Concawe

Environmental Science for European Refining

Report

Report no. 14/18

Review of options for refinery effluent assessment using effect-based tools in combination with passive samplers

ISBN 978-2-87567-095-3
9 782875 670953 >

Review of options for refinery effluent assessment using effect-based tools in combination with passive samplers

Prepared wca Consulting Ltd. for the Concawe Biological Effects Special Taskforce (WQ/STF-32):

G. Whale (Chair)
A. Basseres
L. Camenzuli
M. Comber
N. Djemel
C. Gelber
Á. González Sánchez
I. Keresztenyi
M.C. Laurent
S. Linington
T. Parkerton
K. Mézeth
A. Redman
S.A. Villalobos

M. Spence (Science Executive)
M. Hjort (Research Associate)

Prepared by wca Consulting Ltd.:

B. Brown
D. Leverett
O. Tran

Reproduction permitted with due acknowledgement

© Concawe
Brussels
September 2018

ABSTRACT

The European Commission (EC) is considering the use of "effect-based tools" (EBTs) for environmental monitoring as an alternative to, or in combination with, chemical analysis. In addition, there is interest in the use of passive sampler devices for the time- integrated monitoring of water quality.

In order to better understand the opportunity for using EBTs and passive samplers in combination, Concawe has completed a literature review of relevant EBTs that can be applied to extracts from passive samplers. The study builds on previous work by Concawe in this area, for example the 2012 review of effect-based assays relevant for use in the assessment of refinery effluents and receiving waters (Concawe 2012a). It also takes into account the findings from several literature reviews published over the last 3-4 years that address the potential application of EBTs for screening of effluents, surface waters, sediments and drinking waters (Hamers et al. 2016, CIS 2014, Brack et al. 2016, Di Paolo et al. 2016, Schriks et al. 2015).

Relevant EBTs were identified based on:

- Their commercial availability;
- Their general validation maturity;
- The extent to which they have tested on environmental samples,
- The extent to which they have been applied for screening of petroleum residues
- Their suitability for use with passive sampler extracts.

This initial screening generated a list of 22 assays. An in-depth literature review was then completed on these to obtain a more complete understanding of their performance, interpretation and application. The findings from the review are summarised in a series of fact sheets included in this report.

The short-listed assays were then compared using the information identified in the literature review to develop a suite of bioassays that can be used in combination with passive samplers for refinery effluent assessment. This suite of bioassays is summarised in this report, along with their assay-specific trigger values (identified in the literature review) that can be applied to denote an effect, according to the endpoint under investigation.

It should be noted that the bioassays assessed in the present study are by no means assessed in terms of availability with regard to being "available techniques" as defined by the Industrial Emissions Directive (IED; 2010/75/EU) article 3(10). Moreover, the suite of bioassays are restricted to those assessed in the present study, and are based on the information identified in this review, and the relative advantages and disadvantages of using each of the assays. In addition, it is recognised that new bioassays are continually being developed, and those currently at a relatively early stage of validation are in the process of being standardised and demonstrated to be reproducible.

KEYWORDS

Effect-based tools, bioassays, literature review, Water Framework Directive, passive sampling, refinery effluents, toxicity, whole organism, in vitro, endocrine disruption, genotoxicity, oxidative stress, metabolism

INTERNET

This report is available as an Adobe pdf file on the Concawe website (www.concawe.eu).

NOTE
Considerable efforts have been made to assure the accuracy and reliability of the information contained in this publication. However, neither Concawe nor any company participating in Concawe can accept liability for any loss, damage or injury whatsoever resulting from the use of this information.

This report does not necessarily represent the views of any company participating in Concawe.

CONTENTS

			Page
SUMMARY			**VI**
1.	**INTRODUCTION**		**1**
	1.1.	SCOPE AND DEFINITIONS	2
2.	**METHODOLOGY**		**4**
	2.1.	IDENTIFICATION OF RELEVANT EBTS	4
	2.1.1.	Use of EBTs for assessment of refinery effluents and receiving waters (Concawe 2012a)	4
	2.1.2.	TIPTOP survey (Hamers et al. 2016)	5
	2.1.3.	European Commission (CIS 2014)	6
	2.1.4.	Norman Network publications	7
	2.1.5.	DEMEAU Assessment (Schriks et al. 2015)	8
	2.1.6.	OSPAR	9
	2.1.7.	Industrial Emissions Directive (IED)	10
	2.1.8.	‘Long list’ of EBTs for consideration	11
	2.2.	SCREENING OF EBTS FOR DETAILED ASSESSMENT	11
	2.3.	OPEN LITERATURE SEARCHES	12
3.	**REVIEW OF EFFECT-BASED TOOLS**		**14**
	3.1.	WHOLE ORGANISM TOXICITY BIOASSAYS	14
	3.1.1.	Multi-species microbial toxicity test	16
	3.1.2.	Zebrafish QFET	17
	3.1.3.	*Daphnia Magna* metabolic activity test	18
	3.1.4.	*Caenorhabditis elegans* Growth and Reproduction Test	19
	3.1.5.	Cytotoxicity in a Rainbow Trout cell line (RT Gill-W1)	20
	3.1.6.	Zebrafish Toxarray	21
	3.2.	*IN VITRO* ASSAYS	22
	3.2.1.	Endocrine disruption	23
	3.2.1.1.	ER activation assay	23
	3.2.1.2.	Yeast Oestrogen Screen	25
	3.2.1.3.	AR activation assay using hamster cell line	27
	3.2.1.4.	AR activation assay using human bone marrow cell line	28
	3.2.1.5.	Yeast Androgen Screen	29
	3.2.1.6.	Transthyretin-Binding (TTR-binding)	31
	3.2.2.	Genotoxicity	32
	3.2.2.1.	Ames/ Ames II	33
	3.2.2.2.	umuC Assay	34
	3.2.2.3.	hGADD45 activation assay	36
	3.2.2.4.	p53-pathway activation assay	37
	3.2.2.5.	Micronucleus	38
	3.2.3.	Oxidative stress	40
	3.2.3.1.	Nrf2-pathway activation assay	40
	3.2.3.2.	ARE c32 activation assay	41
	3.2.4.	Metabolism	42
	3.2.4.1.	DR AhR activation assay	42
	3.2.4.2.	AhR activation assay (with more specificity for PAHs)	44
	3.2.4.3.	EROD	45
4.	**DISCUSSION**		**47**
	4.1.	WHOLE ORGANISM ASSAYS	47
	4.2.	ENDOCRINE DISRUPTION (ED)	53
	4.3.	GENOTOXICITY	56
	4.4.	OXIDATIVE STRESS	59
	4.5.	METABOLISM	60

5.	CONCLUSIONS	62
6.	REFERENCES	66
7.	ABBREVIATIONS	74
APPENDIX 1	PASSIVE SAMPLING	77
APPENDIX 2	FULL LIST OF EBTS	80

SUMMARY

As part of an ongoing review of the Water Framework Directive (WFD), the European Commission (EC) is considering “effect-based tools” (EBTs) for use as an alternative to, or in combination with, the monitoring of individual substance concentrations. There are a number of ways in which EBTs may be integrated into environmental monitoring, but one proposal, developed through a European Chemical Industry Council (Cefic) Long Range Initiative (LRi) project, is the use of Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP) (Hamers et al. 2016). In other words, the TIPTOP approach intend to use EBTs are used to assess the toxicity of the extract from the passive sampler.

The EC has formed a working group to examine and document the opportunities for implementation of EBTs for monitoring and assessment under the WFD and Concawe will be a member of this working group. Through its Water, Soil and Waste Management Group (WSWMG), Concawe has already conducted research on EBTs, including a review of assays relevant for use in the assessment of refinery effluents and receiving waters (Concawe 2012a), and look to further develop this research by assessing the use of EBTs alongside passive sampling techniques for the monitoring of waters receiving petroleum refinery discharges. The aim of the project was therefore to provide a critical evaluation of commercially available effect-based assays that may be deployed in conjunction with passive samplers.

A number of comprehensive reviews of EBTs, with an emphasis on their potential use in the environmental assessment of effluents, surface waters, sediments and drinking waters, have been undertaken over the last 3-4 years, (Hamers et al. 2016, CIS 2014, Brack et al. 2016, Di Paolo et al. 2016, Schriks et al. 2015). The EBTs highlighted and assessed in each of these reports/ papers were initially compiled to produce a ‘long list’ of EBTs to be considered in this project. The ‘long list’ of EBTs was subsequently screened based on the information provided in the reports/ papers outlined above, according to a small number of broad criteria:

- Commercial availability;
- General validation maturity;
- Previously application to environmental samples and more specifically, response to PAHs ; and
- Suitability for use with passive sampler extracts.

These criteria were applied relatively loosely, the aim being simply to focus the detailed assessments on those assays likely to be most appropriate and relevant for the future assessment of surface waters receiving refinery effluents, using a passive sampler approach, and to screen out those assays which are unlikely to be suitable for this purpose. These criteria also helped focus on the optimal techniques where a number covering the same type of endpoints are available. The screening resulted in a first ‘short list’ of 22 assays. The shortlisted EBTs were then subject to a detailed review, based on the published scientific literature, to identify relevant information with respect to their performance, interpretation, and application. The detailed review resulted in the production of 22 fact sheets, one for each EBT on the ‘short list’.

Finally, the short-listed assays were compared using the information identified in the literature reviews to develop a suggested suite of bioassays, to cover multiple trophic levels and endpoints, which can be used in combination with passive samplers for the assessment of waters receiving refinery effluents. Commercial availability was not included as part of the fact sheets, but was considered in the discussion in making the suggested suite of bioassays.

It should be noted that the bioassays assessed in the present study are by no means assessed in terms of availability with regard to being "available techniques" as defined by the Industrial Emissions Directive (IED; 2010/75/EU) article 3(10). Moreover, the suite of bioassays are restricted to those assessed in the present study, and are based on the information identified in this review, and the relative advantages and disadvantages of using each of the assays. In addition, it is recognised that new bioassays are continually being developed, and those currently at a relatively early stage of validation are in the process of being standardised and demonstrated to be reproducible.

The suite of bioassays is summarised in the table below (**Table 1**), along with their assay-specific trigger values (identified in the literature review) that can be applied to denote an effect, according to the endpoint under investigation. It is recognised that new bioassays are continually being developed, and those currently at a relatively early stage of validation are in the process of being standardised and demonstrated to be reproducible.

Table 1 Suite of assays that could be used for the assessment of sites receiving refinery effluents (applied to passive sampler extracts)

Assay	Assay type	Trigger value [1 2 3]	Reference
Toxicity to *Allivibrio fischeri* (ISO 11348); e.g. MicroTox	*In vivo*	No trigger value found but a similar approach can be used as for other *in vivo* assays.	-
Multi-species microbial toxicity test; e.g. MARA/ LumiMara	*In vivo*	No trigger value found but a similar approach can be used as for other *in vivo* assays.	-
Miniaturised Daphnia acute test (OECD 202)	*In vivo* (freshwater)	0.05 toxic units.	Van der Oost et al. 2017a
Microplate Algal growth tests (OECD 201 or ISO 10253)	*In vivo*	0.05 toxic units.	Van der Oost et al. 2017a
Bivalve embryo development test (ICES No.54)	*In vivo* (marine)	No trigger value found but a similar approach can be used as for other in vivo assays.	-
Cytotoxicity in a Rainbow Trout cell line (RT Gill-W1)	*In vitro* / adverse effect	0.05 toxic units.	Van der Oost et al. 2017a
Ames test	Genotoxicity	0.005 genotoxic units. Positive or negative for mutagenicity.	Van der Oost et al. 2017a
umuC	Genotoxicity	0.005 genotoxic units.	Van der Oost et al. 2017a

Assay	Assay type	Trigger value [1 2 3]	Reference
Continued Table 1			
ER activation assay; e.g. ER CALUX®	Endocrine Disruption	0.2 – 0.5 ng/L Oestrogen equivalents (EEQ).	Scott et al. 2014; Hamers et al. 2016
AR activation assay; e.g. AR CALUX®	Endocrine Disruption	25 mg Flutamide (FLU) EQ/L.	Van der Oost et al. 2017a
AhR activation assay; e.g. DR CALUX®	Metabolism	50 pg 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD)-EQ/L.	Van der Oost et al. 2017a
AhR activation assay (with more specificity for PAHs); PAH CALUX®	Metabolism	16.2 pg TCDD-EQ/L.	Hamers et al. 2016
AREc32 activation assay	Oxidative stress	EBTV-Effective concentration induction ratio (EC-IR[4]) 1.5 = 6 REF.	Escher et al. 2012

[1] Trigger values derived by Van der Oost et al. (2017a) are for use with passive samplers and conservatively assume that only 50% of the mixture is recovered as part of the passive sampler extract. The trigger values for whole organism assays are based on extrapolation of acute effects to chronic toxicity based on applying a factor of 10.
[2] Toxic unit = 1/Relative Enrichment Factor (REF) where the REF is calculated by dividing the sample concentration from solid phase extraction by the dilution in the assay (Leusch et al. 2014)
[3] Effect based trigger values (EBTV) are not intended to be used in isolation and the results should be considered as part of an overall assessment including results for other EBTs for example as proposed by the SIMONI approach (Van der Oost et al. 2017b).
[4] EC-IR is the concentration causing an induction ratio (IR) of 1.5 (ECIR 1.5)

1. INTRODUCTION

The European Commission (EC) is considering “effect-based tools” (EBTs) for use as an alternative to, or in combination with, the monitoring of individual substance concentrations (which are compared with substance-specific Environmental Quality Standard (EQS) concentrations). The use of effect-based approaches has been reviewed in a recent EC report on “Aquatic Effect-Based Monitoring Tools” (CIS 2014).

There are a number of ways in which EBTs may be integrated into environmental monitoring, but one proposal, developed through a European Chemical Industry Council (Cefic) Long Range Initiative (LRi) project, is the use of Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP) (Hamers et al. 2016). The approach has much in common with Effects Driven Analysis (EDA) approaches because it is focused on the effect assessment of concentrated mixtures of substances extracted from passive samplers which have been deployed to accumulate substances from environmental matrices.

In their review of EBTs the EC identified three types of assays that could be used for water quality monitoring;

- *In vitro* (where cell lines or cell cultures are exposed to an environmental sample providing information of effects at a lower organism level);
- *In vivo* (where whole living organisms (including bacteria) are exposed to an environmental sample providing information on effects at the organism level); and,
- Biomarkers (which are biochemical, physiological, or histological changes or aberrations measured in a whole organism that can be used to estimate either exposure to stressors or resultant effects).

In 2017, the EC formed a working group to examine and document the opportunities for implementation of EBTs for monitoring and assessment under the WFD, and possibly also the Marine Strategy Framework Directive (MSFD). Concawe is part of this working group and this report is aimed to form part of Concawe’s input to the work of this group. Through its Water, Soil and Waste Management Group (WSWMG), Concawe have already funded research on EBTs, including a review of assays relevant for use in the assessment of refinery effluents and receiving waters (Concawe 2012a).

Concawe would like to further develop this research by using EBTs alongside passive sampling techniques for the monitoring of waters receiving petroleum refinery discharges.

The aim of the project is therefore to provide a critical evaluation of commercially available effect-based assays (including biomarkers, where relevant) that may be deployed in conjunction with passive samplers, for use in assessment of any toxicity contribution to receiving waters from hydrocarbons discharged in refinery effluents. The specific aims will be to:

1. Identify a range of commercially available EBTs that can be applied for surface water quality monitoring;
2. Evaluate the assays with respect to their interpretation, validation maturity, pedigree for use in the assessment of environmental samples, applicability for use with passive sampler extracts, and response to hydrocarbons; and
3. Develop a suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents.

In the remainder of this section the scope of this study is outlined; defining some important terms applied throughout the report, and providing some background on the use of passive samplers for environmental assessment. Section 2 describes the methodology applied in identifying the EBTs for assessment, screening the EBTs for detailed evaluation, and searching the published scientific literature for relevant information on the EBTs selected. In Section 3 a detailed review of each of the selected EBTs is presented in the form of factsheets covering (where relevant information was identified); a description of each assay, validation maturity, pedigree for use in the assessment of environmental samples, response to PAHs, suitability for use with passive samplers and result interpretation. The outcomes of the EBT review are discussed in Section 4 by comparing and contrasting assays providing similar outputs, and highlighting the strengths and weaknesses of each. Finally, in Section 5, an assessment of how a suite of EBTs can be applied in combination with passive samplers for the assessment of receiving waters in the vicinity of oil refineries are provided.

1.1. SCOPE AND DEFINITIONS

The overall scope of this project builds on previous reviews of EBTs for use in the assessment of surface waters, sediments, industrial effluents or drinking waters. An exhaustive search for all potentially relevant EBTs is not provided in this report, it has instead taken as the starting point those EBTs highlighted as useful by a series of previous reviews (Section 2.1). The focus of this report is on those assays that have already shown some promise in terms of environmental assessment (albeit possibly with a different emphasis than the objectives in the present study). Nevertheless, many of the EBTs previously reviewed remain relatively novel with respect to use with environmental samples, and therefore the fact that they have been previously reviewed, highlighted or recommended by others does not mean that they are optimal for use in the assessment of surface waters receiving refinery effluents.

Even when restricting the scope of the project in this way, the number and range of potentially useful EBTs is large. While the overall objective of this work is to resolve the available techniques to a suite that is practical for use in assessing surface waters receiving refinery effluents, for the purposes of focussing the review on those assays requiring further evaluation, this report has categorised the available EBTs, as follows:

1. Whole organism assays – comprising ‘traditional’ short-term toxicity assays with algae, invertebrates, fish, bacteria, etc., the majority of which are already widely applied in chemical hazard assessments (e.g. REACH, OSPAR, OCNS, etc.), are well validated, and the results are relatively straightforward to interpret;
2. *In vitro* sub-organismal assays – comprising (usually cell-based) assays, often developed to investigate specific mechanistic endpoints such as oestrogenicity or mutagenicity. Some of these assays have previously been applied for environmental assessment, while others have been developed for other types of hazard assessment but may be adaptable for use in environmental assessment.
3. Relatively novel assays (in both of the categories above), which have not yet been fully developed or validated, but which show promise with respect to the assessment of a specific endpoint, or in being more practical or relevant than currently used assays.

This review is focussed on those assays in categories ii) and iii) above. Well validated and widely applied 'traditional' whole organism assays have only been considered as a group, and in a general sense (e.g. adaptations required for high throughput and use with passive sampler extracts), since their pedigree for use in environmental assessments is well demonstrated for example in the UK Direct Toxicity Assessment Programme (UKWIR 2000) and under OSPAR (OSPAR 2012).

In addition, biomarkers have only been considered in this review in relation to their application in *in vitro* testing systems (e.g. using cell-based assays) or, in some cases *in vivo* in laboratory exposed organisms (i.e. as part of whole organism tests). Assays involving the assessment of biomarkers in whole organisms taken from the receiving environment have not been considered.

Finally, as the overall objective of this project is to resolve a suite of assays that are likely to respond to petroleum-derived substances that may be present in refinery effluents and be recovered by passive sampler devices, detailed assessments of individual assays have been restricted to those likely to show the most promise in this regard. Passive sampling, in the context of this report, describes the use of devices maintained in situ within the water column or that, over time, passively (i.e. by diffusion) accumulate substances that are present in the water. In **Appendix 1**, the principle of passive sampling is described in more detail.

An initial screening of a 'long list' of assays (identified from previous reviews) was been carried out and agreed with Concawe (Section 2.2), resulting in a 'short list' of 22 assays for which specific searches of the scientific literature have been undertaken. For some of the shortlisted assays, little useful information was identified in the scientific literature (using the search terms applied, Section 2.3), and it was therefore difficult to fully evaluate their usefulness with respect to the aims of this work. However, this does not necessarily mean they cannot be utilised for the task in hand, just that they may not been previously applied very extensively in the assessment of environmental samples or are not yet fully mature in terms of validation. Nevertheless, such assays may address a specific endpoint not covered by better validated/ well used assays or may potentially be superior for practical reasons. Thus, while such assays may not be included in the suite of assays at this time, it has been attempted to evaluate them with a view to their potential future use and highlighted the activities required to demonstrate their usefulness (or discount them).

'Commercially available' in the context of this project refers to an assay or test that can be readily obtained and conducted by a competent laboratory. It is recognized that for most of the available tests a level of technical competence and training would be required to perform the assays. It is noted that Concawe are also interested in assays that might be commercially available in the near future (i.e. within 5 years).

'Validation maturity' refers to the level of demonstrated validation of each technique (commercial availability does not necessarily denote validation maturity), according to the validation scheme for environmental assessment techniques established by the NORMAN network (NORMAN Network 2008).

2. METHODOLOGY

2.1. IDENTIFICATION OF RELEVANT EBTS

A number of comprehensive reviews of EBTs have been undertaken over the last 3-4 years, with an emphasis on their potential use in the environmental assessment of effluents, surface waters, sediments and drinking waters, (Hamers et al. 2016, CIS 2014, Brack et al. 2016, Di Paolo et al. 2016, Schriks et al. 2015). The EBTs highlighted and assessed in each of these reports/ papers were compiled to produce an initial 'long list' of EBTs to be considered in this project. A brief summary of these reports/ papers and the EBTs covered by each is provided in the following sub-sections.

2.1.1. Use of EBTs for assessment of refinery effluents and receiving waters (Concawe 2012a)

In 2012, Concawe published a report (Report 1/12) which assessed the use of EBTs in refinery effluents and receiving waters (Concawe 2012a). The report discuss the application of biologically-based effects methods (including ecological monitoring) to refinery discharges and receiving waters, as well as assess the implications of such methods for future regulation of refinery discharges. Furthermore, it provides guidance on good practice that can be used by refineries and the downstream oil industry to carry out and interpret data obtained using biologically-based effects methods. The report includes eight case studies, provided by Concawe member companies, where EBTs were applied to refinery effluents and receiving waters. **Table 2** summarises the EBTs used in at least one of the published case studies.

Table 2 EBTs included in the Concawe Report 1/12 (Concawe 2012a)

Assay	Description	Type
Allivibrio fischeri toxicity test	Reduced bioluminescence of *Aliivibrio fischeri* test (ISO 11348)	*In vivo*
Daphnia magna acute test	Immobility (48h) test (OECD 202) Immobility (24h)	*In vivo*
Daphnia magna chronic test	Chronic toxicity (16d) test	*In vivo*
Algal growth tests	Inhibition of growth rate tests on marine algae: • *Skeletonema costatum* (ISO 10253); • *Phaeodactylum tricornutum* (ISO 10253); • *Selenastrum capricornutum* (OECD 201); • *Tetraselmis sp.*; and/or • *Pseudokirchneriella subcapitata*	*In vivo*
Acartia tonsa acute test	Test of hatching success of *Acartia tonsa*	*In vivo*
Fish acute tests	Lethality tests on fish: • *Brachydanio rerio* (OECD 203); and/or • *Cyprionodon variegatus*	*In vivo*
Mussels growth test	Scope for growth in mussels (*Mytilus edulis*) is assessed	*In vivo*

Assay	Description	Type
Continued Table 2		
Oyster larvae test	Lethality test on oyster larvae	*In vivo*
Whole sediment tests with amphipods	*Corophium volutator* used to assess toxicity of both fresh and marine water sediments.	*In vivo*
Suite of toxicity tests on marine mysid	Lethality, physiological effects (screening for stress proteins) and oxygen consumption tests on marine mysid (*Neomysis integer*)	*In vivo*
Effects of exposure to egg, embryo and larvae of cod	Heart rate, hatching, length/growth and mortality were measured in early life stages of *Gadus morhua L.* (eggs, embryo, and larvae).	*In vivo*
Ethoxy-Resorufin-O-De-ethylase (EROD) activity in flatfish	Cytochrome P450 EROD activity, which shows the (phase I) metabolism of PAH	*Metabolism*
Immunotoxicity test on mussels	Ability to phagocytose assessed, which shows the state of health of the organism.	*Genotoxicity*
Micronuclei test on mussels	Genomic instability assessed via presence of micronuclei. Test detects the frequency of micronuclei formation and nuclear abnormalities in cells.	*Genotoxicity*
Cell damage tests on mussels	Cell damage assessed via the Neutral Red assay, the altered membranes of the lysosomes exude a red colouring: the length of time colouring is retained is measured.	*Genotoxicity*
Oxidative stress test on mussels	Rate of malonedialdhyde assessed, which is a biomarker of oxidizing stress and of lipidic peroxidation (deconstruction of membranes).	*Oxidative stress*

2.1.2. TIPTOP survey (Hamers et al. 2016)

An environmental risk assessment of surface water from six Water Framework Directive (WFD) study sites and two Wastewater Treatment Plants (WWTPs) was conducted using a new approach, TIPTOP, which combined time integrated passive sampling with toxicity profiling using EBTs (Hamers et al. 2016). The study aimed to consider if the bioassay based approach would provide more ecologically relevant results compared with single compound chemical analysis, and also if the approach would be cost effective. The EBTs chosen for use in the survey included whole organism and *in vitro* assays and are summarised in **Table 3**.

The study used several risk assessment approaches to interpret the data including calculating the ‘toxic pressure’ based on both measured analytical concentrations, and whole organism responses. ‘Toxic pressure’ is the concentration in the field which exceeds the critical effect concentration of a species as determined in a single-substance laboratory study. The ‘toxic pressure’ was derived using species sensitivity distributions (SSDs) for the relevant organisms based on acute (EC50) toxicity data and then extrapolated by a factor of 10 for chronic toxicity. The results from *in vitro* assays were compared with mechanism-specific trigger values. The TIPTOP

approach was successful in demonstrating that WWTPs had higher toxicity profiles than surface waters. It was also possible to prioritise surface water sites for further investigation. In general, the environmental risks were found to be low regardless of the risk assessment strategy used. The advantages of the TIPTOP approach are that it reduces uncertainty regarding the presence of unknown substances or of a missed pollution event and is more ecologically relevant and informative regarding actual effects, compared with chemical monitoring. It was also concluded that the approach is cost effective compared with current WFD surveillance monitoring.

Table 3 EBTs included in TIPTOP (Hamers et al. 2016)

Assay	Description	Type
MicroTox™	Reduced bioluminescence of *Aliivibrio fischeri*	*In vivo*
Daphnia magna acute toxicity	Immobility (24/48h) of *Daphnia magna*	*In vivo*
Algae PAM	Photosynthetic activity in alga *Pseudokirchneriella subcapitata*	*In vivo*
Thamnotoxkit F™	Juvenile mortality (24h) of Crustacean *Thamnocephalus platyurus*	*In vivo*
Rotoxkit F™	Juvenile mortality (24h) of rotifer *Brachionus calyciflorus*	*In vivo*
Zebrafish QFET	Zebrafish (*Danio rerio*) embryotoxicity up to 120h	*In vivo*
Zebrafish Tox-array	Freshwater fish	*In vivo*
DR-LUC (H4L1.1c4)	Dioxin like activity through AhR activation	*ED*
ER-LUC (BG1Luc4E2)	Estrogenic activity through ER activation	*ED*
AR-EcoScreen™	Androgenic activity through AR activation; Anti-androgenic activity through AR inactivation in the presence of 5-alpha-Dihydrotestosterone (DHT)	*ED*
TTR-binding	Displacement of thyroid hormone precursor thyroxine(T4) from its plasma transport protein transthyretin (TTR)	*ED*
Ames II	Mutagenic activity in TA98 strain with and without metabolic activation	*Genotoxicity*

2.1.3. European Commission (CIS 2014)

In 2014 the European Commission (EC) published a review of EBTs with a focus on their suitability for use in water quality monitoring under the WFD (CIS 2014). The review lists relevant *in vitro* assays that have previously been used for monitoring purposes. The reviewed assays are summarised in **Table 4**. The Annex to the CIS (2014) report also includes some information on validation, test duration, sample volumes and sensitivity of the assays.

Table 4 *In vitro* assays highlighted for use for monitoring purposes in a European Commission (EC) assessment of EBTs (CIS 2014)

Assay	Description	Type
DR CALUX®	AhR activity through AhR receptor activation	Metabolism
PAH CALUX®	AhR activity through AhR receptor activation	Metabolism
EROD Induction	EROD Induction	Metabolism
ER CALUX® / RYA	Estrogenic / anti-estrogenic activity through ER receptor activation	ED
YES Assay	Yeast oestrogen screening assay	ED
AR CALUX®	Androgenic / anti-androgenic activity through AR receptor activation	ED
YAS Assay	Yeast androgen screening assay	ED
GR CALUX®	Glucocorticoid / anti-glucocorticoid receptor activity	ED
PR CALUX®	Progesterone receptor activity	ED
TTR Binding	Displacement of thyroid hormone precursor thyroxine(T4) from its plasma transport protein transthyretin (TTR)	ED
TRb CALUX®	Thyroid receptor activity	ED
Acetylcholinesterase inhibition assay	Inhibition of acetylcholinesterase activity	Neurotoxicity
Carboxylesterase inhibition assay	Inhibition of carboxylesterase activity	Neurotoxicity
p53 Accumulation	p53-pathway activation (DNA damage response)	Genotoxicity
GreenScreen™	Induction of DNA damage response	Genotoxicity
ABC assay	Antibiotic activity	Pharmaceutical specific

2.1.4. Norman Network publications

The NORMAN Network is a network of organisations focused on the monitoring of emerging environmental substances. Their mission is to share data, validate common methods and monitoring tools, and to conduct problem orientated research into new solutions for monitoring and identifying emerging pollutants. NORMAN collaborators are therefore actively involved in research into the use of both EBTs and passive sampler devices, with recent research highlighted on their website[1]. Two NORMAN publications were identified as relevant for further consideration in terms of identifying EBTs for potential consideration in this project (Brack et al. 2016 and Di Paolo et al. 2016).

[1] http://www.norman-network.net/?q=Publications

Brack et al. (2016) conducted a review of effect directed analysis to support environmental monitoring. As part of this review a summary was made of more than 35 *in vitro* and *in vivo* assays identified as having potential for use in such an approach. The review included some limited assessment of their use in terms of volume of substance required, sample throughput and confounding factors. All of the assays highlighted in this publication were included in the initial 'long list' of EBTs considered in the present study (**Appendix 2**).

Di Paolo et al. (2016) conducted an inter-laboratory investigation of emerging contaminants in spiked water extracts, measured using several EBTs. The EBTs applied in this study are summarised in **Table 5**. The inter-laboratory comparison was conducted at eleven different laboratories using their own internal protocols, and therefore some variability was inevitably introduced due to a lack of method standardisation. The laboratories were, however, successful in identifying samples with estrogenic and mutagenic activity. The paper recommended validation of a basic bioassay battery of tests for use in environmental monitoring.

Table 5 Assays included in inter laboratory investigation of EBTs (Di Paolo et al. 2016)

Assay	Description	Type
Daphnia magna acute toxicity	Immobility (24/48h) of *Daphnia magna*	*In vivo*
Algal growth	Inhibition of microalgae growth with *Pseudokirchneriella subcapitata* (72h)	*In vivo*
Combined algae assay	Inhibition of microalgae growth and photosynthetic activity with *Pseudokirchneriella subcapitata* (24h)	*In vivo*
Zebrafish FET Test	Zebrafish (*Danio rerio*) embryotoxicity up to 72h	*In vivo*
YES Assay	Yeast oestrogen screening assay	*ED*
ER-LUC	Estrogenic activity through ER activation	*ED*
Ames	Mutagenic activity in TA98 and T100 strain with and without metabolic activation	*Genotoxicity*

2.1.5. DEMEAU Assessment (Schriks et al. 2015)

As part of a European Union project on 'Demonstration of promising technologies to address emerging pollutants in water and waste water' (DEMEAU), a review of available EBTs was conducted by Schriks et al. (2015). This review considered a range of possible assays for measuring xenobiotic metabolism, endocrine activity, oxidative stress and genotoxicity in drinking water. Each assay was scored based on a number of criteria including accuracy, reproducibility, robustness, sensitivity, specificity, limit of detection, cytotoxicity control, speed, throughput and clarity of result. The assays recommended by the authors for each endpoint are summarised in **Table 6**.

Table 6 Assays highlighted as most promising for water quality evaluation by Schriks et al. (2015)

Assay	Description	Type
DR CALUX® / AhR GeneBlazer	AhR activity through AhR activation	*Metabolism*
HG5LN PXR / PXR Hep G2	Pregnane X Receptor (PXR) activity through PXR activation	*ED*
ER CALUX® / YES assay	Estrogenic / anti-estrogenic activity through ER activation	*ED*
AR CALUX® / AR MDA -kb2[1]	Androgenic / anti-androgenic activity through AR activation	*ED*
GR CALUX® / GR MDA -kb2	Glucocorticoid / anti-glucocorticoid receptor activity	*ED*
Ames fluctuation	Mutagenic activity in TA98 and T100 strain with and without metabolic activation	*Genotoxicity*
ToxTracker®	Reporter gene activation relating to gene and chromosomal mutations	*Genotoxicity*
Micronucleus assay	Direct measure of frequency of micronuclei formation and nuclear abnormalities in cells	*Genotoxicity*
umuC	Induction of the SOS repair response	*Genotoxicity*
P53 CALUX®	p53-pathway activation (DNA damage response)	*Genotoxicity*
Vitotox®	Induction of the SOS repair response	*Genotoxicity*
BlueScreen™ [2]	Induction of DNA damage response	*Genotoxicity*
NRf2 CALUX®	Induction of Antioxidant Response Element (ARE) pathway	*Oxidative stress*
ARE c32 Assay	Induction of Antioxidant Response Element (ARE) pathway	*Oxidative stress*

[1] AR Ecoscreen™ (41) and YAS assay (37) also scored highly in the DEMEAU assessment compared with AR CALUX® (48) and AR MDA -kb2 (43)
[2] GreenScreen™ is an equivalent yeast based rather than mammalian cell line based assay which also scored highly (36) compared with 38 for BlueScreen™.

2.1.6. OSPAR

The Convention for the Protection of the Marine Environment of the North-East Atlantic (OSPAR) is the current legislative instrument regulating international cooperation on environmental protection in the North-East Atlantic. Many of their guidelines were developed with a focus on the oil and gas industry. The Harmonised Offshore Chemical Notification Format (HOCNF) requires that whole organism bioassays are used for substance registration (OSPAR 2012/05. Update 2015) as outlined in **Table 7**. Alternative assays are available for freshwater based on the standard OECD 201, 202 and 203 guidelines.

Table 7 Whole organism bioassays assays for offshore notification under OSPAR

Assay	Species	Guideline
Algal growth	*Skeletonema costatum* or *Phaeodactylum tricornutum*	ISO 10253
Acute toxicity to crustacea	*Acartia tonsa* or *Tisbe battagliai*	ISO 14669
Acute toxicity to juvenile fish	*Scophthalmus maximus* or *Cyprinodon variegatus*	Part B of the OSPAR Protocols on Methods for the Testing of Chemicals Used in the Offshore
Acute toxicity to a sediment re-worker[1]	*Corophium volutator*	Part A of the OSPAR Protocols on Methods for the Testing of Chemicals Used in the Offshore Industry

[1] For sediment assessment only

OSPAR also recommends whole organism bioassays as part of its risk-based approach to the management of produced water discharges from offshore installations (OSPAR 2012-7). The approach suggests testing at three different trophic levels e.g. bacteria, algae and crustacean using standardised assays. The UK Government specify a preference for the following; bacteria (Microbial Assay for Risk Assessment (MARA and LumiMARA), algae (*Skeletonema costatum*) and crustacea (*Acartia tonsa*) (UK Department for Energy and Climate Change 2014).

2.1.7. Industrial Emissions Directive (IED)

Whole organism bioassays are also considered as part of waste water monitoring under the Industrial Emissions Directive (IED). The best available technique (BAT) conclusions for managing effluents in the chemical sector recommend that toxicity be evaluated using the assays detailed in **Table 8**. It should be noted that the bioassays assessed in the present study are by no means assessed in terms of availability with regard to being "available techniques" as defined by the Industrial Emissions Directive (IED; 2010/75/EU) article 3(10).

Table 8 Whole organism bioassays assays for effluent monitoring under the Industrial Emissions Directive (IED)

Assay	Species	Guideline
Luminescent bacteria	*Aliivibrio fischeri*	ISO 11348
Algal growth	Guideline specified freshwater or marine algae	ISO 8692; ISO 10253; ISO 10710
Acute toxicity to crustacea	*Daphnia magna*	ISO 6341
Fish egg toxicity	*Danio rerio*	ISO 15088
Plant growth	*Lemna minor*	ISO 20079

2.1.8. 'Long list' of EBTs for consideration

Based on the EBTs evaluated in each of the reports/ papers highlighted in sub-sections 2.1.1 to 2.1.6, an initial 'long list' of EBTs for consideration in this project was compiled, and is presented in the **Appendix 2**.

2.2. SCREENING OF EBTS FOR DETAILED ASSESSMENT

The 'long list' of EBTs was subsequently screened based on the information provided in the reports/ papers outlined above, according to a small number of broad criteria:

- Commercial availability;
- General validation maturity;
- Previously application to environmental samples and more specifically, response to PAHs; and
- Suitability for use with passive sampler extracts.

These criteria were applied relatively loosely, the aim being simply to focus the detailed assessments on those assays likely to be most appropriate and relevant for the future assessment of surface waters receiving refinery effluents, using a passive sampler approach. These criteria also enabled the screening out of those assays which are unlikely to be suitable for this purpose, and helped to focus on the optimal techniques where a number covering the same type of endpoints are available. Commonly used whole-organism assays (e.g. acute invertebrate, algae and fish tests) were not subject to this screening, since they are already well proven and no detailed evaluation of their utility for the task in hand was therefore required.

The screening resulted in a first 'short list' of 16 assays. The 'short list' was then shared with Concawe, who requested the addition of a further six EBTs, not initially selected in the screening exercise, giving a total 22 assays to be subject to detailed evaluation. The final 'short list' of EBTs for detailed assessment is shown in **Table 9**.

Table 9 Short list' of EBTs for detailed assessment

Assay	Description	Type
Multi-species microbial toxicity test	Multi-species tests using a battery based on growth of microorganisms (24h). E.g. MARA/ LumiMara.	*In vivo*
Daphnia magna metabolic activity test	Toxicity based on decreased substrate metabolism measured indirectly using a fluorescent marker. E.g. *D. magna* IQ toxicity test™.	*In vivo*
Caenorhabditis elegans assay	*Caenorhabditis elegans* development and reproduction assay 48h	*In vivo*
Zebrafish QFET	Zebrafish (*Danio rerio*) embryotoxicity up to 120h	*In vivo*
Zebrafish Toxarray	Genomics endpoints as an 'add on' to the Zebrafish FET assay	*In vivo*
Cytotoxicity in a Rainbow Trout cell line (RT Gill-W1)	Cytotoxicity assessed using three assays after 24h	*In vitro*
AhR activation assay	AhR activity through AhR receptor activation. E.g. DR LUC/ CALUX®.	Metabolism
AhR activation assay (with more specificity for PAHs)	AhR activity through AhR receptor activation (more specificity for PAHs). E.g. PAH CALUX®.	Metabolism

Assay	Description	Type
Continued Table 9		
EROD Induction	EROD Induction in fish cell lines	Metabolism
ER activation assay	Estrogenic / anti-estrogenic activity through ER activation. E.g. ER LUC/ CALUX®.	Endocrine Disruption
YES assay	Yeast oestrogen screening assay	Endocrine Disruption
AR activation assay AR Ecoscreen	Androgenic / anti-androgenic activity through AR activation. E.g. AR EcoScreen™ and AR LUC/ CALUX®.	Endocrine Disruption
YAS Assay	Yeast androgen screening assay	Endocrine Disruption
TTR Binding	Displacement of thyroid hormone precursor thyroxine (T4) from its plasma transport protein transthyretin (TTR)	Endocrine Disruption
Ames fluctuation	Mutagenic activity in TA98 and T100 strain with and without metabolic activation	Genotoxicity
Micronucleus assay	Direct measure of frequency of micronuclei formation and nuclear abnormalities in cells	Genotoxicity
UmuC	Induction of the SOS repair response	Genotoxicity
P53-pathway activation assay	P53-pathway activation (DNA damage response). E.g. P53 CALUX®.	Genotoxicity
Activation of hGADD45a1 assay	Induction of DNA damage response. E.g. GreenScreen™.	Genotoxicity
Nrf2 pathwayactivation assay	Induction of Antioxidant Response Element (ARE) pathway. E.g. NRf2 CALUX®.	Oxidative stress
ARE c32 activation assay	Induction of Antioxidant Response Element (ARE) pathway	Oxidative stress

2.3. OPEN LITERATURE SEARCHES

Each of the shortlisted EBTs was then subject to a detailed search of the published scientific literature to identify relevant information with respect to its performance, interpretation, and application.

Searches of the published or open literature were conducted using the search engines Toxline and Web of Science (through the Thomson Innovation platform). Date limits of 2010-2016 were applied to all of the searches.

A number of discrete searches for carried out, using the following groups of search terms, as follows:

[‘EBT’*] AND

1. (monitoring OR environment OR effluent OR toxicity OR regulatory OR risk assessment);
2. (oil OR gas OR PAH OR polyaromatic hydrocarbon OR polycyclic aromatic hydrocarbon); and
3. (passive sampler OR passive sampling)

* relates to the term used to describe each individual EBT.

This resulted in three sets of search results for each EBT. The results of the literature searches were, as expected, very varied. Some (widely used) EBTs gave a large number of ‘hits’ using the above three search terms (maximum total 1596 for the Ames assay), while many of the lesser used assays resulted in very small numbers of ‘hits’ (<10). In all but one case, the search results were screened to identify the most relevant studies according to the aims and objectives of the project. For the Ames test and search terms i), which resulted in 1182 ‘hits’, the results were not screened since the large number of ‘hits’ in itself, and information from other papers/ reports (identified during compilation of the ‘long list’) indicated that this assay has been widely applied for the assessment of environmental samples.

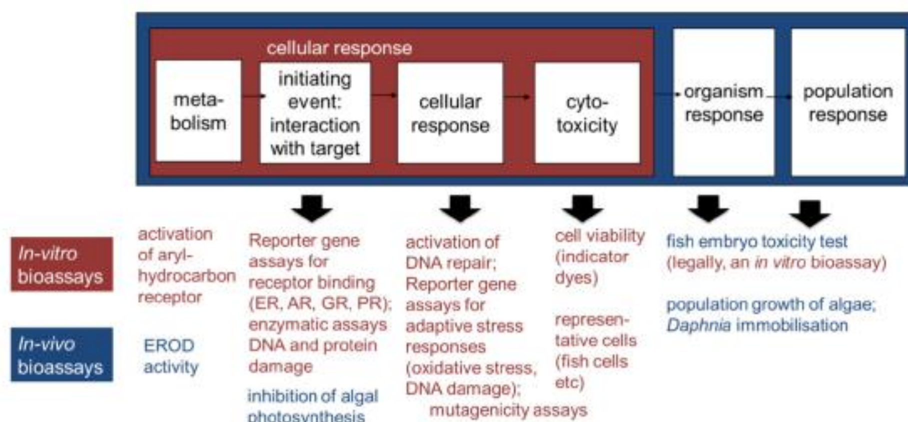
For those assays with small numbers of ‘hits’, all relevant papers were obtained for detailed assessment. For those with large numbers of relevant papers, a second screening was undertaken to reduce the number of papers to be obtained to a manageable number. In this case, the abstracts of the papers were reviewed to highlight those likely to be most relevant to the objectives of this project.

The overall outcome was approximately 80 relevant papers/ studies, which were obtained and reviewed in detail. In addition, a further series of papers and reports, identified during the compilation of the ‘long list’ were utilised in the review.

3. REVIEW OF EFFECT-BASED TOOLS

In this section, the results of the detailed review for the shortlisted EBTs, based on the scientific literature identified under Section 2.3 are presented. In general, the review is organised broadly according to the categorisation of EBTs proposed by Brack et al. (2016) (**Figure 1**); that is whole organism tests (including bacteria), cellular response assays (oxidative stress and genotoxicity), reporter assays for endocrine disruption (ED) and metabolic assays. Each sub-section includes a brief introduction describing the assays in a broad sense, before presenting a series of 'factsheets', which provide detailed information on each shortlisted assay.

Figure 1 EBTs in relation to their position in an adverse outcome pathway (AOP) (Brack et al. 2016)



3.1. WHOLE ORGANISM TOXICITY BIOASSAYS

A large number of freshwater and marine whole organism ecotoxicity tests are available, covering a wide range of species, although primarily focussing on bacteria (including photosynthetic cyanobacteria), plants and algae, invertebrates, and fish. New tests continue to be developed by researchers wishing to assess the effects of chemicals on specific taxa, although in the majority of cases these follow the same general designs, and generate similar endpoints (e.g. mortality, growth, reproduction, etc) as well-established tests, with only species-specific aspects differing. A relatively small sub-set of the available whole organism assays are widely applied in the derivation of environmental hazard data within single-substance regulatory regimes for chemicals (e.g. Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)[2], Offshore Chemical Notification Scheme (OCNS)[3], etc.), and they are generally performed according to mature and well validated internationally standardised test guidelines. A slightly wider range of whole organism assays (also possessing well validated standard guidelines) have been applied in the assessment of environmental samples (e.g. UKWIR, 2000; OSPAR 2012).

For ethical reasons, the testing of vertebrate species in whole organism tests is often (although certainly not always) avoided in environmental monitoring programmes, and in this case, is anyway unlikely to be useful since whole fish tests (even those applying larval stages of relatively small species) generally require large volumes of test solution, precluding the use of passive sample extracts. Therefore, two assays

[2] https://echa.europa.eu/regulations/reach
[3] https://www.cefas.co.uk/cefas-data-hub/offshore-chemical-notification-scheme/

have been included in this section which do not strictly measure effects at the level of the whole organism, but may be helpful nonetheless in elucidating effects on fish. The rainbow trout cytotoxicity assay utilises a fish cell line, and is an *in vitro* assay, but can be used as a surrogate for acute effects on fish. The Fish Embryo Test (FET) utilises fish embryos, rather than fully developed fish, and the Zebrafish Toxarray (which does involve the exposure of whole fish) measures genomic "biomarker" responses, and can be performed as an 'add on' to the FET.

Table 10 lists the whole organism bioassays considered in this project, and highlights which have been reviewed in detail (i.e. a 'factsheet' has been developed in the following sub-sections).

Table 10 Whole organism EBTs

Assay	Matrix and Taxa	Guideline	Factsheet Produced
Toxicity to *Allivibrio fischeri*	Marine bacteria (can be applied to freshwater samples)	ISO 11348-3	N
Multi-species microbial toxicity test	Freshwater and marine bacteria, and fungi (can be applied to freshwater samples)	None	Y
Algal growth: miniaturised, high throughput	Marine or freshwater algae	OECD 201; ISO 10253; Environment Agency/ SCA Blue Books 219 and 225.	N
Combined algae assay (photosynthesis and growth)	Freshwater algae	None	N
Algae pulse-amplitude modulation (PAM)	Freshwater algae	None	N
Daphnia acute: miniaturised, high throughput (including a toxkit)	Freshwater crustacean	OECD 202; Environment Agency/ SCA Blue Book 208	N
Daphnia Magna metabolic activity test	Freshwater crustacean	None	Y
Marine copepod acute tests	Marine crustacean	ISO 14669; ISO 16778; ISO 16778; Environment Agency/ SCA Blue Book 210	N
Freshwater crustacean toxkit (e.g. Thamnotoxkit F™)	Freshwater crustacean	ISO 14380	N
Freshwater or marine rotifer toxkit (e.g. Rotoxkit F™)	Freshwater or marine rotifer	ISO 19820	N
Echinoderm or Oyster Embryo-larval assays	Freshwater/ marine bivalve or echinoderm (marine only)	ISO 17244; ICES TIMES No.54; Environment Agency/ SCA Blue Book 209	N
Caenorhabditis elegans growth and reproduction	Soil dwelling nematode	ISO 10872	Y
Zebrafish QFET	Freshwater fish	ISO 15088	Y
Cytotoxicity in a Rainbow Trout cell line (RT Gill-W1) as a surrogate for acute effects	Freshwater fish	None	Y
Zebrafish Toxarray	Freshwater fish	None	Y

3.1.1. Multi-species microbial toxicity test

Title: Multi-species microbial toxicity test
Description: Multi-species test using a battery of ten bacteria and a fungus with a 24 hour test duration. Growth of the microorganisms is assessed by measuring the reduction of the redox dye tetrazolium red (TTC) spectrophotometrically. The strains show different sensitivities to different chemicals and the resulting array of 11 growth inhibition values gives a 'toxic fingerprint' of the chemicals tested. An example of a multi-species microbial toxicity test is the Microbial Assay for Risk Assessment (MARA) / LumiMARA. LumiMara has nine marine species of bacteria with bioluminescence as an endpoint.
Validation Maturity: Level 3: Routine laboratory. An inter-laboratory ring-test has been performed with acceptable results (Wadhia and Dando 2009).
Application with Environmental Samples: This assay has been applied to raw waters, industrial effluents, sewage sludge and soil leachates (Fai and Grant 2010; Wadhia et al. 2007; Wadhia and Dando 2009).
Response to PAHs: The assay shown to have a high sensitivity to crude oil in wastewater treatment plant effluents (Steliga et al. 2015).
Suitability for use with Passive Sampler Extracts: No specific information identified regarding use with passive sampler extracts, but assay design utilises small sample volumes and provides results relatively rapidly (96 well plate design with test and duration of 24 hours).
Interpretation of Results: Results are reported as a microbial toxic concentration (MTC) whereby the MTC for individual species can be compared to the average MTC for all the species (Steliga et al. 2015). The MTC are more comparable to IC20 values than IC50 values for the individual MARA micro-organisms (Fai and Grant 2010). The assay allows a fingerprint of toxicity based on the response of the eleven different microbes.
Key Publications: Fai and Grant. 2010. An assessment of the potential of the microbial assay for risk assessment (MARA) for ecotoxicological testing. Wadhia et al. 2007. Intra-laboratory evaluation of Microbial Assay for Risk Assessment (MARA) for potential application in the implementation of the Water Framework Directive (WFD). Wadhia and Dando. 2009. Environmental toxicity testing using the Microbial Assay for Risk Assessment (MARA). Steliga et al. 2015. Changes in toxicity during treatment of wastewater from oil plant contaminated with petroleum hydrocarbons.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): Yes, see Discussion-section (section 4) for motivation.

3.1.2. Zebrafish QFET

Title: Zebrafish (*Danio rerio*) embryotoxicity
Description: The QFET is based on a Fish Embryo Acute Toxicity (FET) test adapted for use with low volume samples. The test is conducted in a 24 well microplate with 12-13 embryos (<4 hours post fertilisation) per well in 2ml of test solution and four replicates per treatment. Every 24 hours the Zebrafish embryos are scored for visible malformations up to day 5 of development. LOEC, NOEC and ECx values are reported.
Validation Maturity: Level 3: Routine laboratory. The FET test is based on a standard OECD method but there is limited information on how the reduced test volumes affect the test responses. Test guideline: OECD 236, Fish Embryo Acute Toxicity (FET) Test.
Application with Environmental Samples: No relevant information on the application of this test to environmental samples was identified based on literature searches undertaken for this project.
Response to PAHs: The assay is non-specific but fish tend to be less sensitive to hydrocarbons than *Daphnia* or algae (Concawe 2012a). Not active for B(a)P or benzo(b)fluoranthene in an assessment of relevant river pollutants (Neale et al. 2017).
Suitability for use with Passive Sampler Extracts: The assay has previously been applied in the assessment of passive sampler extracts (Hamers et al. 2016).
Interpretation of Results: LC50 based on indicators of lethality: coagulation of fertilised eggs, lack of somite formation, lack of detachment of the tail-bud from the yolk sac, and lack of heartbeat. The QFET is the term used for the miniaturised assay as used in the TIPTOP study (Hamers et al. 2016). Detectable toxicity in the QFET was observed for approximately 50% of the passive sampler extracts. A Zebrafish Toxarray was used to provide supporting interpretive information on the results observed in the FET test. QFET is not a widely-used term in the scientific literature and therefore limited ‘hits’ were obtained in the literature search. As the FET assay is essentially a well-developed and validated test similar to other whole organism bioassays discussed in this report no further literature search was conducted.
Key Publications: Hamers et al. 2016. Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP): an effect-based strategy for cost-effective chemical water quality assessment.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): Yes, see Discussion-section (section 4) for motivation.

3.1.3. *Daphnia Magna* metabolic activity test

Title: *Daphnia Magna* metabolic activity test

Description:

Toxicity is assessed by observing the cleavage of the fluorometric biomarker methylumbelliferyl galactoside (MUF) *in vivo*. Animals are exposed to the sample for one hour and then a solution of biomarker substrate is added directly to the exposure chamber. Organisms that feed normally with functional galactosidase enzyme systems are able to cleave the marker from the substrate. The fluorescent marker is freed to the haemolymph of the organism and is observed visually using long wave UV light.

An example of a *Daphnia magna* metabolic activity test is the *Daphnia magna* IQ Toxicity Test™

Validation Maturity:

Level 2: Expert laboratory.

An inter-laboratory ring test has been performed (Hayes et al. 1996).

Application with Environmental Samples:

No relevant information on the application of this test to environmental samples was identified based on literature searches undertaken for this project.

Response to PAHs:

No relevant information on the specific responses of this test to PAHs was identified based on literature searches undertaken for this project.

Suitability for use with Passive Sampler Extracts:

The assay was trialled for the TIPTOP survey, but was not used due to technical difficulties in interpreting the assay response (Hamers et al. 2016). However, in general the test design is suitable for use with passive sampler extracts.

Interpretation of Results:

The assay records decreased fluorescence due to decreased substrate metabolism, which indicates toxicity. The results have been correlated (>95%) with standard 48-hour toxicity tests (Hayes et al. 1996).

Key Publications:

Hamers et al. 2016. Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP): an effect-based strategy for cost-effective chemical water quality assessment.

Hayes et al. 1996. Inter- and intra-laboratory testing of the *Daphnia magna* IQ toxicity test™.

Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No):

No, see Discussion-section (section 4) for motivation.

3.1.4. *Caenorhabditis elegans* Growth and Reproduction Test

Title: *Caenorhabditis elegans* Growth and Reproduction Test
Description: Growth (body length) and reproduction (number of juveniles generated from two adult hermaphrodite worms) of *C. elegans* determined after 72 or 96 hour exposure.
Validation Maturity: Level 3: Routine laboratory. A standard method is available and has been ring-tested (ISO 10872) (Clavijo et al. 2016; Höss et al. 2012).
Application with Environmental Samples: Method is reliable for aqueous media, freshwater sediments and soils (Clavijo et al. 2016).
Response to PAHs: PAHs generally show low acute toxicity to *C. elegans* (Sese et al. 2009)**.** The chronic toxicity of PAHs to *C. elegans* (72h EC50 based on reproduction) was less sensitive than acute toxicity to *D. magna* (48h EC50) for four of six PAHs tested (Sese et al. 2009).
Suitability for use with Passive Sampler Extracts: Relatively small volumes of sample are required (0.5 mL water sample in each well of a tissue culture plate), therefore assay likely to be suitable for use with passive sampler extracts.
Interpretation of Results: EC50, NOEC and LOEC based on effects on growth and reproduction. *C. elegans* are soil invertebrates and therefore their environmental relevance for surface water assessments is debatable.
Key Publications: Clavijo et al. 2016. The nematode *Caenorhabditis elegans* as an integrated toxicological tool to assess water quality and pollution. Höss et al. 2012. Interlaboratory comparison of a standardized toxicity test using the nematode *Caenorhabditis elegans* (ISO 10872). Sese et al. 2009. Toxicity of Polycyclic Aromatic Hydrocarbons to the Nematode *Caenorhabditis elegans.*
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): No, see Discussion-section (section 4) for motivation.

3.1.5. Cytotoxicity in a Rainbow Trout cell line (RT Gill-W1)

Title: Cytotoxicity in a Rainbow Trout cell line (RT Gill-W1)
Description: This fish cell line assay is based on rainbow trout gill cells (RT Gill-W1) with cytotoxicity assessed, usually after 24 hours exposure, based on three fluorescent indicator dyes, Alamar Blue™ to measure cell metabolism, 5-carboxyfluorescein diacetate acetoxymethyl ester (CFDA-AM) to measure cell membrane integrity, and neutral red to measure lysosomal activity.
Validation Maturity: Level 2: Expert laboratory. An ISO guideline "Water quality - Determination of acute toxicity of chemicals and water samples to a fish gill cell-line (RT gill-W1)" is in preparation.
Application with Environmental Samples: The assay has been applied to environmental samples including paper mill effluents (Dayeh et al. 2002), oil sands (Samson et al. 2013) and sediments (Amaeze et al. 2014). It has been used to assess oil sands process-affected waters where it was found to be useful at detecting high concentrations of naphthenic acids (Samson et al. 2013). It has also been applied to the assessment of sediment samples, but no consistent relationship between PAH concentration and cytotoxicity in sediment samples from Lagos lagoon, Nigeria was identified (Amaeze et al. 2014).
Response to PAHs: Only two- and three-ring PAHs have been found to be directly cytotoxic (Schirmer et al. 1998) using this assay. Other PAHs may be indirectly cytotoxic following metabolic activation. Water solubility and lipophilicity are the critical properties determining the direct cytotoxicity of PAHs by influencing PAH accumulation in membranes. Only naphthalene was effective at concentrations well below its water solubility limit. Therefore, direct cytotoxicity is likely to be most environmentally relevant with naphthalene. Naphthalene affects the mitochondrial electron transport chain which is measured in the Alamar Blue™ assay. Cytotoxicity of B(a)P is dependent on the xenobiotic metabolism of the cytochrome P450 system but 7-ethoxyresorufin-O-deethylase (EROD) activity is not detectable in the RT Gill-W1 cell line (Schirmer et al. 1998).
Suitability for use with Passive Sampler Extracts: No specific information identified regarding use with passive sampler extracts, but assay design utilises small sample volumes and provides results relatively rapidly (well plate design with test and duration of 24 hours).
Interpretation of Results: Derives an EC50 based on cell viability. The method has been improved to reduce interferences from media and is now predictive of fish acute toxicity for many substances (Tanneberger et al. 2013). Not sensitive to all substances because the cells can lack some receptors. May be insensitive to substances that require significant metabolic activation (including some PAHs) as the relevant transformation enzymes may not be present (Tanneberger et al. 2013).

Key Publications: Amaeze et al. 2014. Cytotoxic and genotoxic responses of the RT Gill-W1 fish cells in combination with the yeast oestrogen screen to determine the sediment quality of Lagos lagoon, Nigeria. Dayeh et al. 2002. Applying whole-water samples directly to fish cell cultures in order to evaluate the toxicity of industrial effluent. Samson et al. 2013. Rapid assessment of the toxicity of oil sands process-affected waters using fish cell lines. Schirmer et al. 1998. Ability of 16 priority PAHs to be directly cytotoxic to a cell line from the rainbow trout gill. Tanneberger et al. 2013. Predicting fish acute toxicity using a fish gill cell line-based toxicity assay.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): Yes, see Discussion-section (section 4) for motivation.

3.1.6. Zebrafish Toxarray

Title: Zebrafish Toxarray
Description: A combination of quantitative Polymerase Chain Reaction (qPCR) and a Zebrafish (*Danio rerio*) toxicity test (Hamers et al. 2016). This system allows the screening of up to 42 target genes, covering several toxicity pathways. The choice of target genes can be adapted to a mode of action of interest. The exposure duration was 5 days, but the test media is not renewed during this period.
Validation Maturity: Level 1: Research laboratory. This is a relatively novel assay, and validation activity is currently restricted to development within research facilities. Interpretation and standardisation remains a barrier for use for regulatory purposes.
Application with Environmental Samples: No relevant information on the application of this test to environmental samples was identified based on literature searches undertaken for this project.
Response to PAHs: No relevant information on the specific responses of this test to PAHs was identified based on literature searches undertaken for this project.
Suitability for use with Passive Sampler Extracts: The assay has previously been applied in the assessment of passive sampler extracts (Hamers et al. 2016).
Interpretation of Results: Provides supporting information on changes in gene expression. As might be expected samples that demonstrated toxicity in the QFET assay, also induced strong responses in this assay. There were also some correlations with *in vitro* assays especially for metabolism endpoints (DR LUC). Non-monotonic dose responses can complicate interpretation of the data. The method of

<table>
<tr><td>exposure and life stage of the organisms were both considered to have affected the genomic response (Hamers et al. 2016).

While a literature search for a Zebrafish Toxarray did not generate additional hits as part of a specific literature search, “omics” endpoints are increasingly being used in environmental monitoring studies.

The Zebrafish model is being increasingly used to better understand the mechanistic basis for Zebrafish toxicity (Goodale et al. 2012; Timme-Laragy et al. 2009).</td></tr>
<tr><td>Key Publications:

Goodale et al. 2012. AHR2 Mutant Reveals Functional Diversity of Aryl Hydrocarbon Receptors in Zebrafish.

Hamers et al. 2016. Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP): an effect-based strategy for cost-effective chemical water quality assessment.

Timme-Laragy et al. 2009. Antioxidant Responses and NRF2 in Synergistic Developmental Toxicity of PAHs in Zebrafish</td></tr>
<tr><td>Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No):

No, see Discussion-section (section 4) for motivation.</td></tr>
</table>

3.2. *IN VITRO* ASSAYS

In vitro assays measure or detect a response in living cells which can be used to indicate exposure to substances inducing a specific type of response or activity. The most common *in vitro* assays are based on genetically modified cell lines (usually human) in which fluorescence or luciferase activity is measured as an indication of receptor mediated activity. The results from the assays are usually reported as Toxic Equivalent Concentrations (TEQs) or Bioanalytical Equivalent Concentrations (BEQ) based on potency (REP) relative to a potent receptor agonist (e.g. 17β-estradiol for estrogenic activity).

Most of the assays are available commercially but can only be performed in accredited laboratories holding a licence for the use of genetically modified organisms, but other *in vitro* assays exist which utilise a more generic cell line (not genetically modified) in which a “biomarker response” is measured (e.g. EROD induction or micronucleus formation).

Most of these tests require a complimentary cytotoxicity assessment to check that the exposed cells are viable and that the exposure is not causing direct toxicity (i.e. depressing cell metabolic activity). The most commonly used assays for such cytotoxicity assessments include the 3-(4,5-dimethyl-2-thiazolyl)-2, 5-diphenyl-2H-tetrazolium bromide (MTT) formazan assay or Alamer Blue™ assay, but it may be sufficient to use the results from bacterial cell based assays (e.g. *Allivibrio fischeri* toxicity test or Ames test) to infer a lack of toxicity (e.g. as undertaken by Hamers et al. 2016 in the TIPTOP study).

Assays have been developed for a range of receptor mediated endpoints, usually associated with hormone pathways. The selected assays was grouped based on their mechanism of action i.e. endocrine disruption, genotoxicity, oxidative stress and metabolism.

3.2.1. Endocrine disruption

Six assays (or assay types) which are indicators of endocrine disruption (ED) were selected for detailed review. Two focus on effects on the oestrogen receptor (assays based on reporter genes within mammalian cell lines and a yeast oestrogen screen (YES)); three for androgenic activity (two based on reporter genes within mammalian cell lines and a yeast androgen screen (YAS)); and finally, a relatively new assay based on transthyretin-binding (TTR-binding) and indicative of effects via thyroid pathways.

3.2.1.1. ER activation assay

Title: Oestrogen Receptor (ER) activation
Description: Oestrogen receptor binding activity is measured in a human bone marrow cell line (U2OS), incorporating the firefly luciferase gene coupled to Oestrogen Responsive Elements (EREs) as a reporter gene for the presence of oestrogens and/or oestrogen-like compounds. The bioassays report binding based on total 17β-estradiol (E2) equivalents or E2 equivalence factors (EEFs). Other similar assays exist but are based on the ERβ receptor. The ERα and ERβ receptors differ in their ligand binding domains and tissue distribution. The ERα is the most commonly recommended (CIS 2014; Kunz et al. 2017). Different ER LUC cell lines also exist, including those based on BG-1 ovarian cancer cells (BG1Luc4E2) (Hamers et al. 2016; Vethaak et al. 2017). The assays can also be used to detect anti-estrogenic activity by co-exposing a substance/sample with E2, and using Tamoxifen as the standard. Example of ER activation assays are ER-CALUX® and other ER LUC assays, with the assays using either ERα or ERβ receptors.
Validation Maturity: Level 3: Routine laboratory. OECD 455: Draft updated TG 455: performance-based test guideline for stably transfected transactivation in vitro assays to detect oestrogen receptor agonists and antagonists
Application with Environmental Samples: These assays have been widely applied to environmental samples in the assessment of both estrogenic and anti-estrogenic activity (Scott et al. 2014).
Response to PAHs: Vrabie et al. (2011) tested four crude oils and seven refined oils and found that there was activity of the ERα receptor and ERβ receptor for most oils with only minor differences in efficacy between the two receptors (Vrabie et al. 2011). The results were highlighted as very different from responses with assays based on yeast cells where estrogenic activity was observed via the ERβ only (Vrabie et al. 2010). The differences may be due to absorption, transport, and metabolism as well as the presence of coactivator proteins. Similar discrepancies are observed regarding the estrogenic activity of B(a)P and benzo(b)fluoranthene which were active in a human cell line but not Zebrafish cell lines (Neale et al. 2017).
Suitability for use with Passive Sampler Extracts: Assay has previously been applied in the assessment of passive sampler extracts (Hamers et al. 2016, Vethaak et al. 2017).

Interpretation of Results: Sensitivity >0.1 ng/l EEQ for oestrogenicity and 5 µg/l tamoxifen equivalent (TMXEQ) for anti-estrogenic effects (Scott et al. 2014). The assay may be affected by anti-estrogenic chemicals and it can be difficult to measure all possible estrogenic substances, and therefore to account for all the estrogenic activity in an environmental sample (Scott et al. 2014). There is no scientific consensus on a safe degree of 'oestrogenicity'. A PNEC for 17β-estradiol of 1 ng/l would be equivalent to 1 ng/l oestrogen equivalents (EEQ) in the ERα-assay (Scott et al. 2014). Alternatively, bioassay-based Environmental Quality Standards (EQS) have been proposed for an ER activation assay; 0.6–2 ng/L EEQ (short-term exposure) or 0.2–0.4 ng/L EEQ (long-term exposure). Low risk trigger values (LR-TV) have also been proposed for ER LUC assays based on a comparison between the 5th percentile of the species sensitivity distribution (SSD) determined for all compounds yielding a specific response in the relevant bioassay (Hamers et al. 2016). A value below this LR-TV for ER agonism of 0.5 ng/L EEQ indicates no risk. More conservatively, an effect based trigger values (EB-TV) of 0.2 ng/L EEQ has been proposed which is designed to be protective for human health (Hamers et al. 2016). Different results can be observed with different types of assays (e.g. yeast based verses mammalian receptor based assays). There are suggestions that the different types of assay should be applied to complement each other and also that any effects should be confirmed using *in vivo* tests (Vrabie et al. 2011). A recent assessment of five *in vitro* bioassays for detecting estrogenic activity considered that an ER activation assay showed the best performance based on precision and repeatability (Kunz et al. 2017). Higher EEQ values for mixtures of substances could be due to effects on cell membranes or metabolism as well as interaction at the receptor (Di Paolo et al. 2016).
Key Publications: Di Paolo et al. 2016. Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts – Towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring. Hamers et al. 2016. Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP): an effect-based strategy for cost-effective chemical water quality assessment. Kunz et al. 2017. Effect-based tools for monitoring estrogenic mixtures: Evaluation of five *in vitro* bioassays. Scott et al. 2014. An assessment of endocrine activity in Australian rivers using chemical and *in vitro* analyses. Vethaak et al. 2017. Toxicity profiling of marine surface sediments: A case study using rapid screening bioassays of exhaustive total extracts, elutriates and passive sampler extracts. Vrabie et al. 2011. Specific *in vitro* toxicity of crude and refined petroleum products: 3. estrogenic responses in mammalian assays.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): Yes, see Discussion-section (section 4) for motivation.

3.2.1.2. Yeast Oestrogen Screen

Title: Yeast Oestrogen Screen (YES)
Description: The YES utilises a recombinant yeast strain used for the identification of substances that can interact with the human oestrogen receptor alpha (hERα). Oestradiol equivalent factors (EEQ) are calculated by measuring the effective concentration required to elicit a 50% response in the exposed organisms (EC50) for the 17β-oestradiol (E2)–positive control and determining the percent of sample required to give an equivalent response (Alvarez et al. 2008; Metcalf et al. 2013). By co-exposing a substance/sample with E2 and using tamoxifen as the standard, the YES also allows testing for hERα-antagonistic activities (anti-oestrogenicity), inhibition of E2 binding to the hERα. Cytotoxicity can be distinguished from effects by measuring growth of the yeast cells (620 nm) in parallel to the enzyme activity (540 nm) (CIS 2014).
Validation Maturity: Level 3: Routine laboratory. The assay is well validated in general terms using a method based on Routledge & Sumpter (1996), and is considered mature. Highly reproducible between laboratories. In the process of being validated for water samples (Schriks et al. 2015). ISO guideline in preparation.
Application with Environmental Samples: The YES assay has been very widely applied to the assessment of environmental samples (Schriks et al. 2015; Metcalf et al. 2013; Nelson et al. 2007; Rastall et al. 2004; 2006; Smith et al. 2015).
Response to PAHs: Several individual PAHs or PAH derivatives have been shown to be either oestrogen agonists or antagonists in this assay (Rastall et al. 2004). Vrabie et al. (2011) tested four crude oils and seven refined oils in a yeast based screen (not YES) and found that there was no activity of the ERα receptor, but that there was activity of all oils to the ERβ receptor (Vrabie et al. 2010). There was evidence for additive effects, antagonistic and synergistic effects. Investigation of the response suggests that the estrogenic effects may be only partially mediated via the receptor. The potency of the oils was much lower than for the reference substance but supermaximal responses (fluorescence formation higher than observed for E2) were observed for some oils, the reasons for this are unclear. Anti-estrogenic responses were also observed for some of the oils. Rastall et al. (2006) found that passive sampler fractions that had estrogenic activity also contained PAHs and their alkylated derivatives, which the authors speculate may have contributed to the activity. ER agonists were detected in oil and gas production water effluents, however 95% of the ER agonists could not be identified (Balaam et al. 2009). Tollefsen et al. (2007) observed ER activity in the dissolved phase of oil and gas production water effluents, whereas no response was detected in filtered oil droplets.
Suitability for use with Passive Sampler Extracts: This assay has routinely been applied to the assessment of passive sampler extracts (Alvarez et al. 2008; Grover et al. 2011; Liscio et al. 2009; Rastall et al. 2004; 2006).
Interpretation of Results: Highly selective for, and sensitive to, estrogenic compounds (Schriks et al. 2015). May be less sensitive than mammalian cell based assays (CIS 2014; Beresford et al. 2016) but it can be used for ER antagonists (CIS 2014).

Anti-oestrogens may interfere with the estrogenic response which may affect the correlation with measured concentrations of oestrogens (Rastall et al. 2006; Alvarez et al. 2008; Grover et al. 2011; Wang et al. 2011; Zhao et al. 2011). Chlorination alters the estrogenic potency of estrogenic compounds (Balaam et al. 2009). Appropriate sample preparation procedures (e.g. collection, extraction) are important in obtaining accurate results (Beresford et al. 2016).

Huggett et al. (2002) compared the estrogenic activity of fractionated effluents from sewage treatment works (STW) in North America and found that the YES assay and *in vivo* medaka vitellogenin induction assay did not always identify oestrogenicity in the same fractions.

The expected estrogenic activity in a sample can be calculated from measured chemical analysis data, based on the concept of concentration addition (calculated EEQs) (Liscio et al. 2009). The measured chemical data cannot always explain the estrogenic activity observed in environmental samples (Metcalf et al. 2013).

LODs as low as 0.02 ng E2/L and 0.03 ng E2/L have been reported in surface water (freshwater) and produced water (marine water), respectively (Grover et al. 2011; Tollefsen et al. 2007).

Fractionation can help to identify estrogenic compounds (Thomas et al. 2004a, 2004b). The results from individual fractions can visualised as an "estrogram" (Rastall et al. 2006). Fractionation can reduce interference from cytotoxic compounds or ER antagonists but has the disadvantage of losing the dose response effect.

A method for classifying oestrogenicity is sometimes used to aid interpretation based on fully, partially, weakly, or not estrogenic if their responses were >75%, 25 to 75%, 10 to 25%, or <10% of the fluorescence formation by E2, respectively (Vrabie et al. 2010).

The estrogenic activity of oils as environmental mixtures in themselves is complex to interpret, even without considering other substances that are also likely to be present in the environment (Vrabie et al. 2010).

Key Publications:

Alvarez et al. 2008. Chemical and toxicologic assessment of organic contaminants in surface water using passive samplers.

Metcalfe et al. 2013. A multi-assay screening approach for assessment of endocrine-active contaminants in wastewater effluent samples.

Balaam et al. 2009. Identification of nonregulated pollutants in North Sea-produced water discharges.

Beresford et al. 2016. Use of a battery of chemical and ecotoxicological methods for the assessment of the efficacy of wastewater treatment processes to remove estrogenic potency.

Di Paolo et al. 2016. Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts – Towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring.

Escher et al. 2008. Monitoring of the ecotoxicological hazard potential by polar organic micropollutants in sewage treatment plants and surface waters using a mode-of-action based test battery.

Fang et al. 2012. Assessment of hormonal activities and genotoxicity of industrial effluents using in vitro bioassays combined with chemical analysis.

Grover et al. 2011. Endocrine disrupting activities in sewage effluent and river water determined by chemical analysis and *in vitro* assay in the context of granular activated carbon upgrade.

Huggett et al. 2002. Comparison of in vitro and in vivo bioassays for oestrogenicity in effluent from North American municipal wastewater facilities.

Liscio et al. 2009. Combining passive samplers and biomonitors to evaluate endocrine disrupting compounds in a wastewater treatment plant by LC/MS/MS and bioassay analyses.

Nelson et al. 2007. The use of *in vitro* bioassays to quantify endocrine disrupting chemicals in municipal wastewater treatment plant effluents.

Osman et al. 2015. Screening of multiple hormonal activities in water and sediment from the river Nile, Egypt, using *in vitro* bioassay and gonadal histology.

Rastall et al. 2004. The identification of readily bioavailable pollutants in Lake Shkodra/Skadar using Semipermeable Membrane Devices (SPMDs), bioassays and chemical analysis.

Rastall et al. 2006. A biomimetic approach to the detection and identification of oestrogen receptor agonists in surface waters using Semipermeable Membrane Devices (SPMDs) and bioassay-directed chemical analysis.

Schriks et al. 2015. Selection criteria to select in vitro bioassays for implementation and use.

Smith et al. 2015. Screening for contaminant hotspots in the marine environment of Kuwait using ecotoxicological and chemical screening techniques.

Thomas et al. 2004b. Identification of in vitro oestrogen and androgen receptor agonists in North Sea offshore produced water discharges.

Tollefsen et al. 2007. Oestrogen receptor (ER) agonists and androgen receptor (AR) antagonists in effluents from Norwegian North Sea oil production platforms.

Vrabie et al. 2010. Specific in vitro toxicity of crude and refined petroleum products. II Oestrogen (α and β) and androgen receptor-mediated responses in yeast assays.

Zhao et al. 2011. Estrogenic activity profiles and risks in surface waters and sediments of the Pearl River system in South China assessed by chemical analysis and in vitro bioassay.

Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No):

No, see Discussion-section (section 4) for motivation.

3.2.1.3. AR activation assay using hamster cell line

Title: AR activation assay using hamster cell line

Description:

Androgen receptor mediated transcription and gene expression is determined in in CHO-K1 hamster ovarian cancer cells. The bioassay reports binding based on 5-alpha-dihydrotestosterone (DHT) equivalents. The assay can also be used to detect anti-androgenic activity by co-exposing a substance/sample with DHT and using flutamide (FLU) as the standard.

An example of an AR activation assay using hamster cell line is AR EcoScreen™.

Validation Maturity:

Level 2: Expert laboratory.

Not validated for water samples (Schriks et al. 2015).

OECD 458: Stably transfected human androgen receptor transcriptional activation assay for detection of androgenic agonist and antagonist activity of chemicals.

Application with Environmental Samples:

This assay has been applied sporadically for environmental research (Schriks et al. 2015), and has been applied to estuarine, coastal and marine sediments (Vethaak et al. 2017).

Response to PAHs: Positive responses were found in estuarine, coastal and marine sediments but they could not be directly attributed to PAHs (Vethaak et al. 2017). Hamers et al. (2016) found that B(a)P contributed to anti-androgenic effects observed in the hamster cell line AR activation assay from passive sampler extracts.
Suitability for use with Passive Sampler Extracts: The assay has previously been applied in the assessment of passive sampler extracts (Hamers et al. 2016, Vethaak et al. 2017).
Interpretation of Results: High sensitivity to, but medium selectivity for, androgenic compounds (Schriks et al. 2015). Sample preparation is important in quantifying responses (Vethaak et al. 2017). Substances causing androgenic and anti-androgenic effects in the same sample may cause antagonistic effects (Hamers et al. 2016). Low risk trigger values (LR-TV) have also been proposed for the hamster cell line AR activation assay based on a comparison between the 5th percentile of the SSD determined for all compounds yielding a specific response in the bioassay (Hamers et al. 2016). A value below a LR-TV for AR antagonism of 25000 ng/L FLU-EQ; indicated no hazard from the passive sampler extract (Hamers et al. 2016).
Key Publications: Hamers et al. 2016. Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP): an effect-based strategy for cost-effective chemical water quality assessment**.** Vethaak et al. 2017. Toxicity profiling of marine surface sediments: A case study using rapid screening bioassays of exhaustive total extracts, elutriates and passive sampler extracts. Schriks et al. 2015. Selection criteria to select in vitro bioassays for implementation and use.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): Yes, see Discussion-section (section 4) for motivation.

3.2.1.4. AR activation assay using human bone marrow cell line

Title: AR activation assay using human bone marrow cell line
Description: Human bone marrow cell line (U2OS), incorporating the firefly luciferase gene coupled to Androgenic Responsive Elements (AREs) as a reporter gene for the presence of androgens and/or androgen-like compounds. An example of an AR activation assay using human bone marrow cell line is AR-CALUX®
Validation Maturity: Level 2: Expert laboratory. A Dutch guideline is available: Rijkswaterstaat RIKZ-Specie-08 guideline.
Application with Environmental Samples: The assay has been applied in a number of environmental monitoring studies (e.g. Leusch et al. 2014; Scott et al. 2014), and has also been applied to industrial and hospital wastewaters and WWTP effluents (van der Linden et al. 2008 cited in Scott et al. 2014).

Response to PAHs: No relevant information on the specific responses of this test to PAHs was identified based on literature searches undertaken for this project.
Suitability for use with Passive Sampler Extracts: No specific information identified regarding use with passive sampler extracts, but assay design utilises small sample volumes and provides results relatively rapidly.
Interpretation of Results: High sensitivity to, but medium selectivity for, androgenic compounds (Schriks et al. 2015). AR activation assay using human bone marrow cell line results can be expressed as bio-equivalents, calculated as EC50 of the reference compound (5α-dihydrotesterone), divided by $EC50_{(REF)}$ of the sample (Leusch et al. 2014). The REF is the Relative Enrichment Factor which takes into consideration enrichment of the sample (i.e. either by concentration steps or from passive sampler extracts). More sensitive than the YAS assay but is more vulnerable to cytotoxicity (Gehrmann et al. 2016). Sensitivity = 7 ng/l dihydrotestosterone equivalent (DHTEQ) for androgenic activity and 60 µg/l flutamide equivalent (FLUEQ) for anti-androgenic activity (Scott et al. 2014). An EQS for anti-AR activity of 25 µg Flu EQ/L is proposed by van der Oost et al. (2017a).
Key Publications: Gehrmann et al. 2016. (Anti-)estrogenic and (anti-)androgenic effects in wastewater during advanced treatment: comparison of three in vitro bioassays. Leusch et al. 2014. Assessment of wastewater and recycled water quality: A comparison of lines of evidence from *in vitro*, *in vivo* and chemical analysis. Scott et al. 2014. An assessment of endocrine activity in Australian rivers using chemical and *in vitro* analyses. Schriks et al. 2015. Selection criteria to select in vitro bioassays for implementation and use.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): Yes, see Discussion-section (section 4) for motivation.

3.2.1.5. Yeast Androgen Screen

Title: YAS assay
Description: The Yeast Androgen Screen is a recombinant yeast strain used for the identification of substances that can interact with the human androgen receptor (hAR). Dihydrotestosterone (DHT) equivalent factors (EEQ) are calculated by measuring the effective concentration required to elicit a 50% response in the exposed organisms (EC50) for the DHT positive control, and determining the percent of sample required to give an equivalent response. By co-exposing a substance/sample with DHT and using Flutamide as the standard, the YAS also allows testing for hAR-antagonistic activities (anti-androgenicity) by inhibition of DHT binding to the hAR. Cytotoxicity can be distinguished from effects by measuring growth of the yeast cells (620 nm) in parallel to the enzyme activity (540 nm) (CIS 2014).

Validation Maturity:

Level 3: Routine laboratory.

Well validated in general terms using a method based on Routledge & Sumpter (1996), and is considered mature. Highly reproducible between laboratories. In the process of being validated for water samples (Schriks et al. 2015). ISO guideline in preparation.

Application with Environmental Samples:

The YAS assay has been widely applied to environmental samples (Schriks et al. 2015; Smith et al. 2015).

The assay has also been used for the assessment of industrial effluents (Fang et al. 2012), sewage treatment works effluents (Thomas et al. 2002), water and sediment samples from a river receiving industrial effluent (Thomas et al. 2002; Urbatzka et al. 2007) and produced water in the North Sea (Thomas et al. 2004b; Tollefsen et al. 2007).

Response to PAHs:

Androgenic antagonist activity was reported in dissolved and oil phases of produced water from the North Sea (Tollefsen et al. 2007), whereas no androgenic agonist activity was observed (Thomas et al. 2004b).

Vrabie et al. (2011) tested four crude oils and seven refined oils and found that there was limited activity of the AR receptor. A single oil induced AR activity 11% greater than testosterone. A lack of androgenic agonist activity was consistent with other data in the literature for PAHs (Thomas et al. 2004b).

A synergistic effect on the AR, has been observed for anti-androgenic effect-based on certain fractions of a commercially-available engine oil for cars (Jonker et al. 2016).

Suitability for use with Passive Sampler Extracts:

No specific information identified regarding use with passive sampler extracts, but assay design utilises small sample volumes and provides results relatively rapidly.

Interpretation of Results:

Highly selective for, androgenic compounds but low sensitivity compared with other assays (Schriks et al. 2015).

A limited number of specific substances have androgenic activity whereas many substances (including PAHs) have been reported as AR antagonists. The large number of substances makes it difficult to attribute cause and effect.

LODs as low as 2 ng DHT/L and 0.01 ng DHT/L have been reported in surface water (freshwater) and produced water (marine water), respectively (Thomas et al. 2002; Thomas et al. 2004b). The influence of other non-androgenic compounds on the mechanism of the YAS assay is unclear (Thomas et al. 2002). Fractionation can help to identify androgenic compounds (Thomas et al. 2002; Urbatzka et al. 2007).

The androgenic activity of oils as environmental mixtures in themselves is complex to interpret, even without considering other substances that are also likely to be present in the environment (Vrabie et al. 2010).

Low sensitivity compared with the AR activation assay using human bone marrow cell line, but more less effected by matrix effects or cytoxicity (Gehrmann et al. 2016)

LOQ is approx. 340 ng DHT/L, by concentration of the sample (i.e. SPE) the LOQ is lowered by the REF (relative enrichment factor) (CIS 2014).

Key Publications: Fang et al. 2012. Assessment of hormonal activities and genotoxicity of industrial effluents using in vitro bioassays combined with chemical analysis. Gehrmann et al. 2016. (Anti-)estrogenic and (anti-)androgenic effects in wastewater during advanced treatment: comparison of three in vitro bioassays. Jonker et al. 2016. Synergistic androgenic effects of a petroleum product caused by the joint action of at least three chemically distinct compounds. Osman et al. 2015. Screening of multiple hormonal activities in water and sediment from the river Nile, Egypt, using *in vitro* bioassay and gonadal histology. Schriks et al. 2015. Selection criteria to select *in vitro* bioassays for implementation and use. Smith et al. 2015. Screening for contaminant hotspots in the marine environment of Kuwait using ecotoxicological and chemical screening techniques. Thomas et al. 2002. An assessment of in vitro androgenic activity and the identification of environmental androgens in United Kingdom estuaries. Thomas et al. 2004b. Identification of in vitro oestrogen and androgen receptor agonists in North Sea offshore produced water discharges. Tollefsen et al. 2007. Oestrogen receptor (ER) agonists and androgen receptor (AR) antagonists in effluents from Norwegian North Sea oil production platforms. Vrabie et al. 2010. Specific in vitro toxicity of crude and refined petroleum products. II Oestrogen (α and β) and androgen receptor-mediated responses in yeast assays. Urbatzka et al. 2007. Androgenic and antiandrogenic activities in water and sediment samples from the river Lambro, Italy, detected by yeast androgen screen and chemical analyses.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): No, see Discussion-section (section 4) for motivation.

3.2.1.6. Transthyretin-Binding (TTR-binding)

Title: Transthyretin-binding (TTR-binding)
Description: This assay involves the displacement of the thyroid hormone precursor thyroxine (T4) from its plasma transport protein transthyretin (TTR). Competitive binding of the sample to human TTR is assessed with ^{125}I-labelled and unlabelled T4. Once binding equilibrium is reached, ^{125}I-T4 binding is determined using a gamma counter and expressed as a percent of the control, yielding a T4-equivalent concentration.
Validation Maturity: Level 1: Research laboratory. This is a relatively novel assay, and validation activity is currently restricted to development within research facilities. Interpretation remains a barrier for use for regulatory purposes.
Application with Environmental Samples: The assay has been sporadically used in studies with environmental samples including effluents (Metcalf et al. 2013) and sediments (Vethaak et al. 2017).

Response to PAHs: PAH quinones and hydroxides have been shown to exhibit strong TTR-binding activity, whereas PAH ketones have no TTR-binding activity (Bekki et al. 2009).
Suitability for use with Passive Sampler Extracts: TTR binding activity assay has been applied in studies with passive samplers (Hamers et al. 2016, Vethaak et al. 2017).
Interpretation of Results: TTR-binding is less specific than other receptor assays and is affected by a range of different types of substance. It is therefore difficult to quantitatively relate the *in vitro* assay data to the concentrations of target analytes in environmental media. Metcalf et al. (2013) regarded this assay solely as a screening tool (Metcalf et al. 2013). TTR-binding activity was found in all sediment extracts collected from silicone passive sampler's *ex situ* (Vethaak et al. 2017), and in all passive sampler extracts from the TIPTOP study (Hamers et al. 2016).
Key Publications: Bekki K et al. 2009. Evaluation of toxic activities of polycyclic aromatic hydrocarbon derivatives using *in vitro* bioassays. Hamers et al. 2016. Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP): an effect-based strategy for cost-effective chemical water quality assessment. Metcalfe et al. 2013. A multi-assay screening approach for assessment of endocrine-active contaminants in wastewater effluent samples. Vethaak et al. 2017. Toxicity profiling of marine surface sediments: A case study using rapid screening bioassays of exhaustive total extracts, elutriates and passive sampler extracts.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): No, see Discussion-section (section 4) for motivation.

3.2.2. Genotoxicity

Five assays were selected which assess genotoxicity; the Ames test, the umuC assay, the micronucleus assay (as measured in a cell line), the activation of hGADD45 assay and the Nrf2-pathway activation assay . The Ames and umuC tests are not strictly *in vitro* tests because they assess effects in bacteria, but they have been considered in this section alongside *in vitro* tests because they evaluate a similar mechanism of action.

3.2.2.1. Ames/ Ames II

Title: Ames/ Ames II
Description: In the Ames test, *Salmonella* strains are exposed to samples for 48 hours in a microplate system, with and without metabolic activation (using liver S9). The induction of reverse mutations and number of revertants are determined. The Ames test measures reversions of frameshift or base pair mutations in the histidine operon. The Ames II test is a microplate based assay.
Validation Maturity: Level 3: Routine laboratory. Well validated in general terms: a standardised guideline exists (ISO 11350; OECD guideline 471, fluctuation method) and is considered mature. Reproducibility between laboratories can be an issue. Validated for use with water samples (Schriks et al. 2015). ISO 11350:2012. Water quality - Determination of the genotoxicity of water and waste water - *Salmonella*/microsome fluctuation test (Ames fluctuation test).
Application with Environmental Samples: The assays has been widely used for the assessment of environmental samples, based on very high number of scientific papers identified in the literature search with terms 'Ames' and (monitoring OR environment OR effluent OR toxicity OR regulatory OR risk assessment). Specific examples include drinking water (Schriks et al. 2015), uncontaminated freshwater (Di Paolo et al. 2016), surface water receiving industrial effluents (Gallampois et al. 2013; Vincent-Hubert et al. 2016) and samples relevant for the oil and gas industry (Zemanek et al. 1997, Lemos et al. 1994; Steliga et al. 2015).
Response to PAHs: The assay is responsive to benzo(a)pyrene with metabolic activation using S9 extracts (Reifferscheid and Grummt 2000; Neale et al. 2017), and the TA98 strain has been shown to detect PAHs with metabolic activation, indicating that reactive intermediate metabolites of PAH are likely causing the mutagenesis (Vincent-Hubert et al. 2016, Zemanek et al. 1997). Positive responses were detected in surface waters receiving industrial effluents using LDPE samplers, indicating the presence of frameshift mutagens (TA98 + S9), however no correlation was observed between PAH concentration and Ames results (Vincent-Hubert et al. 2016). Polar fractions from water-soluble extracts of petroleum-contaminated soils were observed to be weakly mutagenic in the TA98 strain without enzyme activation. A non-toxic response in the combined extract, while its fractionated components reveal a toxic response, may indicate that components within the bulk oil phase inhibit dissolution of potentially toxic components by retaining them within the immobilised oil phase (Zemanek et al. 1997). Sensitive to B(a)P at 60-400 µg/L (depending on version of the assay) (Reifferscheid and Grummt 2000).
Suitability for use with Passive Sampler Extracts: The Ames test has been widely used to assess passive sampler extracts (Rastall et al. 2004; Gallampois et al. 2013; Vincent-Hubert et al. 2016). Sample volumes may limit use with the standard Ames test (Rastall et al. 2004), but the Ames II test can be conducted in 96 well plates.

Interpretation of Results: The Ames test is a mutagenicity tests which detects fixed mutations that are heritable to the next cell generations. The test endpoint is based on number of revertants providing a positive or negative for mutagenicity. The assay is quantifiable based on scoring the number of revertants or percentage effect but it is more difficult to interpret data in terms of relative potency. Different strains of bacteria can be used, and effects observed in each strain can help identify the type of DNA mutation occurring (Brack et al. 2016). Neale et al. (2017) found that only one of 22 water pollutants tested, B(a)P, tested positive in the Ames test even though mode of action analysis would have predicted effects for more of the substances.
Key Publications: Di Paolo et al. 2016. Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts – Towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring. Gallampois et al. 2013. Integrated biological–chemical approach for the isolation and selection of polyaromatic mutagens in surface waters. Lemos et al. 1994. Genotoxicity of river water under the influence of petrochemical industrial complexes. Neale et al. 2017. Development of a bioanalytical test battery for water quality monitoring: Fingerprinting identified micropollutants and their contribution to effects in surface water. Rastall et al. 2004. The identification of readily bioavailable pollutants in Lake Shkodra/Skadar using Semipermeable Membrane Devices (SPMDs), bioassays and chemical analysis. Reifferscheid and Grummt. 2000. Genotoxicity in German surface waters - results of a collaborative study. Schriks et al. 2015. Selection criteria to select *in vitro* bioassays for implementation and use. Steliga et al. 2015. Changes in toxicity during treatment of wastewater from oil plant contaminated with petroleum hydrocarbons. Vincent-Hubert et al. 2016. Use of low density polyethylene membranes for assessment of genotoxicity of PAHs in the Seine River. Zemanek et al. 1997. Toxicity and mutagenicity of component classes of oils isolated from soils at petroleum- and creosote-contaminated sites.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): Yes, see Discussion-section (section 4) for motivation.

3.2.2.2. umuC Assay

Title: umuC assay
Description: This test uses the bacteria *Salmonella typhimurium*, indirectly measuring they activation of the umuC-gene and the SOS repair response via induction of the enzyme β-galactosidase. It is an indirect measure of DNA damage, or genotoxicity. The test is conducted with and without metabolic activation using liver S9.

Validation Maturity: Level 3: Routine laboratory. Well validated in general terms: a standardised guideline exists and is considered mature. Highly reproducible between laboratories. Validated for use with water samples (Schriks et al. 2015). ISO 13829:2000, Water quality - determination of the genotoxicity of water and wastewater using the umu-test.
Application with Environmental Samples: The assay is considered to be a sensitive standard method of estimating the genotoxicity, and has been widely applied to environmental samples, including industrial wastewater, (Escher et al. 2008, Fang et al. 2012, Leusch et al. 2014, Vincent-Hubert et al. 2016), surface waters (Reifferscheid et al. 1991, Reifferscheid and Grummt 2000, Kittinger et al. 2015, Han et al. 2016; Vincent-Hubert et al. 2016; Žegura et al. 2009), drinking water (Schriks et al. 2015) and with oil residues in freshwater and brackish water (Bi et al. 2011).
Response to PAHs: UmuC is responsive to B(a)P with metabolic activation using S9 extracts (Reifferscheid and Grummt 2000). Fang et al. (2012) found that PAHs were minor contributors to the genotoxicity of industrial effluents (Fang et al. 2012). Sensitive to B(a)P at 26-80 µg/L (depending on version of the assay) (Reifferscheid and Grummt 2000).
Suitability for use with Passive Sampler Extracts: The assay has been applied in a number of studies investigating the effects of passive sampler extracts (Muller et al. 2007, Bi et al. 2011; Vincent-Hubert et al. 2016). A miniaturised version of the assay has been developed (Reifferscheid et al. 1991).
Interpretation of Results: The umuC test is an indicator test for genotoxicity, evaluating the ability of a substance to induce DNA damage. It is relatively sensitive, but has low specificity (Leusch et al. 2014; Schriks et al. 2015). The assay has been demonstrated to be sensitive to > 400 chemicals, so specific causes can be difficult to attribute in environmental samples (Kittinger et al. 2015). Weak responses may be linked to the presence of compounds that inhibit the pathway of the assay (Vincent-Hubert et al. 2016). However, can detect mutagenic compounds and mixtures which do not illicit responses in the Ames test (Wieczerzak et al. 2016). According to ISO 13829 an induction rate of ≥1.5 is taken as a signal for mutagenic potency in water samples. The results from the umuC assay can also be normalised to genotoxic units (GTU) where a threshold of 1 GTU calculated as 1/EC50 of a reference substance is considered indicative for DNA damage (Leusch et al. 2014). Genotoxicity can also be expressed using a diuron-based relative enrichment factor (Muller et al. 2007).
Key Publications: Fang et al. 2012. Assessment of hormonal activities and genotoxicity of industrial effluents using in vitro bioassays combined with chemical analysis. Han et al. 2016. Evaluation of genotoxic effects of surface waters using a battery of bioassays indicating different mode of action. Kittinger et al. 2015. Preliminary toxicological evaluation of the river Danube using *in vitro* bioassays.

Leusch et al. 2014. Assessment of wastewater and recycled water quality: A comparison of lines of evidence from in vitro, in vivo and chemical analysis.

Muller et al. 2007. Combining passive sampling and toxicity testing for evaluation of mixtures of polar organic chemicals in sewage treatment plant effluent.

Reifferscheid and Grummt. 2000. Genotoxicity in German surface waters - results of a collaborative study.

Reifferscheid et al. 1991. A microplate version of the SOS/umu-test for rapid detection of genotoxins and genotoxic potentials of environmental samples.

Schriks et al. 2015. Selection criteria to select in vitro bioassays for implementation and use.

Vincent-Hubert et al. 2016. Use of low density polyethylene membranes for assessment of genotoxicity of PAHs in the Seine River.

Wieczerzak et al. 2016. Bioassays as one of the Green Chemistry tools for assessing environmental quality: A review.

Žegura et al. 2009. Combination of in vitro bioassays for the determination of cytotoxic and genotoxic potential of wastewater, surface water and drinking water samples.

Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No):

Yes, see Discussion-section (section 4) for motivation.

3.2.2.3. hGADD45 activation assay

Title: hGADD45 activation assay

Description:

The assay is a Green Fluorescent Protein (GFP) reporter assay based on the expression of the human GADD45a (hGADD45) gene. The hGADD45a gene has a role in the genotoxicity stress response.

A version of the assay (GreenScreen™ EM) based on yeast cultures of *Saccharomyces cerevisiae* has been marketed and trailed for use with environmental samples (Daniel et al. 2004). The same assay is more readily available based on a human cell line (GreenScreen™ HC). A similar assay the BlueScreen™ HC is also available but the test endpoint is based on a luciferase reporter. The tests are conducted with and without metabolic activation using liver S9.

Validation Maturity:

Level 2: Expert laboratory.

The assay based on yeast cultures of *Saccharomyces cerevisiae* appears not to be readily available. The assay based on the human cell line is currently undergoing standardisation, but is not well validated for water samples (Schriks et al. 2015). Further, it is extensively validated for several hundred pure chemical compounds, mainly for pharmaceutical application.

Application with Environmental Samples:

The assay based on the human cell line has been sporadically applied to environmental samples (Schriks et al. 2015) and the assay based on yeast cultures of *Saccharomyces cerevisiae* has been sporadically applied on industrial effluents (Daniel et al. 2004).

Response to PAHs:

No relevant information on the specific responses of this test to PAHs was identified based on literature searches undertaken for this project.

Suitability for use with Passive Sampler Extracts:
No specific information identified regarding use with passive sampler extracts, but assay is specifically designed to be high throughput and use small sample volumes.
Interpretation of Results:
Highly sensitive to, and specific for, compounds causing DNA damage (Schriks et al. 2015), but interferences can occur from particulate matter and bacterial contamination (Daniel et al. 2004). A positive genotoxicity result is indicated by a 1.3 fold induction compared with the control. A dose- response is also preferred (Daniel et al. 2004).
Key Publications:
Daniel et al. 2004. Results of a technology demonstration project to compare rapid aquatic toxicity screening tests in the analysis of industrial effluents. Schriks et al. 2015. Selection criteria to select *in vitro* bioassays for implementation and use.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No):
No, see Discussion-section (section 4) for motivation.

3.2.2.4. p53-pathway activation assay

Title: p53-pathway activation assay
Description:
A human cell line (U2OS) derived pathway selective reporter gene assay. In this assay, a firefly luciferase gene has been coupled to p53 Responsive Elements. The luciferase serves as a reporter gene for the presence of p53-pathway activating compounds. Actinomycin D is used as a positive control. The test is conducted with and without metabolic activation using liver S9. An example of the p53-pathway activation assay is p53 CALUX®.
Validation Maturity:
Level 2: Expert laboratory. This assay is undergoing standardisation, but is not well validated for water samples (Schriks et al. 2015).
Application with Environmental Samples:
A p53-pathway activation assay has been sporadically applied to environmental samples (Schriks et al. 2015).
Response to PAHs:
The assay requires metabolic activation to be sensitive to B(a)P (van der Linden et al. 2014). A false positive result for genotoxicity was reported for phenanthrene (without S9) (van der Linden et al. 2014).
Suitability for use with Passive Sampler Extracts:
No specific information identified regarding use with passive sampler extracts, but assay design utilises small sample volumes and provides results relatively rapidly.

Interpretation of Results:

Highly sensitive to, and specific for, compounds causing DNA damage (Schriks et al. 2015).

A positive response in the p53-pathway activation assay is indicated when at least one concentration of the substance shows an increase of at least 50% (a 1.5-fold induction) or 100% (2 fold induction in the presence of S9 (van der Linden et al. 2014).

The assay (with and without S9) detected 18 of 20 tested genotoxic compounds as positive, with sensitivity and specificity being comparable with a hGADD45 activation assay (van der Linden et al. 2014). A false positive result was reported for phenanthrene (without metabolic activation) while metabolic activation increases the number of false positive results.

Key Publications:

Schriks et al. 2015. Selection criteria to select *in vitro* bioassays for implementation and use.

van der Linden et al. 2014. Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress.

Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No):

No, see Discussion-section (section 4) for motivation.

3.2.2.5. Micronucleus

Title: Induction of Micronucleus in cells (e.g. in *in vitro* cell lines such as V79)

Description:

The micronucleus assay detects the frequency of micronuclei formation and nuclear abnormalities in cells. For environmental testing the assay is often performed with the permanently growing Chinese hamster lung fibroblast cell line V79, but can be performed with many cell types.

Validation Maturity:

Level 3: Routine laboratory.

The assay is well validated in general terms: a standardised guideline exists and is considered mature. Validated for use with water samples (Schriks et al. 2015). Validation with cell lines other than V79 unknown (Brinkmann et al. 2014).

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

Application with Environmental Samples:

The micronucleus assay has been widely applied to environmental samples (Nunes et al. 2011; Schriks et al. 2015; Reifferscheid et al. 2008), including untreated water samples from a river basin influenced by urban and industrial activity (Bianchi et al. 2015), surface water (Garcia et al. 2011; Han et al. 2016) and petroleum refinery effluent (Hara and Marin-Morales 2017).

Response to PAHs:

For derivatives of benzofuran, methylation had a decreasing effect on the genotoxicity. Some PAHs that tested negative in mammalian systems exhibited a positive response in a fish cell line (e.g. dibenzofuran) (Brinkmann et al. 2014).

In a study by Hara and Marin-Morales (2017), micronucleus activity was only detected after the first treatment of petroleum refinery effluent (physico-chemical treatment) and not in the source water, post-biological treatment effluent, final effluent or downstream of the discharge site (Hara and Marin-Morales 2017).

Suitability for use with Passive Sampler Extracts:
No specific information identified regarding use with passive sampler extracts, but assay design utilises small sample volumes. Manual assessment of slides can be a limitation to the assay in terms of throughput (Brack et al. 2016).
Interpretation of Results: The micronucleus test detects non-repairable damage, such as clastogenic[4] and aneugenic[5] lesions rather than recent lesions that can be repaired, such as breaks and alkali-labile sites (as measured in the Comet assay) (Bianchi et al. 2011). Highly sensitive to, and relatively specific for, mutagenic compounds (Schriks et al. 2015). Results can be expressed as relative potency (REP) compared with a known mutagenic substance (Brinkmann et al. 2014). Micronucleus formation *in vitro* was observed in the V79 hamster cell line in following exposure to surface water samples but were not detected *in vivo* (in onion root cells) suggesting that the cell lines are more sensitive for detecting a genotoxic effect (Nunes et al. 2011). The sensitivity of non-standard cell lines may vary depending on the sample type (e.g. water, sediment) (Garcia et al. 2011). For a substance to be considered genotoxic there must be a significant increase in micro-nucleated cells compared with the corresponding negative control, and the number of micro-nucleated cells in treated cultures should exceed the range of the historical control data. The dose response may also be important for biological relevance. (Reifferscheid et al. 2008).
Key Publications: Bianchi et al. 2015. Evaluation of genotoxicity and cytotoxicity of water samples from the Sinos River Basin, southern Brazil. Brinkmann et al. 2014. Genotoxicity of heterocyclic PAHs in the micronucleus assay with the fish liver cell line RTL-W1. Garcia et al. 2011. Micronucleus study of the quality and mutagenicity of surface water from a semi-arid region. Han et al. 2016. Evaluation of genotoxic effects of surface waters using a battery of bioassays indicating different mode of action. Hara RV, Marin-Morales MA. 2017. In vitro and in vivo investigation of the genotoxic potential of waters from rivers under the influence of a petroleum refinery (São Paulo State – Brazil). Nunes et al. 2011. Genotoxic assessment on river water using different biological systems. Reifferscheid et al. 2008. Measurement of genotoxicity in wastewater samples with the *in vitro* micronucleus test—Results of a round-robin study in the context of standardisation according to ISO. Schriks et al. 2015. Selection criteria to select *in vitro* bioassays for implementation and use.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): No, see Discussion-section (section 4) for motivation.

[4] A clastogen is a mutagenic agent giving rise to or inducing disruption or breakages of chromosomes, leading to sections of the chromosome being deleted, added, or rearranged.

[5] An aneugen is a substance that causes a daughter cell to have an abnormal number of chromosomes or aneuploidy.

3.2.3. Oxidative stress

Two oxidative stress assays were selected for assessment, both related to the induction of the Antioxidant Response Element (ARE) pathway: the Nrf2-pathway activation assay and the AREc32 activation assay.

3.2.3.1. Nrf2-pathway activation assay

Title: Nrf2-pathway activation assay
Description: Measure oxidative stress Nrf2 transcriptional activity. The Nrf2-pathway activation assay is composed of a human cell line (U2OS) containing the firefly luciferase gene under control of four Electrophile Responsive Elements (EpREs). The luciferase serves as a reporter gene for activation of the Nrf2 pathway. An example of an Nrf2-pathway activation assay is the Nrf2 CALUX®.
Validation Maturity: Level 2: Expert laboratory. Undergoing standardisation and validation for use with water samples. Moderately reproducible between laboratories (Schriks et al. 2015).
Application with Environmental Samples: Not widely applied to environmental samples (Schriks et al. 2015).
Response to PAHs: Very little relevant information on the specific responses of this test to PAHs was identified based on literature searches undertaken for this project, however, van der Linden et al. (2014) reported that B(a)P did not produce a response in this assay (van der Linden et al. 2014).
Suitability for use with Passive Sampler Extracts: No specific information identified regarding use with passive sampler extracts, but assay is specifically designed to be high throughput and use small sample volumes.
Interpretation of Results: Oxidative stress occurs when there is disturbance in the normal redox state of cells. It can result in the generation of reactive oxygen species which can cause toxic effects through the production of peroxides and free radicals that damage all components of the cell, including strand breaks in DNA which can lead to genotoxicity. Response of the Nrf2-pathway activation assay could provide mechanistic information for observed genotoxic effects in other assays (van der Linden et al. 2014). A positive response in the Nrf2 assay is considered to be when at least one tested concentration of a substance registers an increase in response of at least 50% (a 1.5-fold induction) (van der Linden et al. 2014).
Key Publications: Schriks et al. 2015. Selection criteria to select *in vitro* bioassays for implementation and use. Van der Linden et al. 2014. Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): No, see Discussion-section (section 4) for motivation.

3.2.3.2. ARE c32 activation assay

Title: AREc32 activation assay
Description: The induction of the Nrf2-mediated oxidative stress response pathway is measured in the human breast cancer cell line MCF7. The Nrf2-Antioxidant Response Element (ARE) pathway is responsive to chemicals that cause oxidative stress. The induction of Nrf2 is proportional to the amount of luciferase produced by the cells, which can be assessed by a bioluminescence fluorimeter. The results can be expressed relative to a positive control tert-Butylhydroquinone (tBHQ).
Validation Maturity: Level 2: Expert laboratory. Undergoing standardisation. There has been some inter laboratory comparison (Neale et al. 2017) and is moderately reproducible between laboratories (Schriks et al. 2015).
Application with Environmental Samples: This assay has been widely applied to environmental samples (Schriks et al. 2015), including drinking water and effluents from sewage treatment plants (STPs) (Escher et al. 2012 and 2013).
Response to PAHs: B(a)P and benzo(b)fluoranthene are inducers of the assay response with EC10's of 2.22 x 10^{-7} M and 2.33 x10^{-7} M respectively (Neale et al. 2017). Induction of the response with B(a)P is higher with metabolic activation using S9 extracts but for other substances its presence detoxifies the reactive chemical and therefore S9 extracts are not always used in mixture assessment (Escher et al. 2013).
Suitability for use with Passive Sampler Extracts: No specific information identified regarding use with passive sampler extracts, but assay design utilises small sample volumes and provides results relatively rapidly.
Interpretation of Results: Highly selective for, and sensitive to, compounds causing cellular damage (Schriks et al. 2015 A relatively high number of substances induce the antioxidant response meaning it can be difficult to establish cause and effect (Escher et al. 2012; Neale et al. 2017). The limit of reporting (LOR) is three times the standard deviation of the controls. The concentration causing an induction ratio (IR) of 1.5 ($EC_{IR1.5}$) has been used as an effect benchmark value for drinking water treatment (Escher et al. 2012). A effect based trigger value of $EC_{IR1.5}$ = 6 REF (Relative Enrichment Factor) is proposed, with sample values < 6 REF indicating a concern for further testing. Cytotoxic effects may mask induction of oxidative stress depending on the sample. Concentration addition is an appropriate model for mixture effects for chemicals that are active in AREc32; inactive chemicals do not influence the overall response. Endotoxins may influence oxidative stress response and some chemicals may trigger the synthesis of antioxidants and metabolic enzymes (Escher et al. 2013).
Key Publications: Escher et al. 2012. Water quality assessment using the AREc32 reporter gene assay indicative of the oxidative stress response pathway. Escher et al. 2013. Most oxidative stress response in water samples comes from unknown chemicals: The need for effect-based water quality trigger values. Schriks et al. 2015. Selection criteria to select *in vitro* bioassays for implementation and use.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): Yes, see Discussion-section (section 4) for motivation.

3.2.4. Metabolism

Some hydrocarbons activate the Aryl hydrocarbon receptor (AhR) and induce cytochrome P4501A activity. It is therefore relevant to include assays for AhR mediated activity in this review. Two assays were selected for Dioxin-Responsive (DR) luciferase, and one for 7-ethoxyresorufin-O-deethylase (EROD) induction in a cell line.

3.2.4.1. DR AhR activation assay

Title: DR AhR activation assay
Description: In the Dioxin-Responsive (DR) AhR activation assay, the aryl hydrocarbon receptor (AhR) binding activity in the rat hepatoma H4IIE cell line is measured using a luminometer. Results are expressed as a percentage of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) binding activity. Sample preparation steps are required to limit interferences. The test duration is usually 24 hours which prioritises the detection of more persistent dioxin like substances. For detection of activity of PAHs results are expressed as a percentage of B(a)P binding activity and shorter test durations may be used. Examples of DR AhR activation assays are DR CALUX® (Chemical-Activated Luciferase Gene Expression) and DR LUC in the H4IIE cell line.
Validation Maturity: Level 3: Routine laboratory. AhR activation assays have been well validated in general terms: a standardised guideline exists and is considered mature. Highly reproducible between laboratories. In the process of being validated for water samples (Schriks et al. 2015). Good reproducibility in inter and intralaboratory trials (Eichbach et al. 2014). ICES TIMES No. 55. Protocol for measuring dioxin-like activity in environmental samples using *in vitro* reporter gene DR-LUC assays. Different DR LUC cell lines exist and the exposure duration should be optimised for each assay (Vrabie et al. 2009) and may differ in sensitivity (Eichbach et al. 2014).
Application with Environmental Samples: The assay is widely applied to environmental samples (Schriks et al. 2015), including sediments (Vethaak et al. 2017) and produced water from oil and gas installations (Balaam et al. 2009).
Response to PAHs: Bekki et al. (2009) tested 25 individual PAHs, of which eleven had AhR agonist activity and six had AhR antagonist activity. PAH ketones and quinones had stronger binding activities than hydroxylated PAHs, however PAH derivatives possess weaker binding activity for AhR than parent PAHs overall. Although PAHs can be detected with a DR AhR activation assay, the assay has limitations for detecting PAHs. These include the fact that PAHs are rapidly metabolised by the cells in the reporter assay. Also, during the pre-treatment step with sulphuric acid some PAHs and other less stable compounds are removed (Pieterse et al. 2013). Receptor mediated activity and antagonism was investigated using four crude oils and seven refined petroleum products (Vrabie et al. 2009). The potency of the substances (based on EC50's) was lower than for the reference substances but all crude oils showed AhR activity up to approximately 150% of the maximum B(a)P induction, while refined oils showed a response between 46 -195% of the maximum B(a)P induction. Supermaximal responses can be explained by the presence of a chemical with higher binding affinity for the receptor or by mixture effects. The authors conclude that oils probably contain different types of AhR agonists and that a single chemical is not responsible for all effects observed. Lower potencies are explained by the fact that AhR agonists constitute only a fraction of some of the oils.

Suitability for use with Passive Sampler Extracts:

Assay has previously been applied in the assessment of passive sampler extracts (Hamers et al. 2016, Vethaak et al. 2017).

A linear relationship between bioassay response and deployment time was observed for the Speedisk extracts while in the silicone rubber extracts the response of the two bioassays levelled off towards the end of the test period. Levelling of the response was relatively longer for the DR-LUC assay, which responds to more hydrophobic compounds compared with other assays (Hamers et al. 2016).

Interpretation of Results:

AhR agonists are associated with long-term toxicity to the liver, reproduction system, endocrine system and can cause cancer (Bekki et al. 2009). The assay is highly selective for, and sensitive to, dioxins (Schriks et al. 2015) and dioxin-like substances (chlorinated planar aromatic compounds) (Balaam et al. 2009). While the assay may be relatively specific, ToxCast reports AhR-agonistic potencies for 2132 compounds in the DR-LUC bioassay (Hamers et al. 2016). The presence of AhR antagonists may underestimate the assay results and the solubility of highly non-polar heterocyclic aromatic compounds may affect results (Hinger et al. 2011, Vrabie et al. 2009).

Sample preparation is important in quantifying responses (Vethaak et al. 2017). Cytotoxicity check and dose response curves are recommended. Luciferase induction can be reduced at the highest concentration (even in the absence of measurable cytotoxicity) which may be due to toxicity, interference or quenching of the light detection due to colour in the samples, or a reduction in bioavailability (Vrabie et al. 2009).

Hamers et al. (2016) used a DR LUC in a H4IIE cell line with two sample preparation methods. The stable extract was sulphuric acid treated and used to indicate dioxin exposure while the total extract was used to indicate effects from PAHs.

Alternatively, an AhR activation assay (with more specificity for PAHs) has been optimised for detecting PAHs (Pieterse et al. 2013).

Results can be expressed as B(a)P induction equivalents (IEQ-B(a)P) or TCDD induction equivalents (IEQ-TCDD) per mass of oil. These units can be used to compare concentrations within similar abiotic matrixes and be used for prioritisation (Vrabie et al. 2009). The units have been used as a guide for assessing environmental risks but should not be considered as a definitive result.

Significant correlation was found between DR-LUC responses and the sum of 6 PAH concentrations for sediment extracts collected using silicone passive samplers ex situ, however PAHs only accounted for up to 30% of the response (Vethaak et al. 2017). In another study with silicone rubber extract passive sampler extracts measured PAHs explained a high proportion of the activity in the DR assay (Hamers et al. 2016).

LOQ: 1 pg 2,3,7,8-TCDD equivalents per amount of material processed (CIS 2014).

LOD: 0.3-0.6pM; EC50 7.6 – 14 TCDD (Eichbaum et al. 2014). Generally, more sensitive than other assays (e.g. EROD induction).

Low risk trigger values (LR-TV) have been proposed for DR LUC assays based on a comparison between the 5th percentile of the SSD determined for all compounds yielding a specific response in the relevant bioassay (Hamers et al. 2016). A value below this LR-TV for AhR agonism of 16.2 pg TCDD-EQ/L indicates no risk. An alternative effect based trigger value of 50 pg-TEQ/L has been proposed (van der Oost et al. 2017).

Key Publications:

Balaam et al. 2009. Identification of nonregulated pollutants in North Sea-produced water discharges.

Bekki K et al. 2009. Evaluation of toxic activities of polycyclic aromatic hydrocarbon derivatives using in vitro bioassays.

Eichbaum et al. 2014. In vitro bioassays for detecting dioxin-like activity — Application potentials and limits of detection, a review.

Hamers et al. 2016. Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP): an effect-based strategy for cost-effective chemical water quality assessment.

Hinger et al. 2011. Some heterocyclic aromatic compounds are Ah receptor agonists in the DR-CALUX assay and the EROD assay with RTL-W1 cells.

Schriks et al. 2015. Selection criteria to select *in vitro* bioassays for implementation and use.

Vethaak et al. 2017. Toxicity profiling of marine surface sediments: A case study using rapid screening bioassays of exhaustive total extracts, elutriates and passive sampler extracts.

Vrabie et al. 2009. Specific *in vitro* toxicity of crude and refined petroleum products. 1. Aryl hydrocarbon receptor–mediated responses.

Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No):

Yes, see Discussion-section (section 4) for motivation.

3.2.4.2. AhR activation assay (with more specificity for PAHs)

Title: AhR activation assay (with more specificity for PAHs)

Description:

In a PAH specific AhR activation assay, the aryl hydrocarbon receptor (AhR) binding activity in the rat hepatoma H4IIE cell line is measured using a luminometer. Results are expressed as a percentage of benzo(a)pyrene binding activity. Sample preparation steps are required to limit interferences. The test duration is usually 4-6 hours which prioritises the less persistent PAH substances.

An example of an AhR activation assay with more specificity for PAHs is PAH CALUX®.

Validation Maturity:

Level 2: Expert laboratory.

Method based on DR AhR activation assay. Recently optimised for specific use in detecting PAHs (Pieterse et al. 2013). In the process of being validated for water samples (Schriks et al. 2015).

Application with Environmental Samples:

DR AhR activation assays has been widely applied to environmental samples, often with a shorter duration for detecting PAHs (Schriks et al. 2015).

Response to PAHs:

This is a very specific assay with high predictability, particularly for carcinogenic PAHs (Pieterse et al. 2013). The cell line that is used more specifically responds to AhR ligands (compared with an AhR activation assay), avoiding interference in complex mixtures (Pieterse et al. 2013). There is also a reduction of influence of cellular metabolism of PAHs and a shorter exposure time.

Suitability for use with Passive Sampler Extracts:

The assay is based on a DR AhR activation assays which is widely used with passive samplers**.**

Interpretation of Results:

Relative potencies (REPs) are obtained by dividing the EC50 of B(a)P (the reference compound), by that of the EC50 of the concerning PAH. PAH-activities from samples with unknown contents are quantified as B(a)P equivalents (BEQ). Theoretical total BEQ values based on REP values can be determined for environmental mixtures. Some PAHs (e.g.

fluoranthene, 2-aminoanthracene and phenanthrene) may have an antagonistic effect on the assay but this is not expected to have a major influence on the outcome (Pieterse et al. 2013). Substances detected using this assay are likely to be more readily degradable and therefore less persistent than those detected using the DR-LUC assay (after 24 hours) and therefore may be considered of lower environmental concern. Assay also sensitive to other substances including dioxins and pharmaceuticals (e.g. cyclophosphamide) (van der Oost et al. 2017). An effect based trigger value of 150 ng B(a)P-EQ/L has been proposed (van der Oost et al. 2017) LOQ: 0.45 ng B(a)P equivalents per litre of water.
Key Publications: Pieterse et al. 2013. PAH-CALUX, an optimized bioassay for AhR-mediated hazard identification of polycyclic aromatic hydrocarbons (PAHs) as individual compounds and in complex mixtures. Schriks et al. 2015. Selection criteria to select *in vitro* bioassays for implementation and use. Van der Oost et al. 2017. SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: part I – Model design and effect-based trigger values.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): Yes, see Discussion-section (section 4) for motivation.

3.2.4.3. EROD

Title: 7-ethoxyresorufin-O-deethylase (EROD) Activity measured in *in vitro* cell lines
Description: The ability of a substance to induce EROD activity in rainbow trout (*Onchorhynchus mykiss*) liver cell line (RTL-W1) is measured to determine CYP1A induction. CYP1A induction, either directly or indirectly, can lead to the development of cancer or toxicity.
Validation Maturity: Level 2: Expert laboratory. Standard methods are available for determining EROD activity in fish *in vivo*. *In vitro* assays have been demonstrated to be a suitable substitute for *in vivo* assays (Billiard et al. 2004; Bols et al. 1999) but they are not yet standardised. ISO23893-2:2007. Water quality - Biochemical and physiological measurements on fish - Part 2: Determination of ethoxyresorufin-O-deethylase (EROD).
Application with Environmental Samples: This assay has been applied to water samples (Emelogu et al. 2013), and water and sediment samples associated with the oil and gas industry (Harman et al. 2010; Suares-Rocha et al. 2015).
Response to PAHs: Behrens et al. (2001) tested nine individual PAHs, of which five (B(a)P, 3-methylcholanthrene, benzo(*b*)fluoranthene, chrysene and benzo(*a*)anthracene) consistently induced EROD activity, with EC50 values in the range 57.33 to 455.18 nM. Anthracene, fluoranthene, phenanthrene, and pyrene are not inducers of EROD activity.

No effects were observed in sediment samples associated with an oil refinery, possibly because the PAHs present were not EROD inducers (Suares-Rocha et al. 2015).
Suitability for use with Passive Sampler Extracts: Assay has previously been applied in the assessment of passive sampler extracts (Rastall et al. 2004).
Interpretation of Results: The EROD assay may give false negatives due to cytotoxicity or enzyme inhibition by other substances in the mixture; a dose-response curve is recommended (Behrens et al. 2001; Rastall et al. 2004). Lower sensitivity of the assay to model inducers TCDD and B(a)P was found when using surface or groundwater compared with tissue culture water, possibly due to adsorption to particles (Schirmer et al. 2004). Response can be variable (compared with receptor assays) and vulnerable to effects of other substances which can cause substrate inhibition reducing the sensitivity of the assay (Eichbaum et al. 2014). Theoretical LOD: 1 pM TCDD; EC50 5pM TCDD (Eichbaum et al. 2014). Does not respond to all PAHs (e.g. fluoranthene, phenanthrene) (Rastall et al. 2004). Results expressed as TCCD equivalents (EQs) which can be compared with bioassay-derived TCDD toxicity equivalent factors (TEF).
Key Publications: Behrens et al. 2001. Polycyclic aromatic hydrocarbons as inducers of cytochrome P4501A enzyme activity in the rainbow trout liver cell line, RTL-W1, and in primary cultures of rainbow trout hepatocytes. Eichbaum et al. 2014. In vitro bioassays for detecting dioxin-like activity — Application potentials and limits of detection, a review. Rastall et al. 2004. The identification of readily bioavailable pollutants in Lake Shkodra/Skadar using Semipermeable Membrane Devices (SPMDs), bioassays and chemical analysis. Schirmer et al. 2004. Applying whole water samples to cell bioassays for detecting dioxin-like compounds at contaminated sites. Suares-Rocha et al. 2015. Assessment of cytotoxicity and AhR-mediated toxicity in tropical fresh water sediments under the influence of an oil refinery.
Part of the suite of bioassays that can be used in combination with passive samplers for the assessment of waters receiving refinery effluents (Yes/No): No, see Discussion-section (section 4) for motivation.

4. DISCUSSION

The testing of passive sampler extracts (rather than directly sampled environmental samples) limits the volume of material available for conducting biological assessments, which in turn limits the type of EBTs that can be applied to those that can be conducted with small volumes of sample. If a number of EBTs are to be performed using the same extract (as would be desirable in the comparison of sites receiving refinery discharges), the low extract volumes also limit the total number of assays that can be conducted. The aim, therefore, should be to conduct a small number of targeted assays to screen samples across as many relevant endpoints as possible, rather than attempting to apply a large array of different tests investigating the same type of effects. It should also be noted that the bioassays assessed in the present study are by no means assessed in terms of availability with regard to being “available techniques” as defined by the Industrial Emissions Directive (IED; 2010/75/EU) article 3(10).

It is important to emphasise that the application of EBTs to passive sampler extracts is likely to yield relatively crude outputs, which may be useful for the screening and prioritisation of sites for further investigation, but will not generate results that indicate actual environmental effects. For example, Muller et al. (2007) found that EBTs applied to passive sampler extracts were useful in comparing the toxicity of wastewater treatment effluents at two different sites and sampling occasions, and the results from the tests (*Allivibrio fischeri* toxicity test, umuC assay and Max-I-PAM assay) were able to demonstrate an improvement in the water quality at one of the sites as a result of an upgrade to a UV disinfection system, but they were not sufficient to derive hazard data that could be utilised in risk assessment.

The TIPTOP study was able to prioritise sites for further investigation based toxicity profiles generated using combined information obtained with *in vivo* and *in vitro* assays. The authors use several strategies to interpret the data and argue that the information would be sufficient to conclude on “good ecological status” for a water body and that the measurements provide a more informative conclusion compared with an approach based only on chemical analysis. The passive sampler extracts showed relatively low toxicity, with the observed toxicity being mainly attributed to more polar substances obtained on a Speedisk passive sampler device. The substances causing the effects were however not identified. Nevertheless, the combination of *in vitro* EBTs and passive sampling assays may, at least, provide some insight into the types of substances or substance properties (e.g. genotoxic, endocrine-active) present in environmental samples.

4.1. WHOLE ORGANISM ASSAYS

While most of the well-established whole organism EBTs are well validated and suitable for environmental assessments, they are specifically designed to derive measurements of general apical toxicity to individuals or populations of exposed organisms, and the results from such tests rarely provide much insight into the mechanistic basis of toxicity, or whether the samples to which they are exposed have specific hazardous properties (e.g. oestrogenicity, mutagenicity, genotoxicity, etc.). Nevertheless, such assays are considered to represent (as near as is possible in the laboratory) *actual* population-related effects in the environment, rather than *potential* effects on organisms (as is generally derived in *in vitro* cell-based assays). It is considered sensible, therefore, to utilise both whole organism and cellular EBTs when assessing the hazard of environmental samples (i.e. mixtures of substances) in order to cover both apical, population-relevant and mechanistic responses. In some cases, it may even be possible to assess both general and mechanistic toxicity in the same study (e.g. by conducting cellular assays on cells removed from organisms exposed in the laboratory).

In the context of the present study, the requirement to primarily consider those EBTs that can be applied to passive sampler extracts will, however, mean that the range of suitable whole organism assays is restricted to those that can be conducted over relatively short timeframes (i.e. generally acute or short-term (sub)-chronic tests), and which require only relatively small volumes of sample, or at least that are amenable to miniaturisation to a sufficient degree. The need to potentially assess a large number of extracts concurrently over a relatively short timeframe means that assay testing systems that allow a large number of individual assays to be undertaken concurrently (so called high throughput assays) are also likely to be most appropriate for use in exposures with passive sampler extracts. For these reasons, even some tests that have been recommended for use in effluent testing (such as the *Lemna minor* growth test) are not considered suitable for use with passive samplers.

The general objective when considering any suite of whole organism ecotoxicity assays (regardless of the regulatory regime under which they are applied) is to attempt to obtain as wide a taxonomic coverage as possible, while also considering the matrix of interest (freshwater, marine water, sediment) and accounting for the practical and ethical limitations of any testing programme (which generally restrict the selection to a very small number of ‘representative’ species). Traditionally, this selection has been based loosely on trophic level (e.g. primary producers, primary consumers, and secondary consumers) or very broad taxonomic groupings (algae, invertebrates, fish). Clearly, each of these taxonomic groups contains a great many species, often with very different life histories, feeding strategies, etc. A toxicity test with *Daphnia* (a filter-feeding crustacean) is, for example, unlikely to provide much directly relevant information on the toxicity of the same substance to sensitive freshwater snails or aquatic insect larvae, even if they co-exist at a site. Nevertheless, a balance needs to be struck between attempting to test every potentially sensitive species likely to be present at a site and what is actually practical. For this reason, test species should be selected which are as representative as possible of the species present in the general environment (usually restricted simply to the matrix), have been demonstrated to be sensitive to a wide range of different substances and for which reliable and repeatable toxicity testing methods exist. In some cases, the ‘algae/ invertebrate/ fish’ approach can be expanded to allow further resolution of toxicity within one of these groups, for example by testing both salmonid and cyprinid fish species, or a range of invertebrates with different feeding strategies or life histories.

Given the objectives of this project, the limited range of whole organism assays previously recommended for use in the assessment of environmental samples (Section 2.1), and the need to apply a practicable suite of tests for use in subsequent monitoring programmes of waters receiving refinery effluents, the optimal approach is probably not to stray too far from the traditional approach to the selection of test species. However, the addition of prokaryotic organisms to the test battery may be beneficial in providing a further tier of biological organisation to the environmental assessment process, while an expansion of the range of invertebrate species applied (particularly for marine assessments) may provide additional resolution in screening and prioritising sites, if one test proves under or over sensitive (and therefore prevents differentiation between samples/ sites).

With respect to whole organism assays utilising bacteria, the *Aliivibrio fischeri* toxicity test (*Aliivibrio fischeri* is a fluorescent marine bacterium) is one of the most widely applied toxicity assays in both academic and regulatory settings, although perhaps with an emphasis on the testing of environmental samples (e.g. effluents, waters). The assay is very well-validated, is performed according to standardised methodology, and has been the subject of a number of inter-laboratory ring tests. It has also been widely applied in assessment of oil and gas contamination, with a high degree of sensitivity at the screening level (Concawe 2012a). The *Aliivibrio fischeri* toxicity test system requires only relatively small sample volumes, can generate

results within 15-30 minutes, and allows a number of samples to be tested concurrently.

Multi-species microbial toxicity tests have also been shown to be applicable to the testing of environmental samples (waters, effluents, sewage sludge, soil leachates), and, while a much more recent development than the *Aliivibrio fischeri* toxicity test, is also well validated and results have been shown to be reproducible across laboratories in a ring test. Both multi-species microbial toxicity tests described in this report are 96-cell plate systems which makes them highly amendable to small sample volumes and high sample throughput, and both assays provide results within 24 hours. The assay is commercially available in the form of a kit, and can be performed by any competent ecotoxicology laboratory (possessing a plate-reading spectrophotometer). The primary advantage of a multi-species microbial toxicity test over the *Aliivibrio fischeri* toxicity test is that both include an array of different prokaryotic organisms within the same assay (nine bacteria (freshwater and marine), a yeast and a fungae in the case of MARA assay; nine species of marine bacteria for the LumiMARA assay). This array of species has been applied to effluent samples to 'fingerprint' toxicity (i.e. the relative responses of the different species relate to different types of substance), which can also be useful in determining the types of substance present in a mixed sample. There are also obvious benefits in being able to test a wide range of species, with differing sensitivities, within a single assay.

Three algal assays were considered as part of this project; the traditional 72 hour inhibition of algal growth test, a further version of this assay which adds the measurement of photosynthetic activity to the standard cell number increase endpoints, and the algal Pulse-Amplitude Modulation (PAM) test, which measures the fluorescence of algae when light is delivered in a series of pulses and provides a measurement of the overall photosynthetic health of the algae.

The traditional freshwater inhibition of algal test is obviously extremely well-validated, and is applied is almost all chemical assessment regimes (e.g. for industrial chemicals, pharmaceuticals, plant protection products, biocides, etc.) according to an internationally accepted and long-standing guideline (OECD 201). This test has been demonstrated to be sensitive to a wide-range of different types of test substance, generates growth endpoints that can be applied in both short-term and long-term hazard assessments, and can easily be adapted for use with different species of 'algae' (e.g. cyanobacteria can be employed if anti-microbial effects are suspected). In addition, the assay can successfully be miniaturised and undertaken in 96-well plates, where fluorescence is measured as a surrogate for cell density, and has a long pedigree of being applied in the assessment of environmental samples, at least in the assessment of effluents. The OECD 201 test can also be performed using marine algae, however, marine microalgal tests are more often undertaken according to an ISO guideline (ISO 10253), which is similarly well-validated, ring-tested and sensitive to a wide range of different types of substance.

The combined algal growth and photosynthetic activity assay derives measurements after only 24 hours (compared to 72 hours for the standard algal growth tests), which might provide some advantage if the objectives are to quickly prioritise the phytotoxicity of different samples, and this test employs also employs a miniaturised (96-well plate) design. However, this assay is much more rarely applied than traditional algal growth tests (i.e. primarily only used in research activity) and is therefore less well validated. In addition, 24 (and 48) hour growth measurements can quickly and easily also be generated using standard (miniaturised) algal growth tests, if a fluorescence plate reader is applied to assessing algal growth. Perhaps because measurements are made after only 24 hours exposure, this assay tends to be less sensitive than the standard algal growth tests.

Similarly, the algal PAM assay is less well used, validated, or standardised than the traditional algal tests, but has the advantage of being able to generate results within 6 hours of the start of exposures (although it also appears to be less sensitive to most substances than the standard algal growth tests, with the possible exception of herbicides). The PAM test is a commercially available complete test system, comprising proprietary equipment and software in a manner similar to the *Aliivibrio fischeri* toxicity test system (although only one sample can be assessed at a time), and while the *Aliivibrio fischeri* toxicity test has been developed into a standardised guideline, which can be applied independently of the *Aliivibrio fischeri* toxicity test system, this appears to not be the case for the PAM system. While this assay clearly has its applications (e.g. distinguishing between differently pigmented populations of algae in samples; in-situ assessment of photosynthetic activity), it is perhaps most useful for ecotoxicity testing in allowing toxicity assessments to be made on plant species for which standardised toxicity testing methods do not exist, since any type of plant material can be employed in the PAM system.

A very wide range of whole organism ecotoxicity tests with invertebrates are available, many of them extremely well validated and ring-tested, and shown to be sensitive to different groups of substances. As outlined above, undertaking an array of different invertebrate tests may be beneficial in the hazard and risk assessment of individual substances since this maximises taxonomic diversity and may allow the application of probabilistic PNEC derivation approaches. However, when the objectives are to screen and prioritise receiving water sites, it is probably unrealistic, in terms of time and other resources, to employ many assays in the assessment of each sample.

As for the traditional algal toxicity test, the acute immobilisation test with *Daphnia*, is also applied across many chemical regulatory regimes, is undertaken according to well-validated and ring-tested standardised guidelines, and has been applied successfully in the assessment of freshwater environmental samples. This test can be miniaturised to a sufficient degree to be used with passive sampler extracts, and a ‘toxkit’ version, utilising ephippia, is available which assists in avoiding the need for resource intensive laboratory cultures and in enabling high throughput procedures.

The *Daphnia Magna* metabolic activity test allows the assessment of ‘toxicity’, as indicated by decreased substrate metabolisation measured indirectly using a fluorescent marker. While this test does not appear to be widely used in the assessment of freshwater environmental samples, it has undergone inter-laboratory ring testing and seems to therefore present some promise in this regard. Results have been shown to correlate with 48-hour *Daphnia* immobilisation, although Hamers et al. (2016) reported some difficulties in interpreting the assay response with respect to actual *Daphnia* toxicity. The primary advantage over the traditional *Daphnia* acute test is the ability of the metabolic activity test to derive results after only 75 minutes (60 minutes actual exposure), which can be advantageous if results are required quickly.

‘Toxkit’ type tests are also available for other freshwater species, such as the beavertail fairy shrimp, *Thamnocephalus platyurus* (Thamnotoxkit F™; ISO 1430), and the freshwater rotifer, *Brachionus calyciflorus* (Rotoxkit F™; ISO 19827). Individual factsheets were not prepared for these assays as they are well established methods which have been specifically developed for effluent testing (Daniel et al. 2004). The sensitivity of these assays to most chemicals is reported to be similar to the *Daphnia* acute test, and they can be undertaken over a similar timeframe (24 to 48 hours, although a 1 hour ‘rapidtox’ version of the *T. platyurus* test is also available). The ‘toxkit’ design of such assays allows a relatively high throughput of samples, and makes them amenable to testing with passive sampler extracts. Both assays have been ring-tested and standardised guidelines exist. However, while such ‘toxkit’ approaches may offer advantages over the standard *Daphnia* test in some circumstances (e.g. on-site assessments), most commercial laboratories will be highly

familiar with the conduct of the *Daphnia* acute test, and will be able to undertake this test more readily (and cheaply) than tests with fairy shrimps or rotifers.

The *C. elegans* test is unusual compared to the other invertebrate assays considered in the present study, since *C. elegans* is not an aquatic species. However, its biology and genome are well characterised, the test methods are well validated, and its potential utility as a predictive tool for Developmental and Reproductive Toxicity (DART) in humans is well documented. There test has also been shown to be adaptable to the assessment of environmental samples (including aqueous media), and therefore there appears to be no particular barrier to its application in aquatic environmental assessments. However, based on the available information in the study, this bioassay appears to be relatively insensitive (at least to PAHs) compared to more traditional invertebrate assays (e.g. *Daphnia* acute test), despite the inclusion of chronic endpoints in the test design.

For the assessment of marine waters, a greater number of standardised invertebrate tests are available, perhaps reflecting the more recent development of marine ecotoxicity assays with a primary focus on offshore activities and environmental assessments. The most widely applied acute/ short-term pelagic invertebrate tests comprise tests with marine copepods (using species such as *Acartia tonsa* and *Tisbe battagliai*), embryo-larval tests with bivalve molluscs and sea urchins, and marine rotifers.

Well validated, standardised guidelines are available for all of these tests, and all have been widely applied in the assessment of environmental samples. The embryo-larval tests perhaps have the edge in terms of miniaturisation and high sample throughput, and these early-life stage assays are generally more sensitive than the copepod/ rotifer tests, although this may be less of an issue when testing passive sampler extracts than with 'neat' environmental samples.

Whole fish testing is generally not recommended for water quality monitoring for ethical reasons. Alternatives to using whole fish tests are available, and include the FET test (or miniaturised design, QFET) based on OECD 236, or the measurement of cytotoxicity in fish cell lines. There are arguments that neither is necessary for a screening or prioritisation exercise. For hydrocarbon pollution, there is a body of evidence that suggests that lower organisms are more sensitive than fish (Concawe 2012a). However, a reason for including a measure of fish toxicity is that many of the *in vitro* assays are based on pathways that are most relevant to vertebrates, and therefore a measure of vertebrate effects might be beneficial for comparison.

A miniaturised FET with *Danio rerio* requires a relatively high amount of test solution compared with other whole organism assays, and still uses 'whole fish', even if they are not protected life stages. The advantage is that the assay is well developed as a screening tool and there is considerable scope for including novel biomarker endpoints if it were considered practical (e.g. it is possible to combine the comet assay and the FET assay) or the Zebrafish Toxarray. The Zebrafish Toxarray provides an indication for changes at the genomic level in response to passive sampler extracts. The results from such studies should be treated with caution as the methods have received much less attention with respect to their repeatability and show wider validation compared with other *in vitro tests* discussed in this study. The Zebrafish model is, however, an increasingly important tool in investigating the mechanistic effects of toxic PAHs (Goodale et al. 2012; Timme-Laragy et al. 2009). Genomic data obtained using the Zebrafish Toxarray can be used to cover a broader spectrum of potential mechanistic effects compared with a limited number of bioassays as well as providing supporting data for results observed in other effect based tools.

There have been considerable improvements in the use of cytotoxicity in fish cell lines (gill cell-line RT Gill-W1) as a predictive model for acute toxicity in fish and the assay is now being validated by ISO (Tanneberger et al. 2013). Compared with the FET test, the *in vitro* cell line assay has the advantage of being relatively rapid (24 hours), and requires much smaller volumes of test sample. The assay has some drawbacks for mixtures assessment because it is not sensitive to all substances (including some PAHs) because the cells lack some receptors and metabolic capabilities.

Table 11 summarises the attributes of the whole organism EBTs considered in this study, with respect to water quality monitoring.

Table 11 Whole organism bioassays – summary of attributes for water quality monitoring

Assay	Media and Taxa	Validation maturity [1]	Regulatory use	Use with environmental samples	Test duration	Supporting Information
Aliivibrio fischeri toxicity test (ISO 11348)	Marine bacteria (can be applied to freshwater samples)	3	Applied under several regulatory regimes	Routinely	Up to 30 mins	EC50; data available for range of substance types. Trigger values have been proposed.
Multi-species microbial toxicity test	Freshwater and marine bacteria and fungi	3	Limited regulatory application	Frequently	1 Day	Fingerprint of toxicity based on 11 microbes. Microbial toxic concentration (MTC) for test equivalent to IC50 for single organism.
Algal growth inhibition	Marine or freshwater algae	3	Applied under several regulatory regimes	Routinely	72 hours	EC50, NOEC, LOEC acute and chronic endpoints. Greater sensitivity compared with algal PAM for many substances (Di Paolo et al. 2016).
Algae PAM	Freshwater algae	2	Limited regulatory application	Occasional	4.5 hours	Low sensitivity compared with algal growth due to shorter test duration (Di Paolo et al. 2016). Endpoint can be difficult to interpret.
Daphnia Magna acute immobilisation	Freshwater crustacean	3	Applied under several regulatory regimes	Routinely	48 hours	EC50, NOEC, LOEC acute endpoints.
Daphnia Magna metabolic activity test	Freshwater crustacean	2	None	Rarely	75 minutes	Biomarker endpoint
Marine copepod acute	Marine crustacean	3	Regulatory use for marine environments	Routinely	48 hours	EC50, NOEC, LOEC acute endpoints
Thamnocephalus platyurus toxicity test (ISO 1430)	Freshwater crustacean	3	Limited regulatory application	Frequently	24 hours	EC50, NOEC, LOEC acute endpoints
Brachionus calyciflorus toxicity test (ISO 19827)	Freshwater or marine rotifer	3	Limited regulatory application	Frequently	48 hours	EC50, NOEC, LOEC sub-acute endpoints
Echinoderm or Oyster Embryo larval development	Marine mollusc or echinoderm	3	Regulatory use for marine environments	Routinely	24-48 hours	EC50, NOEC, LOEC sub-acute endpoints

Assay	Media and Taxa	Validation maturity [1]	Regulatory use	Use with environmental samples	Test duration	Supporting Information
Continued Table 11						
C. elegans growth and reproduction	Soil dwelling nematode	3	Standard research tool	Occasional	96 hours	EC50, NOEC, LOEC sub-acute endpoints
Zebrafish QFET	Freshwater fish	3	Applied under several regulatory regimes	Routinely	Up to 120 hours	Acute toxicity, EC50
Rainbow trout cell line cytotoxicity	Freshwater fish	2 (undergoing validation by ISO)	Regulatory uptake may increase once validated	Occasional	72 hours	EC50 based on cell viability in three cytotoxicity assays
Zebrafish Toxarray	Freshwater fish	1	None	Occasional	120 hours	Supporting interpretive information for mechanism of action. Does not provide information on an adverse effect. Used in conjunction with Zebrafish FET.

[1] Validation maturity based on a scale of 1-3 as outlined in the Norman Network validation guidance document (Norman Network 2008)

4.2. ENDOCRINE DISRUPTION (ED)

Receptor mediated assays for detecting ED effects are available for a range of steroid hormone pathways and show some promise for use in water quality monitoring because they can be used to demonstrate the combined effect of substances acting by a similar hormonal mechanism of action, and can be cheaper to use and are potentially more sensitive than direct chemical analysis.

There is particular interest from the EC to develop EBTs to use as an alternative to direct chemical analysis for the monitoring of steroidal oestrogens, 17α-ethinyloestradiol (EE2), 17β-oestradiol (E2) and oestrone (E1), which are on the WFD Watch List. Many versions of receptor based assays for detecting estrogenic activity exist, and most are based on binding to the ERα receptor, although similar assays also exist for the ERβ receptor. The estrogenic effect is reported based on relative potency compared with 17β-oestradiol. These assays are also used to measure anti-estrogenic activity by co-exposing a substance or sample with E2 and using tamoxifen as the standard. Most receptor based assays, including the ER activation assays, are based on human cell lines, but some (e.g. the YES assay) utilise yeast cells. These assays are known to have certain limitations, being affected by cytotoxicity and the presence of competing substances in the sample (i.e. anti-oestrogens), and different assays often generate inconsistent results for the same sample.

In general, the human receptor assays are more sensitive to estrogenic effects than yeast based assays. A study has also recently been conducted to compare the performance of five receptor based assays in terms of their reliability and repeatability for detecting an estrogenic effect (Kunz et al. 2016). The ERα activation assay and YES assay both performed better than other assays in terms of accuracy and repeatability, and were both considered suitable for use in water quality monitoring, with the ERα assay being recommended as superior. The ERα activation assay and YES assay were also both recommended for monitoring of drinking water in an assessment by Schriks et al. (2015). Trigger values of between 0.2-0.5 ng/L oestrogen equivalents have been recommended for use in water quality monitoring (Scott et al. 2014; Hamers et al. 2016).

Estrogenic assays have been widely used in monitoring studies and measured concentrations of known estrogenic substances often fail to account for all the measured estrogenic response. Estrogenic and anti-estrogenic effects have been reported for individual PAHs and mixtures of PAHs, although the results are variable depending on the receptor (α or β) or whether they are measured in a human or yeast based cell line (Vraibie et al. 2010; 2011). Reports of supermaximal responses, i.e. fluorescence formation higher than observed for E2, (at high hydrocarbon concentrations) have been reported for some hydrocarbon mixtures in a yeast based screen, but not in human receptor screens (Vrabie et al. 2010). Similarly results with the YES assay do not always correspond with oestrogenicity measured *in vivo* (Huggett et al. 2002). This variability shows that while receptor based assays may have role in environmental monitoring, further investigation is generally required to confirm a specific estrogenic response.

A relatively low number of substances cause androgenic responses, as measured in receptor assays relative to the response of 5-alpha-dihydrotestosterone (DHT), a potent androgen. For environmental monitoring, anti-androgenic activity measured by co-exposing a substance or sample with DHT, and using flutamide (FLU) as an anti-androgenic standard, is expected to be more relevant (Brack et al. 2016) than androgenic responses alone. A broad range of substances (including PAHs) are known to be anti-androgenic in *in vitro* receptor assays (van der Oost et. al. 2017).

Three androgenic EBTs were considered as part of this review; two types of AR activation assays (using hamster cell line or human bone marrow cell line) and a YAS assay. All assays are based on small sample volumes and would be relevant for use with passive sampler extracts. These assays have many of the same issues as the oestrogen screens, such as being affected by antagonistic substances that may present in an environmental mixture.

The AR activation assay using hamster cell line has been widely used for screening single substances for androgenic activity and is standardised to an OECD guideline. It has also been used in environmental studies, including in the TIPTOP study (Hamers et al. 2016), but to a much lesser extent than the YAS assay. A trigger value of 25 µg/L Flutamide FLU-EQ has been proposed for environmental monitoring using the AR activation assay using hamster cell line for anti-androgenicity.

The AR activation assay using human bone marrow cell line has been applied in monitoring studies in the Netherlands and Australia. Van der Oost et al. (2017) proposes a trigger value of 25 µg/L FLU-EQ for anti-androgenic activity in the AR activation assay using human bone marrow cell line assay (van der Oost et al. 2017), however, this may be unrealistic given that Scott et al. (2014) could only detect anti-androgenic effects at greater than 60 µg/L FLU-EQ in Australian surface water samples. Nevertheless, in their assessment of anti-androgenic assays, Schriks et al. (2015) conclude that the sensitivity of this and other similar assays would be sufficient for detecting anti-androgenic effects in most environmental samples.

The YAS assay has been most widely applied for environmental monitoring of androgenic and anti-androgenic activity over many years, and, as with the human cell based assays, responses are more usually observed for anti-androgenic activity. The assay has a lower sensitivity compared with the human receptor assays but is potentially less vulnerable to cytotoxicity. A comparison of the YES assay and AR activation assay using human bone marrow cell line, for use in monitoring hospital wastewater found that while the assays varied in terms of sensitivity and robustness they were both comparable in terms of classifying samples (Gehrmann et al. 2016). Schriks et al. (2015), in their assessment of assays for use in drinking water, ranked

the AR activation assay using human bone marrow cell line higher than the YAS or AR activation assay using hamster cell line in terms of assay performance.

The TTR binding assay, while a relatively new development, has been adopted for use in water quality monitoring, and has been shown to respond to passive sampler extracts (Hamers et al. 2016; Vethaak et al. 2017). The assay is affected by a wide range of substances, and therefore it may be more useful as marker of potential adverse effects, rather than in the direct identification of substances that may act via the thyroid system.

The main reason for using receptor assays for endocrine disruption is to detect potent estrogenic and androgenic substances such as steroidal oestrogens that cause effects at very low chemical concentrations, and that can be difficult to quantify chemically in environmental samples. Other less potent substances may however also interact with these receptors and can be implicated in the response. There is some evidence for the estrogenic, anti-estrogenic activity and anti- androgenic activity of hydrocarbons or oils *in vitro* although potency is generally very low (more than six orders of magnitude for estrogenic activity) compared with reference substances (Vrabie et al. 2011). Activity in *in vitro* assays does not mean that PAHs are endocrine disrupters, and there is currently limited evidence for them fulfilling the EC definition of an ED, which requires proof that the mechanistic responses cause adverse effects *in vivo*. However, a positive response in an *in vitro* assay in environmental monitoring studies could lead to speculation that PAHs may be contributing to the effect. This could be an issue because it is rare that all the endocrine activity in an environmental sample can be attributed to measured concentrations of specific substances with known endocrine activity, and false positive responses are always a possibility. In the absence of other specific endocrine disrupting substances, PAHs are considered unlikely to exceed the reported trigger values for such assays.

Table 12 summarises the attributes of the EBTs measuring endocrine disruption considered in this study, with respect to water quality monitoring.

Table 12 Endocrine disrupter bioassays – summary of attributes for water quality monitoring

Assay	Validation maturity [1]	Use with environmental samples	Response to PAH	Suitable for use with passive sampler extracts	Supporting information
ER activation assay	3	Frequently	Variable data; evidence for estrogenic, anti-estrogenic and synergistic responses, but potency low.	Yes	Response affected by anti-estrogenic substances. Response affected by anti-estrogenic substances. Trigger values proposed. High precision and repeatability compared with other oestrogen screens including the YES assay.
YES	3	Frequently	Variable data; evidence for estrogenic, anti-estrogenic and synergistic responses, but potency low.	Yes	Response affected by anti-estrogenic substances. Generally, less sensitive compared with mammalian receptor assays. Supermaximal responses for some oils difficult to interpret.

Assay	Validation maturity [1]	Use with environmental samples	Response to PAH	Suitable for use with passive sampler extracts	Supporting information
Continued Table 12					
AR activation assay using hamster cell line	2	Occasional	Anti-androgenic activity	Yes	Anti-androgenic response most used for water quality monitoring, response may be affected by androgenic substances. Trigger value proposed.
AR activation assay using human bone marrow cell line	2	Frequently	Anti-androgenic activity	Yes	Anti-androgenic response most used for water quality monitoring, response may be affected by androgenic substances. Trigger value proposed. High sensitivity compared with YAS assay but effected more by cytoxicity.
YAS	3	Frequently	Anti-androgenic activity	Yes	Anti-androgenic response most used for water quality monitoring, response may be affected by androgenic substances. Low sensitivity compared with AR activation assay using human bone marrow cell line, but less affected by cytotoxic effects.
TTR Binding	1	Rarely	Non-specific interaction	Yes	Lacks validation. Affected by a range of substances difficult to attribute cause and effect.

[1] Validation maturity based on a scale of 1-3 as outlined in the Norman Network validation guidance document (Norman Network 2008)

4.3. GENOTOXICITY

Genotoxicity is relevant for assessment of PAHs because some congeners, such as B(a)P (following metabolic activation), are known carcinogens.

A positive response in the Ames test is a measure of the occurrence of fixed mutations that are heritable to the next cell generation. It demonstrates that an actual mutagenic effect has occurred and indicates the potential for carcinogenicity. The assay has been routinely used as an initial screen for mutagenicity in single substance evaluations, and therefore has regulatory credibility. Standardised protocols exist for water quality monitoring and the assay has been widely applied to environmental samples. A drawback of the assay is that it can be relatively time consuming to conduct, but miniaturised/ high-throughput test designs do exist and these have been used in the assessment of passive sampler extracts (Rastall et al. 2004; Gallampois et al. 2013; Vincent-Hubert et al. 2016). The assay is potentially less sensitive than other EBTs investigating a similar effect (such as umuC), and the outputs are not easily quantifiable, resulting in a challenge in the development of trigger values (van der Oost et al. 2017a). Neale et al. (2017) found that only one of 22 substances tested (B(a)P) responded positively in the Ames test, even though mode of action analysis would have predicted genotoxic responses for a wider range of the substances tested. A lack of sensitivity of the assay may be one explanation for the observed lack of effects, however, the authors also note that for some of the substances tested there

were water solubility issues. To address the issues with water solubility for some substances, a so called Modified AMES test have been developed which can be used as a screening tool for carcinogenicity of highly viscous or semi-solid petroleum hydrocarbon streams (Concawe 2012b). However, the Modified AMES tests is more of relevance for assessing full petroleum hydrocarbon streams than it is for surface waters.

The micronucleus assay detects non-repairable chromosomal damage, and provides an indication of carcinogenicity. Biological relevance is generally improved if a clear dose-response can be demonstrated. The assay is a direct visible measure of micronuclei formation in cells and has been standardised for use both *in vivo* (biomarker response) and *in vitro* (in cell lines such as the Chinese hamster lung fibroblast cell line V79). The *in vitro* assay is more applicable for use with passive sampler extracts because of the low sample volumes required for direct exposure of the cells. The sensitivity of the assay may be variable if cell lines other than the Hamster V79 are used. The results from the micronucleus test can be reported based on the relative potency to known genotoxicity agents, which can aid interpretation, especially if chemical verification of substances of concern is carried out. The manual assessment of the assay does, however, limit the test in terms of the maximal throughput of samples (Brack et al. 2016).

The umuC assay is based on activation of the umuC gene as a measure of the SOS repair response, and is therefore an indirect measure of DNA damage. A miniaturised version of the test is available which is suitable for use with passive sampler extracts, and the assay is widely applied in the assessment of single substances and environmental samples. It is relatively sensitive to DNA damage, but has low specificity compared with the Ames test. Genotoxicity can be expressed based on relative potency and interpretation of effects have been made by deriving a trigger values based on genotoxicity units (GTU) (Leusch et al. 2014). Kittinger et al. (2015) report that it can be difficult to attribute cause and effect to samples responding positively in this assay because a relatively high number of substances can cause a response.

The hGADD45a activation assay and p53-pathway activation assay are both reporter based assays and are similar to the umuC assay in that provide indirect measured of the DNA repair response.

The GADD45a activation GFP reporter assay (based on a yeast cell line) was assessed for use with environmental samples by Daniel et al. (2004), with some limited success, but this assay is no longer readily available. The yeast based assay has been replaced by the hGADD45a activation GFP reporter assay or a luciferase assay (both based on human cell lines), both of which are routinely used for screening single substances for genotoxic effects. These assays are designed to deliver a high throughput of samples, and both provide quantitative and qualitative results based on expression of the hGADD45a gene which is part of the genotoxicity stress response. While the assays show some promise for use as screens for genotoxicity, there is limited data available to support their use or validation with environmental samples.

The p53-pathway activation assay is also a reporter gene assay, using luciferase activity to indicate the presence of p53-pathway activating compounds. The specificity and sensitivity are reported to be similar to the hGADD45a activation assay, however, false positive results have been reported for some substances (e.g. phenanthrene (van der Linden et al. 2014)). The assay is similar to the hGADD45a activation assay in that it requires low sample volumes, and can deliver a high throughput of samples, but its previous use with environmental samples is limited.

Table 13 summarises the attributes of the EBTs measuring genotoxicity considered in this study, with respect to water quality monitoring.

Table 13 Genotoxicity bioassays – summary of attributes for water quality monitoring

Assay	Validation maturity [1]	Use with environmental samples	Response to PAH	Suitable for use with passive sampler extracts	Supporting Information
Ames II	3	Routinely	Sensitive to B(a)P with metabolic activation	Yes	Detects fixed mutations that are heritable to the next cell generations. Endpoint based on number of revertants. Positive or negative for mutagenicity. More labour intensive than assays that can be measured via luminescence
UmuC	3	Routinely	Sensitive to B(a)P with metabolic activation	Yes	Indicates induction of DNA damage response. Sensitive, early warning but low chemical specificity. Trigger values for genotoxicity available. Assay can be normalised to genotoxic units (GTU) for comparison with other assays.
Micronucleus in a cell line	3	Frequently	Sensitive to B(a)P	Yes	Indicative for non-repairable chromosomal damage. Sensitivity of the endpoint in different cell lines may vary. Can be reported based on REP to a known genotoxin. Labour intensive due to manual assessment of slides limiting sample throughput.
hGADD45a activation assay	2	Occasional	Not reported	Yes	Indicates induction of DNA damage response. High throughput. Quantifiable response based on GFP or luciferase activity.
p53-pathway activation assay	2	Occasional (relatively new assay)	Sensitive to B(a)P with metabolic activation. Assay appeared to give a false positive result with phenanthrene.	Yes	Indicates induction of DNA damage response. Quantifiable response based on luciferase activity.

[1] Validation maturity based on a scale of 1-3 as outlined in the Norman Network validation guidance document (Norman Network 2008)

4.4. OXIDATIVE STRESS

Oxidative stress occurs when there is disturbance in the normal redox state of cells. It can result in the generation of reactive oxygen species which can cause toxic effects through the production of peroxides and free radicals that damage all components of the cell, including strand breaks in DNA (which can lead to genotoxicity).

The AREc32 is a luciferase based reporter assay which responds to induction of the Nrf2-mediated oxidative stress response pathway. This assay has been shown to be induced by the PAHs, B(a)P and benzo(b)fluoranthene (Neale et al. 2017). There is, however, considerable inter connection and overlap between the AhR pathway and Nrf2 pathway, so induction in the AREc32 assay by B(a)P is not unexpected (Escher et al. 2012). The assay has undergone some inter laboratory validation, it is suitable for use with passive sampler extracts, is routinely used for environmental samples and has had effect based trigger values derived for its response (Escher et al. 2013). A relatively high number of substances induce the antioxidant response, meaning it can be difficult to establish a substance-specific cause and effect (Escher et al. 2012; Neale et al. 2017).

The Nrf2-pathway activation assay is an alternative assay for detecting oxidative stress based on luciferase activity in response to four Electrophile Responsive Elements (EpREs), and is currently undergoing preliminary validation with single substances. The available data suggests that substances that respond in the AREc32 assay may not all be responsive in the Nrf2-pathway activation assay, although it is not clear why the assays would give a different response. For example, B(a)P is a potent inducer in the AREc32 assay (Escher et al. 2012) as might be expected based on available mechanistic understanding (Goodale et al. 2013) but had no effect in the Nrf2-pathway activation assay (van der Linden et al. 2014).

Table 14 summarises the attributes of the EBTs measuring oxidative stress considered in this study, with respect to water quality monitoring.

Table 14 Oxidative stress bioassays – summary of attributes for water quality monitoring

Assay	Validation maturity [1]	Use with environmental samples	Response to PAH	Suitable for use with passive sampler extracts	Supporting Information
Antioxidant Response Element (ARE)c32 assay	2	Frequently	B(a)P is a potent inducer.	Yes	Induction ratio for response reported. Trigger value proposed for drinking water assessment. Cytotoxicity may mask oxidative stress. Responds to a wide range of substances.
Nrf2-pathway activation assay	2	Occasional	Assay did not respond to B(a)P.	Yes	1.5-fold induction indicates positive response. Assay still under development.

[1] Validation maturity based on a scale of 1-3 as outlined in the Norman Network validation guidance document (Norman Network 2008)

4.5. METABOLISM

Assays indicative of AhR mediated induction of cytochrome P4501A (CYP 1A) were considered as part of this assessment because of their specificity for dioxins and dioxin-like substances, including PAHs. AhR receptor assays such as the AhR activation assay or assays based on the induction of 7-ethoxyresorufin-O-deethylase (EROD) are used to indicate the presence of PAHs, particularly four-five ring PAHs, and they have been widely used for environmental testing with activity being expressed relative to either 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) or B(a)P.

There are several AhR receptor assays based on different human cell lines which respond to dioxin responsive substances with receptor agonism quantified based on luciferase activity. The tests are generally optimised to detect more persistent dioxins and dioxin-like substances, and this requires a pre-treatment step using sulphuric acid to remove interfering substances. More readily degradable AhR receptor agonists are removed via this pre-treatment. The contribution of the more degradable components of the mixture to AhR receptor agonism can, however, be assessed by conducting the assay without the pre-treatment, and by applying a shorter test duration. The pre-treatment step is unlikely to be effective at removing all PAHs that would interact with the AhR receptor ((e.g. benzo(b)fluoranthene and fluoranthene) (van Hattum et al. 2004). In general, responses to the full range of PAHs likely to be present in refinery discharges (which will include both relatively persistent and degradable substances) would require tests both with and without the pre-treatment step.

A PAH-specific AhR receptor assay, based on a AhR activation assay but optimised specifically to respond to the AhR activity from more readily biodegradable PAHs, has been developed by using a cell line with lower metabolic activity and a short test duration (Pieterse et al. 2013). While optimised for PAHs, the assay will also respond to many other substances, including dioxins and pharmaceuticals (e.g., cyclophosphamide) (van der Oost et al. 2017) and many others (Hamers et al. (2016) highlighted that 2132 compounds with AhR-agonistic potencies were reported in the ToxCast database. As with other receptor assays they are also affected by receptor antagonists as well as agonists (and mixtures of PAHs could include both) which can affect interpretation. The DR LUC assays are widely used for environmental monitoring and effect based trigger values are being developed for these tools (van der Oost et al. 2017).

The use of 7-ethoxyresorufin-O-deethylase (EROD) activity as a surrogate for AhR mediated cytochrome CYP 1A induction is long established as a biomarker endpoint measured *in vivo* in fish. It is routinely used in monitoring activities related to the oil and gas industry, despite its well documented limitations which include interferences, and the fact that it only responds to certain PAHs (wca 2017). The assay has been adapted for use in *in vitro* cell lines including a rainbow trout liver cell line (RTL-W1) which is suitable for use with environmental samples including passive sampler extracts. The assay arguably has more relevance than the AhR receptor assays because it measures actual enzyme induction in a cell line from a relevant organism. The response is, however, less sensitive than the DR LUC assays, can produce variable results, and can be affected by substances that cause substrate inhibition (Eichbaum et al. 2014).

Table 15 summarises the attributes of the EBTs measuring metabolism considered in this study, with respect to water quality monitoring.

Table 15 Oxidative stress bioassays – summary of attributes for water quality monitoring

Assay	Validation maturity [1]	Use with environmental samples	Response to PAH	Suitable for use with passive sampler extracts	Supporting information
AhR activation assays	3	Frequently	Specific for AhR receptor agonists including PAHs. Some PAHs show antagonistic activity. Different sample preparation and test durations are required for detection of more readily degradable PAHs.	Yes	Cytotoxicity or colour of sample may affect results at high concentrations. AhR antagonists can inhibit activity. Trigger value proposed.
AhR activation assay (with more specificity for PAHs)	2	Occasional (relatively new assay)	Specific for AhR receptor agonists but optimised for use with more biodegradable PAHs.	Yes	Cytotoxicity or colour of sample may affect results at high concentrations. AhR antagonists can inhibit activity. Sensitive to other substances including dioxins and pharmaceuticals (e.g., cyclophosphamide). Trigger value proposed.
In vitro EROD induction / CYP 1A activity based on cell lines	2	Frequently	Specific for AhR receptor agonists including PAHs. Does not differentiate between Dioxin substances and PAHs.	Yes Result needs to be normalised for protein concentration which is a limiting factor for sample throughput.	Confounded by substances that can inhibit the response. Less sensitive than AhR reporter gene assays. Can produce variable results. More ecological relevance compared with AhR reporter gene assays as it demonstrates upregulation of a protein rather than just receptor mediation.

[1] Validation maturity based on a scale of 1-3 as outlined in the Norman Network validation guidance document (Norman Network 2008)

5. CONCLUSIONS

The assessment in this section are provided with the intent of choosing a suite of bioassays that could be combined with passive samplers to assess waters receiving refinery discharges.

The suite of bioassays are, however, restricted to those assessed in the present study, and are based on the information identified in this review, and the relative advantages and disadvantages of using each of the assays, as detailed in the previous sections. A pilot monitoring activity may be needed, where some assays are found not to provide useful information and can be dropped, or where, for some endpoints, another assay may need to be trialled as a substitute. In addition, it is recognised that new bioassays are continually being developed, and those currently at a relatively early stage of validation are in the process of being standardised and demonstrated to be reproducible.

The selection of bioassays for any environmental assessment will depend largely on the specific objectives of the study, and therefore it is always recommended that a detailed problem formulation assessment is conducted prior to undertaking any monitoring based activity. The bioassays will vary depending on the environmental compartment of concern, the substances of concern and the study objectives (e.g. site prioritisation, comparison with a reference site, or whether there is a threshold of response above which risks are inferred).

In general, while the results of the initial monitoring work are likely to be primarily derived for internal assessments, it is the opinion of the authors that the suite of bioassays applied should also be comprehensible to a broad regulatory audience (or at least those in the regulatory community with the required knowledge and expertise). Therefore, it seems sensible to include both assays that provide a general indication of ‘whole’ toxicity (generally whole organism tests) and a targeted array of more specialist, supporting *in vitro* assays investigating specific endpoints or substance specific properties (e.g. oestrogenicity, genotoxicity, etc.).

In terms of whole organism tests, the suite of tests follows the traditional (limited) hazard assessment approach to taxonomic coverage and include at least a primary producer (i.e. algae), a primary consumer (i.e. an invertebrate), and a secondary consumer (e.g. fish), with the addition of an assay using bacteria. A similar approach is taken by van der Oost et al. (2017a) when proposing whole organism methods for surface water monitoring.

At the primary producer level, the miniaturised (96 well plate using fluorescence as a surrogate for cell density) version of the standardised algal tests (OECD 201 or ISO 10253), is likely to prove the optimal algal assay for use with passive sampler extracts. While the alternatives considered in this study (combined algal growth and photosynthesis, and algal PAM) do have their advantages in specific circumstances, their lack of validation and pedigree for use in environmental assessments, and reduced sensitivity to substances relative to the standard algal test, mean that they are unlikely to offer any advantage over the standard algal test.

Similarly, the standard *Daphnia* immobility test is included in the suite of assays for use in the assessment of extracts from passive samplers deployed in freshwaters in preference to the *Daphnia Magna* metabolic activity test, the ‘toxkit’-type methods (using either *Thamnocephalus platyurus* (ISO 1430) or *Brachionus calyciflorus* (ISO 19827)) or the *C. elegan*s reproduction test. Again, degree of standardisation and validation, and pedigree for use in environmental assessments makes it difficult to discount this test in favour of the others considered in this study. In addition, the

Daphnia Magna metabolic activity test appears to suffer from some interpretational difficulties, and the *C. elegans* test is somewhat environmentally irrelevant for the assessment of water samples, and tends to be less sensitive than the *Daphnia* test.

For marine waters, it is more difficult to distinguish between the tests considered in the present study on the basis of standardisation, validation maturity or environmental assessment pedigree. On balance, the bivalve embryo-larval test is included in the suite of assays, primarily because it is highly amenable to the high throughput of low volume samples, and because of its greater sensitivity to many contaminants compared to the other tests, which may be helpful in providing resolution between the effects inferred by extracts from different sites/ passive sampler deployments.

As highlighted in the earlier sections, whole fish (including embryos) are not included in the suite of assays for the monitoring of waters receiving refinery effluents. Primarily, this is based on ethical considerations, the fact that fish tend to be less sensitive to hydrocarbon pollution than lower organisms, and the relatively high sample volumes required for such tests. Nevertheless, an *in vitro* fish cell line cytotoxicity assay is included, at least in the initial stages of the monitoring programme, in order to support the other *in vitro* assays (see below), since many utilise vertebrate cells and therefore a measure of fish cell toxicity may assist in the interpretation of the results of such assays.

While the *Allivibrio fischeri* toxicity test assay is very widely applied in environmental assessments, in particular for providing a very rapid screening of numerous samples, a multi-species microbial toxicity test assay, although taking longer to produce results than the *Allivibrio fischeri* toxicity test, shows much promise in terms of validation and use for environmental assessments. A multi-species microbial toxicity test also has the advantages of exposing a large number of different species concurrently, and allowing the ‘fingerprinting’ of toxicity to specific groups of substances. This may be particularly useful in ascertaining if effects are likely to be due to hydrocarbons or other types of substances in a sample. For these reasons, both the *Allivibrio fischeri* toxicity test and a multi-species microbial toxicity test are included in the suite of assays, at least for the initial stages of the monitoring programme.

Of the estrogenic EBTs reviewed here, the ERα activation assay and YES assays are considered to be most mature in terms of validation, and both have been widely applied to environmental monitoring. The human cell based assays are, however, generally more sensitive than yeast based assays, and therefore ERα activation assay is included in the suite of assays.

Similarly, for androgenic and anti-androgenic activity the assays based on human receptors appear to have the advantage of greater sensitivity compared with the YAS assay and appear to be increasingly used for water quality monitoring. It is difficult to differentiate between the AR activation assay using hamster cell line and the AR activation assay using human bone marrow cell line in terms of validation maturity, but Schriks et al. (2015) considered the human bone marrow cell line one to be slightly superior to the hamster cell line one for the assessment of drinking water samples based on overall performance. Moreover, the AR activation assay using human bone marrow cell line is more frequently applied to the assessment of environmental samples than the AR activation assay using hamster cell line. The AR activation assay using human bone marrow cell line is therefore considered as the optimal selection for the assessment of (anti-)androgenic activity.

The TTR binding assay is relatively novel and the available information on its use for monitoring purposes suggests that, while it is responsive to environmental samples, the ability of the test to predict specific effects on the thyroid pathway or in identifying substances that might target this pathway is not yet fully developed. There is currently considerable interest in improving the regulatory tests relating to ED effects via the thyroid pathway, and it is anticipated that there will be relatively rapid future development and validation of this, or similar, thyroid pathway screening assays. The TTR binding assay is not, however, considered to be sufficiently developed for inclusion in routine monitoring at this time.

The five genotoxicity assays reviewed in this study provide different measures for genotoxicity and, depending on the purpose of the monitoring, more than one assay could reasonably be applied in order to provide more information on the type of DNA mutation that is occurring (Brack et al. 2017). The Ames test provides a definitive measure for an actual genotoxic effect and is more specific for genotoxicity (less false positives) compared with the assays based on DNA repair. The value of the genotoxicity assays based on the DNA repair response should not, however, be overlooked since these assays are more readily quantifiable and more sensitive than the Ames test. Of the three assays that are based on a DNA repair response, the umuC is the most mature in terms of validation and has considerable pedigree for use with environmental samples. Therefore a combination of the Ames and umuC assays to assess potential genotoxic effects in waters receiving refinery discharges are included in the suite of assays. While the micronucleus assay is well validated and often used in the assessment of environmental samples, it is particularly labour intensive compared to the other assays (owing to the need to visually assess samples for chromosomal damage) and does not lend itself to the high throughput of samples.

Of the two oxidative stress assays evaluated, the AREc32 is more mature in terms of validation, is more commonly applied to environmental samples and has been shown to respond to certain PAHs. Given that the Nrf2 pathway and AhR pathway are known to be linked, there would be value in including this assay in a suite of assays for monitoring hydrocarbon pollution to aid interpretation of the results seen in other tests (e.g. genotoxicity and metabolism assays), however, the Nrf2-pathway activation assay is not yet considered to be sufficiently validated or proven with environmental samples to be a reliable inclusion in the suite assays.

Assays indicative of AhR mediated toxicity are included in a battery of *in vitro* tests for assessing hydrocarbon pollution because they tend to be responsive to PAHs. The AhR activation assay is more sensitive and less vulnerable to confounding factors compared with the direct measurement of EROD induction in a cell line, and has been routinely applied in environmental monitoring. In addition, a supporting assay for detecting the contribution to AhR receptor effects of the more degradable PAH fraction, a PAH specific AhR activation assay, can be applied in tandem with the AhR activation assay. Therefore, both assays are included in the suite of assays.

The full suite of assays that could be used for assessment of sites receiving refinery effluents, along with assay-specific trigger values (identified in the literature review) are summarised in **Table 16**. It should be noted that the bioassays assessed in the present study are by no means assessed in terms of availability with regard to being “available techniques” as defined by the Industrial Emissions Directive (IED; 2010/75/EU) article 3(10). Moreover, the suite of bioassays are restricted to those assessed in the present study, and are based on the information identified in this review, and the relative advantages and disadvantages of using each of the assays. In addition, it is recognised that new bioassays are continually being developed, and those currently at a relatively early stage of validation are in the process of being standardised and demonstrated to be reproducible.

Table 16 Suite of assays that could be used for assessment of sites receiving refinery effluents (applied to passive sampler extracts)

Assay	Assay type	Trigger value [1 2 3]	Reference
Toxicity to *Allivibrio fischeri* (ISO 11348); e.g. MicroTox™	*In vivo*	No trigger value found but a similar approach can be used as for other *in vivo* assays.	-
Multi-species microbial toxicity test; e.g. MARA/ LumiMara	*In vivo*	No trigger value found but a similar approach can be used as for other *in vivo* assays.	-
Miniaturised Daphnia acute test (OECD 202)	*In vivo* (freshwater)	0.05 toxic units [3]	Van der Oost et al. 2017a
Microplate Algal growth tests (OECD 201 or ISO 10253)	*In vivo*	0.05 toxic units	Van der Oost et al. 2017a
Bivalve embryo development test (ICES No.54)	*In vivo* (marine)	No trigger value found but a similar approach can be used as for other in vivo assays.	-
Cytotoxicity in a Rainbow Trout cell line (RT Gill-W1)	*In vitro* / adverse effect	0.05 toxic units	Van der Oost et al. 2017a
Ames test	Genotoxicity	0.005 genotoxic units. Positive or negative for mutagenicity.	Van der Oost et al. 2017a
umuC	Genotoxicity	0.005 genotoxic units.	Van der Oost et al. 2017a
ER activation assay; e.g. ER CALUX®	Endocrine Disruption	0.2 – 0.5 ng/L Oestrogen equivalents (EEQ).	Scott et al. 2014; Hamers et al. 2016
AR activation assay; e.g. AR CALUX®	Endocrine Disruption	25 mg Flutamide (FLU) EQ/L	Van der Oost et al. 2017a
AhR activation assay	Metabolism	50 pg 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD)-EQ/L 16.2 pg TCDD-EQ/L	Van der Oost et al. 2017a Hamers et al. 2016
AhR activation assay (with more specificity for PAHs)	Metabolism	150 ng Benzo(a)pyrene B(a)P-EQ/L	Van der Oost et al. 2017a
AREc32	Oxidative stress	EBTV-Effective concentration induction ratio (ECIR[4]) 1.5 = 6 REF	Escher et al. 2012

[1] Trigger values derived by Van der Oost et al. (2017a) are for use with passive samplers and conservatively assume that only 50% of the mixture is recovered as part of the passive sampler extract. The trigger values for whole organism assays are based on extrapolation of acute effects to chronic toxicity based on applying a factor of 10.
[2] Toxic unit = 1/Relative Enrichment Factor (REF) where the REF is calculated by dividing the sample concentration from solid phase extraction by the dilution in the assay (Leusch et al. 2014)
[3] Effect based trigger values (EBTV) are not intended to be used in isolation and the results should be considered as part of an overall assessment including results for other EBTs for example as proposed by the SIMONI approach (Van der Oost et al. 2017b).
[4] EC-IR is the concentration causing an induction ratio (IR) of 1.5 (ECIR 1.5)

6. REFERENCES

1. Alvarez DA, Cranor WL, Perkins SD, Clark RC, Smith SB. 2008. Chemical and Toxicologic Assessment of Organic Contaminants in Surface Water Using Passive Samplers. Journal of Environmental Quality 37:1024-1033

2. Amaeze NH. Schnell S, Sozeri O, Otitoloju AA, Egonmwan RI, Arlt VM, Bury NR. 2014. Cytotoxic and genotoxic responses of the RTgill-W1 fish cells in combination with the yeast oestrogen screen to determine the sediment quality of Lagos lagoon, Nigeria. Mutagenesis 30:117–127

3. Balaam JL, Chan-Man Y, Roberts PH, Thomas KV. 2009. Identification of nonregulated pollutants in North Sea-produced water discharges. Environmental Toxicology and Chemistry 38(6):1159-1167

4. Behrens A, Schirmer K, Bols NC, Segner H. 2001. Polycyclic aromatic hydrocarbons as inducers of cytochrome P4501A enzyme activity in the rainbow trout liver cell line, RTL-W1, and in primary cultures of rainbow trout hepatocytes. Environmental Toxicology and Chemistry 20(3):632-643

5. Bekki K, Takigami H, Suzuki G, Tang N, Hayakawa K. 2009. Evaluation of toxic activities of polycyclic aromatic hydrocarbon derivatives using in vitro bioassays. Journal of Health Science 55(4):601-610

6. Beresford N, Baynes A, Kanda R, Mills MR, Arias-Salazar K, Collins TJ, Jobling S. 2016. Use of a battery of chemical and ecotoxicological methods for the assessment of the efficacy of wastewater treatment processes to remove estrogenic potency. Journal of Visualized Experiments 115:e54243

7. Bi H, Rissik D, Macova M, Hearn L, Mueller JF, Escher B. 2011. Recovery of a freshwater wetland from chemical contamination after an oil spill. Journal of Environmental Monitoring 13:713-720

8. Bianchi E, Goldoni A, Trintinaglia L, Lessing G, Silva CEM, Nascimento CA, Ziulkoski AL, Spilki FR, Silva LB. Evaluation of genotoxicity and cytotoxicity of water samples from the Sinos River Basin, southern Brazil. Brazilian Journal of Biology 72(2 suppl.):S68-S74

9. Billiard SM, Bols NC, Hodson PV. 2004. In vitro and in vivo comparisons of fish-specific CYP1A induction relative potency factors for selected polycyclic aromatic hydrocarbons. Ecotoxicology and Environmental Safety 59:292-299

10. Bols NC, Schirmer K, Joyce EM, Dixon DG, Greenberg BM, Whyte JJ. 1999. Ability of polycyclic aromatic hydrocarbons to induce 7-ethoxyresorufin-o-deethylase activity in a trout liver cell line. Ecotoxicology and Environmental Safety 44:118-128

11. Brack W, Ait-Aissa S, Burgess RM, Busch W, Creusot N, Di Paolo C, Escher BI, Hewitt LM, Hilscherova K, Hollender J, Hollert H, Jonker W, Kool J, Lamoree M, Muschket M, Neumann S, Rostkowski P, Ruttkies C, Schollee J, Schymanski EL, Schulze T, Seiler T-B., Tindall AJ, Umbuzeiro GDA, Vrana B, Krauss M. 2016. Effect-directed analysis supporting monitoring of aquatic environments—An in-depth overview. Science of the Total Environment 544:1073–1118

12. Brinkmann M, Blenkle H, Salowsky H, Bluhm K, Schiwy S, Tiehm A, Hollert H. 2014. Genotoxicity of heterocyclic PAHs in the micronucleus assay with the fish liver cell line RTL-W1. PLoS ONE 9(1): e85692

13. CIS. 2014. Common Implementation Strategy (CIS) Technical report on aquatic effect-based monitoring tools. Technical Report - 2014 – 077

14. Clavijo A, Kronberg MF, Rossen A, Moya A, Calvo D, Salatino SE, Pagano EA, Morábito JA, Muarriz ER. 2016. The nematode Caenorhabditis elegans as an integrated toxicological tool to assess water quality and pollution. Science of the Total Environment 569-570:252-261

15. Concawe (2012a) Assessment of refinery effluents and receiving waters using biologically-based effect methods. Report No 1/12. Brussels: Concawe

16. Concawe (2012b) Use of the modified Ames test as an indicator of the carcinogenicity of residual aromatic extracts. Report 12/12. Brussels: Concawe

17. Daniel M, Sharpe A, Driver J, Knight AW, Keenan PO, Walmsley RM, Robinson A, Zhang T, Rawson D. 2004. Results of a technology demonstration project to compare rapid aquatic toxicity screening tests in the analysis of industrial effluents. Journal of Environmental Monitoring 6:855-865

18. Dayeh VR, Schirmer K, Bols NC. 2002. Applying whole-water samples directly to fish cell cultures in order to evaluate the toxicity of industrial effluent. Water Research 36:3727-3738

19. Di Paolo C, Ottermanns R, Keiter S, Ait-Aissa S, Bluhm K, Brack W, Breitholz M, Buchinger S, Carere M, Chalon C, Cousin X, Dulio V, Escher BI, Hamers T, Hilscherová K, Jarque S, Jonas A, Maillot-Marechal E, Marneffe Y, Nguyen MT Pandard P, Schifferli A, Schulze T, Seidensticker S, Seiler TB, Tang J, van der Oost R, Vermierssen E, Zounková R, Zwart N, Hollert H. 2016. Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts – Towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring. Water Research 104:473-484

20. Eichbaum K, Brinkmann M, Buchinger S, Reifferscheid G, Hecker M, Giesy JP, Engwall M, van Bavel B, Hollert H. 2014. In vitro bioassays for detecting dioxin-like activity — Application potentials and limits of detection, a review. Science of the Total Environment 15;487:37-48

21. Emelogu ES, Pollar P, Robinson DC, Smedes F, Webster L, Oliver IW, McKinzie C, Seiler TB, Hollert H, Moffat CF. 2012. Investigating the significance of dissolved organic contaminants in aquatic environments: Coupling passive sampling with in vitro bioassays. Chemosphere 90(2):210-219

22. Escher BI, Bramaz N, Quayle P, Rutishauser S, Vermeirssen ELM. 2008. Monitoring of the ecotoxicological hazard potential by polar organic micropollutants in sewage treatment plants and surface waters using a mode-of-action based test battery. Journal of Environmental Monitoring 10:622-631

23. Escher BI, Dutt M, Maylin E, Tang JYM, Toze S, Wolf CR, Lang M. 2012. Water quality assessment using the AREc32 reporter gene assay indicative of the oxidative stress response pathway. Journal of Environmental Monitoring 14:2877-2885

24. Escher BI, van Daele C, Dutt M, Tang JYM, Altenburger R. 2013. Most oxidative stress response in water samples comes from unknown chemicals: The need for effect-based water quality trigger values. Environmental Science & Technology 47:7002-7011

25. Fai PB, Grant A. 2010. An assessment of the potential of the microbial assay for risk assessment (MARA) for ecotoxicological testing. Ecotoxicology 19:1626-1633

26. Fang YX, Ying GG, Zhao JL, Chen F, Liu S, Zhang LJ, Yang B. 2012. Assessment of hormonal activities and genotoxicity of industrial effluents using in vitro bioassays combined with chemical analysis. Environmental Toxicology and Chemistry 31(6):1273-1282

27. Gallampois CMJ, Schymanski EL, Bataineh M, Buchinger S, Krauss M, Reifferscheid G, Brack W. 2013. Integrated biological–chemical approach for the isolation and selection of polyaromatic mutagens in surface waters. Analytical and Bioanalytical Chemistry 405:9101-9112

28. Garcia ACFS, Marcon AE, Ferreira DM, Santos EAB, Amaral VS, Medeiros SRB. 2011. Micronucleus study of the quality and mutagenicity of surface water from a semi-arid region. Journal of Environmental Monitoring 13:3329-3335

29. Gehrmann L, Bielak H, Behr M, Itzel F, Lyko S, Simon A, Kunze G, Dopp E, Wagner M, Tuerk J. 2016. (Anti-)estrogenic and (anti-)androgenic effects in wastewater during advanced treatment: comparison of three in vitro bioassays. Environmental Science and Pollution Research. 1-11.

30. Goodale BC, Jane K. La Du JK, Bisson WH, Janszen DB, Waters KM, Tanguay RL. 2012. AHR2 Mutant Reveals Functional Diversity of Aryl Hydrocarbon Receptors in Zebrafish. PLoS ONE 7(1): e29346.

31. Grover DP, Balaam J, Pacitto S, Readman JW, White S, Zhou JL. 2011. Endocrine disrupting activities in sewage effluent and river water determined by chemical analysis and in vitro assay in the context of granular activated carbon upgrade. Chemosphere 84:1512-120

32. Hamers T, Legradi J, Zwart N, Smedes F, de Weert J, van den Brandhof EJ, van de Meent D, de Zwart D. 2016. Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP): an effect-based strategy for cost-effective chemical water quality assessment. Final report of the LRI-ECO23 project November 2016.

33. Han Y, Li N, Oda Y, Ma M, Rao K, Wang Z, Jin W, Hong G, Li Z, Luo Y. 2016. Evaluation of genotoxic effects of surface waters using a battery of bioassays indicating different mode of action. Ecotoxicology and Environmental Safety 133:448-456

34. Hara RV, Marin-Morales MA. 2017. In vitro and in vivo investigation of the genotoxic potential of waters from rivers under the influence of a petroleum refinery (São Paulo State – Brazil). Chemosphere 174:321-330

35. Harman C, Farmen E, Tollefsen KE. 2010. Monitoring North Sea oil production discharges using passive sampling devices coupled with in vitro bioassay techniques. Journal of Environmental Monitoring 12:1699-1708Hayes KR, Douglas WS, Fischer J. 1996. Inter- and intra-laboratory testing of the Daphnia magna IQ toxicity test™. Bulletin of Environmental Contamination and Toxicology 57:660-666

36. Hinger G, Brinkmann M, Bluhm K, Sagner A, Takner H, Eisenträger A, Braunbeck T, Engwall M, Tiehm A, Hollert H. 2011. Some heterocyclic aromatic compounds are Ah receptor agonists in the DR-CALUX assay and the EROD assay with RTL-W1 cells. Environmental Science and Pollution Research 18:1297-1304

37. Höss S, Ahlf W, Bergtold M, Bluebaum-Gronau E, Brinke M, Donnevert G, Menzel R, Möhlenkamp C, Ratte HT, Traunspurger W, von Danwitz B, Pluta HJ. 2012. Interlaboratory comparison of a standardized toxicity test using the nematode Caenorhabditis elegans (ISO 10872). Environmental Toxicology and Chemistry 31(7):1525-1535

38. Huggett DB, Foran CM, Brooks BW, Weston J, Peterson B, Marsh KE, LaPoint TW, Schlenk D. 2003. Comparison of in vitro and in vivo bioassays for oestrogenicity in effluent from North American municipal wastewater facilities. Toxicological Sciences 72:77-83.

39. ICES. 2012. ICES Techniques in marine environmental sciences No. 52. Guidelines for passive sampling of hydrophobic contaminants in water using silicone rubber samplers, June 2012.

40. Joint Research Centre (JRC). 2017. JRC Reference Report on Monitoring of Emissions to Air and Water from IED installation. Industrial Emissions Directive 2010/75/EU (Integrated Pollution Prevention and Control). Revised Final Draft. June 2017.

41. Jonker MTO, Candido A, Cozmina M, Vrabie CM, Scarlett AG; Rowland SJ. 2016. Synergistic androgenic effects of a petroleum product caused by 3 the joint action of at least three chemically distinct compounds. Chemosphere 144: 1142-1147

42. Kunz PY, Simon E, 1, Creusot N, B. Jayasinghe S, Cornelia Kienle C, Maletz S, Schifferli A, Schönlau C, Aït-Aïssa S, Denslow ND, Hollert H, Werner I, Vermeirssen ELM. 2017. Effect-based tools for monitoring estrogenic mixtures: Evaluation of five in vitro bioassays. Water Research 110:378-388

43. Kittinger C, Baumert R, Folli B, Lipp M, Liebmann A, Kirschner A, Farnleitner AH, Grisold AJ, Zarfel GE. 2015. Preliminary Toxicological Evaluation of the River Danube Using in Vitro Bioassays. Water 7:1959-1968. doi:10.3390/w7051959

44. Lemos CT, Vargas VMF, Henriques JAP, Mattevi MS. 1994. Genotoxicity of river water under the influence of petrochemical industrial complexes. Bulletin of Environmental Contamination and Toxicology 52:848-855

45. Leusch FDL, Khan SJ, Gagnon MM, Quayle P, Trinh T, Coleman H, Rawson C, Chapman HF, Blair P, Nice H, and Tarren Reitsema T. 2014. Assessment of wastewater and recycled water quality: A comparison of lines of evidence from in vitro, in vivo and chemical analysis. Water Research 50: 420-431.

46. Liscio C, Di Carro M, Suter MJF, Vermeirssen ELM. 2009. Combining passive samplers and biomonitors to evaluate endocrine disrupting compounds in a wastewater treatment plant by LC/MS/MS and bioassay analyses. Environmental Pollution 157:2716–2721Metcalfe CD, Kleywegt S, Letcher RJ, Topp E, Wagh P, Trudeau VL, Moon TW. 2013. A multi-assay screening approach for assessment of endocrine-active contaminants in wastewater effluent samples. Science of the Total Environment 454-455:132–140

47. Muller R, Janet Y. M. Tang JYM, Their R, Mueller JF. 2007. Combining passive sampling and toxicity testing for evaluation of mixtures of polar organic chemicals in sewage treatment plant effluent. Journal of Environmental Monitoring 9:104–109

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

49. Nelson J, Bishay F, van Roodselaar A, Ikonomou M, Law FCP. 2007. The use of in vitro bioassays to quantify endocrine disrupting chemicals in municipal wastewater treatment plant effluents. Science of the Total Environment 374:80–90

50. Norman Network. 2008. Protocol for the validation of chemical and biological monitoring methods - Improved version. Network of reference laboratories and related organisations for monitoring and bio-monitoring of emerging environmental pollutants.

51. Norman Network. 2010. Passive sampling of emerging pollutants in the aquatic environment: state of the art and perspectives - Position paper. pp. 44. Vrana B, Vermeirssen ELM, Allan IJ, Kohoutek J, Kennedy K, Mills GA, Greenwood R

52. Nunes E-A, Torres de Lemos C Gavronski L, Moreira TN, Oliveira NCD, da Silva J. 2011. Genotoxic assessment on river water using different biological systems. Chemosphere 84:47–53

53. Osman AGM, Khaled Y. AbouelFadl KY. Krüger A, Kloas W. 2015. Screening of multiple hormonal activities in water and sediment from the river Nile, Egypt, using in vitro bioassay and gonadal histology. Environmental Monitoring and Assessment 187:317

54. OSPAR Commission. 2012. OSPAR Guidelines in support of recommendation 2012/5 for a risk-based approach to the management of produced water discharges from offshore installations. OSPAR Agreement: 2012-7.

55. Olivares A, Quirós L, Pelayo S, Navarro A, Bosch A, Grimalt JO, Fabregat M, Faria M, Benejam L, Benito J, Solé M, Carlos Barata C, Benjamin Piña B. 2010. Integrated biological and chemical analysis of organochlorine compound pollution and of its biological effects in a riverine system downstream the discharge point. Sci. Tot. Environ, 408:5592–5599

56. Pieterse B, Felzel E, Winter R, van der Burg B, Brouwer A. 2013. PAH-CALUX, an optimized bioassay for AhR-mediated hazard identification of polycyclic aromatic hydrocarbons (PAHs) as individual compounds and in complex mixtures. Environmental Science Technology 47:11651−11659

57. Rastall AC, Neziri A, Vukovic Z, Jung C, Mijovic S, Hollert H, Nikcevic S, Erdinger L. 2004. The identification of readily bioavailable pollutants in Lake Shkodra/Skadar using Semipermeable Membrane Devices (SPMDs), bioassays and chemical analysis. Environmental Science and Pollution Research 11:240-253

58. Rastall AC, Getting D, Goddard J, Roberts DR, Erdinger L. 2006. A biomimetic approach to the detection and identification of oestrogen receptor agonists in surface waters using Semipermeable Membrane Devices (SPMDs) and bioassay-directed chemical analysis. Environmental Science and Pollution Research 13:256-267

59. Reifferscheid G, Grummt T. 2000. Genotoxicity in German surface waters - results of a collaborative study. Water, Air, and Soil Pollution 123:67–79

60. Reifferscheid G, Heil J, Oda Y, Zahn RK. 1991. A microplate version of the SOS/umu-test for rapid detection of genotoxins and genotoxic potentials of environmental samples. Mutation Research 253:215-222

61. Reifferscheid G, Ziemann C, Fieblinger D, Dill F, Gminski R, Grummte HJ, Hafner H, Hollert H, Kunz S, Rodrigo G, Stopper H, Selke D. 2008. Measurement of genotoxicity in wastewater samples with the in vitro micronucleus test—Results of a round-robin study in the context of standardisation according to ISO. Mutation Research 649:15–27

62. Routledge EJ, Sumpter JP. 1996. Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. Environmental Toxicology and Chemistry 15:241-248.

63. Samson B, Vo NTK, Kavanagh R, Hanner R, MacKinnon MD. Dixon DG, Lee LEJ. 2013. Rapid assessment of the toxicity of oil sands process-affected waters using fish cell lines. In Vitro Cell Developmental Biology—Animal, 49:52–65

64. Schirmer K. Dixon DG, Greenberg BM, Bols NCl. 1998. Ability of 16 priority PAHs to be directly cytotoxic to a cell line from the rainbow trout gill. Toxicology 127:129–141

65. Schirmer K, Dayeh VR, Bopp S, Russold S, Bols NC. 2004. Applying whole water samples to cell bioassays for detecting dioxin-like compounds at contaminated sites. Toxicology 205:211–221

66. Schriks M, Baken K, Simon E, Besselink H, van der Linden S, Kienle C, van der Burg B. 2015. Selection criteria to select in vitro bioassays for implementation and use. Report from “Demonstration of promising technologies to address emerging pollutants in water and waste water (DEMEAU)” WP41 October 2015.

67. Scott PD, Bartkow M, J. Blockwell SJ, Coleman HM, Khan SJ, Lim R, McDonald JA, Nice H, Nugegoda D, Pettigrove V, Tremblay LA, Warne MSJ, Leusch FDL. 2014. An assessment of endocrine activity in Australian rivers using chemical and in vitro analyses. Environmental Science and Pollution Research 21:12951–12967

68. Sese,B Grant A, Reid BJ. 2009. Toxicity of Polycyclic Aromatic Hydrocarbons to the Nematode Caenorhabditis elegans. Journal of Toxicology and Environmental Health, 72:1-13

69. Smith AJ, McGowan T, Devlin MJ, Massoud MD, Al-Enezi M, AAl-Zaidan AS, Al-Sarawi HA, Lyons BP. 2015. Screening for contaminant hotspots in the marine environment of Kuwait using ecotoxicological and chemical screening techniques. Marine Pollution Bulletin 100:681-688

70. Steliga T, Jakubowicz P, Kapusta P. 2015. Changes in toxicity during treatment of wastewater from oil plant contaminated with petroleum hydrocarbons. Chemical Technology and Biotechnology 90: 408-1418

71. Suares-Rocha P, Braunbeck T, de Francheschi de Angelis D, Marin-Morales MA. 2015. Assessment of cytotoxicity and AhR-mediated toxicity in tropical fresh water sediments under the influence of an oil refinery. Environmental Science and Pollution Research 22:12566–12575

72. Tang JM, McCarty S, Glenn E, Neale PA, Warne MSJ, Escher BI. 2014. Mixture effects of organic micropollutants present in water: values for baseline toxicity. Water Research 47:3300-3314

73. Tanneberger K, Knöbel M, Busser FJM, Sinnige TL, Hermens JLM, Schirmer K. 2013. Predicting fish acute toxicity using a fish gill cell line-based toxicity assay. Environmental Science and Technology 47:1110-1119

74. Thomas KV, Hurst MR, Matthiessen P, McHugh M, Smith A, Waldock MJ. 2002. An assessment of in vitro androgenic activity and the identification of environmental androgens in United Kingdom estuaries. Environmental Toxicology and Chemistry 21(7):1456-1461

75. Thomas KV, Balaam J, Hurst M, Nedyalkova Z, Mekenyan O. 2004a. Potency and characterization of oestrogen-receptor agonists in United Kingdom estuarine sediments. Environmental Toxicology and Chemistry 23(2):471-479

76. Thomas KV, Balaam J, Hurst MR, Thain JE. 2004b. Identification of in vitro oestrogen and androgen receptor agonists in North Sea offshore produced water discharges. Environmental Toxicology and Chemistry 23(5):1156-1163

77. Timme-Laragy AR, Van Tiem LA, Linney EA, Di Giulio RT. 2009. Antioxidant Responses and NRF2 in Synergistic Developmental Toxicity of PAHs in Zebrafish, Toxicological Sciences, 109, Pages 217–227,

78. Tollefsen KE, Harman C, Smith A, Thomas KV. 2007. Oestrogen receptor (ER) agonists and androgen receptor (AR) antagonists in effluents from Norwegian North Sea oil production platforms. Marine Pollution Bulletin 54:277-283

79. United Kingdom Water Industry Research (UKWIR). 2000. UK Direct Toxicity Assessment (DTA) Demonstration Programme Lower Tees Estuary Demonstration Project Part. UKWIR Report 00/TX/02/03

80. United Kingdom Department for Energy and Climate Change. 2014. The United Kingdom risk based approach implementation programme. Version 2 July 2014.Urbatzka R, van Cauwenberge A, Maggioni S, Viganò L, Mandich A, Benfenati E, Lutz I, Kloas W. Androgenic and antiandrogenic activities in water and sediment samples from the river Lambro, Italy, detected by yeast androgen screen and chemical analyses. Chemosphere 67:1080-1087

81. Van der Linden SC, Bergh ARM, van Vught-Lussenburg BMA, Jonker LRA, Teunis M, Cyrille A.M. Krul CAM, van der Burg B. 2014. Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress. Mutation Research/Genetic Toxicology and Environmental Mutagenesis 760:23–32

82. Van der Oost R, Sileno G, Suárez-Munoz M, Nguyen MT, Besselink H, Brouwer A. 2017a. SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: Part I–model design and effect-based trigger values. Environmental Toxicology and Chemistry 36:2385–2399

83. Van der Oost R, Sileno G, Suárez-Munoz M, Nguyen MT, Besselink H, Brouwer A. 2017b. SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: Part II – Field feasibility survey. Environmental Toxicology and Chemistry 36:2400–2416

84. van Hattum B, Senhorst H, Tukker A, Lamoree M, Sanderson T, de Koning A. 2004. Evaluation of current emissions of chlorinated microcontaminants from the Dutch chlorine chain. Final Technical Report of the Chlorine Chain Follow-up Research Programme on Chlorinated Organic Microcontaminants (OVOC) E-04/08, November 2004

85. Vethaak AD, Hamers T, Martínez-Gómez C, Kamstra JH, de Weert J, Leonards PEG, Smedes F. 2017. Toxicity profiling of marine surface sediments: A case study using rapid screening bioassays of exhaustive total extracts, elutriates and passive sampler extracts. Marine Environmental Research 124:81-91

86. Vincent-Hubert F, Uher E, Di Giorgio C, Michel C, De Meo M, Gourlay-France C. 2016. Use of low density polyethylene membranes for assessment of genotoxicity of PAHs in the Seine River. Ecotoxicology 26(2):165-172

87. Vrabie CM, Jonker MTO, Murk AJ. 2009. Specific in vitro toxicity of crude and refined petroleum products. 1. Aryl hydrocarbon receptor–mediated responses. Environmental Toxicology and Chemistry 28:1995–2003.

88. Vrabie CM, Candido A, van Duursen MBM, Jonker MTO, Murk AJ. 2010. Specific in vitro toxicity of crude and refined petroleum products. II Oestrogen (α and β) and androgen receptor-mediated responses in yeast assays. Environmental Toxicology and Chemistry 29:529–1536.

89. Vrabie CM, Candido A, ven den Berg H, Murk AJ, van Duursen MBM, Jonker MTO. 2011. Specific in vitro toxicity of crude and refined petroleum products: 3. estrogenic responses in mammalian assays. Environmental Toxicology and Chemistry 30:973–980

90. Wadhia K, Dando T, Thompson KC. 2007. Intra-laboratory evaluation of Microbial Assay for Risk Assessment (MARA) for potential application in the implementation of the Water Framework Directive (WFD). Journal of Environmental Monitoring 9:953-958

91. Wadhia K and Dando TR. 2009. Environmental toxicity testing using the Microbial Assay for Risk Assessment (MARA). Fresenius Environmental Bulletin 18(2):220-225

92. Wang L, Ying GG, Zhao JL, Liu S, Tang B, Zhou LJ, Tao R, Su HC. 2011. Assessing estrogenic activity in surface water and sediment of the Liao River system in northeast China using combined chemical and biological tools. Environmental Pollution 159:148-156

93. Wca. 2017. Global use of biomarkers. Report to Total, January 2017.

94. Wieczerzak M, Namieśnik J, Kudłak B. 2016. Bioassays as one of the Green Chemistry tools for assessing environmental quality: A review. Environment International 94:341-361

95. Žegura B, Heath E, Černoša A, Filipič M. 2009. Combination of in vitro bioassays for the determination of cytotoxic and genotoxic potential of wastewater, surface water and drinking water samples. Chemosphere 75:1453-1460

96. Zemanek MG, Pollard SJ, Kenefick SL, Hrudey SE. 1997. Toxicity and mutagenicity of component classes of oils isolated from soils at petroleum- and creosote-contaminated sites. Journal of the Air & Waste Management Association 47(12):1250-1258

97. Zhao JL, Ying GG, Chen F, Liu YS, Wang L, Yang B, Liu S, Tao R. 2011. Estrogenic activity profiles and risks in surface waters and sediments of the Pearl River system in South China assessed by chemical analysis and in *vitro* bioassay. Journal of Environmental Monitoring 13(4):813-821

7. ABBREVIATIONS

AB-PS	Adsorption-Based Passive Samplers
AhR	Aryl Hydrocarbon Receptor
AOP	Adverse Outcome Pathway
AR	Androgen Receptor
ARE	Antioxidant Response Element
B(a)P	Benzo(a)Pyrene
BAT	Best Available Technique
BEC or BEQ	Bioanalytical Equivalent Concentrations
CALUX	Chemical-Activated Luciferase Gene Expression
Cefic	European Chemical Industry Council
CFDA-AM	5-Carboxyfluorescein Diacetate Acetoxymethyl Ester
DART	Developmental and Reproductive Toxicity
DEMEAU	Demonstration of promising technologies to address emerging pollutants in water and waste water
DHT	Dihydrotestosterone
DR	Dioxin-Responsive
DTA	Direct Toxicity Assessment
EBT	Effect Based Tools
EB-TV	Effect Based Trigger Value
EC	European Commission
ECx	Effective Concentration
EC-IR	The concentration causing an Induction Ratio (IR) of x
EDA	Effects Driven Analysis
ED	Endocrine Disruption
EEFS	17β-estradiol (E2) Equivalence Factors
EEQ	Equivalent Factors
EpRE	Electrophile Responsive Element
EQS	Environmental Quality Standard
ER	Oestrogen Receptor
ERE	Oestrogen Responsive Elements
EROD	7-ethoxyresorufin-O-deethylase
FLU	Flutamide
GFP	Green Fluorescent Protein
GC-MS	Gas chromatography–mass spectrometry
GR	Glucocorticoid Receptor
GTU	Genotoxic Unit
hGADD45a	Human GADD45a gene

HOCNF	Harmonised Offshore Chemical Notification Format
ICES	International Council for the Exploration of the Sea
ICx	Inhibitory Concentration
IED	Industrial Emissions Directive
IEQ	Induction Equivalents
IR	Induction Ratio
ISO	International Organisation for Standardisation
LC-MS	Liquid chromatography–mass spectrometry
LCx	Lethal Concentration
LDPE	Low Density Polyethylene samplers
LOD	Limit of Detection
LOQ	Limit of Quantification
LOR	Limit of Reporting
LRi	Long Range Initiative
LR-TV	Low Risk Trigger Values
LUC	Luciferase
MARA	Microbial Assay for Risk Assessment
MSFD	Marine Strategy Framework Directive
MTC	Microbial Toxic Concentration
MTT	3-(4,5-dimethyl-2-thiazolyl)-2, 5-diphenyl-2H-tetrazolium bromide formazan
MUF	Methylumbelliferyl Galactoside
NOEC	No Observed Effect Concentration
NORMAN	Network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances
OCNS	Offshore Chemical Notification Scheme
OSPAR	The Convention for the Protection of the Marine Environment of the North-East Atlantic
PAH	Polycyclic Aromatic Hydrocarbons
PAM	Pulse-Amplitude Modulation
PB-PS	Partition-Based Passive Samplers
PR	Progesterone Receptor
PRC	Performance Reference Compounds
POCIS	Polar Organic Chemical Integrative Sampler
PXR	Pregnane X Receptor
qPCR	quantitative Polymerase Chain Reaction
QFET	Q Fish Embryo Toxicity
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
REF	Relative Enrichment Factor

REP	Relative Effect Potency
RT Gill-W1	Rainbow Trout Gill cells
SPMD	Semi-Permeable Membrane Devices
SSD	Species Sensitivity Distribution
tBHQ	Tert-Butylhydroquinone
TCDD	2,3,7,8-Tetrachlorodibenzo-p-Dioxin
TEF	Toxicity Equivalent Factors
TEQ	Toxic Equivalent Concentrations
TIPTOP	Time-Integrative Passive sampling combined with Toxicity Profiling
TMXEQ	Tamoxifen Equivalent
TTC	Triphenyl Tetrazolium Chloride
TTR	Transthyretin
WFD	Water Framework Directive
WSWMG	(Concawe) Water, Soil and Waste Mangement Group
WWTP	Wastewater Treatment Plant
YAS	Yeast Androgen Screen
YES	Yeast Oestrogen Screen
YETI	Yeast Environmental Toxicity Indicator

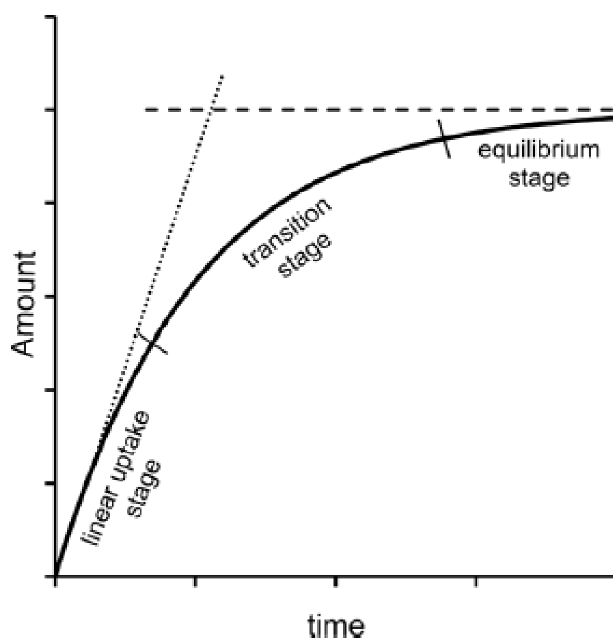
APPENDIX 1 PASSIVE SAMPLING

Passive sampling describes the use of devices maintained *in situ* within the water column or that, over time, passively (i.e. by diffusion) accumulate substances that are present in the water. This process therefore facilitates the analysis of substances (or mixtures of substances) that are diffuse in the environment by concentrating them on the sampler. The accumulated substances can then be extracted from the device and the concentrations of specific substances measured, or the extracts can be utilised in effect-based assays. Because they concentrate substances over time, passive sampler extracts do not directly represent the concentrations of substances present in the deployment environment, and these must be adjusted to account for the deployment time to estimate environmental concentrations (i.e. time weighted average concentrations). Similarly, the concentration of substances within the sampler may mean that passive sampler extracts (i.e. chemical mixtures) can illicit responses in biological assays which would not be sufficiently sensitive to respond to the actual environmental samples (i.e. the same mixture of substances at environmentally realistic concentrations). Therefore, it is important to highlight that effects measured in biological assays applied to passive sampler extracts does not mean that adverse effects are occurring in the environment. Nevertheless, such an approach may be useful in comparing and prioritising sites for further investigation.

There are typically two types of passive sampler, partition-based passive samplers (PB-PS) which accumulate hydrophobic substances, including hydrocarbons, and adsorption-based passive samplers (AB-PS) which adsorb polar organic compounds. AB-PS samplers are less well developed compared with the PB-PS samplers, and are based on diffusion through porous membranes and sorption onto adsorbent materials. Examples include Speedisk (Hamers et al. 2016) and Polar Organic Chemical Integrative Sampler (POCIS). PB-PS samplers are based on diffusion and adsorption, usually to a non-porous polymer such as polyethylene or polydimethylsiloxane (PDMS). Examples include semi-permeable membrane devices (SPMDs), low density polyethylene (LDPE) strip samplers, and silicone strip samplers (ICES 2012).

When deployed, substances will initially be absorbed in the PB-PS device at a rate directly proportional to their aqueous concentration until this approaches equilibrium (linear uptake stage; **Figure A1**). The time to equilibration is usually longer for substances with a high octanol water partition co-efficient.

Figure A1 Contaminant amounts absorbed by passive samplers as a function of time (ICES 2012)



Uptake onto PB-PS samplers generally depends on factors such as exposure time, water flow and temperature, and therefore the concentration of substances accumulated by devices is quantified by assessing the dissipation of Performance Reference Compounds (PRCs), which are spiked onto the sampler prior to deployment (ICES 2012). ICES (2012) guidance exists for the application of silicone rubber-based passive samplers for surface water monitoring, and polymer-water partition coefficients (K_{pw}) have been derived for several groups of substances (including Polycyclic Aromatic Hydrocarbons (PAHs)) for these types of sampling device. These silicone rubber samplers were applied in the TIPTOP survey for detecting non-polar substances, including hydrocarbons.

Accumulated substances are extracted from the sampler and extracts cleaned-up, followed by transfer to non-polar solvents, before chemical analysis (Liquid or Gas Chromatography combined with Mass Spectrometry (LC-MS or GC-MS)). Aqueous concentrations are calculated based on the water sampling rates of the sampler, which are calculated using the ratio of PRC concentration at the beginning and end of the device deployment. These values may be further corrected based on sampling timeframe and uptake models in cases when the sampling has not reached equilibrium. Sampling rates and equilibration time can be enhanced by maximising the sampler surface area to volume ratio and by minimising the resistance to accumulation of substances, which can be enhanced by coating samplers with very thin film sheets made from polymers with rapid chemical diffusion capability (such as silicone rubber) (Brack et al. 2016). In general, up to 1000 times concentration (compared to environmental concentrations) can be achieved using silicone rubber devices for hydrophobic substances (Vethaak et al. 2017).

Extracts from passive samplers are increasingly being tested using bioassays in an effort to link biological and chemical measurements. However, there are technical challenges in this approach because extracts from PRC spiked samplers cannot be used in bioassays, there is a potential for extract solvent toxicity and sampler extract volumes are usually very small. In general, such issues are addressed by deploying separate samplers in the same location for chemical and biological analysis, ensuring that sampler extracts are diluted sufficiently to prevent solvent toxicity (which obviously also affects the maximum extract concentration that can be tested), and maximising sample volume by applying larger sampling devices (e.g. PDMS) since sampler size limits sample volume.

As passive samplers take up individual contaminants at different rates, and the uptake rateis dependent on the variable compound properties of the contaminants, the effect measured in passive sampler extracts can only be considered as an estimate of the actual effects in the environment. Brack et al. (2016) suggest the passive sample extracts broadly reflect the uptake of substances into an organism, but obviously do not account for metabolism. For example, more bioaccumulative substances with higher lipid partitioning properties would be expected to be more concentrated within an extract. Passive sampler extracts have the advantage of providing a method for integrating exposure over time, which has often been a criticism of traditional surface monitoring using bioassays. Overall, studies on the biological effects of passive sampler extracts are generally limited to qualitative assessment (rather than a definitive hazard or risk assessment). The results from such a study could therefore be helpful in the prioritisation of sites that might be of environmental concern for further investigation.

APPENDIX 2 FULL LIST OF EBTS

EBT	Type	Endpoint	Commercially available (Y/N)	Validation level (1-3)[1]	Proven for use with Environmental samples (Y/N)	Can be used with passive sample extracts (Y/N)	Shortlisted [2]
AhR activation assay / LUC (H4L1.1c4)	Metabolism	Dioxin like activity through AhR activation based on luminescence	Y	3	Y	Y	Y
AhR activation assay (with more specificity for PAHs)	Metabolism	AhR receptor activity based on luminescence	Y	2	Y	Y	Y
ER-LUC (BG1Luc4E2)	ED	Estrogenic receptor activity based on luminescence	Y	3	Y	Y	Y
AR activation assay using hamster cell line	ED	Androgenic receptor activity based on luminescence	Y	2	Y	Y	Y
AR activation assay using human bone marrow cell line	ED	Androgen receptor activity based on luminescence	Y	2	Y	Y	Y
GR activation assay	ED	Glucocorticoid receptor activity based on luminescence	Y	2	Y	Y	N
PR activation assay	ED	Progesterone receptor activity based on luminescence	Y	2	Y	Y	N
TRb activation assay	ED	Thyroid receptor activity based on luminescence	Y	1	N	Y	N

EBT	Type	Endpoint	Commercially available (Y/N)	Validation level (1-3)[1]	Proven for use with Environmental samples (Y/N)	Can be used with passive sample extracts (Y/N)	Shortlisted [2]
TTR-binding	ED	Displacement of thyroid hormone precursor thyroxine (T4) from its plasma transport protein transthyretin (TTR)	Y	1	Y	Y	Y
PPAR activation assay (α or δ or γ) (separate assays)	ED	Peroxisome proliferator activated receptor binding based on luminescence	Y	1	N	Y	N
YES	ED	Yeast oestrogen screen	Y	3	Y	Y	Y
YAS	ED	Yeast androgen screen	Y	3	Y	Y	Y
GH3-TRE.Luc assay	ED	Thyroid receptor activity	N	1	N	Y	N
Vitellogenin Induction	ED	Induction of vitellogenin	Y	3	Y	N	N
Cox inhibitor assay	Pharmaceutical specific	Inhibition of cyclooxygenase-1 based on fluorescence	N	1	N	Y	N
ABC assay	Pharmaceutical specific	Antibiotic activity	3	1	Y	Y	N
Ames II	Genotoxicity	Mutagenic activity in TA98 strain based on number of revertants	Y	3	Y	Y	Y

EBT	Type	Endpoint	Commercially available (Y/N)	Validation level (1-3)[1]	Proven for use with Environmental samples (Y/N)	Can be used with passive sample extracts (Y/N)	Shortlisted [2]
umuC	Genotoxicity	SOS response to DNA damage	Y	3	Y	Y	Y
p53-pathway activation assay	Genotoxicity	DNA damage response based on fluorescence	Y	2	N	Y	Y
hGADD45a activation assay	Genotoxicity	DNA damage response based on fluorescence	Y	2	Y	Y	Y
Micronucleus – *in vivo*	Genotoxicity	Frequency of micronuclei indicating chromosomal damage in vivo	Y	3	Y	N	N
Micronucleus – *in vitro*	Genotoxicity	Frequency of micronuclei indicating chromosomal damage in a cell line	Y	3	Y	Y	Y
Comet assay	Genotoxicity	Tail length or moment indicating DNA strand breaks	Y	2	Y	Y	N
DNA adduct formation	Genotoxicity	Number of adducted nucleotides per number of undamaged nucleotides	Y	2	Y	N	N
Nrf2-pathway activation assay	Oxidative stress	Activation of Nrf2 transcriptional activity based on luminescence	Y	2	N	Y	Y
AREc32	Oxidative stress	Activation of Nrf2-Antioxidant Response Element (ARE) activity based on luminescence	Y	2	Y	Y	Y

EBT	Type	Endpoint	Commercially available (Y/N)	Validation level (1-3)[1]	Proven for use with Environmental samples (Y/N)	Can be used with passive sample extracts (Y/N)	Shortlisted [2]
Multi-species microbial toxicity test	In vivo	Microbial growth (e.g. MARA) or bioluminesence (e.g. LumiMara)	Y	3	Y	Y	Y
Algae pulse-amplitude modulation (PAM)	in vivo	Reduced photosynthetic efficiency of green alga *Pseudokirchneriella subcapitata*	Y	2	Y	Y	N
Algae microplate assay	in vivo	Combined growth and photosynthesis of marine or freshwater algae	Y	3	Y	Y	N
Algae microplate assay with PAM	in vivo	Algal growth and photosynthetic efficiency	Y	2	Y	Y	N
Lemna minor growth	In vivo	Plant growth	Y	3	Y	N	N
Thamnocephalus platyurus toxicity test (Thamnotoxkit F™; ISO 1430)	in vivo	Juvenile mortality of crustacean *Thamnocephalus platyurus*	Y	3	Y	Y	N
Daphnia magna acute toxicity test	in vivo	Immobility	Y	3	Y	Y	N
Daphnia magna chronic toxicity test	in vivo	Development and reproduction	Y	3	Y	N	N

EBT	Type	Endpoint	Commercially available (Y/N)	Validation level (1-3)[1]	Proven for use with Environmental samples (Y/N)	Can be used with passive sample extracts (Y/N)	Shortlisted [2]
Brachionus calyciflorus toxicity test (Rotoxkit F™; ISO 19827)	in vivo	Juvenile mortality (24h) of rotifer *Brachionus calyciflorus*	Y	3	Y	Y	N
QFET	in vivo	Zebrafish, *Danio rerio* (24 -120h) embryotoxicity; tailbud and heart beat	Y	3	Y	Y	Y
Zebrafish Toxarray	In vivo	Zebrafish change in gene expression	N	1	N	Y	Y
Daphnia Magna metabolic activity test	In vivo	Decreased substrate metabolization based on fluorescence	Y	3	Y	Y	Y
AChE inhibition	Neurotoxicity	Inhibition of Acetylcholine esterase activity	Y	2	Y	Y	N
AChE inhibition	Neurotoxicity	Inhibition of Acetylcholine esterase activity	Y	2	Y	Y	N
Carboxylesterase inhibition	Neurotoxicity	Inhibition of Carboxylesterase activity	Y	2	Y	Y	N
EROD activity	Metabolism	7-ethoxyresorufin-O-deethylase (EROD) induction as an indication of Cytochrome P450 1A activity	Y	3	Y	N	N

EBT	Type	Endpoint	Commercially available (Y/N)	Validation level (1-3)[1]	Proven for use with Environmental samples (Y/N)	Can be used with passive sample extracts (Y/N)	Shortlisted [2]
EROD activity with RT Liver-W1 cells	Metabolism	7-ethoxyresorufin-O-deethylase (EROD) induction	Y	2	Y	Y	Y
Danio EROD	Metabolism / *in vivo*	EROD induction in *Danio rerio in vivo*	N	1	Y	Y	N
Metallothionein	Metal specific	Metallothionein protein induction	Y	2	Y	N	N
ALA-D	Metal (lead) specific	Inhibition of the enzyme δ-aminolevulinic acid dehydratase (ALA-D)	Y	2	Y	N	N
Caenorhabditis elegans growth and reproduction	in vivo	Growth and reproduction	Y	3	Y	Y	Y
Fish cell lines (RT Gill-W1)	in vitro	Cytotoxicity measured in three assays as a surrogate for acute effects in fish	Y	2	Y	Y	Y
Echinoderm or Oyster Embryo larvae screens	in vivo	Embryo development	Y	3	Y	Y	N
Copepod tests	In vivo	Acute toxicity	Y	3	Y	Y	N

EBT	Type	Endpoint	Commercially available (Y/N)	Validation level (1-3)[1]	Proven for use with Environmental samples (Y/N)	Can be used with passive sample extracts (Y/N)	Shortlisted [2]
Copepod embryo larvae tests	In vivo	Development	Y	3	Y	N	N
Transgenic embryos	In vivo	In vivo nuclear receptor-mediated reporter receptor mediated gene activation e.g. THbZIP-GFP for thyroid hormone (T3)	Specialist	2	Y	Y	N
MTT formazan assay	Cytotoxicity	Cytotoxicity	Y	3		Y	N
Alamer blue	Cytotoxicity	Cytotoxicity / cell viability	Y	3		Y	N
Neutral red retention	Cytotoxicity	Cytotoxicity / cell viability	Y	3	Y	Y	N

[1] Validation maturity based on a scale of 1-3 as outlined in the Norman Network validation guidance document (Norman Network 2008)

Concawe

Environmental Science for European Refining

Concawe
Boulevard du Souverain 165
B-1160 Brussels
Belgium

Tel: +32-2-566 91 60
Fax: +32-2-566 91 81
e-mail: info@concawe.eu
http://www.concawe.eu

ISBN 978-2-87567-095-3