DEMEAU

Selection criteria to select *in vitro* bioassays for implementation and use

Picture (optional)

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no. 308339 for the research project DEMEAU

DEMEAU

Title: Selection criteria to select *in vitro* bioassays for implementation and use

Summary: Bioanalytical tools hold great promise in being introduced and integrated in current water monitoring strategies which mainly utilize chemical analytics at present. The latter focuses mainly on compound identification irrespective of the biological effect. Bioassays are suitable to assess hazards even of complex mixtures of pollutants with limited possibilities of chemical identification and are therefore expected to be highly complementary to modern chemical analytical methods. Smart combinations of chemical- and biological analytics therefore can lead to reduced uncertainty in safety assessments at lower costs. This document is aimed at the selection of a minimal panel of *in vitro* bioassays for cost effective comprehensive toxicity screening for the evaluation of drinking water quality. After determination of selection criteria for such assays and the most relevant toxicological effects of concern, the appropriate bioassay panel is identified to detect human health effects of emerging (micro)pollutants. The project has started with identifying the most relevant toxicological endpoints or modes of action that will be used as a starting point for the effect-based assessment of water quality, in particular for human health. A set of selection criteria has been defined to assess the effect-based assays whether they are suitable to detect activity towards the selected endpoints. Using the selection criteria we established a state-of-the art bioassay panel for comprehensive screening of the water cycle and including drinking water safety assessment.

Grant agreement no:	308339
Work Package:	WP41
Deliverable number:	D41.2
Partner responsible:	BDS
Deliverable author(s):	M.Schriks (KWR), K. Baken (KWR), E. Simon[1] (BDS), H. Besselink (BDS), S. van der Linden[2] (BDS), C. Kienle (EAWAG), B. van der Burg (BDS)
Quality assurance:	Armelle Hebert (VEOLIA)
Planned delivery date:	31 August 2014
Actual delivery date:	October 2015
Dissemination level:	PU = Public

[1]Currently employed at EAWAG
[2]Currently employed at JRC

© 2012 DEMEAU
This Demonstration project 'Demonstration of promising technologies to address emerging pollutants in water and waste water' (DEMEAU) has received funding from the European Union's Seventh Programme for Research, Technological Development and Demonstration under Grant Agreement no. 308330. All rights reserved. No part of this book may be reproduced, stored in a database or retrieval system, or published, in any form or in any way, electronically, mechanically, by print, photograph, microfilm or any other means without prior written permission from the publisher.
This publication reflects only the author's views and the European Union is not liable for any use that may be made of the information contained therein.

Table of contents

LIST OF FIGURES III

LIST OF TABLES IV

SUMMARY 1

1 INTRODUCTION AND PURPOSE OF THE CURRENT DOCUMENT 2

2 IDENTIFICATION OF RELEVANT TOXICITY PATHWAYS TO INCLUDE IN AN ASSAY PANEL FOR SAFETY ASSESSMENT OF DRINKING WATER 4

2.1 Pathway selection 5

3 OVERVIEW OF THE BIOASSAY SELECTION CRITERIA 8

3.1 Assay applicability (maximum 21 points) 10

3.1.1 Applied to environmental samples 10
3.1.2 Validated to water samples 10
3.1.3 Standardized protocol available 10
3.1.4 Service and support available 11
3.1.5 Costs 11
3.1.6 Ease of use (max 6 points) 11

3.2 Assay performance (max 33 points) 12

3.3 Selectivity 12

3.3.1 Accuracy 12
3.3.2 Reproducibility 12
3.3.3 Robustness 13
3.3.4 Sensitivity 13
3.3.5 Specificity 13
3.3.6 Limit of detection 13
3.3.7 Cytotoxicity control 14
3.3.8 Quick 14
3.3.9 Clear/Straightforward read-out 14
3.3.10 High-throughput capacity 14

4 BIOASSAY PANEL SELECTION 15

4.1 Xenobiotic metabolism 15

4.2 Hormone-mediated MoA 17

4.2.1 Estrogenic activity 17
4.2.2 Androgen receptor 19
4.2.3 Glucocorticoid activity 20

4.3 Reactive modes of action 22

4.4 Adaptive stress response and oxidative stress response pathway 25

4.5 Reproductive and developmental toxicity 27

5 DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS 30

REFERENCES 35

List of Figures

Figure 1: Adverse outcome pathway (AOP) structure depicting the realms of *in vitro* and *in vivo* assays, site of interaction with toxicant (initiating event) and site of typical adverse outcome (adapted from OECD, 2012 and Ankley *et al*., 2010). 4

List of Tables

Table 1: Selected endpoints for monitoring health effects of drinking water. 6
Table 2: Set of selected criteria to evaluate bioassays (i.e. scoring matrix) 9
Table 3: Evaluation of a number of available bioanalytical tools for PXR receptor agonists 16
Table 4: Evaluation of a number of available bioanalytical tools for AhR receptor agonists 17
Table 5: Evaluation of a number of available bioanalytical tools for estrogenic activity 19
Table 6: Evaluation of a number of available bioanalytical tools for (anti)androgenic activity 20
Table 7: Evaluation of a number of available bioanalytical tools for glucocorticoid activity 21
Table 8: Evaluation of a number of available bioanalytical tools for gene mutations 23
Table 9: Evaluation of a number of available bioanalytical tools for chromosomal mutations 24
Table 10: Evaluation of a number of available bioanalytical tools for DNA damage response 25
Table 11: Evaluation of a number of available bioanalytical tools for oxidative stress 26
Table 12: Evaluation of a number of available bioanalytical tools for reproductive and developmental toxicity 29
Table 13: Overview of promising *in vitro* bioassays for water quality determination 31

Summary

Bioanalytical tools hold great promise in being introduced and integrated in current water monitoring strategies which mainly utilize chemical analytics at present. The latter focuses mainly on compound identification irrespective of the biological effect. Bioassays are suitable to assess hazards even of complex mixtures of pollutants with limited possibilities of chemical identification and are therefore expected to be highly complementary to modern chemical analytical methods. Smart combinations of chemical- and biological analytics therefore can lead to reduced uncertainty in safety assessments at lower costs. This document is aimed at the selection of a minimal panel of *in vitro* bioassays for cost effective comprehensive toxicity screening for the evaluation of drinking water quality. After determination of selection criteria for such assays and the most relevant toxicological effects of concern, the appropriate bioassay panel is identified to detect human health effects of emerging (micro)pollutants. The project has started with identifying the most relevant toxicological endpoints or modes of action that will be used as a starting point for the effect-based assessment of water quality, in particular for human health. A set of selection criteria has been defined to assess the effect-based assays whether they are suitable to detect activity towards the selected endpoints. Using the selection criteria we established a state-of-the art bioassay panel for comprehensive screening of the water cycle and including drinking water safety assessment.

1 Introduction and purpose of the current document

Currently, the evaluation of water quality mainly relies on the chemical analysis of a selection of single compounds. The actual limit values for these compounds are determined by different methods and may depend on the scope and/or species of interest. At present, the WHO and the US Environmental Protection Agency (US EPA) have derived approximately 125 statutory guideline values for drinking water (US EPA, 2006; WHO, 2011; Schriks *et al*., 2010,). One of the problems with the current methods of evaluating water quality is that the scope is very limited. Many compounds that are present in the aquatic environment are not analyzed and for the compounds that are analyzed, toxicological information is often lacking or insufficient for risk assessment purposes. As a result, limit values are sometimes based on analytical detection limits rather than a toxicological assessment of the compounds under investigation. An representative example is the WHO guideline value of 10 µg/L for arsenic which is partly based on an (obsolete) analytical detection limit. In addition, while drinking water sources contain complex mixtures of chemicals, analytical chemistry does not account for combined effects.

There is an increasing understanding of the pathways (also referred to as modes or mechanisms of action) by which toxic compounds can exert their biological effects. By analyzing the perturbation of a diverse set of pathways, insight can be obtained in the possibility that adverse effects are caused by compounds present in the environment. Bioassays are ideal tools for analyzing these effects in extracts of water samples relatively fast, and cost-effectively, detecting all the compounds affecting the pathways included in the assay panel. *In vivo* assays measure effects of bioactive compounds on parameters such as growth, carcinogenesis, development, reproduction, feeding activity and mortality in test species from different trophic levels (e.g. algae, macrophytes, invertebrates, rodents and fish), while *in vitro* assays measure specific cellular effects of bioactive compounds using cell cultures. To assess ecotoxicological effects, simple *in vivo* models are often used, such as algae, invertebrates and bacterial sensors. For drinking water safety assessment, and thus human risk assessment no such simple *in vivo* models are available. Models used for risk assessment of chemicals generally use mammals such as rodents and rabbits, with a focus on assessment of acute or sub-acute toxic effects, carcinogenesis and mutagenesis and reproductive toxicity (including developmental toxicity). Methods and toxicological endpoints assessed in risk assessment have been laid down in legislation such as REACH (**R**egistration, **E**valuation, **A**uthorisation and Restriction of **Ch**emicals), and preferentially employ OECD (**O**rganisation for **E**conomic **C**o-operation and **D**evelopment) guideline studies. These standard *in vivo* methods used in chemical risk assessment are expensive and time consuming. Therefore, a distinct shift can be recognized towards utilizing *in vitro* models and extrapolating their outcome to *in vivo* responses, avoiding the sacrifice of mammals. An additional advantage of most *in vitro* tests is that they do not require large amounts of sample material (Escher and Leusch, 2012). Moreover, the relevance for human health outcomes are higher when *in vitro* assays are used that focus at key events in pathways that are mechanistically relevant to and predictive of adverse effect in humans (Adler *et al*. 2011). It is, however, increasingly recognized that bioassay thresholds (trigger values) are required to put results into perspective. Although, recent work has made progress in *in vitro* to *in vivo* extrapolation (Punt *et al*., 2013), it is still difficult to predict *in vivo* effects with *in vitro* bioassays. Therefore

in vitro bioassays are better suited to provide insight in potential exposure, thus to be used as quantitative tools for hazard assessment.

2 Identification of relevant toxicity pathways to include in an assay panel for safety assessment of drinking water

Water contaminants can elicit effects by interacting with critical cellular targets such as receptors or other constituents of cells like proteins, DNA or phospholipids. Such interactions can trigger a range of cellular events like the activation of genes, production of proteins, and altered protein signaling. This way, a (series of) pathway(s) can be activated by contaminant exposure. Toxicity pathways are defined as the pathways activated as part of the cellular response after chemical exposure (Collins *et al*., 2008). The concept of toxicity pathways is put into a wider (eco)toxicological perspective as adverse outcome pathways (AOPs) (Ankley *et al*., 2010), linking the toxicity pathway at the cellular level to the response at the organ level and via the response of the organism ultimately to the response at the population level.

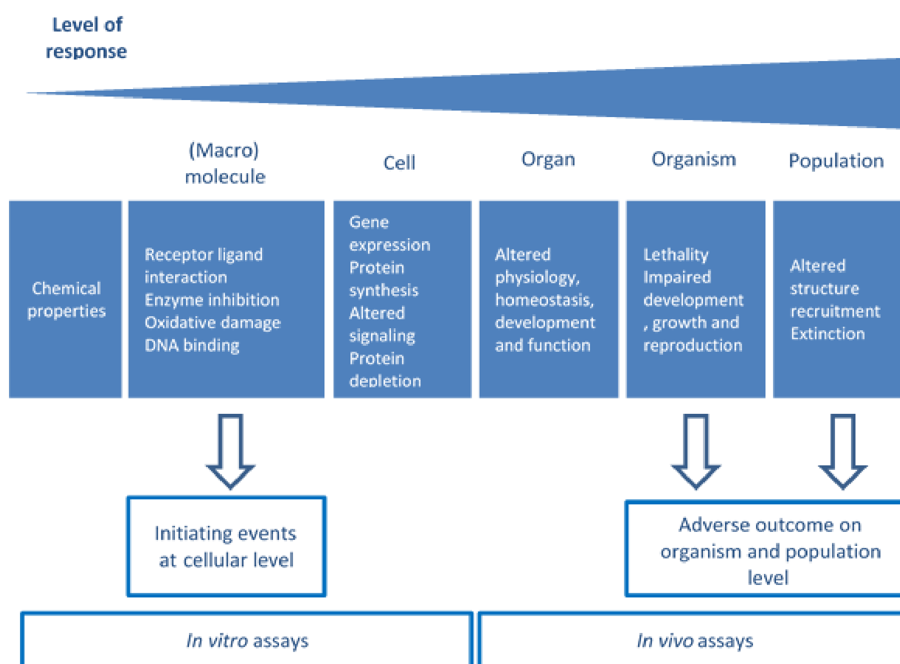

Figure 1: Adverse outcome pathway (AOP) structure depicting the realms of *in vitro* and *in vivo* assays, site of interaction with toxicant (initiating event) and site of typical adverse outcome (adapted from OECD, 2012 and Ankley *et al*., 2010).

The AOP concept has been shown to be applicable for a series of molecular initiating events, including narcosis, photo-activated systemic toxicity, and activation of several receptors including the aryl hydrocarbon receptor (AhR) and the estrogen receptor (ER) (Ankley *et al*., 2010). Many of the toxicity pathways are relatively conserved throughout the animal kingdom (Gunnarsson *et al*., 2008), which allows

a certain level of extrapolation to be made between species, the extent of which may vary with the level of conservation of different pathways. These adverse outcome pathways are currently being developed further to facilitate extrapolation of *in vitro* bioassay results to adverse outcomes as measured in animal experiments. A term almost equivalent to adverse outcome pathways is the mode of action (MoA) of chemicals, a term frequently used in older literature including much of the relevant literature in the area of water quality monitoring. Since most AOPs are currently being constructed we will refer to the older, more generic term pathway of toxicity in the remainder of the document.

Toxicological safety of drinking water (including its sources) and waste water can be assessed using bioassays that measure specific pathways. The pathways can be classified according to the type and degree of interaction taking place between a compound and its biological target (Escher and Hermens, 2002). These targets can be very specific, e.g. specific binding to a nuclear receptor, or more generic by reacting with endogenous molecules, e.g. oxidation of lipids or reaction with proteins or DNA bases. It should be realized, however, that activation of a certain toxicity pathway is not just an intrinsic property of a compound but can differ between species, organs or even tissues in the same organism. In addition, compounds can activate multiple pathways, which may vary with dose, exposure time (chronic versus acute) and timing of exposure (prenatal, infant, adult). In a practical, cost-effective setting, all of these different situations cannot be assessed even in elaborate testing in *in vivo* bioassays. A more modern approach, therefore, which has recently be developed makes use of the knowledge on the pathways of toxicity known to be activated by toxic chemicals. Evidence is accumulating that a limited set of the major pathways can be used to assess toxicity of chemicals (Ankley *et al*., 2010), and chemical mixtures. Activation of such pathways can be assessed in modern *in vitro* bioassays, like the CALUX assays developed in the FP6 integrated project Techneau (Legler *et al*., 2002; Sonneveld *et al*., 2005; van der Linden *et al*., 2008). Many of these assays which have been developed, can be run in an automated high throughput mode, facilitating cost effective and rapid measurements. However, for further cost reduction and efficiency, it is important to select the most relevant pathways activated by potential pollutants in source, and drinking water. Obviously, an important starting point is the inclusion of major toxicological effects and related pathways involved in general toxicity, mutagenesis, carcinogenesis, reproductive and developmental toxicity.

2.1 Pathway selection

A recent inter-laboratory study (Escher *et al*., 2014) performed a comprehensive screening of a representative set of water samples (n=10, including wastewater treatment plant effluent, different recycled waters, storm water, surface and drinking waters) covering a broad range of toxicological effects by using 103 unique *in vitro* bioassays. The main aim of the study was to identify the most relevant modes of action in water quality monitoring. Each water type showed a characteristic bioanalytical profile with particular groups of toxicity pathways either consistently responsive or not responsive across test systems. The most responsive health-relevant toxicity pathways were related to xenobiotic metabolism, hormone-mediated MoA, reactive MoA, and the adaptive stress response. These endpoints were thus demonstrated to be suitable to benchmark water quality. In addition, effects on these toxicity pathways in *in vitro* assays are predictive of adverse human health effects *in vivo* (see Chapter 5). We therefore used the same set of

endpoints to select a purpose-tailored panel of bioassays for routine water quality monitoring (see Table 1). Besides, we evaluated *in vitro* bioassays for developmental and reproductive toxicity, since this is a sensitive endpoint that is relevant for chronic exposure to low doses of chemicals. This is a very complex endpoint involving many pathways. However, it has been shown recently, that panels of relatively simple assays can be used for correct predictions of chemically induced reproductive toxicity (Kroese *et al*., 2015). General (non-specific) toxicity was not included in the evaluation of assays, since this endpoint is usually assessed by measuring cytotoxicity as an control. However, as this endpoint is generally assessed in all assays it can be used as an independent endpoint relevant for acute toxicity, but also for more specific endpoints like reproductive- and developmental toxicity (Van der Burg *et al*., 2015). As stated earlier we used the selection made by Escher *et al*. (2014) as a starting point, but expanded the number of pathways and assays initially, to avoid bias introduced by the relatively small number of water samples used to select pathways by Escher *et al*. We included assays for pathways that were frequently activated (or repressed) by chemicals, using data of an inventory of more than 300 chemicals including a large number of chemicals listed on the Water Framework Directive (WFD), and some additional assays that are used in the standard panel of high throughput CALUX assays. We also evaluated *in vitro* bioassays for developmental toxicity (part of reproductive toxicity), since this is an endpoint that is relevant for chronic exposure to low doses of chemicals, and important in chemical risk assessment. General (non-specific) toxicity was also included in the evaluation of assays, since this pathway is usually assessed by measuring cytotoxicity, but also includes pathways referred to as adaptive stress response pathways by Escher *et al*. (2014).

Table 1: Selected endpoints for monitoring health effects of drinking water.

Toxicity endpoints relevant for drinking water monitoring	Specific pathway
Xenobiotic metabolism	Activation of the pregnane X receptor (PXR) Activation of the aryl hydrocarbon receptor (AhR)
Hormone-mediated modes of action	Estrogenicity (ER) Anti-androgenicity (anti-AR) Glucocorticoid activity (GR)
Reactive modes of action	Gene mutations Chromosomal mutations DNA damage response
Adaptive stress response pathways	ER stress Heat shock Hypoxia Inflammation Metal stress

	Oxidative stress response
Developmental and reproductive toxicity	Pre-implantation toxicity Range of mechanistic assays still in early stage of validation Non-mechanistic assays including early life stages

3 Overview of the bioassay selection criteria

Comparing various bioassays is a demanding exercise due to the variability in availability of information in the (grey) literature. This leads to a degree of uncertainty making individual scores sometimes unreliable. Nevertheless, because of the great detail of analysis and additional factors such as the need of generation of panels with comparable and quality controlled assays, a reasonably well funded choice could be made. To determine whether a bioassay is applicable for the assessment of the chemical water quality, several criteria have to be met. A range of projects (such as ToxCast) have already focused on the - largely overlapping – assay characteristics and selection criteria for bioassays to be used in environmental monitoring. The following criteria are considered to be of high importance when selecting bioassays for water quality monitoring (Table 2). A segregation is made between “applicability” and “performance” of assays, since end-users may assign a different weight to these individual primary criteria. Since “ease of use” is considered as a key-criterion, it is divided in 6 sub-criteria which are individually scored. All criteria are scored for each assay (*c.f.*Chapter 4) and the scores of both sets of criteria are added separately to obtain an impression of the assay applicability and assay performance. The criteria are explained in more detail below. It is possible that certain criterion could not be scored, since adequate information was absent. In that case **Not Available** (NA) was scored. It should be noted that this scoring is quite arbitrary and for certain applications weights may be altered. For instance, for measuring the removal of trace amounts of contaminants in a water treatment process, it is essential that an assay is sensitive enough for reliable measurement. This may preclude certain assays that otherwise score well. In particular, very often yeast-based bioassays are less sensitive that mammalian cell-based, making them less useful or even inappropriate for the application mentioned.

Table 2: Set of selected criteria to evaluate bioassays (i.e. scoring matrix)

Primary criteria	**Sub-criteria**	**Max Points**	**Assay X**	**Assay Y**	**Assay Z**
Assay applicability and ease of use (max 21 points)	Applied to environmental samples	3			
	Validated to water samples	3			
	Standardized protocol available	3			
	Service and support available	3			
	Costs	3			
	Ease of use (*maximum 6 points based on criteria as indicated below*)	6			
	• *Non-GMO*[1]	1			
	• *No specialised equipment/skills required*	1			
	• *Automation possible*	1			
	• *Non-licensed (cell) in vitro model*	1			
	• *Kit available*	1			
	• *Training available*	1			
	Total score	**21**			
Assay performance (max 33 points)	Selectivity	3			
	Accuracy	3			
	Reproducibility	3			
	Robustness	3			
	Sensitivity	3			
	Specificity	3			
	Limit of Detection (LOD)	3			
	Cytotoxicity control	3			
	Quick	3			
	Clear/Straightforward read-out	3			

[1] Genetically Modified Organism

	High-throughput capacity	3			
	Total score	**33**			
	Total maximum score	**54**			

3.1 Assay applicability (maximum 21 points)

An important aspect of using assays within water quality monitoring is related to the applicability of the steps involved. The test has to have a certain level of maturity to gain confidence in the results analyzed. Additionally, a method may be of little use if it can only be performed at a limited number of laboratories because complicated or expensive equipment is needed or it can only be performed by very specialized laboratories because of the complicated procedures involved. This section should provide a summary regarding the applicability of the assays, with special attention to the following topics.

3.1.1 Applied to environmental samples

Most of the today available bioassays are demonstrated to be used and properly working for screening of pure compounds, but might fail in the screening of complex environmental samples. The main aim is to select bioassays for the relevant toxic endpoints to be used in water quality analysis. Therefore, the criterion whether the bioassay has been carried out for the assessment of any environmental samples (especial of waters) is critical. If the assay is applied widely within environmental research (e.g. the ER-CALUX), a maximum score of 3 points can be assigned. If the assay is used sporadically within environmental research, a total of 2 points can be assigned. If there is no track-record of usage within environmental research, only 1 point is assigned.

3.1.2 Validated to water samples

The validation of the selected bioassay to water samples is an important – but sometimes overlooked – step towards the successful application of bioanalytical tools for water quality assessment. The goal of the validation is to demonstrate that the proposed assays produce consistent results, regardless of the type of water and the laboratory performing the assay so that there is enough confidence in the results produced. The validity of the results of an assay can be assessed by a variety of measures of performance, such as accuracy. If the assay has been validated for the use with water samples, a maximum score of 3 points can be assigned. If the assay is in the process of validation, 2 points can be assigned. If the assay did not undergo any formal validation, 1 point can be assigned.

3.1.3 Standardized protocol available

The maturity of an assay is a qualitative assessment of the degree of validation and/or standardization. The assay should be widely used and performed in a standardized way. Preferably, an inter-laboratory validation has been performed (ISO, OECD, DIN). Are historical data available to determine nominative variability, trends, and possibly acceptable and unacceptable conditions? Have test guidelines been agreed

upon internationally, and is the test close to regulatory acceptance? Does the test address important practical needs of end users, concerns of consumers? If the assay completely fulfills the criteria associated with standardization, a maximum of 3 points can be assigned. If the assay is currently undergoing standardization, 2 points can be assigned. If the assay has not been standardized (e.g. scientific publication), 1 point can be assigned.

3.1.4 Service and support available

Along the development, application and/or validation of a bioassay to particular needs/water sample types, the existence and availability of service and support platform is fundamental to overcome any quality, regulatory or security challenges. A maximum of 3 points can be assigned in the case the supplier provides service and support. If there is no service and/or support (e.g. freely available *in vitro* models described in scientific literature), 1 points can be assigned.

3.1.5 Costs

Equipment and reagents required for performing *in vitro* bioassays can range from standard laboratory equipment to highly specialized or custom made materials and equipment. Users may either use their own, preferred equipment or need to buy specific products that suit the assay or kit of choice. In addition, samples may have to be processed separately or can be assessed in bulk. All these aspects contribute to the costs of an assay. If an assay is very cheap to perform (<100 Euro/sample) (such as the E-screen, a maximum of 3 points can be assigned). If the assay costs between 100-1000 Euro/sample, 2 points can be assigned. For relatively expensive assays (>1000 Euro/sample) only 1 point can be assigned.

3.1.6 Ease of use (max 6 points)

This criterion aims to assess how easy it is to use the assay by gaining insight in a number of sub-criteria which are explained in the table below. Since “ease of use” is an important sub-criterion for end-users, is has been sub-divided into 6 individual scoring criteria which together make up the total score (6 points in total).

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

3.2 Assay performance (max 33 points)

This group of criteria is established for evaluating the method design and possible problems that could lead to misleading results.

3.3 Selectivity

To use a bioassay for complex mixtures that are present in the environment, it is important that the test responds specifically to the pathway of interest. The selectivity of an assay quantifies how much a test is affected by the presence of other, non-relevant compounds present in complex mixtures like water sample extracts. For each assay it must be stated what are the known risks of matrix interference and whether the analysis can be affected by the presence of other compounds. Therefore, this has to be determined for different types of water samples. The selectivity is usually expressed as a percentage, with a selectivity close to 100% indicating the assay is only responding to compounds of interest without responding to other compounds. An assay with very high selectivity (>90%) scores a maximum of 3 points. An assay with a relatively poor selectivity (<20%) scores only 1 point.

3.3.1 Accuracy

The accuracy of a test describes the closeness of agreement between test method results and accepted reference values. When validating *in vitro* assays, the accuracy is generally determined by repeated analyses of known concentrations of the reference compound. A highly accurate assay scores a maximum of 3 points.

3.3.2 Reproducibility

The reproducibility describes the agreement among results obtained from testing the same substance or samples (usually 10 or more) using the same test protocol, but with the analysis performed by different people, on different days and even different locations. If samples are tested, the concentrations need to represent the full range of expected concentrations in water samples. The reproducibility can be assessed at different levels:

- The inter-laboratory reproducibility is a measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results. Inter-laboratory reproducibility is determined during validation processes, and indicates the extent to which a test can be successfully transferred between laboratories and hence is also referred to as between-laboratory reproducibility. Generally, the inter-laboratory reproducibility should be below 30%
- The intra-laboratory reproducibility (or within-laboratory reproducibility) is a measure of how well qualified people within the same laboratory can successfully replicate results using a specific protocol at different times. Generally, the intra-laboratory reproducibility should be below 20%. This aspect is related to precision and repeatability, which express how close individual measurements of the same sample are when the analysis is repeated several times under identical

conditions. The value for repeatability should be close to 1 (or 100%). Repeatability and reproducibility are subject to both random and systematic errors (variability). Assays that fulfill all the criteria as indicated above, can score a maximum of 3 points.

3.3.3 Robustness

The robustness of an assay characterizes the sensitivity of a method to operational variation and thus assesses the transferability of a method to other people and laboratories. It gives an indication of the ability of the assay to produce reliable results under slightly varying conditions, e.g. exposure time and temperature. The value for robustness is calculated as inter- and intra-laboratory reproducibility. An assay that scores high on robustness can score a maximum of 3 points.

3.3.4 Sensitivity

The sensitivity of an assay quantifies the proportion of all positive/active substances that are correctly classified by the test. It is a measure of accuracy for a test method when considering categorical results, and its ability to correctly identify positive samples, but it does not take into account the concentrations needed for the positive response. As for bioassays the type and number of positive compounds in samples are generally unknown, this value is usually assessed using a large number of known positive and negative compounds. Assays with low sensitivity may produce false negative results, which is undesirable for screening assays. Ideally, the sensitivity should be close to 1 (or 100%). An assay that correctly identifies (>90%) positive samples can score a maximum of 3 points. An assay with a relatively poor sensitivity (<20%) scores only 1 point.

3.3.5 Specificity

The specificity denotes the proportion of all negative/inactive substances that are correctly classified by the assay. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method. Assays that have a low specificity produce many false positive results, which can be problematic when being used an assay as a screening method. The value for specificity should be close to 1 (or 100%). An assay that has a high specificity (>90%) can score a maximum of 3 points. An assay with a relatively poor specificity (<20%) scores only 1 point.

3.3.6 Limit of detection

The limit of detection (LOD) denotes the minimum amount of activity that can still reliably be detected (within the limits defined for reproducibility and repeatability), but without necessarily being quantified. Generally, the LOD is calculated by interpolating the first significantly different response (signal from the blank + 3x the standard deviation of the blank) in the dose-response curve of the reference compound. Related is the limit of quantification (LOQ), which is similarly calculated but using the response from the blank + 10x standard deviation of the blank. In addition to the LOD, the EC50 value of the reference compound is also indicative for the sensitivity of the assay. The LOQ should be below the trigger value that is proposed for potential human health effects. An assay with a low LOD (<10 ng/L equivalents) can score a maximum of 3 points. If the LOD is relatively poor (>1 µg/L) only 1 point can be assigned.

3.3.7 Cytotoxicity control

Cytotoxicity may mask/hamper the interpretation of the bioassay result. Therefore, the assessment and understanding of the effects of any chemical entity on cellular performance is a necessary and compulsory activity in cell-based screening techniques. If the assay is accompanied by a cytotoxicity control, a maximum of 3 points can be assigned.

3.3.8 Quick

Even if a bioassay produces a result that is very significant from a toxicological point of view, but the analysis itself takes too much time to be used in practice, the assay will not contribute to the safeguarding of the water quality. Therefore, it is important to assess the time to result, i.e. the time frame from taking a sample to having the final results, including and excluding sample pre-treatment. Does the assay provide the information quickly enough to initiate effective management action before unacceptable damage has occurred? The allowed time frame will greatly depend on the type of analysis, endpoint and phase in drinking water preparation. For this criterion, a score of 3 was applied when the test yields results within a day, a score of 2 for results within a week, and 1 if the test takes more than a week to perform.

3.3.9 Clear/Straightforward read-out

The read-out of a bioassay - the recorded observation - can vary from being very general (carcinogenicity, lethality) to very specific (activity on a specific receptor). If the read-out is very straightforward to interpret (e.g. relative light units or optical densities) a score of 3 can be assigned. If the read-out requires a lot of handling before interpretation is possible (e.g. radio ligand binding assays) are score of 1 can be assigned.

3.3.10 High-throughput capacity

In order to perform rapid and cost-effective profiling of the bioactivity of chemicals of unknown toxicity and make predictions about their potential for causing adverse effects, the high-throughput screening capacity of the assays is very important. Using robotics, automated sample workup, miniaturized assay formats, liquid handling devices, sensitive detectors, high-speed plate readers, data processing and control software facilitates the generation of large number of individual assay data points, makes the screening more efficient, and reduces the analyzing costs. The better the assay is suited to be performed in high-throughput, the more points can be a assigned (a maximum of 3 points can be assigned). If the assay is very laborious (such as the classical Ames test), only 1 point can be assigned.

4 Bioassay panel selection

For each relevant endpoint (*c.f.* Chapter 2, Table 1) a number of *in vitro* bioassays were compared using the above described criteria. A set of assays that are representative for the different types of available *in vitro* tools for human health assessment, but not entirely exhaustive, were evaluated. Applicability and performance of bioassays were compared within each endpoint, and all bioassays were ranked from 1 (poor), 2 (good) to 3 (excellent) for each of the selection criteria. Scores were based on information obtained from literature or producers or else on expert judgment. The scores for assay applicability and assay performance are added separately. Bioassays (maximum three per endpoint) with the highest total score were selected accordingly and described in more detail below. It should be noted, however, that in practical use, a panel with similar assays generally is superior to one with more heterogeneous selected assays. Procedures with a homogeneous panel can be made efficient and automated, and quality control is facilitated in many ways.

4.1 Xenobiotic metabolism

The liver and the intestine are sites of major metabolic activity for both endogenous and exogenous chemicals (Lemaire *et al*., 2004). Xenobiotic metabolism gives an indication of exposure to bioactive chemicals, which are recognized for their ability to induce the transcription of genes encoding biotransformation enzymes and xenobiotic transporters in mammalian organisms. These biotransformation process are mainly regulated by steroidal (such as estrogen and androgen) and non-steroidal (such as constitutive androstane receptor=CAR, pregnane X receptor=PXR, and the peroxisome proliferator-activated receptor=PPARs) nuclear receptors. The aryl hydrocarbon receptor (AhR) also functions to regulate a battery of genes encoding biotransformation function. (Omiecinski *et al*., 2011; Kohle and Bock, 2009). The most common nuclear receptors responsive to typical water contaminants and involved in the induction of drug metabolizing enzymes include the PXR (CYP3A4 enzyme) and AhR (CYP1A2 enzyme) (Escher *et al*., 2014; Omiecinski *et al*., 2011).

The role of PXR is ever-expanding as ligands are continuing to be added to an already wide range of structurally diverse lipophilic ligands, which include steroids, vitamins, oxysterols, bile acids, and numerous triazin pesticides (promethryn, terbuthryn, terbutylazine), pharmaceuticals (fenofibrate, bezafibrate, clonazepam, medazepam) and non-coplanar polychlorobiphenyls (PCBs; PCB101, 138, 180) (Creusot *et al*., 2010; Lin *et al*., 2009).

Table 3: Evaluation of a number of available bioanalytical tools for PXR receptor agonists

		PXR-cisFACTORIAL	PXR-transFACTORIAL	HG5LN PXR	PXR-CALUX	PXR reporter gene assay HepG2
Assay applicability	**Applied to environmental samples**	2	2	3	NA	2
	Validated to water samples	1	1	2	1	2
	Standardized protocol available/maturity	2	2	3	1	2
	Service and support available	1	1	2	3	2
	Costs	1	1	2	2	2
	Ease of use TOTAL	2	2	3	4	3
	Non-GMO	0	0	0	0	0
	No specialised skills/equipment required	1	1	1	1	1
	Automation possible	NA	NA	1	1	1
	Non-licensed (cell) in vitro model	1	1	1	1	1
	Kit available	0	0	0	0	0
	Training availabilities	0	0	NA	1	NA
Score		9	9	15	11	13
Assay performance	**Selectivity**	NA	NA	2	NA	2
	Accuracy	NA	NA	NA	NA	2
	Reproducibility	NA	NA	3	NA	3
	Robustness	NA	NA	3	NA	3
	Sensitivity	NA	NA	3	NA	NA
	Specificity	1	2	2	NA	NA
	LOD	NA	NA	NA	NA	NA
	Cytotoxicity control	3	3	3	3	3
	Quick	2	2	2	2	2
	Clear/Straightforward read-out	1	1	3	3	3
	High-throughput capacity	2	2	3	3	3
Score		9	10	24	11	21
Total NA		7	7	3	8	4
Total score		25	26	42	30	38

The most prominent AhR agonists are dioxins and furans. The binding of these coplanar compounds to the AhR ultimately may result in various adverse effects, such as hepatotoxicity, immunotoxicity, neurotoxicity, dermal and reproductive toxicity, as well as teratogenicity and carcinogenicity (Behnisch *et al*., 2001; Wahl *et al*., 2010). Further on, PCBs, polycyclic aromatic hydrocarbons (PAHs), polybrominated dipneyl ethers (PBDEs) and a wide range of other halogenated compounds are known to interact with the AhR.

Table 4: Evaluation of a number of available bioanalytical tools for AhR receptor agonists

		DR/PAH-CALUX	AhR yeast	AhR-GeneBlazer	AhR-cisFACTORIAL	DART Cyp 1a induction
Assay applicability	Applied to environmental samples	3	2	3	1	3
	Validated to water samples	2	1	1	1	2
	Standardized protocol available/maturity	3	2	2	2	3
	Service and support available	3	2	2	2	1
	Costs	2	3	1	1	2
	Ease of use TOTAL	4	4	4	2	3
	Non-GMO	0	0	0	0	0
	No specialised skills/equipment required	1	1	1	1	1
	Automation possible	1	1	1	NA	1
	Non-licensed (cell) in vitro model	1	1	1	1	1
	Kit available	0	0	1	0	0
	Training availabilities	1	1	NA	0	NA
Score		17	14	13	9	14
Assay performance	Selectivity	3	2	NA	NA	2
	Accuracy	3	2	3	NA	NA
	Reproducibility	3	2	3	NA	NA
	Robustness	3	2	2	NA	NA
	Sensitivity	3	1	NA	NA	1
	Specificity	3	2	NA	NA	NA
	LOD	3	1	NA	NA	2
	Cytotoxicity control	3	3	3	3	3
	Quick	2	2	2	2	2
	Clear/Straightforward read-out	3	3	3	1	1
	High-throughput capacity	3	2	3	1	1
Score		32	22	19	7	12
Total NA		0	0	5	8	5
Total score		49	36	37	24	31

4.2 Hormone-mediated MoA

The endocrine system, which regulates and plays a crucial role in the maintenance of homeostasis, sexual development, metabolism, growth and behavior, is known to be vulnerable to water contamination. Among all hormone-mediated modes of action estrogenicity, anti-androgenicity and glucocorticoids activity seem to be the most relevant endpoints for water quality monitoring. At present, thyroid receptor interactions were not observed, which does not include the interferences with relevant binding proteins involved in the thyroid hormone pathway.

4.2.1 Estrogenic activity

Estrogens are the female steroid sex hormones that are involved (among other functions) in the development of female secondary sexual characteristics. In females, they are mainly produced by the

ovaries, although other organs like liver, adrenal gland and fat cells produce estrogens as well. Estrogens are produced by all vertebrates and have an essential role in fetal development. The actions of estrogens are mediated by the estrogen receptor (ER), which is activated by compounds that bind properly to the receptor inducing hormonal effects in animals and humans. Estrogenic chemicals interfere with the synthesis, metabolism, binding or cellular responses of the natural estrogens. Many compounds exist in the environment that are known to influence the estrogen pathway, either agonistically or antagonistically. Known active compounds include natural (17β-estradiol [E2] and estrone [E1]) and synthetic hormones (17α-ethinylestradiol [EE2] and estriol [E3]), pharmaceuticals (e.g. carbamazepine, sulfamethoxazole) industrial chemicals (e.g. nonylphenol, bisphenol A, 4-t octylphenol and benzyl butyl phthalate, personal care products (e.g. triclosan) and pesticides (e.g. bentazone and mecoprop) (Leusch *et al*., 2010; Brand *et al*., 2013). Recent profiling of over 3000 compounds (mainly pesticides) by the US EPA ToxCast project showed that approximately 3% of the compounds were ERα agonists, while over 5% of the compounds showed signs of antagonistic activity (Huang *et al*., 2011). Exposure to estrogenic chemicals may lead to abnormalities in the development and maintenance of feminine characteristics *in vivo* (Colborn *et al*., 1993; Hotchkiss *et al*., 2007; Creusot *et al*., 2013).

Table 5: Evaluation of a number of available bioanalytical tools for estrogenic activity

		ERα-CALUX	MELN-(MCF-7)-ERE	Yeast Estrogen Screen (YES)	E-Screen	ERE-cisFACTORIAL	hERα-HeLa-9903	ATG ERα-transFACTORIAL	Erα-GeneBLAzer	T47DKBluc	BG1luc ER TA
Assay applicability	**Applied to environmental samples**	3	2	3	3	3	3	3	3	3	3
	Validated to water samples	2	2	2	2	2	NA	1	1	2	NA
	Standardized protocol available/maturity	3	3	3	2	2	3	2	2	2	3
	Service and support available	3	NA	2	NA	1	2	1	3	NA	2
	Costs	2	NA	2	NA	1	2	1	1	2	2
	Ease of use TOTAL	4	3	4	4	2	3	2	4	3	3
	Non-GMO	0	0	0	0	0	0	0	0	0	0
	No specialised skills/equipment required	1	1	1	1	1	1	1	1	1	1
	Automation possible	1	1	1	1	NA	1	NA	1	1	1
	Non-licensed (cell) in vitro model	1	1	1	1	1	1	1	1	1	1
	Kit available	0	0	0	0	0	0	0	1	0	0
	Training availabilities	1	NA	1	1	0	NA	0	NA	NA	0
Score		17	10	16	11	11	13	10	14	12	13
Assay performance	**Selectivity**	3	NA	3	1	NA	NA	NA	NA	NA	NA
	Accuracy	3	NA	3	NA	NA	NA	NA	3	NA	NA
	Reproducibility	3	2	3	2	NA	3	NA	3	NA	3
	Robustness	3	3	3	2	NA	2	NA	2	2	3
	Sensitivity	3	2	1	3	NA	3	NA	3	2	3
	Specificity	2	NA	NA	NA	NA	3	NA	NA	NA	3
	LOD	3	NA	3	NA	NA	NA	NA	NA	NA	NA
	Cytotoxicity control	3	3	3	3	NA	3	NA	3	3	3
	Quick	2	2	2	1	2	2	2	2	2	2
	Clear/Straightforward read-out	3	3	3	2	1	3	1	3	3	3
	High-throughput capacity	3	3	3	2	2	3	2	3	3	3
Score		31	18	27	16	5	22	5	22	15	23
Total NA		0	7	1	5	9	5	9	4	7	4
Total score		48	35	44	32	25	40	24	40	34	40

4.2.2 Androgen receptor

Androgens are the male steroid sex hormones that are responsible for the development of secondary sexual characteristics and the maintenance of libido. In males, androgens are also responsible for the initiation and stimulation of spermatogenesis. Androgens are also intermediates in the production of estrogens and can be readily converted to estrogens by aromatase. Androgens play an important role in fetal sexual development and the placenta secretes significant amount of the androgen testosterone. The traditional model of the function of androgens is that they exert their action via binding to the androgen receptor, a steroid hormone receptor that regulates the transcription of specific genes. Many compounds exist in the environment that can influence the androgen receptor pathway, either agonistically or antagonistically. Known active compounds include natural (e.g. testosterone, 5α-dihydrotestosterone, androsterone) and synthetic hormones (e.g. anabolic steroids), pharmaceuticals (e.g. trenbolone, boldenone), industrial chemicals (e.g. morine dye, perfluorooctanesulfonamide) and pesticides (e.g. tetrachlorophenol, tributyltin benzoate). Recent profiling of over 3000 compounds (mainly pesticides) by

the US EPA ToxCast project showed that approximately 2% of the compounds were AR agonists, while over 10% of the compounds showed signs of antagonistic activity (Huang *et al*., 2011). With reprotoxic chemicals, mainly antagonistic effects are found in the AR-CALUX (van der Burg *et al*., 2015). Similarly to estrogenic compounds, exposure to anti-androgenic compounds may lead to reproductive and infertility problems *in vivo*.

Table 6: Evaluation of a number of available bioanalytical tools for (anti)androgenic activity

		YAS	AR CALUX	AR-EcoScreen	CHO-AR-luc	GeneBLAzer	MDA-kb2	PALM	A-SCREEN
Assay applicability	**Applied to environmental samples**	3	3	2	1	3	3	2	1
	Validated to water samples	2	2	1	1	1	2	1	1
	Standardized protocol available/maturity	3	3	2	2	2	3	3	2
	Service and support available	1	3	2	NA	3	3	2	NA
	Costs	2	2	1	1	1	2	1	NA
	Ease of use TOTAL	4	4	3	3	4	2	4	4
	Non-GMO	0	0	0	0	0	0	0	0
	No specialised skills/equipment required	1	1	1	1	1	1	1	1
	Automation possible	1	1	1	1	1	NA	1	1
	Non-licensed (cell) in vitro model	1	1	1	1	1	1	1	1
	Kit available	0	0	0	0	1	0	0	0
	Training availabilities	1	1	NA	NA	NA	NA	1	1
Score		15	17	11	8	14	15	13	8
Assay performance	**Selectivity**	3	3	2	1	NA	1	2	1
	Accuracy	3	3	NA	NA	3	2	2	NA
	Reproducibility	3	3	2	3	3	3	2	2
	Robustness	2	3	2	2	2	3	2	2
	Sensitivity	1	3	3	3	3	2	2	3
	Specificity	NA	2	NA	NA	NA	1	3	NA
	LOD	2	3	3	NA	NA	2	NA	NA
	Cytotoxicity control	3	3	3	3	3	3	3	3
	Quick	2	2	2	2	2	2	2	1
	Clear/Straightforward read-out	3	3	3	3	3	3	3	2
	High-throughput capacity	3	3	3	2	3	3	3	2
Score		25	31	23	19	22	25	24	16
Total NA		1	0	3	5	4	2	1	5
Total score		41	48	37	32	40	42	38	29

4.2.3 Glucocorticoid activity

Glucocorticoids regulate a variety physiological processes including glucose and fat metabolism and anti-inflammatory and immunosuppressive actions. They are secreted by the adrenal cortex, with cortisol being the most important and active glucocorticoid. Glucocorticoids act by binding to the intracellular glucocorticoid receptor (GR). As they play a pivotal role in energy metabolism, glucocorticoid receptors are expressed by most vertebrate cells. Because of their immunosuppressive action, glucocorticoids are among the most widely used pharmaceuticals worldwide. The glucocorticoid receptor is closely related to the

mineralocorticoid receptor, which is involved in the ionic pressure and water transport. Many ligands can act on both receptors. For environmental relevance, it is important to note that some species do not have differentiated receptors. The most prominent glucocorticoids detected in various water bodies are cortisol, cortisone, prednisolone, prednisone, dexamethasone and triamcinolone acetonide (Schriks *et al.*, 2010). Given the complex and important functions of glucocorticoids, environmental chemicals interfering with the glucocorticoid homeostasis may cause a wide spectrum of diseases, such as diabetes, obesity, cardiovascular, inflammatory and immune diseases.

Table 7: Evaluation of a number of available bioanalytical tools for glucocorticoid activity

		H295R steroid production	GR-CALUX	HG5LN GAL4-GR	HMLN-MMTV-Luc-SVNeo GR	Yeast glucocorticoid bioassay	GR-transFACTORIAL	GR-geneBLAZER	GR-MDA-kb2
Assay applicability	Applied to environmental samples	1	3	NA	3	2	3	3	3
	Validated to water samples	NA	1	NA	NA	1	1	1	1
	Standardized protocol available/maturity	3	2	2	2	2	2	2	2
	Service and support available	NA	3	NA	NA	2	1	3	3
	Costs	2	2	2	2	3	1	1	2
	Ease of use TOTAL	2	4	3	3	4	2	4	4
	Non-GMO	0	0	0	0	0	0	0	0
	No specialised skills/equipment required	1	1	1	1	1	1	1	1
	Automation possible	NA	1	1	1	1	NA	1	1
	Non-licensed (cell) in vitro model	1	1	1	1	1	1	1	1
	Kit available	0	0	0	0	0	0	1	0
	Training availabilities	0	1	NA	NA	1	0	NA	1
Score		8	15	7	10	14	10	14	15
Assay performance	Selectivity	2	3	2	2	1	NA	NA	2
	Accuracy	2	3	NA	NA	3	NA	3	2
	Reproducibility	2	3	NA	NA	3	NA	3	3
	Robustness	2	2	2	2	1	NA	2	3
	Sensitivity	3	3	3	2	2	NA	3	2
	Specificity	NA	2	NA	NA	3	NA	NA	1
	LOD	NA	3	NA	NA	1	NA	NA	NA
	Cytotoxicity control	3	3	3	3	3	NA	3	3
	Quick	2	2	2	2	2	2	2	2
	Clear/Straightforward read-out	3	3	3	3	3	1	3	3
	High-throughput capacity	3	3	3	2	3	2	3	3
Score		22	30	18	16	25	5	22	24
Total NA		5	0	8	7	0	9	4	1
Total score		35	45	33	33	39	24	40	40

4.3 Reactive modes of action

Direct reactivity of a chemical with a macromolecule may be the molecular initiating event resulting in a toxic outcome. Chemicals that act through a reactive mode of action may cause oxidative stress, protein damage, or genotoxicity. In the study of Escher *et al*. (2014), genotoxicity was demonstrated to be a health-relevant endpoint that was responsive to water contaminants. Genotoxic compounds affect the integrity of the genome by interacting with DNA and/or DNA replication processes, thereby altering structure, information content, or segregation of DNA. This DNA damage may be repaired by cellular DNA damage response mechanisms before cell division has occurred. Some of these repair systems are however error-prone, causing changes in the DNA. Mutations are permanent changes in the amount or structure of the genetic material in a cell, that persist because they are not restored adequately. Gene mutations include base-pair substitutions and frameshift mutations. Chromosomal mutations are large-scale structural and numerical changes in the DNA, e.g. deletions, insertions, breakage (clastogenicity), or chromosome loss or gain (aneuploidy). Examples of genotoxic environmental contaminants are arsenic, benzene, benzo[a]pyrene, and vinyl chloride. Exposure to genotoxic compounds may cumulatively lead to distorted cell function, which contributes to ageing and could eventually (when favourable conditions are present and cell death does not occur) result in carcinogenesis. Since DNA is the carrier of hereditary information, mutations in germ cells can lead to errors in the development of the offspring, which may cause congenital disease.

Table 8: Evaluation of a number of available bioanalytical tools for gene mutations

		Gene mutations (bacteria/yeast)					mutations (mamma		DNA replication	
		Ames test	Ames II/fluctuation test	Vibrio harveyi	MutaGen	Mitotic gene conversion assay	Mammalian gene mutation assays	Mouse Lymphoma Assay	Polymerase inhibition assay	ToxTracker
Assay applicability	Applied to environmental samples	3	3	3	1	3	2	2	1	1
	Validated to water samples	3	3	1	1	1	1	1	1	1
	Standardized protocol available/maturity	3	3	2	1	2	3	3	1	2
	Service and support available	1	3	1	1	1	1	2	1	3
	Costs	2	2	1	2	2	2	2	3	1
	Ease of use TOTAL	2	5	2	2	2	3	5	5	3
	Non-GMO	0	0	0	0	0	1	1	1	0
	No specialised skills/equipment required	1	1	1	1	1	1	1	1	0
	Automation possible	0	1	0	0	0	0	1	1	1
	Non-licensed (cell) in vitro model	1	1	1	1	1	1	1	1	0
	Kit available	0	1	0	0	0	0	1	1	1
	Training availabilities	0	1	0	0	0	0	0	0	1
Score		14	19	10	8	11	12	15	12	11
Assay performance	Selectivity	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Accuracy	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Reproducibility	1	2	NA	NA	NA	NA	NA	NA	3
	Robustness	3	3	NA	NA	NA	NA	NA	NA	3
	Sensitivity	2	2	2	2	NA	NA	3	NA	NA
	Specificity	1	1	NA	2	NA	1	1	NA	3
	LOD	1	2	1	NA	1	NA	2	NA	NA
	Cytotoxicity control	3	3	1	1	2	2	3	3	2
	Quick	2	2	2	3	2	1	1	3	2
	Clear/Straightforward read-out	2	3	2	3	2	2	1	3	3
	High-throughput capacity	1	3	1	3	1	1	2	3	3
Score		16	21	9	14	8	7	13	12	19
Total NA		2	2	5	5	6	6	4	7	4
Total score		32	42	24	27	25	25	32	31	34

Table 9: Evaluation of a number of available bioanalytical tools for chromosomal mutations

		DNA strand breaks				Clastogenicity, Aneupoidy			DNA replication	
		Comet assay	Alkaline yeast comet	Sister chromatid exchange	DEL	Micronucleus test	Chromosome abberation assay	RadarScreen	Polymerase inhibition assay	ToxTracker
Assay applicability	**Applied to environmental samples**	3	2	2	1	3	2	1	1	1
	Validated to water samples	2	1	1	1	3	1	1	1	1
	Standardized protocol available/maturity	3	3	3	1	3	3	1	1	2
	Service and support available	3	3	1	1	3	2	1	1	3
	Costs	1	1	2	2	2	1	2	3	1
	Ease of use TOTAL	4	3	3	2	3	2	3	5	3
	Non-GMO	1	1	1	0	0	1	0	1	0
	No specialised skills/equipment required	0	0	1	1	0	0	1	1	0
	Automation possible	1	1	0	0	1	0	1	1	1
	Non-licensed (cell) in vitro model	1	1	1	1	1	1	1	1	0
	Kit available	1	0	0	0	1	0	0	1	1
	Training availabilities	0	0	0	0	0	0	0	0	1
Score		16	13	12	8	**17**	11	9	12	11
Assay performance	**Selectivity**	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Accuracy	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Reproducibility	NA	NA	NA	NA	NA	NA	NA	NA	3
	Robustness	NA	NA	NA	NA	NA	NA	NA	NA	3
	Sensitivity	2	2	NA	NA	3	2	2	NA	NA
	Specificity	3	3	NA	2	2	2	3	NA	3
	LOD	2	NA	NA	NA	2	NA	NA	NA	NA
	Cytotoxicity control	NA	NA	2	2	3	3	3	3	2
	Quick	2	2	1	2	2	1	3	3	2
	Clear/Straightforward read-out	1	NA	1	2	1	1	3	3	3
	High-throughput capacity	2	2	1	2	2	1	3	3	3
Score		12	9	5	10	15	10	17	12	19
Total NA		5	7	7	6	4	5	5	7	4
Total score		33	29	24	24	36	26	31	31	34

Table 10: Evaluation of a number of available bioanalytical tools for DNA damage response

		Response to DNA damage (bacteria)					Response to DNA damage (mammalian)											
		Mutatox assay	SOS Chromo	SOS lux	SOS umu/umuC	Vitotox	HepG2 reporter gene assay	GreenScreen GC (Gentronix)	GreenScreen HC (Gentronix)	BlueScreen HC (Gentronix)	MCF-7-p53R2	p53 CALUX (BDS)	p53-cisFACTORIAL (Attagene)	CellSensor p53RE-bla (GeneBlazer, Invitrogen)	CellCiphr p53 (Cellumen)	Cignal p53 pathway (Qiagen)	NCGC p53 (GAMA-Bio)	NF-kB assay
Assay applicability	Applied to environmental samples	3	3	2	3	3	2	2	1	1	1	2	2	1	1	1	1	1
	Validated to water samples	2	2	1	3	2	1	1	1	1	1	1	1	1	1	1	1	1
	Standardized protocol available/maturity	2	3	1	3	2	1	2	2	2	1	2	2	2	2	2	1	1
	Service and support available	3	3	1	3	3	1	3	3	3	1	3	1	2	1	3	1	3
	Costs	2	2	2	3	2	2	2	2	2	2	2	1	2	1	1	2	2
	Ease of use TOTAL	4	4	4	5	4	3	4	4	5	3	4	2	1	2	3	3	4
	Non-GMO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	No specialised skills/equipment required	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1
	Automation possible	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Non-licensed (cell) in vitro model	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	1
	Kit available	1	1	1	1	1	0	1	1	1	0	1	0	0	1	0	1	1
	Training availabilities	0	0	0	1	0	0	1	1	1	0	1	0	0	0	1	0	0
Score		16	17	11	20	16	10	14	13	14	9	14	9	9	8	11	9	12
Assay performance	Selectivity	NA	NA	NA	NA	NA	NA	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Accuracy	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Reproducibility	1	3	NA	3	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Robustness	NA	NA	NA	NA	NA	NA	NA	3	3	NA	NA	NA	3	NA	NA	NA	NA
	Sensitivity	1	1	2	2	2	3	3	2	3	3	3	NA	NA	NA	NA	NA	NA
	Specificity	2	2	NA	1	3	3	3	3	3	3	3	NA	NA	NA	NA	NA	NA
	LOD	1	3	3	3	3	NA	NA	NA	NA	NA	2	1	1	NA	NA	NA	NA
	Cytotoxicity control	2	3	3	3	3	2	3	3	3	NA	3	NA	1	3	3	NA	1
	Quick	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	NA	2
	Clear/Straightforward read-out	NA	3	3	3	3	3	3	3	3	3	3	1	3	1	3	NA	3
	High-throughput capacity	NA	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Score		10	21	17	21	23	16	18	20	20	14	19	7	13	9	11	3	9
Total NA		5	3	5	3	3	5	4	3	4	6	4	7	5	7	7	10	7
Total score		31	41	33	44	42	31	36	36	38	29	37	23	27	24	29	22	28

4.4 Adaptive stress response and oxidative stress response pathway

As pointed out by Molendijk (2013), adaptive stress responses include the cellular reactions that occur after exposure to various stressors. Adaptive stress responses typically occur before more holistic endpoints, such as genotoxicity can be detected (Christmann and Kaina, 2013). Oxidative stress is often described as a disturbance in the balance between Reactive Oxygen Species (ROS) and the antioxidant reaction which can be caused by free radicals and a wide variety of chemicals (Leonard *et al*., 2004). Oxidative stress can directly or indirectly damage components of the cell, including proteins, lipids and DNA. The two most important regulators of the adaptive stress response to counteract oxidative stress are (i) nuclear factor E2 P45-related factor 2 (Nrf2) and (ii) Kelch-like ESH associated protein (keap1). When an excess of electrophilic chemicals or reactive oxygen species is present in any cell type, the oxidative stress response pathway will be activated, eventually resulting in activation of the antioxidant response element (ARE) and in the production of cytoprotective proteins with antioxidant and detoxifying capacity. The cellular response to oxidative stress is an important part of the cellular defense against different electrophilic chemicals and reactive oxygen species. The activation of this cellular signaling pathway constitutes an early stage of toxicity and is triggered at lower concentrations than apical endpoints like cytotoxicity or systems malfunctions. According to Escher and co-workers, oxidative stress response appears to be a highly sensitive and yet selective indicator of environmental pollution that responds to a wide range of chemical as well as to transformation products and disinfection by-products (Escher *et al*.,

2013). Such compounds include pharmaceuticals (e.g. fluoxetine, propranolol, atorvastin) and to a lesser extend pesticides such as dichlorvos, fipronil and propiconazole) (Escher *et al*., 2013). At present, there is no direct quantitative relationship between the induction of the oxidative stress response and adverse effects (on human health).

Table 11: Evaluation of a number of available bioanalytical tools for oxidative stress

		Assays for adaptive stress response														
		ER stress			Hypoxia		Inflammation			Metal Stress	Oxidative stress				
		CellSensor ESRE-bla	Attagene Factorial XBP1	Attagene Factorial HSE	Cellsensor HRE-bla	Attagene Factorial HIF1a	Cellsensor NFkB-bla	NF-kB CALUX	Attagene Factorial NF-kB	Attagene Factoral MRE	AREc32 assay	Attagene Factorial NRF2/ARE	CellSensor ARE-bla	nrf2-CALUX	Promega GSH-Glo Gluthatione
Assay applicability	Applied to environmental samples	1	1	1	1	1	1	1	1	1	3	1	1	1	1
	Validated to water samples	1	1	1	1	1	1	1	1	1	3	1	1	2	1
	Standardized protocol available/maturity	3	2	2	3	3	2	3	2	2	2	2	2	2	3
	Service and support available	3	2	2	3	2	3	3	2	2	1	2	3	3	3
	Costs	1	1	1	1	1	1	1	1	1	3	1	1	1	1
	Ease of use TOTAL	1	1	1	1	1	1	3	1	1	1	1	1	4	2
	Non-GMO	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	No specialised skills/equipment required	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Automation possible	1	1	1	1	1	1	1	1	1	1	1	1	1	0
	Non-licensed (cell) in vitro model	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	Kit available	0	0	0	0	0	0	1	0	0	0	0	0	1	0
	Training availabilities	0	0	0	0	0	0	1	0	0	0	0	0	1	1
Score		10	8	8	10	9	9	12	8	8	13	8	9	13	11
Assay performance	Selectivity	NA	NA	NA	NA	NA	NA	2	NA	NA	NA	NA	NA	NA	NA
	Accuracy	NA	NA	NA	NA	NA	NA	2	NA	NA	NA	NA	NA	NA	NA
	Reproducibility	2	NA	NA	2	NA	2	2	NA	NA	2	NA	2	2	NA
	Robustness	3	NA	NA	3	NA	3	3	NA	NA	3	NA	3	3	2
	Sensitivity	NA	NA	NA	NA	NA	NA	2	NA	NA	3	NA	NA	NA	NA
	Specificity	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	NA	NA	NA	NA
	LOD	2	NA	NA	1	NA	1	NA	NA	NA	2	NA	2	2	NA
	Cytotoxicity control	1	NA	NA	1	NA	1	2	NA	NA	3	NA	2	3	NA
	Quick	3	NA	NA	3	NA	3	2	NA	NA	2	NA	3	2	3
	Clear/Straightforward read-out	3	1	1	3	1	3	2	1	1	3	1	3	3	3
	High-throughput capacity	3	1	1	3	1	3	3	1	1	3	1	3	3	3
Score		17	2	2	16	2	16	20	2	2	24	2	18	18	11
Total NA		4	9	9	4	9	4	2	9	9	2	9	4	4	7
Total score		31	19	19	30	20	29	34	19	19	39	19	31	35	29

4.5 Reproductive and developmental toxicity

Developmental toxicity refers to adverse effects on the developing organism that are induced prior to conception, during pregnancy, or postnatal up to the time of sexual maturity. Since the early developmental phase is the most vulnerable phase of human life with regard to exposure to hazardous substances, developmental effects of toxicants will occur at lower doses than are required for effects in adults. Effects of toxicant exposure during critical windows of exposure may appear at any point in the human life span. Developmental toxicity may be elicited by numerous environmental factors and through many mechanisms. For some environmental contaminants, convincing epidemiological evidence of adverse pregnancy or developmental outcomes is present, whereas other agents are suspected of associations with developmental toxicity on the basis of limited human data or laboratory studies. Contaminants present in the aquatic environment and associated with developmental toxicity are found within the categories of inorganic compounds (arsenic for instance), organochlorines (such as PCB's), pesticides (e.g. atrazine), solvents (e.g. benzene), disinfection by-products (such as trihalomethanes), endocrine disrupters (bisphenol A, certain phthalates and others), disease medications (carbamazepine for instance), and lifestyle compounds (such as caffeine and ethanol). Available *in vitro* assays for developmental toxicity make use of primary or immortalized cell cultures and whole embryo cultures, and assess processes rather than pathways pre-implantation toxicity (sperm cell or oocyte function, maturation and fertilisation), developmental toxicity (embryonic development), or placental toxicity and transport. Typically, in comparison to mechanistic *in vitro* assays such assays are relatively time consuming, laborious, costly and sensitive to disturbances, making them unsuitable for routine testing of environmental samples.

Because of the complexity of the process of mammalian reproduction intact organisms are often regarded to be essential in assessing reproductive toxicity of compounds. Even then, it has been shown that large species differences exist and interspecies extrapolation of developmental toxicity typically is not higher than 60% when using apical endpoints in animals (Carney *et al*., 2011). It has been argued that the use of mechanistic information may improve the possibility to extrapolate between species since pathways of toxicity share many similarities between different species (Krewski *et al*., 2010). This mechanism-based approach of toxicity testing was one of the cornerstones of the ChemScreen program. In 2005 the consortium developed a panel of mechanism-based CALUX assays to assess hormonal activity of compounds (Sonneveld *et al*., 2005), a panel which has shown to be highly predictive for such activities in experimental animals (Sonneveld *et al*., 2006; 2011). Some of these mechanistic assays also formed a part of a battery of tests used in the Framework program (FP)6 program ReProTect that showed promising results predicting developmental toxicity of chemicals (Schenk *et al*., 2010). This study also very clearly showed that an *in vitro* test battery covering only part of the reproductive cycle processes can provide very promising result with respect to reproductive toxicity testing.

Using these assays a round robin trial among ChemScreen partners employing the battery was carried out, analysed as to predictability and kinetic extrapolation, showing promising results with a panel of 12 chemicals (Piersma *et al*., 2013). A second feasibility study focused on the possibility to distinguish between effects on sex organs (SO) and neural tube development (NTD). Analysis of the results of the CALUX high throughput panel shows that it is very suitable to identify not only estrogenic and androgenic

compounds, but also can predict chemically induced sex organ deformations with over 80% accuracy (van der Burg *et al*., 2015). This screening panel can be used in combination with more apical tests, such as the EST test or early life stage tests (zebra) to efficiently predict developmental toxicity of chemicals (Piersma *et al*., 2013; van der Burg *et al*., 2015).

Table 12: Evaluation of a number of available bioanalytical tools for reproductive and developmental toxicity

		DEVELOPMENTAL TOXICITY																						
		Preimplantation toxicity								Embryonic development						Endocrine disruption						Placenta		
		Computer-assisted Male fertility Computer-assisted sperm analysis (CASA)	Hamster Egg Penetration Test	Leydig cell systems	ReProComet assay	fertility follicle culture bioassay (FBA)	bovine oocyte maturation assay (bIVM)	bovine fertilisation test (bIVF)	mouse peri-implantation assay (MEPA)	rodent postimplantation whole embryo culture (WEC)	zebrafish embryo teratogenicity assay	frog embryo teratogenesis assay xenopus (FETAX)	chicken embryotoxicity screening test (CHEST)	rat limb bud micromass test (MM)	embryonic stem cell test (EST)	Estrogen and androgen receptor binding assays	Other receptor binding assays	MCF-7 cell proliferation assay	Ishikawa cell test	hormone receptor transcriptional activation assays	synthesis of sex steroid hormones (H295R)	Ex vivo placental perfusion assays	trophoblast cell assay	placental aromatase assay
Assay applicability	**Applied to environmental samples**	2	1	2	1	1	1	1	1	1	2	2	1	1	1	3	1	2	1	3	3	1	1	1
	Validated to water samples	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1
	Standardized protocol available/maturity	3	1	1	1	1	2	1	1	3	2	2	2	3	3	3	2	2	1	3	3	1	1	2
	Service and support available	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	1	1	1	3	1	1	1	1
	Costs	2	1	3	1	2	2	2	2	1	2	2	3	2	2	3	3	3	3	3	3	1	2	2
	Ease of use TOTAL	2	2	3	3	2	2	2	2	2	2	2	2	2	2	5	4	3	3	5	4	2	3	3
	Non-GMO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
	No specialised skills/equipment required	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1
	Automation possible	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
	Non-licensed (cell) in vitro model	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Kit available	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0
	Training availabilities	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Score		11	7	11	8	8	9	8	8	9	10	10	10	10	11	19	12	12	10	18	15	7	9	10
Assay performance	**Selectivity**	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Accuracy	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Reproducibility	NA	NA	NA	NA	3	NA	NA	3	NA	NA	3	3	NA	NA	3	NA	1	NA	NA	NA	NA	NA	1
	Robustness	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA
	Sensitivity	2	2	NA	NA	NA	3	NA	NA	2	3	2	1	2	2	3	NA	NA	NA	NA	3	NA	NA	NA
	Specificity	NA	1	NA	NA	NA	3	NA	NA	2	3	2	2	2	2	3	NA	NA	NA	NA	3	NA	NA	NA
	LOD	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Cytotoxicity control	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Quick	2	2	2	3	1	2	2	1	1	2	2	2	NA	NA	2	2	2	2	2	2	3	2	3
	Clear/Straightforward read-out	2	2	2	1	1	2	2	1	2	2	2	1	3	2	3	3	3	3	3	3	NA	2	2
	High-throughput capacity	3	1	2	2	1	1	1	1	1	1	1	1	2	3	3	3	2	2	3	3	1	2	2
Score		9	8	7	6	6	11	5	6	8	11	14	12	9	9	18	8	8	7	8	14	4	6	8
Total NA		7	6	7	8	7	6	8	7	6	6	4	3	7	7	4	8	7	8	8	6	9	8	7
Total score		27	21	25	22	21	26	21	21	23	27	28	25	26	27	41	28	27	25	34	35	20	23	25

5 Discussion, conclusions and recommendations

Bioanalytical tools hold great promise in being introduced in current monitoring strategies which mainly utilize analytical chemistry at present. Such tools could potentially be of great benefit for e.g. drinking water companies since they are confronted with an exponentially expanding list of chemicals which require expensive analytical method development and monitoring programs. However, the number of (*in vitro*) bioassays is ever expanding and there is a large variation in validation, acceptation and relevance as illustrated in the present report. Some bioassays are incidentally used for specific studies related to the mode of action of certain chemicals, whereas others are thoroughly validated using a large set of reference compounds by e.g. ECVAM, ICCVAM or ISO. The present document is aimed at selecting a panel of bioassays that is relevant for (drinking) water quality assessment. The rapport has focused on state-of–the-art (*in vitro*) bioassays for human health and is not covering bioassays intended to assess ecosystem health. For the latter bioassays the reader is referred to Kienle *et al*. (2011, 2012). In addition a comprehensive overview of commercially available (mainly *in vivo*) bioassays for assessing chemical toxicity in aqueous samples has been presented by Kokkali and van Delft (2014).

In a first tier the rapport has attempted to focus on a selection of human health related endpoints that are considered relevant for water quality assessment. The basis for the selection is a comprehensive study as carried out by Escher *et al* (2014). The latter authors quantitatively benchmarked a total of 103 bioassays for performance in water quality determination. Although, a large number of human health related endpoints was covered only a smaller selection was responsive for water relevant contaminants namely xenobiotic metabolism, hormone mediated modes of action, reactive modes of action and adaptive stress response pathway. Recognizing that the responsiveness of the latter endpoints is biased to a certain extent because local water was utilized (sampled in Australia), the present report has used these endpoints as a point of departure for further exploration of promising bioassays. It is also recognized that a number of pathways are not covered such as the interference with the thyroid hormone axis and the retinoic acid (RAR/RXR) signaling pathway. Also classification of these pathways to these modes-of-action is debatable. However, the present selection of endpoints is a basis for establishing a panel and extension can obviously take place in a later stage. The present report also attempts to utilize quantitative criteria to evaluate the respective bioassays that are relevant for the respective human health relevant endpoints. The total score of criteria should be considered as indicative and not too much weight should be given to scores of individual criteria. The major reasons are that (i) relatively unknown (but promising) bioassays inherently receive a lower score and (ii) it is sometimes difficult to quantitatively score an individual criterion since information is scattered or partly absent, and (iii) the scoring weights are relatively arbitrary. However, it is unlikely that very promising bioassays are “under-scored” thus the approach serves it purpose, in particular to design a final panel of promising assays. This panel may also be evaluated as a whole, which practical and financial aspects being considered, leading to a practical and cost-effective panel. As such, in the context of this program a comprehensive CALUX assay panel was evaluated by the project partners. Results of this evaluation will be presented elsewhere.

Table 13 presents the overview of selected *in vitro* bioassays.

Table 13: Overview of promising *in vitro* bioassays for water quality determination

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)

For xenobiotic metabolism (PXR/AhR activation) a number of assays are selected which are suitable for routine application. It should be noted that PXR pathway assays and AhR pathways are not mutually exchangeable and that for instance the AhR pathway is much more linked to a range of health effects (developmental, immune, reprotoxic, cancer) than the PXR pathway. PXR rather is a xenosensor involved in breakdown of exogenous ligands and in that context a protective pathway that can be placed in the class of the adaptive stress response pathways. The advantage of both the HG5LN PXR assay and the DR CALUX is that they have been successfully applied for water extracts and can be carried out in a high throughput mode (e.g. Creusot *et al.*, 2009) thus making them suitable candidates for routine implementation. The

advantage of the AhR geneblazer system is simultaneous cytotoxicity and target analysis which saves materials and time particularly when used in a standalone setup. In the CALUX panel, this cytotoxicity/specificity control is run in a single assay, parallel with the other CALUX assays which in a panel setting is an efficient approach as well, avoiding artefact due to expression of multiple constructs in a single cell line.

The most promising bioassays are available for the endocrine/hormone mediated mode of action. Illustrative examples (van der Linden *et al*., 2008; Mehinto *et al*., 2015; Kunz *et al*., 2015) have already demonstrated that panel application (AR, GR, ER, PR CALUX) is feasible even for complex environmental waters such as hospital effluent. It is beyond the scope of the present document to cover each EDC bioassay individually, but a discrimination can be made between reporter gene assays (CALUX, T47DKBluc), yeast based systems (YES/YAS) and proliferation assays (E-screen/A-screen). The most important advantage of reporter gene assays is that they are generally more sensitive (Leusch *et al*., 2010) as compared to other type bioassays. Advantages of other categories are that they are robust and do not require advanced equipment/skills and/or specific licenses. In the present report the AR/GR-MDA-kb2 bioassay has also been explored. A difference with the CALUX panel is that this (stable) transgenic cell line incorporates a reporter gene construct (MMTV.luc.neo) that allows interference of AR and GR pathways (Wilson *et al*., 2002). When specific information is required related to the mode of action of compounds, this bioassay may not be suitable, while interferences may also lead to unwanted interferences of pathways. The latter is important for deriving trigger values for specific modes of action.

As illustrated by Escher *et al*. (2013), oxidative stress response appears to be highly sensitive and yet a selective indicator of environmental pollution that responds to a wide range of chemicals as well as to transformation products and disinfection by-products. For oxidative stress response, there are a number of promising bioassays available namely the AREc32 bioassay and the nrf2 CALUX. The AREc32 bioassay has been thoroughly validated for the use in combination with water extracts and an indicative trigger value have been established (Escher *et al*., 2012, 2013). On the other hand the nrf2 CALUX has only been sporadically applied to environmental waters (Schriks *et al*., unpublished data) and a trigger value is absent, but the performance of both assays may be in the same range due to a comparable mechanism (namely nrf2 reporter based). The latter would require additional validation in order to assess the most promising alternative.

Reactive mode of action is a multidimensional endpoint consisting out of three major classes namely (i) gene mutations, (ii) chromosomal mutations and (iii) DNA damage response. The most well-known example is the Ames reverse gene mutation assay. Although validated in much detail and also applied to environmental waters, it remains laborious and thus inherently expensive. An interesting alternative could be the ToxTracker assay which makes use of mouse embryonic stem cells stably transfected with various reporter genes (Hendriks at al., 2012). The advantage is that this assay can be applied for both gene/chromosomal mutations. However, it has primarily been developed to facility single-compound testing (to reduce animal experiments) and has not been developed a priori for complex extracts from

surface water, drinking water and other environmental matrices so further validation would be required. To address chromosomal mutations, the micronucleus assay may be presented as a promising complementary bioassay compared to the Toxtracker. The micronucleus has been formally validated by the OECD (test no 487, 2010), a standard operating protocol is freely available and it has been applied numerously to water extracts (Sobol *et al*., 2012). The disadvantage of the latter bioassay is that it is not very suited for high-throughput application and read-out (counting of micronuclei) is laborious (or expensive to automate by means of dedicated software. Although there are many options for DNA damage response, promising models are the UMU chromotest and the Vitotox assay. Both assays make use of Salmonella bacteria (Oda *et al*. 1985; Verschaeve *et al*., 1999) but the difference is the promotor and the reporter gene (RecN/luciferase versus UmuC/b-galactosidase). A difference between the UmuC assay is its free availability whereas the Vitotox is distributed under a commercial license. In addition to the UMUc test and the Vitotox assay, there are two mammalian assays available, namely the p53 CALUX and the Bluescreen assay. The p53 CALUX has extra potential since it can be incorporated in a broad panel of assays sharing the same standard operating procedure.

As mentioned by Piersma (2004), developmental toxicity is a very complex endpoint since it incorporates many delicate physiological processes which vary in time and individual. Therefore, it is not possible to capture the complete array of processes that underlay developmental physiology in a single panel of bioassays, nor with *in vivo* animal models that have a limited level of cross-predictivity only. However, it has been clearly shown in previous validations that it is not needed to evaluate all possible processes and mechanisms to still be able to accurately predict toxicity of chemicals (Schenk *et al*., 2010; Piersma *et al*., 2013; Uibel *et al*., 2010; van der Burg *et al*, 2015). In particular, when mechanistic bioassays are used the predictions can be extrapolated to various processes since many mechanisms are reused throughout development. Taking this into account it has been shown that a panel of mechanistic bioassays for developmental toxicity focussing on a relatively narrow aspect within development such as the involvement and perturbation of endocrine signaling or wnt/TCF signaling can give much better predictions than anticipated a priori (Uibel *et al*., 2010; van der Burg *et al*., 2015). This is also illustrated in so-called read-across procedures were three classes of developmental toxicants were screened in a panel of CALUX bioassays. The results show that the CALUX panel of bioassays can successfully predict the developmental toxicity as established with the EST[2] ZET[3] and ReProGlo[4] assay (Kroese *et al*., 2015). The H295R steroidogenesis assay has potential since is well-characterized and thoroughly validated by the OECD (test no 456). In addition, it has been applied to various coastal waters and sewage effluent (Gracia *et al*., 2008). A future challenge to increase this bioassay merits would be to develop a trigger value which would facilitate interpretation of the results.

In conclusion, the best bioassay(panel) would be a simple, sensitive multiplex system suitable for high-throughput application incorporating various priority (human relevant) endpoints. However, at present such system is not available although limited multiplex systems (incorporating <3 endpoints) exist.

[2] Embryonic Stem Cells
[3] Zebrafish Emrbryotoxicity assay
[4] Stem cell based reporter gene assay

Therefore, a panel of bioassays would still be required for water quality assessment, preferably suitable for high-throughput application. This also has the advantage to be able to quantify more accurately and to establish straightforward trigger values. In practical use, a panel with similar assays generally is superior to one with more heterogeneous selected assays. Procedures with a homogeneous panel can be made efficient and automated, and quality control is facilitated in many ways. No single panel of assays currently is available that is sufficiently validated to serve as a robust routine screening tool for water quality assessment.

References

Adler, S., Basketter, D., Creton, S., *et al*. (2011). Alternative (non-animal) methods for cosmetics testing: Current Status and future prospects – 2010. Arch. Toxicol. 85, 367-485.

Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrrano, J.A., Tietge, J.E., Villeneuve, D.L., 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem. 29, 730-741.

Behnisch, P.A., Hosoe, K., Sakai, S., 2001. Bioanalytical screening methods for dioxins and dioxin-like compounds - a review of bioassay/biomarker technology. Environ. Int. 27, 413-439.

Brand, W., de Jongh, C.M., van der Linden, S.C., *et al*., 2013. Trigger values for investigation of hormonal activity in drinking water and its sources using CALUX bioassays. Environ. Int. 55, 109-118.

Carney, E.W., Ellis, A.L., Tyl, R.W., Foster, P.M., Scialli, A.R., Thompson, K., Kim J., 2011. Critical evaluation of current developmental toxicity testing strategies: a case of babies and their bathwater. Birth Defects Res. B Dev. Reprod. Toxicol. 92, 395-403.

Colborn, T., Saal, F.S.V., Soto, A.M., 1993. Developmental effects of endocrine-disrupting chemicals in wildlife and humans. Environ. Health. Perspect. 101, 378-384.

Collins, F.S., Gray, G.M., Bucher, J.R., 2008. Toxicology. Transforming environmental health protection. Science 319, 906-907.

Creusot, N, Budzinski, H., Balaguer, P., *et al*. 2013. Effect-directed analysis of endocrine-disrupting compounds in multi-contaminated sediment: identification of novel ligands of estrogen and pregnane X receptors. Anal. Bioanal. Chem. 405, 2553-2566.

Creusot, N., Kinani, S., Balaguer, P., *et al*., 2010. Evaluation of an hPXR reporter gene assay for the detection of aquatic emerging pollutants: screening of chemicals and application to water samples. Anal. Bioanal. Chem. 396, 563.

Christmann, M., Kaina, B., 2013. Transcriptional regulation of human DNA repair genes following genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation. Nucleic Acids Res. 41, 8403-8420.

Escher, B.I., Van Daele, C., Dutt, M., Tang, J.Y.M., Altenburger, R., 2013. Most oxidative stress response in water samples comes from unknown chemicals: The need for effect-based water quality trigger values. Environ. Sci Technol. 47, 7002-7011.

Escher, B.I., Hermens, J.L.M., 2002. Modes of action in ecotoxicology: Their role in body burdens, species sensitivity, QSARs, and mixture effects. Environ. Sci. Technol. 36, 4201-4217.

Escher, B.E., Leusch, F.D.L., 2012. Bioanalytical tools in water quality assessment. IWA Publishing, London, UK. 253pp.

Escher, B.I., Mayumi, A., Altenburger, R., *et al*., 2014. Benchmarking Organic Micropollutants in Wastewater, Recycled Water and Drinking Water with *In vitro* Bioassays. Environ. Sci. Technol. 48, 1940-1956.

Farre, M.J., Day, S., Neale, P.A., *et al*., 2013. Bioanalytical and chemical assessment of the disinfection by-product formation potential: Role of organic matter. Water Res. 47, 5409-5421.

Gracia, T., Jones, P.D., Higley, E.B., Hilscherova, K., Newsted, J.L., Murphy, M.B., Chan, A.K., Zhang, X., Hecker, M., Lam, P.K., Wu, R.S., Giesy, J.P., 2008. Modulation of steroidogenesis by coastal waters and sewage effluents of Hong Kong, China, using the H295R assay. Environ. Sci. Pollut. Res. Int. 15, 332-343.

Gunnarsson, L., Jauhiainen, A., Kristiansson, K., Nerman, O., Larsson, D.G.J., 2008. Evolutionary Conservation of Human Drug Targets in Organisms used for Environmental Risk Assessments. Environ. Sci. Technol. 42, 5807-5813.

Hendriks G., Atallah M., Morolli, B., Calléja F., Ras-Verloop, N., Huijskens, I., Raamsman, M., van de Water B., Vrieling H., 2012. The ToxTracker assay: novel GFP reporter systems that provide mechanistic insight into the genotoxic properties of chemicals. Toxicol Sci. 125, 285-298

Huang R., Xia, M., Cho, M.H., Sakamuru, S., Shinn, P., Houck, K.A., *et al*., 2011. Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. Environ. Health. Perspect. 119, 1142–1148.

Hotchkiss, A.K., Nelson, R.J., 2007. An environmental androgen, 17 beta-trenbolone, affects delayed-type hypersensitivity and reproductive tissues in male mice. J. Toxicol. Environ. Health Part A 70, 138-140.

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

Kienle C, Kunz P.Y, Vermeirssen E, Homazava N, Werner I, (2012). Evaluation von Methoden für den effektbasierten Nachweis von Östrogen aktiven Substanzen in Abwasserreinigungsanlagen und Fliessgewässern. Studie im Auftrag des BAFU. Schweizerisches Zentrum für angewandte Ökotoxikologie, Eawag-EPFL, Dübendorf.

Koehle, C., Bock, K. W., 2009. Coordinate regulation of human drug-metabolizing enzymes, and conjugate transporters by the Ah receptor, pregnane X receptor and constitutive androstane receptor. Biochem Pharmacol. 77, 689-699.

Kokkali, V. and van Delft, W. Overview of commercially available bioassays for assessing chemical toxicity in aqueous samples. TRAC 61, 133-155.

Krewski, D., Acosta, D. Jr, Andersen, M., Anderson, H., Bailar, J.C. 3rd, Boekelheide, K., Brent, R., Charnley, G., Cheung, V.G., Green, S. Jr, Kelsey, K.T., Kerkvliet, N.I., Li, A.A., McCray, L., Meyer, O., Patterson, R.D.,

Pennie, W., Scala, R.A., Solomon, G.M., Stephens, M., 2010. Yager J, Zeise L. Toxicity testing in the 21st century: a vision and a strategy. J. Toxicol. Environ. Health B Crit. Rev. 13, 51-138.

Kroese, E.D., Bosgra, S., Buist, H.E., *et al*., 2015. Evaluation of an alternative *in vitro* test battery for detecting reproductive toxicants in a grouping context. Reprod. Toxicol. 55, 11-19.

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

Legler, J., Jonas, A., Lahr, J. *et al*., 2002. Biological measurement of estrogenic activity in urine and bile conjugates with the *in vitro* ER-CALUX reporter gene assay. Environ. Toxicol. Chem. 21, 473-479.

Leonard, S.S., Harris, G.K., Shi, X.L., 2004. Metal-induced oxidative stress and signal transduction. Free Radic. Biol. Med. 2004, 1921-1942.

Lemaire, G., de Sousa, G., Rahmani, R., 2004. A PXR reporter gene assay in a stable cell culture system: CYP3A4 and CYP2B6 induction by pesticides. Biochem. Pharmacol. 68, 2347-2358.

Lin, Y.S., Kazuto, Y., Mahfoud., A., *et al*., 2009. The Major Human Pregnane X Receptor (PXR) Splice Variant, PXR.2, Exhibits Significantly Diminished Ligand-Activated Transcriptional Regulation. Drug. Metab. Dispos. 37, 1295-1304.

Leusch, F.D.L., De Jager, C., Levi, Y., Lim, R., Puijker, L., Sacher, F., Tremblay, L.A., Wilson, V.S., Chapman, H.F., 2010. Comparison of five *in vitro* bioassays to measure estrogenic activity in environmental waters. Environ. Sci Technol. 44, 3853-3860.

Mehinto, A.C., Jia, A., Snyder, S.A., Jayasinghe, B.S., Denslow, N.D., Crago, J., Schlenk, D., Menzie, C., Westerheide, S.D., Leusch, F.D.L., Maruya, K.A., 2015. Interlaboratory comparison of *in vitro* bioassays for screening of endocrine active chemicals in recycled water. Water Res. 83, 303-309.

Molendijk, J., 2013. Improving water and sediment quality assessment using adaptive stress response assays and implementing more efficient methodologies. Thesis Utrecht University, The Netherlands.

OECD, 2011. Detailed review paper on the state of the science on novel *in vitro* and *in vivo* screening and testing methods and endpoints for evaluating endocrine disruptors. ENV/JM/MONO(2012)23. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)23&doclanguage=en. Accessed September 2015.

Oda, Y., Nakamura, S., Oki, I., Kato, T., Shinagawa, H., 1985. Evaluation of the new system (umu-test) for the detection of environmental mutagens and carcinogens. Mut. Res. 147, 219–229.

Omiecinski, C.J., van den Heuvel, J.P., Perdew, G.H., *et al*., 2011. Xenobiotic Metabolism, Disposition, and Regulation by Receptors: From Biochemical Phenomenon to Predictors of Major Toxicities. Toxicol. Sci. 120, S49-S75.

Piersma, A.H., 2004. Validation of alternative methods for developmental toxicity testing. Toxicol. Lett. 149, 147-53.

Piersma, A.H., Bosgra, S., van Duursen, M.B., Hermsen, S.A., Jonker, L.R., Kroese, E.D., van der Linden, S.C., Man, H., Roelofs, M.J., Schulpen, S.H., Schwarz, M., Uibel, F., van Vugt-Lussenburg, B.M., Westerhout, J., Wolterbeek, A.P., van der Burg B., 2013. Evaluation of an alternative *in vitro* test battery for detecting reproductive toxicants. Reprod. Toxicol. 38, 53-64.

Punt, A., Brand, W., Murk, A.J., van Wezel, A.P., Schriks, M., Heringa, M.B., 2013. Effect of combining *in vitro* estrogenicity data with kinetic characteristics of estrogenic compounds on the *in vivo* predictive value. Toxicol. *In vitro* 27, 44-51.

Schenk, B., Weimer, M., Bremer, S., van der Burg, B., Cortvrindt, R., Freyberger, A., *et al*., 2010. The ReProTect Feasibility Study, a novel comprehensive *in vitro* approach to detect reproductive toxicants. Reprod. Toxicol. 30, 200–218.

Schriks, M., van Leerdam, J.A., van der Linden, S.C, *et al*., 2010. High-Resolution Mass Spectrometric Identification and Quantification of Glucocorticoid Compounds in Various Wastewaters in The Netherlands. Environ. Sci. Technol. 44, 4766-4774.

Schriks, M., Heringa, M.B., van der Kooi, M.M.E, de Voogt, P., van Wezel, A.P., 2010. Toxicological relevance of emerging contaminants for drinking water quality, Water Res. 44, 461-476.

Sobol, Z., Homiski, M.L., Dickinson, D.A., Spellman, R.A., Li, D., Scott, A., Cheung, J.R., Coffing, S.L., Munzner, J.B., Sanok, K.E., Gunther, W.C., Dobo, K.L., Schuler, M., 2012. Development and validation of an *in vitro* micronucleus assay platform in TK6 cells. Mut. Res. 746, 29-34

Sonneveld, E, Jansen, H.J., Riteco, J.A.C., *et al*., 2005. Development of androgen- and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid-responsive bioassays. Toxicol. Sci. 83, 136-148.

Sonneveld, E., Pieterse, B., Schoonen, W.G., *et al*., 2010. Validation of *in vitro* screening models for progestagenic activities: Inter-assay comparison and correlation with *in vivo* activity in rabbits. Toxicol. *In vitro* 25, 545-554.

Sonneveld, E., Riteco, J.A.C., Jansen, H.J., *et al*., 2006. Comparison of *in vitro* and *in vivo* screening models for androgenic and estrogenic activities. Toxicol. Sci. 89, 173-187.

Uibel, F., Mühleisen, A., Köhle, C., Weimer, M., Stummann, T.C., Bremer, S., Schwarz, M., 2010. ReProGlo: a new stem cell-based reporter assay aimed to predict embryotoxic potential of drugs and chemicals. Reprod. Toxicol. 30, 103-112.

US EPA, 2006. Drinking Water Health Advisories 2006 Edition. http://water.epa.gov/action/advisories/drinking/dwstandards.cfm. Accessed September 2015.

Van der Linden, S.C., Heringa, M.B., Hai-Yen, M., *et al*., 2008. Detection of multiple hormonal activities in wastewater effluents and surface water, using a panel of steroid receptor CALUX bioassays. Environ. Sci. Technol. 42, 5814-5820.

Van der Burg, B., Pieterse, B., Buist, H., Lewin, G., Van der Linden, S.C., Man, H.Y., Piersma, A.H., Rorije E., Mangelsdorf, I., Wolterbeek, A.P., Kroese, E.D., Van Vugt-Lussenburg, B.M., 2014. A high throughput

screening system for predicting chemically-induced endocrine disruption and sex organ deformities. Reprod. Toxicol., pii: S0890-6238.

van der Burg, B., Pieterse, B., Buist, H., Lewin, G., van der Linden, S.C., Man, H.-Y., Rorije, E., Piersma, A.H., Mangelsdorf, I., Wolterbeek, A.P.M., Kroese, E.D., van Vugt-Lussenburg, B., 2015. A high throughput screening system for predicting chemically-induced reproductive organ deformities. Reprod. Toxicol. 55, 95-103.

Van der Burg, B., Wedebye, E.B., Dietrich, D.R., Jaworska, J., Mangelsdorf, I., Paune, E., Schwarz, M., Piersma, A.H., Kroese, E.D., 2015. The ChemScreen project to design a pragmatic alternative approach to predict reproductive toxicity of chemicals. Reprod. Toxicol. 55, 114-123.

Wahl, M., Guenther, R. Yang, L., *et al*., 2010. Polybrominated diphenyl ethers and arylhydrocarbon receptor agonists: Different toxicity and target gene expression. Toxicol. Lett. 198, 119-126.

WHO, 2011. Guidelines for drinking water quality, 4th edition. http://www.who.int/water_sanitation_health/publications/dwq_guidelines/en/. Accessed September 2015.

Wilson, V.S., Bobseine, K., Lambright, C.R., Gray, L.E. Jr., 2002. A novel cell line, MDA-kb2, that stably expresses an androgen- and glucocorticoid-responsive reporter for the detection of hormone receptor agonists and antagonists. Toxicol. Sci. 66, 69-81.