baseline} < 1$. As for the $SR_{cytotoxicity}$ analysis, more chemicals were highly specific in the hormone receptor assays AR-BLA and ERα-BLA, in a few cases even exceeding an $SR_{baseline}$ of $>10^6$. Estriol, a metabolite of estradiol and estrone and known to be a good ER-ligand, had an $SR_{baseline}$ of 45,000 in ERα-BLA and fluorometholone, a synthetic glucocorticoid, had an $SR_{baseline} > 106$ in AR-BLA. For ARE-BLA and AhR-CALUX, the range of specific chemicals were narrow, with only very few chemicals exceeding an $SR_{baseline}$ of 10^4.

Discussion

Specific Toxicity

We found that 32% to 56% of all chemicals in Tox21 had TR > 10 and could therefore be classified as specifically acting. This is a higher proportion than in the experimental data set of 15 compounds, where most chemicals were classified as baseline toxicants, with 7 of them specifically selected because they are known to be baseline toxicants (Vaes et al. 1998). However, when comparing with classification of a large set of ecotoxicity data using different mode-of-action classification tools (Kienzler et al. 2017), depending on the method, 27% to 69% of chemicals were assigned as baseline toxicants. The remainder are not necessarily expected to have a TR > 10 but also included chemicals that could not be assigned to a mode-of-action class. There exist no such estimates for cytotoxicity, but the ranges in the present analysis seem realistic in comparison with the analysis of ecotoxicity data (Kienzler et al. 2017).

Identification of the Cytotoxicity Burst

Some of the reporter gene assays could be activated by the known baseline toxicants, but, as expected from baseline toxicants, the $SR_{cytotoxicity}$ ranged within an order of magnitude around one (Figure 2B), with the only exception being 2,4,5-Trichloroaniline. Thus, the proposed classification ($1 \leq SR < 10$ moderately specific, $10 \leq SR < 100$ specific, and $100 \leq SR$ highly specific) appeared reasonable because the confirmed baseline toxicants fell into the range of SR < 10. Of course, it is possible that the initial

Table 3. Specificity ratio $SR_{cytotoxicity}$ analysis for the six Tox21 *in vitro* reporter gene assays.

Reporter gene assay	n (chemicals)[a]	n ($SR_{cytotoxicity}$)[b]	n ($SR_{cytotoxicity} > 10$)[c]		n ($1 \leq SR_{cytotoxicity} \leq 10$)[c]		n ($SR_{cytotoxicity} < 1$)[c]	
ARE-BLA	7,214	332	29	(8.7%)	180	(54.2%)	123	(37.0%)
AhR-CALUX	7,968	305	29	(9.5%)	115	(37.7%)	161	(52.8%)
PPARγ-BLA	6,968	75	2	(2.7%)	12	(16.0%)	61	(81.3%)
AR-BLA	7,959	204	29	(14.2%)	42	(20.6%)	133	(65.2%)
ERα-BLA	7,947	181	13	(7.2%)	46	(25.4%)	122	(67.4%)
GR-BLA	7,963	121	2	(1.7%)	14	(11.6%)	105	(86.8%)

Note: The underlying data is given in Tables S1–S6.
[a]Number (n) of the evaluated chemicals [n (chemicals)].
[b]Number of chemicals n for which a $SR_{cytotoxicity}$ could be derived n ($SR_{cytotoxicity}$).
[c]Binning into the $SR_{cytotoxicity}$ categories ($SR_{cytotoxicity} > 10$, $1 \leq SR_{cytotoxicity} \leq 10$, $SR_{cytotoxicity} < 1$). In parentheses are the percentages of categories in each bin.

Table 4. Specificity ratio $SR_{baseline}$ analysis for the six Tox21 *in vitro* reporter gene assays.

Reporter gene assay	n (chemicals)[c]	n ($SR_{baseline}$)[b]	n ($SR_{baseline} > 10$)[c]		n ($1 \leq SR_{baseline} \leq 10$)[c]		n ($SR_{baseline} < 1$)[c]	
ARE-BLA	7,214	1,077	700	(65.0%)	341	(31.7%)	36	(3.3%)
AhR-CALUX	7,968	837	495	(59.1%)	270	(32.3%)	72	(8.6%)
PPARγ-BLA	6,968	145	86	(59.3%)	45	(31.0%)	14	(9.7%)
AR-BLA	7,959	318	186	(58.5%)	94	(29.6%)	38	(11.9%)
ERα-BLA	7,947	307	114	(37.1%)	117	(38.1%)	76	(24.8%)
GR-BLA	7,963	187	106	(56.7%)	43	(23.0%)	38	(20.3%)

Note: The underlying data is given in Tables S1–S6.
[a]Number (n) of the evaluated chemicals [n (chemicals)].
[b]Number of chemicals (n) for which a $SR_{baseline}$ could be derived n ($SR_{baseline}$).
[c]Binning into the $SR_{baseline}$ categories ($SR_{baseline} > 10$, $1 \leq SR_{baseline} \leq 10$; $SR_{baseline} < 1$). In parentheses are the percentages of categories in each bin.

classification of baseline toxicants according to Verhaar et al. (Verhaar et al. 1992, 2000) was faulty, but with respect to cytotoxicity they were all at TR < 10. Furthermore, these chemicals are not reactive and too small to bind to hormone receptors.

The observed reporter gene activation by baseline toxicants supports our earlier proposed data evaluation strategy (Escher et al. 2018): All chemical concentrations $>IC_{10}$ should be excluded from analysis of reporter gene activation to ensure that the derivation of effect concentrations is not influenced by the cytotoxicity burst (SR < 1). Considering that the present analysis demonstrated that the cytotoxicity burst is a critical confounding factor when using *in vitro* reporter gene assays, we suggest an even more cautious approach when using reporter gene assay data for risk assessment: Any chemical with $1 < SR_{cytotoxicity} < 10$ should be further scrutinized. A particularly striking example is butoxyethanol, with a $SR_{cytotoxicity}$ of 0.31 in AREc32, which is even below SR 1 and would not have been identified as an active chemical if the data were analyzed only up to IC_{10} however, if all activity data were considered, it would have been mistaken as specifically acting (false positive).

The cytotoxicity burst was even more pronounced for the selected environmental chemicals where an even larger proportion of chemicals triggered the cytotoxicity burst ($SR_{cytotoxicity} < 1$) or were moderately specific ($1 < SR_{cytotoxicity} < 10$) (Figure 3B).

Not all literature data are subject to such a rigorous data treatment; hence, it is very likely that some of the reported specific activity in the literature are false positives caused by the cytotoxicity burst. Cytotoxicity does not equate to baseline toxicity, but we can use baseline toxicity as a reference and $SR_{baseline}$ (Equation 7) can be a proxy for identification of a potential cytotoxicity burst in absence of $SR_{cytotoxicity}$. $SR_{baseline}$ may even serve better to identify specific effects in cases when a different specific mechanism than the receptor/pathway targeted by the reporter gene assay had led to cell death, i.e., for chemicals with TR ≥ 10.

We have not included p53-BLA in our analysis because it is an example of a bioassay where it is hard to derive an EC_{10} because, in almost all cases, the activation of the p53 pathway is close to cell death, and therefore cytotoxicity masks activation, which is hard to differentiate from the cytotoxicity burst.

Comparison of Experimental EC and Tox21 Database

To make our own data comparable with the Tox21 database, we reevaluated the concentration–response curves and derived EC_{10} and IC_{10} values for Tox21 assays, which are typically reported as ACC (Filer et al. 2016; Judson et al. 2016). Neither ACC nor the previously used AC10 was suitable for the TR and SR analyses because the baseline toxicity QSARs are available only for IC_{10}, and the ACC is not associated to a fixed effect level but rather a measure of the lowest concentration that is statistically robust to show an effect. In this respect, ACC is rather similar to the lowest observed effect concentration LOEC. The EC_{10} was derived from the absolute 10% effect in relation to the maximum effect triggered by a potent reference compound, and these types of raw data could be extracted from the Tox21 Concentration–Response Browser as outlined in "Materials and Methods."

Our goal was not to evaluate the uncertainty of Tox21 data or to propose an alternative effect measure but to extract robust data for the TR and SR analyses. Therefore, we evaluated all data sets for one compound together, even if they stemmed from different laboratories. If they differed too much (relative standard errors >50%), the entire data set was excluded and classified as "inconclusive," but for most chemicals, there was high consistency among labs, and the joint evaluation of all data sets yielded robust and representative EC_{10} values, as is demonstrated for the example of ERα-BLA in Figure S20.

The EC_{10} and IC_{10} values agreed well between our measurements and the Tox21 database for ARE-BLA (Figure S22A),

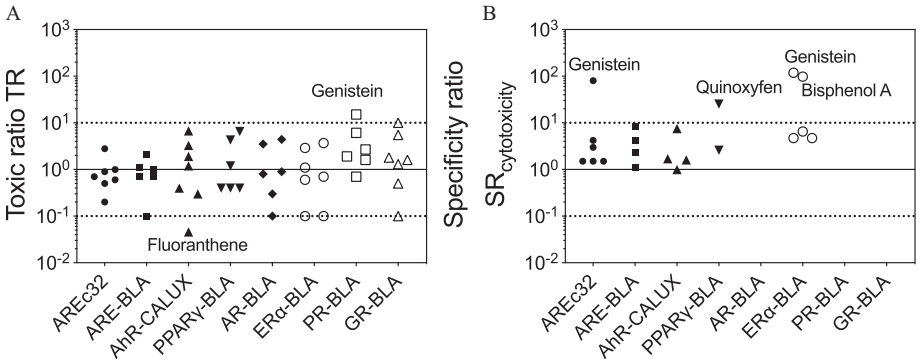

Figure 3. (A) Toxic ratios and (B) specificity ratios of the environmental chemicals (Bisphenol A, Quinoxyfen, Fluoranthene, Genistein, Coumarin, 8-Gingerol, Zingerone) in all reporter gene assays. Underlying data are in Tables S10–S17. The solid line is a $SR_{cytotoxicity}$ of 1, and the dashed lines are the thresholds of 10 and 0.1. If the effects did not exceed 10%, then no IC_{10} and/or EC_{10} could be derived, and there is no symbol in the figure. Underlying data are in Tables S2–S9.

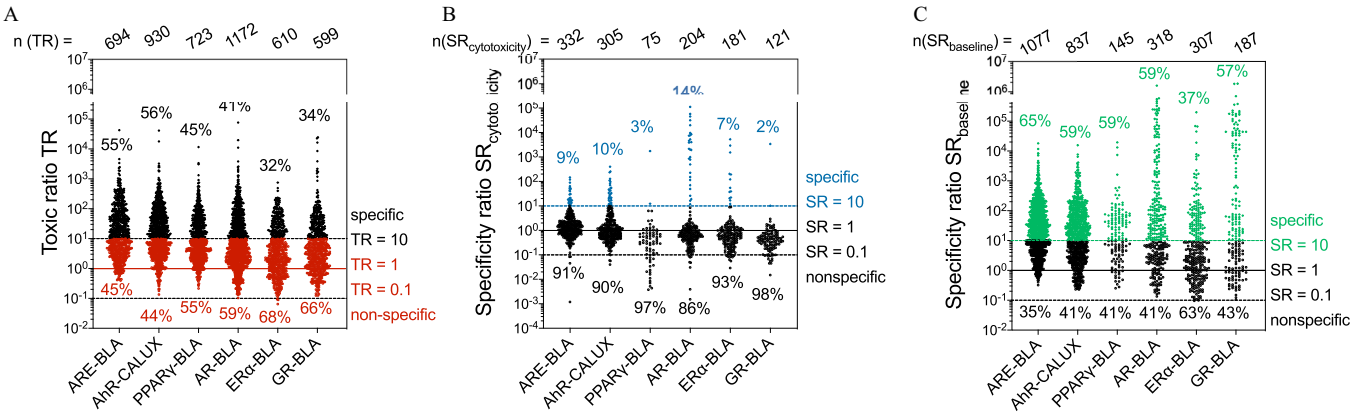

Figure 4. (A) Toxic ratio TR, (B) specificity ratio $SR_{cytotoxicity}$, and (C) specificity ratio $SR_{baseline}$ of chemicals that triggered the specific effect (EC_{10}) and/or were cytotoxic (IC_{10}) within the measured concentration range in the six Tox21 reporter gene assays. The underlying data are in the Tables S1–S6. The total number of chemicals included in the analysis is given in the top row [number of chemicals *n* for which a TR could be derived *n* (TR); number of chemicals *n* for which an $SR_{cytotoxicity}$ could be derived *n* ($SR_{cytotoxicity}$); number of chemicals *n* for which an $SR_{baseline}$ could be derived *n* ($SR_{baseline}$)]. The percentages in the top refer to all data with $TR/SR > 10$ (diamond symbols); the percentages in the bottom refer to all data with $TR/SR \leq 10$ (circle symbols).

AhR-CALUX (Figure S22B), and PPARγ-BLA (Figure S22C). The only exception was ERα-BLA, but here the SR were so low that the evaluated ERα-receptor is clearly not of relevance for those chemicals that showed no agreement, such as Quinoxyfen with a $SR_{cytotoxicity}$ of 5. The good agreement of most data (Figure S23) indicated that, despite different assay protocols and different plate formats, the responses are fairly robust. In the present study we worked in 384-well plates with larger medium volumes (40 μL) in comparison with the Tox21 reporter gene assays that were performed in 1,536-well plates with 4–6 μL of medium. The application of 384-well plates and larger medium volumes typically reduced the chemical losses from the system, in comparison with 1,536-well plates, due to higher storage capacity of the medium proteins and slower uptake in well-plate plastic (Fischer et al. 2018), but the difference appeared to be rather small.

Analysis of Tox21 Database

Approximately 50% of the chemicals triggered cytotoxicity at 10 times lower concentrations than their predicted $IC_{10,baseline}$. This finding means that a specific effect or reactive toxicity led to premature cell death. The mechanisms leading to cytotoxicity must not necessarily be the same as the receptor/pathway associated with the reporter gene. A larger proportion of chemicals were specifically toxic ($TR \geq 10$) for ARE-BLA (55%) and AhR-CALUX (56%) than for PPARγ-BLA (45%) and the hormone receptors (32%–41%). Well-known toxicants stood out, such as digitoxin ($TR = 43{,}000$ in ARE-BLA, no data in AhR-CALUX, 3,400 in PPARγ-BLA, 115 in AR-BLA, 99 in ERα-BLA, and 2,200 in GR-BLA), which is cardiotoxic and highly cytotoxic with the proposed mechanisms related to oxidative stress response and interferon-related pathways, which explains the much higher TR in AREc32 (Prassas et al. 2011). The next highest TR chemical was the reactive dye 1,8-Dihydroxy-4,5-dinitroanthraquinone with a TR of 4,600 in ARE-BLA and again much more moderate TRs in the other assays.

The cytotoxicity burst phenomenon may have led to a significant number of false positives in all evaluated Tox21 reporter gene assays. The $SR_{cytotoxicity}$ analysis clearly showed that a large proportion (86% to 97%) of the measured EC_{10} were above cytotoxic concentrations (IC_{10}) and thus impacted by the cytotoxicity burst. Only 1.7%–14.2% of the chemicals classified as specifically active with another large fraction (12%–54%) in the range of $1 \leq SR < 10$. Given the variability of $SR_{cytotoxicity}$ in our own experimental data of confirmed baseline toxicants (Figure 2B), it is likely that not only chemicals of low specificity were included in the range of $1 \leq SR < 10$ but also some with effects triggered by the cytotoxicity burst.

Due to practical reasons, all Tox21 chemicals were tested in the same concentration range (typically between ~0.001 and ~100 μM), which in turn resulted in only a small proportion of the chemicals reaching IC_{10} concentrations. Thus, the total number of chemicals with both experimental IC_{10} and EC_{10} was reduced. A larger proportion of chemicals were classified as specific in the $SR_{baseline}$ analysis than $SR_{cytotoxicity}$, because this analysis could be performed for more chemicals in the absence of experimental cytotoxicity data (up to three times more data points included) but also because chemicals with a cytotoxicity triggered by a different pathway had smaller $SR_{cytotoxicity}$. A large proportion of the active chemicals (~60%, exception ERα-BLA) were likely truly specific (Fig. 4C). Nevertheless, for 3% to 25% of the active chemicals with an EC_{10}, the cytotoxicity burst presumably led to a false positive EC_{10} and a further 23%–30% was only moderately specific, some of which, given the experimental variability could also have been impacted by the cytotoxicity burst. For ERα-BLA, 25% of the EC_{10} fell below $IC_{10,baseline}$ (Fig. 4C); thus, the cytotoxicity burst was particularly impeding the outcome of this assay.

Conclusion

Our findings clearly confirm previous reports of the cytotoxicity burst phenomenon and show a systematic way to identify and interpret the cytotoxicity burst. For chemicals with a high TR (i.e., those that are more toxic than baseline toxicity), one needs to be especially cautious because the specific effect may in some cases also cause enhanced cytotoxicity, e.g., in the two assays of oxidative stress response, AREc32 and ARE-BLA. For these assays, the activation of defense mechanisms might go hand-in-hand with the higher cytotoxicity caused by specific effects. In contrast, the activation of the hormone receptors ERα-BLA, AR-BLA, PR-BLA, and GR-BLA is not likely to be associated with increased cytotoxicity, so any activation of hormone receptors at cytotoxic concentration is likely to be a result of the cytotoxicity burst phenomenon.

For routine applications of cell-based bioassays, we recommend performing the evaluation of activity data only at

concentrations smaller than IC_{10} to avoid false-positive responses ($SR < 1$). Likewise, measures can be taken to avoid false negatives with respect to highest tested concentration. The baseline toxicity QSARs can help to predict, at which concentrations we expect the minimal toxicity (baseline toxicity). We previously suggested to dose chemicals up to their solubility limit in medium or to their predicted $IC_{10,baseline}$, whichever is lower (Fischer et al. 2019). Although it is challenging for HTS experiments to adjust each concentration, it might be possible to group chemicals according to their physicochemical properties (hydrophobicity, speciation, etc.) and test them groupwise at fixed concentrations depending on their expected IC_{10}-range. Both medium solubility limit and $IC_{10,baseline}$ can be predicted for neutral chemicals by the log K_{ow} of the chemicals as sole descriptor (Escher et al. 2019; Fischer et al. 2019). For ionizable chemicals, the speciation additionally needs to be included. By ensuring that the tested concentration range reached cytotoxicity, one can calculate both the $SR_{cytotoxicity}$ and the $SR_{baseline}$ as a diagnostic tool for evaluation of the specificity of a chemical. This approach will also help to reduce the uncertainty of HTS data. Watt and Judson (2018) identified data with high variability by an uncertainty analysis of the Tox21 data. It is conceivable that the elimination of data triggered by nonspecific effects will decrease uncertainty.

Although we have validated the classification method with a small selection of reporter gene assays only, the approach is easily transferable to other reporter gene assays because the baseline toxicity QSARs are based on constant critical membrane concentrations and can be predicted for any cell-based bioassay, provided that lipid and protein composition of cells and medium are known (Escher et al. 2019).

For application of *in vitro* test methods in risk assessment, it is important that mode-of-action specific QIVIVE are performed only with data that are not compromised by the cytotoxicity burst.

Acknowledgments

The authors gratefully acknowledge the financial support by the CEFIC Long-Range Research Initiative (LRI), project ECO36. L.H. was funded by Unilever. The robotic HTS systems are a part of the major infrastructure initiative Chemicals in the Environment Profiler (CITEPro), funded by the Helmholtz Association with cofunding by the States of Saxony and Saxony-Anhalt.

The authors thank L. Glauch and M. Mühlenbrink for experimental support. The authors also thank W. Oesterheld for support in the MATLAB programming, P. Mayer (Technical University of Denmark), T. Gouin (TG Environmental Research), and B. Nichols (Unilever) for helpful discussions.

References

Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, et al. 2013. The Tox21 robotic platform for the assessment of environmental chemicals – from vision to reality. Drug Discov Today 18(15–16):716–723, PMID: 23732176, https://doi.org/10.1016/j.drudis.2013.05.015.

Bittermann K, Spycher S, Goss KU. 2016. Comparison of different models predicting the phospholipid-membrane water partition coefficients of charged compounds. Chemosphere 144:382–391, PMID: 26383265, https://doi.org/10.1016/j.chemosphere.2015.08.065.

Brennan JC, He G, Tsutsumi T, Zhao J, Wirth E, Fulton MH, et al. 2015. Development of species-specific ah receptor-responsive third generation CALUX cell lines with enhanced responsiveness and improved detection limits. Environ Sci Technol 49(19):11903–11912, PMID: 26366531, https://doi.org/10.1021/acs.est.5b02906.

Burden N, Mahony C, Müller BP, Terry C, Westmoreland C, Kimber I. 2015. Aligning the 3Rs with new paradigms in the safety assessment of chemicals. Toxicology 330:62–66, PMID: 25932488, https://doi.org/10.1016/j.tox.2015.01.014.

Endo S, Escher BI, Goss K-U. 2011. Capacities of membrane lipids to accumulate neutral organic chemicals. Environ Sci Technol 45(14):5912–5921, PMID: 21671592, https://doi.org/10.1021/es200855w.

Escher BI, Abagyan R, Embry M, Klüver N, Redman AD, Zarfl C, et al. 2020. Recommendations for improving methods and models for aquatic hazard assessment of ionizable organic chemicals. Environ Toxicol Chem 39(2):269–286, PMID: 31569266, https://doi.org/10.1002/etc.4602.

Escher BI, Baumer A, Bittermann K, Henneberger L, König M, Kühnert C, et al. 2017. General baseline toxicity QSAR for non-polar, polar and ionisable chemicals and their mixtures in the bioluminescence inhibition assay with *Aliivibrio fischeri*. Environ Sci: Processes Impacts 19(3):414–428, PMID: 28197603, https://doi.org/10.1039/C6EM00692B.

Escher BI, Dutt M, Maylin E, Tang JYM, Toze S, Wolf CR, et al. 2012. Water quality assessment using the AREc32 reporter gene assay indicative of the oxidative stress response pathway. J Environ Monit 14(11):2877–2885, PMID: 23032559, https://doi.org/10.1039/c2em30506b.

Escher BI, Glauch L, König M, Mayer P, Schlichting R. 2019. Baseline toxicity and volatility cutoff in reporter gene assays used for high-throughput screening. Chem Res Toxicol 32(8):1646–1655, PMID: 31313575, https://doi.org/10.1021/acs.chemrestox.9b00182.

Escher BI, Neale PA, Villeneuve D. 2018. The advantages of linear concentration–response curves for in vitro bioassays with environmental samples. Environ Toxicol Chem 37(9):2273–2280, PMID: 29846006, https://doi.org/10.1002/etc.4178.

Fay KA, Villeneuve DL, Swintek J, Edwards SW, Nelms MD, Blackwell BR, et al. 2018. Differentiating pathway-specific from nonspecific effects in high-throughput toxicity data: a foundation for prioritizing adverse outcome pathway development. Toxicol Sci 163(2):500–515, PMID: 29529260, https://doi.org/10.1093/toxsci/kfy049.

Filer DL, Kothiya P, Setzer RW, Judson RS, Martin MT. 2016. tcpl: the ToxCast pipeline for high-throughput screening data. Bioinformatics 33(4):618–620, https://doi.org/10.1093/bioinformatics/btw1680.

Filer D, Patisaul HB, Schug T, Reif D, Thayer K. 2014. Test driving ToxCast: endocrine profiling for 1858 chemicals included in phase II. Curr Opin Pharmacol 19:145–152, PMID: 25460227, https://doi.org/10.1016/j.coph.2014.09.021.

Fischer FC, Cirpka O, Goss K-U, Henneberger L, Escher BI. 2018. Application of experimental polystyrene partition constants and diffusion coefficients to predict the sorption of organic chemicals to well plates in in vitro bioassays. Environ Sci Technol 52(22):13511–13522, PMID: 30298728, https://doi.org/10.1021/acs.est.8b04246.

Fischer FC, Henneberger L, König M, Bittermann K, Linden L, Goss K-U, et al. 2017. Modeling exposure in the Tox21 *in vitro* bioassays. Chem Res Toxicol 30(5):1197–1208, PMID: 28316234, https://doi.org/10.1021/acs.chemrestox.7b00023.

Fischer FC, Henneberger L, Schlichting R, Escher BI. 2019. How to improve the dosing of chemicals in high-throughput in vitro mammalian cell assays. Chem Res Toxicol 32(8):1462–1468, PMID: 31328914, https://doi.org/10.1021/acs.chemrestox.9b00167.

He GC, Tsutsumi T, Zhao B, Baston DS, Zhao J, Heath-Pagliuso S, et al. 2011. Third-generation Ah receptor-responsive luciferase reporter plasmids: amplification of dioxin-responsive elements dramatically increases CALUX bioassay sensitivity and responsiveness. Toxicol Sci 123(2):511–522, PMID: 21775728, https://doi.org/10.1093/toxsci/kfr189.

Henneberger L, Mühlenbrink M, Fischer FC, Escher BI. 2019. C18-coated solid-phase microextraction fibers for the quantification of partitioning of organic acids to proteins, lipids, and cells. Chem Res Toxicol 32(1):168–178, PMID: 30585484, https://doi.org/10.1021/acs.chemrestox.8b00249.

Henneberger L, Mühlenbrink M, Heinrich D, Teixeira A, Nicol B, Escher BI. 2020. Experimental validation of mass balance models for in vitro cell-based bioassays. Environ Sci Technol 54(2):1120–1127, PMID: 31852189, https://doi.org/10.1021/acs.est.9b06144.

Hsieh J-H, Huang R, Lin J-A, Sedykh A, Zhao J, Tice RR, et al. 2017. Real-time cell toxicity profiling of Tox21 10K compounds reveals cytotoxicity dependent toxicity pathway linkage. PLoS One 12(7):e0181291, https://doi.org/10.1371/journal.pone.0181291.

Hu B, Gifford E, Wang H, Bailey W, Johnson T. 2015. Analysis of the ToxCast chemical-assay space using the Comparative Toxicogenomics Database. Chem Res Toxicol 28(11):2210–2223, PMID: 26505644, https://doi.org/10.1021/acs.chemrestox.5b00369.

Huang R, Sakamuru S, Martin MT, Reif DM, Judson RS, Houck KA, et al. 2014. Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. Sci Rep 4:5664–5664, PMID: 25012808, https://doi.org/10.1038/srep05664.

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. 2010. *In vitro* screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118(4):485–492, PMID: 20368123, https://doi.org/10.1289/ehp.0901392.

Judson RS, Houck KA, Martin MT, Richard AM, Knudsen TB, Shah I, et al. 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(2):323–339, PMID: 27208079, https://doi.org/10.1093/toxsci/kfw092.

Kienzler A, Barron MG, Belanger SE, Beasley A, Embry MR. 2017. Mode of action (MOA) assignment classifications for ecotoxicology: an evaluation of approaches. Environ Sci Technol 51(17):10203–10211, PMID: 28759717, https://doi.org/10.1021/acs.est.7b02337.

Kleinstreuer NC, Allen D, Casey WM. 2015. Using ToxCast/Tox21 HTS assays to assess endocrine disruption. Birth Defects Research Part A-Clinical and Molecular Teratology. 103:395, https://doi.org/10.1002/bdra.23387.

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, https://doi.org/10.1016/j.aquatox.2018.12.003.

König M, Escher BI, Neale PA, Krauss M, Hilscherová K, Novák J, et al. 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, PMID: 27884472, https://doi.org/10.1016/j.envpol.2016.11.011.

Krewski D, Andersen ME, Tyshenko MG, Krishnan K, Hartung T, Boekelheide K, et al. 2020. Toxicity testing in the 21st century: progress in the past decade and future perspectives. Arch Toxicol 94(1):1–58, PMID: 31848664, https://doi.org/10.1007/s00204-019-02613-4.

Kwon JH, Liljestrand HM, Katz LE. 2006. Partitioning of moderately hydrophobic endocrine disruptors between water and synthetic membrane vesicles. Environ Toxicol Chem 25(8):1984–1992, PMID: 16916015, https://doi.org/10.1897/05-550r.1.

Maeder V, Escher BI, Scheringer M, Hungerbuhler K. 2004. Toxic ratio as an indicator of the intrinsic toxicity in the assessment of persistent, bioaccumulative, and toxic chemicals. Environ Sci Technol 38(13):3659–3666, PMID: 15296318, https://doi.org/10.1021/es0351591.

Martin MT, Dix DJ, Judson RS, Kavlock RJ, Reif DM, Richard AM, et al. 2010. Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. Chem Res Toxicol 23(3):578–590, PMID: 20143881, https://doi.org/10.1021/tx900325g.

Murk AJ, Legler J, Denison MS, Giesy JP, van de Guchte C, Brouwer A. 1996. Chemical-activated luciferase gene expression (CALUX): a novel in vitro bioassay for Ah receptor active compounds in sediments and pore water. Fundam Appl Toxicol 33(1):149–160, PMID: 8812260, https://doi.org/10.1093/toxsci/33.1.149.

National Research Council. 2007. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC: National Academies Press.

Neale PA, Altenburger R, Aït-Aïssa S, Brion F, Busch W, de Aragão Umbuzeiro G, et al. 2017. 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, PMID: 28728110, https://doi.org/10.1016/j.watres.2017.07.016.

Pham N, Iyer S, Hackett E, Lock BH, Sandy M, Zeise L, et al. 2016. Using ToxCast to explore chemical activities and hazard traits: a case study with ortho-phthalates. Toxicol Sci 151(2):286–301, PMID: 26969370, https://doi.org/10.1093/toxsci/kfw049.

Prassas I, Karagiannis GS, Batruch I, Dimitromanolakis A, Datti A, Diamandis EP. 2011. Digitoxin-induced cytotoxicity in cancer cells is mediated through distinct kinase and interferon signaling networks. Mol Cancer Ther 10(11):2083–2093, PMID: 21859838, https://doi.org/10.1158/1535-7163.MCT-11-0421.

Reif DM, Martin MT, Tan SW, Houck KA, Judson RS, Richard AM, et al. 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. Environ Health Perspect 118(12):1714–1720, PMID: 20826373, https://doi.org/10.1289/ehp.1002180.

Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, et al. 2016. ToxCast chemical landscape: paving the road to 21st century toxicology. Chem Res Toxicol 29(8):1225–1251, PMID: 27367298, https://doi.org/10.1021/acs.chemrestox.6b00135.

Sipes NS, Martin MT, Kothiya P, Reif DM, Judson RS, Richard AM, et al. 2013. Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. Chem Res Toxicol 26(6):878–895, PMID: 23611293, https://doi.org/10.1021/tx400021f.

Sipes NS, Wambaugh JF, Pearce R, Auerbach SS, Wetmore BA, Hsieh JH, et al. 2017. An intuitive approach for predicting potential human health risk with the Tox21 10k library. Environ Sci Technol 51(18):10786–10796, PMID: 28809115, https://doi.org/10.1021/acs.est.7b00650.

Thomas RS, Bahadori T, Buckley TJ, Cowden J, Deisenroth C, Dionisio KL, et al. 2019. The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. Toxicol Sci 169(2):317–332, PMID: 30835285, https://doi.org/10.1093/toxsci/kfz058.

Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, et al. 2013. Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. Toxicol Sci 136(1):4–18, PMID: 23958734, https://doi.org/10.1093/toxsci/kft178.

Tice RR, Austin CP, Kavlock RJ, Bucher JR. 2013. Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 121(7):756–765, PMID: 23603828, https://doi.org/10.1289/ehp.1205784.

U.S. EPA (U.S. Environmental Protection Agency). 2019. Chemistry Dashboard. https://comptox.epa.gov/dashboard [accessed 22 March 2019].

Vaes WH, Ramos EU, Hamwijk C, van Holsteijn I, Blaauboer BJ, Seinen W, et al. 1997. Solid phase microextraction as a tool to determine membrane/water partition coefficients and bioavailable concentrations in in vitro systems. Chem Res Toxicol 10(10):1067–1072, PMID: 9348427, https://doi.org/10.1021/tx970109t.

Vaes WHJ, Ramos EU, Verhaar HJM, Hermens J. 1998. Acute toxicity of nonpolar versus polar narcosis: is there a difference? Environ Toxicol Chem 17(7):1380–1384, https://doi.org/10.1002/etc.5620170723.

van der Heijden SA, Jonker M. 2009. Evaluation of liposome-water partitioning for predicting bioaccumulation potential of hydrophobic organic chemicals. Environ Sci Technol 43(23):8854–8859, PMID: 19943657, https://doi.org/10.1021/es902278x.

van Wezel AP, Opperhuizen A. 1995. Narcosis due to environmental pollutants in aquatic organisms: residue-based toxicity, mechanisms, and membrane burdens. Crit Rev Toxicol 25(3):255–279, PMID: 7576154, https://doi.org/10.3109/10408449509089890.

Verhaar HJM, Solbé J, Speksnijder J, van Leeuwen CJ, Hermens J. 2000. Classifying environmental pollutants: part 3. External validation of the classification system. Chemosphere 40(8):875–883, PMID: 10718581, https://doi.org/10.1016/S0045-6535(99)00317-3.

Verhaar HJM, Van Leeuwen CJ, Hermens J. 1992. Classifying environmental pollutants. Chemosphere 25(4):471–491, https://doi.org/10.1016/0045-6535(92)90280-5.

Vinken M, Blaauboer BJ. 2017. *In vitro* testing of basal cytotoxicity: establishment of an adverse outcome pathway from chemical insult to cell death. Toxicol In Vitro 39:104–110, PMID: 27939612, https://doi.org/10.1016/j.tiv.2016.12.004.

Watt ED, Judson RS. 2018. Uncertainty quantification in ToxCast high throughput screening. PLoS One 13(7):e0196963, PMID: 30044784, https://doi.org/10.1371/journal.pone.0196963.

Wetmore BA. 2015. Quantitative *in vitro*-to-*in vivo* extrapolation in a high-throughput environment. Toxicology 332:94–101, PMID: 24907440, https://doi.org/10.1016/j.tox.2014.05.012.