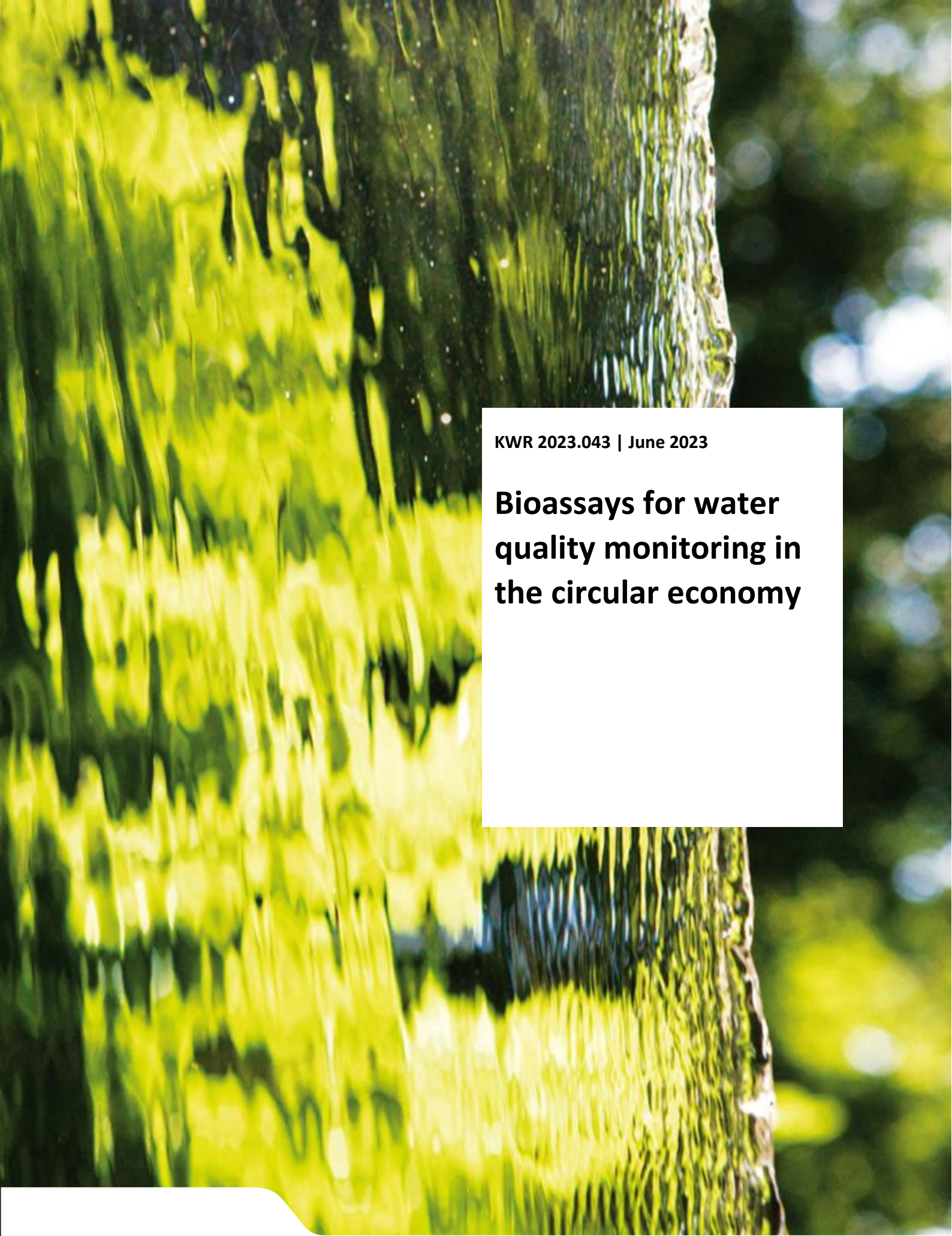

KWR 2023.043 | June 2023

Bioassays for water quality monitoring in the circular economy

KWR

Bridging Science to Practice

Collaborating Partners

Report

Bioassays for water quality monitoring in the circular economy

KWR 2023.043 | June 2023

Project number
403616

Project manager
Dr. Patrick Bäuerlein

Client
Core group WiCE

Author(s)
Astrid Reus MSc., dr. Sanah Majid Shaikh ERT, dr. Milo de Baat

Quality Assurance
Dr. Milou Dingemans, ERT

Sent to
This report is distributed to the clients of this research and is public.

Keywords
in vitro bioassays, effect-based methods, genotoxicity, water quality monitoring, circular economy, water reuse

Year of publishing
2023

More information
Astrid Reus
T +3130 606 9703
E Astrid.Reus@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

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

Summary

Freshwater from groundwater and surface water sources is becoming scarce in dry periods, which raises interest in the (re)use of alternative sources such as rainwater and treated domestic and industrial wastewater (effluent). This requires consideration of the potential risks resulting from circulating and potentially accumulating microbial and chemical contaminants in different water cycles, depending on the nature of water reuse. Therefore, it is important to gain insight into potential microbiological and chemical hazards. This knowledge is needed to safely and responsibly reuse water.

The current project focused on chemical hazards. Effect-based methods using *in vitro* cell models and bacteria (bioassays) offer a powerful approach to investigate relevant hazards of complex low-level mixtures in the water cycle without prior information on their chemical composition. Panels of bioassays covering endpoints relevant for ecosystem health and drinking water safety have been compiled and implemented in the water sector. In particular bioassays for investigating DNA damage or endocrine disruption are increasingly implemented within the water sector. These complement chemical analyses to assess the quality of (drinking) water sources and the effectiveness of treatment processes.

KWR has developed the Water Wise concept, an easy-to-use, transparent and consistent evaluation framework as a first step toward quantitative evaluation of microbiological and chemical risks inherent to new water cycles. Research and expert judgement on new water cycle-related hazards and risks are needed to further develop the framework (tool under development). Depending on the information fed into the tool and the Water Wise risk assessment outcome, the output may include a recommendation for the application of bioassays to assess the quality of alternative freshwater sources or to confirm the outcome of the tool.

In the current project, a set of bioassays assessing different types of DNA damage was selected (Ames fluctuation test, umu chromotest, micronucleus test, comet assay, p53 CALUX and ToxTracker) and applied to different types of water samples from water cycles (rainwater and wastewater treatment plant (WWTP) effluent from three different locations). DNA damage was selected as the primary biological effect studied here because of its relevance for human health by inducing mutations and tumor formation and because well-established *in vitro* assay protocols for this endpoint are available. As the interpretation of bioassay responses regarding the potential effects on human health is not straightforward, results were compared, where possible, to assay-specific effect-based trigger values that identify the bioassay response level where a human health risk cannot be excluded.

The bioassays responses for the defined water cycles are presented, alongside a proposed strategy for data interpretation and water quality assessment in a water reuse context. This will support responsible water reuse through the implementation of bioassays in the circular economy and by creating awareness on the need for water quality assessment for existing and new, alternative freshwater sources.

Contents

1 Introduction

1.1 Challenges in freshwater (re)use

Over the last decade, access to clean and safe water has become one of the major challenges facing modern society (Jackson et al. 2001). Currently, two thirds of the world's population live in areas that experience scarcity of water for at least one month a year (Conner et al. 2017). Large urban areas in Australia, the USA, Asia and Europe have seen a shrinkage of groundwater and surface water sources, caused by the growing imbalance between water availability and consumption. This imbalance is a consequence of extensive droughts and increased population (Jackson et al. 2001; Leusch et al. 2014). Water for food and irrigation accounts for up to 70% of the freshwater extraction in fast-growing economies (Conner et al. 2017). Access to clean and safe water is not only essential for human health, but also for ecosystems health, as the earth's ecosystems do not function without sufficient water supplies. However, as the human population expands, the water supply by rivers, lakes and wetlands will further decrease in quantity (Chicharo et al. 2015). As a result of the decreasing water supply quantities, large cities began to import water from less populated areas. Coastal cities have also begun to desalinate seawater into fresh water. Exploiting these alternative water supplies is costly, energy intensive and not available in all areas. Therefore, other alternatives such as the use of reclaimed water (water reuse, e.g. using waste water treatment plant (WWTP) effluent and rainwater) have become increasingly important (Leusch et al. 2014; UNWWAP 2015).

Water reuse refers to the process of using treated wastewater (effluent) for beneficial purposes such as agricultural and landscape irrigation, industrial processes, non-potable urban applications (such as watering the garden and car washing), groundwater recharge (sub-irrigation) and potable use (Angelakis and Snyder 2015; Ternes et al. 2017). However, the contamination of water sources with chemical pollutants is a major threat to water quality worldwide (Malaj et al., 2014). Anthropogenic (human-made) substances, such as pharmaceuticals, personal care products and household care products that are discharged into the sewage by households and hospitals (de Jongh et al. 2012; Oliveira et al. 2015) are generally not effectively removed by conventional wastewater treatment plants (WWTPs) (Aydin et al. 2019; Golovko et al. 2021; Oliveira et al 2015). Consequently, treated wastewater still contains chemical contaminants that elicit a chemical burden on aquatic systems to which the effluent is discharged. Consequently, indirect water reuse is the status quo (de facto) and human health needs to be considered when surface water is a drinking water source (Beard et al. 2019; Houtman 2010). Pronk et al. 2021). Due to the presence of many, different chemical contaminants, various biological processes may be affected, including DNA damage and disturbances to the hormone balance (Dopp et al., 2019). The onset of potential human and ecological health risks resulting from circulating and accumulating microbial and chemical contaminants should be prevented (Grundmann et al. 2013; Voulvoulis, 2018). Prevention of contamination of water sources is essential to reduce the effort needed in water treatment processes and treatment-related hazards and risks need to be considered. The more contaminated the water, the more advanced water treatment is required, which may result in the formation of potentially toxic disinfection byproducts and transformation products. Some of these newly formed compounds can damage the DNA and negatively affect other cellular processes (Dieter 2010, Escher and Fenner 2011, Srivastav et al. 2020). Thus, the evaluation of (potential) chemical hazards and associated human health and environmental risks is essential for safe and responsible use of (recycled) water.

In the last decades, the prevention of further pollution of surface water and measurements to determine the concentration of pollutants in the surface water gained more attention. This led to the EU water framework directive (WFD) (2000/60/EC) and surface water quality standards for priority substances (2013/09/EU). Water quality is generally evaluated by comparing actual measurements to environmental and drinking water quality standards, in which a series of physicochemical parameters are regulated. However, quality standards mainly exist for drinking water, but to a lesser extent for other applications such as irrigation and industrial purposes. For irrigation and industrial purposes, less strict quality standards compared to drinking water standards may be needed (fit-for-purpose standards). Environmental quality standards are available for only a limited number of substances of interest, while surface water typically contains thousands of substances in complex mixtures and mixture effects remain unknown (Escher et al., 2014). This emphasizes the need for additional guidance on water quality assessment of existing and new water cycles.

1.2 Decision-support system for safe water (re)use cycles

KWR has developed the Water Wise concept, an easy-to-use, transparent and consistent evaluation framework as a first step toward the quantitative evaluation of microbiological and chemical risks inherent to new water cycles. Research and expert judgement on new water cycle-related risks are needed to further develop the framework (Hockin et al. 2020).

The Water Wise concept evaluates the microbial and chemical water quality, compares with relevant national and international guidelines, and offers users a basic risk assessment process. The goal of Water Wise is to support users in making safer and smarter decisions concerning their water cycles, in order to identify solutions for optimal water (re)use. It represents a first step toward a tool to quantitatively evaluate the risks inherent to new and existing water cycles, and creates a better understanding of how end-users can avoid and manage these risks a priori, e.g. by using advanced treatment processes. More research, including further development of the quantitative chemical risk assessment (QCRA) concept developed by Cantoni et al. (2021), is needed to develop the framework further into a user-friendly tool (Hockin et al. 2020). Depending on the information fed into the tool and the outcome of the Water Wise risk assessment, it is foreseen that the output may include a recommendation for the use of effect-based methods. This also requires additional guidance and support on which specific methods should be applied, which can be fed by the results of the current project.

1.3 Bioassays for safe water (re)use cycles

Currently, there are different methods to assess chemical water quality, of which quantitative targeted compound methods, targeted and non-targeted (chemical) screening and effect-based methods (bioassays) are most frequently used (Prasse et al. 2015). Hazards of each individual substances identified by chemical analysis can be assessed to get an idea on the risk for human health or the environment. However, for many substances toxicological data will be lacking and the collection of all required data is very time-consuming. Moreover, when assessing the hazard of individual substances, mixture effects are not considered. Results of chemical analysis can therefore not easily be used for a direct hazard and risk assessment of complex chemical mixtures such as those found in the water cycle. Effect-based methods using *in vitro* cell models and bacteria (bioassays) offer a powerful approach to investigate relevant hazards of complex low-level mixtures in the water cycle without prior information on their chemical composition (Dingemans et al. 2019, Di Paolo et al. 2016, Leusch et al. 2014, Xu et al. 2019). *In vitro* bioassays based on bacteria, yeast or mammalian cells are designed to respond to specific toxicity mechanisms and are frequently used for safety assessment of (treated) wastewater, drinking water and environmental waters (Escher et al. 2014; Leusch et al. 2018; Macova et al. 2010; Neale et al. 2017; Välitalo et al. 2017). Panels of bioassays covering endpoints relevant for ecosystem health and drinking water safety have been compiled and implemented in the water sector, in addition to chemical analyses, to assess the quality of (drinking) water sources and the effectiveness of treatment processes (Escher et al. 2014, Macova et al. 2011, Leusch et al. 2014a, Schriks et al. 2015, Tang et al. 2013).

Since bioassays are generally rather specific, a single bioassay cannot assess the complex toxicity of (reclaimed) water, and a battery of bioassays involving multiple biological effects or *in vitro* tests with specific toxicity mechanisms is recommended to cover multiple potential health effects (Xu et al. 2019). Moreover, within a biological effect (e.g., DNA damage), multiple cellular and molecular mechanisms can lead to an adverse effect (e.g., gene mutations and cytogenetic damage), emphasizing the need of using multiple bioassays for the same biological effect in some cases (Mohrenweiser and Jones 1990, Sommer et al. 2020). The availability of a variety of bioassays and differences in terms of sensitivity and specificity complicate the selection of suitable bioassays for water quality assessment even if the endpoint of interest is defined.

1.4 Relevant biological effects for water quality assessment

Disrupting effects on DNA (genotoxicity), hormone system (endocrine disruption) and nervous system (neurotoxicity) can occur after exposure to low concentrations of chemicals. Based on known effects after exposure to water relevant substances these endpoints are commonly prioritized for water quality assessment (Dingemans et al. 2019; Escher et al. 2014; Keifer and Firestone, 2007; Kortenkamp, 2014; Kuckelkorn et al. 2018; Nohmi, 2018).

1.4.1 Genotoxicity

Genotoxicity is defined by the potential of a chemical to damage the genetic information in a living cell. Genetic damage can either arise through gene mutations or via cytogenetic damage. Gene mutations are alterations in the nucleotide sequence of the DNA, including insertions, deletions and substitutions of base pairs (Mohrenweiser and Jones 1990). Cytogenetic damage includes partial loss of chromosomes (i.e., clastogenicity) and loss of complete chromosomes (i.e., aneugenicity) (Sommer et al. 2020). Many *in vitro* bioassays for genotoxicity are available. Both gene mutations and cytogenetic damage were investigated in the experimental part of the project. In addition, a literature search was performed on the application of genotoxicity bioassays in a water reuse context.

1.4.2 Endocrine disruption

Endocrine disrupting chemicals (EDCs) are exogenous compounds that alter the function(s) of the endocrine system, of which 17α-ethinylestradiol in oral contraceptives is a well-known example (Marty et al. 2018). EDCs consist of a very heterogenous group of chemicals that can interfere with biological processes by mimicking hormones, activating or blocking the body's hormone receptors, disrupting the synthesis of hormones, or altering their degradation (Unüvar & Büyükgebiz, 2012). Endocrine disruption may result in adverse health effects in an intact organism, or its progeny, or (sub)populations, including developmental malformations, reproductive impairments, increased cancer risk, and disruption of immune and nervous system function (USEPA, 2022; WHO/IPCS, 2002). Different types of wastewater, including municipal, agricultural, industrial, hospital/pharmaceutical, stormwater and landfill lechate, generally contain a cocktail of different EDCs such as, atrazine, nonylphenol, polyethoxylates, bisphenol A, triclosan, ethinylestradiol, progesterone, testosterone and others (Benotti et al., 2009; Marty et al. 2018). Estrogenic compounds are major contributors to endocrine disrupting activity in water samples (Chen et al., 2016). Many bioassays are available for endocrine disruption. In the current project a literature search was performed on the application of bioassays for endocrine disruption in a water reuse context.

1.4.3 Neurotoxicity

Specific adverse health effects, such as effects to the nervous system (neurotoxicity), are also relevant in the context of water quality assessment, but more challenging to assess due to their complexity (Reus et al. 2020a, 2020b). Neurotoxicology is the study of adverse effects on the structure or function of the developing or mature (adult) nervous system following exposure to chemical, biological, or physical agents (Crofton et al., 2022). It is well established that neurotoxicity can be induced by a variety of neuroactive compounds including illicit drugs, pharmaceuticals, and environmental pollutants (Bellot et al., 2021). A clear understanding of the links between these substances and the way they affect the nervous system remains largely unknown. Limited testing capacity, a lack of deeper mechanistic knowledge, combined with the need to assess the hazards of hundreds of thousands of chemicals have led regulatory bodies [including the European Food Safety Authority (EFSA), the European Chemicals Agency (ECHA), and the United States Environmental Protection Agency (EPA)] to promote the development and use more mechanistically informative testing methods, including *in vitro, in silico* and alternative *in vivo* approaches (invertebrates or vertebrate embryos) to predict the developmental neurotoxicity (DNT) or adult neurotoxicity (NT) of chemicals (Paparella et al., 2020; EFSA; 2021; Crofton et al., 2022; EFSA, 2022). It has been estimated that up to 30% of all commercially used chemicals may have neurotoxic potential (Tilson et al. 1995). Additionally, in a literature study looking at the known modes of action (MoA) of organic contaminants detected in freshwater monitoring studies, neurotoxicity was identified as the MoA linked to nearly 30% of all detected chemicals (Busch et al. 2016). This shows the relevance of detecting neurotoxic compounds in the environment, increasing the demand for bioanalytical tools capable of identifying and possibly quantifying neurotoxic effects (Legradi et al. 2018). Despite the fact that neuroactive substances are found in the aquatic environment, tests to measure the effect of chemical water quality in the area of neurotoxicity are not frequently applied. The current project briefly discusses the application of neurotoxicity bioassays in a water reuse context.

1.5 Project description

The overall aim of the current project was to obtain a benchmark for water quality regarding genotoxic activity to classify current and future water sources and reuse applications based on toxicological data. The project included a literature study, in which the state of the art of the application of bioassays for genotoxic, endocrine and neurotoxic hazard and risk assessment in a water reuse context was summarized, and an experimental part, in which a set of bioassays assessing different types of DNA damage was selected and applied to different types of water samples from alternative freshwater sources. The aim of the experimental research was to develop a test battery for genotoxicity whilst considering practical considerations and sensitivity and specificity. DNA damage was selected as the primary biological effect for the experimental part because of its relevance for human health by inducing mutations and tumor formation, and the availability of well-established *in vitro* assays. In addition, suitable bioassays to assess endocrine disruption and neurotoxicity were defined for future research in the context of safeguarding water quality of different sources and use in the circular economy. As the interpretation of bioassay responses regarding the potential effects on human health is not straightforward, methods that facilitate bioassay interpretation (e.g., effect-based trigger values) were additionally evaluated.

2 Materials and methods

2.1 Collection of water samples

Samples from Dutch and Belgian water systems (rainwater, rainwater after treatment, WWTP effluent (at three different locations: A, B and C) and WWTP B effluent subjected to additional treatment) were kindly made available from existing KWR collaborations within the Water in the Circular Economy (WiCE) program and the Joint Research Programme with the water utilities (BTO). The sampled rainwater was stored in an open pond. Technically, it is ground water, but it was considered as surface water based on the higher total organic carbon (TOC) content compared to rainwater. The treatment was ultra-filtration (UF) and reverse osmosis (RO). WWTP A was located on high sandy soils in the south of the Twente region and treats wastewater from small towns and villages (about 29,000 inhabitants) by conventional biological activated sludge treatment. WWTP B was in an agricultural area near the City of Amsterdam and treats wastewater from small cities and villages (approximately 100.000 inhabitants). A pilot plant was built at this WWTP to remove micropollutants from the WWTP effluent by ozonation and active carbon filtration. The WWTP C effluent was sampled at an industrial WWTP in the province of Noord-Brabant, which contains the following processing steps: upflow anaerobic sludge blanket digestion (UASB), aerobic sludge treatment, followed by secondary clarifier and polishing pond.

A volume of 1L of the six different water samples was collected up to 14-fold in XAD rinsed bottles according to standard sample collection procedures. Samples were transported to the KWR Laboratory for Materials Research and Chemical Analysis while cooled. Samples were stored at 2-10°C until further processing for a maximum of 7 days. Samples that were not processed for the preparation of water extracts were kept as backup sample and stored at -18°C.

2.2 Preparation of water extracts

Water samples were processed for the bioassays by solid-phase extraction (SPE) using OASIS® HLB cartridges within 7 days after collection as described previously (Heringa et al. 2011) with minor modifications (Timmers et al. 2022). Evian mineral water was processed simultaneously to serve as negative control in the bioassay experiments. After the extraction procedure, 10,000-fold concentrated extracts in dimethyl sulfoxide (DMSO) were obtained. The required volume of each extract for each laboratory was aliquoted into vials. All extracts were stored at -18°C until transport to the different laboratories. Extracts were shipped overnight while cooled with ice packages and upon receipt stored at -18°C until analysis.

2.3 Bioassays

Concentrated extracts in DMSO were diluted in culture medium for the exposure in the bioassay. The maximum dilution rate varies for each of the bioassays depending on the sensitivity of the test system to DMSO. Except for the comet assay, multiple dilutions of each water extract were tested. Considering the concentration factor of the water samples (10.000x) and the dilution rate of the sample (varying from 50x up to 2000x), the relative enrichment factor (REF) was calculated. The REF was subsequently used to express the concentrations tested in each assay and to mutually compare results of different bioassays.

Six *in vitro* genotoxicity bioassays that were compatible with water samples were selected in the current project based on availability within the collaborative partners, including the Ames fluctuation test (Reifferscheid et al. 2012), umu test (Oda et al. 1985), micronucleus test (Reifferscheid et al. 2008), comet assay (Singh et al. 1988), p53 CALUX (Van der Linden et al. 2014) and ToxTracker (Hendriks et al. 2012). Harmonized protocols of the

Ames fluctuation test, micronucleus test and umu test have been embedded in ISO standards (ISO 11350:2012, ISO 21427(2):2006 and ISO 13829:2000). The umu test, p53 CALUX and ToxTracker have in research practice proven to be applicable to environmental samples and mixtures (Hendriks et al. 2019, Liu et al. 2021, Reifferscheid et al. 1991). The bioassay methods are described in more detail in Annex I and differences in characteristics of the six bioassays are summarized in Table 1.

Table 1: Differences in characteristics of the bioassays selected for genotoxicity assessment of water samples

	Ames fluctuation test	Comet assay	Micronucleus test	Umu test	p53 CALUX	ToxTracker
Endpoint	Mutagenicity	Chromosome damage		DNA damage response		
Mode of action	Frameshift mutations, base-pair substitutions	Single and double strand DNA breaks, alkali labile sites	Clastogenicity (chromosome breaks), aneugenicity (chromosome loss)	SOS response	p53 induction	ATR-Chk1, NF-kB, p53, Nrf2, HMOX1 signaling pathways, protein damage
Test system	*S. typhimurium* TA98+TA100	HepG2 cells	HepG2 cells	*S. typhimurium* TA1535	U2OS cells	mES cells
Wildtype or modified?	Modified	Wildtype	Wildtype	Modified	Modified	Modified
Test system type	Bacteria	Human liver tumour cell line	Human liver tumour cell line	Bacteria	Human bone tumour cell line	Mammalian (mouse) embryonic stem cells
Exposure time	90 min	3h, 24h	72h	4h in total	24h	24h
Xenobiotic metabolism (S9) included?	Yes, performed +/-S9	No, performed -S9 only	No, performed -S9 only	Yes, performed +/-S9	Yes, performed +/-S9	Yes, performed +/-S9
Maximum tested concentration (REF)	200	100	200	200	100	100
Read-out genotoxicity	Number of revertant wells	%tail DNA	Micronuclei formation	β-galactosidase activity	Luminescence	Fluorescence
Direct or indirect measure of DNA damage?	Direct	Indirect (reversible)	Direct	Indirect (induction of repair mechanism)	Indirect (induction of repair mechanism)	Indirect (induction of cellular path-ways involved in DNA repair, oxidative stress and protein damage)
Read-out cytotoxicity	Bacterial growth	Colour conversion	Cell proliferation	Bacterial growth	Cell viability	Cell viability
Cytotoxicity cut-off	>50%	>50%	>50%	Growth factor <0.5	>20%	>25%
Available guidelines for water quality assessment	ISO 11350:2012	No	ISO 21427-2:2006[2]	ISO 13829:2000	No	No
Available guidelines for regulatory testing	OECD 471 (classical Ames test)	OECD 489 (*in vivo* comet assay)[1]	OECD 487 (various cell types)[2]	No	No	No

[1] No *in vivo* OECD guideline available
[2] HepG2 cells are not recommended either by ISO 21427-2:2006 or OECD 487, but are frequently used for water quality assessment

2.4 Literature study

2.4.1 Genotoxicity and endocrine disruption

A literature study was performed to summarize the current developments on bioassay test batteries for water quality assessment and to generate an overview of bioassays that were applied for assessment of genotoxicity and endocrine effects in a water reuse context.

For the latter, the following search term was used in PubMed (https://pubmed.ncbi.nlm.nih.gov/) using the following search term:

(((genotoxic[Title/Abstract] OR genotoxicity[Title/Abstract]) OR endocrine disrupting[Title/Abstract] OR endocrine disruptors[Title/Abstract])) AND (Effect-based[Title/Abstract] OR in vitro[Title/Abstract] OR assay[Title/Abstract]) AND (Water reuse[Title/Abstract] OR reclaimed water[Title/Abstract] OR recycled water[Title/Abstract] OR wastewater effluent[Title/Abstract] OR water recovery[Title/Abstract] OR water recycling[Title/Abstract] OR water treatment[Title/Abstract] OR treated wastewater[Title/Abstract] OR recycling wastewater[Title/Abstract]) AND (2010:2021[pdat])

On March 15, 2021 this resulted in 82 articles, of which 49 were considered relevant for an overview of applied bioassays for assessment of genotoxicity and endocrine effects in a water reuse context. Suitable articles were selected based on the abstract and duplicates were removed.

In addition, the following search team was applied in Scopus (https://www.scopus.com):

TITLE-ABS-KEY(''Genotoxic'') OR TITLE-ABS-KEY(''genotoxicity'') OR TITLE-ABS-KEY(''endocrine disrupting'') OR TITLE-ABS-KEY(''endocrine disruptors'') AND TITLE-ABS-KEY(''effect-based'') OR TITLE-ABS-KEY(''in vitro assay'') OR TITLE-ABS-KEY(''in vitro bioassay'') AND TITLE-ABS-KEY(''water reuse'') OR TITLE-ABS-KEY(''reclaimed water'') OR TITLE-ABS-KEY(''recycled water'') OR TITLE-ABS-KEY(''wastewater effluent'') OR TITLE-ABS-KEY(''water recovery'') OR TITLE-ABS-KEY(''water recycling'') OR TITLE-ABS-KEY(''water treatment'') OR TITLE-ABS-KEY(''treated wastewater'') OR TITLE-ABS-KEY(''recycling wastewater'') AND PUBYEAR > 2009

On March 15, 2021, in total 262 articles were found, of which finally 78 were considered relevant for an overview of applied bioassays for assessment of genotoxicity and endocrine effects in a water reuse context. Suitable articles were selected based on the abstract or content and duplicates (also from PubMed) were removed, finally resulting in 45 relevant articles for the overview of applied bioassays for assessment of genotoxicity and endocrine effects in a water reuse context. Eventually, one article was removed from the selection, after careful reading it did not contain relevant information for the literature research.

On January 6, 2023, the following search term was applied in PubMed to add articles related to the application of bioassays on hospital wastewater, stormwater, rainwater and grey water.

(((genotoxic[Title/Abstract] OR genotoxicity[Title/Abstract]) OR endocrine disrupting[Title/Abstract] OR endocrine disruptors[Title/Abstract])) AND (Effect-based[Title/Abstract] OR in vitro[Title/Abstract] OR assay[Title/Abstract]) AND (hospital wastewater[Title/Abstract] OR stormwater[Title/Abstract] OR rainwater[Title/Abstract] OR grey water[Title/Abstract]) AND (2010:2022[pdat])

This resulted in the addition of 7 articles and 52 articles in total for the overview of applied bioassays for assessment of genotoxicity and endocrine effects in a water reuse context. All articles were summarized and listed by water type studied in Table II. The results of the literature search were used to support the discussion of the results of the current project.

2.4.2 Neurotoxicity

An additional literature study was performed to explore the application of bioassays for neuroactive substances in a water reuse context using the following search term in PubMed:

((neurotoxic[Title/Abstract] OR neuroactive[Title/Abstract])) AND ((Effect-based[Title/Abstract] OR in vitro[Title/Abstract] OR assay[Title/Abstract])) AND ((water reuse[Title/Abstract] OR recycled water[Title/Abstract] OR reclaimed water[Title/Abstract]))

On February 1, 2023 this resulted in 3 relevant articles. Simplification of the search term resulted in a higher number of articles, but did not gain more relevant articles. A fourth relevant paper was found while reading one of the papers that were identified as relevant. The articles are discussed in section 4.1.3.

3 Experimental results

Results of the six water samples obtained with the bioassays are presented below. Cytotoxicity measurements are presented first, as these data aid in the interpretation of genotoxicity responses by excluding concentrations of the water samples inducing severe cytotoxicity from a conclusion on genotoxicity. In all figures, concentrations of the water samples are expressed as relative enrichment factor (REF) and each bar colour represents a REF.

3.1 Ames fluctuation test

According to the ISO standard 11350:2012, cytotoxicity was only determined for *S. typhimurium* strain TA98, both in the presence and absence of S9 for metabolic activation by measuring bacterial density. Positive controls and negative controls showed the expected responses both in the presence and absence of S9 (data not shown) and acceptance criteria for a valid test were met. In the absence of S9, severe cytotoxicity (>50%) was observed for WWTP A effluent and WWTP C effluent at the maximum REF of 200 (Figure 1A). In the presence of S9, severe cytotoxicity was observed for WWTP A effluent, WWTP B effluent, WWTP B effluent after treatment and WWTP C effluent, all again at the maximum REF of 200 (Figure 1B). Overall, cytotoxicity showed a dose-related trend. Results of severe cytotoxic concentrations were not taken into consideration for mutagenicity evaluation to exclude false negative responses.

No mutagenic response was observed for rainwater, treated rainwater, WWTP A effluent, WWTP B effluent and treated WWTP B effluent, in any of the bacterial strains tested both in the absence and presence of S9 up to the maximum REF of 200. WWTP C effluent showed a mutagenic response in TA100 in the absence of S9 from above a REF of 50, but not in the presence of S9, and not in TA98 both in the presence and absence of S9 up to the maximum REF of 200 (Figures 2A-2D). These results indicate that the enzymes present in S9 may be capable of detoxifying potential mutagenic substances present in WWTP C effluent. This is however in contrast with the cytotoxicity results, where overall the responses were higher in the presence of S9. The difference in response between cytotoxicity and mutagenicity can be related to different micropollutants in the water sample extract with different modes of action.

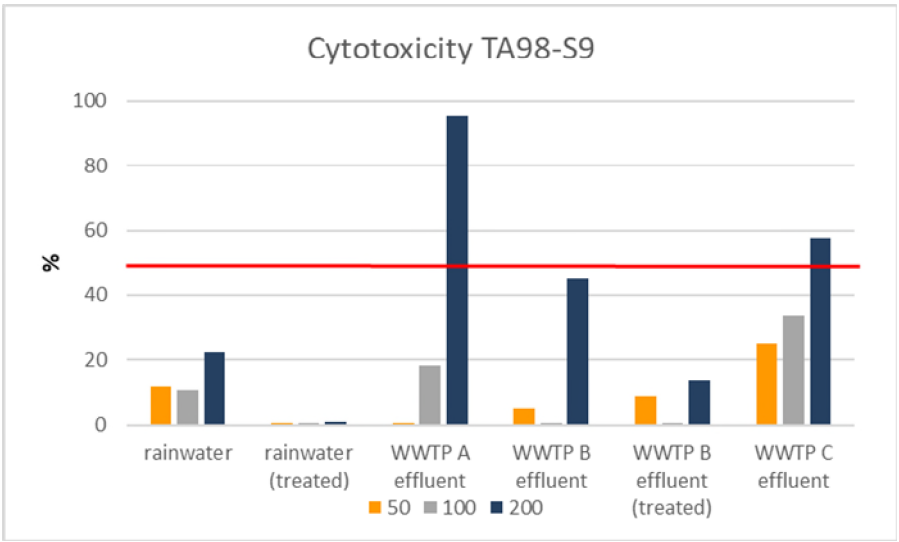

Figure 1A: Percentage cytotoxicity caused by water samples as compared to the Evian mineral water negative control in the Ames fluctuation test with TA98 in the absence of S9 after 90 min exposure. The red line indicates the cut-off of 50% for severe cytotoxicity. Concentrations of the water samples during exposure are expressed as REF.

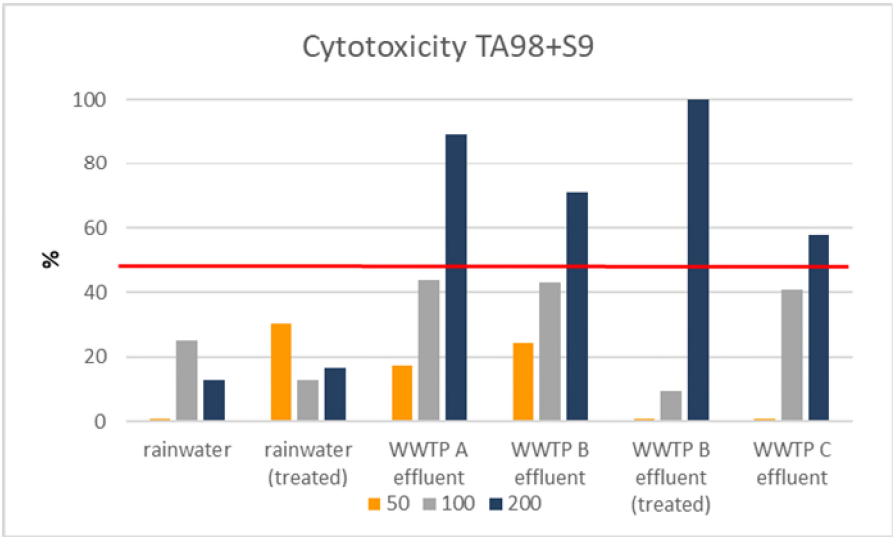

Figure 1B: Percentage cytotoxicity caused by water samples as compared to the Evian mineral water negative control in the Ames fluctuation test with TA98 in the presence of S9 after 90 min exposure. The red line indicates the cut-off of 50% for severe cytotoxicity. Concentrations of the water samples during exposure are expressed as REF.

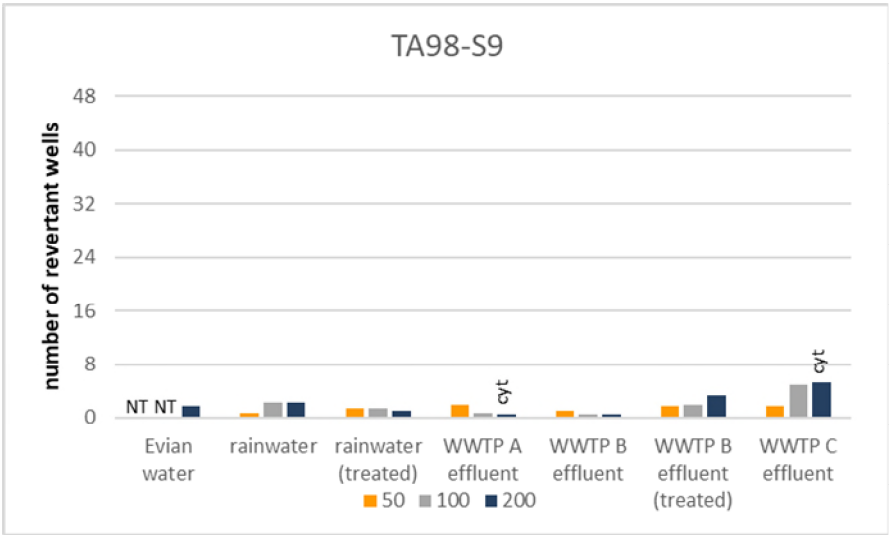

Figure 2A: Number of revertant wells induced by water samples in the Ames fluctuation test with TA98 in the absence of S9 after 90 min exposure. NT indicates that a concentration is not tested, cyt indicates that the concentration showed severe cytotoxicity (see Figure 1A). Concentrations of the water samples during exposure are expressed as REF. None of the water samples showed a positive response for mutagenicity.

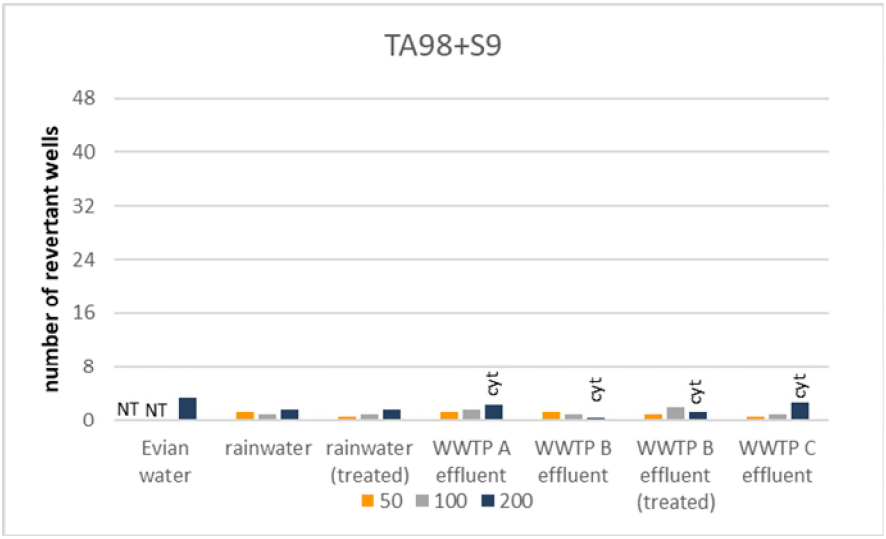

Figure 2B: Number of revertant wells induced by water samples in the Ames fluctuation test with TA98 in the presence of S9 after 90 min exposure. NT indicates that a concentration is not tested, cyt indicates that the concentration showed severe cytotoxicity (see Figure 1B). Concentrations of the water samples during exposure are expressed as REF. None of the water samples showed a positive response for mutagenicity.

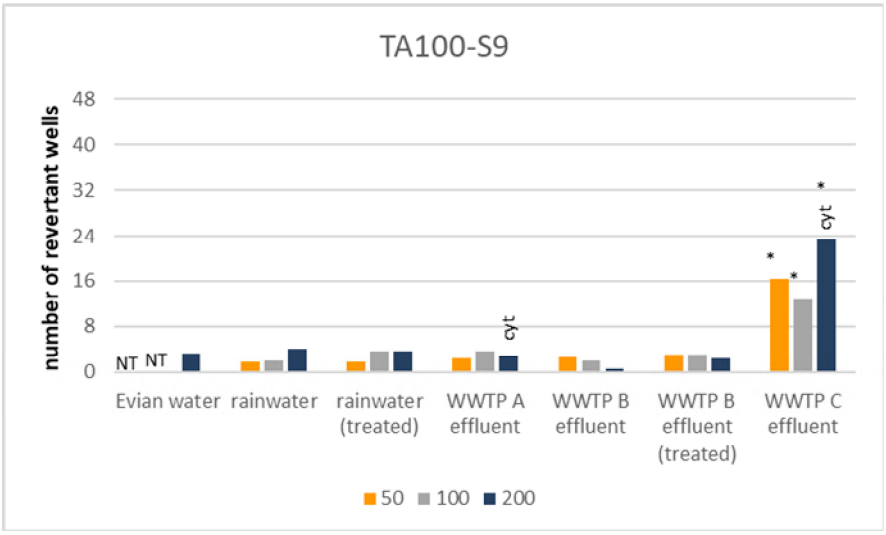

*Figure 2C: Number of revertant wells induced by water samples in the Ames fluctuation test with TA100 in the absence of S9 after 90 min exposure. NT indicates that a concentration is not tested, cyt indicates that the concentration showed severe cytotoxicity in TA98 (see Figure 1A) and * indicates a positive response for mutagenicity (based on statistical significance). Concentrations of the water samples during exposure are expressed as REF.*

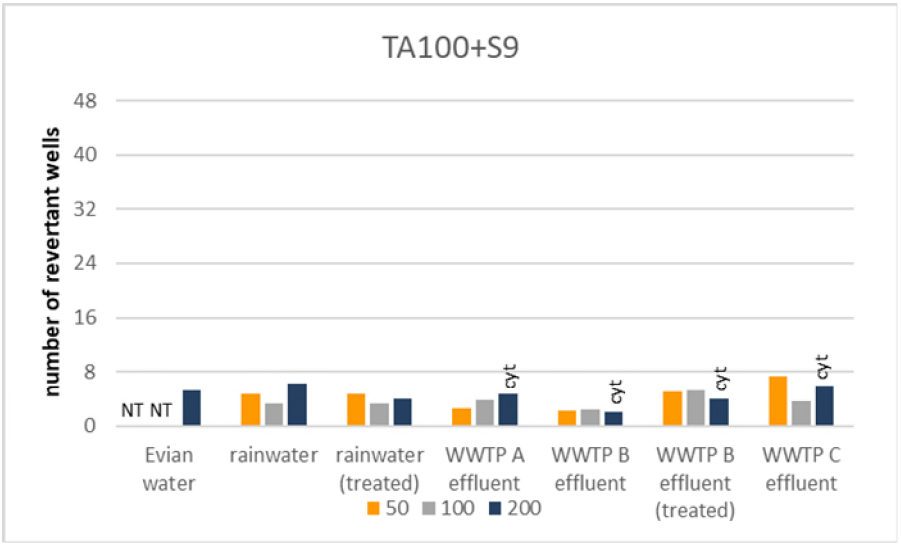

Figure 2D: Number of revertant wells induced by water samples in the Ames fluctuation test with TA100 in the presence of S9 after 90 min exposure. NT indicates that a concentration is not tested, cyt indicates that the concentration showed severe cytotoxicity in TA98 (see Figure 1B). Concentrations of the water samples during exposure are expressed as REF. None of the water samples showed a positive response for mutagenicity.

3.2 Umu test

In the absence of S9, positive controls and negative controls showed the expected responses (data not shown). In the presence of S9, the positive control did not show the expected response (data not shown). Since the positive control in the presence of S9 showed a positive response for genotoxicity in an umu test that was performed simultaneously and the negative control showed the expected response (data not shown), the results of the current project were considered reliable. No cytotoxicity was observed in the umu test for any of the water samples at any of the concentrations tested up to the maximum REF of 200 during exposure, as reflected by growth factors > 0.5 both in the absence and presence of S9 (data not shown). In the absence of S9, WWTP A effluent, WWTP B effluent and WWTP C effluent exceeded the IR of 1.5 and were therefore considered genotoxic from a REF of 50 (WWTP B effluent) or 100 (WWTP A effluent and WWTP C effluent) during exposure. Rainwater, rainwater after treatment and treated WWTP B effluent did not show a genotoxic response in the absence of S9 (Figure 3A). The absence of a genotoxic response in treated WWTP B effluent indicates that the treatment is effective in removing substances that can induce the SOS response in bacterial mutagenesis. In the presence of S9, a genotoxic response was observed for WWTP C effluent only, from a REF of 100 during exposure. Rainwater, treated rainwater, WWTP A effluent, WWTP B effluent and treated WWTP B effluent did not show a genotoxic response in the umu test in the presence of S9 (Figure 3B).

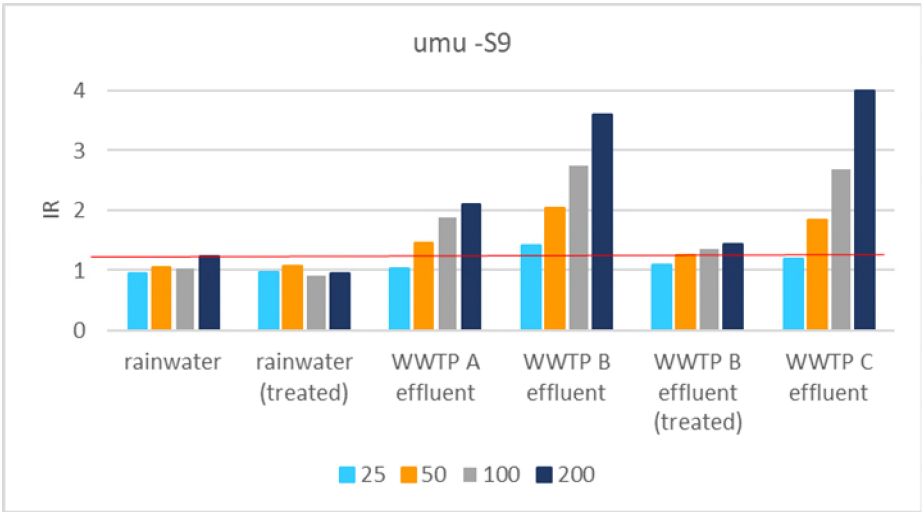

Figure 3A: Responses of the different water samples expressed as induction factor in the umu test in the absence of S9 after in total 4h exposure. The red line indicates the criterion of induction ratio (IR) ≥ 1.5 for a positive response. Concentrations of the water samples are expressed as REF.

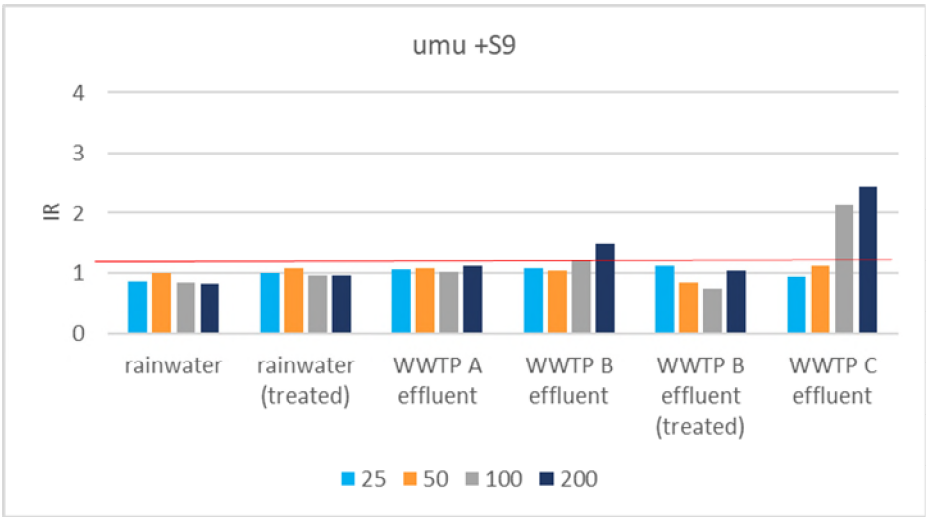

Figure 3B: Responses of the different water samples expressed as induction factor in the umu test in the presence of S9 after in total 4h exposure. The red line indicates the criterion of induction ratio (IR) ≥ 1.5 for a positive response. Concentrations of the water samples are expressed as REF.

3.3 Micronucleus test

Three independent micronucleus experiments were performed to obtain robust results for each water sample at non-cytotoxic concentrations. Positive controls and negative controls showed the expected responses (data not shown) and acceptance criteria for a valid test were met in all cases. Cytotoxicity was determined based on cell proliferation and was observed for all water samples, except for treated rainwater up to the maximum REF of 200. Rainwater was cytotoxic from a REF of 50 during exposure, WWTP A effluent was cytotoxic from a REF of 12.5, WWTP B effluent was cytotoxic from a REF of 33.3, treated WWTP B effluent was cytotoxic from a REF of 28.5 and WWTP C effluent was cytotoxic from a REF of 16.7. Overall, cytotoxicity showed a dose-related trend (Figures 4A-4C). The treatment process applied to rainwater seems to be effective in removing substances that are cytotoxic to HepG2 cells after 72 hours. However, the treatment process of WWTP B effluent did not seem to be effective to remove cytotoxic activity in this cell model. From the bioassay results it cannot be distinguished if cytotoxic substances are not removed, or that they are removed, but also being formed during the treatment process. Results of severe cytotoxic concentrations were not taken into consideration for genotoxicity evaluation to exclude false positive responses.

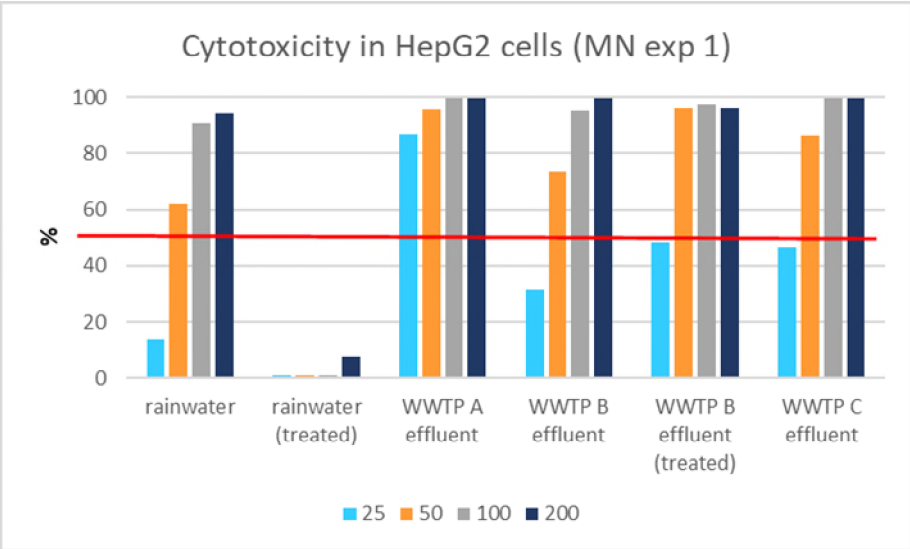

Figure 4A: Percentage cytotoxicity caused by water samples as compared to the concurrent Evian mineral water negative control in the micronucleus test with HepG2 cells after 72h exposure (first experiment). The red line indicates the cut-off of 50% for severe cytotoxicity. Concentrations of the water samples during exposure are expressed as REF.

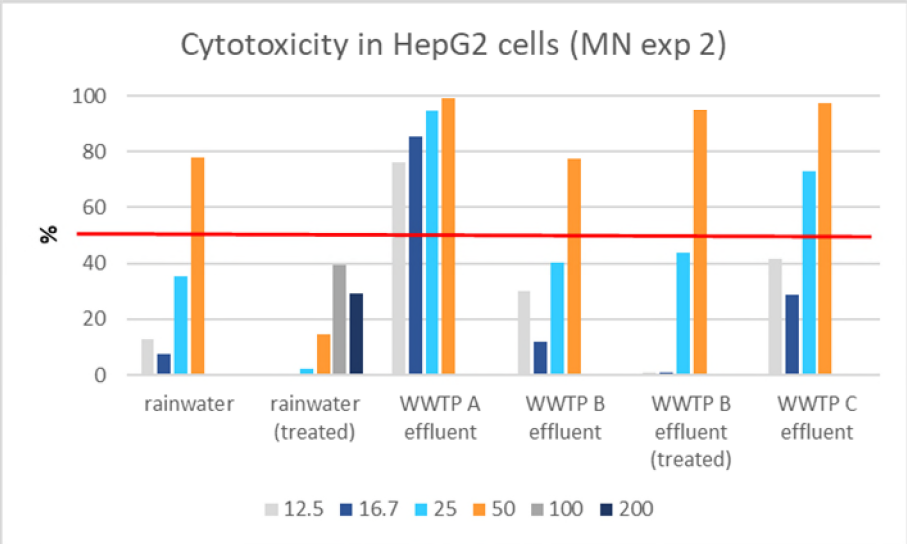

Figure 4B: Percentage cytotoxicity caused by water samples as compared to the concurrent Evian mineral water negative control in the micronucleus test with HepG2 cells after 72h exposure (second experiment). The red line indicates the cut-off of 50% for severe cytotoxicity. Concentrations of the water samples during exposure are expressed as REF.

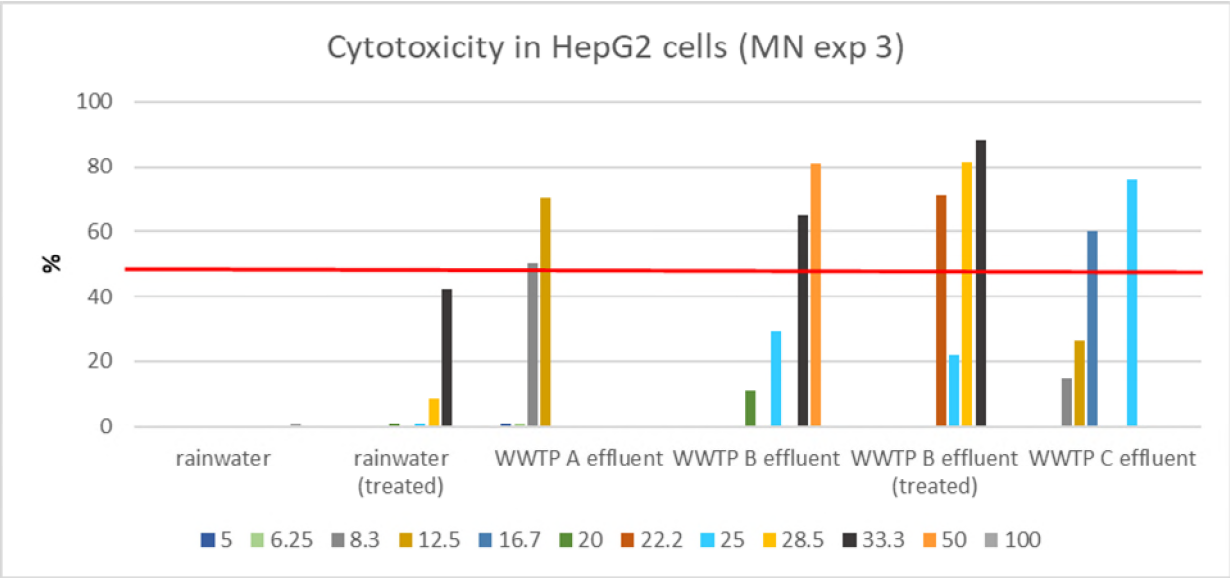

Figure 4C: Percentage cytotoxicity caused by water samples as compared to the concurrent Evian mineral water negative control in the micronucleus test with HepG2 cells after 72h exposure (third experiment). The red line indicates the cut-off of 50% for severe cytotoxicity. Concentrations of the water samples during exposure are expressed as REF.

None of the water samples showed a genotoxic response in the micronucleus test with HepG2 cells at non-cytotoxic concentrations, except for rainwater which showed negative as well as positive results at non-cytotoxic concentrations, and was therefore considered equivocal (Figures 5A-5C).

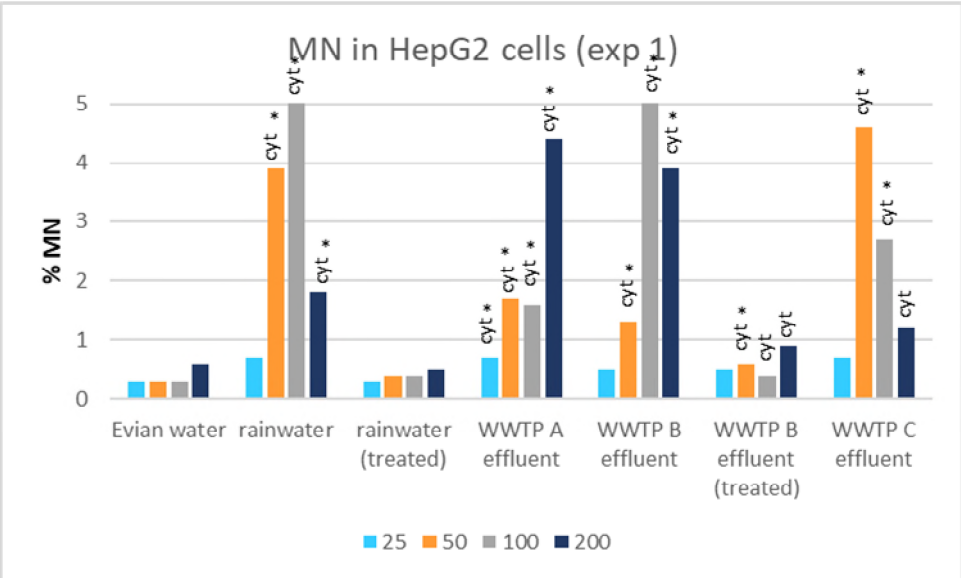

*Figure 5A: Percentage micronuclei induced by water samples as compared to the concurrent Evian mineral water negative control in the micronucleus test with HepG2 cells after 72h exposure (first experiment). Cyt indicates cytotoxicity, * indicates a genotoxic response. Results of severe cytotoxic concentrations were not taken into consideration for genotoxicity evaluation to exclude false positive responses Concentrations of the water samples during exposure are expressed as REF.*

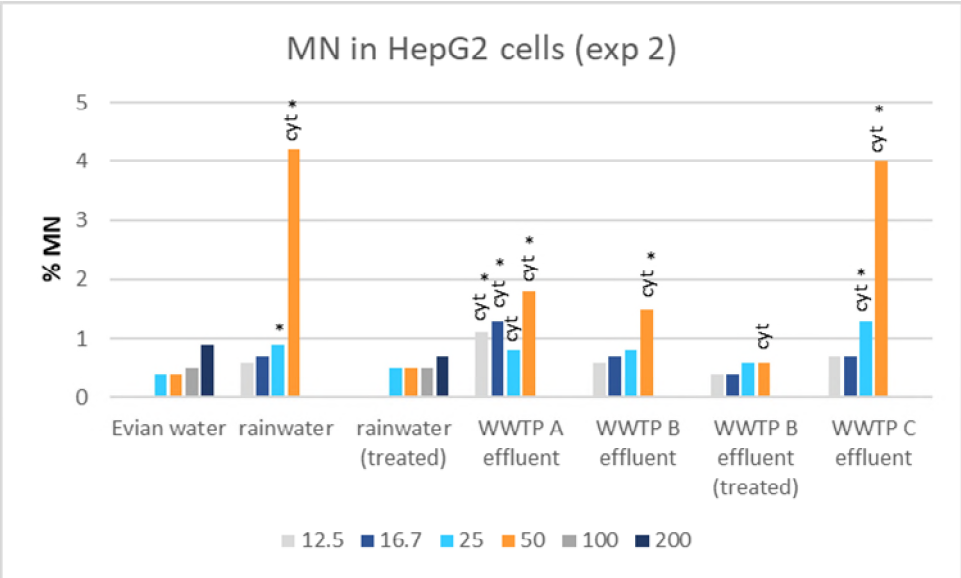

*Figure 5B: Percentage micronuclei induced by water samples as compared to the concurrent Evian mineral water negative control in the micronucleus test with HepG2 cells after 72h exposure (second experiment). Cyt indicates cytotoxicity, * indicates a genotoxic response. Results of severe cytotoxic concentrations were not taken into consideration for genotoxicity evaluation to exclude false positive responses Concentrations of the water samples are expressed as REF.*

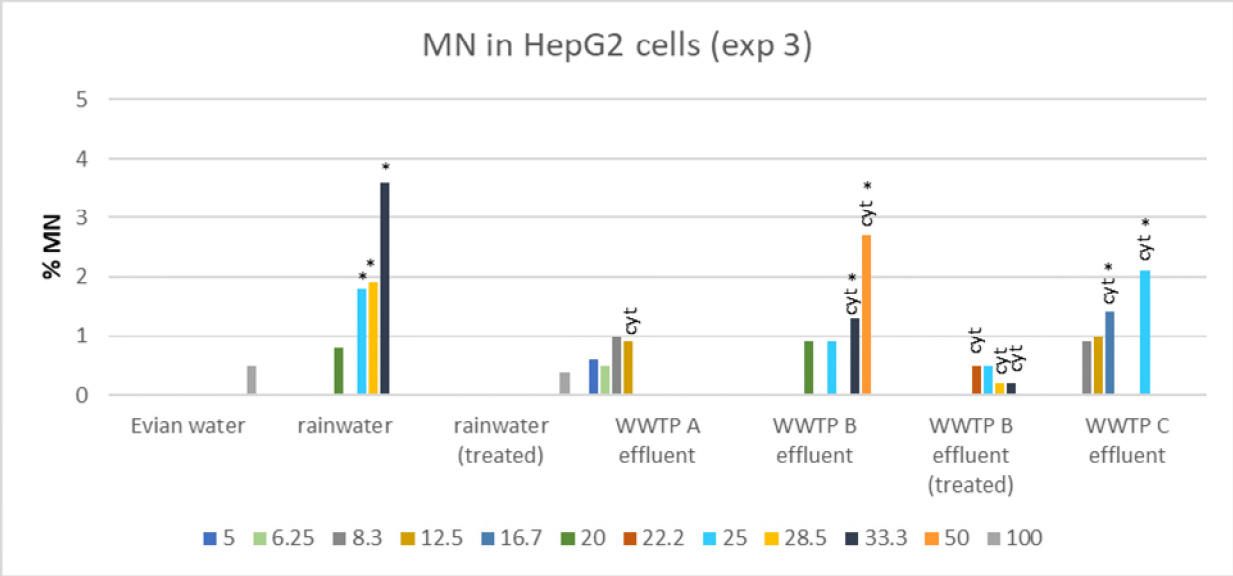

*Figure 5C: Percentage micronuclei induced by water samples as compared to the concurrent Evian mineral water negative control in the micronucleus test with HepG2 cells after 72h exposure (third experiment). Cyt indicates cytotoxicity, * indicates a genotoxic response. Results of severe cytotoxic concentrations were not taken into consideration for genotoxicity evaluation to exclude false positive responses. Concentrations of the water samples are expressed as REF.*

3.4 Comet assay

For practical reasons, in the comet assay a single concentration of the water samples of a REF of 100 was tested in HepG2 cells using a 3h and 24h exposure period. Positive controls and negative controls showed the expected responses (data not shown) and acceptance criteria for a valid test were met in both cases. Severe cytotoxicity was observed after 24h exposure for rainwater, WWTP A effluent, WWTP B effluent and WWTP C effluent, whereas rainwater after treatment and WWTP B effluent after treatment did not show severe cytotoxicity. None of the water samples caused severe cytotoxicity after 3h exposure (Figure 6). From these bioassay results it cannot be distinguished if the more pronounced results after 24h exposure are due to metabolization of substances into more cytotoxic ones (assuming metabolic competence of HepG2 cells) or related to a longer exposure period. The lower cytotoxicity observed for treated rainwater and treated WWTP B effluent indicates that the treatment processes used are effective in removal of substances that are cytotoxic to HepG2 cells. Severe cytotoxic concentrations were not taken into consideration for genotoxicity evaluation to exclude false positive responses.

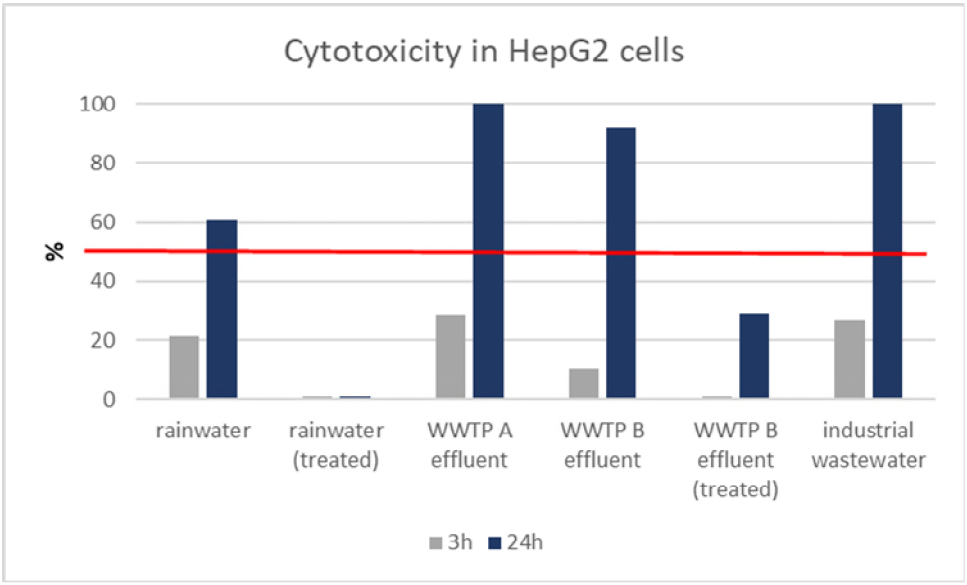

Figure 6: Percentage cytotoxicity caused by water samples as compared to the Evian mineral water negative control in the comet assay with HepG2 cells after 3h and 24h exposure. The red line indicates the cut-off of 50% for severe cytotoxicity. Samples were tested at a concentration REF of 100.

A genotoxic response was observed for WWTP A effluent after 3h exposure at the maximum REF of 100. There was no reason to repeat the experiment with lower concentrations of WWTP A effluent after 24h exposure, because the response at lower concentrations will not change the conclusion on genotoxicity. Rainwater, treated rainwater, WWTP B effluent, treated WWTP effluent and WWTP C effluent did not show a genotoxic response, neither after 3h nor after 24h exposure, up to the maximum REF of 100 (Figure 7). Since the severe cytotoxic response of WWTP B effluent and WWTP C effluent was not associated with a higher %tail DNA or higher number of ghost cells, it was considered that repeating the 24h exposure with lower concentrations of WWTP B effluent and WWTP C effluent was not needed to confirm the absence of genotoxicity.

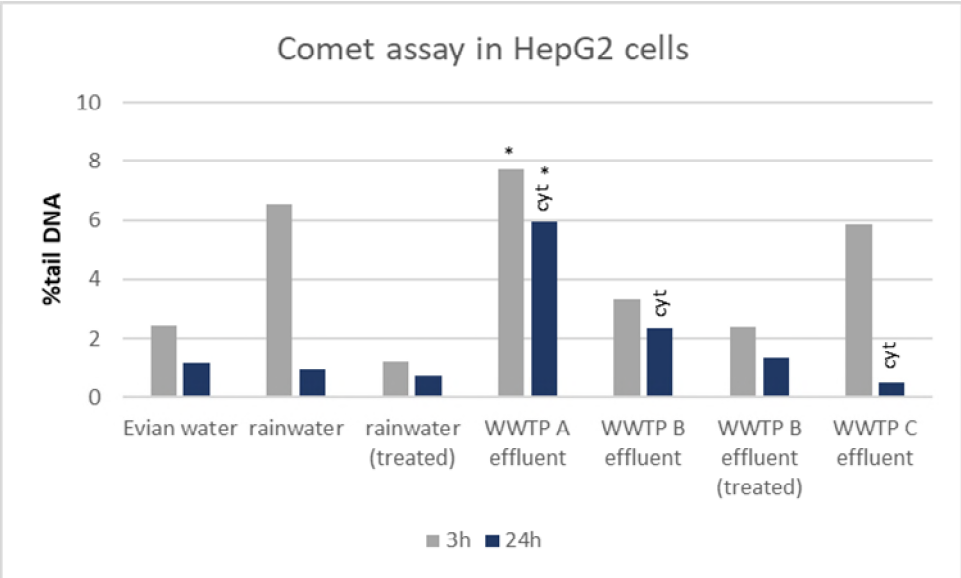

*Figure 7: Percentage tail DNA induced by water samples in the comet assay with HepG2 cells after 3h and 24h exposure. Cyt indicates cytotoxicity, * indicates a genotoxic response based on the cut-off of a 3-fold increase over the Evian mineral water negative control. Samples were tested at a concentration REF of 100.*

3.5 p53 CALUX

Positive controls and negative controls showed the expected responses both in the absence and presence of S9 (data not shown) and acceptance criteria for a valid test were met. Cytotoxicity was observed in the cytotox CALUX for rainwater and treated WWTP B effluent at the maximum REF of 100, for WWTP effluent A from a REF of 3 and for WWTP B effluent from a REF of 33 (Figure 8).

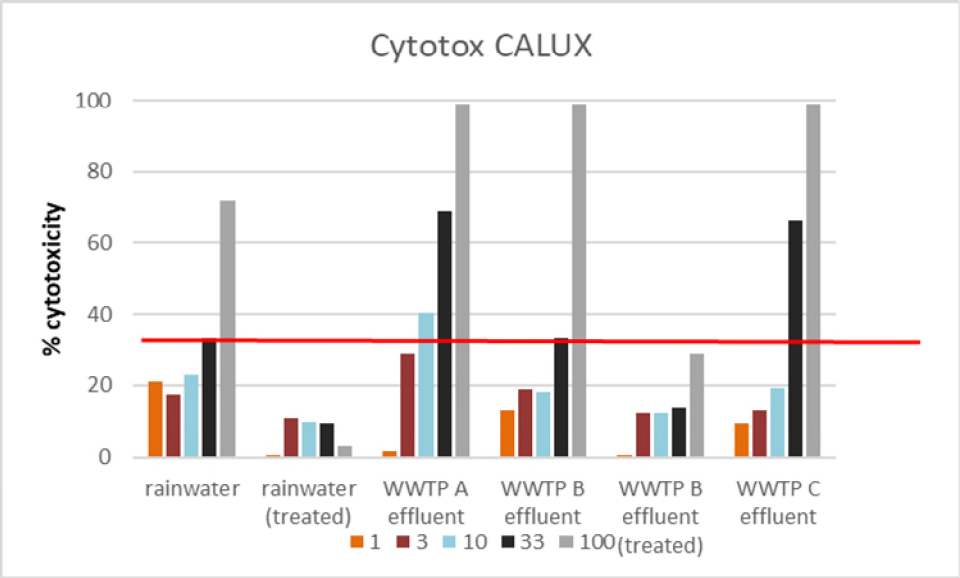

Figure 8: Percentage cytotoxicity caused by water samples as compared to the concurrent Evian mineral water negative control in the Cytotox CALUX after 24h exposure. The red line indicates the cut-off of 20% for severe cytotoxicity. Concentrations of the water samples during exposure are expressed as REF.

None of the water samples showed a response above the detection limit in the p53 CALUX and hence, no genotoxicity was observed for any of the water samples tested up to the maximum REF of 100, both in the presence and absence of S9.

3.6 ToxTracker

Cytotoxicity in the ToxTracker assay was determined based on cell survival. In the absence of S9, WWTP A effluent was found to be severely cytotoxic from a REF of 12.5 during exposure, WWTP effluent B was found to be severely cytotoxic at the maximum REF of 100 and WWTP C effluent was found to be severely cytotoxic from a REF of 50 (Figure 9A). In the presence of S9, severe cytotoxicity was observed for WWTP A effluent from a REF of 50, for WWTP B effluent at the maximum REF of 100 and for WWTP C effluent from a REF of 12.5 (Figure 9B). Severe cytotoxic concentrations were not taken into consideration for genotoxicity evaluation to exclude false positive responses.

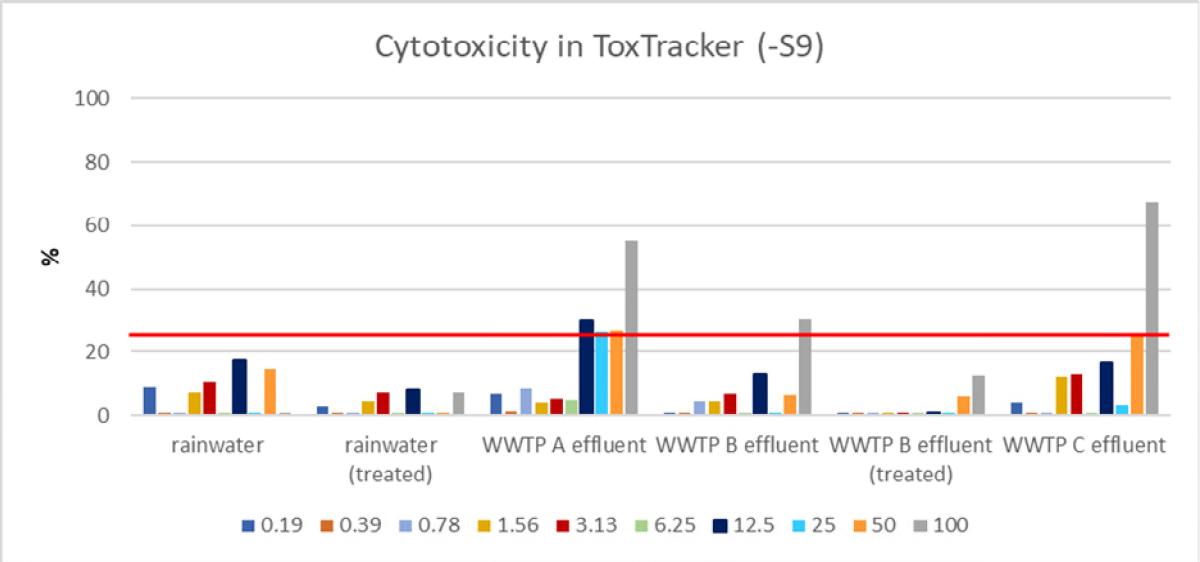

Figure 9A: Percentage cytotoxicity caused by water samples as compared to the concurrent Evian mineral water negative control in the ToxTracker assay in the absence of S9 after 24h exposure. The red line indicates the cut-off of 25% for severe cytotoxicity. Concentrations of the water samples during exposure are expressed as REF.

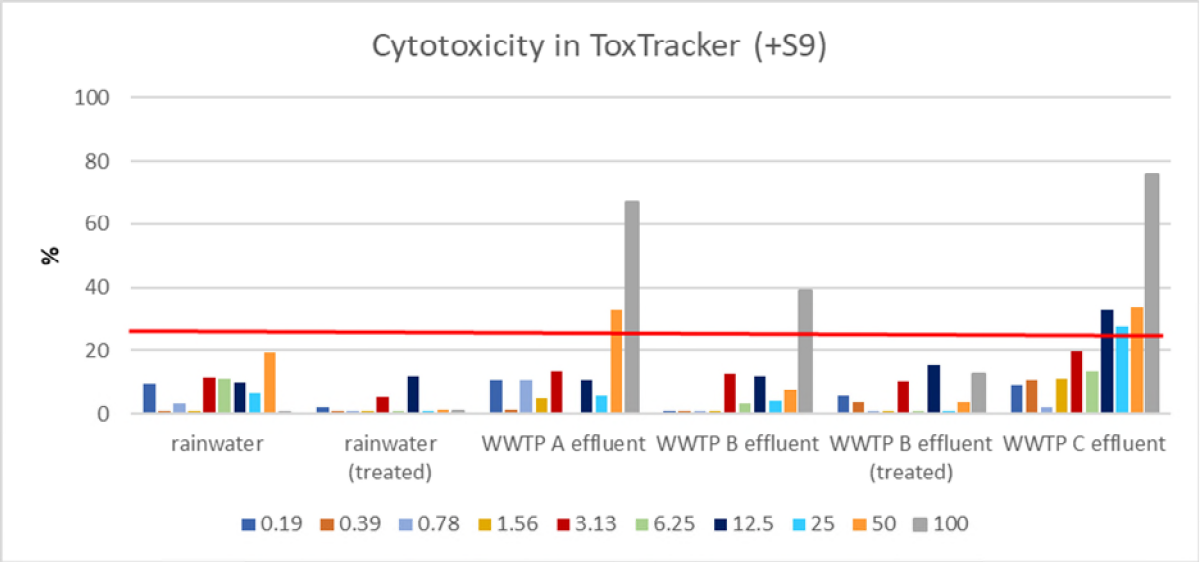

Figure 9B: Percentage cytotoxicity caused by water samples as compared to the concurrent Evian mineral water negative control in the ToxTracker assay in the presence of S9 after 24h exposure. The red line indicates the cut-off of 25% for severe cytotoxicity. Concentrations of the water samples during exposure are expressed as REF.

None of the water samples showed a genotoxic response in the ToxTracker assay at non-cytotoxic concentrations (Figure 10).

	Evian mineral water	Rainwater	Rainwater after treatment	WWTP A effluent
Classification	Non-genotoxic	Non-genotoxic	Non-genotoxic	Non-genotoxic
Direct DNA reactive genotoxin	□□□□	□□□□	□□□□	□□□□
Indirect genotoxin: Oxidative stress	□□□□	□□□□	□□□□	□□□□
Indirect genotoxin: Aneugen	□□□□	□□□□	□□□□	□□□□
Oxidative stress	□□□□	■■□□	□□□□	■□□□
Protein stress	□□□□	□□□□	□□□□	■■□□
S9 metabolism	No	No	No	No

	WWTP B effluent	WWTP B effluent after treatment	WWTP C effluent
Classification	Non-genotoxic	Non-genotoxic	Non-genotoxic
Direct DNA reactive genotoxin	□□□□	□□□□	□□□□
Indirect genotoxin: Oxidative stress	□□□□	□□□□	□□□□
Indirect genotoxin: Aneugen	□□□□	□□□□	□□□□
Oxidative stress	■□□□	■□□□	■■□□
Protein stress	■■□□	■□□□	■■□□
S9 metabolism	No	No	No

Figure 10: Summary of ToxTracker assay results after 24h exposure of mES cells to the water samples. For the classification 'Direct DNA reactive genotoxin', 'Indirect genotoxin: Oxidative stress' and 'Indirect genotoxin: Aneugen', the number of coloured squares indicates the likelihood of a certain mode of action, based on the pattern of reporter activation. For 'Oxidative stress' and 'Protein stress', the number of coloured squares indicates the strength of the reporter induction in such a way that 1 square is low level but above 2, 2 or 3 squares are more activation, and 4 squares is strong activation comparable to the positive control (ToxTracker Test Report, R. Derr and I. Brandsma, 2022 and accompanying correspondence)[3].

[3] In the report, WWTP C effluent was considered equivocal for genotoxicity. Based on updated criteria for classification of test substances based on a OECD validation study, the responses led to a non-genotoxic conclusion for WWTP C effluent.

3.7 Integrated results of the six bioassays

WWTP A effluent, WWTP B effluent and WWTP C effluent showed a genotoxic response in at least one bioassay, which can be explained by the higher levels of chemical micropollutants in these samples compared to, for example, rainwater. WWTP C effluent showed two positive responses for genotoxicity (Ames fluctuation test and umu test, at an REF of 100 and 50, respectively). WWTP A effluent also showed two positive responses (comet assay and umu test, at an REF ≤ 100), and WWTP B effluent showed a positive response for genotoxicity in the umu test only (at an REF of 50). Rainwater showed an equivocal response in the micronucleus test. Treated rainwater and treated WWTP effluent B did not show a genotoxic response in any of the bioassays performed (Table 2). Interpretation of the genotoxicity results is further discussed in sections 5.1 and 5.3.

All water samples, except for rainwater after treatment, showed severe cytotoxicity in three or more bioassays at varying concentrations (expressed as REF). Rainwater was found to be severely cytotoxic in three bioassays (comet assay, micronucleus test and p53 CALUX), treated WWTP B effluent showed severe cytotoxicity also in three bioassays (Ames fluctuation test, micronucleus test and p52 CALUX), and WWTP A effluent, WWTP B effluent and WWTP C effluent were found to be severely cytotoxic in five bioassays (Ames fluctuation test, comet assay, micronucleus test, p53 CALUX and ToxTracker). Concentrations where positive responses were obtained for the different water samples varied between the assays (Table 3). The absence of a severe cytotoxic response for treated rainwater can be explained by an absence or very low presence of chemical micropollutants in the sample. Interpretation of the cytotoxicity results is further discussed in section 5.1.

Table 2: Summary of genotoxicity responses of the six bioassays applied to six water samples from various water cycles. Genotoxicity responses shown in the table occurred at non-cytotoxic concentrations. Genotoxicity responses were regarded as positive if at least one condition (with or without S9 for exogenous metabolic activation) was found to be positive based on the concurrent assay-specific criteria. The lowest concentration during exposure expressing a genotoxic response is indicated between brackets (expressed as REF). Red: positive response, orange: equivocal response, green: negative response for genotoxicity.

	Genotoxicity					
	Mutagenicity		Chromosome damage	DNA damage response		
	Ames fluctuation test	Comet assay	Micronucleus test	umu test	p53 CALUX	Tox Tracker
Rainwater	Negative	Negative	Equivocal (25)	Negative	Negative	Negative
Rainwater (after treatment)	Negative	Negative	Negative	Negative	Negative	Negative
WWTP A effluent	Negative	Positive (≤100)	Negative	Positive (100)	Negative	Negative
WWTP B effluent	Negative	Negative	Negative	Positive (50)	Negative	Negative
WWTP B effluent (after additional treatment)	Negative	Negative	Negative	Negative	Negative	Negative
WWTP C effluent	Positive (100)	Negative	Negative	Positive (50)	Negative	Negative

Table 3: Summary of cytotoxicity responses of the six bioassays applied to six water samples from various water cycles. The lowest concentration during exposure that caused severe cytotoxicity is indicated between brackets (expressed as REF). Red: (severe) cytotoxic, green: non-cytotoxic.

	Cytotoxicity					
	Ames fluctuation test	Comet assay	Micronucleus test	umu test	p53 CALUX	ToxTracker
	Bacterial growth	Colour conversion	Cell proliferation	Bacterial growth	Cell viability	Cell viability
Rainwater	Not cytotoxic	Cytotoxic (≤100)	Cytotoxic (33)	Not cytotoxic	Cytotoxic (10)	Not cytotoxic
Rainwater (after treatment)	Not cytotoxic	Not cytotoxic	Not cytotoxic	Not cytotoxic	Not cytotoxic	Not cytotoxic
WWTP A effluent	Cytotoxic (200)	Cytotoxic (≤100)	Cytotoxic (8.3)	Not cytotoxic	Cytotoxic (3)	Cytotoxic (25)
WWTP B effluent	Cytotoxic (200)	Cytotoxic (≤100)	Cytotoxic (33)	Not cytotoxic	Cytotoxic (33)	Cytotoxic (50)
WWTP B effluent (after additional treatment)	Cytotoxic (200)	Not cytotoxic	Cytotoxic (28.5)	Not cytotoxic	Cytotoxic (100)	Not cytotoxic
WWTP C effluent	Cytotoxic (200)	Cytotoxic (≤100)	Cytotoxic (16.7)	Not cytotoxic	Cytotoxic (33)	Cytotoxic (100)

4 Results of the literature study

4.1 Literature search output

From the output produced by the literature search, 51 papers were identified as relevant for the state-of-the-art overview on the application of *in vitro* bioassays to investigate genotoxicity and endocrine disruption in a water reuse context. Neurotoxicity was not part of this overviews since the use of bioassays for this endpoint is less well established at this moment. After careful reading of the abstracts of these 51 papers, it was concluded that the number of papers that specifically focused on water reuse was relatively low, *i.e.,* in most case the reuse purpose of the water investigated was not specified, e.g., WWTP effluent was investigated as a potential alternative water source. Nevertheless, all papers that were considered relevant for the application of *in vitro* bioassays to investigate genotoxicity and endocrine disruption in a water reuse context were summarized (Table II).

Wastewater/WWTP effluent appeared to be the most investigated water type (n=33 papers), followed by surface water (n=13 papers), hospital wastewater/effluent (n=5 papers), stormwater (n=2 papers) and drinking water (n=2 paper). For three out of the 51 papers the water type (source) was not specified (Table A). Details of the studies can be found in Table II.

Table A: Water types studied in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate genotoxicity and endocrine disruption in a water reuse context.

Total	Wastewater/ WWTP effluent	Hospital wastewater/ effluent	Stormwater	Drinking water	Not specified
51	33	5	2	2	3

In most cases, the application (water reuse purpose) was not entirely clear from the literature search (n=36 papers). Indeed, as mentioned above, not all papers identified specifically focused on water reuse. Drinking water (n=9 papers) was the most identified application, which was produced from surface water (n=7 papers) or an unspecified source (n=2 papers), thus representing de facto reuse (Beard et al. 2019, Houtman et al. 2010). It is acknowledged that the number of papers on drinking water quality assessment using bioassay is notably higher, but these papers did not come up in the current literature search due to the inclusion of water reuse related search terms in the string. Specific water reuse applications included groundwater recharge (n=3 papers), infiltration in superficial aquifer with downstream abstraction for irrigation (n=1 paper) and urban non-potable reuse and agriculture[4] (n=1 paper) (Table B). One paper specifically mentioned that the wastewater treatment that was investigated aimed at improvement of surface water quality. Although not specifically mentioned in the other papers, it can be assumed that in line with the Water Framework Directive the ultimate goal of all WWTP is be to protect and, where necessary, restore water bodies in order to reach good status, and to prevent deterioration protect or even improve surface water quality (2000/60/EC). Details of the studies can be found in Table II.

[4] Classes A+, A and B according to https://www.epa.vic.gov.au/.

Table B: Purposes of water reuse studied in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate genotoxicity and endocrine disruption in a water reuse context.

Total	Unspecified	Drinking water production	Groundwater recharge	Infiltration superficial aquifer[b]	Urban non-potable reuse and agriculture[3]
51	36	9[a]	3	1	1

[a] Produced from surface water (n=7 papers) or unspecified source (n=2 papers), thus representing *de facto* reuse.
[b] With downstream abstraction for irrigation

4.1.1 Genotoxicity

From the 51 papers that were identified as relevant, 26 papers were considered relevant for the state-of-the-art overview on the application of *in vitro* genotoxicity bioassays in a water reuse context. In those 27 papers, seven different *in vitro* bioassays (Ames test (Reifferscheid et al. 2012), micronucleus test (Reifferscheid et al. 2008), chromosomal aberration (Galloway et al. 1987), comet assay (Singh et al. 1988), umu test (Oda et al. 1985), SOS chromo (Quillardet et al. 1982), p53 CALUX (Van der Linden et al. 2014) were used. These six bioassays were used to measure three endpoints (mutagenicity, chromosomal aberrations and adaptive response to DNA damage). The most frequently used bioassay was the Ames (fluctuation) test (n=16 papers), followed by the micronucleus test (n=12 papers), umu (n=10 papers) and Comet assay (n=8 papers). Both the SOS chromo and p53 CALUX tests were described in one paper only (Table C). Details of the studies can be found in Table II.

Table C: Bioassays used in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate genotoxicity in a water reuse context.

Total	Ames (fluctuation) test	Micronucleus test	Umu test	Comet assay	SOS[a] chromotest	p53[b] CALUX
26	16	12	10	8	1	1

[a] Response-pathway of DNA damage present in prokaryotic cells (unicellular organisms such as bacteria) (Galloway et al. 1987)
[b] Protein that initiates pathway to protect cells from DNA damage in vertebrates (Van der Linden et al. 2014))

The number of identified papers and specific water types was too low to make a relevant distinction between water type/water reuse purpose and specific genotoxicity bioassays (*i.e.,* to investigate if a certain bioassay was performed specifically for a certain water type). Generally, it can be concluded that the bioassays described are technically applicable to all water types, but relevance of a bioassay for specific types of water and reuse applications can be defined. This discussion was considered out of scope of the current project because more data with water samples would be needed, e.g., information on the limit of quantification, expected responses based on sample composition etc. Moreover, there are developments on test batteries for water quality assessment (section 4.2), of which the most recent developed test strategies (STF2 and GWRC) define different strategies for different water types, also related to genotoxicity (section 4.2.1).

4.1.2 Endocrine effects

From the 51 papers that were identified as relevant, 41 papers were considered relevant for the state-of-the-art overview on the application of *in vitro* bioassays for endocrine disruption in a water reuse context. Several bioassays have been developed to investigate hormone-mediated adverse effects, such as chemical activated luciferase expression (CALUX) assays (Sonneveld et al. 2005, Van der Burg et al. 2013) and yeast-based screening assays for estrogenicity and androgenicity (Arnold et al. 1996, Kolle et al. 2010). The application of these assays to several samples of the water cycle, including waste water, surface water and rainwater, has been described (Hamers et al. 2003, Murk et al. 2002, van der Linden et al. 2008).

Endocrine disruption includes multiple modes of action (MOA), including estrogen receptor (ER)-activity, androgen receptor (AR)-activity, glucocorticoid receptor (GR)-activity, progesterone receptor (PR)-activity and

thyroid receptor (TR)-activity. From these MOA, based on the literature search, ER-activity is the best studied biological effect (n=39 papers), followed by AR-activity (n=20 papers) (Table D). With regards to ER-activity, the ER CALUX was most-frequently used (n=12 papers), closely followed by the E-screen (n=11 papers) and YES (n=9 papers), and two-hybrid ER (n=5 papers), anti-ER CALUX (n=4 papers), T47D-Kbluc (n=3 papers), VM7Luc4E2 (n=2 papers) and ER GeneBLAzer (n=2 papers) (Table E). There were some assays that were described in one paper only. Details of the studies can be found in Table II.

Table D: MOA measured in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate endocrine disruption in a water reuse context.

Total	ER-activity	AR-activity	PR-activity	GR-activity	TR-activity
41	39	20	10	8	6

Table E: Bioassays used in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate estrogenic activity in a water reuse context.

Total	ER CALUX	E-screen[a]	YES[b]	Two-hybrid ER[c]	Anti-ER CALUX	T47D-Kbluc[d]	VM7Luc4E2[d,e]	ER Gene-BLAzer
39	12	11	9	5	4	3	2	2

[a] Estrogen-screen
[b] Yeast estogen screen
[c] Molecular biology technique to study protein-protein or protein-DNA interacties
[d] Modified breast cancer cell lines modified with a luciferase gene (Rogers and Denison, 2000; Wilson et al. 2004)
[e] Formerly designated as BG1Luc4E2 (Dvořáková et al. 2016)

Likewise, the AR-CALUX (n=12 papers) was the most frequently used bioassay to study AR-activity, followed by YAS (n=5 papers), anti-AR CALUX (n=4 papers), AR GeneBLAzer (n=2 papers) (Table F). There were some assays that were described in one paper only. Details of the studies can be found in Table II. It must be noted that for investigation of AR-activity, bioassays for agonism (e.g. AR CALUX) as well as antagonism (e.g. anti-AR CALUX) were applied. More details regarding the methodology of assessing agonism and antagonism can be found in Neale and Leusch (2015).

Table F: Bioassays used in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate androgenic activity in a water reuse context.

Total	AR CALUX	YAS[a]	Anti-AR CALUX	AR GeneBLAzer
20	12	5	4	2

[a] Yeast androgen screen

PR-activity was studied in 10 out of 41 papers, of which 7 used the PR CALUX. The PR GeneBLAzer (n=3 papers), anti-PR CALUX (n=3 papers), PR-Laz (n=2 papers) and UEL zfPR (1 paper) were also used to determine PR-activity (Table G). Details of the studies can be found in Table II.

Table G: Bioassays used in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate progesteronic activity in a water reuse context.

Total	PR CALUX	Anti-PR CALUX	PR GeneBLAzer	Two hybrid PR[a]	UEL zfPR[b]
10	7	3	3	2	1

[a] Molecular biology technique to study protein-protein or protein-DNA interacties (Fields and Sternglanz 1994)
[b] Reporter cell line to study GR-activity (Neale et al. 2020)

GR-activity was studied in 8 out of the 41 of the identified papers relevant for endocrine disruption, of which 7 used the GR CALUX. The GR GeneBLAzer (n=5 papers), anti-GR CALUX (n=3 papers), GR Switchgear (n=2 papers), anti-GR GeneBLAzer (n=1 paper) and UMLN zfGR (n=1 paper) were also used to determine GR-activity (Table H). Details of the studies can be found in Table II.

Table H: Bioassays used in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate glucortocoid activity in a water reuse context

Total	GR CALUX	GR Gene-BLAzer	Anti-GR CALUX	GR Switchgear	Anti-GR GeneBLAzer	UMLN zfGR[a]
8	7	5	3	2	1	1

[a] Reporter cell line to study GR-activity (Neale et al. 2020)

From the papers identified, TR-activity was the least studied endocrine disruption MOA in a water reuse context (n=6 papers), for which the TR/TRβ CALUX (n=3 papers), T-screen (n=2 papers), two-hybrid TR (n=1 paper) and PC-DR-Luc (n=1 paper) were used (Table I). Details of the studies can be found in Table II.

Table I: Bioassays used in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate thyroidic activity in a water reuse context

Total	TR/TRβ CALUX	T-screen[a]	Two hybrid TR[b]	PC-DR-Luc[c]
6	3	2	1	1

[a] Thyroid screen
[b] Molecular biology technique to study protein-protein or protein-DNA interacties (https://en.wikipedia.org)
[c] Avian PC12 cells modified with luciferase gene to study thyroid receptor activity

Similar to the genotoxicity endpoint, the number of identified papers and specific water types was too low to make a relevant distinction between water type/water reuse purpose and specific bioassays to study endocrine effects (*i.e.,* to investigate if a certain bioassay was performed specifically for a certain water type). Generally it can be concluded that the bioassays described are technically applicable to all water types, but relevance of a bioassay for specific types of water and reuse applications can be defined. This discussion was considered out of scope of the current project because more data with water samples would be needed, e.g., information on the limit of quantification, expected responses based on sample composition etc. (Recent) developments on test batteries for water quality assessment also include various assays for endocrine disruption (section 4.2). From the literature search, the following was concluded:

– The amounts of material needed for the assays varied between the assays studied, e.g. ER-binding assay >> YES > ER-CALUX (Murk et al. 2002)
– Generally, mammalian cell lines are considered more sensitive towards endocrine disrupting chemicals than yeast cell lines (Dopp et al., 2021; Gehrmann et al., 2018; Leusch, et al. 2017). The mammalian cell lines are also closer to human physiology. For robustness the opposite is true as yeast bioassays were considered more robust (less subjective to variability) than mammalian cell lines (Dopp et al., 2021).

Since development of a test battery for estrogenic activity was out of scope of the current project, results and conclusions were not investigated in further detail, but provide a starting point for subsequent research.

4.1.3 Neurotoxicity

From the 3 papers that were obtained and identified as relevant (Tables J and K), all papers used the acetylcholine esterase (AchE) assay for evaluation of neurotoxicity (He et al. 2018; Macova et al. 2010; Macova et al. 2011). An additional relevant paper (Hamers et al. 2000) on the application of the AchE assay on rainwater samples was referred to and found during reading of these papers. The AchE assay quantifies the potency of the compounds present to inhibit the acetylcholine esterase enzyme. AchE hydrolyses the neurotransmitter

acetylcholine and effects are measured by a colorimetric reaction and is specifically designed for the detection of organophosphates and carbamate insecticide (He et al. 2018; Macova et al. 2010). Details regarding the AchE assay can be found in Ellman et al. (1961). A German standard method is available for the examination of water, waste water and sludge (DIN 38415-1, 1995). Application of the AchE has been explored in previous KWR research to investigate the robustness of drinking water treatment processes, but was not selected for the experiments because further optimization was considered to be required (Schriks et al. 2016).

Table J: Water types studied in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate neurotoxicity in a water reuse context

Total	Wastewater/WWTP effluent	Surface water	Rainwater
4	3	1	1

Table K: Purposes of water reuse studied in abstracts that were identified as relevant in the context of the application of *in vitro* bioassays to investigate neurotoxicity in a water reuse context

Total	Discharge to surface water	Drinking water	Unspecified
4	3	1	1

Since the group of neuroactive substances is not limited to organophosphates and carbamate insecticides, and because neurotoxicity is a complex toxicological endpoint in which multiple molecular and cellular processes are involved, additional neurotoxicity bioassays are required to cover this toxicological endpoint in water quality assessment (Reus et al. 2020a). A testing strategy specific for neurotoxicity in the context of water quality assessment is discussed in section 4.2.3.

4.2 (Recent) developments in test batteries for water quality assessment

Regulatory test batteries for genotoxicity as defined by European authorities for safety assessment of chemicals, pharmaceuticals and food ingredients include one or two tests for gene mutations and one for cytogenetic damage (ECHA 2017, EFSA 2011, EMA 2012). In line with this, for water quality assessment a genotoxicity test battery including 2-3 genotoxicity bioassays would also be considered as a minimum requirement to cover the different mechanisms of genotoxicity. It is known that adding additional genotoxicity assays may only marginally increase the sensitivity (defined as the ability to produce a positive response with a rodent carcinogen), but it may decrease the specificity (defined as the ability to give negative results with known rodent non-carcinogens), which may lead to more false positive results (Kirkland et al. 2005). For complex mixtures such as water samples, however, it is difficult to identify whether the observed response in such a test battery is in concordance with the expected results. The availability of multiple bioassays for different endpoints in water quality assessment, and within the same endpoint, raises the question which ones can be applied best in terms of feasibility, accuracy and cost-efficiency.

The DEMEAU project that was performed within the European Union Seventh Framework Programme aimed at selecting a panel of bioassays that is relevant for (drinking) water quality assessment. The selection was based on various selection criteria including assay applicability, ease of use and assay performance and resulted in an overview of promising *in vitro* bioassays for water quality determination (Figure 11, Schriks et al. 2015). The number of bioassays (for each specific pathway) in this overview is relatively high and the project did not establish a testing battery, but the overview provided a good starting point for follow-up research.

Toxicity endpoints relevant for drinking water monitoring	Specific pathway	Most promising bioassay(s)
Xenobiotic metabolism	PXR receptor agonists AhR receptor agonists	HG5LN PXR assay, PXR HepG2 assay DR CALUX, AhR geneblazer
Hormone-mediated mode of action	(anti)estrogenic activity (anti)androgenic activity (anti)glucocorticoid activity	ERα CALUX, YES assay AR CALUX, AR-MDA-kb2 GR CALUX, GR-MDA-kb2
Reactive mode of action	Gene mutations Chromosomal mutations DNA damage response	Ames fluctuation assay, ToxTracker Micronucleus assay, ToxTracker UMUc assay, Vitotox, p53 CALUX, BlueScreen
Adaptive stress response	Oxidative stress pathway	Nrf2 CALUX, AREc32 assay
Developmental toxicity	Focus point endocrine distruption	Various nuclear receptor activation assays, H295R assay)

Figure 11: Overview of promising in vitro *bioassays for water quality determination from the DEMEAU project, categorized by toxicity endpoints for water quality monitoring and subcategorized by specific pathway (Schriks et al. 2015). AhR: aryl hydrocarbon receptor, AR: androgen receptor, ARE: antioxidant responsive element, DR: dioxine receptor, ERα: estrogen receptor α, GR: glucocorticoid receptor, H295R: human adrenocortical carcinoma cell line, HG5LN: human cervical carcinoma cell line, MDA-kb2: human breast cancer cell line, Nrf2: nuclear factor erythroid 2–related factor 2 p53: tumour suppressor protein, PXR: pregnane X receptor, YES: yeast estrogen assay.*

For the application of bioassays on surface water, the Smart Integrated Monitoring (SIMONI) strategy was the first bioanalytical tool to be applied in surface water quality monitoring programs in the Netherlands. The first tier of the SIMONI strategy includes bioanalytical hazard identification of sites and the second tier focuses on identifying the risk drivers and confirming *in vitro* results with *in vivo* assays at the sites with the highest hazard indication only. The SIMONI strategy applies bioassays to detect specific modes of action, including the p53 CALUX for genotoxicity (Van der Oost et al. 2017a, Van der Oost et al. 2017b). For the interpretation of bioassay results in terms of environmental hazards and risks, effect-based trigger values (EBT) were developed for all bioassays. For an overall estimation of the environmental risks all risk quotients (bioassay effect divided by EBT) were used to calculate the SIMONI Risk Indication (SRI, formerly known as SIMONI score).

Since then, other bioassay testing strategies have been developed. In a Global Water Research Coalition (GWRC) project, a decision-making tool was developed that groups bioassays into three test batteries based on assay sensitivity, with test battery selection depending on the sampling campaign context and purpose (Figures 12 and 13) (Neale et al. 2021). A comparable approach where water context is considered in the bioassay strategy, is the basis set of bioassays (SFT2) that has been defined in the Dutch KIWK[5]-project, a collaboration between the Dutch government, provinces, water authorities, drinking water companies and knowledge institutes (De Baat et al. 2022). The SFT2 (*Key factor Toxicity 2.0*) is a follow-up of the ESF Toxicity (*Ecological key factor Toxicity*) which proposed a testing and interpretation strategy for the use of bioassays in water quality assessment, largely based upon the SIMONI strategy (Posthuma et al. 2016). The EBT for the interpretation of bioassay results are similar in SFT2 and the SIMONI model, except for the one used for PAH CALUX that is almost four times lower in SFT2.

[5] 'Kennisimpuls Waterkwaliteit', in English: knowledge impulse water quality

More recent, commissioned by the Dutch Ministry of Infrastructure and Waterways (I&W) a 'guidance for conducting biological effect monitoring for advanced treatment of WWTP effluents' was developed by several stakeholders from the water sector. In this context, conduct of the ERα CALUX, PAH CALUX, PXR CALUX, Cytotox CALUX and Microtox are proposed as mandatory bioassays for those water boards that on a voluntary basis participate in the contribution plan 'Purification of pharmaceutical residues'[6] from I&W and the innovation program 'Micropollutants from WWTP effluent' (IMPV)[7] that is being implemented together with STOWA[8] and the water boards. The p53 CALUX is only mandatory in case advanced treatment processes such as ozone, H_2O_2 and UV, that are being used to treat the WWTP effluent due to concerns of the formation of genotoxic transformation products (Figure 15) (Ecofide, 2023). The bioassays of this guidance show to some extent overlap with the SFT2 bioassays.

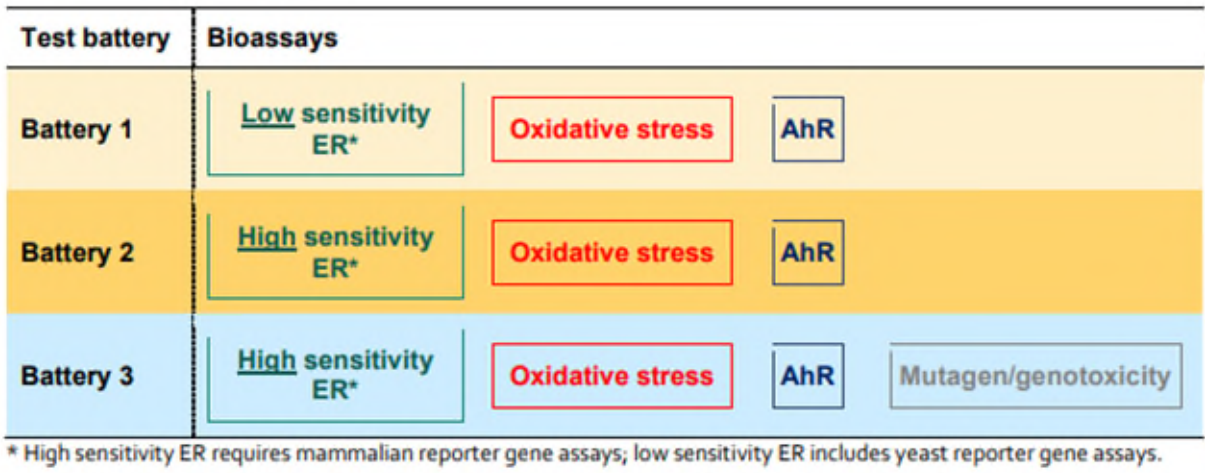

Figure 12: GWRC recommended endpoints in the different test batteries to apply for water quality monitoring (Neale et al. 2021). For battery selection depending on the context and purpose of the sampling campaign, see Figure 13. AhR: aryl hydrocarbon receptor, ER: estrogen receptor.

Water context	Purpose		
	Assess product quality	Assess treatment efficacy	Understand treatment processes (eg, CCP)
Wastewater treatment	Battery 1	Battery 1	Battery 2
Water reuse (non-potable)	Battery 2	Battery 2	Battery 2
Drinking water (incl potable reuse)	Battery 3	Battery 3	Battery 3

Figure 13: GWRC battery selection depending on sampling campaign context and purpose (Neale et al. 2021). For description of the different test batteries, see Figure 12. CCP: critical control points.

[6] In Dutch: Bijdrageregeling 'Zuivering medicijnresten'
[7] In Dutch: Innovatieprogramma 'Microverontreinigingen uit RWZI-afvalwater' (IPMV)
[8] Foundation for Applied Water Research, https://www.stowa.nl/english

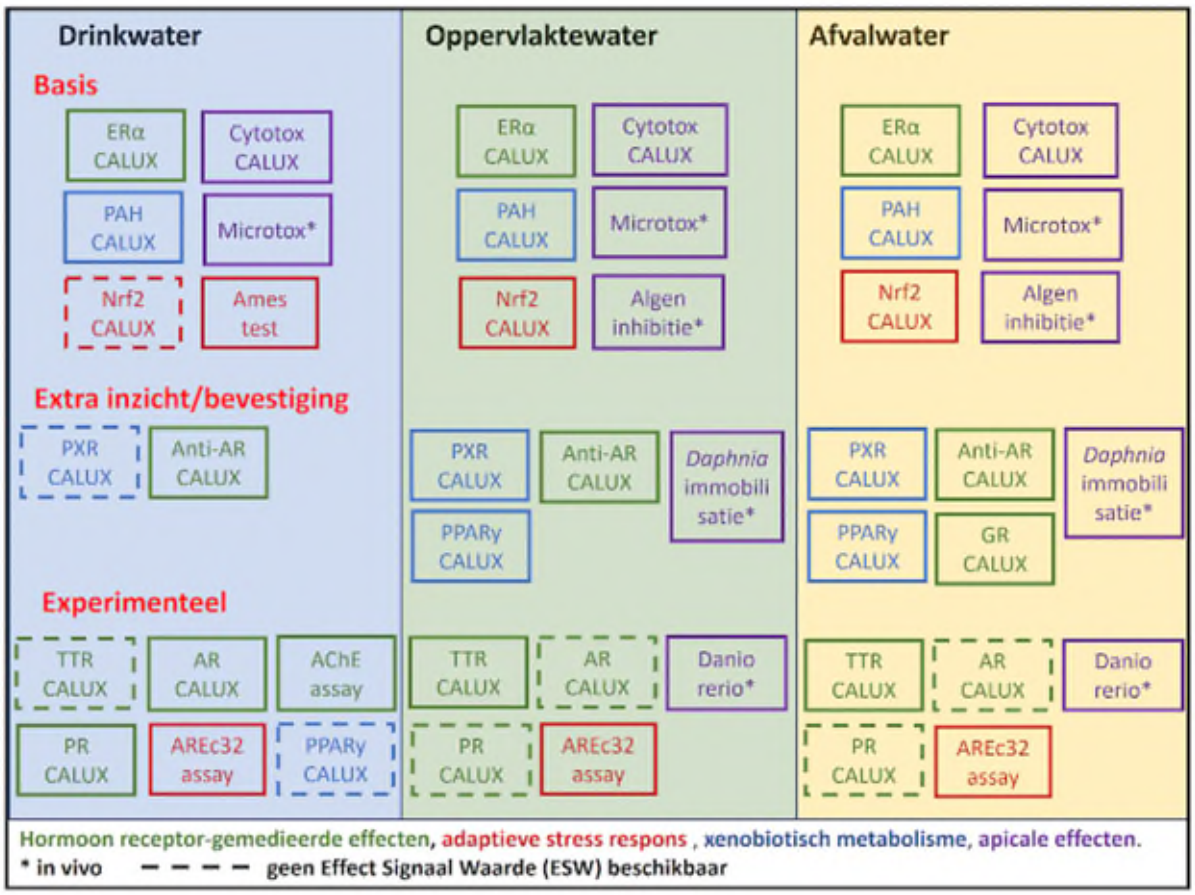

Figure 14: SFT2 bioassays developed to detect toxic substances and mixtures in a water quality context (De Baat et al. 2022). AchE: acetyl choline esterase, AR: androgen receptor, ARE: antioxidant responsive element, ERα: estrogen receptor α, GR: glucocorticoid receptor, PAH: polyaromatic hydrocarbon, PPARγ: peroxisome proliferator-activated receptor, PR: progesterone receptor, PXR: pregnane X receptor, Nrf2: nuclear factor erythroid 2–related factor 2, TTR: transthyretin. Drinkwater: drinking water (blue background), oppervlaktewater: surface water (green background), afvalwater: wastewater (yellow background). Basis: basic set, extra inzicht/bevestiging: additional information/confirmation, experimenteel: experimental. Bioassays in green text: hormone receptor mediated effects, red text: adaptive stress response, blue text: xenobiotic metabolism, purple text: apical effects (i.e. cytotoxicity). Bioassays surrounded by a dashed line: no effect-based trigger (EBT) value available.

Altijd verplicht	Alleen verplicht bij oxidatieve technieken
ERα-Calux PAH-Calux Microtox PXR-Calux Cytotox-Calux	P53-Calux (met en zonder S9) [a)]

Figure 15: Bioassay battery for biological effect monitoring for advanced treatment of WWTP effluents , AR: androgen receptor, ERα: estrogen receptor α, PAH: polyaromatic hydrocarbon, PXR: pregnane X receptor. a) Oxidative techniques such as ozone, H2O2 or UV can result in the formation of by-products. Some of these will not be environmentally hazardous, but substances with genotoxic properties, for example, may also be formed. To determine this effect, the p53-Calux without metabolic S9 activation is mandatory. In addition, substances can be formed, which are not genotoxic in themselves, but which (once in surface water) can still be converted into a genotoxic substance (precursor). This conversion is mimicked by performing the p53-Calux also with metabolic activation by S9. Once it has been determined for the WWTP and advanced treatment technique in question that the likelihood of genotoxic effects is limited (no/limited increase in response), follow-up research can be limited to either implementation method. This choice will be based on the results obtained (Ecofide, 2023)

4.2.1 Genotoxicity test batteries for water quality assessment

In both the GWRC and SFT2 decision making tools, genotoxicity testing is only considered for drinking water (section 4.2). The SFT2 solely recommends the Ames test (type not specified) for genotoxicity testing. The Ames test is categorized as an assay to investigate the adaptive stress response, together with the Nrf2 CALUX in the basic set and with AREc32 in additional experimental setting. Both Nrf2 CALUX and AREc32 are oxidative stress assays that also are recommended for surface water and wastewater (either in the basic set or in additional experimental setting) (Figure 14). The GWRC test battery does not provide a recommendation for a specific genotoxicity bioassay in the testing battery itself, however the report mentions that either the umu test for genotoxicity or the Ames assay for mutagenicity could be applied as both are commonly applied to water samples. Like the SFT2, the GWRC test battery also recommends oxidative stress for water contexts other than drinking water (Neale et al. 2021).

Specifically, for genotoxicity assessment, the section 'Toxicology of Drinking Water and Swimming Pool Water' at the German Environment Agency (UBA) developed a strategy within the Tox-Box consortium to investigate potential adverse effects of water samples. Their genotoxicity battery includes the Ames, umu and *in vitro* micronucleus test using V79 cells (Figure 16), and the combined results should give an indication of possible adverse effects on human health. In terms of risk assessment and management, the Tox-Box strategy describes the derivation of health-related indicator values (HRIVs) for single substances (Grummt et al. 2020), but the bioassay testing strategy can also be applied to complex mixtures such as water samples. However, it must be noted that single substances can be tested up to their solubility limit within the HRIV/Tox-Box concept, while the water samples contain substances in lower concentrations, even after concentration, e.g. using SPE (personal communication J. Kuckelkorn). Currently, only the ToxBox strategy meets the requirement of including the two different mechanisms of genotoxicity, i.e. mutagenicity and chromosome damage.

Ames Fluctuation Assay	Umu Test	Micronucleus Assay
ISO 11350: 2012	ISO 13829: 2000	ISO 21427-2: 2009
End point: Gene mutation in bacteria	End point: Induction of DNA Repair in bacteria	End point: Chromosome mutations in mammalian cells
Standard test strain:: *Salmonella typhimurium* TA98 *Salmonella typhimurium* TA100	Standard test strain:: *Salmonella typhimurium* TA1535/ psk 1002	Standard cell line: V79 cells (Chinese hamster fibroblasts)

Figure 16: ToxBox basic test battery consisting of three in vitro tests, which together represent the assessment-relevant endpoints "gene mutation" and "chromosome mutation" (Grummt et al. 2020).

Although differences may exist between water quality assessment and safety assessment for the authorization of substances, it is relevant to follow new developments in hazard assessment of regulatory testing to align with state-of-the-art methods and testing strategies for individual substances. The European Partnership for the Assessment of Risks from Chemicals (PARC) (PARC, 2022), for example, is a recent initiative that has a subtask on the development of an integrated approaches to testing and assessment (IATA) for individual substances in a regulatory context. In the work package on the development of Integrated approaches to testing and assessment (IATA) of chemicals new approach methodologies (NAM), such as the Ames fluctuation test and ToxTracker, are considered along with various other methodologies (Marx-Stoelting et al. 2023).

4.2.2 Test batteries for endocrine effects in water quality assessment

Assays to investigate endocrine disruption (hormone-mediated effects), as well as oxidative stress, metabolism and general toxicity have been included in recently developed test batteries for water quality assessment (STF2 and GWRC, Figures 12 and 14). Based on the literature study of the current project, estrogenic effects were investigated most, followed by (anti-)androgenic effects, supporting the main focus on reproduction toxicity within the field of endocrine disruption. Since there are different endpoints within endocrine disruption related to reproduction toxicity (e.g. estrogenic, (anti-)androgenic effects, (anti-)progestogenic effects (Houtman et al. 2021) inclusion of multiple assays in a test battery is recommended.

The ToxBox strategy describes a test battery specific for endocrine disruption, which is more extensive than the few reporter gene assays suggested by GWRC and SFT2. The ToxBox test battery for endocrine effects includes a tiered approach, including a receptor-mediated test (e.g. CALUX or Yeast screen) and the H295R steroidogenesis assay[9]. The second tier includes an in vivo reproduction test (Grummt et al. 2020) (Figure 17).

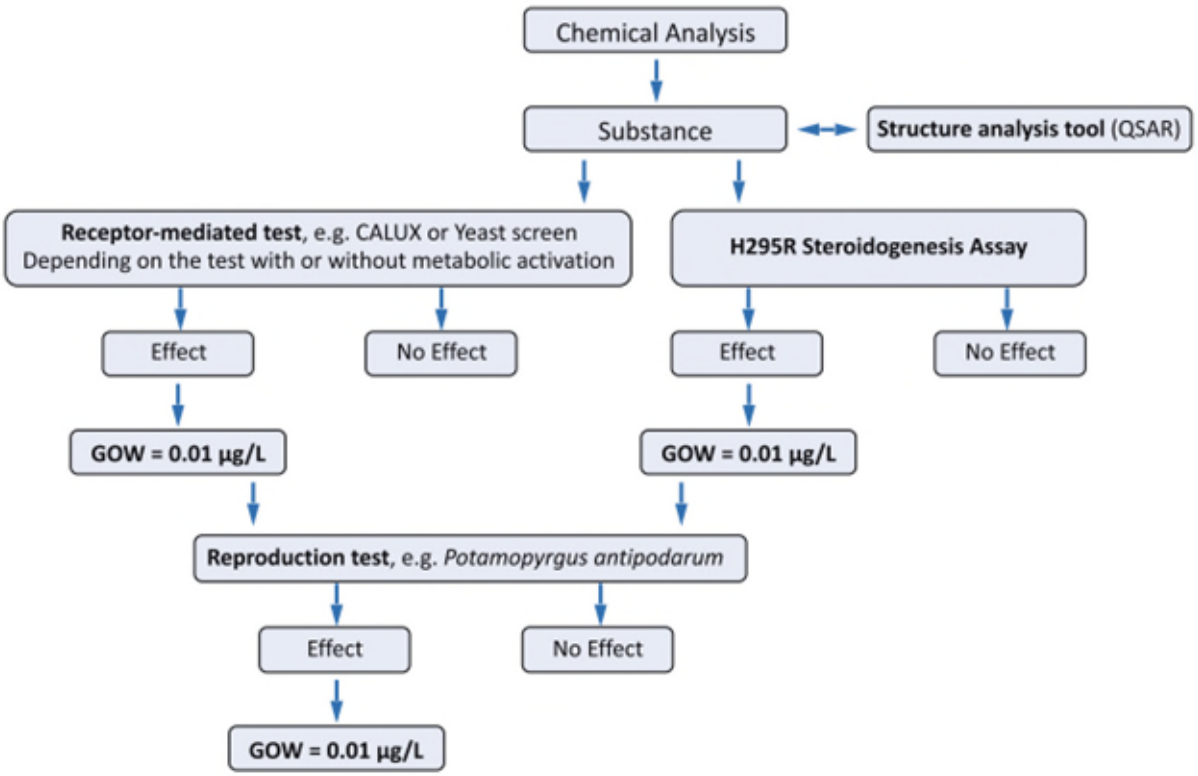

Figure 17: Schematic overview of the test strategy for endocrine effects of individual substances. The concentration is given an equivalent of the respective hormone (Grummt et al. 2020). GOW, Gesundheitlicher Orientierungswert (health benchmark).

[9] The H295R steroidogenesis assay is an *in vitro* test system for the identification of non-receptor-bound effects of endocrine disruptors (Hecker et al. 2011) and OECD Test no. 456: H295R Steroidogenesis Assay (2022).

4.2.3 Neurotoxicity test batteries for water quality assessment

To date, there are no *in vitro* regulatory neurotoxicity assays or *in vivo* bioassays in non-vertebrates that have been accepted for authorization of substances. Legradi et al. (2018) provided an ecotoxicological view on neurotoxicity assessment and concluded that neurotoxicity assessment is challenging due to the increasing number of neuroactive pollutants and complexity of the endpoint. Developments on *in vitro* neurotoxicity bioassays and their applicability for water quality assessment in the context of human health were also summarized previously (Reus et al. 2020a). Many bioassays for neurotoxicity are labour-intensive, costly, of limited relevance for humans, depending on limitedly available test systems and/or associated with ethical considerations. There is a need for a test battery for neurotoxicity in water quality assessment given the substances that may be present in different types of water for which this endpoint is relevant (e.g. pesticides). Information on occurrence and effects of such substances in the environment is needed as long term exposure to low concentrations may be associated with potential adverse effects to the nervous system (Davies, 1990). Development and validation of such a test battery, however, requires a lot of research. It is therefore recommended to continue following the current developments on neurotoxicity assessment.

Currently, neurotoxicity is not included in the GWRC test battery. The SFT2 test battery includes AchE assay for neurotoxicity assessment, but not in the basic set. The Tox-Box strategy includes a tiered approach including multiple neurotoxicity bioassays. A crucial step is the distinction of neurotoxicity from cytotoxic effects. Therefore, in the first test stage (tier), cytotoxicity studies are carried out. This also serves to determine the (maximum) test concentrations for subsequent specific neurotoxicity studies. This concept is a first important step in the evaluation of neurotoxicity (Grummt et al. 2020) and was validated in the follow-on project “Neuro-Box” (Grummt and Kuckelkorn, 2020). Based on the results, the existing test strategy could be further adapted and optimized (Figure 18). However, the investigation of new cell lines (microglial cells or induced human pluripotent stem cell (ihPSC) neurons) showed that practice-oriented use does not seem to make sense at this point in time. Nevertheless, the MTT test with neuronal stem cells, for example, was able to detect neurotoxicity with great sensitivity. The same applies to the use of -omics, which have been able to show various influences of the test substances at the molecular and behavioral level, but the corresponding biomarkers have not (yet) been identified. Both the extended EST (neuronal differentiation) and the studies on the embryo model were able to demonstrate the neurotoxic effect of hormone-active substances, which emphasizes the great interactive importance of these substances on the (early) developmental stages. The retinal assay and the olfactory epithelial assay proved to be good methods to detect effects on the sensory organs of fish. However, the transfer or linking of these findings to human neurotoxic effects still needs to be validated and is still pending. In addition, in the case of complex environmental samples, the additive effect with a similar direction of action or the antagonistic effect with different mechanisms of action, which could be predicted with CA and IA models, must be considered.

Nevertheless, the results also showed the further development potential for overcoming research-oriented possibilities and practice-oriented needs. The use of primary cells, pluripotent stem cells or highly specific cell lines; the use of -omics; the transfer of results from fish embryo tests to human toxicology; the evaluation of complex mixtures compared to the existing individual substance evaluation; all these are examples that can better map the known complexity of neurotoxic effects.

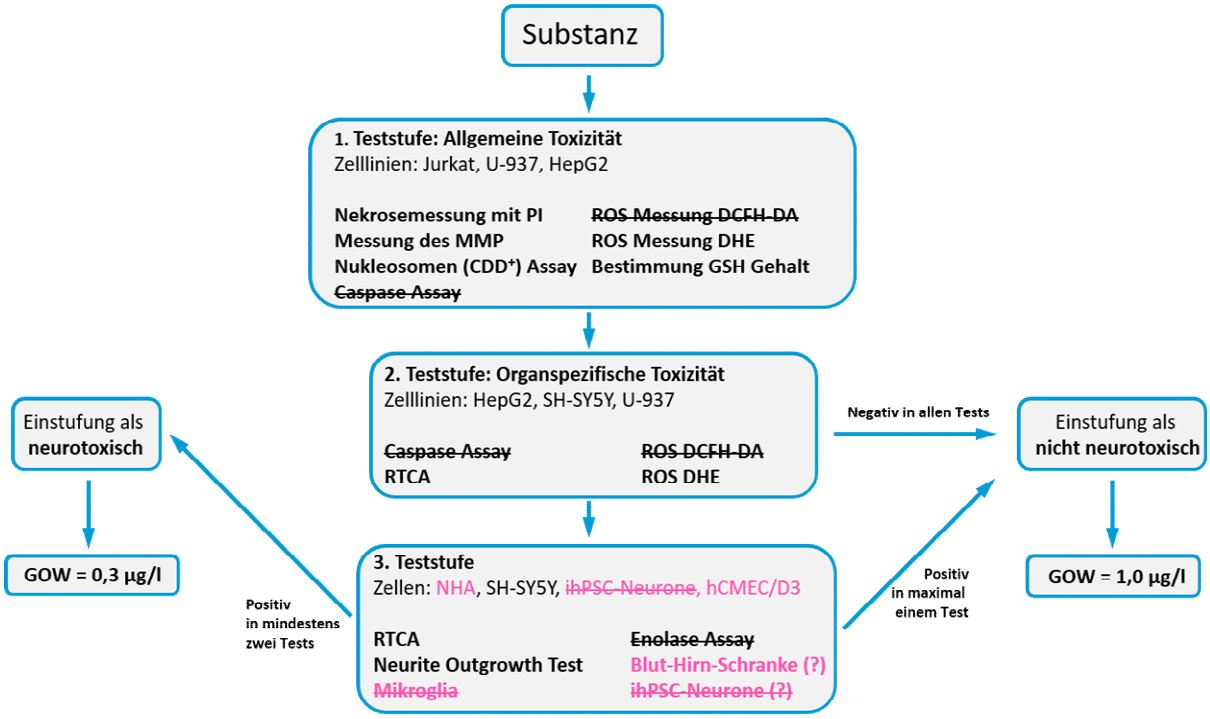

Figure 18: Scheme for in vitro testing for potential neurotoxicity of individual substances (Grummt and Kuckelkorn, 2020). Abbreviations: PI: propidium iodide, MMP: matrix metalloproteinase, ROS: reactive oxygen species, DCFH-DA: Dichloro-dihydro-fluorescein diacetate assay, GSH: glutathion, RTCA: real-time cell analysis, HRIV: health related indicator value. Jurkat, U-937, HepG2, SH-SY5Y, hCMEC/D3 and NHA are cell lines.

5 Discussion

A set of bioassays assessing different types of DNA damage were applied to different types of water samples from defined water cycles. DNA damage was selected as the primary biological effect for the experimental part because of the availability of well-established *in vitro* assays and its relevance for human health by inducing mutations and tumor formation. According to the standardized (ISO) protocols, cytotoxicity was measured simultaneously to assess if the observed genotoxicity response was not associated with excessive cytotoxicity. The aim was to develop a genotoxicity test battery for water quality assessment whilst considering cost-efficiency (considering labour-intensity and material costs) and sensitivity and specificity (based on the experimental data obtained in the present study and data available from literature). In addition, the conduct of bioassays for endocrine disruption and neurotoxicity has been explored in a literature search.

5.1 Reflection on experimental results

The bioassays selected for the experimental part of the current project highly aligned with the bioassays found to be used for water quality monitoring in the literature. In addition, the SOS chromotest was also found regularly in literature, but not included for the experimental part of the current project because the SOS chromotest and umu test show high similarity in endpoint. Moreover, the preference of the umu test over the SOS chromotest has been reported earlier as the umu test was demonstrated to be more sensitive and more cost-effective, and therefore more suitable for screening large numbers of environmental samples (McDaniels et al. 1990). Based on literature, the p53 CALUX was not often used for water quality assessment in a water reuse context, but this bioassay was selected for the experimental part of the current project because of its availability and common use for water quality monitoring in the Netherlands in practice. No literature was found on the use of the ToxTracker with water samples, however this assay was selected because of its availability and promising preliminary studies for water quality (Baken and Dingemans, 2017).

Based on the bioassay results of the current project, it is concluded that bioassays are able to discriminate between contaminated (untreated) samples and less contaminated (treated) samples. The treatment process applied to rainwater (UF-RO) was found to reduce the induction of cytotoxicity, which is associated with reduced amounts and/or concentrations of chemical micropollutants that are capable of inducing cytotoxicity, as reflected by absence of a severe cytotoxic response in the comet assay, micronucleus test and p53 CALUX observed for treated rainwater, whereas for untreated rainwater cytotoxicity was observed in the same assays. The additional treatment process applied to WWTP B effluent (a combination of ozonation and active carbon filtration) seemed also effective in removing chemical micropollutants that cause genotoxic responses, associated with the removal of substances with genotoxic properties, as reflected by absence of a positive response for treated WWTP B effluent in the umu test, absence of severe cytotoxicity in the comet assay and ToxTracker and a lower concentration where severe cytotoxicity occurred in the p53 CALUX when compared to WWTP B effluent. Reduction of genotoxicity after treatment of WWTP effluent has been observed previously (Macova et al. 2010, Mišík et al. 2011, Ternes et al. 2017, Zhang et al. 2019). However, increases in genotoxicity during ozonation or medium pressure UV treatment of WWTP effluent have been reported as well, but active carbon filtration genotoxicity or other subsequent treatment steps generally reduced the genotoxicity (Leusch et al. 2014a, Ternes et al. 2017, Zhang et al. 2019). This corresponds to the results of the present study in which samples of treated rain water and treated WWTP B effluent were collected at the end of the treatment process where genotoxicity and cytotoxicity were expected to be reduced. However, in case of the Ames fluctuation test and the micronucleus test, there was no, or hardly any, difference in severe cytotoxicity concentration levels between WWTP B effluent and treated WWTP B effluent (see Section 3.6, Tables 2 and 3). This could be related to the assay-specific characteristics of the bioassays, including the parameter used for cytotoxicity assessment

(bacterial growth/cell proliferation in Ames fluctuation test/micronucleus test vs. colour conversion/cell viability in comet assay/p53 CALUX and ToxTracker).

Most of the positive responses for genotoxicity were obtained in the umu test (three in total: WWTP A effluent, WWTP B effluent and WWTP C effluent), followed by the Ames fluctuation test (WWTP C effluent only) and comet assay (WWTP A only). The micronucleus test showed an equivocal response for genotoxicity for rainwater. None of the water samples showed a genotoxic response in the p53 CALUX (Table 2). Absence of p53 induction was confirmed by absence of activation of the Btg2-GFP reporter in the ToxTracker assay (ToxTracker Test Report and raw data, R. Derr and I. Brandsma, 2022). In the ToxTracker, none of the samples was found to induce aneugenicity (loss of complete chromosomes). Aneugenic substances are considered to exhibit non-linear dose-response curves and in terms of risk assessment a threshold is usually estimated, including identification of reference points for establishing a health-based guideline value or use in a margin of exposure approach (EFSA, 2021). The micronucleus test (Hashimoto et al. 2010, Benameur et al. 2011) and ToxTracker ACE[10] (Brandsma et al. 2020) can both be used to distinguish an aneugenic response from a clastogenic response (partial loss of chromosomes) by using a modified protocol or scoring method, however this was beyond the scope of this project. Since water samples are mixtures of compounds with different modes of action, estimation of a threshold based on aneugenicity may be less applicable in water quality assessment, unless aneugenicity is the only type of DNA damage caused (*i.e.* the sample only contains (predominantly) aneugenic substances e.g., pharmaceuticals such as colchicine, vinblastine sulphate and paclitaxel, but not other genotoxic compounds). Thresholds for aneugenicity are expected to differ from the current thresholds (e.g., EBT) that are applied for bioassay data interpretation in water quality assessment. More information on bioassay data interpretation can be found in section 5.3.

The results of the present study are in line with previous studies which reported positive responses for wastewater or WWTP effluent obtained with the Ames fluctuation test (Mišík et al. 2011, Papa et al. 2016), umu test (Escher et al. 2014, Jia et al. 2015, Leusch et al. 2014a, Watson et al. 2012) and comet assay (Mišík et al. 2011, Papa et al. 2016). Negative responses with wastewater have been reported as well for the Ames fluctuation test (Ragazzo et al. 2017), which is in line with the observation that not all WWTP effluent samples showed positive responses in these bioassays in the present study. The micronucleus test with WWTP effluent showed only negative responses in the present study, which is in line with results from literature (Magdaleno et al. 2014, Mišík et al. 2011, Papa et al. 2016). Responses above the LOQ have been reported for wastewater in the p53 CALUX (without S9), but not with WWTP effluent (Välitalo et al. 2017), in line with the responses below LOQ for WWTP effluent in the present study. Since the number of samples and publications on wastewater, WWTP effluent in combination with the p53 CALUX is limited, comparison of the results of the present study with literature is difficult.

Available literature on rainwater was limited. A couple of studies in literature compared results of stormwater (which can be compared to results obtained with rainwater in the present study) with WWTP effluent and concluded that stormwater had a slightly different toxicity pattern to WWTP effluents based on algae photoinhibition, but that it was also dominated by pesticides (Escher et al. 2014). Low or no genotoxicity of stormwater samples has been observed with the umu test, with responses around or slightly above the level typically found in secondary treated WWTP effluents (Tang et al. 2013). This indicates that positive genotoxic responses with stormwater can be obtained, which is in the current study reflected by the equivocal response of the micronucleus test.

[10] ACE:Aneugen and Clastogen Evaluation

While most of the positive responses for genotoxicity were obtained for the umu test, none of the water samples showed severe cytotoxicity in this bioassay. Three water samples were severely cytotoxic in the ToxTracker assay (WWTP A effluent, WWTP B effluent and WWTP C effluent), four samples were severely cytotoxic in the Ames fluctuation test (WWTP A effluent, WWTP B effluent, treated WWTP B effluent and WWTP C effluent) and comet assay (rainwater, WWTP A effluent, WWTP B effluent and WWTP C effluent) and five samples showed severe cytotoxicity in the micronucleus test and p53 CALUX (all water samples except treated rainwater; Table 3). The relatively high number of samples for which severe cytotoxicity was observed in the micronucleus test and p53 CALUX is not reflected by a higher number of positive responses for genotoxicity. From these data it seemed that there is an inversely proportional relationship between severe cytotoxicity and a positive genotoxic response for the micronucleus test and p53 CALUX. This can be related to differences in the duration of the exposure and recovery times (with the umu test having the shortest exposure/recovery time and the micronucleus test the longest), in the sensitivity of the cell types towards genotoxicity and cytotoxicity (bacteria vs. human cell lines vs. mouse embryonic stem cells) and in the different endpoints (e.g. mutagenicity, chromosome damage and cytotoxicity) and parameters (e.g. number of yellow wells, micronuclei, bacterial growth, cell proliferation and viability) (see also Table 1).

Although no cytotoxicity was observed for any of the samples in the umu test, growth factors below 0.5 have been reported previously for some wastewater samples (influent) (Välitalo et al. 2017). It must be noted that all effluent samples of the present study have undergone treatment, resulting in less contaminated samples which explains absence of cytotoxicity in the umu test.

In the current project, the results of the p53 CALUX and micronucleus test were highly comparable in terms of qualifying water samples genotoxic or cytotoxic. The only difference was that rainwater showed an equivocal response in the micronucleus test and a negative response in the p53 CALUX and the lowest concentrations showing genotoxic or cytotoxic responses varied between the two assays. A relationship between p53 induction and micronuclei formation was demonstrated by Salazar et al. (2009), who indicated that the level of p53 is associated with chromosomal damage human RKO cells[11]. Further research is warranted to confirm this and to investigate the relationship between p53 induction and micronuclei formation in water quality monitoring in practice to conclude if the p53 CALUX and micronucleus test can be considered interchangeable in a test battery.

In the present study, samples showed positive responses in the absence of S9, and not in the presence of S9. This is in line with previous observations for the Ames fluctuation test (Zhang et al. 2019), umu test (Macova et al. 2010, Tang et al. 2014) and is likely to represent detoxification of substances in the water samples.

5.2 Provisional genotoxicity test battery for water samples

Based on the performance of the bioassays and practical considerations, an attempt was made to develop a testing battery for genotoxicity testing of water samples.

5.2.1 Performance of the bioassays

In contrast to individual substances for which *in vivo* genotoxicity and carcinogenicity data can be available, it is difficult to investigate the performance of the bioassays for water samples, as it is not known what to expect as the samples are mixtures of (usually) varying, unknown chemicals. Even with mixtures of known chemicals it is difficult to define the expected response due to interactions between the chemicals (e.g., synergism, addition, antagonistic). In general, it can be assumed that with samples that are known to contain a high amount of contaminants (such as wastewater and effluent) there is a higher chance of a positive response, whereas for relatively clean samples (rainwater and drinking water) a positive response is unlikely.

[11] RKO is a poorly differentiated human colon carcinoma cell line. It can be used as the control cell line for investigating the effects of p53 and gadd45 on cellular parameters (https://www.atcc.org)

Based on the results obtained in the current project, the most conservative option would be to use the umu test as first tier in a testing battery, as it showed the most positive responses (3 out of 6 samples) compared to the other bioassays. Reifferscheid et al. (1996) concluded a relatively low sensitivity and specificity of the umu test. In the present study, the positive responses obtained in the umu test were found in the samples that are expected to contain a relatively high amount of contaminants, showing that the umu test appears to be able to discriminate between contaminated and less contaminated to clean samples. Based on data available from literature on individual substances (Table 4), the ToxTracker shows the highest sensitivity and specificity, followed by the p53 CALUX. It must be noted, however, that sensitivity and specificity calculations are highly dependent on the number of carcinogens and non-carcinogens used and the properties of the substances (e.g. genotoxicity mechanism, water solubility, test concentrations) that are included. In addition, sensitivity and specificity of bioassays using (human) cell lines may be higher that bioassays using bacteria because (human) cell lines are physiologically closer to animals (to which the data of *in vitro* tests often is compared) and humans. Moreover, the SOS response, which is principle of the umu test, is not present in eukarytic cells (Janion, 2008). This may explain the relatively low sensitivity and specificity calculated for the Ames (fluctuation) and umu test.

Table 4: Sensitivity and specificity of current genotoxicity assays for prediction of carcinogenicity[12]. The number of chemicals showing a response and the total number of chemicals used for the calculation of the sensitivity and specificity is shown between brackets.

Test name	Sensitivity (%)	Specificity (%)	References
Regulatory			
Bacterial reverse mutation (Ames) test[1]	60 (326/541)	77 (136/176)	Kirkland et al. 2005
Mammalian cell mutation test[1]	81 (198/245)	48 (50/105)	Kirkland et al. 2005
In vitro micronucleus test[1]	81 (72/89)	54 (14/26)	Kirkland et al. 2005
In vitro chromosome aberration test[1]	70 (245/352)	55 (75/136)	Kirkland et al. 2005
Comet assay	88 (74/84) 89 (31/35)	64 (7/11) 78 (6/27)	Anderson et al. 1998[2] Kirkland et al. 2006[3]
Screening			
Ames fluctuation test	58 (28/48)	63 (5/8)	Kamber et al. 2009
Umu test	62 (93/149)	72 (18/25)	Reifferscheid et al. 1996
p53 CALUX	82 [4,5]	90 [4,5]	Van der Linden et al. 2014
ToxTracker	95 [4]	94 [4]	Hendriks et al. 2016

[1] If equivocal responses are counted positive
[2] Not taking account of the difference between in vitro and in vivo responses, species differences or organ and tissue differences.
[3] In vivo comet assay
[4] Number of chemicals unknown
[5] Sensitivity and specificity for prediction of in vivo genotoxicity

5.2.2 Practical considerations

Although it is acknowledged that sensitivity and specificity is important in the selection of bioassays, for water quality assessment it is important to consider practical considerations as well. Water extracts are costly and consequently generally available at low quantities. In addition, since water quality monitoring requires frequent testing of multiple water samples, a cost-effective test battery is preferred.

[12] This table was adopted from the ToxTracker Test Report (R. Derr and I. Brandsma, 2022) and expanded with additional bioassays

Based on personal experience of the authors, in terms of *cost efficiency*, the umu test may be favored as it is a relatively fast and cost-effective bioassay (results are available within one working day). The Ames fluctuation test and p53 CALUX can be regarded as moderately cost-efficient (requiring a couple of days to obtain results, taking into account initiation of cell culture, exposure and recovery times). The comet assay and micronucleus test are more labour-intensive than the umu test, Ames fluctuation test and p53 CALUX. In the current study, for the comet assay a separate experiment for cytotoxicity assessment was conducted, while for the other assays cytotoxicity was measured within the genotoxicity assay. While the data analysis of the umu test, Ames fluctuation test and p53 CALUX can be done within a couple of minutes or hours after the test is completed, scoring of comets and micronuclei can take days. However, there are high-throughput options available for the comet assay and micronucleus test, which makes these bioassays more attractive for cost-efficient water quality monitoring than the classic versions (Shibai-Ogata et al. 2011; Stang and Witte 2009; Sykora et al 2018). The ToxTracker is comparable to the Ames fluctuation test and p53 CALUX in terms of duration.

Costs for performance of a bioassay are also associated with the generation of water sample extracts. Many different protocols are used worldwide to concentrate water samples into extracts for *in vitro* bioassays, and the optimum protocol depends on the water matrix and toxicological endpoint investigated with the assay (Abbas et al. 2019). The concentration factor also affects the bioassay response, with higher concentration factors increasing the chance of an effect, but at the same time this may increase the number of false positive responses. However, this should not be the case when the bioassay responses are compared to an EBT taking into account correction for the REF. In order to reduce the number of false positives in genotoxicity assessment of individual substances in a regulatory context, lowering the maximum concentration is recommended (Corvi et al. 2017). Indeed, several international organisations have updated their guidance regarding top dose selection for genotoxicity testing in the last decade (Galloway et al 2011, ICH 2012, OECD, 2022). However, testing of individual chemicals is notably different from testing of complex mixtures such as those occurring in water samples, and hence, it is difficult to establish a top concentration that is sufficiently high to detect a response and at the time avoids a false positive response. In the current project, the maximum REF tested varied between 100 and 200 for different assays, and for most of the samples either a genotoxic or cytotoxic response was observed, indicating that the concentration factor of 10.000x applied was sufficiently high to induce an effect. The exceptions are treated rainwater (neither genotoxicity nor cytotoxicity was found in any of the tests), and rainwater and treated WWTP B effluent in case of the Ames fluctuation test, comet assay and/or ToxTracker (negative for both genotoxicity and cytotoxicity). These samples can be assumed the cleanest samples after treated rainwater, thus the absence of a cytotoxic response is not unexpected. Higher concentration factors will increase the chance of an effect, but at the same time increase the chance of false positive responses.

In terms of the *volume of water extract* that is required for bioassay analyses, from the six genotoxicity assays used in the current project, the p53 CALUX is the most efficient assay, followed by the umu test, comet assay and micronucleus test (the latter depending on the plate type used, number of replicate cultures and experiment), while the Ames fluctuation test and ToxTracker require the highest volumes of water extract. Indeed, the Ames fluctuation test and ToxTracker include multiple bacterial strains or cell lines, thus requiring more volume of water extract is logical.

5.2.3 Possible position of genotoxicity bioassays in a test battery for water quality assessment

It can be argued that in particular water samples for which long-term human exposure is foreseen, e.g. drinking water, should be tested for genotoxicity. However, also with respect to other (reuse) purposes and environmental health, it is acknowledged that exposure to genotoxic substances, also via water, should be avoided as much as possible. The Ames test was added to the SFT2 testing battery for drinking water because of the formation of potential genotoxic transformation products that can be induced by specific drinking water treatment processes (De Baat et al. 2022), e.g. ozone oxidation and chlorination (Han et al. 2018, Zhang et al. 2016). Since advanced treatment of WTTP effluent is increasingly applied, also within the context of water reuse, genotoxicity testing may be recommended for treated WWTP effluent as well. If needed, additional treatments

(e.g. additional of a granular activated carbon (GAC) filtration) steps can be included to remove such activity (Heringa et al. 2011). In the current study, the genotoxic responses of WWTP B were removed after the additional treatment process, as demonstrated by absence of a genotoxic response in all genotoxicity assays performed.

Important to note is that the DNA damage detected by the comet assay can be reversible (Tice et al. 2000) and that the umu test, p53 CALUX and ToxTracker indicate activation of the DNA damage response (Hendriks et al. 2016, Oda et al. 1985, Van der Linden et al. 2014), and thus can be considered as indicator assays. The mutations and cytogenetic damage detected with the Ames fluctuation test and micronucleus test represent irreversible DNA damage.

Although the ToxTracker can be applied as genotoxicity screen in the (non-regulatory preclinical) development phase of pharmaceuticals and personal care products, it is foreseen that for water quality assessment the ToxTracker has a potential in following up a battery of genotoxicity bioassays to provide insight into mode of action of (genotoxic compounds in) samples and herewith support a weight of evidence approach[13]. Moreover, it can be considered a more advanced assay as it uses six cell lines and covers multiple relevant mechanisms of genotoxicity, including clastogenicity and aneugenicity.

To date, the Ames (fluctuation) test is the only bioassay applied in water quality assessment that is designed to detect mutagenicity. There are other *in vitro* bioassays available, such as the mammalian cell gene mutation test (OECD 476, 2016; OECD 490, 2015) but these are less suitable for water quality monitoring for practical reasons (time consuming and requiring relatively high amounts of sample volume). Therefore, it is acknowledged that a test battery should include at least the Ames fluctuation test for mutagenicity assessment.

In addition to the Ames fluctuation test to detect mutagenicity, a test battery for genotoxicity should include a bioassay to detect chromosomal damage. The micronucleus test, preferably a medium to high-throughput version e.g., using flow cytometry analysis, is a good candidate. Similar to the Ames test, this test has been accepted and implemented in testing strategies for regulatory purposes. Further research is warranted to investigate if the p53 CALUX and micronucleus test can be considered interchangeable in a test battery, as there may be a relationship between p53 induction and micronuclei formation (Salazar et al. 2009). Further research is also warranted to investigate if the ToxTracker and micronucleus test can be considered interchangeable due to overlapping mechanism (clastogenicity/aneugeniciy). In addition, the ToxTracker may be a valuable assay to follow-up a positive responses for genotoxicity.

Prantl. et al. (2018) proposes an in vitro genotoxicity test battery consisting of the Ames fluctuation test with two tester strains (ISO 11350), the umu test and the micronucleus test, or the Ames test with five tester strains (OECD 471) and the micronucleus test (*i.e.*, with more strains in the Ames test the umu is considered not needed). The Tox-Box strategy (Grummt et al. 2018) also includes Ames fluctuation test, micronucleus test and umu. Indeed, the umu test is complementary to the Ames fluctuation test and micronucleus test and could be included in a genotoxicity test battery as a third bioassay. The umu test is considered *cost efficient* (section 4.3.2), which can be helpful in case of emergencies such as sudden contamination of a drinking water source. However, similar to the Ames fluctuation test, the umu test has the disadvantage of having a relatively low sensitivity and specificity compared to the other assays in Table 4. Further research is warranted to investigate if the umu test is of added value for genotoxicity assessment of water samples.

Using a test battery with 2-3 genotoxicity bioassays may often result in conflicting results. Positive responses in one bioassay cannot be neglected, even when one or two other bioassays show negative responses. For regulatory purposes, chemicals, pharmaceuticals and food and feed ingredients are subjected to *in vivo* (rodent)

[13] Inspired by https://www.criver.com/products-services/safety-assessment/toxicology-services/genetic-toxicology/toxtracker-assay?region=3696.

genotoxicity testing when one or more positive results are obtained with *in vitro* genotoxicity tests (ECHA 2017, EFSA 2011, EMA 2012). However, besides ethical issues that would arise with testing water samples in rodents, *in vivo* testing is expensive and time-consuming, and therefore not recommended as follow-up option. If positive responses are obtained with water samples in bioassays, follow-up research is recommended to investigate the cause of the positive response by effect-directed analysis (Zwart et al. 2020) or chemical composition identification (Béen et al. 2021a) and integrated human health risk assessment (Baken, 2018).

5.3 Interpretation of bioassay responses for different water types

With regards to interpretation of positive responses for genotoxicity and cytotoxicity, it should be emphasized that the water samples selected for the current project were chosen to include different water qualities, and that risk assessment requires more data (including repeat measurements and exposure data). It should also be emphasized that the samples chosen are not necessarily representative of rainwater and WWTP effluent in general, and that outcomes may vary per location and season. In addition, it must be considered that the samples tested here are concentrated before application in the bioassay to increase sensitivity of the analyses, which is in line with the requirements of regulatory testing (ECHA 2017, EFSA 2011, EMA 2012), which can be seen as a worst-case scenario. Furthermore, it is known that *in vitro* assays are generally considered oversensitive with respect to animal models and humans (Kirkland et al. 2007, Lynch et al. 2011). Bioassay results of the current project are interpreted to demonstrate methods rather than making conclusions on risks. Nevertheless, the results obtained in the current project demonstrate that the application of bioassays in a water (re)use context is of added value to prioritize water samples for further evaluation. When there is reason to further investigate possible risks to humans and the environment on the basis of effect-based methods (bioassay responses), effect-directed analysis (EDA) can be used to conduct a targeted search for the chemical identity of the substances that caused this response (Brack et al. 2016, Houtman et al. 2020, Zwart et al. 2018). For those identified substances with missing toxicity data, *in silico* approaches, such as quantitative structure-activity relationships (QSARs) and read-across using *in silico* tools (e.g., QSAR Toolbox[14] (Dimitrov et al. 2015)), and existing experimental databases (e.g., ToxCast[15] (Dix et al. 2007)) can be used to link bioassay responses to specific substances and subsequent estimation of potential human and environmental health risks (Shao et al. 2019). This may direct selection and deployment of measures and decision-making processes in (drinking) water production to mitigate any health or environmental risks in targeted ways (Reus et al. 2022a, Reus et al. 2023a).

5.3.1 Effect-based trigger values

Extrapolation from *in vitro* test results to the relevance for human health and the environment is challenging. Effect-based trigger (EBT) values may aid in this context, where responses above the EBT value should indicate that a human or environmental health risk cannot be excluded (*i.e.* require further investigation) and responses below the EBT value indicate that there is a low risk. EBT values for human and environmental health risks are available for multiple reporter gene assays including the ER-CALUX (Béen et al. 2021b, Brand et al. 2013, Escher et al. 2015, Escher et al. 2018, Van der Oost et al. 2017a). Currently, for the six bioassays used in the current project, no EBT values are available. Projects to develop an EBT value for the Ames fluctuation test, umu test and p53 CALUX are ongoing at KWR. Since the responses of the p53 CALUX in the present study were found to be below the limit of quantification, comparison to the EBT value is not relevant.

Experimental Ames fluctuation test and umu results of the current study were compared to preliminary EBTs derived for these tests (Reus et al. 2023b). Responses of the water samples of both bioassays were expressed as equivalents of the reference compound 4-nitroquinoline-n-oxide (4-NQO). The WWTP C effluent sample exceeded the least conservative EBT value of the Ames fluctuation test derived by Reus et al. (2023b). For the umu test, the least conservative EBT value derived by Reus et al. (2023b) was exceeded by WWTP effluent A,

[14] https://qsartoolbox.org
[15] https://www.epa.gov

WWTP effluent B and WWTP C effluent. Exceedance of the EBT indicates that a health risk after consumption of this water cannot be excluded. However, it is not likely that WWTP effluent and WWTP C effluent will be consumed directly by humans or animals (except in cases where the effluent is discharged to small surface waters in periods of drought). In addition to the considerations made in the first paragraph of section 5.3, it must be noted that the currently developed EBT values for the Ames fluctuation test and umu test are developed for human risk assessment, not animal or ecological risk assessment. Since the absolute responses in the bioassays were a between 100-10000 and 10-100 times higher than the EBT value, for the Ames fluctuation test and umu test, respectively, further evaluation of the EBT using additional individual chemical and validation data is warranted.

5.3.2 Toxic equivalency factor

EBT values have not (yet) been developed for all bioassays. In absence of an EBT value, the response of the bioassay can also be expressed as equivalents of a reference compound, followed by applying a threshold based on the toxic equivalency factor (TEF). In this approach, based on the obtained reference compound equivalent concentrations for the tested water samples and carcinogenicity data of the reference compound (in case of genotoxicity bioassays), an indication of the associated risk of the tested water samples is obtained via the margin of exposure (MOE). The MOE is the ratio of its no-observed-adverse-effect level to its theoretical, predicted, or estimated dose or concentration of human intake (Benford et al. 2010). Martijn et al (2016) developed this TEF approach for the Ames fluctuation test in in UV/H_2O_2 water treatment samples using the reference compound 4-NQO. Based on a rodent carcinogenicity study, a body weight of 70 kg and assuming a drinking water consumption of 2 L per day, it was concluded that the 4-NQO equivalent concentration should not exceed 80 ng/L associated with a negligible risk. Application of this approach on samples from MP UV/H2O2 treated water of a full-scale drinking water production facility, a 4-NQO equivalent concentration of 107 ng/L was established. These results indicate a safety concern in case this water would be distributed as drinking water without further post treatment (Martijn et al. 2016). In principle, this approach can also be applied to other (genotoxicity) bioassays and overcomes the limitation of requiring data availability of sufficient compounds to be included in the EBT derivation, but requires thorough risk assessment based on existing toxicity data of the reference compound.

When the positive responses of the Ames fluctuation test of the current study are compared with the 80 ng 4-NQO equivalents/L threshold derived by Martijn et al. (2016), the threshold is exceeded only for the WWTP C effluent sample. When the positive responses of the umu test are compared to the same threshold, it is exceeded only for the WWTP effluent B and WWTP C effluent sample. However, because of the high concentration of anthropogenic substances as indicated by the positive response of the WWTP A effluent in the umu test, direct consumption of WWTP A effluent should also be avoided. The same considerations for interpretation of the results should also be considered here. Furthermore, it is recommended to further evaluate the applicability of the threshold value of 80 ng 4-NQO eq/L for the umu test by generating more data. It should also be considered that the TEF approach is based on a single compound, whereas the EBT approach is based on multiple chemicals with a different effect potency. For the TEF approach it is extremely important to select an appropriate, relevant reference compound.

6 Conclusions and future perspectives

Raising interest in the (re)use of alternative water sources requires consideration of potential risks resulting from circulating and potentially accumulating microbial and chemical contaminants in water cycles. Effect-based methods (bioassays) offer a powerful approach to investigate relevant hazards of complex low-level chemical mixtures in the water cycle. Available test batteries are useful decision-making tools for bioassay selection for investigation of water treatment efficacies and quality assessment of various types of water, including wastewater, surface water and drinking water. Guidance on bioassay selection for other types of water use, including non-potable use such as (sub-)irrigation is not yet available. In these cases, recommendations for water types that are comparable to the water use of interest can be followed (e.g. surface water or WWTP-effluent in case of sub-irrigation), but this should be further elaborated. Initiatives in this field have started (e.g., TKI project 'Borging van effluent RWZI voor glastuinbouwsector').

Genotoxicity bioassays used in the present study were able to discriminate between untreated and treated samples in water cycles. The responses between the bioassays differed due to differences in assay-specific characteristics, e.g., genotoxicity mechanism, sensitivity and specificity. Based on the experimental results of the water samples obtained in this study, it was not possible to define a test battery for genotoxicity. Based on literature, the Ames fluctuation test and micronucleus test are most commonly used for genotoxicity assessment in water quality monitoring. These bioassays cover both mechanisms of genotoxicity (base pair mutations and chromosomal damage) and are obliged for regulatory testing of chemicals, pharmaceuticals and food and feed ingredients. Recommendations regarding individual genotoxicity bioassays were made in the current research, along with suggestions for further research. Guidance on the selection of genotoxicity bioassays supplements existing bioassay batteries (such as GWRC and SFT2). Genotoxicity testing is considered most important for water types associated which long-term human exposure is foreseen (e.g. drinking water), though for other water types and in the context of environmental health, the presence of genotoxic substances should also be avoided as much as possible. Suggestion of future research include further investigation of the performance of bioassays by testing additional water samples and mixtures with known genotoxic and/or non-genotoxic compounds. Current guidance gaps including neurotoxicity, immunotoxicity and microbiological hazards, and their data interpretation should also be considered in future research to expand the applicability of effect-based methods to safeguard the chemical quality of water resources.

For interpretation of the bioassay results, the use of assay-specific criteria are recommended. Benchmarking of the bioassay response by comparison to an EBT provide insight in potential risks for human health and the environment, where environmental EBT can be used for non-potable (re)use. However, the derivation of preliminary EBT for genotoxicity bioassays warrants further research.

In a water (re)use context, bioassays are of added value to evaluate safety and prioritize water samples for further evaluation, including chemical identification and integrated risk assessment. Using existing methods such as effect-directed analysis, *in silico* tools to predict toxicity and evaluation of available experimental databases, observed bioassay responses can be linked to specific substances, which may direct selection and deployment of measures and decision-making processes in (drinking) water production and to mitigate any health or environmental risks in targeted ways. This approach supports decisions on applicability of alternative water sources while safeguarding human and environmental health.

Acknowledgement

The authors would like to thank René van Doorn and Leonie Pap from KWR (Nieuwegein, The Netherlands), Tim Verbruggen (former student of Utrecht University), Corine Houtman and Yvonne van Oorschot from Het Waterlaboratorium (Haarlem, The Netherlands), Raymond Pieters, Evie Egelmeers, Ruben Addie and Laila Kuipers from the Hogeschool Utrecht (Utrecht, The Netherlands), Giel Hendriks, Remco Derr and Inger Brandsma from Toxys (Oegstgeest, The Netherlands), Jochen Kuckelkorn, Ralf Junek, Andrea Sehr from the Umweltbundesamt (Bad Elster, Germany), Ron van der Oost, Manon Bechger and Christa Morgenschweis from Waternet (Amsterdam, The Netherlands) and Evelyn the Meyer from De Watergroep (Brussel, Belgium) for their valuable contributions to the project. Furthermore, the authors would like to thank the people that supplied the water samples.

Tables

Table I – Overview of bioassays recommended in different test batteries applied for water quality monitoring (abbreviations on next page)

	DEMEAU (Schriks et al. 2015)	Tox-Box Guideline (Grummt et al. 2018b)	SIMONI (Van der Oost et al. 2017a, Van der Oost et al. 2017b)	GWRC (Neale et al. 2021)	SFT2 (De Baat et al. 2022)[3]	Guidance WWTP effluent (Ecofide, 2023)
Genotoxicity						
Mutagenicity	Ames fluctuation test, ToxTracker	Ames fluctuation test	-	Ames fluctuation assay	Ames test[4]	-
Cytogenetic damage	Micronucleus test, ToxTracker	Micronucleus test	-	-	-	-
DNA damage response	Umu test, Vitotox, p53 CALUX, BlueScreen	Umu test	p53 CALUX	Umu test	-	p53 CALUX
(Anti-)estrogenic activity	ERα CALUX, YES	ERα CALUX, RYES	ER-CALUX	ERα CALUX[1], YES, E-screen, EASZY	ERα CALUX	ERα CALUX
(Anti-)androgenic activity	AR CALUX, AR-MDA-kb2	AR CALUX, RYAS	Anti-AR CALUX	-	AR CALUX, anti-AR CALUX	-
Other hormone-mediated effects	GR CALUX, GR-MDA-kb2	H295R Steroidogenesis Assay for non-receptor bound effects of endocrine disruptors, reproduction test with *Potamopyrgus antipodarum*	GR-CALUX (immune response, metabolism and development)	-	GR CALUX (immune response, metabolism and development), PR CALUX, TTR CALUX (thyroid hormone-mediated effects)	
Oxidative stress	Nrf2 CALUX, AREc32 assay	-	Nrf2 CALUX	AREc32 assay, Nrf2 CALUX[2]	Nrf2 CALUX, AREc32 assay	

	DEMEAU (Schriks et al. 2015)	Tox-Box Guideline (Grummt et al. 2018b)	SIMONI (Van der Oost et al. 2017a, Van der Oost et al. 2017b)	GWRC (Neale et al. 2021)	SFT2 (De Baat et al. 2022)[3]	Guidance WWTP effluent (Ecofide, 2023)
Other toxicological endpoints	H5GLN PXR assay, PXR HepG2 assay, DR CALUX, AhR geneblazer (xenobiotic metabolism), various nuclear receptor assays, H295R assay for developmental toxicity	A test battery for neurotoxicity has been proposed. Details can be found on page 23 of Grummt et al. 2018b.	Microtox, Algaltoxkit, Daphniatoxkit, cytotox CALUX (general toxicity) PXR-CALUX, DR-CALUX, PAH-CALUX, PPARγ-CALUX (metabolism), RIKILT WaterSCAN (antibiotic activity),	AhR CALUX, PAH CALUX (xenobiotic metabolism)	AchE assay (neurotoxicity), algae inhibition, *Dahnia* immobilization, Zebrafish, Microtox, cytotox CALUX (general toxicity), PAH CALUX, PXR CALUX (xenobiotic metabolism), PPARγ CALUX (metabolism)	Microtox (general toxicity) PAH CALUX, PXR CALUX (xenobiotic metabolism)

Abbreviations to Table I:
AhR: aryl hydrocarbon receptor, ARE: antioxidant responsive element, AR: androgen receptor, EASZY: zebrafish embryo assay, ERα: estrogen receptor α, GR: glucocorticoid receptor, Nrf2: nuclear factor erythroid 2-related factor 2, PAH: polycyclic aromatic hydrocarbon, PR: progesterone receptor, PXR: pregnane X receptor, RYAS: (recombinant) yeast androgen screen, IYES: (recombinant) yeast estrogen screen, TTR: transthyretin receptor.

[1] And other mammalian reporter gene assays such as ERα GeneBLAzer, Hela-9903, MELN, T47D-Kbluc

[2] And similar assays such as ARE GeneBLAzer, Nrf2 reporter gene assay, Nrf2-MDA-MB

[3] Recommended assays for specific water types (drinking water, surface water and wastewater) and level (basic set, confirmatory and experimental) can be found in De Baat et al. 2022.

[4] Type not specified

Table II: Bioassays applied for the assessment of genotoxicity and endocrine disruption (search terms of the literature search described in section 2.4.1).

Reference	Water source	Reuse application	Genotoxicity assays	Endocrine activity assays
Bain et al. 2014	Wastewater	Unspecified		ER/AR/GR/PR-CALUX, Anti-ER/AR-CALUX
Chaves et al. 2020	Wastewater	Unspecified		YES
Dopp et al. 2021	Wastewater	Unspecified	Ames, Micronucleus, Comet	ER-CALUX, A-YES
Giebner et al. 2018	Wastewater	Unspecified	Ames	YES, YAS, YAES, YAAS
Hamilton et al. 2016	Wastewater	Unspecified		ERBA, E-screen, Two-hybrid-ER
Houtman et al. 2018	Wastewater	Unspecified		ER/AR/GR/PR-CALUX
Inoue et al. 2011	Wastewater	Unspecified		Two-hybrid-ER/TR
Jia et al. 2015	Wastewater	Unspecified	Ames, Umu	ER/AR/GR/PR-CALUX, Anti-ER/AR/GR/PR-CALUX, YES, YAS, GR-GeneBLAzer, GR-Switchgear
Kusk et al. 2011	Wastewater	Unspecified		YES
Leusch et al. 2014a	Wastewater	Unspecified	Umu	E-screen, AR-CALUX
Leusch et al. 2014b	Wastewater	Groundwater recharge, infiltration in superficial aquifer with downstream abstraction for irrigation	Ames, Micronucleus	ER/AR/GR/PR-CALUX
Lundqvist et al. 2019	Wastewater	Unspecified		VM7Luc4E2, AR-EcoScreen
Mišík et al. 2011	Wastewater	Unspecified	Ames, Micronucleus, Comet	
Papa et al. 2016	Wastewater	Unspecified	Ames, Micronucleus, Comet	E-screen
Ragazzo et al. 2017	Wastewater	Unspecified	Ames, Micronucleus, Comet	
Rao et al. 2014	Wastewater	Unspecified		Two-hybrid-ER/AR/PR
Schilirò et al. 2012	Wastewater	Unspecified		E-screen, MELN
Tang et al. 2014	Wastewater	Grondwater recharge	Umu	
Ternes et al. 2017	Wastewater	Unspecified	Ames	YES, YAS, YAES, YAAS
Välitalo et al. 2017	Wastewater	Unspecified	Umu, p53-CALUX	ER/AR-CALUX
Watson et al. 2012	Wastewater	Urban non-potable reuse and agriculture	Umu	ER/AR-CALUX
Westlund et al. 2017	Wastewater	Unspecified		YES, YAS, YAES, YAAS
Zhang et al., 2019	Wastewater	Unspecified	Ames	

Reference	Water source	Reuse application	Genotoxicity assays	Endocrine activity assays
Etteieb et al., 2015	Wastewater, surface water	Unspecified		E-screen
Neale et al., 2020	Wastewater, surface water	Unspecified		HELN zfER[1], UALH zfAR[1], UMLN zfGR[1], UEL zfPR[1]
Escher et al., 2011	WWTP effluent	Unspecified		E-screen
Li et al., 2011	WWTP effluent	Unspecified		Two-hybrid-ER/AR/TR/PR
Mehinto et al., 2015	WWTP effluent	Unspecified		ER/AR/GR/PR-GeneBLAzer
Jia et al. 2015	WWTP effluent	Groundwater recharge, influent AOP pilot plant	Ames, Umu	ER/AR/GR/PR/TRb-CALUX, Anti-ER/AR/GR/PR-CALUX, YES, GR-GeneBLAzer, Anti-GR-GeneBLAzer, GR-Switchgear, T-screen
Leusch et al. 2014a	WWTP effluent	Unspecified	Ames, Micronucleus	ER/AR/GR/PR/TRβ-CALUX
Comtois-Marotte et al. 2017	WWTP effluent, surface water	Unspecified		YES
Macova et al. 2010	WWTP effluent, surface water	Industrial water	Umu	E-screen
Medlock Kakaley et al. 2020	WWTP effluent, surface water	Drinking water		T47D-Kbluc, MDA-kb2, CV1-hGR
Sun et al. 2016	WWTP effluent, surface water	Surface water quality improvement		MDA-kb2, YES
Leusch et al., 2018	WWTP effluent, surface water, drinking water	Unspecified		ER/AR/GR/PR-GeneBLAzer, GR/PR-CALUX
Escher et al., 2014	WWTP effluent, surface water, stormwater, drinking water	Unspecified	Ames, Micronucleus, SOS chromotest, Umu	ER/AR/GR/PR/TR-CALUX, Anti-ER/AR/GR/PR-CALUX E/T-screen, YES, YAS, ER/AR/GR/PR-GeneBLAzer and more
Janousek et al. 2021	Hospital effluent	Unspecified	Ames, Micronucleus, Comet assay, Chicken egg genotoxicity assay	
Jírova et al. 2019	Hospital effluent	Unspecified	Ames, Micronucleus, Comet	YES/YAS assay
Gehrmann et al. 2018	Hospital wastewater	Unspecified		ER/AR-CALUX, YES, YAS, A-YES/YAS
Perrodin et al. 2013	Hospital wastewater, surface water	Unspecified	Ames, Comet	
Magdaleno et al. 2014	Hospital wastewater, WWTP effluent	Unspecified	Chromosomal abberation, Micronucleus test	
Dias et al. 2015	Surface water	Unspecified		YES
Feretti et al. 2020	Surface water	Drinking water	Ames, Micronucleus, Comet	

Reference	Water source	Reuse application	Genotoxicity assays	Endocrine activity assays
Gou et al. 2016	Surface water	Drinking water		T47D-Kbluc, E-screen
Oskarsson et al. 2021	Surface water	Drinking water	Micronucleus	VM7Luc4E2
Valbonesi et al. 2021	Surface water	Drinking water	Micronucleus	E-Screen
Valcárcel et al. 2018	Surface water	Drinking water		HeLa-9903, PC-DR-Luc
Tang et al. 2013	Stormwater	Unspecified	Umu	E-screen
Neale et al. 2012	Unspecified	Drinking water	Umu	
Shen et al. 2018	Unspecified	Drinking water		Two-hybrid-ER/AR
Van Zijl et al. 2017	Unspecified	Drinking water		T47D-Kbluc, YES

AR: androgen receptor, A-YES: Arxula yeast estrogen assay, CALUX: chemical activated luciferase expression assay, CV1 cell line transduced with the human glucocorticoid receptor (Medlock Kakaley et al. 2020), E-screen: estrogen screen, ER: estrogen receptor, ERBA: estrogen receptor binding assay, GR: glucocorticoid receptor, Hela9903: cervical carcinonoma cell line modified with luciferase gene to study estrogenic receptor activity, MDA-kb2: breast cancer cell line, MELN: estrogen-sensitive human breast cancer cells (MCF-7), PC-DR-Luc: avian PC12 cells modified with luciferase gene to study thyroid receptor activity, PR: progesterone receptor, T47D-Kbluc: breast cancer cell line modified with luciferase gene to study estrogenic receptor activity, T-screen: thyroid screen, TR: thyroid receptor, VM7Luc4E2: breast cancer cell line modified with luciferase gene to study estrogenic receptor activity, YAAS: yeast anti-androgen assay, YAS: yeast androgen assay, YAES: yeast anti-estrogen assay, YES: yeast estrogen assay,

[1] Reporter cell lines to study specific endocrine activities (Neale et al. 2020)

References

2000/60/EC. Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for the Community action in the field of water policy. Official Journal of the European Communities 2000, L327, 1-72.

2013/19/EU. Directive 2013/39/EU of the European Parliament and the Council of August 2013 amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy. Official Journal of the European Communities 2013, L226, 1-17.

Abbas A, Schneider I, Bollmann A, Funke J, Oehlmann J, Prasse C, Schulte-Oehlmann U, Seitz W, Ternes T, Weber M, Wesely H, Wagner M. What you extract is what you see: Optimising the preparation of water and wastewater samples for in vitro bioassays. Water Res. 2019 Apr 1;152:47-60.

Anderson D, Yu TW, McGregor DB. Comet assay responses as indicators of carcinogen exposure. Mutagenesis. 1998 Nov;13(6):539-55.

Angelakis A, Snyder S. 2015. Wastewater Treatment and Reuse: Past, Present, and Future. Water 7(12):4887–95

Arnold SF, Robinson MK, Notides AC, Guillette LJ Jr, McLachlan JA. A yeast estrogen screen for examining the relative exposure of cells to natural and xenoestrogens. Environ Health Perspect. 1996 May;104(5):544-8.

Aydin S, Aydin ME, Ulvi A, Kilic H. Antibiotics in hospital effluents: occurrence, contribution to urban wastewater, removal in a wastewater treatment plant, and environmental risk assessment. Environ Sci Pollut Res Int. 2019 Jan;26(1):544-558.

Bain PA, Williams M, Kumar A. Assessment of multiple hormonal activities in wastewater at different stages of treatment. Environ Toxicol Chem. 2014 Oct;33(10):2297-307.

Baken K, Dingemans M. Effect-based monitoring with bioassays – a roadmap. BTO report 2017.008, KWR Water Research Institute, October 2017.

Beard JE, Bierkens MFP, Bartholomeus RP. Following the Water: Characterising de facto Wastewater Reuse in Agriculture in the Netherlands. Sustainability. 2019; 11(21):5936.

Been F, Kruve A, Vughs D, Meekel N, Reus A, Zwartsen A, Wessel A, Fischer A, Ter Laak T, Brunner AM. Risk-based prioritization of suspects detected in riverine water using complementary chromatographic techniques. Water Res. 2021a Oct 1;204:117612.

Been F, Pronk T, Louisse J, Houtman C, van der Velden-Slootweg T, van der Oost R, Dingemans MML. Development of a framework to derive effect-based trigger values to interpret CALUX data for drinking water quality. Water Res. 2021 Apr 1;193:116859.

Benameur L, Orsière T, Rose J, Botta A. Detection of environmental clastogens and aneugens in human fibroblasts by cytokinesis-blocked micronucleus assay associated with immunofluorescent staining of CENP-A in micronuclei. Chemosphere. 2011 Jul;84(5):676-80.

Benford D, Bolger PM, Carthew P, Coulet M, DiNovi M, Leblanc JC, Renwick AG, Setzer W, Schlatter J, Smith B, Slob W, Williams G, Wildemann T. Application of the Margin of Exposure (MOE) approach to substances in food that are genotoxic and carcinogenic. Food Chem Toxicol. 2010 Jan;48 Suppl 1:S2-24.

Brack W, Ait-Aissa S, Burgess RM, Busch W, Creusot N, Di Paolo C, Escher BI, Mark Hewitt L, Hilscherova K, Hollender J, Hollert H, Jonker W, Kool J, Lamoree M, Muschket M, Neumann S, Rostkowski P, Ruttkies C, Schollee J, Schymanski EL, Schulze T, Seiler TB, Tindall AJ, De Aragão Umbuzeiro G, Vrana B, Krauss M. Effect-directed analysis supporting monitoring of aquatic environments–An in-depth overview. Sci Total Environ. 2016 Feb 15;544:1073-118.

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.

Brandsma I, Moelijker N, Derr R, Hendriks G. Aneugen Versus Clastogen Evaluation and Oxidative Stress-Related Mode-of-Action Assessment of Genotoxic Compounds Using the ToxTracker Reporter Assay. Toxicol Sci. 2020 Sep 1;177(1):202-213.

Buchner EM, Happel O, Schmidt CK, Scheurer M, Schmutz B, Kramer M, Knauer M, Gartiser S, Hollert H. Approach for analytical characterization and toxicological assessment of ozonation products in drinking water on the example of acesulfame. Water Res. 2019 Apr 15;153:357-368.

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.

Cantoni B, Penserini L, Vries D, Dingemans MML, Bokkers BGH, Turolla A, Smeets PWMH, Antonelli M. Development of a quantitative chemical risk assessment (QCRA) procedure for contaminants of emerging concern in drinking water supply. Water Res. 2021 Apr 15;194:116911.

Chaves FP, Gomes G, Della-Flora A, Dallegrave A, Sirtori C, Saggioro EM, Bila DM. Comparative endocrine disrupting compound removal from real wastewater by UV/Cl and UV/H2O2: Effect of pH, estrogenic activity, transformation products and toxicity. Sci Total Environ. 2020 Dec 1;746:141041.

Chen, T., Xu, Y., Liu, Z., Zhu, S., Shi, W., & Cui, F. (2016). Evaluation of drinking water treatment combined filter backwash water recycling technology based on comet and micronucleus assay. Journal of Environmental Sciences (China), 42, 61–70.

Chicharo, Luis, Felix Müller, and Nicola Fohrer. 2015. “Local Pressures and Drivers of Aquatic Ecosystem Change in: Ecosystem Services and River Basin Ecohydrology.” 1:341.

Comtois-Marotte S, Chappuis T, Vo Duy S, Gilbert N, Lajeunesse A, Taktek S, Desrosiers M, Veilleux É, Sauvé S. Analysis of emerging contaminants in water and solid samples using high resolution mass spectrometry with a Q Exactive orbital ion trap and estrogenic activity with YES-assay. Chemosphere. 2017 Jan;166:400-411.

Corvi R, Madia F. In vitro genotoxicity testing-Can the performance be enhanced? Food Chem Toxicol. 2017 Aug;106(Pt B):600-608.

Conner, R., A. Renata, and C. Ortigara. 2017. “The United Nations World Water Development Report, 2017: Wastewater: The Untapped Resource – UNESCO Digital Library.” Retrieved March 23, 2021 (https://unesdoc.unesco.org/ark:/48223/pf0000247153).

Davies JE. Neurotoxic concerns of human pesticide exposures. Am J Ind Med. 1990;18(3):327-31. doi: 10.1002/ajim.4700180314. PMID: 2220839. Dias AC, Gomes FW, Bila DM, Sant’Anna GL Jr, Dezotti M. Analysis of estrogenic activity in environmental waters in Rio de Janeiro state (Brazil) using the yeast estrogen screen. Ecotoxicol Environ Saf. 2015 Oct;120:41-7.

de Baat M. van de Berg S, Pronk T. Het toepassen van bioassays binnen Nederland. Deltafact, versie 1.0, 20 december 2022.

Dingemans MM, Baken KA, van der Oost R, Schriks M, van Wezel AP. Risk-based approach in the revised European Union drinking water legislation: Opportunities for bioanalytical tools. Integr Environ Assess Manag. 2019 Jan;15(1):126-134.

Dieter HH. The relevance of “non-relevant metabolites” from plant protection products (PPPs) for drinking water: the German view. Regul Toxicol Pharmacol. 2010 Mar;56(2):121-5.

Dimitrov SD, Diderich R, Sobanski T, Pavlov TS, Chankov GV, Chapkanov AS, Karakolev YH, Temelkov SG, Vasilev RA, Gerova KD, Kuseva CD, Todorova ND, Mehmed AM, Rasenberg M, Mekenyan OG. QSAR Toolbox – workflow and major functionalities. SAR QSAR Environ Res. 2016 Mar;27(3):203-219.

DIN 38415-1. German standard methods for the examination of water, waste water and sludge – Bio-assays with microorganisms (group T) Determination of organophosphate and carbamate pesticides by the cholinesterase inhibition test (T 1). 1995-02, Berlin, Germany.

Di Paolo C, Ottermanns R, Keiter S, Ait-Aissa S, Bluhm K, Brack W, Breitholtz M, Buchinger S, Carere M, Chalon C, Cousin X, Dulio V, Escher BI, Hamers T, Hilscherová K, Jarque S, Jonas A, Maillot-Marechal E, Marneffe Y, Nguyen MT, Pandard P, Schifferli A, Schulze T, Seidensticker S, Seiler TB, Tang J, van der Oost R,

Vermeirssen E, Zounková R, Zwart N, Hollert H. 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.

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci. 2007 Jan;95(1):5-12.

Dopp, E., Pannekens, H., Itzel, F., & Tuerk, J. (2019). Effect-based methods in combination with state-of-the-art chemical analysis for assessment of water quality as integrated approach. International Journal of Hygiene and Environmental Health, 222(4), 607–614.

Dopp E, Pannekens H, Gottschlich A, Schertzinger G, Gehrmann L, Kasper-Sonnenberg M, Richard J, Joswig M, Grummt T, Schmidt TC, Wilhelm M, Tuerk J. Effect-based evaluation of ozone treatment for removal of micropollutants and their transformation products in waste water. J Toxicol Environ Health A. 2021 May 19;84(10):418-439. Doi: 10.1080/15287394.2021.1881854. Epub 2021 Feb 24. PMID: 33622194.

Dutch drinking water directive, valid from October 13, 2021. wetten.nl – Regeling – Drinkwaterbesluit – BWBR0030111 (overheid.nl).

Dvořáková M, Kejlová K, Bendová H, Rucki M, Kohout P, Vavrouš A, Jírová D. Alternative methods in vitro for screening of endocrine disruptors. Neuro Endocrinol Lett. 2016 Dec 18;37(Suppl1):123-131.

ECHA 2017. Guidance on Information Requirements and Chemical Safety Assessment Chapter R.7a: Endpoint specific guidance, Version 6.0, July 2017.

Ecofide (2023). Handreiking voor het uitvoeren van biologische effectmonitoring bij vergaande zuivering van RWZI-effluenten. Versie 0.8, 8 juni 2023.

EFSA 2011. Scientific opinion on genotoxicity testing strategies applicable to food and feed safety assessment EFSA Journal 2011;9(9):2379.

EFSA 2021. Guidance on aneugenicity assessment. EFSA Journal 2021;19(8):6770.

Ellman GL, Courtney KD, Andres V Jr, Feather Stone RM. A new and rapid colorimetric determination of acetylcholinesterase activity. Biochem Pharmacol. 1961 Jul;7:88-95.

EMA 2012. ICH S2 (R1) Genotoxicity testing and data interpretation for pharmaceuticals intended for human use – Step 5. June 2012.

Escher BI, Aït-Aïssa S, Behnisch PA, Brack W, Brion F, Brouwer A, Buchinger S, Crawford SE, Du Pasquier D, Hamers T, Hettwer K, Hilscherová K, Hollert H, Kase R, Kienle C, Tindall AJ, Tuerk J, van der Oost R, Vermeirssen E, Neale PA. Effect-based trigger values for in vitro and in vivo bioassays performed on surface water extracts supporting the environmental quality standards (EQS) of the European Water Framework Directive. Sci Total Environ. 2018 Jul 1;628-629:748-765.

Escher, B.I., Allinson, M., Altenburger, R., Bain, P.A., Balaguer, P., Busch, W., Crago, J., Denslow, N.D., Dopp, E., Hilscherova, K., Humpage, A.R., Kumar, A., Grimaldi, M., Jayasinghe, B.S., Jarosova, B., Jia, A., Makarov, S., Maruya, K.A., Medvedev, A., Mehinto, A.C., Mendez, J.E., Poulsen, A., Prochazka, E., Richard, J., Schifferli, A., Schlenk, D., Scholz, S., Shiraishi, F., Snyder, S., Su, G., Tang, J.Y., van der Burg, B., van der Linden, S.C., Werner, I., Westerheide, S.D., Wong, C.K., Yang, M., Yeung, B.H., Zhang, X. and Leusch, F.D. (2014) Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays. Environ Sci Technol 48(3), 1940-1956.

Escher BI, Fenner K. Recent advances in environmental risk assessment of transformation products. Environ Sci Technol. 2011 May 1;45(9):3835-47.

Escher BI, Lawrence M, Macova M, Mueller JF, Poussade Y, Robillot C, Roux A, Gernjak W. Evaluation of contaminant removal of reverse osmosis and advanced oxidation in full-scale operation by combining passive sampling with chemical analysis and bioanalytical tools. Environ Sci Technol. 2011b Jun 15;45(12):5387-94.

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

Etteieb S, Kawachi A, Han J, Tarhoun J, Isoda H. Estrogenic activity assessment in environmental matrices by means of in vitro cell-based bioassays coupled with molecular investigation. Clean-Soil, Air, Water 2015, July 2,43(9):1279–1288.

Fields S, Sternglanz R. The two-hybrid system: an assay for protein-protein interactions. Trends Genet. 1994 Aug;10(8):286-92.

Galloway SM, Armstrong MJ, Reuben C, Colman S, Brown B, Cannon C, Bloom AD, Nakamura F, Ahmed M, Duk S, et al. Chromosome aberrations and sister chromatid exchanges in Chinese hamster ovary cells: evaluations of 108 chemicals. Environ Mol Mutagen. 1987;10 Suppl 10:1-175.

Galloway S, Lorge E, Aardema MJ, Eastmond D, Fellows M, Heflich R, Kirkland D, Levy DD, Lynch AM, Marzin D, Morita T, Schuler M, Speit G. Workshop summary: Top concentration for in vitro mammalian cell genotoxicity assays; and report from working group on toxicity measures and top concentration for in vitro cytogenetics assays (chromosome aberrations and micronucleus). Mutat Res. 2011 Aug 16;723(2):77-83.

Gehrmann L, Bielak H, Behr M, Itzel F, Lyko S, Simon A, Kunze G, Dopp E, Wagner M, Tuerk J. (Anti-)estrogenic and (anti-)androgenic effects in wastewater during advanced treatment: comparison of three in vitro bioassays. Environ Sci Pollut Res Int. 2018 Feb;25(5):4094-4104. Erratum in: Environ Sci Pollut Res Int. 2019 Apr 26.

Golovko O, Örn S, Sörengård M, Frieberg K, Nassazzi W, Lai FY, Ahrens L. Occurrence and removal of chemicals of emerging concern in wastewater treatment plants and their impact on receiving water systems. Sci Total Environ. 2021 Feb 1;754:142122.

Gómez L, Niegowska M, Navarro A, Amendola L, Arukwe A, Ait-Aissa S, Balzamo S, Barreca S, Belkin S, Bittner M, Blaha L, Buchinger S, Busetto M, Carere M, Colzani L, Dellavedova P, Denslow N, Escher BI, Hogstrand C, Khan EA, König M, Kroll KJ, Lacchetti I, Maillot-Marechal E, Moscovici L, Potalivo M, Sanseverino I, Santos R, Schifferli A, Schlichting R, Sforzini S, Simon E, Shpigel E, Sturzenbaum S, Vermeirssen E, Viarengo A, Werner I, Lettieri T. Estrogenicity of chemical mixtures revealed by a panel of bioassays. Sci Total Environ. 2021 Sep 1;785:147284.

Grummt et al. 2013. Tox-Box: securing drops of life – an enhanced health-related risk approach for risk assessment of drinking water in Germany. Environmental Sciences Europe 2013, 25:27.

Grummt et al. 2018a. Editorial: Special Issue “Effect-related evaluation of anthropogenic trace substances - concepts for genotoxicity, neurotoxicity and endocrine effects”. Environmental Science and Pollution Research, 25: 3945-3950.

Grummt et al. 2018b. Tox-Box Guideline Hazard-based risk management of anthropogenic trace substances in drinking water to secure a long-term drinking water supply. Leitfaden_TOXBOX_engl.indd (umweltbundesamt.de).

Grummt and Kuckelkorn, 2020. RiSKWa – Verbundprojekt NeuroBox Methodische Weiterentwicklung zur Bewertung von neurotoxischen Effekten im Wasserkreislauf. Methodische Weiterentwicklung der dreistufigen In-vitro-Testbatterie zur Erfassung neurotoxischer Wirkungen und Koordination des Verbundes. https://www.umweltbundesamt.de/sites/default/files/medien/5620/dokumente/bmbf_fkz_02wrs1419a-f_neurobox_schlussbericht_0.pdf.

Grundmann, V., Bilitewski, B., Zehm, A. et al. Risk-based management of chemicals and products in a circular economy at a global scale- Impacts of the FP7 funded project RISKCYCLE. Environ Sci Eur 25, 14 (2013). https://doi.org/10.1186/2190-4715-25-14.

Gyori BM, Venkatachalam G, Thiagarajan PS, Hsu D, Clement MV. OpenComet: an automated tool for comet assay image analysis. Redox Biol. 2014 Jan 9;2:457-65. doi: 10.1016/j.redox.2013.12.020. Erratum in: Redox Biol. 2021 Apr;40:101876.

Hamers T, van den Brink PJ, Mos L, van der Linden SC, Legler J, Koeman JH, Murk AJ. Estrogenic and esterase-inhibiting potency in rainwater in relation to pesticide concentrations, sampling season and location. Environ Pollut. 2003;123(1):47-65.

Han Y, Ma M, Li N, Hou R, Huang C, Oda Y, Wang Z. Chlorination, chloramination and ozonation of carbamazepine enhance cytotoxicity and genotoxicity: Multi-endpoint evaluation and identification of its genotoxic transformation products. J Hazard Mater. 2018 Jan 15;342:679-688.

Hashimoto K, Nakajima Y, Matsumura S, Chatani F. An in vitro micronucleus assay with size-classified micronucleus counting to discriminate aneugens from clastogens. Toxicol In Vitro. 2010 Feb;24(1):208-16.

He Y, Nurul S, Schmitt H, Sutton NB, Murk TAJ, Blokland MH, Rijnaarts HHM, Langenhoff AAM. Evaluation of attenuation of pharmaceuticals, toxic potency, and antibiotic resistance genes in constructed wetlands treating wastewater effluents. Sci Total Environ. 2018 Aug 1;631-632:1572-1581.

Hecker M, Hollert H, Cooper R, Vinggaard AM, Akahori Y, Murphy M, Nellemann C, Higley E, Newsted J, Laskey J, Buckalew A, Grund S, Maletz S, Giesy J, Timm G. The OECD validation program of the H295R steroidogenesis assay: Phase 3. Final inter-laboratory validation study. Environ Sci Pollut Res Int. 2011 Mar;18(3):503-15.

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, Brandsma I, Moelijker N, Rasz PI, Bogaard PJ, Ketelslegers K. Adaptation of the ToxTracker reporter assay for the genetic toxicology assessment of petroleum products. Society of Toxicology 58th Annual Meeting, Baltimore, March 10-14, 2019.

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.

Hockin, A., Reus, A., Verschoor, A. Water Wise -A decision support system for safe water use and reuse. BTO report 2020.030, KWR Water Research Institute, Nieuwegein, June 2020.

Houtman, CJ. 2010. Emerging Contaminants in Surface Waters and Their Relevance for the Production of Drinking Water in Europe. Journal of Integrative Environmental Sciences 7(4):271–95.

Houtman CJ, Brewster K, Ten Broek R, Duijve B, van Oorschot Y, Rosielle M, Lamoree MH, Steen RJCA. Characterisation of (anti-)progestogenic and (anti-)androgenic activities in surface and wastewater using high resolution effectdirected analysis. Environ Int. 2021 Aug;153:106536.

Houtman CJ, Ten Broek R, van Oorschot Y, Kloes D, van der Oost R, Rosielle M, Lamoree MH. High resolution effect-directed analysis of steroid hormone (ant)agonists in surface and wastewater quality monitoring. Environ Toxicol Pharmacol. 2020 Nov;80:103460.

ISO 11350:2012; Water quality — Determination of the genotoxicity of water and waste water — Salmonella/microsome fluctuation test (Ames fluctuation test), June 2012.

ISO 13829:2000; Water quality — Determination of the genotoxicity of water and waste water using the umu-test, March 2000.

ISO 21427-2:2006; Water quality — Evaluation of genotoxicity by measurement of the induction of micronuclei — Part 2: Mixed population method using the cell line V79, November 2016.

Jackson, Robert B., Stephen R. Carpenter, Clifford N. Dahm, Diane M. McKnight, Robert J. Naiman, Sandra L. Postel, and Steven W. Running. 2001. “WATER IN A CHANGING WORLD.” Ecological Applications 11(4):1027–45.

Janion C. Inducible SOS response system of DNA repair and mutagenesis in Escherichia coli. Int J Biol Sci. 2008 Sep 23;4(6):338-44.

Janousek S, Vlkova A, Jirova G, Kejlova K, Krsek D, Jirova D, Kandarova H, Wittlingerova Z, Heinonen T, Mannerstrom M, Maly M. Qualitative and Quantitative Analysis of Certain Aspects of the Cytotoxic and Genotoxic Hazard of Hospital Wastewaters by Using a Range of In Vitro Assays. Altern Lab Anim. 2021 Jan-Mar;49(1-2):33-48

de Jongh CM, Kooij PJ, de Voogt P, ter Laak TL. Screening and human health risk assessment of pharmaceuticals and their transformation products in Dutch surface waters and drinking water. Sci Total Environ. 2012 Jun 15;427-428:70-7.

Jia A, Escher BI, Leusch FD, Tang JY, Prochazka E, Dong B, Snyder EM, Snyder SA. In vitro bioassays to evaluate complex chemical mixtures in recycled water. Water Res. 2015 Sep 1;80:1-11.

Jírova G, Vlkova A, Wittlerova M, Dvorakova M, Kasparova L, Chrz J, Kejlová K, Wittlingerova Z, Zimova M, Hosíkova B, Jiravova J, Kolarova H. Toxicity of wastewater from health care facilities assessed by different bioassays. Neuro Endocrinol Lett. 2019 Jan;39(6):441-453.

Keifer MC, Firestone J. Neurotoxicity of pesticides. J Agromedicine. 2007;12(1):17-25.

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, Speit G. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens III. Appropriate follow-up testing in vivo. Mutat Res. 2008 Jul 31;654(2):114-32.

Kolle SN, Kamp HG, Huener HA, Knickel J, Verlohner A, Woitkowiak C, Landsiedel R, van Ravenzwaay B. In house validation of recombinant yeast estrogen and androgen receptor agonist and antagonist screening assays. Toxicol In Vitro. 2010 Oct;24(7):2030-40.

Kortenkamp A. Low dose mixture effects of endocrine disrupters and their implications for regulatory thresholds in chemical risk assessment. Curr Opin Pharmacol. 2014 Dec;19:105-11.

Kuckelkorn J, Redelstein R, Heide T, Kunze J, Maletz S, Waldmann P, Grummt T, Seiler TB, Hollert H. A hierarchical testing strategy for micropollutants in drinking water regarding their potential endocrine-disrupting effects-towards health-related indicator values. Environ Sci Pollut Res Int. 2018 Feb;25(5):4051-4065.

Legradi JB, Di Paolo C, Kraak MHS, van der Geest HG, Schymanski EL, Williams AJ, Dingemans MML, Massei R, Brack W, Cousin X, Begout ML, van der Oost R, Carion A, Suarez-Ulloa V, Silvestre F, Escher BI, Engwall M, Nilén G, Keiter SH, Pollet D, Waldmann P, Kienle C, Werner I, Haigis AC, Knapen D, Vergauwen L, Spehr M, Schulz W, Busch W, Leuthold D, Scholz S, Vom Berg CM, Basu N, Murphy CA, Lampert A, Kuckelkorn J, Grummt T, Hollert H. An ecotoxicological view on neurotoxicity assessment. Environ Sci Eur. 2018;30(1):46.

Leusch FD, de Jager C, Levi Y, Lim R, Puijker L, Sacher F, Tremblay LA, Wilson VS, Chapman HF. Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters. Environ Sci Technol. 2010 May 15;44(10):3853-60.

Leusch FD, Khan SJ, Laingam S, Prochazka E, Froscio S, Trinh T, Chapman HF, Humpage A. Assessment of the application of bioanalytical tools as surrogate measure of chemical contaminants in recycled water. Water Res. 2014a Feb 1;49:300-15.

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 FD, Khan SJ, Gagnon MM, Quayle P, Trinh T, Coleman H, Rawson C, Chapman HF, Blair P, Nice H, Reitsema T. Assessment of wastewater and recycled water quality: a comparison of lines of evidence from in vitro, in vivo and chemical analyses. Water Res. 2014 Mar 1;50:420-31.

Liu X, Park M, Beitel SC, Lopez-Prieto IJ, Zhu NZ, Meng XZ, Snyder SA. Exploring the genotoxicity triggers in the MP UV/H2O2-chloramination treatment of bisphenol A through bioassay coupled with non-targeted analysis. Sci Total Environ. 2021 May 15;769:145218.

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. Bioanalytical tools for the evaluation of organic micropollutants during sewage treatment, water recycling and drinking water generation. Water Res. 2011 Aug;45(14):4238-47.

Magdaleno A, Juárez AB, Dragani V, Saenz ME, Paz M, Moretton J. Ecotoxicological and genotoxic evaluation of Buenos Aires city (Argentina) hospital wastewater. J Toxicol. 2014;2014:248461.

Malaj, E., Von Der Ohe, P. C., Grote, M., Kühne, R., Mondy, C. P., Usseglio-Polatera, P., … Schäfer, R. B. (2014). Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. Proceedings of the National Academy of Sciences of the United States of America, 111(26), 9549–9554.

Martijn BJ, Van Rompay AR, Penders EJ, Alharbi Y, Baggelaar PK, Kruithof JC, Rietjens IM. Development of a 4-NQO toxic equivalency factor (TEF) approach to enable a preliminary risk assessment of unknown genotoxic compounds detected by the Ames II test in UV/H_2O_2 water treatment samples. Chemosphere. 2016 Feb;144:338-45.

Marty MS, Borgert C, Coady K, Green R, Levine SL, Mihaich E, Ortego L, Wheeler JR, Yi KD, Zorrilla LM. Distinguishing between endocrine disruption and non-specific effects on endocrine systems. Regul Toxicol Pharmacol. 2018 Nov;99:142-158.

Marx-Stoelting P, Rivière G, Luijten M, Aiello-Holden K, Bandow N, Baken K, Cañas A, Castano A, Denys S, Fillol C, Herzler M, Iavicoli I, Karakitsios S, Klanova J, Kolossa-Gehring M, Koutsodimou A, Vicente JL, Lynch I, Namorado S, Norager S, Pittman A, Rotter S, Sarigiannis D, Silva MJ, Theunis J, Tralau T, Uhl M, van Klaveren J, Wendt-Rasch L, Westerholm E, Rousselle C, Sanders P. A walk in the PARC: developing and implementing 21st century chemical risk assessment in Europe. Arch Toxicol. 2023 Mar;97(3):893-908.

Mišík M, Knasmueller S, Ferk F, Cichna-Markl M, Grummt T, Schaar H, Kreuzinger N. Impact of ozonation on the genotoxic activity of tertiary treated municipal wastewater. Water Res. 2011 Jun;45(12):3681-91.

Mohrenweiser HW, Jones IM. Review of the molecular characteristics of gene mutations of the germline and somatic cells of the human. Mutat Res. 1990 Jul;231(1):87-108.

Murk AJ, Legler J, van Lipzig MM, Meerman JH, Belfroid AC, Spenkelink A, van der Burg B, Rijs GB, Vethaak D. Detection of estrogenic potency in wastewater and surface water with three in vitro bioassays. Environ Toxicol Chem. 2002 Jan;21(1):16-23.

Neale PA, Altenburger R, Aït-Aïssa S, Brion F, Busch W, de Aragão Umbuzeiro G, Denison MS, Du Pasquier D, Hilscherová K, Hollert H, Morales DA, Novák J, Schlichting R, Seiler TB, Serra H, Shao Y, Tindall AJ, Tollefsen KE, Williams TD, Escher BI. Development of a bioanalytical test battery for water quality monitoring: Fingerprinting identified micropollutants and their contribution to effects in surface water. Water Res. 2017 Oct 15;123:734-750.

Neale P, Leusch F, Escher B. Effect Based Monitoring in Water Safety Planning, WP3.2: Medium-to-high throughput bioanalytical tools and decision-making tool for selection of bioassays. Reports - GWRC (globalwaterresearchcoalition.net).

Neale PA, Leusch FD. Considerations when assessing antagonism in vitro: Why standardizing the agonist concentration matters. Chemosphere. 2015 Sep;135:20-3.

Nohmi T. Thresholds of Genotoxic and Non-Genotoxic Carcinogens. Toxicol Res. 2018 Oct;34(4):281-290

Oda Y, Nakamura S, Oki I, Kato T, Shinagawa H. Evaluation of the new system (umu-test) for the detection of environmental mutagens and carcinogens. Mutat Res. 1985 Oct;147(5):219-29.

OECD 2022. OECD Test Guidelines for Chemicals. OECD Test Guidelines for Chemicals - OECD

OECD 456. H295R Steroidogenesis Assay. 30 June 2022.

OECD 471. Bacterial Reverse Mutation Test. 29 June 2020.

OECD 476. In Vitro Mammalian Cell Gene Mutation Tests using the Hprt and xprt genes. 29 July 2016.

OECD 487. In Vitro Mammalian Cell Micronucleus Test. 29 July 2016.

OECD 489. In Vivo Mammalian Alkaline Comet Assay. 29 July 2016.

OECD 490. In Vitro Mammalian Cell Gene Mutation Tests Using the Thymidine Kinase Gene. 28 July 2015.

Oliveira TS, Murphy M, Mendola N, Wong V, Carlson D, Waring L. Characterization of Pharmaceuticals and Personal Care products in hospital effluent and waste water influent/effluent by direct-injection LC-MS-MS. Sci Total Environ. 2015 Jun 15;518-519:459-78.

PARC 2022. European Partnership for the Assessment of Risks from Chemicals (PARC) | Anses - Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail.

Perrodin Y, Christine B, Sylvie B, Alain D, Jean-Luc BK, Cécile CO, Audrey R, Elodie B. A priori assessment of ecotoxicological risks linked to building a hospital. Chemosphere. 2013 Jan;90(3):1037-46. Posthuma L, de Zwart D, Osté L, van der Oost R, Postma J. Ecologische sleutelfactor toxiciteit. Deel 1: Methode voor het in beeld brengen van de effecten van giftige stoffen in oppervlaktewater. September 2016. STOWA 2016-15A.pdf

Prantl EM, Kramer M, Schmidt CK, Knauer M, Gartiser S, Shuliakevich A, Milas J, Glatt H, Meinl W, Hollert H. Comparison of in vitro test systems using bacterial and mammalian cells for genotoxicity assessment within the "health-related indication value (HRIV) concept. Environ Sci Pollut Res Int. 2018 Feb;25(5):3996-4010.

Prasse, C., Stalter, D., Schulte-Oehlmann, U., Oehlmann, J., & Ternes, T. A. T. A. (2015). Spoilt for choice: A critical review on the chemical and biological assessment of current wastewater treatment technologies. Water Research, 87, 237–270.

Pronk, G.J., Stofberg, S.F., Van Dooren, T.C.G.W. *et al.* Increasing Water System Robustness in the Netherlands: Potential of Cross-Sectoral Water Reuse. *Water Resour Manage* 35, 3721–3735 (2021) Pronk, G.J., Stofberg, S.F., Van Dooren, T.C.G.W. et al. Increasing Water System Robustness in the Netherlands: Potential of Cross-Sectoral Water Reuse. Water Resour Manage 35, 3721–3735 (2021)

Quillardet P, Huisman O, D'Ari R, Hofnung M. SOS chromotest, a direct assay of induction of an SOS function in Escherichia coli K-12 to measure genotoxicity. Proc Natl Acad Sci U S A. 1982 Oct;79(19):5971-5.

Reifferscheid G, Heil J. Validation of the SOS/umu test using test results of 486 chemicals and comparison with the Ames test and carcinogenicity data. Mutat Res. 1996 Aug 12;369(3-4):129-45..

Reifferscheid G, Heil J, Oda Y, Zahn RK. A microplate version of the SOS/umu-test for rapid detection of genotoxins and genotoxic potentials of environmental samples. Mutat Res. 1991 Dec;253(3):215-22.

Reifferscheid G, Maes HM, Allner B, Badurova J, Belkin S, Bluhm K, Brauer F, Bressling J, Domeneghetti S, Elad T, Flückiger-Isler S, Grummt HJ, Gürtler R, Hecht A, Heringa MB, Hollert H, Huber S, Kramer M, Magdeburg A, Ratte HT, Sauerborn-Klobucar R, Sokolowski A, Soldan P, Smital T, Stalter D, Venier P, Ziemann C, Zipperle J, Buchinger S. International round-robin study on the Ames fluctuation test. Environ Mol Mutagen. 2012 Apr;53(3):185-97.

Reifferscheid G, Oepen Bv. Genotoxicity and mutagenicity of suspended particulate matter of river water and waste water samples. ScientificWorldJournal. 2002 Apr 18;2:1036-9..

Reifferscheid G, Ziemann C, Fieblinger D, Dill F, Gminski R, Grummt HJ, Hafner C, Hollert H, Kunz S, Rodrigo G, Stopper H, Selke D. Measurement of genotoxicity in wastewater samples with the in vitro micronucleus test: results of a round-robin study in the context of standardisation according to ISO. Mutat Res. 2008 Jan 8;649(1-2):15-27.

Reus A, Hoondert R, Yanagihara M. Evaluation of QSAR tools in combination with bioassays for transformation products and emerging substances. BTO report 2023.015, KWR Water Research Institute, Nieuwegein, February 2023a.

Reus A, Hoondert R, Majid S. Effectsignaalwaarden voor eenduidige interpretatie van genotoxiciteitstesten. BTO report 2023.052, KWR Water Research Institute, Nieuwegein, June 2023b.

Reus A, Dingemans M, Hoondert R, Houtman C, van der Oost, R. QSAR en read-across-modellen voor waterkwaliteit. H2O Online, 18 maart 2022a.

Reus A, Pronk T, Dingemans M. Bioassays voor neuroactieve stoffen in water. BTO report 2020.035, KWR Water Research Institute, Nieuwegein, August 2020a.

Reus A, Pronk T, ter Laak T, Dingemans M. Bioassays voor neuroactieve stoffen in water. H2O Online, 21 August 2020b.

Reus et al. Poster presentation International Congress of Toxicology 2022, Effect-based monitoring of water quality and human health risk assessment of water in the circular economy, Toxicology letters 368S1 (2022b) S14-S79 pp. 546.

Rogers JM, Denison MS. Recombinant cell bioassays for endocrine disruptors: development of a stably transfected human ovarian cell line for the detection of estrogenic and anti-estrogenic chemicals. In Vitr Mol Toxicol. 2000 Spring;13(1):67-82.

Rosenmai AK, Lundqvist J, le Godec T, Ohlsson Å, Tröger R, Hellman B, Oskarsson A. In vitro bioanalysis of drinking water from source to tap. Water Res. 2018 Aug 1;139:272-280Shao Y, Hollert H, Tarcai Z, Deutschmann B, Seiler TB. Integrating bioassays, chemical analysis and in silico techniques to identify genotoxicants in surface water. Sci Total Environ. 2019 Feb 10;650(Pt 2):3084-3092.

Salazar AM, Sordo M, Ostrosky-Wegman P. Relationship between micronuclei formation and p53 induction. Mutat Res. 2009 Jan 31;672(2):124-8.

Schriks M, Murk T, Hofs B, Wols B. DPWE robuustheid: toepassen van bioassays. KWR report 2016/018, April 2016.

Schriks M, Baken K, Simon E, Besselink H, van der Linden S, Kienle C, van der Burg B. Selection criteria to select in vitro bioassays for implementation and use. October 2015. D41 1_FINAL 19102015 (MS).pdf (demeau-fp7.eu)

SFT2, 2022. Basis-set bioassays NL (sleutelfactortoxiciteit.nl).

Shibai-Ogata A, Kakinuma C, Hioki T, Kasahara T. Evaluation of high-throughput screening for in vitro micronucleus test using fluorescence-based cell imaging. Mutagenesis. 2011 Nov;26(6):709-19.

Singh NP, McCoy MT, Tice RR, Schneider EL. A simple technique for quantitation of low levels of DNA damage in individual cells. Exp Cell Res. 1988 Mar;175(1):184-91.

Sommer S, Buraczewska I, Kruszewski M. Micronucleus Assay: The State of Art, and Future Directions. Int J Mol Sci. 2020 Feb 24;21(4):1534.

Sonneveld E, Jansen HJ, Riteco JA, Brouwer A, van der Burg B. Development of androgen- and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid-responsive bioassays. Toxicol Sci. 2005 Jan;83(1):136-48.

Srivastav AL, Patel N, Chaudhary VK. Disinfection by-products in drinking water: Occurrence, toxicity and abatement. Environ Pollut. 2020 Dec;267:115474.

Stang A, Witte I. Performance of the comet assay in a high-throughput version. Mutat Res. 2009 Apr 30;675(1-2):5-10.

Sykora P, Witt KL, Revanna P, Smith-Roe SL, Dismukes J, Lloyd DG, Engelward BP, Sobol RW. Next generation high throughput DNA damage detection platform for genotoxic compound screening. Sci Rep. 2018 Feb 9;8(1):2771.

Tang JY, Aryal R, Deletic A, Gernjak W, Glenn E, McCarthy D, Escher BI. Toxicity characterization of urban stormwater with bioanalytical tools. Water Res. 2013 Oct 1;47(15):5594-606.

Ternes, Thomas A., Carsten Prasse, Christian Lütke Eversloh, Gregor Knopp, Peter Cornel, Ulrike Schulte-Oehlmann, Thomas Schwartz, Johannes Alexander, Wolfram Seitz, Anja Coors, and Jörg Oehlmann. 2017. “Integrated Evaluation Concept to Assess the Efficacy of Advanced Wastewater Treatment Processes for the Elimination of Micropollutants and Pathogens.” Environmental Science and Technology 51(1):308–19.

Tice RR, Agurell E, Anderson D, Burlinson B, Hartmann A, Kobayashi H, Miyamae Y, Rojas E, Ryu JC, Sasaki YF. Single cell gel/comet assay: guidelines for in vitro and in vivo genetic toxicology testing. Environ Mol Mutagen. 2000;35(3):206-21.

Tilson HA, MacPhail RC, Crofton KM. Defining neurotoxicity in a decision-making context. Neurotoxicology. 1995 Summer;16(2):363-75.

Timmers PHA, Slootweg T, Knezev A, van der Schans M, Zandvliet L, Reus A, Vughs D, Heijnen L, Knol T, El Majjaoui J, van der Wielen P, Stuyfzand PJ, Lekkerkerker-Teunissen K. Improved drinking water quality after adding advanced oxidation for organic micropollutant removal to pretreatment of river water undergoing dune infiltration near The Hague, Netherlands. J Hazard Mater. 2022 May 5;429:128346.

Unüvar, T., & Büyükgebiz, A. (2012). Fetal and neonatal endocrine disruptors. Journal of Clinical Research in Pediatric Endocrinology, 4(2), 51–60.

UNWWAP. 2015. “The United Nations Wold Water Development Report: Water for a Sustainable World.” UNESCO.

Valentin-Severin I, Le Hegarat L, Lhuguenot JC, Le Bon AM, Chagnon MC. Use of HepG2 cell line for direct or indirect mutagens screening: comparative investigation between comet and micronucleus assays. Mutat Res. 2003 Apr 20;536(1-2):79-90.

van der Burg B, van der Linden S, Man H, Winter R, Jonker L, van Vugt-Lussenburg, Brouwer A. A Panel of Quantitative Calux® Reporter Gene Assays for Reliable High-Throughput Toxicity Screening of Chemicals and Complex Mixtures. High-Throughput Screening Methods in Toxicity Testing. Chapter 28. First published 13 February 2013.

van der Linden SC, Heringa MB, Man HY, Sonneveld E, Puijker LM, Brouwer A, Van der Burg B. Detection of multiple hormonal activities in wastewater effluents and surface water, using a panel of steroid receptor CALUX bioassays. Environ Sci Technol. 2008 Aug 1;42(15):5814-20.

van der Linden SC, von Bergh AR, van Vught-Lussenburg BM, Jonker LR, Teunis M, Krul CA, van der Burg B. Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress. Mutat Res Genet Toxicol Environ Mutagen. 2014 Jan 15;760:23-32.

van der Oost R, Sileno G, Suárez-Muñoz M, Nguyen MT, Besselink H, Brouwer A. 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.

Voulvoulis, N. Water reuse from a circular economy perspective and potential risks from an unregulated approach. Curr. Opin. Environ. Sci. Health 2, 32 (2018). https://doi.org/10.1016/j.coesh.2018.01.005

Wilson VS, Bobseine K, Gray LE Jr. Development and characterization of a cell line that stably expresses an estrogen-responsive luciferase reporter for the detection of estrogen receptor agonist and antagonists. Toxicol Sci. 2004 Sep;81(1):69-77.

Xu J, Wei D, Wang F, Bai C, Du Y. Bioassay: A useful tool for evaluating reclaimed water safety. J Environ Sci (China). 2020 Feb;88:165-176.

Zhang S, Wang X, Yang H, Xie YF. Chlorination of oxybenzone: Kinetics, transformation, disinfection byproducts formation, and genotoxicity changes. Chemosphere. 2016 Jul;154:521-527.

Zwart N, Nio SL, Houtman CJ, de Boer J, Kool J, Hamers T, Lamoree MH. High-Throughput Effect-Directed Analysis Using Downscaled in Vitro Reporter Gene Assays To Identify Endocrine Disruptors in Surface Water. Environ Sci Technol. 2018 Apr 3;52(7):4367-4377

Zwart N, Jonker W, Broek RT, de Boer J, Somsen G, Kool J, Hamers T, Houtman CJ, Lamoree MH. Identification of mutagenic and endocrine disrupting compounds in surface water and wastewater treatment plant effluents using high-resolution effect-directed analysis. Water Res. 2020 Jan 1;168:115204.

Annex I – Description of bioassay methods

Ames fluctuation test (KWR)

The potential of a substance to induce gene mutations can be detected by the bacterial reverse mutation (Ames) test, for which a modified version of the classical Ames test has been developed and validated for water samples (Reifferscheid et al. 2002). In the current project, the Ames fluctuation test was performed as previously (Heringa et al., 2011), with minor modifications (Timmers et al. 2022). In brief, *Salmonella typhymurium* strains TA98 and TA100 (Xenometrix, Switzerland) were exposed to three dilutions of the water extracts in culture medium (REF 50, 100 and 200). Dilutions of the water extracts were tested in the presence and absence of an exogenous metabolic activation system (rat liver S9 mix), resulting in four different test conditions (TA98-S9, TA98+S9, TA100-S9 and TA100+S9). Negative controls (extracts of Evian mineral water), solvent controls (DMSO) and appropriate positive controls were run in parallel. A single experiment was performed in triplicate cultures. After an incubation period of 90 minutes at 37°C and gently shaking on a rocking platform (250 rpm), cultures were incubated with indicator medium for another 48 hours at 37°C. The number of yellow wells per 48 wells per culture were counted manually. Cytotoxicity was measured in parallel by measuring optical density at 595 nm before and after the 90 minute-exposure according to the International Organization for Standardization (ISO) protocol 11350:2012. An individual response was considered positive for genotoxicity if the response of the sample was different from the procedure control with a certainty of 99%, based on a binominal distribution (Heringa et al., 2011). In addition to statistical significance, biological relevance was taken into account to interpret the results. This implies evaluation of a concentration-response and cytotoxicity, where sample dilutions showing >50% cytotoxicity compared to the negative control were regarded as severely cytotoxic. Severely cytotoxic concentrations were not taken into consideration for mutagenicity evaluation to exclude false negative responses.

Umu test (KWR)

The umu test is based on the principle of activation of the umuC gene as a response to genetic damage, where the umuC gene is coupled to a gene coding for the β-galactosidase enzyme in the genetically engineered *S. typhimurium strain* TA1535/pSK1002. Expression of the β-galactosidase enzyme is quantified using a colorimetric measurement (Oda et al. 1985). In the current project, the umu test was performed based on the protocol provided with the test kit (EBPI Environmental Bio-Detection Products Inc., Canada) and based on ISO 13829:2000. In brief, *S. typhimurium* strain TA 1535 pSK1002 were exposed to four dilutions of the water extracts in sterile water (REF 25, 50, 100 and 200). Dilutions of the water extracts were tested in the presence and absence of S9, resulting in two different test conditions (-S9 and +S9). Negative controls (extracts of Evian mineral water), solvent controls (DMSO) and appropriate positive controls were run in parallel. A single experiment was performed in triplicate cultures. After an incubation period of in total 4 hours at 37°C, bacterial growth was determined by measuring optical density at 620 nm. Then the plates were incubated for 30 min at 37°C and the induction of the umuC-gene was determined by measuring optical density at 405 nm. An individual response was considered positive if the induction ratio (IR) was >1.5. Cytotoxicity was assessed in parallel by evaluating the growth factor. Both the IR and growth factor were obtained from a calculation sheet provided by the supplier of the test kit. In addition to numerical significance, biological relevance was taken into account to interpret the results. This implies evaluation of a concentration-response and cytotoxicity, where sample dilutions showing a growth factor ≤0.5 were regarded as severe cytotoxic to avoid interpretation of false positive responses.

Micronucleus test (UBA)

Cytogenetic damage can be detected with the micronucleus test (Sommer et al. 2020), for which a modified version has been developed and validated for water samples (Reifferscheid et al. 2008). In the current project, the micronucleus test was performed according to OECD TG 487 (2016) and ISO 21427-2 (2006). In brief, HepG2 liver cells were exposed to dilutions of the water samples in culture medium (REF varying from 5 to 200, depending on the response of the sample in the first experiment). Negative controls (extracts of Evian mineral water), solvent controls (DMSO) and appropriate positive controls were run in parallel. Three biologically independent replicates (experiments) were performed and each replicate was carried out in quadruple technical replicates. After an incubation period of 72 hours at 37°C, cell nuclei/micronuclei were sequentially stained with two different dyes (chromatin of dead/dying cells and vital cells). Fluorescence was detected and cell nuclei and micronuclei were separately counted with flow cytometry and the number of micronuclei was used as a parameter for genotoxicity. A response was considered genotoxic if there was a 2-fold increase of micronuclei compared to the negative control. Cytotoxicity was assessed in parallel by calculation of the cell proliferation rate based on cell nuclei counts and in comparison with the negative control. In addition to numerical significance, biological relevance was considered to interpret the results. This implies the assessment of a concentration-response relationship and cytotoxicity, with sample dilutions showing a cell proliferation rate <50% of the negative control being considered as highly cytotoxic to avoid interpretation of false positive responses.

Comet assay (HU)

The comet assay or single cell electrophoresis assay provides another method to investigate both gene mutations and cytogenetic damage (Singh et al. 1988) and can also be applied to water samples (Heringa et al. 2011, Rosenmai et al. 2008). The principle of the comet assay is the migration of DNA fragments in an agarose gel during electrophoresis, where a tail is formed by small DNA fragments and intact DNA of the cell nuclei forms a head. The comet assay can be performed with any cell type and allows rapid detection of genetic damage such as single and double DNA strand breaks, alkali labile sites and incomplete excision repair sites (Tice et al. 2000). In the current project, the comet assay was performed as previously (Heringa et al. 2011). In brief, HepG2 liver cells were exposed to a single dilution of the water extract (REF 100). Negative controls (extracts of Evian mineral water), solvent controls (DMSO) and appropriate positive controls were run in parallel. A single experiment was performed in duplicate cultures. The comet assay was performed both after 3h and 24h exposure to be able to detect direct as well as indirect acting (metabolism requiring) genotoxins, assuming that HepG2 cells are metabolically competent and that a 24h exposure period is sufficient for pro-genotoxins to be metabolized into genotoxins. These two exposure periods are in line with comet assays in HepG2 cells performed by others (Heringa et a. 2011, Valentin-Severin et al. 2003) and are also in line with the OECD guideline of the *in vivo* comet assay (OECD 489). In the present study, after an incubation period for 3 hours or 24 hours (*i.e.* two independent experiments) at 37 °C, a single cell suspension was prepared, mixed 0.5% low melting agarose and added to a microscopic slide pre-coated with 1.5% normal melting agarose. After overnight lysis at 2-10 °C and 30 minutes unwinding in ice-cold electro-phoresis buffer, electrophoresis was performed in ice-cold electrophoresis buffer for 30 minutes at 25 V (0.83 V/cm) and ± 300 mA. After neutralization of the slides, slides were dehydrated in ethanol and allowed to dry. For comet analysis, slides were stained with SYBR Green and evaluated using a fluorescence microscope and Comet Score (TriTek Corp, USA). At least 50 cells per slide and two slides per culture were evaluated, resulting in a total of at least 200 analyzed cells per water sample. The median tail intensity (%DNA in the tail) was calculated per slide and averaged for the two slides per culture, followed by calculation of the mean of the two slides per water sample. The water extracts were considered positive for genotoxicity when a three-fold increase in tail intensity over the negative control was observed. Cytotoxicity of the water samples to HepG2 cells was assessed after 3 hours and 24 hours of exposure by fluorometric determination of alamarBlue™ conversion. In addition to numerical significance, biological relevance was taken into account to interpret the results. This implies evaluation of cytotoxicity, where samples showing a >50% cytotoxicity compared to the negative control were regarded as severely cytotoxic to avoid interpretation of false positive responses. The 50% cut-off was adopted from Buchner et al. (2019), Prantl et al. (2018),

p53 CALUX (HWL)

The p53 CALUX is a human osteosarcoma cell line (U2OS)-derived pathway selective reporter gene assay. In this assay, a luciferase gene has been coupled to p53 responsive elements to detect activation of the tumor suppressing gene TP53, which play a role in regulation or progression through the cell cycle, apoptosis, and genomic stability by means of several mechanisms. Levels of p53 are indicative of genotoxicity, as the p53 protein responds to DNA damage, and is a transcription factor for genes related to DNA-damage repair, cell-cycle arrest and apoptosis. In the p53 CALUX, the p53 responsive elements have been coupled to a luciferase gene (Van der Linden et al. 2014). In the current project, the p53 CALUX® reporter gene assay (BioDetection Systems B.V., Amsterdam, the Netherlands) was performed as described by Van der Linden (2014) and as laid down in the protocols of the supplier. In the experiments, p53 CALUX cells were exposed to five dilutions of the water samples (REF 1, 3, 10, 33, 100) for 24 hours in the presence and absence of S9, resulting in two different test conditions (-S9 and +S9). Negative controls (extracts of Evian mineral water), solvent controls (DMSO) and appropriate positive controls were run in parallel. A single experiment was performed in triplicate cultures. The reference compounds were actinomycin D (-S9) and cyclophosphamide (+S9) and a ten-point calibration curve of the reference compound was included on each plate. Responses of dilutions of the water were expressed as fold induction above the response of the solvent control (DMSO). A sigmoidal curve was fitted for the dilution response (induction ratio) series. With this curve, the dilution was calculated corresponding to an induction ration of 1.5. With this value, the genotoxic activity in a sample was calculated and expressed as equivalents of the reference compounds per liter water. Cytotoxicity was measured in parallel with the cytotox CALUX® (Van der Linden et al. 2014), for which tributyltin acetate (TBT) was used as reference compound. Sample dilutions were considered positive for cytotoxicity if they induced <80% of the luminescence of that induced by the solvent control (DMSO).

ToxTracker (Toxys)

ToxTracker is a green fluorescent protein based genotoxicity assay consisting of different genetically engineered mouse embryonic stem (mES) reporter cell lines having the green fluorescent protein (GFP) gene coupled to a variety of reporter genes involved in oxidative stress and the cellular response to different direct and indirect genotoxic effects. ToxTracker is also able to provide insight into the primary toxic properties of compounds through integrated evaluation of the results from the different reporter cells in the test (Hendriks et al. 2012). In the current project, the ToxTracker assay was performed as previously described (Baken and Dingemans, 2017; Hendriks et al. 2012). In brief, the six different reporter cell lines were exposed to five dilutions of the water samples (REF 6.25, 12.5, 25, 50 and 100) in the presence and absence of 0.25% S9 and required co-factors (RegenSysA+B, Moltox), resulting in two different test conditions (-S9 and +S9). Negative controls (extracts of Evian mineral water), solvent controls (DMSO) and the standard relevant positive controls were run in parallel. Three independent repeat experiments were performed. Induction of the green fluorescent protein (GFP) reporters was determined after 24 hours exposure using a flow cytometer. Only GFP expression in intact single cells was determined. Mean GFP fluorescence was measured and used to calculate GFP reporter induction compared to a vehicle control treatment. Cytotoxicity was estimated by cell count after 24 hours exposure using a flow cytometer and was expressed as percentage of intact cells after 24 hours exposure compared to solvent exposed controls. Relative GFP induction was calculated compared to the solvent control and a 2-fold increase or more was considered as positive response for each marker. If activation of the Bscl2-GFP and/or Rtkn-GFP marker above 2 was observed, the test substance was classified as genotoxic. In addition to numerical significance, biological relevance was taken into account to interpret the results, where sample dilutions showing >25% cytotoxicity compared to the negative control were regarded as cytotoxic to avoid interpretation of false positive responses. For cytotoxicity assessment in the ToxTracker assay, the relative cell survival for the six different reporter cell lines was averaged, because the levels were very similar, and normalized to Evian mineral water. Cytotoxicity (%) was calculated by subtracting the relative survival rate from 1 and multiplying by 100.