Joint Research Programme
KWRW 2025.004| February 2025

Health-based trigger values for data-poor bioassays

Joint Research Programme

KWR

Bridging Science to Practice

Colophon

Health-based trigger values for data-poor bioassays

KWRW 2025.004| February 2025

This research is part of the Joint Research Programme of KWR, the water utilities and Vewin.

Project number
404300/026

Project manager
Dr. Renske P.J. Hoondert, Patrick S. Bäuerlein

Client
BTO - Thematic research - Chemical safety

Author(s)
Dr. ir. Tessa E. Pronk, Dr. Renske P.J. Hoondert

Quality Assurance
Dr. Milou Dingemans.

Sent to
This report is distributed to BTO-participants.

After establishing the definitive EBT with additional tested compounds in the investigated bioassays, the publication can be made public.

Keywords

in vitro assays, trigger values, toxicity, chemical water quality

Year of publishing
2025

More information
Dr. ir. T.E Pronk
T 06 29337806
E Tessa.Pronk@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

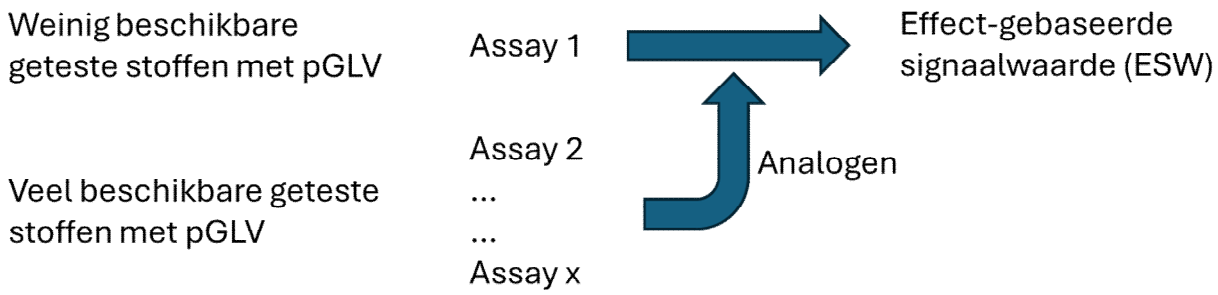
T +31 (0)30 60 69 511
E info@kwrwater.nl
I www.kwrwater.nl

Management samenvatting

Gezondheidskundige bioassay signaalwaarden afleiden met weinig data

Auteurs Dr. ir. Tessa E. Pronk, Dr. Renske P.J. Hoondert

Bioassays kunnen worden gebruikt om de chemische waterkwaliteit te testen. Een effect-gebaseerde signaalwaarde (ESW) kan worden gebruikt om onderscheid te maken tussen een waterkwaliteit waarbij mogelijk gezondheidsrisico niet kan worden uitgesloten en een veilig geachte waterkwaliteit. In eerder onderzoek is een methode ontwikkeld om per bioassay een ESW te berekenen op basis van bioactiviteit van stoffen waarvan de humane gezondheidskundige richtwaarde bekend is. Voor sommige bioassays zijn echter maar weinig gegevens van geteste stoffen beschikbaar. In deze gevallen kunnen ESW's worden berekend uit gegevens van stoffen die zijn getest in vergelijkbare bioassays ('analogen') die voldoende gelijk reageren op relevante stoffen. In dit onderzoek is deze methodiek toegepast om ESW's af te leiden voor relatief nieuwe bioassays waarin nog weinig individuele stoffen getest zijn (Nrf2 CALUX en PXR CALUX). Deze bioassays kunnen na validatie worden ingezet in de drinkwatersector om waterkwaliteit te interpreteren.



Overzicht van de methodologie om ESW's af te leiden voor data-arme bioassays met behulp van informatie van vergelijkbare (analoge) bioassays. pGLV staat voor (provisional) guideline value (voorlopige gezondheidskundige richtwaarde).

Belang: Effect-signaalwaarden zijn belangrijk voor het interpreteren van waterkwaliteit

Een signaalwaarde maakt het mogelijk met een bioassay onderscheid te maken tussen water met of zonder risico op gezondheidseffecten. Hoe meer data van op bioactiviteit geteste stoffen worden betrokken bij de afleiding van de signaalwaarde, hoe betrouwbaarder. Het is dan waarschijnlijker dat de signaalwaarde ook beschermend is voor andere in water aangetroffen stoffen. Voor sommige bioassays zijn (nog) weinig data van geteste stoffen beschikbaar. Zulke bioassays kunnen toch potentieel hebben voor waterkwaliteitsonderzoek, omdat ze bijvoorbeeld een relevant eindpunt onderzoeken, of makkelijk zijn in het gebruik, gevoelig of goedkoop. Als ze worden ingezet voor waterkwaliteitsbewaking zijn er echter signaalwaarden nodig.

Aanpak: Informatie lenen uit vergelijkbare bioassays

Veelal bestaan er meerdere bioassays waarmee vergelijkbare effecten kunnen worden onderzocht. De respons van een bioassay wordt uitgedrukt in equivalente concentraties van een vast mee geteste referentiestof. In sommige gevallen lijken de responsen van geteste stoffen uit zulke bioassays op elkaar. In dit onderzoek is uitgezocht welke bioassays dat zijn voor de PXR CALUX en de Nrf2 CALUX door in

de ToxCast database te zoeken naar bioassays die hetzelfde eindpunt testen en eenzelfde respons vertonen bij blootstelling aan chemicaliën. Dit is bepaald met een correlatieanalyse. Vervolgens is gekeken naar stoffen die in deze analoge bioassays de ESW bepalen. Deze zullen naar verwachting dan dezelfde respons in de PXR of Nrf2 CALUX veroorzaken. Op deze manier kan voor deze bioassays, waarvoor weinig informatie beschikbaar is, gegevens 'geleend' worden bij zulke vergelijkbare 'analoge' bioassays om een ESW te berekenen.

Resultaat: Lagere signaalwaarden voor data-arme bioassays

Voor beide voorbeeld bioassays werden een of meerdere voldoende gelijkende bioassays met veel geteste stoffen gevonden in de ToxCast database. Stoffen die in de analoge bioassays tot een lage signaalwaarde leiden, bleken veelal niet in de CALUX-bioassays getest te zijn. Naar verwachting zal de signaalwaarde in de CALUX-bioassays dus lager liggen dan alleen op basis de binnen de CALUX-bioassays geteste stoffen. De verwachting is daarom dat na het testen van relevante stoffen uit de analoge bioassays de ESW naar beneden zal worden bijgesteld, zodat de ESW ook voor deze stoffen beschermend is. Na het testen van deze stoffen kan de ESW definitief worden gemaakt.

Naast beperkte overlap in geteste stoffen tussen CALUX en analoge bioassays waren ook de bekende referentiestoffen voor de twee CALUX bioassays niet in alle (analoge) ToxCast-bioassays getest. Dit beperkte de identificatie van analoge bioassays. Het gebruik van referentiestof(fen) kan beter worden afgestemd door de partijen binnen het domein van het effect gebaseerd meten.

Een andere beperking was de beschikbaarheid van (provisionele) gezondheidskundige richtwaarden (pGLV). Binnen dit project zijn daarom relevante data bij elkaar gebracht, wat heeft geresulteerd in een database met pGLVs voor 920 unieke stoffen.

Toepassing: Beter risico inschatten voor meer bioassays

Met de ontwikkelde methodiek is het mogelijk om voorlopige ESW af te leiden voor bioassays met (nog) weinig beschikbare data rond individueel geteste stoffen met een pGLV. Dit biedt kansen voor snellere implementatie van nieuwe bioassays.

Rapport

Dit onderzoek is beschreven in het rapport *Health-based trigger values for data-poor bioassays* (KWRW 2025.004).

Gerelateerde publicaties zijn:

- BTO 2020.053 Development of a framework to derive effect-based trigger values to interpret CALUX data for drinking water quality
- BTO 2023.052 Effectsignaalwaarden voor eenduidige interpretatie van genotoxiciteitstesten

.

Contents

Meer informatie
dr. ir. Tessa Pronk, Tessa
T 030 606 9681
E Tessa.Pronk@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

1 Introduction

In vitro bioassays are used in water quality assessment to measure combined effects of chemicals on certain toxicity indicating biological processes. An effect-based 'trigger value' (EBT) for human health can be applied to assess whether an observed effect exceeds a threshold above which risks are not negligible (Been et al., 2021; Brand et al., 2013; Escher et al., 2015). When exceeding this value (mixtures of) chemicals may be present at concentrations that are harmful to human health. As such, the EBT can distinguish between poor and acceptable water quality in terms of risks for human health. Hence, ideally, if the *in vitro* bioassay response is lower than the EBT, there is no risk of adverse effects on consumer health for the endpoint that the *in vitro* bioassay addresses, while an exceedance of the EBT signifies a health risk, although this is not easily quantifiable (Pronk et al., 2024).

In this report we derive EBTs for two *in vitro* bioassays (in short 'assays' in the remainder of the report) that are used in practice in the Netherlands, associated respectively with xenobiotic metabolism and oxidative stress: the PXR CALUX® and Nrf2 CALUX®. The Pregnane X receptor (PXR) CALUX measures the activation of the Pregnane X receptor, which activates the cyp3A4 enzyme triggering xenobiotic metabolism (Houtman et al., 2017). Xenobiotic metabolism is the chemical transformation by a biological system (metabolism) of xenobiotic compounds (e.g. drugs). Here, lipophilic chemicals are converted to hydrophilic metabolites, which are more easily excreted (Oladimeji & Chen, 2018). The nuclear factor erythroid 2-related (Nrf2) CALUX measures oxidative stress through the Nrf2 pathway (Van Der Linden et al., 2014). This pathway is the principal cellular defence mechanism against oxidative stress by activating phase II detoxification enzymes (Battino et al., 2018). It is known that innate immune responses are also induced by Nrf2 activation, and lead to the suppression or elimination of several bacterial and viral pathogens (Battino et al., 2018).

Both these assays are currently used in drinking water testing (Bertelkamp et al., 2019) and surface water testing (Pronk et al., 2021). An EBT is essential for a better interpretation of the responses of these assays. The assays however have limited data availability on the response that is induced in the assays by individually tested chemicals ('bioactivity data'), needed for the derivation of an EBT.

In general, for assays, more data means a more solid derived EBT. There is often less data available than required to derive such EBTs. Data are not (publicly) available or difficult to find. There is as of yet still no central place or fixed protocol to share such data. The NORMAN network is in the process of assembling such a database. For now this database contains data for several *in vitro* assays, but still has to acquire more data. The ToxCast database (US EPA, 2023) does contain sufficiently tested assays, often with hundreds of chemicals tested and their activity concentrations to induce 50% of the maximal effect of the assay (AC_{50}) registered ('bioactivity data'). Responses of assays are often recalculated to the effect relative to a single reference compound, the relative effect potency. If the assays show similar relative effect potencies for chemicals compared to assays used for water quality monitoring, these should, in theory, have a similar EBT. Therefore, in this report, the information from other, similar (or in other words 'analogous') assays from the ToxCast database is used to derive an EBT for the Nrf2 and PXR CALUX assays. For this purpose the similarity between the assays needs to be established and quantified as well, and a method to do this is also described in this report.

2 Deriving EBTs for data-poor assays

2.1 Methodology for deriving EBT

To interpret assay responses that are indicative of human health risks a link must be made between a health-wise safe concentration in (drinking)water and an available assay responses. For the assay response, the active concentration at which 50% of the total activity of the *in vitro* assay is observed (AC_{50}) is chosen, or if this is not available, the lowest observed effect concentration (LOEC). In general we refer to such data as 'bioactivity data'. For determining health risks, provisional guideline values (pGLV) are used. A pGLV for an individual chemical is the provisional, health-based and non-binding maximum advised daily dose in drinking water (Baken et al., 2018; Brand et al., 2013). The pGLV for a chemical is established based on measures as acceptable daily intake (ADI), Reference dose (RfD) or tolerable daily intake (TDI) under the assumption of a lifelong consumption of 2 litres of water a day by a person of average body weight (70 kg) (WHO, 2022). This guideline value is based on the most sensitive ('critical') reported endpoint for the chemical in literature. This is however not necessarily associated to the target endpoint of the assay for which a EBT is to be derived based on the pGLV. In future research a consideration is to revise the pGLV for the endpoint associated to the assay endpoint. An indication that a pGLV is not relevant for a assay is if the AC_{50} taken for the assay endpoint of interest deviates strongly (e.g. more than 10 times) from the pGLV derived in literature. Future research with help of physiologically based pharmacokinetic (PBPK) modelling and quantitative in vitro to in vivo extrapolation (QIVIVE) may be used to better estimate if these values significantly differ from a toxicological perspective. If the pGLV is low compared to the AC_{50} of the *in vitro* assay, the pGLV is likely not associated to the specific activity/response that the *in vitro* assay represents. Therefore, a condition for the inclusion of the pGLV was applied as proposed by Béen et al. (2021), in which pGLVs were only included if they were at least 10 times the AC_{50} included in the study:

$$\text{Select if: } \frac{pGLV_i}{AC50_i} > 0.1 \qquad \text{Equation 1}$$

where the $pGLV_i$ is the provisional guideline value of chemical i and $AC_{50}i$ is the concentration at which half of the maximum activity is reached in the assay for chemical i. Health-based EBTs for the selected *in vitro* assays were derived using the method outlined in Béen et al. (2021) which is briefly explained below. Firstly, the effect potencies of chemicals are calculated relative to a (potent) reference chemical:

$$REP_i = \frac{AC50_y}{AC50_i} \qquad \text{Equation 2}$$

where REP_i represents the relative effect potency of chemical of interest *i*, AC_{50-i} is the concentration at which half of the maximum activity is reached for chemical *i*, and AC_{50-y} represents this value for the reference chemical (see Table 1 for an overview of reference chemicals per assay). A low REP indicates a non-potent substance. Any response of the assay in reference substance equivalents will mean a higher concentration of the non-potent substance was present. In the second step, the health-related bioanalytical equivalent pGLV of each chemical is calculated. This involves determining the expected pGLV related assay activity expressed as reference chemical equivalents (based on the *REP*):

$$BpGLV_i = pGLV_i \cdot REP_i \qquad \text{Equation 3}$$

where $BpGLV_i$ is the bioanalytical pGLV equivalent of chemical *i*, and REP_i is the relative effect potency of chemical *i*. In Béen et al., *BpGLV* is referred to as pGLV BEQ. The lowest *BpGLV* is used to derive the EBT instead of the lower 5% limit of a distribution (Béen et al., 2021).

$$EBT = \text{minimum } (BpGLV_{i-n}) \qquad \text{Equation 4}$$

For data-poor assays, it is possible that the tested compounds are not amongst the substances potentially having a low BpGLV in this assay. For this reason, the derived health-based EBTs are checked for robustness by comparing EBT and bioactivity data from similar (or, in other words: analogous) assays. Assays can be regarded as 'similar' if they have a similar REP for commonly tested compounds and a similar biological endpoint. This is further explained in paragraph 2.4.

From Equation 3 it is clear that the *REP* and *pGLV* proportionally determine the EBT. Because the *pGLV* remains the same for compounds between *in vitro* assays, solely the *REP* will determine differences in *BpGLV* for compounds between analogous assays. Derived EBTs may differ between analogous assays with the same endpoint (e.g. PXR-receptor binding) because:

1. A compound (or several compounds) with very low potency (in other words; high AC_{50} or $LOEC/AC_{10}$) and/or low pGLV was tested in one assay, that was not tested or selected (Equation 1) in the others;
2. Specifically the reference compound has a different relative potency with respect to chemicals in one assay compared to another assay. If the reference compound is relatively one of the more potent compounds, the REP of other tested compounds in the assay will be overall low. The other way around, if the reference compound is relatively not very potent, the REP of other tested compounds will be overall high (Equation 2).
3. Inherent variability in bioactivity of compounds even if they are tested in the same assay, or an analogous assay.

Based on the above points, for any data-poor assay, its EBT can be adjusted to a lower value if there is an analogous assay for which a lower EBT is calculated, caused by the availability of data for different compounds (1). Correcting for any differences in REP caused by a difference in relative potency of the reference compound (2) is tricky because the difference can also lie in the potency of other tested compounds. The inherent variability (3) cannot be remedied, however it is good to know that this variability exists.

2.2 Data compilation

2.2.1 Health values, (p)GLV data

To facilitate the derivation of EBTs, available provisional guideline values (pGLVs) for active chemicals in the assays were collected. Provisional guideline values for drinking water were collated from several sources. Some sources provide a direct pGLV, for others this was calculated.

- Pesticide database from the European Commission (calculated)
- IRIS database (calculated)
- Guidelines for Canadian Drinking Water Quality (2023) (pGLV)
- Australian drinking water guidelines (2023) (pGLV)
- WHO (2022) (pGLV)
- ITER database (Calculated)
- EFSA (Calculated)

The pGLV was calculated by allocating a 20% proportion of the reference dose (RfD), acceptable daily intake (ADI) or tolerable daily intake (TDI) to drinking water to make allowance for exposure from other sources, e.g. food and subsequently by multiplying this proportion by the average weight of an adult (70 kg) and dividing by the average drinking water intake (adults: 2 L/day) (WHO, 2022). No check was done if the 20% allocation was justified for all chemicals, this is a potential under- or overestimation of what the pGLV should be. Some of the pGLVs used in the present study were assembled from Béen et al. (2021), Baken et al. (2018) and an unpublished project report that used EFSA and WHO-advised daily intake values, amongst others (van den Berg, 2021). This resulted in an inhouse dataset of 920 chemicals with a pGLV.

For some chemicals, the used sources provided different pGLVs. This occurs when there is a difference in the data sources used to derive the pGLV or if allocation factors have been applied differently. To derive effect-based trigger values (EBT) in this report for the assays according to the method described in Béen et al. (2021) (supplemented with the method developed in this report) only the lowest values per substance were maintained to get a conservative EBT.

2.2.2 Nrf2 CALUX

LOEC and/or EC_{50} data for individual chemicals tested in the Nrf2 CALUX were collected from literature. Nrf2 CALUX data included data from Been et al. (2021), based on a dataset by Van Der Linden et al. (2014), reporting LOECs for 14 chemicals. Additional data were added from a dataset from Escher et al. (2018). In total, bioactivity data for 38 chemicals were collected. pGLVs were available for 28 chemicals. In total nine unique chemicals could be used for the derivation of the health-based EBT, after application of the criterion in Equation 1. Potent chemicals are generally preferred as a reference compound because this yields low REP values for other compounds. The reference compound reported in the study of Van Der Linden et al. (2014) was tert-butylhydroquinone and the study from Escher (2018) used dichlorvos as a reference compound. It is known that the chemical curcumin is currently used as the reference compound for Nrf2 CALUX by the vendor, Biodetection systems (BDS). Carbendazim is mentioned in van der Oost et al. (2017).

Note: The Nrf2 CALUX data were an assembly of three different sources and hence, the experimental set-up could differ between results. It would be best to normalize the individual assay responses in the different sources based on common (reference) chemical(s) prior to the calculation of REP. However, there are no chemicals that were tested in all three sources. Instead, we normalized the results of each source to a common bioactivity median (divide values per source by its own median, times the median over all three sources).

2.2.3 PXR CALUX

PC_{10} and PC_{20} data (concentrations inducing a 10% and 20% response relative to the positive control, respectively) for individual chemicals for the PXR CALUX were provided by BioDetection Systems, Amsterdam (BDS). Data was available for 226 individual chemicals. For 72 of these chemicals pGLVs were available. Several of these chemicals were excluded based on the criterion described in Equation 1 ($pGLV/AC_{10} > 0.1$), resulting in 32 chemicals that could be used for the derivation of a health-based EBT. Nicardipine (55985-32-5) was mentioned as the reference compound by BDS. Metolachlor (CAS: 51218-45-2) and DEHP (117-81-7) are mentioned as reference compounds in other studies (Neale et al., 2023). Azoxystrobin (CAS: 131860-33-8) was identified as a suitable reference compound in the current report, because it is a potent substance in PXR assays and has been tested frequently.

2.3 ToxCast analogous *in vitro* assays

AC_{50}s for a broad selection of *in vitro* assays were extracted from the ToxCast database (Feshuk et al., 2023). Only AC_{50}s with active hit-calls (reflecting a good quality of the data) were included in the dataset. AC_{10}s (active concentrations at which 10% of the response is observed; Figure 1) were calculated based on the reported AC_{50} (inflection point of the Hill curve), top response value, and slope of the curve. This calculation can be found in the Appendix I. This recalculation was done to get the AC values mor in line with those of the Nrf2 CALUX (LOEC) and PXR CALUX (p10/p20).

The similarity (or analogy) between any two assays is determined based on having a biological similar endpoint and the correlation between the REP values of commonly tested chemicals between the CALUX and any ToxCast assay. Some additional conditions are set (see Figure 1). For eligible assays, the reference compound is required to be one of the tested chemicals. And, at least 10 chemicals need to overlap between two assays to ensure the REP correlation is based on sufficient observations. No requirement was set for the chemicals to belong to

different chemical, toxicological or structural groups. A test for ranking is a spearman correlation test. We assume that, if the spearman correlation is high (> 0.5) and significant (p value < 0.05), the REPs between *in vitro* assays from ToxCast are similar enough to assume that they act as each other's analogues. Therefore, any new compound can be assumed to have a similar REP between the assays. A second selection is made for the overall similarity in REP values between assays for commonly tested chemicals. If the average percentage likeness in REP values is really low overall (less than 0.25 with 1 the highest, towards 0 the lowest) the assays are not considered as analogues after all.

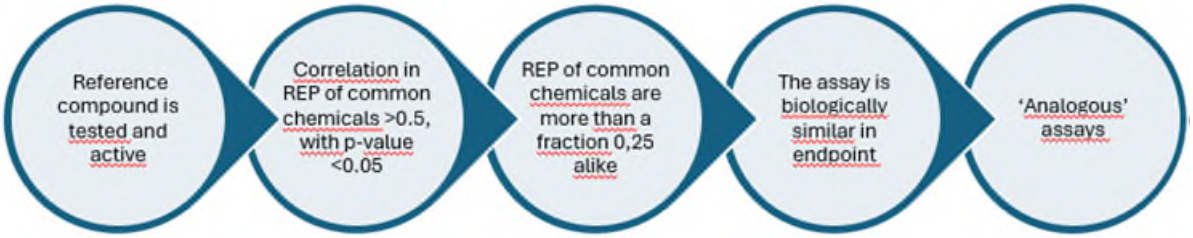


Figure 1: Selection criteria for determining analogy between assays.

The method cannot be used when the assays are NOT similar/analogous even if the assays in theory test the same endpoint: when the potency ranking between compounds (i.e. the sequence of compounds when ordering the compounds based on REP) is not similar, it cannot be assumed that any untested compound will have a similar response in the other assay. Assays that test the same endpoint but are not analogous, may be caused for instance by a slight difference in induction of the mechanism of PXR activation in the assay, or a difference in cell-line.

Note: There is an inherent variability in the measured REP values of any chemical tested in a single assay, possibly due to experimental circumstances or chance effects. On average over all assays, a REP value of a chemical that is tested on separate occasions (and present as a separate AC_{10} in the data) is a fraction of 0.60 alike within the same assay. This means that a REP value of 0.7 can, on average, be anywhere between 0.42 and 1. The reason for (extra) variability in REP between a single compound tested in different assays can be the lenient correlation requisite of 0.5 ($p<0.05$) which means there can be specific outliers to this correlation.

2.4 Tackling the different reference compound problem

For the analyses in this report, reference compounds were chosen based on two practical considerations: they should have been tested in potential analogous assays and they should be relatively potent (e.g. relevant) in the assays (= low AC_{10} value). In assays applied to water quality testing, however, the reference compound is not flexible because it is chosen by the laboratory that performed the analyses. In theory, these reference compound equivalent concentrations can be recalculated for an individual assay, based on a different reference compound, resulting in reference chemical equivalent concentrations. However, for this recalculation the effect potencies of both reference chemicals in the assay should be known. The effect concentration (for analogous assays only) can in that case be recalculated by:

$$Ref.Eq_y = Ref.Eq_x \cdot \left(\frac{REPx}{REPy}\right) \qquad Equation\ 6$$

in which $Ref.eq_y$ refers to the reference equivalent concentrations based on reference compound y, and $Ref.eq_x$ refers to the reference equivalent concentrations based on reference compound x. The REPx is, by default, 1. The same equation can be used to recalculate one EBT value of a single assay into another EBT value for the assay, with another reference compound. Ref.Eq can be replaced by EBT in Equation 6, in that case.

Two factors are in practice limiting to apply this calculation. For any assay, both (or 'all' for more flexibility) reference compounds will have to be tested. It is therefore recommended that a fixed list of relevant chemicals is tested in any assay with a common endpoint so that the bioactivities of chemicals can be easily compared across assays. This list can be 4-6 substances long. Secondly, individual chemicals are mostly tested in only a single time in a single study in any assay. To reduce uncertainty and get a more robust REPs – and thus EBTs - chemicals should be tested in assays more often in independent studies, on different occasions and in different laboratories (with slightly different experimental set-ups) to get information on the variability and median of the REP. In that case REP can be relied on or at least the variability is clear.

3 Results

3.1 Nrf2 CALUX

3.1.1 Nrf2 CALUX and analogue *in vitro* assays from ToxCast

For the Nrf2 CALUX, AC_{10} data for analogous *in vitro* assays are available in ToxCast after re-calculation (see Appendix I). For the four different potential reference compounds, analogous *in vitro* assays have been identified in ToxCast, all of which include this reference compound (Table 1).

Several assays are similar in REP values of common tested chemicals, however, only ATG_NRF2_ARE_CIS_up (Table 1) is a biological analogue, targeting the NRF2 receptor. The other assays with similar REP values (not shown) are not considered analogous assays: BSK_hDFCGF_IP10_down is an assay that tests for target gene CXCL10 (Chemokine ligand 10) and this can be used to screen for chemicals with anti-inflammatory effects (Fay et al., 2018). CCTE_Shafer_MEA_dev_LDH_dn, BSK_hDFCGF_Proliferation_down and Tox21_DT40 (DNA repair deficient isogenic chicken DT40 cell viability) are non-specific endpoints (Ji et al., 2009, Naidenko et al., 2021, Carstens et al., 2022). In general, assay non-specific endpoints are cytotoxicity, proliferation, or viability (Naidenko et al., 2021). BSK_LPS_VCAM1_down is an assay with the target gene involved in adhesion of molecules that mediate the interaction between leukocytes and other cell types. These mediate the immune cell cross-talk and the orchestration of the innate and adaptive response against pathogens (Naidenko et al., 2021). CCTE_Padilla_ZF_144hpf_TERATOSCORE_up is a zebrafish embryo assay that is used to identify developmental effects.

Table 1: BpGLV of substances in assays analogous to Nrf2 CALUX. The rank activity for reference compound in the last column is for illustration. A potency rank closer to zero indicates a relatively potent reference chemical. The BpGLV in bold is selected as the lowest (adjusted) EBT for Nrf2 CALUX in reference chemical equivalents.

	BpGLV	CHEM	CHEM Ac10 (ug/L)	CHEM pGLV (ug/L)	CHEM REP	potency rank ref. chem.
Reference compound Carbendazim						
ATG_NRF2_ARE_CIS_up	143,8	2,4-Dimethylphenol*	734	140	1,03	0,44
ATG_NRF2_ARE_CIS_up	95,5	Dimethomorph	2765	350	0,27	0,44
ATG_NRF2_ARE_CIS_up	169,8	Spiroxamine	777	175	0,97	0,44
ATG_NRF2_ARE_CIS_up	141,2	Bifenazate	374	70	2,02	0,44
ATG_NRF2_ARE_CIS_up	185,2	Alachlor	81	20	9,26	0,44
ATG_NRF2_ARE_CIS_up	186,6	Prometon	424	105	1,78	0,44
ATG_NRF2_ARE_CIS_up	134,6	Thiophanate-methyl	784	140	0,96	0,44
ATG_NRF2_ARE_CIS_up	139,4	Bentazone	1136	210	0,66	0,44
ATG_NRF2_ARE_CIS_up	95,5	Etridiazole	790	100	0,95	0,44
ATG_NRF2_ARE_CIS_up	186,7	1,2-Benziso-thiazolin-3-one	424	105	1,78	0,44
ATG_NRF2_ARE_CIS_up	100,4	Thiobencarb	300	40	2,51	0,44
ATG_NRF2_ARE_CIS_up	85,7	Iprodione	880	100	0,86	0,44

ATG_NRF2_ARE_CIS_up	135,3	Triadimefon	501	90	1,50	0,44
ATG_NRF2_ARE_CIS_up	137,0	Vinclozolin*	963	175	0,78	0,44
ATG_NRF2_ARE_CIS_up	179,3	Metolachlor	42	10	17,93	0,44
ATG_NRF2_ARE_CIS_up	192,7	Carboxin	219	56	3,44	0,44
ATG_NRF2_ARE_CIS_up	135,2	Dazomet	390	70	1,93	0,44
ATG_NRF2_ARE_CIS_up	**79,6**	2,3,4,6-Tetrachlorophenol	948	100	0,80	0,44
ATG_NRF2_ARE_CIS_up	106,7	Fenarimol	283	40	2,67	0,44
ATG_NRF2_ARE_CIS_up	178,5	Hexachlorocyclopentadiene	177	42	4,25	0,44
ATG_NRF2_ARE_CIS_up	206,4	2,4,5-Trichlorophenol	2557	700	0,29	0,44
NRF2_CALUX	236,1	Anilazine*	3297	700	0,34	0,41
Reference compound Curcumin						
ATG_NRF2_ARE_CIS_up	429,5	2,4-Dimethylphenol*	734	140	3,07	0,70
ATG_NRF2_ARE_CIS_up	285,2	Dimethomorph	2765	350	0,81	0,70
ATG_NRF2_ARE_CIS_up	422,0	Bifenazate	374	70	6,03	0,70
ATG_NRF2_ARE_CIS_up	402,2	Thiophanate-methyl	784	140	2,87	0,70
ATG_NRF2_ARE_CIS_up	416,5	Bentazone	1136	210	1,98	0,70
ATG_NRF2_ARE_CIS_up	285,3	Etridiazole	790	100	2,85	0,70
ATG_NRF2_ARE_CIS_up	300,0	Thiobencarb	300	40	7,50	0,70
ATG_NRF2_ARE_CIS_up	256,0	Iprodione	880	100	2,56	0,70
ATG_NRF2_ARE_CIS_up	404,4	Triadimefon	501	90	4,49	0,70
ATG_NRF2_ARE_CIS_up	409,2	Vinclozolin*	963	175	2,34	0,70
ATG_NRF2_ARE_CIS_up	403,9	Dazomet	390	70	5,77	0,70
ATG_NRF2_ARE_CIS_up	**237,7**	2,3,4,6-Tetrachlorophenol	948	100	2,38	0,70
ATG_NRF2_ARE_CIS_up	318,8	Fenarimol	283	40	7,97	0,70
NRF2_CALUX	457,4	Anilazine*	3297	700	0,65	0,49
Reference compound Dichlorvos						
ATG_NRF2_ARE_CIS_up	142,1	Dimethomorph	2765	350	0,41	0,53
ATG_NRF2_ARE_CIS_up	142,1	Etridiazole	790	100	1,42	0,53
ATG_NRF2_ARE_CIS_up	127,5	Iprodione	880	100	1,27	0,53
ATG_NRF2_ARE_CIS_up	**118,4**	2,3,4,6-Tetrachlorophenol	948	100	1,18	0,53
NRF2_CALUX	146,0	Anilazine*	3297	700	0,21	0,31
Reference compound TBH						
ATG_NRF2_ARE_CIS_up	24,9	2,4-Dimethylphenol*	734	140	0,18	0,17
ATG_NRF2_ARE_CIS_up	16,5	Dimethomorph	2765	350	0,05	0,17
ATG_NRF2_ARE_CIS_up	29,4	Spiroxamine	777	175	0,17	0,17
ATG_NRF2_ARE_CIS_up	43,4	Endothal	301	100	0,43	0,17
ATG_NRF2_ARE_CIS_up	24,5	Bifenazate	374	70	0,35	0,17

ATG_NRF2_ARE_CIS_up	32,1	Alachlor	81	20	1,60	0,17
ATG_NRF2_ARE_CIS_up	32,3	Prometon	424	105	0,31	0,17
ATG_NRF2_ARE_CIS_up	23,3	Thiophanate-methyl	784	140	0,17	0,17
ATG_NRF2_ARE_CIS_up	24,1	Bentazone	1136	210	0,11	0,17
ATG_NRF2_ARE_CIS_up	16,5	Etridiazole	790	100	0,17	0,17
ATG_NRF2_ARE_CIS_up	32,3	1,2-Benziso thiazolin-3-one	424	105	0,31	0,17
ATG_NRF2_ARE_CIS_up	17,4	Thiobencarb	300	40	0,43	0,17
ATG_NRF2_ARE_CIS_up	41,2	Triclosan	1044	329	0,13	0,17
ATG_NRF2_ARE_CIS_up	14,8	Iprodione	880	100	0,15	0,17
ATG_NRF2_ARE_CIS_up	23,4	Triadimefon	501	90	0,26	0,17
ATG_NRF2_ARE_CIS_up	23,7	Vinclozolin*	963	175	0,14	0,17
ATG_NRF2_ARE_CIS_up	31,1	Metolachlor	42	10	3,11	0,17
ATG_NRF2_ARE_CIS_up	33,4	Carboxin	219	56	0,60	0,17
ATG_NRF2_ARE_CIS_up	23,4	Dazomet	390	70	0,33	0,17
ATG_NRF2_ARE_CIS_up	**13,8**	2,3,4,6-Tetrachlorophenol	948	100	0,14	0,17
ATG_NRF2_ARE_CIS_up	18,5	Fenarimol	283	40	0,46	0,17
ATG_NRF2_ARE_CIS_up	30,9	Hexachlorocyclopentadiene	177	42	0,74	0,17
ATG_NRF2_ARE_CIS_up	44,8	Bisphenol A	1020	350	0,13	0,17
ATG_NRF2_ARE_CIS_up	44,1	Clomazone	2756	931	0,05	0,17
ATG_NRF2_ARE_CIS_up	35,8	2,4,5-Trichlorophenol	2557	700	0,05	0,17
NRF2_CALUX	46,1	Anilazine*	3297	700	0,07	0,10

** Compound was tested in Nrf2 CALUX*

In Table 1, each lowest EBT per reference compound is indicated in bold. The corrected EBT for the analogous ToxCast assays is lower than the EBT derived for Nrf2 CALUX for three of the reference compounds (Table 1). The cause for a lower EBT in the analogous assay compared to the NRf2 CALUX lies in the fact that the lowest REP-pGLV combination (See Eq. 3 and 4) in these analogous assays is based on a compound that was not tested in the Nrf2 CALUX. It can be assumed that this chemical would be potent in Nrf2 CALUX as well, because the potencies of commonly tested compounds have shown to be correlated (Figure 3). This is why, from a precautionary perspective, it is proposed to take into account that the EBT value can be as low as these values. The EBT to take into account per reference compound are summarized and listed in Table 2. Two compounds with lower BpGLV than the EBT for Nrf2 CALUX were tested in the Nrf2 CALUX: vinclozolin (not selected; LOEC 108276.5 ug/L, REP 0.0002 TBH) and 2,4-Dimethylphenol (selected; LOEC 462 ug/L, REP 0.47 TBH).

Figure 2 shows, as an illustration, the (natural log-transformed) REPs of tested compounds in the Nrf2 CALUX assay compared to REPs of the same compounds tested in the ATG_NRF2_ARE_CIS_up assay from ToxCast, indicating the variability of REPs of compounds between these assays.

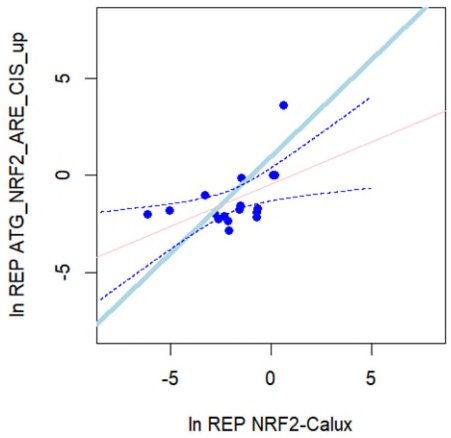


Figure 2: Ln Relative Effect Potencies (REP) of compounds tested in both assays (Nrf2 and ToxCast analogue). Light blue line: optimal correspondence (perfect diagonal). Dotted dark blue lines: 90 percentile confidence interval. Red line: regression line.

It may occur that a reference compound is not tested in analogous ToxCast assays, but other reference compounds do lead to a lower EBT than the Nrf2 CALUX itself. With Equation 6 we can still derive an EBT to take into account for that reference compound for Nrf2 CALUX. As an example we show the recalculation of the EBT in TBH equivalents to Curcumin equivalents. When the reference compound is curcumine for the Nrf2 CALUX, the REP of curcumine is 1 and the REP of tert-butylhydroquinone is 9.93 (it is more potent). The REP of all four reference compounds in Nrf2 CALUX can be found in Appendix II. The calculation of the TBH-EBT equivalent for curcumine is:

$$35.5\ (EBT\ TBH) = 352.6\ (EBTcurcumin) \cdot \left(\frac{1\ (REP\ curcumin}{9.93\ (REP\ TBH}\right) \qquad Equation\ 6$$

This means that if we have a new tentative EBT for TBH (Table 1), we can calculate the corresponding tentative EBT in curcumin equivalents:

$$13.8 = Ref.Eq_x \cdot \left(\frac{1}{9.93}\right) \qquad Equation\ 6$$

$$Ref.Eq_x = \left(\frac{13.8\ (EBTc\ TBH)}{0.1007}\right) = 137.0\frac{ug}{L}CEQ \qquad Equation\ 6$$

In Table 2 we provide the calculated EBTs for each reference compound. This leaves in total four EBT per reference compound; one from the analogous assay and three calculated from the other reference compounds. The calculated EBT are not exactly the same per reference compound. This is caused by a new tested chemical causing a lower EBT or by differences in REP of substances in another assay. To have one EBT, the EBT can be adjusted to the lowest of these EBT of all reference compounds.

Table 2: Overview of the tentative EBT in reference compound equivalents for the Nrf2 CALUX assay calculated for different reference compounds. EBTc is the lowest EBT in an analogue. Calculated EBT is the recalculated EBTc from another reference

compound (Equation 6) (reference compound in brackets). Tentative EBT is the lowest over EBTc and equivalent calculated EBTc.

CALUX ASSAY	REFERENCE COMPOUND	CAS NUMBER	EBT (UG/L)	EBT_c (UG/L)	CALCULATED EBT (UG/L) (EQ. 8)	TENTATIVE EBT (UG/L)
NRF2-CALUX	Dichlorvos	62-73-7	146.0	118.4	43.7 (TBH) 49.4 (Carb) 76.1 (Curc)	**43.7**
NRF2-CALUX	Tert-butylhydroquinone	123-31-9	46.1	13.8	15.9 (Carb) 23.8 (Curc) 37.9 (DiV)	**13.8**
NRF2-CALUX	Curcumine	458-37-7	457.4	237.7	137.0 (TBH) 154.4 (Carb) 370.6 (DiV)	**137.0**
NRF2-CALUX	Carbendazim	10605-21-7	236,1	79.6	123.6 (Curc) 70.8 (TBH) 191.8 (DiV)	**70.8**

3.2 PXR CALUX

3.2.1 PXR CALUX and analogues *in vitro* assays from ToxCast

For the PXR CALUX, AC_{10} data for analogous *in vitro* assays are available in ToxCast. Table 3 lists per reference compound the analogous assays and the chemicals with lower BpGLV (Equation 4) than the PXR CALUX EBT. Only for reference compound Azoxystrobin analogous assays were found. Metalachlor was not tested in the PXR CALUX and neither Nicardipine or DEHP was tested or fit the criteria of analogy (see Figure 1) in any of the ToxCast assays.

Two other assays (in addition to the assays described for the NRF2 CALUX) show similarity in REP but do not test the PXR receptor binding (not shown). This is the APR_HepG2_CellLoss_72h_dn; it measures cell proliferation. The other one is ATG_NRF2_ARE_CIS_up, which measures NRF2 receptor binding (oxidative stress) and is the Nrf2 CALUX analogue.

Table 3: BpGLV of substances in assays analogous to PXR CALUX. The rank activity for reference compound in the last column is for illustration. A potency rank closer to zero indicates a relatively potent reference chemical. The BpGLV in bold is selected as the lowest (adjusted) EBT for PXR CALUX in reference chemical equivalents.

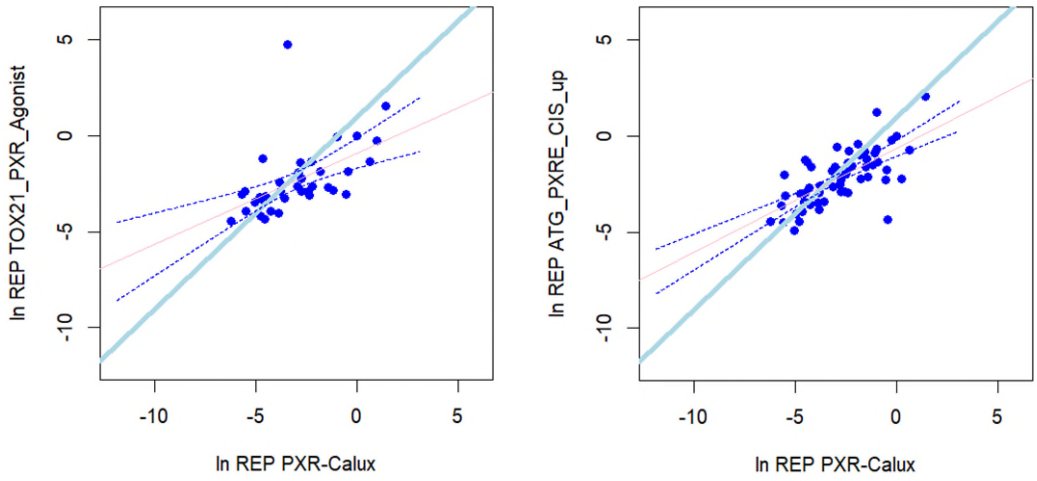
	EBT	CHEM	CHEM AC10 (ug/L)	CHEM pGLV (ug/L)	CHEM REP	potency rank reference chemical
Reference compound Nicardipine						
PXR_CALUX	**3.5**	Oxadiazon*	217	25.2	0.14	0.01
Reference compound Azoxystrobin						
ATG_mPXR_XSP1_up	6.4	Oxadiazon*	98	25.2	0.25	0.26
ATG_mPXR_XSP1_up	5.0	Triclosan*	1640	329	0.02	0.26
ATG_mPXR_XSP1_up	5.9	Acetochlor	106	25.2	0.23	0.26
ATG_mPXR_XSP1_up	8.7	Vinclozolin*	497	175	0.05	0.26
ATG_mPXR_XSP1_up	3.6	Fenthion	48	7	0.52	0.26
ATG_mPXR_XSP1_up	6.8	Metformin	1435	392	0.02	0.26
ATG_mPXR_XSP2_up	8.1	Tebuconazole*	319	210	0.04	0.08
ATG_mPXR_XSP2_up	6.3	Bromuconazole	137	70	0.09	0.08
ATG_mPXR_XSP2_up	4.0	Metconazole	217	70	0.06	0.08
ATG_mPXR_XSP2_up	2.9	Triclosan*	1385	329	0.01	0.08
ATG_mPXR_XSP2_up	**2.0**	Iprodione	600	100	0.02	0.08
ATG_mPXR_XSP2_up	3.2	Pendimethalin	76	20	0.16	0.08
ATG_mPXR_XSP2_up	6.8	Fenthion	13	7	0.97	0.08
ATG_mPXR_XSP2_up	2.2	4-tert-Butylphenol*	2733	490	0.00	0.08
ATG_PXRE_CIS_up	8.7	8-Hydroxyquinoline	2411	350	0.02	0.05
ATG_PXRE_CIS_up	6.9	Thiabendazole	6034	700	0.01	0.05
ATG_PXRE_CIS_up	8.1	Bifenazate	516	70	0.12	0.05
ATG_PXRE_CIS_up	6.8	Propyzamide	616	70	0.10	0.05
ATG_PXRE_CIS_up	6.3	Thiobencarb	381	40	0.16	0.05

ATG_PXRE_CIS_up	8.4	Azamethiphos	1245	175	0.05	0.05
ATG_PXRE_CIS_up	7.3	Iprodione	818	100	0.07	0.05
ATG_PXRE_CIS_up	7.5	Iodoform	801	100	0.07	0.05
ATG_PXRE_CIS_up	7.8	Myclobutanil	1335	175	0.04	0.05
PXR_CALUX	9.3	Oxadiazon*	217	25.2	0.37	0.04
Reference compound DEHP						
PXR_CALUX	**18.1**	Oxadiazon*	217	25.2	0.72	0.07

** Compound was tested in PXR CALUX*

From the analogous assays (Table 3) for the reference compound Azoxystrobin, ATG_mPXR_XSP2_up has the lowest EBTc. Iprodione was not tested in the PXR CALUX but because the assays are correlating in REP, it can be assumed that this compound will react similarly in the PXR CALUX. For that reason it is proposed to tentatively set the EBT for the PXR CALUX to 2.0 ug/L Azoxystrobin equivalents, until iprodione and other compounds with lower BpGLV are tested in the PXR CALUX to confirm this lower EBT. EBT to take into account per reference compound are summarized in Table 4.

Figure 3 shows the (natural log-transformed) REPs of tested compounds in the PXR CALUX assay compared to REPs of the same compounds tested in analogous assays from ToxCast, indicating the variability of REPs of compounds deviating from their analogous assays in ToxCast.



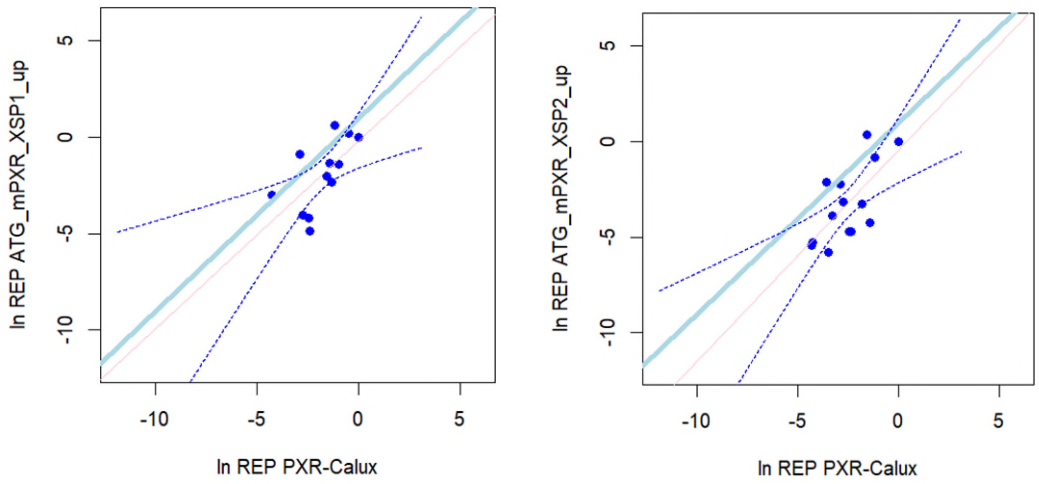


Figure 3: Ln Relative Effect Potencies (REP) of compounds tested in both assays (PXR and ToxCast) for reference compound Azoxystrobin. Light blue line: optimal correspondence (perfect diagonal). Dotted dark blue lines: 90 percentile confidence interval. Red line: regression line.

In Table 4 the Nicardipine and DEHP-based equivalent calculated concentrations (column 'Calculated EBT (EQ. 8) (UG/L)') and the original EBT for the reference compound Axozystrobin are logically the same (9.3, although a small rounding difference is present) because these are derived from the same assay, PXR CALUX. The Azoxystrobin-based calculated EBT are different, because this EBT is calculated via the EBTc of the analogous assay.

Table 4: Tentative EBT in Reference compound equivalents for the PXR CALUX assays. EBTc is the lowest EBT in an analogue, corrected for the potency of the reference compound. Calculated EBT is the recalculated EBTc from another reference compound (Equation 6) (reference compound in brackets). Tentative EBT is the lowest over EBTc and equivalent calculated EBTc.

CALUX ASSAY	REFERENCE COMPOUND	CAS NUMBER	EBT (UG/L)	EBT_c (UG/L)	CALCULATED EBT (EQ. 8) (UG/L)	TENTATIVE EBT
PXR-CALUX	Metolachlor*	51218-45-2	n/a	n/a	n/a	n.a
PXR-CALUX	Azoxystrobin	131860-33-8	9.3	2.0	9.3 (Nic) 9.4 (DEHP)	**2.0**
PXR-CALUX	Nicardipine	55985-32-5	3.5	3.5	0.74 (Azo) 3.4 (DEHP)	**0.7**
PXR-CALUX	DEHP**	117-81-7	18.1	18.1	3.9 (Azo) 18.1 (Nic)	**3.9**

** This chemical was not reported in the PXR CALUX dataset*

*** This chemical was not reported in the ToxCast dataset in analogous assays*

3.3 Comparison to bioactivity in water quality monitoring

In research programs DPWE and TKI (Bertelkamp et al., 2019; Bertelkamp et al, 2020; Hofman-Caris et al., 2023) bioactivity values were collected for drinking water in different stages of treatment, and sewage treatment. For Nrf2 CALUX these are reported in curcumine equivalent concentrations and for PXR CALUX in nicardipine equivalent concentrations. In Figure 4 the collected response values and how they relate to the derived EBT are visualized, for both assays.

The derived Nrf2 EBT in this report of 137 ug/L Ceq is above of what is generally measured, even in waste water influent (category STP in Figure 4). This means that for the collected data, generally water is considered safe regarding the mode of action that mediates response in the Nrf2 CALUX ('oxidative stress'), except in relatively rare cases. The Nrf2 CALUX has twice been observed to have a response above 200 ug/L Ceq (Figure 4). This happens in surface water with only pre-treatment and also after oxidation treatment (both in category 'DWp' in Figure 4). Groundwater is less polluted as compared to surface water and this is also apparent from the response values (category 'GWp' in Figure 4). Should the Nrf2 CALUX ever exceed the derived EBT, it will be worthwhile to follow up because this will be an exceptional event.

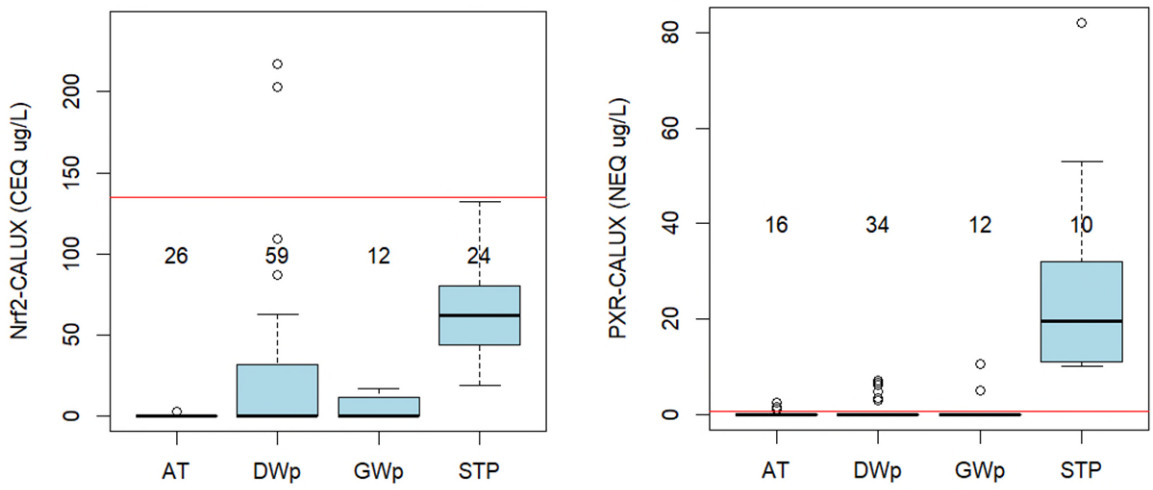


Figure 4. PXR and Nrf2 CALUX values in practice. The red line is the derived tentative EBT. AT stands for 'After treatment' for drinking water (e.g. after filtering over granular activated carbon, or membrane filtering). DWp stands for surface water as source and in production. GWp stands for ground water as source and in production. STP stands for sewage treatment water. The numbers above the boxplots indicate the number of observations in each category.

For the PXR CALUX, slightly less data is available (Figure 4). The PXR CALUX does exceed the EBT derived in this report for all categories, and in all cases for STP (Figure 4). This means that in three of sixteen instances the PXR is activated by substances after drinking water treatment (category AT in Figure 4) at concentrations for which risks cannot be excluded in these samples. During treatment (categories DWp and GWp in Figure 4) these exceedances are higher. In sewage treatment water the response of the PXR CALUX is always above the derived tentative EBT.

In a separate non-public dataset, responses in Dutch surface waters were collected by BDS, including waters influenced by sewage treatment effluent. This dataset (not shown) indicates that the NRF2 CALUX EBT in Ceq is surpassed in 10% of all 378 measurements in various surface waters. For PXR, on the other hand, the EBT in Neq

is surpassed in 85% of the 723 tested waters. This means that in these waters, PXR activation is in many cases at such a level that in theory the concentration of the chemical used for the derivation of the EBT (and other chemicals like it) could be at toxic values. This does not always mean that this most toxic chemical is actually causing the response. However, a risk cannot be excluded. It is recommended that the chance on encountering a compound at toxic concentrations at a PXR assay response above EBT is calculated (Pronk et al., 2024) once the EBT has been confirmed after testing of extra substances (Table 1, Table 3) in PXR and Nrf2 CALUX.

4 Discussion

In this report a method is described to derive tentative EBT for data-poor assays by including information from similar, analogous *in vitro* assays from ToxCast. The method is based on the notion that assays for a similar biological endpoint and in which similar chemicals elicit a sufficiently similar response can be used to fine-tune derived EBT for assays with only a low available number of tested chemicals with provisional guideline values for drinking water. Namely, if two assays have a similar response, a tested chemical in one assay can be expected to have a similar effect in the other assay. To account for variability in relative potency of the reference compound, the EBT in reference compound equivalents is calculated from EBT of different reference compounds, and the lowest is selected as a tentative EBT.

For the Nrf2 CALUX a tentative EBT of 137 µg Ceq/L was derived in this report. For the PXR CALUX a tentative EBT of 0.7 µg Neq/L was derived. Other EBTs have recently been derived for the PXR CALUX and the Nrf2 CALUX. STOWA (unpublished results) has chosen practical EBTs based on clean reference waters. If the assay response is above the signal found in such clean reference waters, it is a sign that these have been influenced by chemicals. This warrants further investigation. For the PXR CALUX a practical EBT of 16 µg Neq/L was derived. As the EBT derived in this report (0.7 Neq/L, based on laboratory experiments) is lower than the EBT derived based on reference waters, this means that chemicals in the clean reference waters already show a response at which risk for human health cannot be excluded. For the Nrf2 CALUX the STOWA-derived EBT was 110 µg Ceq/l. The Nrf2 EBT in the present study was 137 µg Ceq/l and this means that no threshold is surpassed in clean reference waters for this endpoint. Van der Oost et al. (2017) earlier also derived EBTs for signaling risk in ecotoxicity, for Nrf2 and PXR CALUX. These EBTs were also practical EBT based on the background BEQ, because the derived ecotoxicity trigger values based on Species Sensitivity Distributions (SSDs) were too low for practical use. Since responses measured at polluted sites were close to the background BEQ, the proposed EBT for PXR CALUX from van der Oost et al. (2017) was 3 µg Neq/L, which is approximately twice the background BEQ. Again, this EBT of van der Oost et al. (2017) does not indicate low micropollutant risks and should be considered as an indicator of overall chemical stress. Since responses measured at polluted sites were close to the background BEQ for the Nrf2 CALUX, the proposed EBT for overall oxidative stress activity was 10 µg Ceq/L, which is approximately twice the background BEQ value (van der Oost et al., 2017).

Although the Nrf2 CALUX EBT at its current tentative derived value cannot distinguish between water types (e.g. drinking water and sewage treatment water) this is not the purpose of a health-based trigger value. The purpose is to signal water quality that is either of low risk and water quality in which risks cannot be excluded. Values are high enough to expect that treated drinking water will only on rare occasions give a response above EBT.

Some shortcomings in the data were identified. A number of compounds were tested several times in the same ToxCast assay and the results (AC_{10}s and thus REPs) were variable. The different values may be caused by separate experiments, in different laboratories, using a different experimental set-up. In ToxCast, it is not possible to readily discern separate experiments and this means the REP of chemicals cannot be calculated with their 'own' reference compound AC_{10} (i.e. the reference compound chosen by the laboratory in the unique experiment). For our calculations of relative effect potencies we took the mean of all the reported AC_{10} per compound tested in the assay. In principle it is a good thing if compounds are tested more often in the same assay on different occasions. Because the variability in REP was established in this report as a mean fraction of 0.6 alike, more tests will average out such variability and the REP can be trusted more readily. pGLV availability was a limiting factor for EBT derivation. In this report several pGLV are calculated based on collected established maximum daily intake values. It was not investigated if the standard allocation factor of 20% needed adjustment. Moreover, in general, pGLV represent the most critical health-adverse endpoint and in future research an effort can be made on the explicit

alignment of pGLV relevant to the assay endpoint. In addition, physiologically based pharmacokinetic (PBPK) modelling and quantitative in vitro to in vivo extrapolation (QIVIVE) may be used to better estimate if the AC_{10} and pGLV values significantly differ from a toxicological perspective.

The choice of effect size reporting is also different between studies. Some report AC_{50}, others AC_{10} or AC_{20}. In this report the AC_{10} values were re-calculated from the AC_{50} values with help of the slope. Logically AC_{10} values are lower than AC_{50} values. Because REPs are calculated with respect to the reference compound and this is a relative measure, such differences do not influence the derived EBT in principle. However, AC_{50} values will tend to have a wider range of values compared to AC_{10} values.

The choice of reference compound for the derivation of EBT can be variable, and in practice based on practical considerations like 'most tested' and 'potent'. The EBT, according to the calculation followed in this report, can be derived quite easily for different reference compound equivalent concentrations, as long as they have been tested. However, for water quality testing, generally only one reference compound is tested by individual laboratories. We provided a formula to recalculate assay responses in different reference compound equivalents, that can be used in a single assay. Different results of effect-based water quality tests between analogous/similar assays can be compared most accurate based on a concurrent tested unique reference compound. The assay community should come together to decide on reference chemicals of choice, and a habit should be adopted by laboratories to test several reference compounds with any batch of tested compounds in any assay, and report all underlying results for these compounds.

An open question is what similarity is enough to use an assay for deriving EBT values for an assay of interest. In this report values were chosen that can be considered the most 'lenient' possible. With more stringent values (i.e. deeming assays analogues based on higher correlation thresholds), very few assays were found analogous to these two CALUX assays in this report. Similarity will increase if REP of compounds can be determined based on the reference chemical tested in the same experiment. Again, to provide flexibility and comparability in the choice of reference compound, this requires that laboratories test different reference compounds. A better way to assess similarity of REP between assays is to use a linear regression. The error between REP can be determined more accurately in the cases where the REP relation is not completely diagonal (1:1). Moreover, the REP of one assay can be more accurately recalculated to the expected REP of the other assay based on the equation of the linear regression. In follow up research, this can be tested.

The compounds used in the derivation of the EBT were not checked on their water-relevance. Also, the compounds were not checked on their actual permitted use in Europe. The assumption here is that any compound can end up in the water system at some point in time in a location. By taking into account all compounds the EBT is extra prudent. For correlation analysis in general, more data means the more robust a correlation can be established. However, because we aim to compare trigger values for data poor assays this means by default that data availability is generally limiting. This is why the requirement of having at least 10 overlapping substances is the minimum. If the assay of interest is very data-poor, and less than 10 substances were tested, no similarity can be calculated. The requirement of having biological similarity will prevent 'false positive' similarities. A last check if the EBT are justly lowered based on the data from analogous assays is to test the substances of the analogous assays on which the lower EBT were based on. Although it is expected that these will give a similar response in the CALUX assays, experimental evidence will prove that.

If an EBT is exceeded, it means that risk cannot be excluded. However, it was shown previously that the lower the concentrations, the lower the predictive value of assays is (Pronk et al., 2024). If an EBT is exceeded, it does not necessarily mean that a toxic concentration of chemicals is reached. A practical solution to prevent too many unnecessary follow ups when exceeding an EBT is to take into account the severity of the exceedance. With higher exceedances, there is a higher predictive value. With the current derived EBT for the Nrf2 CALUX follow ups are expected to be few. The response is generally far below the derived EBT. For PXR CALUX the number of follow ups

will be high. It is a practical choice to select the EBT exceedance level to follow up on, and this depends on the risks involved versus the costs and effort involved in a follow-up.

Recommendations for drinking water companies

The current report results in several concrete recommendations to drinking water companies. These are listed below.

- Bioassay monitoring in drinking water with Nrf2 and PXR CALUX can be implemented with the expectation of a tentative EBT. Substances that were tested in analogous assays that caused a lowering of the EBT for the CALUX assays should be tested in the CALUX assays to confirm their expected relevance. These are 15 substances for the PXR CALUX and 18 substances for the Nrf2 CALUX (Appendix III).
- Effort should be directed, or obligations should be negotiated, to report bioactivity of individually tested compounds in (water)laboratories for assays used in water quality testing in an open accessible database, like the NORMAN bioactivity database (BADB) (https://www.norman-network.com/nds/badb/).
- When testing individual substances in biologically analogous assays, several common reference compounds should at least be tested. For the Nrf2 and PXR endpoints, four compounds each are suggested in this report. Ideally the four compounds should be re-tested with each batch of new compounds tested, to confirm their REP values.
- In any case, substances should all ideally be tested multiple times in different experiments/laboratories to get a robust average REP of a substance in an assay.
- Analogy for assays of other relevant endpoints can be established to be able to compare these, as was done for the Nrf2 and PXR CALUX. With knowledge on the analogy of assays an informed choice can be made for using any particular assay that is available and performs well in predicting risks based on exceeding the EBT. Moreover, with the establishment of analogous assays per endpoint, candidate substances to elicit a response in tested water samples can be checked against all compounds tested in such analogous assays for a broader view of candidates.
- Depending on costs and frequency of exceedances of the EBT, combined with a calculation of the predictive value of the assay (Pronk et al., 2024), a rule for determining if a follow-up is worth the effort can be decided (for instance follow-up above a certain response threshold that is higher than the EBT) for the PXR CALUX.
- A further improvement of the methodology for establishing analogy between assays is to use the fit of a linear regression between two assays rather than correlation. This enables to put a threshold on the acceptable deviation of REP between the assays, and the formula can be used to re-calculate REP of one assay into the REP of the other assay if the slope of the correlation is not exactly 1.
- The possibility of selecting only pGLV for substances with some proven relevance for the tested assay endpoint can be investigated.

Acknowledgements
With many thanks for constructive feedback received from Corine Houtman (HWL) and Jaimy de Schepper (HWL) in their role as linking-pins from the Theme group Chemical Safety.

5 References

Baken, K., Sjerps, R. M., Schriks, M., & van Wezel, A. P. (2018). Toxicological risk assessment and prioritization of drinking water relevant contaminants of emerging concern. *Environment international, 118*, 293-303. doi:10.1016/j.envint.2018.05.006

Battino, M., Giampieri, F., Pistollato, F., Sureda, A., de Oliveira, M. R., Pittalà, V., Fallarino, F., Nabavi, S. F., Atanasov, A. G., & Nabavi, S. M. (2018). Nrf2 as regulator of innate immunity: A molecular Swiss army knife! *Biotechnology Advances, 36*(2), 358-370. doi:https://doi.org/10.1016/j.biotechadv.2017.12.012

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

Bertelkamp C, Brunner A, Dingemans M, Houtman CJ, Kolkman A, Siegers WG, Arondeus K, Harmsen D, Wols BA (2019) DPWE Robuustheid uitvoering doseerproeven 2017/2018 –Resultaten van doelstof-analyses, non-target screening en bioassays Herziene versie 2020 KWR 2019.040 https://livelink.kwrwater.nl/livelink/livelink.exe?func=ll&objaction=overview&objid=61197043

Bertelkamp, C. Dingemans, M.M.L. Roest, K. Hornstra, L.M. Hofman-Caris, C.H.M. Reus, A.A. (2020) TKI Sluiten watercyclus Noord-Holland. KWR rapport - KWR 2020.027

Brand, W., de Jongh, C. M., van der Linden, S. C., Mennes, W., Puijker, L. M., van Leeuwen, C. J., van Wezel, A. P., Schriks, M., & Heringa, M. B. (2013). Trigger values for investigation of hormonal activity in drinking water and its sources using CALUX bioassays. *Environ Int, 55*, 109-118. doi:10.1016/j.envint.2013.02.003

Carstens KE, Amy F Carpenter, Melissa M Martin, Joshua A Harrill, Timothy J Shafer, Katie Paul Friedman, Integrating Data From In Vitro New Approach Methodologies for Developmental Neurotoxicity, Toxicological Sciences, Volume 187, Issue 1, May 2022, Pages 62–79, https://doi.org/10.1093/toxsci/kfac018

Escher, B., Ait-Aissa, S., Behnisch, P.A., Brack, W., Brion, F., Buchinger, S., Crawford, S., Hamers, T., Hettwer, K., Hilscherova, K., Hollert, H., Kase, R., Kienle, C., Legradi, J., Tuerk, J., van der Oost, R., Vermeirssen, E. and Neale, P.A. (2018) Effect-based trigger values for in vitro and in vivo bioassays performed on surface water extracts supporting the environmental quality standards (EQS) of the European Water Framework Directive. Science of The Total Environment 628–629: 748-765 https://doi.org/10.1016/j.scitotenv.2018.01.340

Escher, B. I., Neale, P. A., & Leusch, F. D. (2015). Effect-based trigger values for in vitro bioassays: Reading across from existing water quality guideline values. *Water Res, 81*, 137-148. doi:10.1016/j.watres.2015.05.049

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(2):500-515. doi: 10.1093/toxsci/kfy049.

Feshuk, M., Kolaczkowski, L., Dunham, K., Davidson-Fritz, S. E., Carstens, K. E., Brown, J., Judson, R. S., & Paul Friedman, K. (2023). The ToxCast pipeline: updates to curve-fitting approaches and database structure. *Front Toxicol, 5*, 1275980. doi:10.3389/ftox.2023.1275980

Hofman-Caris, C.H.M. Siegers, W.G. Harmsen, D.J.H. Reus, A.A. (2023) Sluiten van de watercyclus. Synergistisch effect van de combinatie van O3 en CMF. KWR rapport - KWR 2023.026

Houtman, C., Kroesbergen, J., Behnisch, P., Brouwer, A., & Felzel, E. (2017). Bioassays als alternatief voor chemische monitoring van PCB's en PAK's: H2O-online.

Ji K, Kogame T, Choi K, Wang X, Lee J, Taniguchi Y, Takeda S. (2009) A novel approach using DNA-repair-deficient chicken DT40 cell lines for screening and characterizing the genotoxicity of environmental contaminants. Environ Health Perspect. 117(11):1737-44. doi: 10.1289/ehp.0900842

van der Linden, S. C., Von Bergh, A. R., van Vught-Lussenburg, B. M., Jonker, L. R., Teunis, M., Krul, C. A., & Van Der Burg, B. (2014). Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 760*, 23-32. doi:10.1016/j.mrgentox.2013.09.009

Naidenko OV, Andrews DQ, Temkin AM, Stoiber T, Uche UI, Evans S, Perrone-Gray S (2021) Investigating Molecular Mechanisms of Immunotoxicity and the Utility of ToxCast for Immunotoxicity Screening of Chemicals Added to Food. Int J Environ Res Public Health 18(7):3332. doi: 10.3390/ijerph18073332.

Neale PA, Escher BI, de Baat ML, Enault J, Leusch FDL. Effect-Based Trigger Values Are Essential for the Uptake of Effect-Based Methods in Water Safety Planning. Environ Toxicol Chem. 2023 Mar;42(3):714-726. doi: 10.1002/etc.5544. Epub 2023 Feb 8. PMID: 36524849.Oladimeji, P. O., & Chen, T. (2018). PXR: More Than Just a Master Xenobiotic Receptor. *Mol Pharmacol, 93*(2), 119-127. doi:10.1124/mol.117.110155

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

Pieterse, B., Rijk, I. J. C., Simon, E., van Vugt-Lussenburg, B. M. A., Fokke, B. F. H., van der Wijk, M., Besselink, H., Weber, R., & van der Burg, B. (2015). Effect-based assessment of persistent organic pollutant and pesticide dumpsite using mammalian CALUX reporter cell lines. *Environmental Science and Pollution Research, 22*(19), 14442-14454. doi:10.1007/s11356-015-4739-5

Pronk TE, Hoondert RPJ, Kools SAE, Kumar V, de Baat ML. Bioassay predictive values for chemical health risks in drinking water. Environ Int. 2024 Jun;188:108733. doi: 10.1016/j.envint.2024.108733. Epub 2024 May 10. PMID: 38744044.

Pronk, T.E., M.L. de Baat, S.J.P. van den Berg, R. van der Oost (2021). Achtergronddocument Basis-Set Bioassay Selectie. Achtergronddocument beschikbare kennis bij de sleutelfactor Toxiciteit. Versie 1, 21 december 2021. KIWK-Toxiciteit Notitie. Amersfoort, Nederland. Kennis Impuls Water Kwaliteit https://www.sleutelfactortoxiciteit.nl/sites/default/files/2023-01/1640087700-basis-set-bioassays-selectiev6%283%29.pdf

Reus A, Hoondert R, Shaikh S (2023) Effectsignaalwaarden voor eenduidige interpretatie van genotoxiciteitstesten BTO 2023.052 | Juni 2023 https://livelink.kwrwater.nl/livelink/livelink.exe?func=ll&objaction=overview&objid=70614230&logStopConditionID=1119085_2121151497_2_open

U.S. EPA (2023) ToxCast & Tox21 Summary Files from invitrodb_v3. Retrieved from https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data on October 28, 2023. Data released October 2018.

WHO (2022) Guidelines for drinking-water quality: fourth edition incorporating the first and second addenda. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO.

I Calculation of AC50 to AC10

The underlying fitted Hill curve used by the ToxCast algorithm to obtain the AC_{50} is as follows:

$$y = \frac{tp}{1 + 10^{(AC_{50} - x) \cdot gw}} \qquad Equation\ I1$$

in which *y* is the observed bioactivity response, *tp* is the top response level (the top of the curve), *x* is the chemical concentration in the assay, and *gw* is the slope. This equation can be rewritten to obtain the AC_{10} (the value at which 10% activity occurs in the assay):

$$AC_{10} = AC_{50} - \frac{log10(10\% - 1)}{gw} \qquad Equation\ I2$$

These recalculated AC_{10} data were formatted to only include chemicals for which the AC_{50} does not exceed the solubility of the chemical. Figure I1 shows calculated AC_{10}s for all ToxCast assays included in this study, plotted against their respective reported AC_{50}s. Basing our EBTs on AC_{10}s instead of AC_{50}s will likely lead to an increased number of chemicals included, due to the selection criterion in Equation 1 ($pGLV/AC_{10} > 0.1$), as AC_{10}s are always lower than AC_{50}s. However this $pGLV/AC_{10}$ ratio may differ considerably between compounds across assays compared to $pGLV/AC_{50}$ ratios due to variability in the slopes of the fitted Hill models (*gw*). This means that other compounds may be excluded depending on the use of AC_{10} or AC_{50} data.

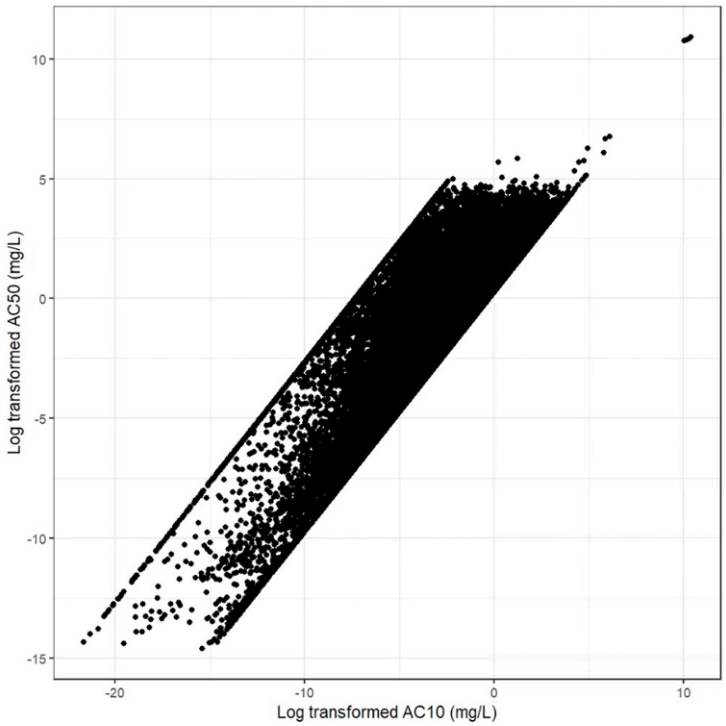


Figure I1: Log transformed AC_{50} values (mg/L) from the in vitro *assay analogues (for Nrf2 CALUX and PXR CALUX) from ToxCast compared to the calculated AC_{10} values, based on the reported slope (and ceiling value) of the dose-response curves.*

II REP values of substances

Table IIa. REP values (Equation 2) in the Nrf2 CALUX for potential reference compounds, per reference compound.

	REP CURCUMINE	REP TERT-BUTYLHYDROQUINONE	REP CARBENDAZIM	REP DICHLORVOS
CURCUMINE	1	9.93	1.94	3,13
TERT-BUTYLHYDROQUINONE	0.10	1	0.20	0.32
CARBENDAZIM	0,52	5.13	1	1.62
DICHLORVOS	0.32	3.17	0.62	1

Table IIb. REP values (Equation 2) in the PXR CALUX for potential reference compounds, per reference compound.

	REP AZOXYSTROBIN	REP NICARDIPINE	REP DEHP	REP METOLACHLOR
AZOXYSTROBIN	1	2,67	0,52	n.a.
NICARDIPINE	0,37	1	0,19	n.a.
DEHP	1,94	5,17	1	n.a.
METOLACHLOR	n.a.	n.a.	n.a.	n.a.

III Substances to test in PXR and Nrf2 CALUX

CAS	CHEM	Assay	Use
2634-33-5	1,2-Benzisothiazolin-3-one	Nrf2 CALUX	biocide
58-90-2	2,3,4,6-Tetrachlorophenol	Nrf2 CALUX	Fungicide
95-95-4	2,4,5-Trichlorophenol	Nrf2 CALUX	Fungicide
15972-60-8	Alachlor	Nrf2 CALUX	Herbicide
101-05-3	Anilazine	Nrf2 CALUX	Fungicide
25057-89-0	Bentazone	Nrf2 CALUX	Herbicide
149877-41-8	Bifenazate	Nrf2 CALUX	acaricide
80-05-7	Bisphenol A	Nrf2 CALUX	Industrial
5234-68-4	Carboxin	Nrf2 CALUX	Fungicide
81777-89-1	Clomazone	Nrf2 CALUX	Herbicide
533-74-4	Dazomet	Nrf2 CALUX	Pesticide
110488-70-5	Dimethomorph	Nrf2 CALUX	Fungicide
145-73-3	Endothal	Nrf2 CALUX	Herbicide
2593-15-9	Etridiazole	Nrf2 CALUX	Fungicide
60168-88-9	Fenarimol	Nrf2 CALUX	Fungicide
77-47-4	Hexachlorocyclopentadiene	Nrf2 CALUX	precursor
36734-19-7	Iprodione	Nrf2 CALUX	Fungicide
51218-45-2	Metolachlor	Nrf2 CALUX	Herbicide
1610-18-0	Prometon	Nrf2 CALUX	Herbicide
118134-30-8	Spiroxamine	Nrf2 CALUX	Fungicide
28249-77-6	Thiobencarb	Nrf2 CALUX	Insecticide
23564-05-8	Thiophanate-methyl	Nrf2 CALUX	Fungicide
43121-43-3	Triadimefon	Nrf2 CALUX	Fungicide
3380-34-5	Triclosan	Nrf2 CALUX	Bactericide
148-24-3	8-Hydroxyquinoline	PXR CALUX	Bactericide
34256-82-1	Acetochlor	PXR CALUX	Herbicide
35575-96-3	Azamethiphos	PXR CALUX	Insecticide
149877-41-8	Bifenazate	PXR CALUX	Acaricide
116255-48-2	Bromuconazole	PXR CALUX	Fungicide
55-38-9	Fenthion	PXR CALUX	Insecticide
75-47-8	Iodoform	PXR CALUX	Desinfectant
36734-19-7	Iprodione	PXR CALUX	Fungicide
125116-23-6	Metconazole	PXR CALUX	Fungicide
657-24-9	Metformin	PXR CALUX	Pharmaceutical
88671-89-0	Myclobutanil	PXR CALUX	Fungicide

40487-42-1	Pendimethalin	PXR CALUX	Herbicide
23950-58-5	Propyzamide	PXR CALUX	Herbicide
148-79-8	Thiabendazole	PXR CALUX	Fungicide
28249-77-6	Thiobencarb	PXR CALUX	Insecticide