BTO 2017.008 | October 2017

BTO report

Effect-based monitoring with bioassays – a roadmap

KWR Watercycle Research Institute

BTO

Effect-based monitoring with bioassays – a roadmap

BTO 2017.008 | October 2017

Project number
400554-180

Project manager
Stefan Kools

Client
BTO - Thematical research - New monitoring concepts

Quality Assurance
Annemarie van Wezel

Author(s)
Kirsten Baken, Milou Dingemans

Sent to
BTO participants

Meer informatie
Milou Dingemans

T (030) 606 9637
E milou.dingemans@kwrwater.nl

KWR
PO Box 1072
3430 BB Nieuwegein
The Netherlands

KWR Watercycle Research Institute

BTO *Managementsamenvatting*

Effect-gestuurd monitoren met bioassays – een wegwijzer voor de selectie en het gebruik van bioassays voor chemische waterkwaliteit

Auteur: dr. Kirsten Baken, dr. Milou Dingemans

Een stappenplan beschrijft de selectie en evaluatie van *in vitro* bioassays voor de chemische waterkwaliteit, dat verder ontwikkeld kan worden tot een standaardprotocol om de prestaties en toepasbaarheid van nieuwe kandidaat *in vitro* bioassays empirisch te testen. Daarmee kan de toegevoegde waarde van nieuwe bioassays gestructureerd worden vergeleken met analysemethoden die al in gebruik zijn. *In vitro* bioassays vormen een waardevolle aanvulling op de analytische chemie bij het beoordelen van de chemische waterkwaliteit omdat ze op een efficiënte manier inzicht geven in het gezamenlijke effect van mengsels microverontreinigingen. Omdat er veel verschillende *in vitro* bioassays beschikbaar zijn, is gestructureerde selectie van groot belang. Voor dit onderzoek is een verkennende studie met de ToxTracker assay voor genotoxiciteit uitgevoerd.

Belang: *in vitro* bioassays veelbelovend voor monitoring van chemische waterkwaliteit

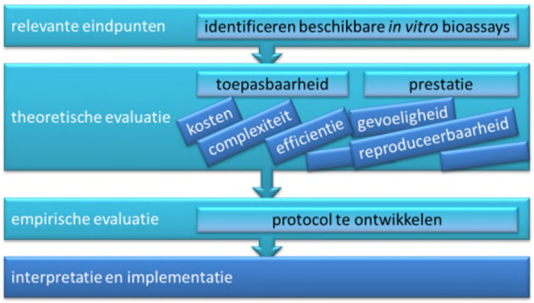
In vitro bioassays geven een biologische respons op de aanwezigheid van mengsels microverontreinigingen in bijvoorbeeld drinkwater(bronnen). Daarvoor hoeven die microverontreinigingen niet geïdentificeerd te zijn. *In vitro* bioassays vormen daarmee een aanvulling op de bekende analytisch-chemische methoden. De Nederlandse watersector erkent de toegevoegde waarde van *in vitro* bioassays en past een beperkte selectie van dergelijke testen toe voor beoordeling van de chemische waterkwaliteit. Recent is ook in de EU Kader Richtlijn Water ruimte ontstaan om *in vitro* bioassays in te zetten voor monitoring van de waterkwaliteit. Er is volgens de Nederlandse drinkwatersector een plek voor *in vitro* bioassays in risico-gebaseerde monitoring, bijvoorbeeld bij het prioriteren van stoffen voor meetprogramma's, bij het bepalen van trends in ruimte en tijd en bij het beoordelen van de efficiëntie van de waterbehandeling. Er is behoefte aan een gestructureerde aanpak voor de keuze, interpretatie en implementatie van *in vitro* bioassays bij de bewaking van drinkwaterkwaliteit.

Aanpak: samenwerking, literatuuronderzoek, enquête en een pilot experiment

Binnen de context van internationale onderzoeksconsortia EDA-EMERGE en DEMEAU is onderzoek gedaan naar de selectie van monstervoorbewerkingstechnieken en *in vitro* bioassays. Op deze resultaten is voortgebouwd binnen dit BTO-onderzoek. Via literatuuronderzoek en bezoek van (inter)nationale expertbijeenkomsten is kennis verzameld over de cruciale aspecten bij de keuze en uitvoering van *in vitro* bioassays. Een enquête bij verschillende partijen binnen de Nederlandse watersector gaf inzicht in de visies over toepassing van *in vitro* bioassays. In de loop van het project werd het ToxTracker genotoxiciteit-assay gekozen om de toepasbaarheid daarvan empirisch te toetsen. Relevante aspecten voor de beoordeling en implementatie van *in vitro* bioassays in waterkwaliteitsonderzoek zijn door onderzoek naar de ToxTracker geïllustreerd.

Resultaten: stappenplan voor succesvolle inzet van *in vitro* bioassays

Resultaat van het onderzoek is een stappenplan voor inzet van *in vitro* bioassays (zie figuur):



Stappenplan voor de implementatie van in vitro bioassays

Stap a. Bioassay selectie. Er zijn veel verschillende (soorten) *in vitro* bioassays beschikbaar, gericht op diverse biologische effecten die verontreinigingen kunnen veroorzaken: de "biologische eindpunten". De meest relevante biologische eindpunten voor (drink)watercontaminanten zijn samengevat. Het gaat dan bijvoorbeeld om genotoxiciteit en hormoonverstoring. Ook is een scoringmethodiek ontwikkeld op basis van toepasbaarheid en prestaties om de *in vitro* bioassays te vergelijken en te selecteren. Hoeveel benodigde informatie daarvoor beschikbaar is, blijkt sterk te variëren tussen de verschillende bioassays.

Stap b. Empirische evaluatie van de kandidaat in vitro bioassay. Een pilotstudie met de ToxTracker liet zien dat deze test voldoende gevoelig is om effecten van voorbeeldstof arseen in concentraties beneden diens wettelijke norm aan te tonen. Ook was de test in staat om de effecten van mengsels van microverontreinigingen te meten. Dit leverde handvatten op voor een gestructureerde vormgeving van de evaluatie van *in vitro* bioassays.

Stap c. Interpretatie van bioassay effect data. Momenteel zijn meerdere methoden in omloop voor de interpretatie van de resultaten van *in vitro* bioassays op watermonsters van onbekende samenstelling (met effect-grenswaarden/trigger values). Het ontwikkelen van een uniforme aanpak kan zorgen voor onbetwistbare conclusies over de chemische waterkwaliteit en bijdragen aan wettelijke verankering van het gebruik van bioassays in de waterkwaliteitsmonitoring.

Naar verwachting zullen doorlopend nieuwe *in vitro* bioassays worden ontwikkeld. Met het verschijnen van nieuwe stoffen in drinkwaterbronnen kunnen ook aanvullende biologische eindpunten relevant worden. Dit noopt tot continue verkenning van de bruikbaarheid van *in vitro* bioassays voor de drinkwatersector.

De Nederlands/Vlaamse drinkwaterbedrijven passen *in vitro* bioassays op dit moment in variërende mate toe en hebben in verschillende mate interesse in deze methoden. Er worden mogelijkheden gezien voor het toepassen van *in vitro* bioassays bij onderzoek naar de gezamenlijke effecten van mengsels van stoffen.

Implementatie: ontwikkel standaardprotocol voor selectie-, implementatie- en interpretatie

Aanbevolen wordt het hier ontwikkelde stappenplan verder te ontwikkelen tot een standaard protocol om de prestaties en toepasbaarheid van nieuwe kandidaat *in vitro* bioassays empirisch te testen. Daarmee kan de toegevoegde waarde van aanvullende bioassays gestructureerd worden vergeleken met analysemethoden die al in gebruik zijn. Om *in vitro* bioassays te implementeren in de monitoring van chemische waterkwaliteit, moet een respons van een watermonster van onbekende samenstelling vertaald worden naar de mogelijke relevantie voor de humane gezondheid. Ook daarvoor moet een uniforme methode worden opgenomen in het standaard protocol. Ook is verder onderzoek nodig naar het effect van verschillende monstervoorbehandelingsmethoden op de resultaten van *in vitro* bioassays.
Op het gebied van de ontwikkeling van *in vitro* bioassays zelf kan de watersector voordeel hebben van verdergaande miniaturisering en automatisering, omdat dit de snelheid en kosten-effectiviteit vergroot en de koppeling tussen analytische chemie en *in vitro* bioassays vergemakkelijkt.

Rapport

Dit onderzoek is beschreven in het rapport BTO 2017.008 *Effect-based monitoring with bioassays – a roadmap*

Meer informatie
Milou Dingemans

T (030) 606 9637
E milou.dingemans@kwrwater.nl

KWR
PO Box 1072
3430 BB Nieuwegein
The Netherlands

KWR Watercycle Research Institute

Summary

There is growing support for the implementation of *in vitro* bioassays, parallel to chemical analyses, in the context of (drinking) water quality monitoring. Clear advantages to include *in vitro* bioassays is the provided insight in the risks associated with exposure to complex low-level mixtures of pollutants in water. This report describes the state-of-the-art for different crucial steps in the selection, implementation and interpretation of *in vitro* bioassays for chemical water quality monitoring in the context of human health related risks and drinking water production, and proposes clear procedures and protocols. These steps are illustrated by a case study concerning application of the ToxTracker assay.

A plethora of *in vitro* bioassays is available that can be used to test effects on different biological processes and in different types of models. The most relevant toxicological endpoints considered in this perspective are carcinogenesis, adverse effects on reproduction and development, effects on xenobiotic metabolism, modulation of hormone systems, reactivity and adaptive stress responses. Only few *in vitro* bioassays to study mechanisms related to effects on reproduction and development are available. A scoring matrix has been designed to evaluate (candidate) bioassays for their applicability and performance, including the coverage of different toxic mechanisms, cost-effectiveness, performance (sensitivity with regard to realistic environmental mixtures of chemicals in low concentrations, reproducibility), possibilities for high-throughput and ease of implementation with regard to laboratory requirements or specialist knowledge. Data for a list of reference chemicals could be shared in a database to allow comparisons between *in vitro* bioassay methods. Further miniaturization and automation may allow a more efficient collection and connection of analytical chemistry and bioassay information. Also, the impact of different SPE approaches on the presence of different (types of) chemicals in the concentrate needs to be more firmly established as it is known that different types of chemicals behave differently in SPE extraction.

Due to the high sensitivity of *in vitro* bioassays, responses can be expected (far) below exposure concentrations that are relevant for human health. Effect-based trigger values are therefore needed for bioassay interpretation. Different approaches are currently being followed including trigger values based on relative ecotoxicity potency, health-based threshold values for chronic exposure in humans and kinetics of reference chemicals, and read-across from guidelines. It is important that trigger values are sufficiently but not too conservative, to serve as indicators of potential health effects. Mechanistic manners of interpreting *in vitro* bioassay data based on the link between cellular and molecular effects *in vitro* and potentially associated adverse outcomes in the intact organism are expected to become feasible in the (near) future. It is expected that a common understanding of crucial steps in the selection, implementation and interpretation of *in vitro* bioassays including clear procedures and protocols, will facilitate legal embedding in the context of water quality monitoring for drinking water production. To support a statutory basis for the legal implementation of *in vitro* bioassays, it is recommended that an overview of effect-monitoring data is collected in a national database.

Contents

1 Introduction 4

2 Effect-based monitoring: bioassay selection 6

3 Empirical evaluation of a candidate bioassay 11

4 Effect-based trigger values 15
Effect-trigger values based on relative potency (Van der Oost et al., 2017a) 16
Effect-based trigger values based on reference chemical potency and kinetics (Brand et al., 2013) 18
Effect-based trigger values based on TTC (Mennes, 2004) 20
Effect-based trigger values read-across from existing guidelines (Escher et al., 2015) 20

5 Implementation in Dutch water sector 25

6 Regulations for risk-based monitoring using bioassays 26

7 Conclusions and the road ahead 30

8 References 35

1 Introduction

There are clear advantages to include *in vitro* bioassays in water quality monitoring strategies in parallel to chemical analysis.

There are several situations in which the inclusion of bioassays is informative and reduces uncertainty in risk and safety assessment. For instance when chemical identification is challenging, in particular in non-targeted analyses in which complex molecules observed may not be readily identified and thus substance-specific risk assessment is not possible. Bioassay responses of water extracts with known origin but unknown composition give insight in the risks associated with complex low-level mixtures of pollutants in water. This is another advantage over substance-specific monitoring and risk assessment, since it cannot be ruled out that the effect of a mixture is different compared to the combined effects of the individual constituents of the mixture. Bioassays have also been demonstrated to be sufficiently sensitive to detect effects of chemicals, in some cases below the detection limits of chemical analyses (e.g. Chapman et al., 2011). Bioassay measurements are also critical for new and emerging chemicals for which health based guidance values indicating safe exposure levels may not be available. Parallel chemical analysis prevents that other (new) chemicals, that do not give a response in the selected *in vitro* bioassays, are overlooked. The complementary contributions for human health risk assessment has been demonstrated in several case studies. In general there is consensus[1] that effects measured in bioassays are a valuable addition in the hazard and risk characterization of chemical exposures via water.

This report describes the state-of-the-art of the different crucial steps that are required in the selection, implementation and interpretation process of *in vitro* bioassays for water quality monitoring in the context of drinking water production. As there is a plethora of *in vitro* bioassays available that can be used to test effects on different biological processes in different types of models (e.g. cells from different organs or species), it is necessary to select the most relevant set of *in vitro* bioassays for water quality monitoring in the context of drinking water production (Chapter 2). Every new candidate *in vitro* bioassay should be empirically tested in more detail for its applicability, sensitivity with regard to realistic environmental mixtures of chemicals in low concentrations, reproducibility, coverage of different toxic mechanisms, cost-effectiveness, possibilities for high-throughput and ease of implementation with regard to laboratory requirements or specialist knowledge (Chapter 3). Due to the high sensitivity of *in vitro* bioassays, responses can be expected (far) below exposure concentrations that are relevant for potential effects on human health. To determine whether a bioassay response is relevant for human health, effect-based trigger values can be used for data interpretation (Chapter 4). Several of the Dutch water utilities already implemented *in vitro* bioassays in parallel to the routine monitoring, or consider it a valuable approach when further developed (Chapter 5), and there are several national and international initiatives to include *in vitro* bioassays in water safety regulations (Chapter 6). In the final chapter (7), the main conclusions are listed, as well as the main data gaps and challenges in the implementation of *in vitro* bioassays for water quality monitoring.

[1] KWR researchers contributed to the 2017 workshop organized by Global Water Research Coalition (GWRC; globalwaterresearchcoalition.net) and the NORMAN network (norman-network.net) on bioassays. This workshop was organized to get insight in the developments in bioassay applications and the aims and needs of researchers, stakeholders and end-users. Within the workshop contributors there was clear consensus on the advantages of the addition of bioassays to the water quality toolbox. Nevertheless, it is clear that the implementation of bioassays in water quality monitoring can be supported by the development of confidence in bioassays and guidance with regard to the interpretation of such data.

In the Attachments, additional efforts and detailed information collected in the context of this project have been included. For efficiency and to allow a quantitative comparison between the collected data, a water sample concentration method that can be used for both chemical-analytical analysis and effect-based testing in in vitro bioassays is preferred. Attachment I describes efforts to design such solid phase extraction methods. Attachment II includes the instructions for the bioassay scoring matrix. Attachments III and IV contain ToxTracker data for individual chemicals and water samples, respectively. Attachment V contains the abstract of a poster presentation of ToxTracker mixture data. Attachment VI includes the survey response and Attachment VII is the report of the KWR contribution to the GWRC workshop on bioassays.

In several chapters, the presented aspects of *in vitro* bioassay implementation are illustrated by a case study on the application of the ToxTracker assay.

2 Effect-based monitoring: bioassay selection

In *in vitro* bioassays, effects of chemical exposure on biological processes are tested in cell or tissue models. The results can be used as an indicator for the presence of particular (groups of) chemicals that may cause adverse health effects.

Exposure to chemicals may result in various biological effects on cells and molecules, that may ultimately result in adverse effects on human health. Such relationships between the interaction of a chemical on a cellular target and the events that are triggered on organ or organism level are described in toxicity pathways and, in a wider perspective, Adverse Outcome Pathways (Ankley et al., 2010; Villeneuve et al., 2014; Vinken 2013). Depending on the route of exposure and types of effects, different organ or physiological systems may be affected. For rapid screening of potential toxicological effects of chemicals in water, *in vitro* models are preferred over testing in intact organisms (*in vivo*). The preference for *in vitro* assays is based on cost- and time efficiency and ethical considerations, but also on the fact that smaller sample volumes are sufficient and that *in vitro* assays give information on specific toxicity pathways (National Research Council, 2007).

Toxicity pathways can be classified according to the type of interaction between a chemical and biological targets. A target can be a specific molecular entity, e.g. a (nuclear) receptor, or a wide range of molecules with which the chemical has a generic interaction, e.g. reaction with DNA bases.

TABLE 2.1. HEALTH EFFECTS AND RELATED TOXICITY MECHANISMS RELEVANT FOR DRINKING WATER SAFETY ASSESSMENT. SOURCE: SCHRIKS ET AL., 2015.

Health effects	**Toxicity mechanism**
xenobiotic metabolism	PXR activation
	AhR activation
modulation of hormone systems	estrogenicity
	anti-androgenicity
	glucocorticoid activity
reactivity	gene mutation
	chromosomal mutation
	DNA damage response
adaptive stress responses	ER stress
	heat shock
	hypoxia
	inflammation
	metal stress
	oxidative stress
reproductive toxicity	pre-implantation toxicity
	non-mechanistic assays (*in vivo*) including early life stages
	mechanistic assays to be included (several in early R&D stage)
developmental toxicity	non-mechanistic assays (*in vivo*) including early life stages
	mechanistic assays to be included (several in early R&D stage)

Cellular interactions can result in effects on cellular processes, resulting for example in oxidative stress. A chemical can only activate a toxicity pathway if the involved molecular and cellular entities are present in the exposed organ or species, and chemicals may activate multiple toxicity pathways simultaneously. The activation of a toxicity pathway depends on the exposure level, exposure duration and timing of exposure (e.g. during sensitive windows in development).

A wide range of mode of action categories has been observed for water-relevant chemicals (Busch et al., 2016). For a feasible day-to-day water quality monitoring approach, a selection was made for the main (relevant) toxicological endpoints (Table 2.1) based on 1) relevant human health effects and 2) bioactivity of chemicals present in drinking water and its sources (Escher et al., 2014). Toxicological effects that are considered to have the largest impact on human health and quality of life are carcinogenesis and adverse effects on reproduction and development. For the second aspect, results of a large inter-laboratory study were used in which effects of different types of water samples were tested in 103 different *in vitro* bioassays studying a broad range of toxicity pathways (Escher et al., 2014).

TABLE 2.2. SCORING MATRIX FOR *IN VITRO* BIOASSAYS. BRIEF INSTRUCTIONS ON THE USE OF THE SCORING MATRIX ARE INCLUDED IN ATTACHMENT I. SOURCE: SCHRIKS ET AL., 2015.

A. Assay applicability and ease of use (maximum combined score: 21)	
Criteria	**maximum score[a]**
applied to environmental samples	3
validated to water samples	3
standardized protocol available	3
service and support available	3
costs	3
ease of use	6[b]
non-GMO[c]	
no specialized equipment or skill required	
automation possible	
non-licensed *in vitro* (cell) model	
commercial kit available	
training available	
B. Assay performance (maximum combined score: 33)	
Criteria	**maximum score**
selectivity	3
accuracy	3
reproducibility	3
robustness	3
sensitivity	3
specificity	3
limit of detection	3
cytotoxicity control	3
speed	3
clear/straightforward read-out	3
high-throughput capacity	3

[a]Number of points represent poor (1 point), good (2 points) or excellent (3 points) score. [b]Score based on sub criteria (1 point each). [c]Genetically Modified Organism

The results demonstrated that the most responsive toxicity pathways were related to xenobiotic metabolism, modulation of hormone systems, reactivity and adaptive stress responses. Both modulation of hormone systems and reactivity may underlie carcinogenesis, while modulation of hormone systems may also be related to developmental effects. Reactivity may also impact reproduction and development if DNA in germ cells is affected. New endpoints (such as developmental toxicity, neurotoxicity or immunotoxicity) or emerging chemicals with alternative mechanisms-of-action may urge expansion of the *in vitro* bioassay battery (see Chapter 7).
A scoring matrix with a maximum score of 54 (Table 2.2) was designed to evaluate bioassays for their applicability and performance. The scores were based on information obtained from scientific literature, bioassay suppliers or expert judgement. Candidate bioassays per toxicity mechanism were identified based on their scores (Table 2.3).

TABLE 2.3. OVERVIEW OF SELECTED CANDIDATE *IN VITRO* BIOASSAYS FOR WATER QUALITY SCREENING. SOURCE: SCHRIKS ET AL., 2015.

Health effects	**Toxicity mechanism and candidate bioassays[ab]**
xenobiotic metabolism	PXR activation
	• HG5LN PXR assay (score: 42)
	• PXR HepG2 assay (score: 38)
	AhR activation
	• DR CALUX (score: 49)
	• AhR GeneBlazer (score: 37)
modulation of hormone systems	(anti-)estrogenicity
	• ERα CALUX (score: 48)[c]
	• Yeast Estrogen Screen (YES) assay (score: 44)
	(anti-)androgenicity
	• AR CALUX (score: 48)[c]
	• AR-MDA-kb2 (score: 42)
	(anti)glucocorticoid activity
	• GR CALUX (score: 45)
	• GR-MDA-kb2 (score: 40)
reactivity	gene mutation
	• Ames fluctuation assay (score: 42)[c]
	• ToxTracker (score: 34)
	chromosomal mutation
	• micronucleus assay (score: 36)
	• ToxTracker (score: 34)
	DNA damage response
	• UMUc assay (score: 44)
	• Vitotox (score: 42)
	• p53 CALUX (score: 37)
	• BlueScreen (score: 38)
adaptive stress responses	oxidative stress
	• Nrf2 CALUX (score: 35)
	• AREc32 assay (score: 39)
developmental toxicity	modulation of hormone systems
	• various nuclear receptor activation assays
	• H295R assay (score: 35)

[a]The individual assays are described in more detail in Schriks et al., 2015. [b]Other assays (with lower scores) are included in Schriks et al., 2015. [c]This assay is routinely included in water quality screenings at or for KWR Watercycle Research Institute.

Based on the collected scores (Table 2.3) the ToxTracker assay was selected to explore applicability for water quality testing of this assay. In this selection process it was also considered that a bioassay panel with similar assays is preferable for efficiency and quality control considerations.

The selection process and proposed panel of *in vitro* bioassays is described in more detail in the report 'Selection criteria to select *in vitro* bioassays for implementation and use' (Schriks et al., 2015). This report also includes descriptions of the different chemical exposure-related health effects and related mechanisms. In its current form, the scoring matrix can be used to compare *in vitro* bioassays. The scoring matrix can be further refined in the theoretical part of an evaluation process of *in vitro* bioassays prior to implementation by ranking the subcriteria on their impact (the maximum number of points could depend on the impact). Moreover, varying deviations between freely dissolved concentration and nominal concentration depending on *in vitro* bioassay format (e.g. medium constituents) may explain differences in sensitivity between *in vitro* bioassays (Fischer et al., 2017), and this should be included in the evaluation of *in vitro* bioassays. It is recommended that the scores and the underlying information are collected in a shared database. In the DEMEAU research it has also become clear that *in vitro* bioassays to study potential effects on neurotoxicity, immunotoxicity, reproduction and development that fulfil the criteria for applicability and performance (Table 2.2) are currently lacking and developments and innovations in these particular fields are of interest for water quality monitoring.

The tables can be used to evaluate individual *in vitro* bioassays for particular mechanisms, but when combining *in vitro* bioassays in a panel for water quality monitoring, it should also be taken into account that the bioassays cover the various types of toxic action, i.e. non-specific, specific and reactive toxicity, that the panel is cost-effectiveness in terms of equipment and consumables, and that the included *in vitro* bioassays perform well and can be implemented without high-tech laboratory requirements or specialist knowledge (Van der Oost et al., 2017a).

The ToxTracker bioassay (Hendriks et al., 2012, 2016) can be used to study the effects of chemicals exposure on biomarkers for direct (via DNA reactivity) and indirect (e.g. via oxidative stress) genotoxicity. In artificial chromosomes in the reporter cell lines, the gene code for GFP (green fluorescent protein) is fused to the biomarker genes. Induction of the biomarker genes therefore results in a bioluminescence signal directly related to the cellular response of interest. The reporter genes were selected in the development phase of ToxTracker, in which it was demonstrated that the activation of the reporter genes are associated with specific (geno-)toxic mechanisms. The ToxTracker bioassay was included in the DEMEAU evaluation of bioanalytical tools for 1) gene mutations and 2) chromosomal mutations. The ToxTracker received points for assay applicability and for assay performance. With regard to assay applicability, the maximum number of points were awarded for service and support, and points were also awarded for the availability of a standardized protocol and single points for 'applied for environmental samples', 'validated for environmental samples' and costs. Points were also awarded for 'ease of use' (automation is possible, a kit and training is available, however this is a GMO, specialized equipment is required and the cell model is licenced). With regard to assay performance, the maximum number of points were awarded for reproducibility, robustness, specificity, clear/straightforward read-out and high-throughput capacity. For cytotoxicity control and speed the ToxTracker was also awarded

points. No points could be awarded for selectivity, accuracy, sensitivity or limit of detection (due to lack of information).

Based on the combined score in the scoring matrix (Table 2.3), the ToxTracker was selected as a candidate *in vitro* bioassay as an alternative for the Ames fluctuation assay for gene mutations and the micronucleus test for chromosomal mutations. Further empirical evaluation (Chapter 3) has been performed to determine whether the ToxTracker is a candidate for implementation in water laboratories to screen chemical water quality (Schriks et al., 2015).

3 Empirical evaluation of a candidate bioassay

Based on the selection criteria (Chapter 2) a specific bioassay can be marked as a candidate bioassay to use for the screening of water quality, or to replace another assay that measures comparable endpoints. For example, there are numerous protocols and methodological approaches to be applied when environmental samples are tested for mutagenicity (reviewed in Umbuzerio et al., 2017).

The applicability of a candidate bioassay for implementation in water laboratories should be confirmed in an empirical evaluation process. In this proposed empirical evaluation process, different aspects should be included. These aspects are listed below, but it remains to be decided in what order these should be tested (and if go/no-go decisions are to be included). Moreover, it is recommended that the results of such empirical evaluations are collected in a shared data-base.

It needs to be established whether the candidate bioassay is compatible to test the effects of realistic low-level mixtures in (concentrated) water sample extracts without disturbing matrix effects, and if the sensitivity of the bioassay is sufficient to detect possible effects of chemicals at concentrations present in drinking water and its sources. For example, a recent study by Leusch et al. (2017) has demonstrated that sensitivity of *in vitro* bioassays for endocrine activity ranges widely. Also, concentration-dependent effects of relevant chemicals at relevant exposure concentrations can be studied in the candidate bioassay. The occurrence of concentration-dependent effects confirms a causal relationship between the exposure and an effect (Fedak et al., 2015). It is critical that a no-effect concentration and a lowest-observed effect concentration (LOEC) are included in these evaluations. The evaluation of an *in vitro* bioassay can also include an assessment of the intra- and inter-day variability in *in vitro* bioassay results. Such an analysis, using environmentally relevant environmental mixtures of estrogenic compounds, was performed for a set of *in vitro* bioassays for estrogenicity, demonstrating differences in precision and repeatability (Kunz et al., 2017). *In vitro* bioassay results can differ due to differences in experimental protocols, model organisms and data analysis. Harmonized standard procedures can improve the reproducibility (whether comparable results are obtained) which can be confirmed in interlaboratory studies (e.g. Di Paolo et al., 2016). The observed LOEC for a chemical in a particular assay can be placed in the context of 1) effect concentrations in other bioassays; 2) concentrations that correspond to water regulations, legislative parameters or health based guideline values. In the empirical parts of the evaluation process prior to implementation of a specific *in vitro* bioassay, a protocol to establish the sensitivity to detect effects of (relevant) regulated chemicals at their respective limit values can be included. This can also be expanded to include specific chemicals of concern at concentrations corresponding to provisional guideline values based on health effects.

The ToxTracker reporter cell lines were exposed following standard protocols (Hendriks et al., 2012, 2016) to seven relevant chemicals and extracts of drinking water, surface water and waste water treatment plant effluent. All experiments were conducted both in the absence and

presence of S9 liver extract to assess the impact of potential metabolic activation of chemicals. Potential effects of these exposures were tested on the expression of biomarker genes (between brackets) for general cell stress (Btg2), oxidative stress (Srxn1 and Blvrb), protein damage (Ddit3) and DNA damage (Bscl2 and Rtkn). The individual chemicals that were tested include arsenic, benzo(a)pyrene, bromate, bromoform, bromodichlormethane, N-nitrosodimethylamine (NDMA), and tetrachloroethene. Concentration-responsive effects were detected on the expression of biomarker genes. Different potencies are observed for the different chemicals, as demonstrated clearly for the effects on the most sensitive biomarker, oxidative stress biomarker Srxn1 (Figure 3.1). The lowest effective concentrations of all chemicals with regard to their effects on cell survival and the expression of all ToxTracker biomarker genes are included in Table 3.1 (concentration-response curves are included in Attachment III).

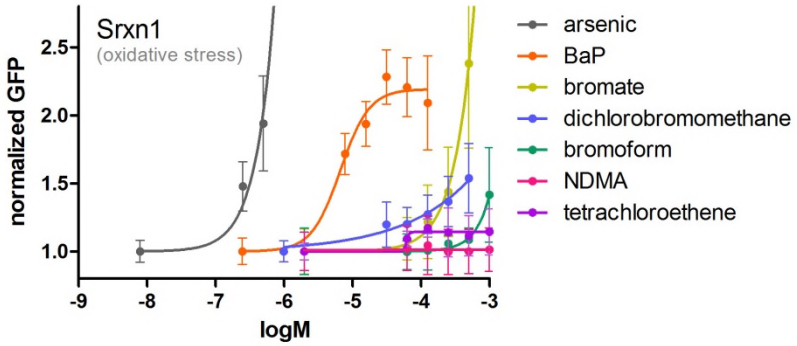


FIGURE 3.1. CONCENTRATION-DEPENDENT EFFECTS ON THE EXPRESSION OF OXIDATIVE STRESS MARKER SRXN1 (NORMALIZED GPF MARKER EXPRESSION) IN THE TOXTRACKER *IN VITRO* BIOASSAY.

For the two most potent chemicals (arsenic and benzo(a)pyrene), we aimed to evaluate the sensitivity of the ToxTracker by 1) comparing the LOEC with LOECs observed in other *in vitro* bioassays for genotoxicity, and 2) by comparing the concentrations at which effects are observed with those that reflect the guideline exposure in hypothetical water samples.

To compare the sensitivity of the ToxTracker assay for arsenic and BaP based on a comparison of the LOECs in other assays, data was collected from publically available databases with genotoxicity data, such as the a) GENETox database from US-EPA, b) EURL ECVAM Genotoxicity & Carcinogenicity Consolidated Database of Ames Positive Chemicals (Kirkland et al., 2014) and c) the ISSTOX Chemical Toxicity Databases of the Istituto Superiore di Sanità (Italy). Unfortunately, adequate data to compare LOECs in Ames was not available as these databases report mainly qualitative codes and scores. Also in literature, adequate data could not be found due to large differences in experimental set-ups and a lack of complete concentration-response curves. For optimal comparison it is recommended to test the effects of a set of model chemicals,

including arsenic and benzo(a)pyrene, in the Ames fluctuation test implemented for water quality monitoring at KWR.

TABLE 3.1. LOWEST OBSERVED EFFECT CONCENTRATIONS OBSERVED FOR DRINKING WATER RELEVANT CHEMICALS[A] ON CELL SURVIVAL AND THE EXPRESSION OF BIOMARKER GENES IN TOXTRACKER.

	arsenic	benzo(a) pyrene	bromate	bromo dichlor methane
DNA damage				
BScl2	4 µM	n.e.	1000 µM	n.e.
Rtkn	n.e.	n.e.	1000 µM	n.e.
Oxidative stress				
Srxn1	0.5 µM	7.8 µM	500 µM	500 µM
Blvrb	1 µM	n.e.	500 µM	n.e.
Protein damage				
Ddit3	4 µM	7.8 µM	n.e.	n.e.
General stress				
Btg2	4 µM	62.5 µM	500 µM	n.e.
Cell survival	2 µM	7.8 µM	250 µM	31.3 µM

[a]*Exposure to bromoform, NDMA en tetrachloorethene (up to 1000 µM) did not result in positive responses on biomarker gene expression. N.e.: no effect observed*

The ToxTracker was also used to test blindly the possible effects of a limited number of random water samples: drinking water, surface water and waste water treatment plant (wwtp) effluent. These water samples were enriched using large-volume solid-phase extraction (drinking water and surface water were concentrated 2000x, and wwtp effluent 1000x). These concentrates were diluted in the assay medium, with final enrichment factors ranging between 0.625 and 20. The concentrates induced concentration-dependent reporter gene responses. Figure 3.2 is an overview of the lowest final concentration factors in the assay at which responses are observed. The final concentration factor inducing effect decreases as expected with the number of chemicals (and levels thereof) in the extracts (unpublished) although a concise overview of the exposure data collected in the BE-BASIC project is pending. Exposure to surface water and wwtp effluent also resulted in reductions in cell survival, but most gene responses occur at concentrations not affecting cell survival. Based on these results, it can be concluded that the ToxTracker assay can be used to test effects of realistic mixtures of micropollutants in water concentrates. Moreover, it should be noted that (drinking) water samples are generally concentrated at least 10.000x for bioassay studies (Heringa et al., 2011), which would result in higher final enrichment factors (i.e. higher exposure concentrations) in *in vitro* bioassay such as the ToxTracker.

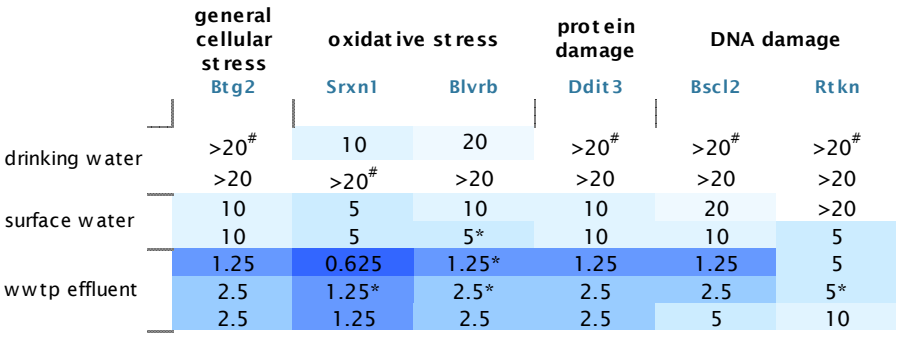


FIGURE 3.2. TOXTRACKER RESPONSES INDUCED BY A LIMITED NUMBER OF SAMPLES OF DRINKING WATER, SURFACE WATER AND WASTE WATER TREATMENT PLANT EFFLUENT, IN ABSENCE OF METABOLIC S9 MIXTURE. VALUES ARE THE LOWEST FINAL ENRICHMENT FACTORS IN THE ASSAY, AT WHICH POSITIVE RESPONSES ARE OBSERVED (WITH CORRESPONDING COLOR INTENSITY). MAXIMUM ENRICHMENT FACTORS ARE 10 (FOR WASTE WATER TREATMENT PLANT EFFLUENT) AND 20 (FOR DRINKING AND SURFACE WATER). # INDICATES THAT CONCENTRATION-DEPENDENT INCREASE IS OBSERVED (<1.5). * INDICATES THAT A RESPONSE WAS ALSO INDUCED IN THE PRESENCE OF METABOLIC S9 MIXTURE (AT THE HIGHEST ENRICHMENT ONLY). THE INDIVIDUAL DATA ARE INCLUDED IN ATTACHEMENT IV.

Considering that (drinking) water samples are generally concentrated at least 10.000x for bioassay studies (Heringa et al., 2011), and that these are diluted at least 100x in cell culture medium in the ToxTracker bioassay resulting in a final enrichment of 100x in the experiments, it can be calculated which theoretical concentration corresponds to Dutch limit values (Drinkwaterbesluit, 2011). For arsenic, the concentration that corresponds to the limit value is thus 13 µM (Dutch limit value: 10 µg/L ≈ 0.13 µM, enriched 100x). Effects of arsenic in the ToxTracker were already observed at a LOEC of 0.5 µM. For benzo(a)pyrene, the concentration that corresponds to the Dutch limit value is 0.004 µM (Dutch limit value: 0.01 µg/L ≈ $4*10^{-5}$ µM, enriched 100x). This is lower than the lowest concentration of benzo(a)pyrene tested in the ToxTracker (7.8 µM). Since clear effects of benzo(a)pyrene were observed at 7.8 µM, effects at lower concentrations cannot be excluded and this should be tested. This analysis indicates that the sensitivity of ToxTracker is sufficient to detect effects at and below the legislative parameter for arsenic.

4 Effect-based trigger values

In current water monitoring strategies, chemical water quality is mostly assessed using analysis of individual chemicals (chapter 1). Measured concentrations are related to (provisional) guideline values for individual chemicals in water. When derived for health safety purposes, such guideline values are based on health-based threshold values such as the acceptable daily intake of a chemical. Such threshold values are usually derived by regulatory agencies based on exposure levels resulting in adverse effects in experimental animal toxicity studies. Health-based threshold values are extrapolated from these exposure levels using a number of extrapolation factors to compensate for species differences, inter-individual differences, exposure duration and uncertainties or data gaps in toxicological effects. Many efforts are currently being undertaken in academia and research institutes to support the derivation of human health-based threshold values derived from toxicological and kinetic data obtained in *in vitro* bioassays (Allen et al., 2014; NRC, 2007; Wetmore et al., 2015).

Recently many developments in quantitative *in vitro* to *in vivo* extrapolation (Q-IVIVE) and reverse dosimetry have been reported (e.g. Groothuis et al., 2015; Louisse et al., 2017; Punt et al., 2013). Biologically effective (freely dissolved) concentrations can be modeled as they can deviate from nominal effect concentrations (the concentration added to the *in vitro* system) due to differences in chemical partitioning in *in vitro* bioassays (Fischer et al., 2017; Wetmore et al., 2015). The freely dissolved concentration in high-throughput assays can be used for high-throughput or more quantitative IVIVE approaches to derive oral equivalent doses (reviewed in Wetmore 2015; Yoon et al., 2015). These insights can be applied to use *in vitro* bioassay data as a first tier for screening for human health risk assessment. However, it is currently not (yet) feasible to use such approaches for routine risk monitoring of exposure to (mixtures of) chemicals via drinking water as this requires detailed exposure and *in vitro* bioassay effect information. It is however expected that in the near future, detailed reverse dosimetry analysis can be used to determine to which external exposure dose an effective *in vitro* concentrations corresponds. The main challenges relate to the analysis of effects of combined exposure and exposure to environmental mixtures of unknown composition.

The exact composition of a drinking water or environmental water sample concentrate is generally unknown. Nevertheless, responses of water samples observed in *in vitro* bioassay can be used to obtain insight in links to potential adverse health outcomes. A number of approaches have been proposed to interpret bioassay results (Escher and Leusch 2011, Leusch et al., 2017). The starting point is that it is only required to investigate further when a response of an undiluted water sample is observed above the level of quantification (e.g. deviation from negative control). When an effect is observed, effect-based trigger values can be used to establish whether the response observed in a bioassay is linked to a potential adverse health outcome. For this interpretation, different approaches have been proposed. In this chapter, these different approaches to derive effect-based trigger values and their applicability for interpretation of bioassay data are described. These include effect-based trigger values based on relative potency, the TTC, on reference chemical potency and kinetics, and read-across from guidelines.

By aiding in the interpretation of the responses induced by water samples of unknown chemical composition, effect-based trigger values can be used for prioritization for further

analysis and risk assessment or abatement processes. In this manner, trigger values unlock bioassay data for risk management and the management of abatement processes in drinking water production processes and regulation. It is important that trigger values are sufficiently conservative to serve as indicators of potential health effects, but should not be too conservative, to prevent that unnecessary studies are conducted to investigate further if preventive or remediatory actions are needed.

Effect-trigger values based on relative potency (Van der Oost et al., 2017a)

Effect-based trigger values (EBTs) have been designed for the Smart Integrated Monitoring (SIMONI) strategy, that is included in the conceptual framework of the Ecological Key Factors for the ecological assessment of water quality issues (Van der Oost et al., 2017a, 2017b). In the SIMONI framework (Figure 4.1) the outcome of a bioassay can be an indication for 'low risk' or 'potential risk'. The latter outcome is reason to proceed with further and more detailed analysis of the situation that may include chemical analysis, identification of potentially hazardous chemicals or more refined risk assessment.

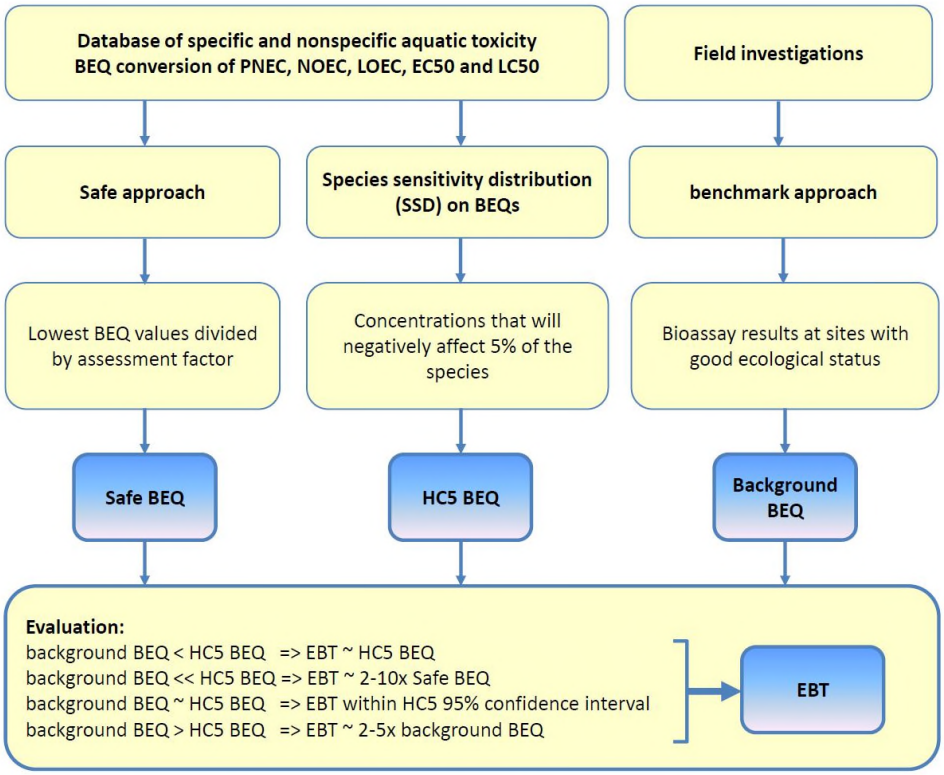


FIGURE 4.1. SCHEMATIC PRESENTATION OF THE SIMONI APPROACH FOR THE DERIVATION OF EFFECT-BASED TRIGGER VALUES (EBT); << MEANS AT LEAST 100 TIMES LESS (VAN DER OOST ET AL, 2017A).

For a selection of bioassays evaluating the most relevant toxicity mechanisms (Table 4.1), EBTs were derived to allow rapid risk identification of adverse effects of chemical exposure on aquatic organisms with the goal to protect ecosystems. These trigger values were derived using a bioanalytical equivalency (BEQ) approach. A chemical's BEQ is the exposure concentration of the reference chemical that would cause the same effect size as the chemical of interest, and can be calculated by multiplying the exposure concentration with its relative effect potency (REP) value. This approach can be used to calculate the combined BEQ of known mixtures. For unknown mixtures, for example those in water samples, the BEQ can be experimentally determined by investigating which amount of the reference compound

in this bioassay causes the same effect. To this aim, a concentration-response curve of the reference chemical should always be measured in parallel to the uncharacterized samples. To interpret whether the BEQ of such an unknown mixture indicates an environmental health hazard, trigger value BEQs (EBT BEQs) were derived as described below.

For each bioassay, a reference chemical and other relevant chemicals were identified based on their response in the assay at low concentrations. The relative potencies of these chemicals (REPs > 0.001) as measured in the bioassay of interest were used to calculate toxic equivalents for observed ecotoxic effects found in literature.

TABLE 4.1. BIOASSAYS AND CORRESPONDING EFFECT-BASED TRIGGER VALUES (EBTS) ESTABLISHED IN THE SIMONI FRAMEWORK (MODIFIED FROM VAN DER OOST ET AL., 2017A)[1].

Endpoint	Bioassay	Safe BEQ[2]	Low risk EBT BEQ[2]
estrogenic activity	ERα CALUX[3]	0.0066 ng EEQ/L	0.5 ng EEQ/L
anti-androgenic activity	anti-AR CALUX[3]	0.00005 µg FluEQ/L	25 µg FluEQ/L
glucocorticoid activity	GR CALUX[3]	20 ng DexEQ/L	100 ng DexEQ/L
dioxin-like activity	DR CALUX[3]	0.4 pg TEQ/L	50 pg TEQ/L
PPAR γ receptor activity	PPAR γ CALUX	0.00014 ng RosEQ/L	10 ng RosEQ/L
toxic PAHs activity	PAH CALUX	0.04 ng BaPEQ/L	150 ng BaPEQ/L
oxidative stress	Nrf2 CALUX[3]	0.000006 µg CurEQ/L	10 µg CurEQ/L
pregnane X receptor activity	PXR CALUX	0.000004 µg NicEQ/L;	3 µg NicEQ/L;

1. Activity of different classes of antibiotics were also included for ecological risk assessment. 2. Reference chemicals: ERα CALUX: 17-β estradiol; anti-AR CALUX: flutamide; GR CALUX: dexamethasone; DR CALUX: 2,3,7,8-TCDD; PPARγ CALUX: rosiglitazone; PAH CALUX: Benzo[a]pyrene; Nrf2 CALUX: curcumin; PRX CALUX: nicardipine. 3. In vitro bioassays also included in the DEMEAU evaluation of in vitro assays for drinking water safety assessment (Table 2.3).

The lowest BEQs of all toxic effects found in literature for the selected chemicals (where needed converted to realistic safe levels using extrapolation factors depending on the reported unit) is considered as the 'safe BEQ'. A BEQ value representing an exposure situation that will negatively impact at maximally 5% of the species in an ecosystem ('HC5 BEQ') was derived from species sensitivity distributions (SSDs) which included toxicity data for a specific group of chemicals converted to BEQ values. While for ecological purposes differences in species sensitivity need to be taken into account, this is not needed when EBTs are derived for human health assessment (as the BEQs are already developed for relevant species only). The derived EBTs were also benchmarked to bioassay responses measured for water from ecologically clean sites (Waternet), to ensure that a realistic EBT BEQ was derived. Hypothetically, these experimentally derived 'background BEQs'for these unpolluted water samples were in between the 'safe BEQ' and the 'HC5 BEQ'. If this was not the case, the EBT BEQ was further refined, mainly based on the background responses at unpolluted sites. The trigger values included in Van der Oost et al. (2017a) are derived specifically for CALUX assays (as relative potency of reference substances in a specific assay is included in the BEQ derivation). The REP values can however be adjusted in order to derive trigger values for other bioassays for the same activity. It is recommended also by Van der Oost et al. (2017ab) that validation/calibration studies are undertaken to ensure that the EBT BEQs are exceeded only at polluted sites, which may require further optimization. A variant of this approach based on toxicity data, possibly with the inclusion of biokinetic parameters, could also be developed in the near future to estimate effects on human health.

Effect-based trigger values based on reference chemical potency and kinetics (Brand et al., 2013)

In this approach the trigger value is based on health-based threshold values for chronic exposure in humans, corrected for bioavailability based on absorption from the gastrointestinal tract and protein binding in blood. CALUX bioassays detect the total specific endocrine activity of individual chemicals or complex mixtures. Endocrine activity is expressed as ng equivalents of a reference compound per L. The observed effect cannot directly be used to predict risks or effects in humans, as the internal exposure to chemicals after oral exposure depends on kinetics, including uptake and first pass metabolism (reducing absorption), binding to protein (reducing distribution), biotransformation reactions (metabolism) by e.g. the liver, and excretion. Brand and co-workers (2013) therefore extended on the previously published approach by Mennes (2004) and derived trigger values for hormone-receptor mediated activity in water by combining reference values for chronic exposure to humans with realistic worst-case kinetic factors (derived from literature or estimated using *in silico* tools) and exposure assumptions. Responses exceeding these trigger values indicate that human health risk cannot be waived *a priori* and additional examination of specific endocrine activity may be warranted.

The trigger values (Table 4.2) were derived as followed (Figure 4.2). The acceptable daily intake (ADI) of the reference chemical of the CALUX assay of interest was corrected for bioavailability (by multiplying with the estimated oral bioavailable fraction which passes the intestinal transport barriers and that escapes first pass metabolism by the intestine and liver) and free internal concentration (by multiplying with the fraction unbound to plasma proteins). The resulting internal ADI of the reference chemical is used to calculate the safe external (oral) equivalent exposure to water-relevant chemicals with the same endocrine activity (by dividing by the highest observed values for oral bioavailable fraction and fraction unbound to plasma proteins in the relevant set of chemicals). In cases where the used ADI was not similar to the ADI of the usual CALUX assay reference chemical this was corrected for using their respective relative potencies. The safe external equivalent exposure was used to calculate the safe trigger value in equivalents of the reference chemical (by multiplying with average body weight, dividing by average water consumption and taking the default allocation factor for drinking water into account). For an estimation of the risk for adverse health effects resulting from exposure to endocrine modulating chemicals via water, margins of exposure (in the article: Benchmark Quotient values) were calculated by dividing the measured hormonal activity in water samples by the derived trigger. When applying this approach on Dutch water samples, the observed levels of endocrine activity were 1-2 orders of magnitude below the respective trigger values (Brand et al., 2013).

TABLE 4.2. TRIGGER VALUES DERIVED FOR CALUX ASSAYS (BRAND ET AL., 2013).

CALUX	Trigger value
ERα	3.8 ng 17 β -estradiol (E2)-equivalents (eq)/L
AR	11 ng dihydrotestosterone (DHT)-eq/L
PR	21 ng dexamethasone (DEX)-eq/L
GR	333 ng Org2058-eq/L

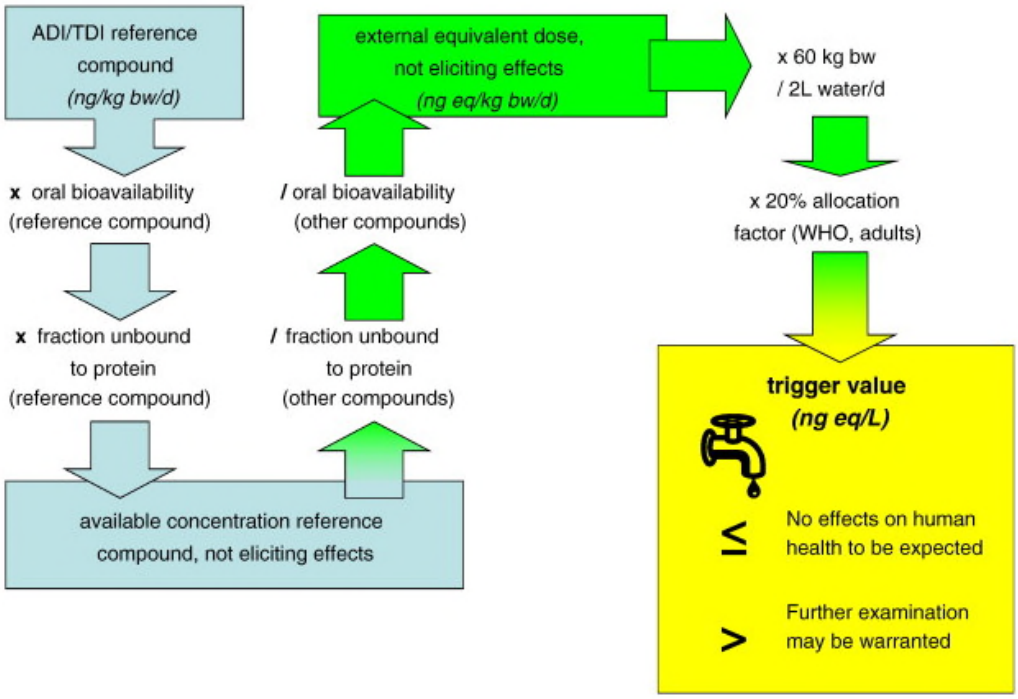


FIGURE 4.2. FLOW DIAGRAM OF DERIVING TRIGGER VALUES FOR ADDITIONAL INVESTIGATION FOR HORMONAL ACTIVITIES IN WATER FROM ACCEPTABLE OR TOLERABLE DAILY INTAKE (ADI OR TDI) VALUES OF REFERENCE COMPOUNDS (BRAND ET AL., 2013).

$$\text{Trigger value} = 1000 \times \frac{BW \times Af}{V} \times REP \times \frac{ADI_{E2} \times fa_{E2} \times fu_{PE2}}{fa_x \times fu_{P_x}} \text{ ng E2-eq/L}$$

EQUATION 4.1. PARAMETERS: *BW*: DEFAULT ADULT BODY WEIGHT (60 KG); *AF*: DEFAULT DRINKING WATER ALLOCATION FACTOR (0.2); *V*: DEFAULT ADULT DAILY DRINKING WATER CONSUMPTION (2 L/D); ADI_{E2}: ADI OF E2 (0.050 MG/KG BW/D); FA_{E2}: FRACTION ABSORBED OF E2 (0.05); FU_{PE2}: FRACTION UNBOUND TO PROTEIN OF E2 (0.02); FA_X: FRACTION ABSORBED OF UNKNOWN ESTROGENIC SUBSTANCE IN WATER SAMPLES (0.5); FU_{PX}: FRACTION UNBOUND TO PROTEIN FOR UNKNOWN ESTROGENIC SUBSTANCE IN WATER SAMPLES (0.16); *REP*: RELATIVE POTENCY OF ADI REFERENCE SUBSTANCE COMPARED TO THE CALUX BIOASSAY REFERENCE SUBSTANCE (FOR E2 THIS IS EQUAL TO 1).

The trigger values included in Brand et al. (2013) are derived specifically for CALUX assays (as relative potency of reference substances in a specific assay is included in the equation to calculate a trigger value; Equation 4.1). The values of the parameters in the equation can however be modified for deriving trigger values for other bioassays for the same agonistic endocrine activity. The trigger values mainly depend on the information used for the calculation. Therefore, it needs to be followed whether the used reference values for chronic exposure to humans remain valid or if new values are proposed based on new toxicological information. Specific trigger values may be derived for specific sensitive groups, such as young children (based on specific body weight and drinking water consumption patterns and exposure via other exposure routes). Therefore, due to the possibility that the trigger value

may be more or less valid for a particular exposure situation, a certain degree of uncertainty remains in the interpretation of *in vitro* bioassay data. However, it should be noted that these also exist in a framework using analytical chemistry. These include the impact of metabolites that may not be included in the analyses and loss of chemicals during sample preparation.

Effect-based trigger values based on TTC (Mennes, 2004)

The Threshold of Toxicological Concern (TTC) approach can be used for the risk assessment of chemicals for which insufficient toxicological information is available. For different groups of chemicals, exposure threshold levels have been established, based on distributions of reference values for chemicals with known chronic toxicity, below which the probability of a risk of adverse effects on human health is very low. Mons and co-workers (2013) derived a drinking water target value of 0.01 µg/L for steroid hormones from established TTC values. This threshold corresponds to the trigger value of 7 ng EEQ/L (Estradiol Equivalents) for the ER-CALUX assay (Mennes, 2004). It should be investigated further how drinking water TTC levels relate to other bioassay trigger values. The main uncertainties in this approach are caused by the fact that kinetic differences are not taken into account. When considering this approach for the derivation of trigger values for other *in vitro* bioassays, it should be noted that there are also groups of chemicals that are excluded from the TTC approach. The TTC concept can also be used to refine the Brand et al., (2003) trigger value approach by deriving a critical exposure threshold of bioanalytical concern by high-througput toxicokinetic modelling (e.g. Wambaugh et al., 2015) based on an extensive list of chemicals relevant for a particular mechanism (key event).

Effect-based trigger values read-across from existing guidelines (Escher et al., 2015)

Escher and co-workers (2015) propose an approach (Figure 4.3) in which effect-based trigger values (EBT) are based on existing water quality guidelines. Effect concentrations in bioassays are matched to existing chemical guideline values and the relevant reference chemicals. Bioanalytical (toxic) equivalents integrate the effects of groups of chemicals with the same mode of action. Statistical distribution methods are used to derive specific effect-based trigger bioanalytical equivalent concentration (EBT-BEQ) for bioassays for receptor-mediated toxicity. EBT-BEQ is the concentration of the reference chemical that would elicit the same effect as the water sample of unknown composition. Effect concentrations in bioassays and guidance values were collected for regulated chemicals. Relative effect potencies (REPs) were calculated based on effect concentrations within an order of magnitude of the guideline values (corrected for enrichment of water samples). For the calculation of REPs, reference chemicals were selected that were preferably also included in the same guideline. The REPs were used to convert guideline values to BEQs and these were included in cumulative distributions per bioassay. The 5th percentile in the distribution were selected as the EBT-BEQ for that assay. As the sensitivity of a bioassay has an impact on its EBT-BEQ, trigger values can differ between bioassays for the same mechanism of action.

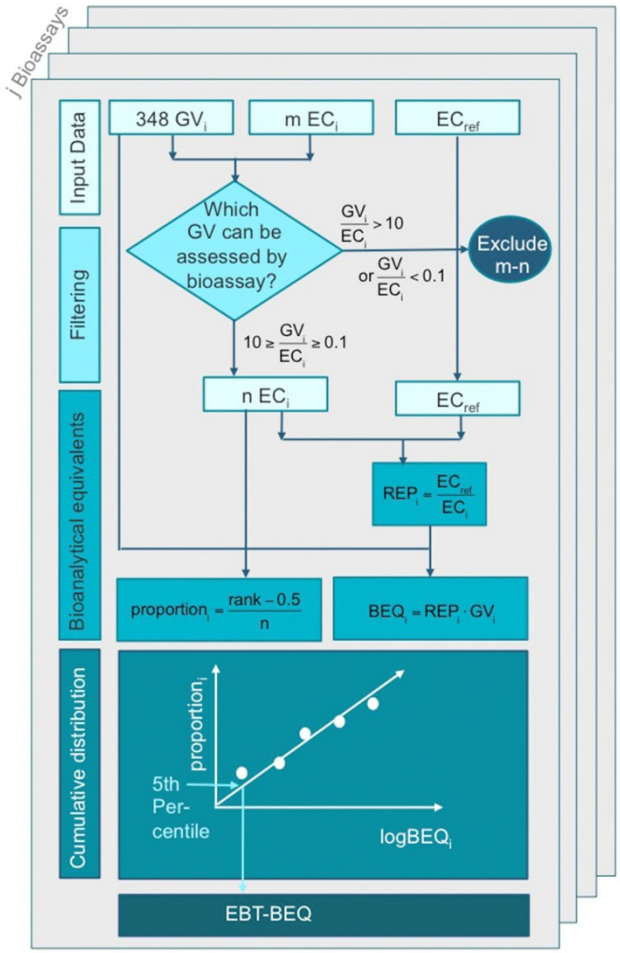


FIGURE 4.3. APPROACH TO DERIVE EBT-BEQS FOR CELL-BASED BIOASSAYS (GV = GUIDELINE VALUE, EC = EFFECT CONCENTRATION, BEQ = BIOANALYTICAL EQUIVALENT CONCENTRATION, REP = RELATIVE EFFECT POTENCY, N = NUMBER OF CHEMICALS I, M = NUMBER OF AVAILABLE EC VALUES PER BIOASSAY J, REF = REFERENCE COMPOUND) (ESCHER ET AL., 2015).

TABLE 4.3. BIOASSAYS, REFERENCE CHEMICALS AND EBT-BEQS (ESCHER ET AL., 2015)

Bioassay	Reference chemical	EBT-BEQ
PXR-cisFACTORIAL	metolachlor	59 μg/L
AhR-cisFACTORIAL	carbaryl	18 μg/L
Algae photosynthesis inhibition	diuron	0.6 μg/L
Acetylcholinesterase Inhibition	parathion	26 μg/L
AR-GeneBLAzer	testosterone	14 ng/L
ER-GeneBLAzer	17β-Estradiol	1.8 ng/L
ER-CALUX	17β-Estradiol	0.2 ng/L
E-SCREEN	17β-Estradiol	0.9 ng/L
YES	17β-Estradiol	12 ng/L
hERα-HeLa-9903	17β-Estradiol	0.6 ng/L
GR-CALUX	dexamethasone	150 ng/L

By directly relating regulatory guideline values to EBTs based on effective concentrations of the regulated chemicals (not necessarily with causal physiological relationships), *in vitro* bioassays are purely used as an analytical tool to capture combined effects of complex low-level mixtures in water. In the presented proof-of-concept study, an EBT-BEQ could not be derived for every bioassay due to a lack of available data on regulated chemicals, but 11 provisional EBT-BEQs (Table 4.3) could be derived based on guideline values from the Australian Guidelines for Water Recycling, and the applicability was demonstrated using a diverse set of water samples.

Different parameters are used for the calculation of trigger values in the different approaches and toxicokinetics are not taken into account in every approach. Therefore, it can be expected that the different approaches derive (slightly) different trigger values. When comparing effect-based trigger values for estrogenicity responses in the ERα CALUX (expressed as 17β-estradiol-equivalents) calculated using different methods (Table 4.4), these vary between 0.2-7 ng/L. Although this is a relatively small range, the effect curve of 17β-estradiol is steep (e.g. Leusch et al., 2017) resulting in potentially large differences in effect sizes between near-equal concentrations.

TABLE 4.4. COMPARISON OF DIFFERENT TRIGGER VALUES FOR THE ERA CALUX.

		Trigger value (17β-estradiol-equivalents)
guideline values	Escher et al., 2012	0.2 ng/L
relative ecotoxicity potency	van der Oost et al., 2017a	0.5 ng/L*
chronic exposure reference values and kinetics	Brand et al., 2013	3.8 ng/L
threshold for toxicological concern	Mennes, 2004	7 ng/L

* This is the 'low risk' ecotoxicological trigger value, not related to human health.

When establishing a protocol for the derivation of trigger values, it is proposed to apply the 'Brand et al., 2013' approach and proceed with trigger values based on reference values for chronic exposure to humans, based on the biological link between a molecular mechanism and the potential adverse outcome in humans, and taking kinetic factors into account). Further refinement of this approach is possible and as a next step it can be explored if these *in vitro* bioanalytical thresholds meet the requirements of the TTC approach and existing drinking water guidelines (see Chapter 7).

A selection of water samples was tested in the ToxTracker assay (Chapter 4). It was explored if the approaches described above could be used to determine the relevance of the observed effects for human health.

A pragmatic approach is to first establish whether the observed effects are above the level of quantification (Escher and Leusch, 2011), defined here as statistical deviation from the solvent control (using Students' t-test). This analysis demonstrated that only exposures to surface water or wwtp effluent resulted in activation of the Bscl2 marker gene for DNA damage in the ToxTracker, while drinking water samples did not activate markers for DNA damage. To establish whether these are relevant (acceptable or non-acceptable) effects, the measured responses need to be analysed using effect-based trigger values.

Effect-based trigger values are generally based on the relative potency of a suitable reference chemical. It was possible to convert the ToxTracker

responses of environmental water samples to reference equivalent concentrations as a range of positive controls for assay performance were included and one of the tested chemicals (BaP) could be used as reference chemical. However, the resulting equivalent concentrations are highly uncertain as the responses were outside of the range of those measured in the reference curves and therefore highly dependent on the concentration-response fitting. Therefore it was not further pursued to derive a effect-based trigger value based on health-based guidance values of the reference compound, corrected for uptake from the gastrointestinal tract and protein binding based on an inventory of other genotoxic compounds and their pharmacokinetic factors (Brand et al., 2013).

In the approach by Escher and coworkers (2015), relating regulatory guideline values to effective concentrations of regulated chemicals directly places *in vitro* bioassay effects in the context of practical water quality monitoring. It is possible to use this method to calculate effect-based trigger values for the ToxTracker assay. However, the number of effective concentrations of regulated chemicals (Table 4.5) is too low to reliably calculate effect-based trigger values (the 5th percentile of the distribution of available bioanalytical equivalent concentrations) for the induction of the different biomarker genes.

In conclusion, it is currently not feasible to derive reliable effect-based trigger values for the ToxTracker *in vitro* bioassay. Although the data collected in this research is promising, further evaluation is necessary before effect-based trigger values can be derived and calibrated, which is required before the ToxTracker assay may be implemented for water quality monitoring.

TABLE 4.5. EFFECTIVE CONCENTRATIONS (LOWEST-OBSERVED EFFECT CONCENTRATIONS) IN THE TOXTRACKER *IN VITRO* BIOASSAY AND GUIDELINE VALUES OF REGULATED CHEMICALS (DRINKWATERBESLUIT, 2011). THE LOEC OBSERVED IN THE PRESENCE OF S9 METABOLIC MIXTURE IS ONLY INCLUDED IF LOWER.

Chemical	**Effect concentration (LOEC; µM)[a]**		**Drinkwaterbesluit Guideline**	
	Bscl2	Srxn1	(µM)	(µg/L)
arsenic	4	0.5	0.13	10
benzo(a)pyrene	>125 (S9: 31)	7.8	0.000040	0.01
bromate	1000	500	0.01	1
bromoform	>1000 (S9: 1000)	>1000 (S9: 1000)	0.1	25
dichlorobromomethane	>500	500	0.09	15
NDMA	>1000	>1000	0.00016	0.012
tetrachloroethene	>1000	>1000	0.06	10

[a] It is not possible to interpolate EC_{50} values as full concentration-response curves (with a maximum response) are not generated.

Further data that needs to be collected includes the responses of reference and regulated chemicals, non-linear fits to full concentration response curves (including lowest-observed and maximum effect concentrations) to extrapolate activity of water samples in relation to the reference chemical activity, and the final enrichment in an *in vitro* bioassay which is considered relevant (based for example on safety factors for risk assessment). Moreover, it is critical in *in vitro* bioassay data to distinguish specific toxicity from non-specific cytotoxicity, in particular when a decrease in the endpoint value is the read-out. The ToxTracker cell lines were exposed to concentrated extracts of water samples, and with increasing final enrichment, cytotoxicity occurs. In this case this is less of a problem as the read-out of the assay is the activation (increased expression of) marker genes in viable cells. Finally, proposed trigger values should always be calibrated by comparison with the effects observed for adequate numbers of realistic water samples with known water quality.

5 Implementation in Dutch water sector

Experts in the field of *in vitro* bioassays at Dutch water companies and water laboratories have been approached to share information on the current application of *in vitro* bioassays in their day-to-day work, and share their views on remaining needs and expectations, also with regard to the use of the collected data (e.g. the collection of *in vitro* bioassay data in a combined database).

Waternet has been active in the derivation and publication of ecotoxicological trigger values (Van der Oost et al., 2017ab) and applies bioassays in several monitoring campaigns each year. A panel of *in vitro* bioassays is included in the Smart Integrated Monitoring (SIMONI) strategy for water quality assessment (Van der Oost, 2017a). The selection of *in vitro* bioassays in the SIMONI panel was based on Waternet research and earlier studies in which *in vitro* bioassays were evaluated for their use in water quality monitoring (Escher et al., 2014; Macova et al., 2011 ; van der Linden et al., 2008; Willemsen et al.,1995). This approach has also been tested to assess risk related to micropollutants in the field (van der Oost, 2017b).
At water laboratory HWL, a number of *in vitro* bioassays have already been implemented and are presently used for water quality monitoring for water companies. The effect-directed analysis (EDA) platform at HWL is implemented for research on the presence of the active components in complex mixtures of micropollutant in surface water and wwtp effluent. At Dunea, PWN, Waternet, effect-directed analysis is also considered as a key step for the identification of active micropollutants in cases where trigger values are exceeded in *in vitro* bioassays. HWL researchers have recently also proposed to use *in vitro* bioassays for specific toxic mechanisms (activation of the arylhydrocarbonreceptor and androgen receptor antagonism) to monitor the presence of all dioxines en dioxine-like compounds by their total activity instead of measuring concentration of a limited number of such chemicals (Houtman et al., in preparation). In this manner, implementation of *in vitro* bioassays to monitor the presence of large groups of chemical has the potential to reduce the amount of chemical analyses. With this aim, RIWA has already replaced the analysis of estrogenic compounds in surface water with the ER-CALUX bioassay (Van der Hoek et al., 2015), and other CALUX assays are considered to be implemented to monitor the water quality of the Rhine and Meuse. There are representatives from other water companies (e.g. de Watergroep, BE) that do not have any experience with *in vitro* bioassay in their work, although some of them express their specific interest in this topic (Oasen), some related in particular to mixture toxicology (Evides). Also other stakeholders besides the drinking water companies see possibilities for the application of *in vitro* bioassays, in particular for the evaluation of mixture effects (RIWA). Responses (in Dutch) have been collected in Attachment VI.

In conclusion, the value of *in vitro* bioassays is generally acknowledged in the Dutch water sector, but the actual use is currently limited to a few drinking water companies and laboratories, mainly because there is a knowledge gap with regard to the practice and interpretation of *in vitro* bioassay data. The implementation of *in vitro* assays can thus be further supported by the development of protocols and workflows for the evaluation, use and implementation of *in vitro* bioassays for water quality monitoring and guidance for the interpretation of data (i.e. effect-based trigger values).

6 Regulations for risk-based monitoring using bioassays

With regard to chemical water quality, regulations only include guidelines that specify permitted concentrations of individual chemicals or groups of structurally very similar chemicals. Regulations for chemical water quality assessment using human health based *in vitro* bioassays have not yet been established. In this chapter the current status of regulations for water quality assessment using *in vitro* bioassays is described. This was reported earlier in more detail, together with examples on the application of *in vitro* bioassays for food quality monitoring and chemical legislation, in the report '*In vitro* bioassays for prediction of human health hazard in international regulatory frameworks' by Sjerps et al. (2016). In this chapter, expected near-future developments have also been included.

The regulatory framework of drinking water quality concerns the quality of water intended for human consumption. In the regulations quality standards are set. Examples of regulations are the Drinking Water Directive in the European Union, The Safe Drinking Water Act (SDWA) in the United States of America, the Drinking Water Protection Act in Canada and the Australian Drinking Water Guidelines in Australia. Currently, in these drinking water frameworks there is no use of *in vitro* bioassays for water quality assessment for human health. Nevertheless, in the recent revision of the EU Drinking Water Directive (EC, 2017) amendments are included that allow risk-based monitoring approaches, as well as the use of *in vitro* bioassays, providing this ensures full protection of public health. Member states are expected to transpose these amendments in national legislation. It is expected that effect-based tools will also be included in the imminent revision of the EU Water Framework Directive (Brack et al., 2017) and Guidelines on Integrating Water Reuse into Water Planning and Management in the context of the WFD, which is currently an informal consensus position between EU member states and relevant stakeholders. A practical framework to apply bioanalytical tools for routine and recycled water quality monitoring has been proposed (Leusch and Snyder 2015).

Battery tests have been included in specific research in water quality assessment, e.g., efficiency of sewage treatment plants and the chemical quality of drinking water (Kienle et al., 2011; Macova et al., 2010; Pablos et al., 2009; Macova et al., 2011; Zěgura et al., 2009). For example, safe environmental concentrations of estrogenic equivalents were determined in wwtp effluent based on *in vitro* potencies of steroid estrogens (specific for particular in bioassay and testing protocols), *in vivo* predicted no effect concentration (PNECs), and relative contributions of estrogenic compounds to the overall estrogenicity detected in municipal WWTP effluents (Jarosova et al., 2014). Endocrine activities measured in US stream water demonstrated that concentrations of estrogens were well-predicted by the estrogenic agonistic activity. The detected androgenic and glucocorticoid activities did not correspond to target-chemical predictions indicating that other unknown chemicals in the water are responsible for the bioactivity (Conley et al., 2017a). These studies clearly demonstrate that the incorporation of *in vitro* bioassays in parallel to chemical analyses in water quality monitoring allows for a more complete assessment of chemical mixtures in the aquatic environment. *In vitro* bioassays can also play a role in evaluating water treatment efficacy, as was demonstrated in the study of Conley and coworkers (2017b) who used *in vitro* bioassays

for estrogenicity to demonstrate that estrogenic activity was effectively reduced in drinking water treatment processes. *In vivo* (aquatic) bioassays with intact organisms are being used as standard tools for characterising effluent quality and are well accepted by water quality regulators in many regions of the world, including the EU (Power and Boumphrey 2004). This is not yet the case for *in vitro* bioassays for human health effects, but their advantages (see Chapter 1) have been recognized by the Directorate-General (DG) Environment of the European Commission and effect-based tools and methods are currently being evaluated in the Common Implementation Strategy (CIS) Work Programme. A particular advantage of *in vitro* bioassays is that they can be used to elucidate the consequences of combined exposure. Nevertheless, *in vitro* bioassays for water quality monitoring are currently mainly used as a diagnostic research tools for hazard identification and testing of treatment efficiency of (novel) drinking water treatment methodologies.

There are currently no formal guideline values in water quality regulation for measurements resulting from *in vitro* bioassays. Regulatory acceptance is defined as the formal adoption of a (validated) test method by a regulatory agency/authority. Implementation of *in vitro* bioassays for human health in national water quality regulatory frameworks is supported in the EU Drinking Water Directive. To optimize progress in this matter collaboration between researchers and policy makers is needed to obtain 1) more knowledge on the correlation between results from chemical analyses and *in vitro* bioassays (see Chapter 2), 2) validation and harmonization of candidate *in vitro* bioassays and 3) consensus on the derivation and application of effect-based trigger values (guidance values) (Chapter 4) to interpret results that are obtained in *in vitro* bioassays.

Formal validation of *in vitro* bioassays to be accepted by regulators and policy makers is performed by several international institutes for the Validation of Alternative Methods. In Europe this is JRC-ECVAM, the EU Reference Laboratory for alternatives to animal testing. The task of this institute is to validate methods which reduce, refine or replace the use of animals for safety testing and efficacy/potency testing of chemicals, biologicals and vaccines. Currently, this work is much focused on the testing in the context of registration and authorization of individual chemicals to gain market access, however not focused on the use of these assays in the context of environmental quality assessment and management. The validation of a test, to demonstrate that it is reliable and reproducible, can occur on request from scientific institutes, industries or other stakeholders. If validation criteria are met, the OECD can adopt a standard protocol. It should be noted however that not all tests included in regulatory frameworks have been formally validated, although these mainly include *in vivo* tests and long existing *in vitro* bioassays that have proven their value in practice. It is also to be noted that not all validated tests are included in regulatory frameworks. This is often a consequence of insufficient collaboration with the regulatory authorities (Schiffelers et al., 2012) stressing that early involvement of regulatory authorities is critical. Another important step towards regulatory acceptance is the harmonization of *in vitro* bioassay test methods by establishing standards (ISO) and guidelines (OECD). A standard can assure that a method is safe, reliable, and of good quality. Harmonization is a complex process, because all member states of the responsible organization must reach consensus. More details on validation and harmonization procedures, lists of validated methods, and lists of standards and guidelines are included in Sjerps et al., (2016).

Stimulating factors for the inclusion of a particular *in vitro* bioassay in regulatory frameworks are 1) regulator needs (guidelines), 2) the derivation of effect-based trigger values to facilitate the interpretation of bioassay results and 3) a clear link between the results obtained in a bioassay and the related adverse outcome on human health. Factors that need to be overcome for an *in vitro* bioassay to be included in regulatory frameworks

are the formal validation procedures and the fact that established *in vivo* tests cannot be replaced by only a single *in vitro* bioassay. Communication with and education of regulators also supports the implementation process, as has been demonstrated in Australia where bioassays are now specifically mentioned in the Australian Guidelines for Water Recycling.

In vitro bioassays, mainly for genotoxicity, have already been implemented for the regulation of industrial chemicals, cosmetics, biocides and plant protection products (for overview, see Sjerps et al., 2016). In most cases these *in vitro* bioassays are included in tiered approaches in which *in vivo* studies can be performed to confirm results from *in vitro* bioassays. This is not the case for cosmetics ingredients and products, for which only *in vitro* bioassays are allowed. *In vitro* bioassays for cytotoxicity and genotoxicity are also included in the REACH legislation and CLP regulation (classification, labelling and packaging). Also in food and feed safety regulations, *in vitro* bioassays have been implemented mainly in tiered approaches and they are also used to assess effects of groups of chemicals (using the TEQ concept for dioxins and dioxin-like compounds). The TEQ concept is already occasionally applied in water quality legislation. It is thus recommended to stay informed on developments in food safety and chemical regulation that may be used as an example for the future acceptance and implementation of *in vitro* bioassays in the framework of water quality guidelines.

Already since before the revision of the EU Drinking Water Directive, Dutch water companies support the implementation of *in vitro* bioassays for human health assessment in water quality monitoring, and already apply a number of *in vitro* bioassays for water quality monitoring. Official inclusion of *in vitro* bioassays is however in the hands of Dutch policy makers. To support their decision-making process, a guidance document is being drafted by branch organization VEWIN together with scientists from the Dutch water companies and from KWR Watercycle Research Institute. In this guidance document, a national strategy is proposed for risk-based monitoring of chemical water quality. The chemicals for which risk-based monitoring will be pursued are included in Table 6.1. The strategy includes the identification and prioritization of chemicals (using the PRIO model described in Velzeboer et al., 2015), which can be included in a monitoring program. In particular in the prioritization step, *in vitro* bioassays can be applied to determine relative toxicity of candidate chemicals. It is also proposed to periodically evaluate the selection of chemicals that are included in the monitoring program. An initiative in which chemicals that are rarely encountered in target chemical analyses are monitored using *in vitro* bioassays (in which they induce a combined response) is currently being evaluated.

TABLE 6.1. ANTHROPOGENIC CHEMICALS AND CHEMICALS GROUPS THAT ARE PROPOSED TO BE INCLUDED IN RISK-BASED MONITORING (TABLE IIIC FROM DRINKWATERBESLUIT 2011.

adsorbed organic halogens (AOX)
aromatic amines
(chloro)phenols
diglyme(s)
ethyl tert-butyl ether (ETBE)
halogenated monocyclic hydrocarbons
halogenated aliphatic hydrocarbons
methyl tert-butyl ether (MTBE)
aromatic monocyclic hydrocarbons
other anthropogenic chemicals

The ToxTracker assay has not yet been validated for use in regulatory frameworks, and is thus not yet included in the recommended battery of genotoxic assays in the OECD guidelines. Steps taken by ToxTracker developer Toxys© for regulatory acceptance include an ECVAM application for validation, the submission of a draft standard project submission form for OECD (which will be officially submitted by Dutch representatives at a later stage), and extensive validation of the assay using ECVAM and ToxCast compound libraries. Toxys© scientists are involved in a specialized workgroup on mechanistic *in vitro* bioassays of ILSI Hesi. A large interlaboratory study is currently being organized (G. Hendriks, personal communication).

7 Conclusions and the road ahead

In vitro bioassays are valuable tools to include in water quality monitoring in parallel to chemical analysis (Chapter 1). The current selection of recommended *in vitro* models is based on an evaluation of 1) the most relevant potential health effects of exposure to chemicals via water and 2) applicability and use of available *in vitro* bioassays (Chapter 2). As there is a wide range of possible endpoints, models and technologies, it is critical to appropriately evaluate candidate bioassays for applicability and sensitivity (Chapter 3), and to establish trigger values to determine the relevance of potentially observed effects of water samples of unknown composition (Chapter 4). *In vitro* bioassays are being used by several of the Dutch drinking water companies (Chapter 5) and there are a number of developments that support a legislative framework (Chapter 6). This chapter includes a discussion of possible future developments as well as recommendations for efforts needed for the successful implementation of *in vitro* bioassays for water quality monitoring.

Which assay?
The choice of *in vitro* bioassays used for water quality monitoring (Chapter 2) is an ongoing process. There can be several reasons to consider the implementation of new *in vitro* bioassays, including emerging compounds (with specific effects), or the availability of new bioassays (models and methods). A wide range of mode of action categories has been observed for water-relevant chemicals (Busch et al., 2016). Although a broad selection of toxic endpoints will gain a more complete picture of possible effects of unexpected or unknown chemicals (Escher and Leusch, 2011), time and cost-effectiveness should also be included in these considerations. Therefore, it is recommended that the exploration of candidate bioassays is a continuous (or repeated) effort.

The scoring matrix for bioassays developed in the DEMEAU project can be used to compare candidate bioassays for implementation in water quality monitoring. Nevertheless, it should be noted that the availability of information varies widely which results in a degree of uncertainty on the individual scores. This urges the need for recurring evaluations, e.g. on the availability of validation studies of certain bioassays. New information may boost the score for suitability for environmental monitoring of water quality. This work is focussed on *in vitro* bioassays to assess potential threats to human health. Bioassays to assess ecosystem health (Kienle et al., 2011; Kokkali and van Delft 2014) were thus not included but their availability, applicability and performance may be evaluated in a similar manner as used for *in vitro* bioassays in DEMEAU.

In a recent study by Leusch et al. (2017), the sensitivity of *in vitro* bioassays for different types of endocrine activity was evaluated. For androgenic, estrogenic, progestagenic and glucocorticoid activity, available assays have sufficient sensitivity, while this is less the case for assays to detect thyroid activity. Moreover, it is concluded that there is a current lack of standardization in the methods used to test antagonistic activity.

Possible effects on reproduction and development are currently not included in water quality monitoring strategies as there are few *in vitro* bioassays to study the relevant mechanisms. The amount of convincing evidence for causal relationships between exposure to specific chemicals and adverse effects on human development obtained in experimental and/or epidemiological studies varies widely. However, such health effects can occur as the result of chronic exposure to low levels of chemicals in the environment, including in water (Baken

2013; Gezondheidsraad, 2014) and could have a considerable impact on the quality of life due to their disease burden.

Developmental toxicity refers to adverse effects on the developing organism prior to conception, during prenatal development, and/or during the postnatal phase until maturity. Early development is a highly vulnerable phase with regard to the effects of chemical exposures. Such adverse effects during early development can be caused through a multitude of mechanisms at lower exposures in comparison to those in adults, in particular during specific critical windows of exposure. It should also be noted that there are also specific organ systems that are particularly vulnerable to the effects of chemicals during development. These include the endocrine system, the nervous system and the immune system (Gezondheidsraad, 2014). Specific developmental processes may be targeted by chemicals, and apparently organ-specific mechanisms may be related. Such an example is the impact of thyroid hormone deficiency (hypothyroidism) on brain development. Due to the complexity of mammalian reproduction and development, many studies used intact organisms to study developmental toxicity. It has been demonstrated however that interspecies differences are considerable with regard to developmental processes (Carney et al., 2011), which argues for the use of (human) *in vitro* bioassays to study the potential of chemicals for developmental toxicity. Several available *in vitro* assays for developmental toxicity are designed to assess highly specific processes during development, such as sperm cell function. Mechanistic *in vitro* bioassays are likely more appropriate and feasible for routine testing of environmental samples, although a number of approaches with intact organisms are being explored (e.g. tests in zebrafish embryos or *c. elegans*). Pathways of toxicity are known to share similarities between species (Krewski et al., 2010) and it has been demonstrated that effects of chemicals could be predicted to a considerable extent using a set of *in vitro* bioassay (Piersma et al., 2013; van der Burg et al., 2015). In the DEMEAU research was encountered that several bioassays for effects on reproduction and development where in the development phase when the scoring matrix was applied. These bioassays may become available in the near future.

The ToxTracker pilot study

In this project, the ToxTracker bioassay for genotoxicity mechanisms had a high score in the DEMEAU approach to evaluate available bioassays and was therefore selected for further empirical evaluation (Chapter 3). It was demonstrated that the assay was sufficiently sensitive to detect genotoxic activity in different types of water. The collected data did however not allow a detailed comparison with other *in vitro* bioassays for genotoxicity. It is therefore recommended that a standard protocol is developed for the theoretical and empirical evaluation of new candidate *in vitro* bioassays. Including a list of reference chemicals (e.g. based on the list included in Bush et al., 2017) can aid in the comparison between *in vitro* bioassays. This evaluation can also include a cost-effectiveness analysis and the possibilities for high-throughput.

In vitro to in vivo extrapolation

A more mechanistic manner of interpreting *in vitro* bioassay data can be based on a better understanding (also quantitatively) of the physiological link between cellular and molecular

effects *in vitro* and potentially associated downstream adverse outcomes in the intact organism. This knowledge may be used to determine when an effect in an *in vitro* bioassay can be considered acceptable. This is in particular relevant for adaptive stress responses that are the cellular reactions that occur after exposure to various stressors. These responses are detected at lower concentrations in comparison to adverse effects, and can be reversible. An example of such a stressor is oxidative stress. A wide range of chemicals can cause cellular oxidative stress by causing an imbalance between the amount of reactive oxygen species and the protective antioxidant reaction. To prevent reactive damage on proteins, lipids and DNA, oxidative stress response pathways (regulated by nuclear factor proteins Nrf2 and keap1) are activated resulting in the production of proteins with antioxidant and detoxifying capacity. The oxidative stress response has been demonstrated to be a sensitive indicator of environmental exposure to a wide range of chemicals, as well as transformation products and disinfection-by-products (Escher et al., 2013). The quantitative relationship between adaptive stress responses and adverse effects on human health (either reversible or irreversible) is not yet clear. In particular for these types of effect, it is interesting in which effect range the intact organism (humans) may adapt with protective mechanisms, and at what effect size (unacceptable) irreversible adverse effects can occur. It is therefore recommended that new innovations and developments that support the feasibility of quantitative *in vitro* to *in vivo* extrapolations (including for example the inclusion of interactive tools in US-EPA's Integrated Chemical Environment; Bell et al., 2017) are continuously explored. Additionally, ADME (absorption, distribution, metabolism, elimination) processes are critical factors in the potential for health effects of chemicals, and the inclusion of an assessment of the impact of ADME is well-established in the pharmaceutical industry and can be used to improve the predictivity of *in vitro* bioassays (reviewed in Tsaioun et al., 2016).

Trigger values

For the interpretation of the effects of unknown (environmental) mixtures in *in vitro* bioassays it is critical to establish widely supported effect-based trigger values. As there is not yet any statutory basis, different approaches are currently being followed to interpret the relevance of bioassay effect data (Chapter 4). Some approaches develop trigger values that can be used in all *in vitro* bioassays studying a similar endpoint, while in other case assay-specific trigger values are developed.

Toxic equivalency approaches are based on the theory that chemicals act via a single well-defined mechanism such as receptor activation. In reality, this is however almost never the case. Even when studying the effects of a chemical mixture on a single receptor, agonistic and antagonistic mechanisms occur simultaneously (Leusch et al., 2017). New approaches may therefore be needed for more complex integrated *in vitro* bioassays. It needs to be further established and agreed upon which approaches (calculation and calibration of the trigger values) should be used for either interim of definite trigger values.

The trigger values based on reference values for chronic exposure to humans are based on the biological link between a molecular mechanism and the potential adverse outcome in humans (taking also ADME factors into account), and can be easily modified to apply to other *in vitro* bio-assays with the relevant read-out (Brand et al., 2013). It is therefore recommended to proceed with the derivation of new trigger values based on this approach. Possibilities for refinement are optimization of the combination of molecular mechanism and potential health effect (based on Adverse Outcome Pathways). If necessary, health reference values for specific effects can be derived from toxicological information. Moreover, this approach is developed for receptor-based assays and it needs to be re-evaluated and potentially adapted for its applicability to genotoxicity. Another possibility for refinement is the use of high-througput toxicokinetic modelling (e.g. Wambaugh et al., 2015) to derive a

critical threshold of bioanalytical concern to based on an extensive list of guideline exposure values for chemicals relevant for a particular mechanism (key event).

Uncertainties will remain

Chemical risk assessment is the estimation of risk related to exposure to chemicals, for which different types of data can be used. In the risk assessment process, some assumptions will have to be made that correspond to a certain degree of uncertainty. For example, the guideline exposure values for exposure to humans are based on experimental data in other organisms, and effects in cell-based systems (*in vitro* bioassays) are extrapolated to risk in intact organisms (*in vitro* to *in vivo* extrapolation). Limit values of chemicals in water (which are monitored with analytical chemistry) are designed to prevent adverse health effects, but this generally does not take into account potential effects of combined exposure. Although the use of *in vitro* bioassays is therefore a valuable addition to the available tools for monitoring water quality (as these measure the combined effect of the chemicals in a complex environmental mixture), a number of uncertainties remain when studying effects of unknown in *in vitro* bioassays. In most cases is unknown which component(s) in a complex mixture of micropollutants is responsible for the observed response in an *in vitro* bioassay. Depending on the *in vitro* bioassay model and the mechanisms of the constituents of the mixture, the complete toxic impact of a mixture can also be underestimated (if different mechanisms are summarized in an organ response) or overestimated (if adaptive mechanisms are in place). This argues for the uses of a wide range of *in vitro* bioassays (with regard to mechanisms) and the pursuit of the connection between analytical chemistry and *in vitro* bioassays for effect-directed analysis in which the complexity of tested mixtures is reduced.

High-throughput / high-content and sample preparation: further development for implementation

In vitro cellular models in bioassays are increasingly being used for human health hazard and risk characterization of exposure to chemicals via food and the environment. Besides ethical reasons, the use of *in vitro* cellular models in bioassays is more time and cost efficient and gives more specific mechanistic information in comparison to traditional regulatory studies with animal models. This has given rise to a rapid development of new *in vitro* bioassays in academic research and biotechnology business. Not all *in vitro* bioassays are compatible with high-throughput methods (which was also a parameter included in the DEMEAU scoring matrix). However, automation of an *in vitro* bioassay to achieve a high-throughput mode for cost effective and rapid measurements may have been outside the scope or expertise of the researchers that are or have been developing a bioassay. It is therefore recommended to explore possibilities to optimize the efficiency of *in vitro* bioassays by innovative methods such as miniaturization and automation. These methods may also allow a more efficient connection of analytical chemistry and bioassay, for example for effect-directed analysis (e.g. Jonker et al., 2015, 2017).

To have optimal insight in chemical water quality, it is recommended to combine analytical chemistry and *in vitro* bioassays. For an efficient workflow and the integration of exposure and effect data, it is recommended that chemical analyses and bioassays are performed using the same concentrates obtained using solid-phase extraction (SPE). It has been demonstrated that this is feasible (Attachment I). Nevertheless, the developed method should be further optimized (for example with regard to column preparation and the applicability for non-target screening) to be efficiently implemented in day-to-day practice. Furthermore, it is known that different types of chemicals respond differently to SPE extraction (some are concentrated to allow analysis using chemical or bioassay methods, while others may be lost in the process). It is therefore recommended to investigate the impact of different SPE approaches on the resulting availability in the concentrate for different (types of) chemicals.

Implementation, legislation and regulation

In vitro bioassays are included in water quality monitoring by Dutch water companies by varying degree (Chapter 5). This may be related to the fact that only recently there have been developments in the legislative frameworks for water quality that give room for the use of *in vitro* bioassays (Chapter 6). A clear overview of effect-monitoring data may support statutory basis for the legal implementation of *in vitro* bioassays, but such an overview of effect-monitoring data of Dutch surface waters, industrial effluents, and abatement and drinking water production processes is currently lacking. It is therefore recommended that this data, available at KWR, laboratoria of drinking water companies and water boards, academia and other research institutes is collected in a national database. Further statutory basis can also be supported by an increased confidence in bioassay data for human health and water quality relevance. To this aim, it is critical to adequately and reproducibly demonstrate the applicability of the selected set of *in vitro* bioassays and to develop guidelines including trigger values that aid in the indisputable interpretation of bioassay data. Newly introduced bioassays (models or methods) should also be evaluated in a coordinated manner. Collecting the results of such theoretical and empirical evaluations of *in vitro* bioassays in a shared database allows comparisons between (implemented and new) methods and ensures that evaluations are not unnecessarily repeated.

In analogy with the ToxTracker, Dutch biotech company Toxys B.V. is currently developing an *in vitro* reporter bioassay to detect effects on early embryonic development. This is a human stem cell line based assay which visualizes the various stages of early embryonic development by using fluorescent reporters for specific and different phases of differentiation of stem cells. This bioassay has the potential to rapidly identify the potential of novel compounds to adversely affect developmental processes. Whether this is a candidate bioassay for water quality monitoring depends on the pending results from interlaboratory validation studies with reference compounds.

8 References

Allen TE, Goodman JM, Gutsell S, Russell PJ. Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. Chem Res Toxicol. 2014 Dec 15;27(12):2100-12.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem. 2010 Mar;29(3):730-41.

Baken K. 2013. BTO 2013.226(s). Screening of water samples for developmental toxicity. An overview of relevant compounds and available assays.

Bell SM, Phillips J, Sedykh A, Tandon A, Sprankle C, Morefield SQ, Shapiro A, Allen D, Shah R, Maull EA, Casey WM, Kleinstreuer NC. An Integrated Chemical Environment to Support 21st-Century Toxicology. Environ Health Perspect. 2017 May 24;125(5):054501.

Brack W, Dulio V, Ågerstrand M, Allan I, Altenburger R, Brinkmann M, et al., Towards the review of the European Union Water Framework Directive: Recommendations for more efficient assessment and management of chemical contamination in European surface water resources. Sci Total Environ. 2017 Jan 15;576:720-737.

Brand W, de Jongh CM, van der Linden SC, Mennes W, Puijker LM, van Leeuwen CJ, van Wezel AP, Schriks M, Heringa MB. Trigger values for investigation of hormonal activity in drinking water and its sources using CALUX bioassays. Environ Int. 2013 May;55:109-18.

Busch W, Schmidt S, Kühne R, Schulze T, Krauss M, Altenburger R. Micropollutants in European rivers: A mode of action survey to support the development of effect-based tools for water monitoring. Environ Toxicol Chem. 2016 Aug;35(8):1887-99.

Carney EW, Ellis AL, Tyl RW, Foster PM, Scialli AR, Thompson K, Kim J. Critical evaluation of current developmental toxicity testing strategies: a case of babies and their bathwater. Birth Defects Res B Dev Reprod Toxicol. 2011 Oct;92(5):395-403.

Chapman H.F., Leusch F.D.L., Prochazka E., Cumming J., Ross V., Humpage A., Froscio S., Laingam S., Khan S.J., Trinh T., McDonald J.A. (2011). A national approach to health risk assessment, risk communication and management of chemical hazards from recycled water. Waterlines report 48.

Conley JM, Evans N, Cardon MC, Rosenblum L, Iwanowicz LR, Hartig PC, Schenck KM, Bradley PM, Wilson VS. Occurrence and *In vitro* Bioactivity of Estrogen, Androgen, and Glucocorticoid Compounds in a Nationwide Screen of United States Stream Waters. Environ Sci Technol. 2017a May 2;51(9):4781-4791.

Conley JM, Evans N, Mash H, Rosenblum L, Schenck K, Glassmeyer S, Furlong ET, Kolpin DW, Wilson VS. Comparison of *in vitro* estrogenic activity and estrogen concentrations in source

and treated waters from 25 U.S. drinking water treatment plants. Sci Total Environ. 2017b Feb 1;579:1610-1617.

Di Paolo C, Ottermanns R, Keiter S, Ait-Aissa S, Bluhm K, Brack W, et al., Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts - Towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring. Water Res. 2016 Nov 1;104:473-484.

Drinkwaterbesluit. Besluit van 23 mei 2011, houdende bepalingen inzake de productie en distributie van drinkwater en de organisatie van de openbare drinkwatervoorziening (Drinkwaterbesluit) . Staatsblad van het Koninkrijk der Nederlanden 2011 [293]. 2011. 's-Gravenhage.

Escher BI, Allinson M, Altenburger R, Bain PA, Balaguer P, Busch W, et al. Benchmarking organic micropollutants in wastewater, recycled water and drinking water with *in vitro* bioassays. Environ Sci Technol. 2014;48(3):1940-56.

Escher, B.; Leusch, F. Bioanalytical Tools in Water Quality Assessment; IWA Publishing: London, UK, 2011.

Escher BI, Neale PA, Leusch FD. Effect-based trigger values for *in vitro* bioassays: Reading across from existing water quality guideline values. Water Res. 2015 Sep 15;81:137-48.

Escher BI, van Daele C, Dutt M, Tang JY, Altenburger R. Most oxidative stress response in water samples comes from unknown chemicals: the need for effect-based water quality trigger values. Environ Sci Technol. 2013 Jul 2;47(13):7002-11.

European Commision. Review of the Directive. 2017. URL:http://ec.europa.eu/environment/water/water-drink/review_en.html [accessed 11-07-2017].

Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerg Themes Epidemiol. 2015 Sep 30;12:14.

Fischer FC, Henneberger L, König M, Bittermann K, Linden L, Goss KU, Escher BI. Modeling Exposure in the Tox21 *in vitro* Bioassays. Chem Res Toxicol. 2017 May 15;30(5):1197-1208.

Gezondheidsraad. Risico's van prenatale blootstelling aan stoffen. Den Haag: Gezondheidsraad, 2014; publicatienr. 2014/05. ISBN 978-90-5549-995-3.

Groothuis FA, Heringa MB, Nicol B, Hermens JL, Blaauboer BJ, Kramer NI. Dose metric considerations in *in vitro* assays to improve quantitative *in vitro-in vivo* dose extrapolations. Toxicology. 2015 Jun 5;332:30-40.

Hendriks G, Atallah M, Morolli B, Calléja F, Ras-Verloop N, Huijskens I, Raamsman M, van de Water B, Vrieling H. The ToxTracker assay: novel GFP reporter systems that provide mechanistic insight into the genotoxic properties of chemicals. Toxicol Sci. 2012 Jan;125(1):285-98.

Hendriks G, Derr RS, Misovic B, Morolli B, Calléja FM, Vrieling H. The Extended ToxTracker Assay Discriminates Between Induction of DNA Damage, Oxidative Stress, and Protein Misfolding. Toxicol Sci. 2016 Mar;150(1):190-203.

Heringa MB, Harmsen DJ, Beerendonk EF, Reus AA, Krul CA, Metz DH, Ijpelaar GF. Formation and removal of genotoxic activity during UV/H(2)O(2)-GAC treatment of drinking water. Water Res. 2011 Jan;45(1):366-74.

Houtman C, Behnisch P, Brouwer B, Felzel E, Kroesbergen J. CALUX-bioassays als alternatief voor de chemische monitoring van PCBs en PAKs voor het Drinkwaterbesluit? *In preparation*.

Jarošová B, Bláha L, Giesy JP, Hilscherová K. What level of estrogenic activity determined by *in vitro* assays in municipal waste waters can be considered as safe? Environ Int. 2014 Mar;64:98-109.

Jonker W, Lamoree MH, Houtman CJ, Hamers T, Somsen GW, Kool J. Rapid activity-directed screening of estrogens by parallel coupling of liquid chromatography with a functional gene reporter assay and mass spectrometry. J Chromatogr A. 2015 Aug 7;1406:165-74.

Jonker W, Stöckl JB, de Koning S, Schaap J, Somsen GW, Kool J. Continuous fraction collection of gas chromatographic separations with parallel mass spectrometric detection applied to cell-based bioactivity analysis. Talanta. 2017 Jun 1;168:162-167.

Kienle, C., Kase, R., Werner, I. 2011. Evaluation of bioassays and wastewater quality: *In vitro* and *in vivo* bioassays for the performance review in the Project "Strategy MicroPoll". Swiss Centre for Applied Ecotoxicology, Eawag-EPFL, Duebendorf.

Kirkland D, Aardema M, Henderson L, Müller L. Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. Mutat Res. 2005 Jul 4;584(1-2):1-256.

Kirkland D, Zeiger E, Madia F, Corvi R. Can *in vitro* mammalian cell genotoxicity test results be used to complement positive results in the Ames test and help predict carcinogenic or *in vivo* genotoxic activity? II. Construction and analysis of a consolidated database. Mutat Res Genet Toxicol Environ Mutagen. 2014 Dec;775-776:69-80.

Kokkali, V., van Delft, W. Overview of commercially available bioassays for assessing chemical toxicity in aqueous samples. Trends in Analytical Chemistry 2014, 61, 133-155.

Krewski D, Acosta D Jr, Andersen M, Anderson H, Bailar JC 3rd, Boekelheide K, Brent R, Charnley G, Cheung VG, Green S Jr, Kelsey KT, Kerkvliet NI, Li AA, McCray L, Meyer O, Patterson RD, Pennie W, Scala RA, Solomon GM, Stephens M, Yager J, Zeise L. Toxicity testing in the 21st century: a vision and a strategy. J Toxicol Environ Health B Crit Rev. 2010 Feb;13(2-4):51-138.

Kroese ED, Muller JJA, Mohn GR, Dortant PM, Wester PW. Tumorigenic effects in Wistar rats orally administered benzo[a] pyrene for two years (gavage studies). Implications for human cancer risks associated with oral exposure to polycyclic aromatic hydrocarbons. RIVM Rapport 658603010.

Kunz PY, Simon E, Creusot N, Jayasinghe BS, Kienle C, Maletz S, Schifferli A, Schönlau C, Aït-Aïssa S, Denslow ND, Hollert H, Werner I, Vermeirssen EL. Effect-based tools for monitoring

estrogenic mixtures: Evaluation of five *in vitro* bioassays. Water Res. 2017 Mar 1;110:378-388.

Leusch FD, Neale PA, Hebert A, Scheurer M, Schriks MC. Analysis of the sensitivity of *in vitro* bioassays for androgenic, progestagenic, glucocorticoid, thyroid and estrogenic activity: Suitability for drinking and environmental waters. Environ Int. 2017 Feb;99:120-130.

Leusch FDL and Snyder SA. Bioanalytical tools: half a century of application for potable reuse. Environ. Sci.: Water Res. Technol., 2015,1, 606-621.

Louisse J, Beekmann K, Rietjens IM. Use of Physiologically Based Kinetic Modeling-Based Reverse Dosimetry to Predict *in vivo* Toxicity from *in vitro* Data. Chem Res Toxicol. 2017 Jan 17;30(1):114-125.

Macova M, Escher BI, Reungoat J, Carswell S, Chue KL, Keller J, Mueller JF. Monitoring the biological activity of micropollutants during advanced wastewater treatment with ozonation and activated carbon filtration. Water Res. 2010 Jan;44(2):477-92.

Macova M, Toze S, Hodgers L, Mueller JF, Bartkow M, Escher BI, 2011. Bioanalytical tools for the evaluation of organic micropollutants during sewage treatment, water recycling and drinking water generation. Water Res. 45: 4238-4247.

Mennes W. Assessment of human health risks for oestrogenic activity detected in water samples, using the ER-CALUX assay. Memo. National institute for public health and the environment, Bilthoven, The Netherlands, 2004.

National Research Council. 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press. https://doi.org/10.17226/11970.

Pablos MV, Fernández C, del Mar Babín M, María Navas J, Carbonell G, Martini F, García-Hortigüela P, Vicente Tarazona J. Use of a novel battery of bioassays for the biological characterisation of hazardous wastes. Ecotoxicol Environ Saf. 2009 Jul;72(5):1594-600.

Piersma AH, Bosgra S, van Duursen MB, Hermsen SA, Jonker LR, Kroese ED, van der Linden SC, Man H, Roelofs MJ, Schulpen SH, Schwarz M, Uibel F, van Vugt-Lussenburg BM, Westerhout J, Wolterbeek AP, van der Burg B. Evaluation of an alternative *in vitro* test battery for detecting reproductive toxicants. Reprod Toxicol. 2013 Jul;38:53-64.

Power EA, Boumphrey RS. International trends in bioassay use for effluent management. Ecotoxicology. 2004 Jul;13(5):377-98.

Punt A, Brand W, Murk AJ, van Wezel AP, Schriks M, Heringa MB. Effect of combining *in vitro* estrogenicity data with kinetic characteristics of estrogenic compounds on the *in vivo* predictive value. Toxicol *In vitro*. 2013 Feb;27(1):44-51.

Schiffelers MJ, Blaauboer BJ, Hendriksen CF, Bakker WE. Regulatory acceptance and use of 3R models: a multilevel perspective. ALTEX. 2012;29(3):287-300.

Schriks M., Baken K., Simon E., Besselink H., van der Linden S., Kienle C., van der Burg B. 2015. Selection criteria to select *in vitro* bioassays for implementation and use. DEMEAU (FP7) report. URL: http://demeau-fp7.eu/content/d411-selection-criteria-select-vitro-bioassays-implementation-and-use [accessed 11-07-2017].

Sjerps R., Schriks M., Baken K. 2016. *In vitro* bioassays for prediction of human health hazard in international regulatory frameworks. DEMEAU (FP7) report. URL: http://demeau-fp7.eu/news/257 [accessed 22-09-2017].

Tsaioun K, Blaauboer BJ, Hartung T. Evidence-based absorption, distribution, metabolism, excretion (ADME) and its interplay with alternative toxicity methods. ALTEX. 2016;33(4):343-358.

Umbuzeiro GA, Heringa M, Zeiger E. *In vitro* Genotoxicity Testing: Significance and Use in Environmental Monitoring. Adv Biochem Eng Biotechnol. 2017;157:59-80.

Van der Burg B, Pieterse B, Buist H, Lewin G, van der Linden SC, Man HY, Rorije E, Piersma AH, Mangelsdorf I, Wolterbeek AP, Kroese ED, van Vugt-Lussenburg B. A high throughput screening system for predicting chemically-induced reproductive organ deformities. Reprod Toxicol. 2015 Aug 1;55:95-103.

Van der Hoek C, Bannink A, Slootweg T. An update of the lists with compounds that are relevant for the drinking water production from the river Meuse. RIWA 2015.

Van der Linden S, Heringa M, Man H-Y, Sonneveld E, Puijker LM, Brouwer A and Van der Burg B, 2008. Detection of Multiple Hormonal Activities in Wastewater Effluents and Surface Water, Using a Panel of Steroid Receptor CALUX Bioassays. Environ. Sci. Technol. 42: 5814–5820.

Van der Oost R, Sileno G, Suárez-Muñoz M, Nguyen MT, Besselink H, Brouwer B. SIMONI (smart integrated monitoring) as a novel bioanalytical strategy for water quality assessment: Part i-model design and effect-based trigger values. Environ Toxicol Chem. 2017a Sep;36(9):2385-2399.

Van der Oost R, Sileno G, Janse T, Nguyen MT, Besselink H, Brouwer A. SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: Part II-field feasibility survey. Environ Toxicol Chem. 2017b Sep;36(9):2400-2416.

Velzeboer I, Houtman CJ, Slootweg T, Kroesbergen J. 2015. Organische stoffen in het Drinkwaterbesluit: 1. Hoe geven we invulling aan de eisen voor 'overige' antropogene stoffen? H_2O 12 oktober 2015.

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, Landesmann B, Lettieri T, Munn S, Nepelska M, Ottinger MA, Vergauwen L, Whelan M. Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol Sci. 2014 Dec;142(2):312-20.

Vinken M. The adverse outcome pathway concept: a pragmatic tool in toxicology. Toxicology. 2013 Oct 4;312:158-65.

Wambaugh JF, Wetmore BA, Pearce R, Strope C, Goldsmith R, Sluka JP, Sedykh A, Tropsha A, Bosgra S, Shah I, Judson R, Thomas RS, Setzer RW. Toxicokinetic Triage for Environmental Chemicals. Toxicol Sci. 2015 Sep;147(1):55-67.

Wetmore BA. Quantitative *in vitro*-to-*in vivo* extrapolation in a high-throughput environment. Toxicology. 2015 Jun 5;332:94-101.

Wetmore BA, Wambaugh JF, Allen B, Ferguson SS, Sochaski MA, Setzer RW, Houck KA, Strope CL, Cantwell K, Judson RS, LeCluyse E, Clewell HJ, Thomas RS, Andersen ME. Incorporating high-throughput exposure predictions with dosimetry-adjusted *in vitro* bioactivity to inform chemical toxicity testing. Toxicol Sci. 2015 Nov;148(1):121-36.

Willemsen A, Vaal MA, De Zwart D, 1995. Microbiotests as tools for environmental monitoring. RIVM report 607042005.

Yoon M, Blaauboer BJ, Clewell HJ. Quantitative *in vitro* to *in vivo* extrapolation (QIVIVE): An essential element for *in vitro*-based risk assessment. Toxicology. 2015 Jun 5;332:1-3.

Attachment I

Specific extraction methods for parallel chemical and *in vitro* bioassay analysis

a. Mixed-mode solid-phase extraction

Within the EU-funded project EDA-emerge (ufz.de/eda-emerge), research was performed on sample preparation methods for effect-based and chemical identification monitoring of organic pollutants with adverse effect potential in European surface waters. The main aims of EDA-EMERGE were to develop a simplified protocol for effect directed analysis (EDA), link biological effects to target compounds and estimate their risk to aquatic biota. An earlier study at KWR demonstrated the suitability of a specific extraction method for parallel chemical analysis and bioassays for genotoxicity and endocrine receptor activation (Kolkman et al., 2013). The EDA-emerge contribution of KWR Watercycle Research was to further optimize a solid-phase extraction (SPE) method to allow the parallel study of effects in *in vitro* bioassays of surface water extracts in parallel to chemical analysis. An overview of the EDA-emerge project and the work on sample preparation for effect directed analysis of emerging organic micro-contaminants in surface waters is described below and is currently being prepared for publication.

For *in vitro* bioassay testing sample preparation techniques for pre-concentration and sample clean-up are required to detect effects in environmental water samples, which is less critical in current analytical methods. A wide range of sorbent materials is available for SPE, the choice influencing selectivity, specificity, affinity, and capacity based on both the characteristics of the target chemicals and the sample matrix. A number of criteria were defined for combined sample preparation. Firstly, this method should extract chemicals within the wide range of physico-chemical properties present in complex environmental samples. Secondly, the method should enrich compounds to a sufficient extent, to match the sensitivity limit of both analysis types. The enrichment for chemical analysis is commonly 1000 times, while the required enrichment for *in vitro* bioassays can be much higher (Leusch et al., 2017). Thirdly, the extraction methods should not include solvents or other components that interfere in the biological response.

For a large set of water relevant chemicals (n=117) with a wide range of log K_{ow} values (-4 to 5), solid phase extraction (SPE) was followed by liquid chromatography coupled to high resolution tandem mass spectrometry LC-(HR)MS/MS and *in vitro* bioassays. To this aim, a SPE procedure combining 4 different sorbent materials in a single mixed-bed multilayer cartridge (Figure S1.1) was optimized. Target compounds were analyzed using an LTQ FT Orbitrap interfaced with a HPLC-pump and autosampler Accela (Thermo Electron GmbH, Bremen, Germany). Target chemicals were identified and confirmed based on accurate mass measurements. The hypothesis was that this SPE approach would provide high enrichment

efficiency and sufficient recoveries for a wide range of compounds with different physicochemical properties.

A mixed-bed multilayer SPE cartridge is made by filling an empty glass column (6mL) with 200 mg Oasis HLB as a first material in enrichment flow direction. As a second material, 350 mg of a mixture of Strata X-AW, Strata X-CW and Isolute ENV+ in a ratio of 1/1/1.5 (X-AW/X-CW/ENV+) was used. One liter of surface water buffered with ammonium acetate 0.1 M at pH 6.5 was loaded over the mixed-bed cartridges, and test compounds were eluted sequentially with 5 mL of 20% methanol in acetonitrile, 6 mL of 0.5% ammonium hydroxide in 20% methanol in acetonitrile and 4 mL 1.7% formic acid in 20% methanol in acetonitrile. The eluates were collected together in glass test tubes and evaporated under a gentle stream of nitrogen up to 50 µL of DMSO. For further chemical analysis, 10 μL of 20,000 times concentrated DMSO extracts were pipetted and diluted in 200 μL of ultrapure water, yielding 1,000-fold concentrated analytical samples with a 5% DMSO. All extracts were stored at -18° C until analysis.

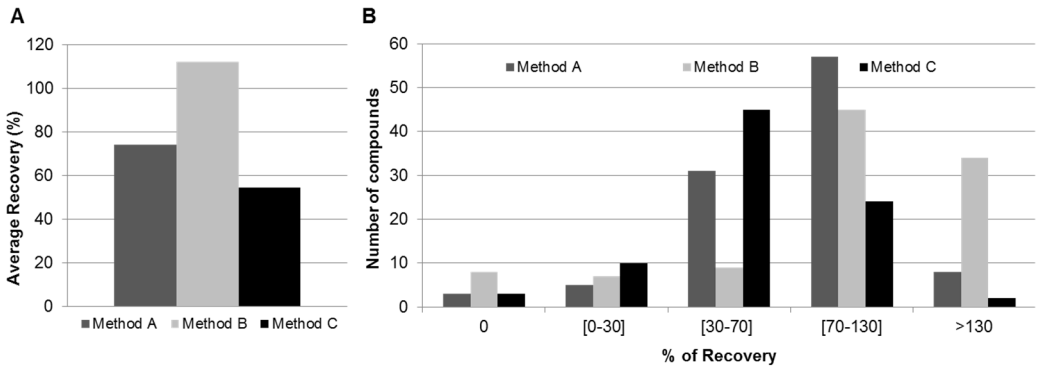


FIGURE S1.1. COMPARISON OF SPE PROCEDURES TESTED (I.E. METHODS A: MIXED-BED, B: OASIS HLB AND C: MCX) FOR THE EXTRACTION OF 117 DIFFERENT COMPOUNDS BY MEANS OF: (A) AVERAGE PERCENTAGE OF ABSOLUTE RECOVERY OF ALL COMPOUNDS TESTED; AND (B) NUMBER OF COMPOUNDS RECOVERED AT: 0 %, [0-30] %, [30-70] %, [70-130] % AND ≥130 % WITH A MAXIMUM RSD OF 30% IN ALL CASES.

To meet these criteria the enrichment of the selected chemicals was optimized and validated using further quantitative chemical analysis in combination with responses in CALUX *in vitro* bioassays. The performance of the mixed-bed SPE procedure was compared to other standardized SPE methods using a single sorbent. Based on recovery rates (70-130%) and precision (<30%), it was determined that the use of this mixed-mode SPE procedure is preferred over the other evaluated SPE protocols, although the impact of different SPE approaches on the resulting availability in the concentrate for different (types of) chemicals needs further study.

b. Large-volume solid phase extraction

Concentrates of a limited number of water samples (drinking water, surface water, and waste water treatment plan effluent) were an in kind contribution of KWR (BTO 2013.208(s)) to the subproject 'Environmental impact of chemicals, bio-based molecules & processes' of the BE-BASIC project (be-basic.org; Bio-Based, Ecologically Balanced Sustainable Industrial Chemistry). A special large-volume extraction SPE method was developed by KWR for this purpose (reported in BTO report 2013.208(s)).

To optimize comparability of the results, on the one hand between results in different *in vitro* bioassays at different partner institutes, and on the other hand between bioassays results and parallel chemical analyses, a large volume of concentrate was generated from a large volume of water. The concentrate was split and divided between BE-BASIC partners, and these concentrates were also tested in the ToxTracker assay. An integrated analysis of the generated data in BE-BASIC is pending.

References

Brandt A, Bäuerlein PS, Kolkman A. 2013. BTO 2013.208(s). Evaluatie van grootvolume-extractie als hulpmiddel voor de structuur opheldering van onbekende stoffen in water.

Kolkman A, Schriks M, Brand W, Bäuerlein PS, van der Kooi MM, van Doorn RH, Emke E, Reus AA, van der Linden SC, de Voogt P, Heringa MB. Sample preparation for combined chemical analysis and *in vitro* bioassay application in water quality assessment. Environ Toxicol Pharmacol. 2013 Nov;36(3):1291-303.

Leusch FD, Neale PA, Hebert A, Scheurer M, Schriks MC. Analysis of the sensitivity of *in vitro* bioassays for androgenic, progestagenic, glucocorticoid, thyroid and estrogenic activity: Suitability for drinking and environmental waters. Environ Int. 2017 Feb;99:120-130.

Attachment II

Scoring matrix instructions

The information included in this attachment has been reported earlier in the report 'Selection criteria to select *in vitro* bioassays for implementation and use' (Schriks et al., 2015).

TABLE S2.1. SCORING MATRIX FOR *IN VITRO* BIOASSAYS. SOURCE: SCHRIKS ET AL., 2015.

A. Assay applicability and ease of use (maximum combined score: 21)	
Criteria	**maximum score[a]**
applied to environmental samples	3
validated to water samples	3
standardized protocol available	3
service and support available	3
costs	3
ease of use	6[b]
non-GMO[c]	
no specialized equipment or skill required	
automation possible	
non-licensed *in vitro* (cell) model	
commercial kit available	
training available	
B. Assay performance (maximum combined score: 33)	
Criteria	**maximum score**
selectivity	3
accuracy	3
reproducibility	3
robustness	3
sensitivity	3
specificity	3
limit of detection	3
cytotoxicity control	3
speed	3
clear/straightforward read-out	3
high-throughput capacity	3

a. Number of points represent a poor (1 point), good (2 points) or excellent (3 points) score. b. score based on sub criteria (1 point each). c. Genetically Modified Organism

Assay applicability (maximum 21 points)

An important aspect of using assays within water quality monitoring is related to the feasibility of the practical steps involved. The test has to have a certain level of maturity to allow confidence in the results analyzed. Moreover, a very specific method may be of little use if it can only be performed at a limited number of laboratories due to the requirement of highly specialized equipment or specific skills of the researcher. The assay applicability part

of the scoring matrix will provide insight in the practical applicability of a bioassay. Instructions for the scoring of the individual criteria are included below.

Applied to environmental samples. The study of effects of pure individual chemicals is commonly performed in many bioassays. The applicability to study complex environmental samples is however less common. As the main aim of this effort is to select candidate bioassays for water quality analysis, it is an advantage if the bioassay is already demonstrated to be used for environmental samples. If the assay is applied widely to study effects of environmental samples, the maximum score of 3 points can be assigned. If the assay is used sporadically to study effects of environmental samples, a total of 2 points can be assigned. If there is no track-record of the use of the bioassay to study effects of environmental samples, only 1 point is assigned. If information on this criterion is absent, no points are assigned.

Validated to water samples. Further confidence can be given to the results of a bioassay if consistent output are produced, regardless of the origin of the water or the laboratory performing the bioassay. If the assay has been evaluated in an interlaboratory study for the study of the effects of water samples, the maximum score of 3 points can be assigned. If the assay is currently being evaluated in an interlaboratory study, 2 points can be assigned. If the assay was not evaluated in an interlaboratory study, 1 point is assigned. If information on this criterion is absent, no points are assigned.

Standardized protocol available. The availability of a standardized protocol demonstrates that the bioassay may be thoroughly evaluated and/or standardized, and that the bioassay is widely used and performed in a standardized manner. Aspect of this criteria include 1) whether a interlaboratory validation has been performed (ISO, OECD, DIN), 2) the availability of historical data to determine variability and trends in the data, 3) whether test guidelines have been agreed upon internationally, 4) is the test being considered for regulatory acceptance and 5) does the bioassay address the needs of end users and concerns of consumers. If the bioassay completely meets these subcriteria, the maximum of 3 points can be assigned. If the assay is currently in the process to meet these subcriteria, 2 points can be assigned. If the assay has not been standardized, 1 point is assigned. If information on this criterion is absent, no points are assigned.

Service and support available. During the development, application and/or validation of a bioassay to study effects of water samples, the availability of a service and support platform is critical if quality, regulatory or security challenges need to be addressed. The maximum score of 3 points can be assigned in the case the supplier of a bioassay provides service and support. If there is no service and/or support, 1 point is assigned. If information on this criterion is absent, no points are assigned.

Costs. Equipment and reagents required for performing *in vitro* bioassays can range from standard laboratory equipment to highly specialized or custom made materials and equipment. Per-sample costs of a bioassay mainly depends on the need to buy specific material for the bioassay and the number of replicates that can be studied simultaneously. If an assay is very cheap to perform (<100 Euro/sample), the maximum score of 3 points can be assigned. If the assay costs between 100-1000 Euro/sample, 2 points can be assigned. For more expensive assays (>1000 Euro/sample) only 1 point is assigned. If information on this criterion is absent, no points are assigned.

Ease of use. A number of practical considerations with considerable impact on the ease of use are included in this combined score. Since this criterion is of high relevance for the

implementation in routine monitoring by end-users, each subcriterion can be assigned 1 points and the maximum score is thus 6 points. If information on this criterion is absent, no points are assigned.

- Non-GMO. Does the assay make use of wild type cells instead of Genetically Modified Organisms? (GMO) (yes=1 point, no=0 points)
- No specialized equipment/skills required. Can the assay be carried out with relative straight forward equipment? (yes=1 point, no=0 points)
- Automation possible. Can the assay be automated? (yes=1 point, no=0 points)
- Non-licensed *in vitro* model. Is the cell line freely available? (yes=1 points, no=0 points)
- Kit available. Is the assay offered in standardized kit format? (yes=1 point, no=0 points)
- Training available. Does the supplier provide specific training? (yes=1 point, no=0 points)

Assay performance (maximum 33 points)

The criteria below give insight in the performance of a bioassay by evaluating the methodology and conclusions that can be based on the generated data.

Selectivity. The selectivity of an assay quantifies how much the response in a bioassay is impacted by matrix interference or the presence of non-relevant chemicals in contrast to the chemicals that specifically affect the pathway of interest. This is in particular important in the study of environmental samples such as complex low-level mixtures in water extracts. Selectivity may differ for different types of water samples. Usually, selectivity is expressed as a percentage, with a selectivity close to 100% indicating the assay is only responding to chemicals of interest without responding to other chemicals (that are known to be of less relevance for the pathway of interest). An assay with very high selectivity (>90%) scores the maximum of 3 points. An assay with low selectivity (<20%) scores only 1 point. If information on this criterion is absent, no points are assigned.

Accuracy. The accuracy of a test describes the degree of agreement between test method results and accepted reference values. When validating *in vitro* bioassays, the accuracy is generally determined by repeated analyses of known concentrations of one or more reference chemicals. A highly accurate assay scores the maximum score of 3 points. If information on this criterion is absent, no points are assigned.

Reproducibility. The reproducibility of a bioassay describes the agreement among results obtained from testing the same substance or samples (usually 10 or more) using the same test protocol, performed by different people, on different days and preferably in different laboratories. Evaluation of reproducibility should preferably be performed with test samples representing the full range of expected concentrations in water sample extracts. Reproducibility can be assessed at different levels. Inter-laboratory reproducibility (between-laboratory) is a measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results. Inter-laboratory reproducibility is determined during validation processes, and indicates the extent to which a test can be successfully transferred between laboratories. Inter-laboratory reproducibility should be below 30%. Intra-laboratory reproducibility (within-laboratory reproducibility) is a measure of how qualified people within the same laboratory can successfully replicate results using a specific protocol at different times. Intra-laboratory reproducibility should be below 20%. This aspect is related to precision and repeatability, which express how similar individual measurements of the same sample are when the

analysis is repeated several times under identical conditions. The value for repeatability should be close to 1 (or 100%). Repeatability and reproducibility are subject to both random and systematic errors (variability). Assays that fulfill all criteria as indicated above, can score the maximum of 3 points. If information on this criterion is absent, no points are assigned.

Robustness. The robustness of an assay characterizes the sensitivity of a method to operational variation and thus assesses the transferability of a method to other people and laboratories. It gives an indication of the ability of the assay to produce reliable results under slightly varying conditions, e.g. exposure time and temperature. The value for robustness is calculated as inter- and intra-laboratory reproducibility. An assay that scores high on robustness can score the maximum of 3 points. If information on this criterion is absent, no points are assigned.

Sensitivity. The sensitivity of a bioassay quantifies the proportion of all known positive/active substances that are correctly classified by the test. It is a measure of accuracy for a test method when considering categorical results, and its ability to correctly identify positive samples, but it does not take into account the concentrations needed for the positive response. As for bioassays the type and number of positive compounds in samples are generally unknown, this value is usually assessed using a large number of known positive and negative chemicals. Assays with low sensitivity may produce false negative results, which is negatively impacts the usability of a bioassay for screening purposes. Ideally, sensitivity should be close to 1 (or 100%). An assay that correctly identifies (>90%) positive samples can score the maximum of 3 points. An assay with low sensitivity (<20%) scores only 1 point. If information on this criterion is absent, no points are assigned.

Specificity. The specificity of a bioassay denotes the proportion of all negative/inactive substances that are correctly classified in the assay. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method. Assays that have a low specificity produce many false positive results, which is a problem when the bioassay is used for screening purposes. The value for specificity should be close to 1 (or 100%). An assay that has a high specificity (>90%) can score the maximum of 3 points. An assay with low specificity (<20%) scores only 1 point. If information on this criterion is absent, no points are assigned.

Limit of detection. The limit of detection (LOD) denotes the minimum amount of activity that can still reliably be detected in a particular bioassay. Such detection should lie within the limits defined for reproducibility and repeatability, although quantification may not be possible. In general, the LOD is calculated by interpolating the first significantly different response (e.g. compared to the signal from the blank + 3x the standard deviation of the blank) in the dose-response curve of the reference compound. The limit of quantification (LOQ) is similarly calculated based on the response from the blank (e.g. blank + 10x standard deviation of the blank). In addition to the LOD, the EC_{50} value of the reference compound is also indicative for the sensitivity of the assay. For water quality screening using a bioassay, it is critical that the LOQ is lower than health-based trigger value in this bioassay. A bioassay with a low LOD (<10 ng/L equivalents) scores the maximum of 3 points. If the LOD is high (>1 μg/L) only 1 point is assigned. If information on this criterion is absent, no points are assigned.

Cytotoxicity control. Cytotoxic effects of a chemical may mask the effect and/or hamper the interpretation of bioassay results. Therefore, the assessment and understanding of the effects of any chemical entity on cell health (viability) is critical to include in all cell-based screening techniques. If a cytotoxicity control method is included in a bioassay, the

maximum of 3 points is assigned. If information on this criterion is absent, no points are assigned.

Speed. Even if a bioassay produces information on the effects of a chemical that is very significant from a toxicological point of view or very relevant to human health, if the analysis itself takes a lot of time the bioassay cannot be used in practice to safeguard water quality. In other words, does the assay provide the information rapidly enough to initiate effective management action before unacceptable damage has occurred? The speed of an assay can be defined as the time it takes to produce a result, i.e., the time frame from taking a water sample to having the final results, including and excluding sample pre-treatment. This timeframe depends on the type of analysis, endpoint and phase in drinking water preparation (in which intervention may be necessary). For this criterion, the maximum score of 3 is when the test yields results within a day, a score of 2 points for results within a week, and only 1 point if the test takes more than a week to perform. If information on this criterion is absent, no points are assigned.

Clear/Straightforward read-out. The read-out of a bioassay - the recorded observation - can vary from being very general (e.g. cell death) to very specific (activity on a specific receptor resulting on a fluorescent marker read-out). If the read-out is very straightforward to interpret (e.g. relative light units or optical densities) the maximum score of 3 can be assigned. If the read-out requires a lot of handling before interpretation is possible (e.g. radio ligand binding assays) a score of 1 point is assigned. If information on this criterion is absent, no points are assigned.

High-throughput capacity. Rapid and cost-effective profiling of the bioactivity of chemicals of unknown toxicity and assess the associated hazard and risk for adverse effects can be optimized if the bioassay has high-throughput screening capacity. The possibility to apply robotics, automated sample workup, miniaturized assay formats, liquid handling devices, sensitive detectors, high-speed plate readers, data processing and control software facilitates the generation of large number of individual assay data points. This will make the screening more efficient, and reduces the per-sample analyzing costs. High-throughput bioassays are assigned the maximum of 3 points. If a bioassay may be compatible with high-throughput optimization, 2 points are assigned. If the assay is very laborious and automation is not expected, only 1 point is assigned. If information on this criterion is absent, no points are assigned.

Attachment III

ToxTracker data of individual chemicals

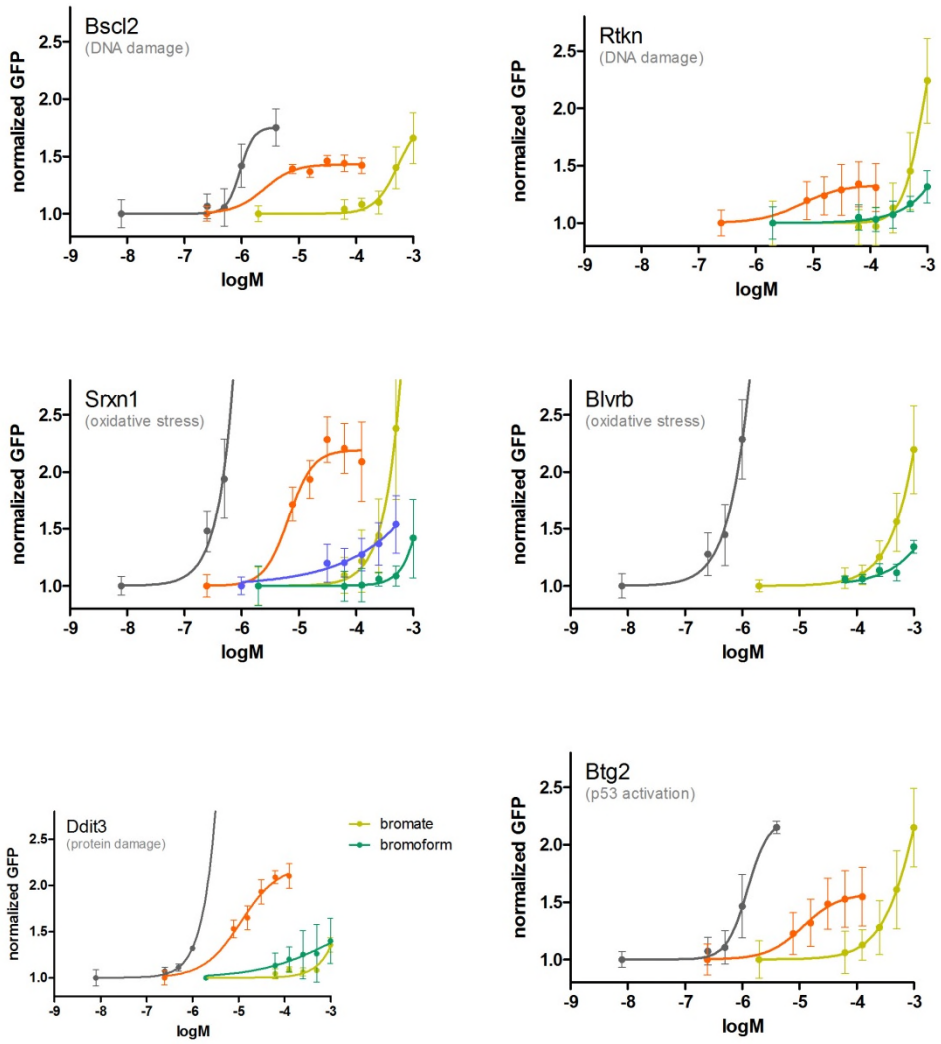


FIGURE S3.1. CONCENTRATION-DEPENDENT RESPONSES (MEAN±STDEV) INDUCED BY INDIVIDUAL CHEMICALS ON TOXTRACKER BIOMARKER GFP. ARSENIC: GREY; BENZO[A]PYRENE: ORANGE; BROMATE: LIGHT GREEN; BROMOFORM: DARK GREEN; DICHLOROBROMOMETHANE: BLUE; NO EFFECT DATA IS NOT INCLUDED (RESPONSES OF ALL CHEMICALS ON ALL MARKERS WERE INVESTIGATED).

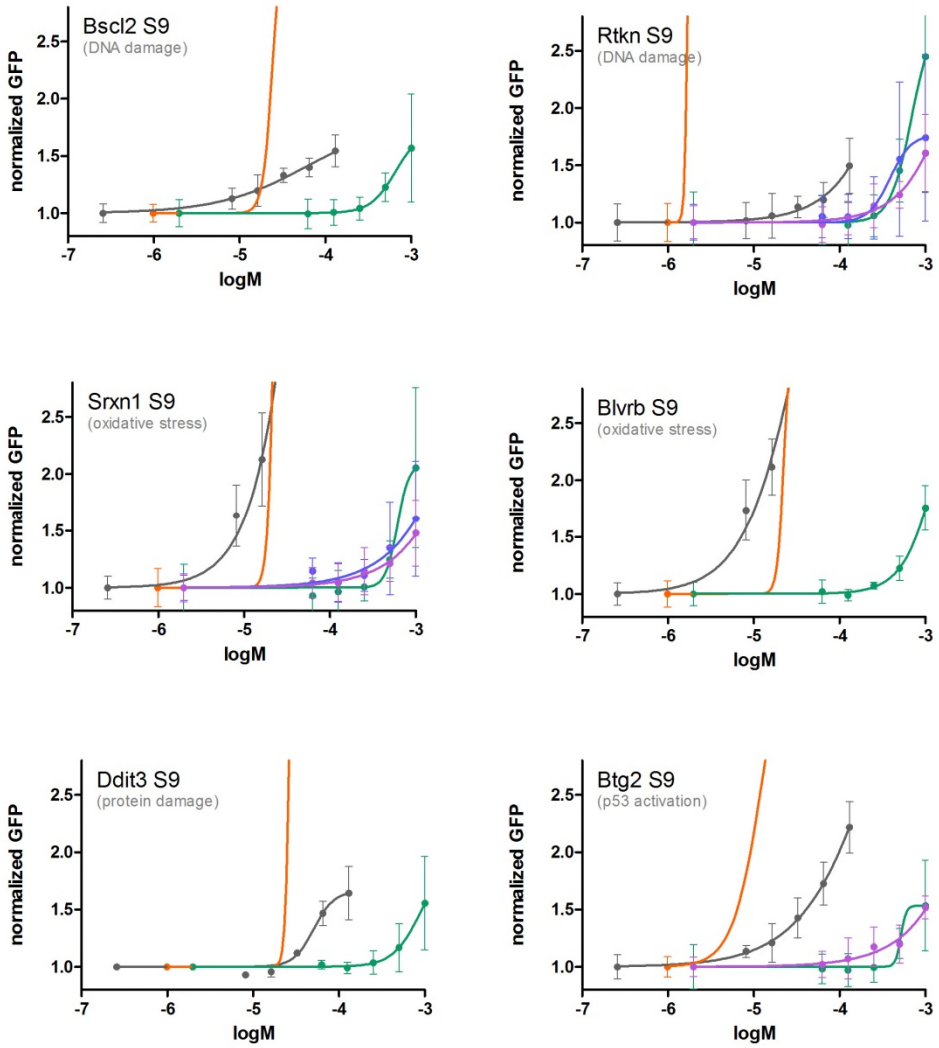


FIGURE S3.2. CONCENTRATION-DEPENDENT RESPONSES (MEAN±STDEV) INDUCED BY INDIVIDUAL CHEMICALS ON TOXTRACKER BIOMARKER GFP EXPRESSION IN THE PRESENCE OF METABOLIC S9 MIXTURE. ARSENIC: GREY; BENZO[A]PYRENE: ORANGE; BROMOFORM: DARK GREEN; ; DICHLOROBROMOMETHANE: BLUE; TETRACHLOROETHENE: PURPLE. NO EFFECT DATA IS NOT INCLUDED (RESPONSES OF ALL CHEMICALS ON ALL MARKERS WERE INVESTIGATED).

Attachment IV

ToxTracker data water samples

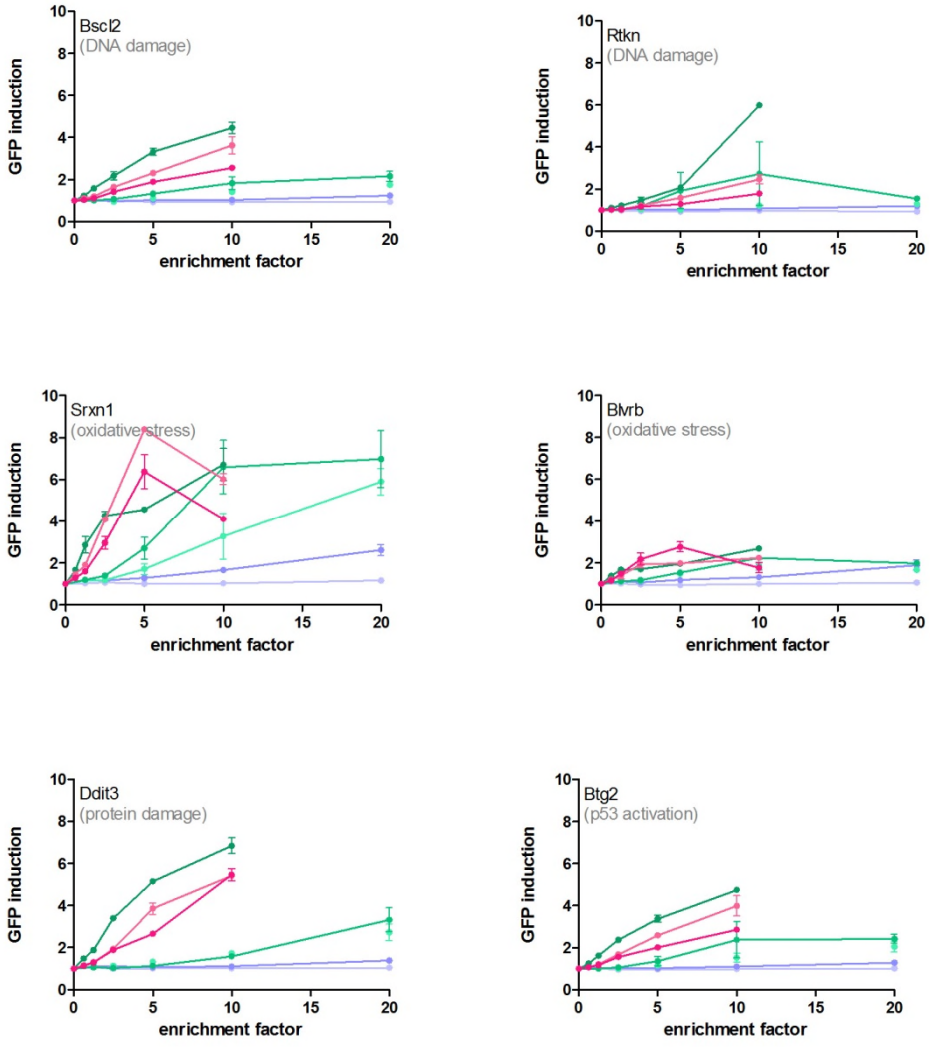


FIGURE S4.1. RESPONSES (MEAN±STDEV) INDUCED BY A SELECTION OF WATER SAMPLE CONCENTRATES (AT DIFFERENT FINAL ENRICHMENTS IN THE ASSAY) ON TOXTRACKER BIOMARKER GFP. BLUE: DRINKING WATER; GREEN: SURFACE WATER; PINK: WASTE WATER TREATMENT PLANT EFFLUENT.

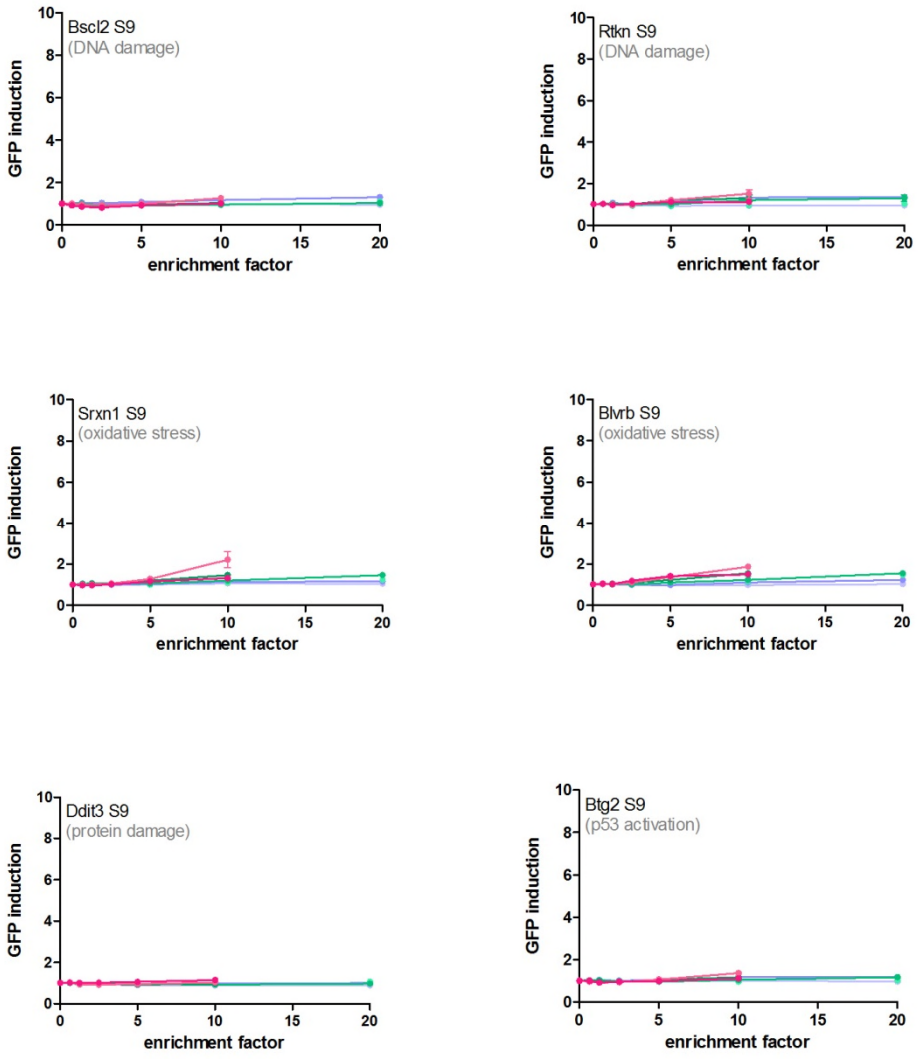


FIGURE S4.2. RESPONSES (MEAN±STDEV) INDUCED BY A SELECTION OF WATER SAMPLE CONCENTRATES (AT DIFFERENT FINAL ENRICHMENTS IN THE ASSAY) ON TOXTRACKER BIOMARKER GFP IN THE PRESENCE OF METABOLIC MIXTURE S9. BLUE: DRINKING WATER; GREEN: SURFACE WATER; PINK: WASTE WATER TREATMENT PLANT EFFLUENT.

Attachment V

ToxTracker mixture data

Abstract prepared for the annual meeting of the Dutch Toxicology Association (Nederlandse Vereniging voor Toxicologie). Data will be presented in the form of a poster which will be made available via BTO Net.

Mixture effects of chemicals included in Dutch drinking water regulations in the ToxTracker

Milou M.L. Dingemans[1*], Kirsten A. Baken[1], Merijn Schriks[2], Giel Hendriks[3], Annemarie P. van Wezel[1,4]
[1] KWR Watercycle Research Institute [kwrwater.nl], Groningenhaven 7, P.O. Box 1072, 3430 BB Nieuwegein, the Netherlands
[2] Vitens [vitens.nl], Oude Veerweg 1, 8019 BE Zwolle, the Netherlands
[3] Toxys B.V. [toxys.com], Robert Boyleweg 4, 2333 CG Leiden, the Netherlands
[4] Copernicus Institute [www.uu.nl/copernicus], Utrecht University, the Netherlands
[*] Corresponding author: milou.dingemans@kwrwater.nl

Dutch legislation requires routine monitoring of a selection of chemicals in order to safeguard drinking water quality. Human health relevance of individual water contaminants is assessed based on their measured concentrations and health-based guideline values. There is increasing interest for the use of *in vitro* bioassays for water quality monitoring as these tests can detect a combined biological response induced by chemicals in mixtures.

The aim of this research was to evaluate the applicability of the ToxTracker assay (a stem cell-based reporter assay for genotoxicity and cancer hazard assessment) for water quality monitoring and to explore possible mixture effects. DMSO concentrates of surface water and waste water treatment plant (wwtp) effluent samples were generated using solid-phase extraction. Effects of these environmental mixtures and binary mixtures of arsenic (0.004-2 µM), benzo[a]pyrene (0.1-50 µM), bromoform (2-1000 µM) and bromate (2-1000 µM) were investigated.

In these pilot experiments, ToxTracker responses were observed for surface water and wwtp effluent extracts. While surface water extracts induced responses when concentrated ≥5x, wwtp effluent extracts induced responses when (much) less concentrated. All tested binary mixtures induced reported gene expression at concentrations below the lowest-observed effect concentration of the individual chemicals (for different reporter genes).

The presented data demonstrate that the ToxTracker assay can be used to detect effects of environmental mixtures and that mixture toxicity can occur between chemicals that are currently assessed for their health impact on an individual basis.

Funding. Joint Research Programme of the Dutch Water utilities (BTO2013-2017).

Keywords: water, *in vitro* bioassay, genotoxicity, mixture

Attachment VI

Survey responses

1. Welke *in vitro* bioassays heeft u tot uw beschikking en in hoeverre past u deze toe in de dagelijkse praktijk?

'Microtox, algentoxkit, daphniatoxkit en RIKILT antibiotica assay (CALUX assays via BDS) enkele campagnes per jaar' – Waternet(RvdO)

'ER-CALUX, GR-CALUX, anti-AR-CALUX, p53-CALUX, Nrf2-CALUX, PAH-CALUX' – HWL(TvdVS)

'Vitens heeft zelf geen bioassays tot haar beschikking, maar werkt wel aan de inrichting voor basisvereisten van een ML1 laboratorium. In het kader van twee toepassingen is er wel ervaring gemaakt met de toepassing van bioassays: 1) Automatische monstervoorbewerking ten behoeve van de ER CALUX (project voormalige collega Varvara Kokkali). Punt op de horizon was/is automatische toepassing in het kader van bronbewaking. Dit project is nog niet afgerond; 2) Controle van Vitens humuszuur extract (humvi) op dioxineachtige stoffen met behulp van de DR CALUX bioassay (partner BDS).' – Vitens (MS/JS)

2. Welke effect trigger values (drempelwaarden) past u toe om gemeten effecten in *in vitro* bioassays te duiden?

'SIMONI EBT (Van der Oost et al, 2017, in press)' – Waternet(RvdO)

'De door Brand et al. bepaalde triggervalues van 3,8 ng/L E2-eq (ER) en 21 ng/L DEX-eq (GR) (Environment International 55 (2013) 109–118). Voor de anti-AR CALUX is nog geen triggervalue beschikbaar, maar die zouden we wel heel graag hebben.' - HWL(TvdVS)

'Vooropgesteld dat de referentie stof (zoals EE2 in ER-Calux) (een van) de meest potente is zou je kunnen concluderen dat de werkelijke concentratie (ug/l) van de veroorzakende stof vele malen hoger is. In dat geval is een trigger waarde van 0,1 ug/l het maximum dat ik zou willen toestaan. We hebben dan te maken met een biologisch actieve stof waarvan de werkelijke waarde hoger ligt dan wat vereist is door het Europees Rivieren Memorandum (zie bijlage)' –RIWA(GS)

'Zoals genoemd past Vitens zelf nog geen bioassays toe. Wel is het onderscheid in triggervalues voor ecosystem health of human health duidelijk en ook naar welke literatuur gekeken moeten worden. In het kader van humvi controle (met DR CALUX) wordt gebruikt gemaakt van internationale richtlijnen voor dioxine achtige stoffen (TEF/TEQs).' – Vitens (MS/JS)

3. Zijn er recente ontwikkelingen of heeft u verwachtingen voor de nabije toekomst met betrekking tot de toepassing van *in vitro* bioassays?

'Waternet gaat steeds meer bioassay analyses gebruiken als het aan mij ligt' - Waternet(RvdO)

'Een eerste toepassing is het routinematig meten van de relevante eindpunten die benoemd zijn in het recente BTO-rapport van o.a. Kirsten Baken en Merijn Schriks. Dit als aanvulling op

de huidige monitoring met doelstoffenanalyses en screenings. Daarnaast hebben we bij HWL het ht-EDA platform geïmplementeerd dat door PhD studenten van de VU is ontwikkeld. Dit jaar willen we deze gaan inzetten om in watermonsters waar activiteit waargenomen wordt in de bioassays proberen te identificeren welke componenten de oorzaak zijn van de activiteit. Verder willen we kijken of het mogelijk is om enkele doelstoffenanalyses te vervangen met geschikte bioassays, zoals de PAK analyses en de genotoxische amines.' - HWL(TvdVS)

'Combinatie met chromatografische scheiding geeft direct inzicht in mengsel-tox. Je test het totale monster en ook de fracties en vergelijkt de responsen. Daarnaast zijn er indicaties dat natuurlijke niet toxische stoffen de totale respons ook kunnen beïnvloeden, zie bijlage. In dat geval kan de tox hoger zijn dan verwacht op basis van chem analyse.' – RIWA(GS)

'Niet bij Evides. Ik hoop wel dat ze een rol gaan spelen in de beoordeling van complexe mengsels' – Evides(HK)

'De ontwikkeling van trigger values lijkt een verdere vlucht te nemen in de diverse gremia. Vooralsnog levert dit voor de eindgebruikers bescheiden resultaat. Verwachting is dat de bioassays een rol krijgen bij risk based monitoring (RBM) vanaf 2019.' – Vitens (MS/JS)

4. Welke kansen en mogelijkheden ziet u met betrekking tot verdere analyses van de verzamelde *in vitro* bioassay data, en het faciliteren daarvan door het centraal verzamelen van zulke data in een database?

'Kan een goede kans zijn om de relevantie van bepaalde bioassays en de EBT daarvan te evalueren' - Waternet(RvdO)

'Een mogelijkheid die ik vooral zie voor de analyse van de bioassay data is om op basis van een trigger waarde of op basis van afwijkingen/verhogingen tov eerdere metingen op dezelfde locatie te beslissen of verder onderzoek noodzakelijk is. Zo ja, dan kunnen doelstoffenanalyses uitgevoerd worden of het ht-EDA platform ingezet worden.' - HWL(TvdVS)

'Met name identificatie van nieuwe/onbekende stoffen bergt veel inspanning. Het zou goed zijn te overwegen om niet alleen resultaten maar ook de analyse/identificatie inspanning te delen door bijvoorbeeld inkoop van analysetijd bij universiteitslabs.' - RIWA(GS)

'Dat kan helpen bij standaardisatie en begrip van de uitkomsten' – Evides(HK)

'Het lijkt erop dat dit vooralsnog niet realiseerbaar is omdat er geen concensus is over methoden. Wel is het nuttig om TEFs te verzamelen van de diverse referentiestoffen.' – Vitens (MS/JS)

5. Wat ontbreekt nog / welke wensen en verwachtingen heeft u voor de toepassing van *in vitro* bioassays in de (nabije) toekomst?

'Betere *in vivo – in vitro* extrapolatie, indicatie over adverse outcome pathways, meer EBT voor drinkwater' - Waternet(RvdO)

'Het belangrijkste wat nog ontbreekt zijn de trigger values voor alle aanbevolen relevante eindpunten. En wat nog voor een deel ontbreekt is ervaring/begrip/vertrouwen m.b.t. wat de resultaten in een bioassay betekenen bij de mensen die nog nooit met bioassays gewerkt hebben of betrokken zijn geweest bij projecten. Dat vraagt meestal een goede uitleg' - HWL(TvdVS)

'Wees voorzichtig met de boodschap dat niet-bewezen-toxisch betekent dat er geen probleem is. We weten meer over een monster maar nog steeds niet alles.' - RIWA(GS)

'Juiste interpretatie. Standaardisatie tussen labs. Juiste trigger values' – Evides(HK)

'Vitens ontbreekt een dataset (gericht op winningen) om een analyse te kunnen maken over de nut/noodzaak van bioassays. Er zijn twee sporen mogelijk: 1) een (paar) productielocatie(s) -zoals Vechterweerd- van puttenveld tot reinwater doormeten om te beoordelen of sprake is van een afnemende bioassay respons in het productieproces. Hierbij zijn de triggervalues minder belangrijk omdat een afname van de bioassay respons inzichtelijk maakt dat de zuivering goed werkt; 2) Een selectie van bioassays -specifiek voor reactive mode of action (MOA) (genotox, mutatox, oxidatieve stress respons) opnemen in het meetprogramma om een beeld te krijgen op ruwwaterkwaliteit van een aantal kwetsbare winningen (bijvoorbeeld Engelse werk, Vechterweerd, Epe etc). Wat nog ontbreekt is een koppeling tussen non-target screening en bioassay respons. Dit zou verder uitgewerkt kunnen worden in nieuwe meetprogramma's. Vitens wil graag inzetten op bioassays met een reactieve MOA en gelooft dat bioassays die specifiek zijn voor hormonen een lagere prioriteit hebben (in het kader van grondwater).' – Vitens (MS/JS)

6. Wat zou u nog willen meegeven in het kader van nieuw onderzoek naar de toepassing van *in vitro* bioassays?

'Zie hierboven (vraag 5) en voorstellen GWRC workshop' - Waternet(RvdO)

'Focus iig op de triggerwaarden voor de relevante eindpunten die al geselecteerd zijn en waarvoor nog geen waarde beschikbaar is. Probeer te blijven verkennen wat geschikte assays zijn om de geselecteerde eindpunten te onderzoeken en of er aanvullende relevante eindpunten zijn die getest kunnen worden met bioassays (zoals er nu aandacht is voor ontwikkelingstox en neurotox).' - HWL(TvdVS)

'Bundel en deel de inspanning, zoek ook contact via NWO en STW met andere kennisinstituten, probeer niet alles zelf te doen.' - RIWA(GS)

'Hou het heel dicht bij de praktijk van het beoordelen van het totale effect van de aanwezige stoffen op effecten voor de mens (en niet bijv. op foetussen).' – Evides(HK)

'Een belangrijk aspect voor drinkwaterbedrijven is efficiëntie behalen in de uitvoering van meetpakketten. Op dit moment zijn drinkwaterbedrijven wettelijk verplicht om verschillende somparameters te meten zoals PCB's, genotoxische amines, PAK's en pesticides. Het zou toegevoegde waarde hebben om met bioassays deze somparameters te bepalen in plaats van met klassieke (relatief kostbare) methoden. Een goed voorbeeld kan genomen worden aan dioxine-achtige stoffen in o.a. eieren en de bepaling daarvan met de DR CALUX. In aanvulling blijft de sensitiviteit van bioassays een heikel punt. Onderzoek zou zich moeten concentreren op het overbodig maken van een concentratiestap zodat de vergelijking met (non-)target methden beter te maken is. Tenslotte is uiteraard het beoordelingskader een belangrijk ontwikkelveld.' – Vitens (MS/JS)

Attachment VII

GWRC bioassay workshop report

Blog written to report the contribution of KWR in the GWRC Bioassay Workshop (February 2017, Paris). Published February 23, 2017 on the KWR website (kwrwater.nl).

KWR contribution to the GWRC bioassay workshop

GWRC organized a workshop to get insight in the developments in bioassay applications and the aims and needs of researchers, stakeholders and end-users. There was clear consensus on the advantages of the addition of bioassays to the water quality toolbox. Clear common goals were defined and discussed, and concrete proposals were written for research efforts to support the implementation of bioassays. The GWRC workshop gave us a great opportunity to communicate and discuss the KWR/Dutch view on the implementation of bioassays in water quality monitoring.

Bioassays: many applications, a few hurdles to overcome
There is a major interest of numerous international stakeholder in water quality research and management in the potential application of bioanalytical tools, also known as *in vitro* bioassays, or, in short, bioassays. The Global Water Research Coalition (GWRC), is a network of researchers from academia, research institutes and private water companies. This week GWRC organized a workshop (Feb21-22) to get insight in the developments in bioassay applications and the aims and needs of researchers, stakeholders and end-users. Although I am relatively new in water research, as a toxicologist I am already very familiar with the application of *in vitro* cellular models in research on the effects of chemical exposures. This gave me confidence that I could contribute to the discussions, despite the fact that this was my first introduction in this international network for water research. Stefan (Kools) also attended the workshop, and he noticed in particular that both 'operators' and 'regulators' express the expectation that the application of bioassays will aid in the demonstration of and communication on safer and cleaner water. Not only between people working in the water sector, but also with the public and clients of drinking water companies. In the workshop discussion groups on the firsts day, there was clear consensus on the advantages of the addition of bioassays to the water quality toolbox. Nevertheless, there were also worries expressed about the hurdles that need to be overcome, in particular related to interpretation of and confidence in bioassay results. On the second day, clear common goals were defined and discussed, and concrete proposals were written for efforts to efficiently overcome these hurdles.

Paris and bioassays by night
During the workshop diner, discussions on bioassays continued. Despite or maybe thanks to the beautiful view on Paris by night from the diner boat, discussions were not avoided on the potential contributions of different stakeholders to support implementation of bioassays for water quality monitoring. All in all, the GWRC workshop gave us a great opportunity to communicate and discuss the KWR/Dutch view on the implementation of bioassays in water quality monitoring.