Technical Proposal for Effect-Based Monitoring and Assessment under the Water Framework Directive

Report to the Common Implementation Strategy (CIS) Working Group Chemicals on the outcome of the work performed in the subgroup on Effect-Based Methods (EBM)

MANDATE 2016-2018

October 2021

Disclaimer:

This technical proposal has been developed through a collaborative framework (the Common Implementation Strategy (CIS)) involving the Member States, EFTA countries, and other stakeholders including the European Commission. The document is a working draft and does not necessarily represent the official, formal position of any of the partners. To the extent that the European Commission's services provided input to this technical document, such input does not necessarily reflect the views of the European Commission. Neither the European Commission nor any other CIS partners are responsible for the use that any third party might make of the information contained in this document. The technical document is intended to facilitate the implementation of Directive 2000/60/EC and is not legally binding. Any authoritative reading of the law should only be derived from Directive 2000/60/EC itself and other applicable legal texts or principles. Only the Court of Justice of the European Union is competent to authoritatively interpret Union legislation.

Co-chairs:

IT: Mario Carere

JRC, European Commission: Teresa Lettieri

SE: Ann-Sofie Wernersson, Niklas Hanson

DE: Sebastian Buchinger, Robert Kase-Pasanen (former representative for Switzerland, then co-representative for Germany)

EBM Drafting group:

DE: Georg Reifferscheid, Anja Duffek

SE: D. G. Joakim Larsson, Maria Kahlert, Willem Goedkoop, Lars Forlin

BE: Carole Chalon, Yves Marneffe, Sofie Van Volsem, Lut Hoebeke

FR: Olivier Perceval, Selim Ait-Aissa

CZ: Premysl Soldan, Petr Tusil

IT: Ines Lacchetti, Stefania Marcheggiani, Monica Potalivo, Sara Valsecchi, Camilla Puccinelli

NL: Eric Verbruggen, Dorien Ten Hulscher

CH: Inge Werner, Etienne Vermeirssen, Eszter Simon, Cornelia Kienle

UK: Joanna Katsadiaki, Brett Lyons, Philipp Antczak

JRC, European Commission, Ispra: Magdalena Niegowska, Livia Gomez Cortes

JRC, European Commission, Geel: Marina Ricci

DG Environment, European Commission: Stephanie Schaan, Helen Clayton

Cefic: Steven Van De Broeck

Concawe: Graham Whale, Mike Spence, Carolina Di Paolo, Markus Hjort

Eurometaux: Chris Cooper, Stijn Baken

Eurelectric: Marie Jo Booth, Frank Marolleau

Syngenta: Mick Hamer

ECPA: Stuart Rutherford

EEB: Ismene Jager

Veolia: Armelle Hebert

Norman Network: Henner Hollert (RWTH Aachen University, DE), Beate Escher (UFZ Leipzig, DE), Valeria Dulio (Ineris, FR), Werner Brack (UFZ Leipzig, DE)

SOLUTIONS FP7 project partners: Werner Brack (UFZ Leipzig, DE) et al.

CONTENTS

1. GLOSSARY AND TERMS AND DEFINITIONS

AA	Annual average
A-YES	Yeast estrogen screen assay using *Arxula adeninivorans*
AChE	Acetylcholinesterase
ADI	Acceptable daily intake
AhR	Aryl hydrocarbon receptor
ALA-D	Delta-amino levulinic acid dehydratase
AMR	Antimicrobial resistance
AOP	Adverse outcome pathway
AR	Androgen receptor
ARG	Antibiotic Resistance Genes
BAC	Background assessment criteria
BaP	Benzo(a)pyrene
BAC	Background Assessment Criteria
BAT	Best Available Techniques
BLM	Biotic Ligand Model
BREFs	BAT Reference Documents
BEQ	Biological equivalence concentrations
BQE	Biological quality elements
CA	Concentration addition
CALUX	Chemical Activated LUciferase gene eXpression
CAT	Catalase
CEMP	Coordinated Environmental Monitoring Programme
CIRCABC	Communication and Information Resource Centre for Administrations, Businesses and Citizens
CIS	Common Implementation Strategy
CMEP	Chemical Monitoring and Emerging Pollutants
Comet	Single cell gel electrophoresis assay (SGGE)
DDT	Dichlorodiphenyltrichloroethane
dl-PCBs	dioxin-like PCBs
DNA	Deoxyribonucleic acid
DPSIR	Drivers – Pressures – State – Impact and Response approach
E1	Estrone

E2	17-beta-estradiol
EAC	Environmental assessment criteria
EBM	Effect-based methods
EBT	Effect-based trigger value
EC	Effect concentration
EC50	Median-effect concentration
ECHA	European Chemicals Agency
EDA	Effects-Directed Analysis
EE2	17-alpha-ethinylestradiol
ER	Estrogen receptor
EROD	Ethoxyresorufin-O-deethylase
EQS	Environmental quality standard(s)
EQSD	EQS Directive on priority substances
FDI	Fish disease index
GES	Good environmental status (MSFD), good ecological status (WFD)
GPx	Glutathione peroxidase
GST	Glutathione-S-transferase
GR	Glucocorticoid receptor
GRed	Glutathione reductase
HELCOM	Helsinki Commission
ICES	International Council for the Exploration of the Sea
IED	Industrial Emissions Directive
JAMP	Joint Assessment and Monitoring Programme
LH	Liver histopathology
LMS	Lysosomal membrane stability
LOQ	Limit of quantification
MAC	Maximum Allowable Concentration
MELN	Luciferase-transfected human breast cancer cell line gene-reporter assay
MIC	Minimum inhibitory concentration
MLN	Liver macroscopic neoplasms
MN	Micronucleus
MoA	Mode of action
MS	Member State(s)
MSFD	Marine Strategy Framework Directive
MT	Metallothionein

NOEC	No observed effect concentration
OECD	Organisation for Economic Co-operation and Development
OOAO	One-out-all-out
OSPAR	Convention for the Protection of the Marine Environment of the North-East Atlantic
PAH	Polycyclic aromatic hydrocarbon
PBDE	Polybrominated diphenylether
PBTs	Persistent, bioaccumulative and toxic substances
PCB	Polychlorinated biphenyl
PCDD	Polychlorinated dibenzodioxins
PCDF	Polychlorinated dibenzofurans
PFOS	Perfluorooctanesulfonic acid
PNEC	Predicted no effect concentration
PoM	Programme of Measures
PPARγ	Peroxisome proliferator-activated receptor
PS	Priority substance
PXR	Pregnane x receptor
QA	Quality assurance
QC	Quality control
RBSP	River Basin Specific Pollutant(s)
REP	Relative effect potency
ROS	Reactive oxygen species
RQ	Risk quotient
RSC	Regional Seas Convention
SCCS	Scientific Committee on Consumer Safety
SCENIHR	Scientific Committee on Emerging and Newly Identified Health Risks
SCHER	Scientific Committee on Health and Environmental Risks
SfG	Scope for Growth
SOP	Standard Operating Procedure
SoS	Stress on Stress
TBT	Tributyltin
TEF	Toxic equivalency factor
TEQ	Toxic equivalents
TIE	Toxicity Identification Evaluation
TIMES	ICES Techniques in Marine Environmental Sciences

ToR	Terms of Reference
TR	Thyroid receptor
uPBT	Ubiquitous persistent bioaccumulative and toxic substance(s)
VDSI	Vas Deference Sequence Index
Vtg	Vitellogenin
WEA	Whole Effluent Assessments
WFD	Water Framework Directive
WG	Working Group
WGBEC	Working Group on the Biological Effects of Contaminants
WHO	World Health Organization
WL	Watch list
YES	Yeast estrogen screen assay using *Saccharomyces cerevisiae*

2. SUMMARY

A specific sub-group for Effect-Based Methods (EBM) was established with representatives from nine Member States (MS), Switzerland and several stakeholders in the context of the Common Implementation Strategy (CIS) for the Water Framework Directive (WFD), specifically the Working Group Chemicals. The Main Objective of the activity of the group has been to examine and further document the possible implementation of effect-based methods for monitoring and assessment in the WFD context, alongside traditional chemical analysis, bearing in mind their possible application also under the Marine Strategy Framework Directive (MSFD). It has built on all scientific evidence and practical knowledge available to-date, including the conclusions of the Chemical Monitoring and Emerging Pollutants (CMEP) work (European Commission 2014-Technical Report on Aquatic Effect-Based Tools) and the estrogen monitoring project. Three meetings (Rome, Prague, Ispra) have been organised during the activity of the sub-group. The activity presented in this report is in line with the Commission Communication on mixtures (EC, 2012) and with the objectives of the 7th Environment Action Programme.

The report is a “Proposal for effect-based monitoring and assessment under the WFD”, it is a further step after the publication in 2014 of the Technical Report on Aquatic Effect-Based Tools because it gives concrete proposals for the application of EBM under the WFD. The report gives clear recommendations for the possible use of EBM in different contexts and scenarios and in Chapter 6 there are examples of these possible applications under the WFD

3. INTRODUCTION

3.1. Current legislative framework and approach

The European Water Framework Directive (WFD; 2000/60/EC) objectives (Art 4 of the WFD) include the aim to achieve and ensure "good ecological and chemical status" of all water bodies throughout Europe through the updating and implementation of management plans at the river-basin level. The Directive employs the DPSIR approach: Drivers – Pressures – State – Impact and Response (Pirrone et al. 2005).

The analysis of important drivers and identification of significant pressures forms the basis for the elaboration of monitoring programmes and programmes of measures (see article 5 and Annex II 1.4. and 1.5. of the WFD). The identified categories of significant pressures, such as urban waste water, agriculture, waste disposal sites, IED and non IED plants and atmospheric deposition need to be reported by the Member States (MS) to the European Commission (WFD reporting guidance 2016). The Common Implementation Strategy (CIS) guidance 3 describes how the "analysis of pressure and assessment of impact" can be performed.

WFD monitoring programmes need to be established by the MS to ensure that sufficient data is generated to assess status and to identify cost efficient measures (art 8 of the WFD). In WFD terms, the monitoring programmes are divided into three categories: surveillance monitoring, operational monitoring and investigative monitoring. Surveillance and operational monitoring programmes should be established on the basis of the water-body characterisation and pressures and impacts assessment required according to WFD art 5 and Annex II (see WFD annex V 1.3.)[1]. There is also a mechanism in place in the WFD triggering investigative monitoring in certain cases[2]. CIS documents 19 and 25 provide further guidance on the establishment of WFD chemical monitoring programmes. The Directive 2009/90/EC ("QA/QC-directive") provides further requirements regarding Quality Assurance (QA) and Quality Control (QC) for the chemical analysis to be used in operational monitoring.

The WFD assessment of quality of surface water bodies is based on an integrated approach, taking into account the following aspects (Annex V of the WFD):

- biological effects observed at population and community level, defined in terms of the values of the Biological Quality Elements (BQE), being phytoplankton, macroalgae, angiosperms, benthic invertebrate fauna and fish and the use of specific indices and ecological quality ratios;
- hydrological and morphological conditions;
- physico-chemical elements (such as pH and nutrient concentrations);

[1] "On the basis of the characterisation and impact assessment carried out in accordance with Article 5 and Annex II, Member States shall for each period to which a river basin management plan applies, establish a surveillance monitoring programme and an operational monitoring programme. Member States may also need in some cases to establish programmes of investigative monitoring".

[2] Investigative monitoring shall according to WFD Annex V. 1.3.3. more specifically be carried out "where the reason for any exceedances is unknown", where surveillance monitoring "indicates that the objectives set out in Article 4 for a body of water are not likely to be achieved and operational monitoring has not already been established, in order to ascertain the causes of a water body or water bodies failing to achieve the environmental objectives", or "to ascertain the magnitude and impacts of accidental pollution, and shall inform the establishment of a programme of measures for the achievement of the environmental objectives and specific measures necessary to remedy the effects of accidental pollution".

- concentrations of toxic substances (such as PFOS, cadmium and dioxins).

The assessment of dangerous chemical substances is regulated in two ways: by way of a separate "chemical status" for currently 45 EU priority substances (PS), and by way of quality elements ("river basin specific pollutants, RBSP") that are part of the "ecological status" (on average, each MS has around 60 RBSP). Environmental Quality Standards (EQS) for these substances are set at EU level for "chemical status"and at MS level for RBSP. The former EQS are included in the EQS Directive on priority substances (EQSD) (2008/105/EC as amended by 2013/39/EU and the PS are specified in WFD Annex X). At national level, MS may also develop and apply standards for alternative environmental compartments than those specified for a particular PS in the EQSD (see Art 3 in the EQSD), including sediment. A prerequisite is however that the alternative EQS corresponds to at least the same level of protection.

As will be described in more detail below, there are currently very few biological indices, applied under the WFD, that respond to toxic chemicals. Thus, the Biological indices established today in general would not respond to the toxic action of chemicals but rather other types of stressors, such as low oxygen levels. Nevertheless, the EQS for both PSs and RBSP are designed to protect the aquatic environment (pelagic and benthic organisms), human health through dietary intake of fish and seafood or drinking water as well as birds and mammals that are exposed through aquatic food webs ("secondary poisoning"). In some cases, drinking water protection can also be the main driver of a water EQS, and applied for water bodies that are used for drinking water extraction. The methodology used to establish such EQS for water, biota and sediment is described in detail in TGD CIS guidance 27.

For water, there are two types of EQS:

- The annual average (AA) EQS[3] is normally set as a water concentration based on chronic effects data for direct toxicity but it can also be based on recalculation from other compartments, in particular biota, depending on which protection goal is the most sensitive. Thus, the overall purpose of this standard is to ensure long-term water quality to protect pelagic organisms, and to protect human health and fish-eating birds and mammals from secondary poisoning.
- The maximum allowable concentration (MAC) is based on acute effects data for direct toxicity. The purpose of this standard is to protect pelagic organisms from short-term concentration peaks.

Sediment EQS aim at protecting benthic organisms from substances accumulating in sediment. Biota EQS are established when the main driver[4] is to protect human health (when exposed to substances in fish and seafood) and/or predators (e.g. fish-eating birds) from the risk of secondary poisoning from substances accumulating in prey.

---

[3] Please note that although the standard is expressed as an "annual" average, i.e. a one year period, shorter time periods can sometimes be more appropriate e.g. for pesticides where their use and exposure tend to be seasonal. See also CIS 27 foot note 6: "When the exposure pattern for a substance is known to be episodic e.g. many pesticides, the averaging period may be a shorter period than a year. This is case-specific but is determined by the expected exposure pattern, not toxicology"

[4] According to CIS 27 procedures, quality standards (QS) developed for different compartments and protection purposes (e.g. QSsec pois to protect predators from secondary poisoning) are recalculated into water concentrations. The lowest water concentration indicates which protection goal is the most critical/sensitive. The recalculated water concentrations can sometimes also be included in the EQSD. This is e.g. the case for PBDE and PFOS and why the EQSD (Art 3.2.) suggests that the "primary" EQS to use is the biota standard. Thus, the water EQS in these cases do not primarily refer to "safe levels" for pelagic organisms (although they are protected too), but rather – indirectly – indicate which are the safe levels to also protect against adverse human health effects and/or secondary poisoning.

CIS guidance 32 describes different aspects to take into account in the assessment of status using biota data and CIS guidance 33 provides guidance on analytical methods for biota.

In most cases, the EQSrelate to single substances. However, in some cases the EQS refer to groups of substances, as in the case of the EQS for dioxins and dioxin-like PCBs, which uses an approach based on TEQs and the potency of the individual substances.

Chemical status has only two status classes: "good chemical status" and "not good chemical status" whereas ecological status is divided into five classes. In the latter case, the EQS defined for RBSP are used to distinguish between "good" and "moderate" ecological status. CIS guidance no. 13 describes in more detail how ecological status assessment is performed. In general and for the PS and RBSP in particular, a "one-out-all-out" (OOAO) approach is used in the classification, meaning that it is sufficient that one single PS or RBSP is present in concentrations above its EQS for the status to "fail" ("not good chemical status" or "moderate ecological status").

The revised EQSD identifies a certain category of PS often referred to as "ubiquitous PBTs" (uPBTs), specified in article 8.a.1.a. in the EQSD. Given their widespread distribution and long recovery times such substances may be monitored less intensively (art 8.a.2. in the EQSD)[5].

Besides the objectives included in WFD art 4, stating that good status should be achieved by 2015[6], status may not deteriorate (often called the "no deterioration principle" and referring to deterioration from one status class to a lower status class for an individual quality element[7]). In the EQSD there is also a specific objective stating that concentrations of PS with a tendency to accumulate in sediment or biota may not increase significantly. The concentrations of those PS (in particular those specified in art 3.6. of the EQSD) therefore need to be monitored in sediment and/or biota. The concentration trend is not taken into account in the assessment of status ("compliance checking") but "Member States shall take measures aimed at ensuring, subject to Article 4 of Directive 2000/60/EC, that such concentrations do not significantly increase in sediment and/or relevant biota" (EQSD art 3.6.). In addition, an overall objective of the WFD is to eliminate pollution of surface water by the PS (see e.g. art 1), and EQSD requires MS to establish an inventory of emissions, discharges and losses of substances included in the EQSD.

[5] See also preamble 13 of directive 2013/39/EC: "Monitoring should be adapted to the spatial and temporal scale of the expected variation in concentrations. Given the widespread distribution and long recovery times expected for substances behaving like ubiquitous PBTs, Member States should be allowed to reduce the number of monitoring sites and/or the frequency of monitoring for those substances to the minimum level sufficient for reliable long-term trend analysis, provided that a statistically robust monitoring baseline is available."

[6] For PS added in 2013, the objective is rather to reach good status by 2027.

[7] See also the conclusions made in the Weser case (C-461/13).

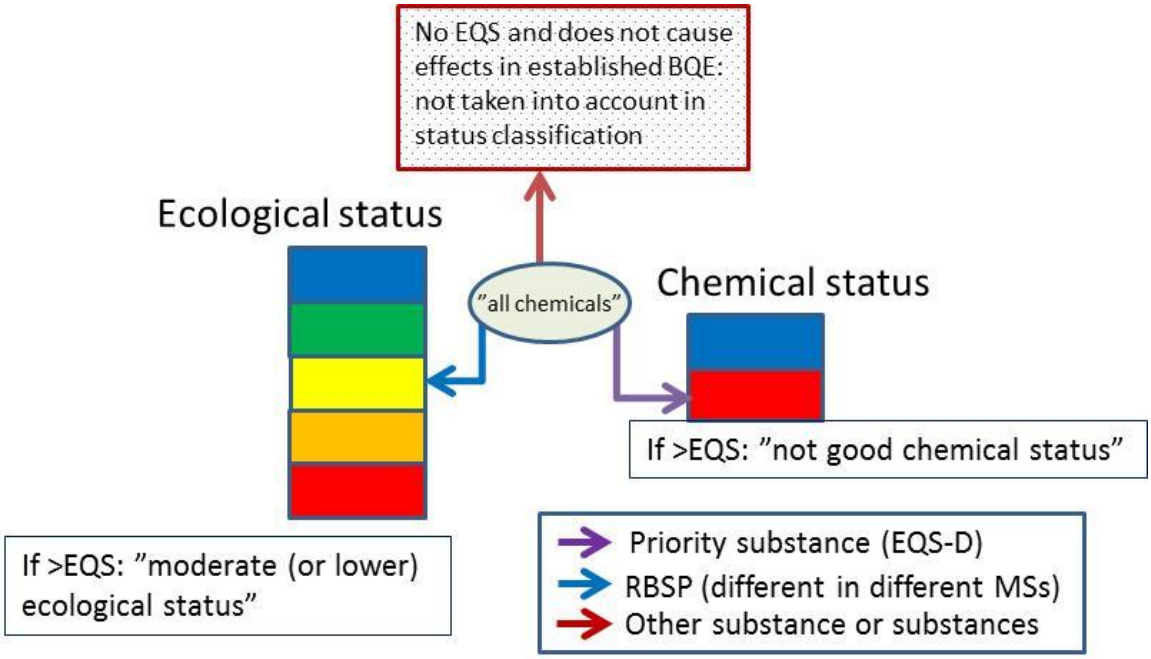

Figure 1. Simplified illustration of the current WFD strategy to take toxic chemicals into account in the status classification. For chemicals included in the EQSD and in concentrations above their EQSs, chemical status will be classified as "not good". For RBSP occurring in concentrations above their EQS in the respective MS, ecological status will be classified as "moderate" (at most). Other chemicals, for which no EQS has yet been established and for which no response is observed in a recognised biological index will not be taken into account in the WFD status classification.

3.2. Need for a holistic approach

3.2.1. Limitations of the current WFD approach to regulate toxic chemicals

As was evident from the previous section, under the WFD the toxicity of chemical substances is currently taken into account using mainly a substance-by-substance approach. Thus, the WFD status assessments are largely based only on chemical analytical data and the limited availability of valid (eco)toxicity datasets for setting EQS. The purpose is to protect aquatic organisms, human health and predators exposed via the aquatic environmentby applying EQS that take these pathways into account. However, this classical single-chemical risk-assessment approach for the management of chemical pollution of water bodies has some limitations as follows (Altenburger et al 2015; Brack et al 2015):

- It is not possible to analyse, detect and quantify all substances that are present in the aquatic environment. Thus, the environmental impact of substances not yet regulated and/or monitored under the WFD will not be considered. Under the Regulation on chemicals (REACH), more than 100.000 chemical substances have been registered;
- The effects caused by the mixtures of substances present in the aquatic environment may not be predictable on the basis of chemical analyses alone.

To reach the protection goal we also must understand the potential for effects caused by the sum of the chemical substances in the aquatic environment (including emerging pollutants, metabolites and transformation products) and link the observed effects with cost-effective management options. As was pointed out earlier, WFD assessment criteria for chemicals (EQS) are generally developed substance-by-substance, based on laboratory studies, and

usually do not consider the consequences of exposure to multiple chemicals or cumulative effects from several stressors or modifying factors. Furthermore, to derive EQS and to establish monitoring programmes for all these substances is highly challenging and for the RBSP different MS have so far frequently established in some cases quite different values for the same substance.

3.2.2. Effect-based methods (EBM) – General information

The use of EBMs for monitoring in the WFD context can overcome some of the challenges identified above. The history behind several "legacy" substances shows that they were first identified to be of major concern after observations of adverse effects were made in the environment. For instance, the effects from tributyltin (TBT) were documented about a decade before the effects could be linked to TBT (Blaber 1970, Féral 1980, Smith 1971, Smith 1981). Also, effects from dichlorodiphenyltrichloroethane (DDT) and polychlorinated biphenyl (PCB) were discovered through observations in the aquatic environment and on birds. Thus, these substances were not identified to be of concern through pro-active risk assessments but rather in retrospect. Estrogenic effects have also been observed in the aquatic environment, and several estrogenic substances, such as EE2, that can explain field-observations such as intersex in fish have been identified (Jobling et al. 1995, Harries et al. 1997, Matthiesen and Sumpter 1998; Vos et al. 2000, Kidd et al. 2014, Adeel et al. 2017, Arlos et al. 2018).

A more systematic monitoring of effects would potentially be able to discover additional substances of concern posing a potential threat to ecological systems and/or human health.

Moreover, the use of EBMs in the WFD context could overcome some of the identified challenges with the current WFD approach (see previous section). Several such methods have already been developed and used, not only in research but also in regular monitoring programmes or screening campaigns.

In the 2010-2012 mandate of the CMEP expert group, a specific task was foreseen for the elaboration of a technical report on aquatic effect-based tools. The activity was chaired by Sweden and co-chaired by Italy and progressively involved several MS and stakeholders in an EU-wide group (47 experts). The Technical Report on Aquatic Effect Based Monitoring Tools (European Commission 2014) aimed at presenting the state of the art of aquatic effect-based monitoring methods and at describing how these methods might help EU MS to establish more efficient monitoring programmes (including to reduce monitoring costs) and at the same time cover the aspects described above.

The report published in 2014 described the state of the art of the use of EBMs in Europe, gave a series of recommendations for their use under the WFD and included an annex with 14 case studies and several fact sheets for different EBMs. The report was also published in the Springer Nature open access journal *Environmental Science Europe* (Ref. Wernersson et al, 2015) and has been disseminated across the scientific community through different channels (for example Springer international).
https://enveurope.springeropen.com/articles/10.1186/s12302-015-0039-4

The Technical Report concluded that the main use of effect-based monitoring tools in the current WFD context would be:

- As screening tools, as part of the pressures and impacts assessment to aid in the prioritisation of water bodies to study further;

- To establish early warning systems, to prioritise further studies in areas that are not concluded to be at risk because they are located far from known local sources;
- To take the effects from mixtures of pollutants or not routinely analysed chemicals ("unknowns") into account (e.g. to support investigative monitoring where causes of a decline of specific species are unknown);
- To provide additional support in water and sediment quality assessment, though not as a replacement for conventional chemical and ecological monitoring under the WFD.

It was also concluded that EBMs are at the moment particularly suitable as part of investigative monitoring programmes, for which the regulatory requirements are less formally determined.

3.2.3. EBMs in regulations and guidance documents

The concept of using EBMs is not new and the usefulness of EBMs in a regulatory context has been shown through their inclusion in various guidance documents and pieces of legislation.

In the European food legislation, EBMs (referred to as "screening methods") can be used to assess the level of dioxin contamination of food (589/2014/EG)[8]. For the assessment of the EBM-outcome, an action value is defined, and exceedance of this triggers a further chemical analysis of the sample. The aim for the use of EBM in this context is to focus the effort involved in chemical analysis on suspect samples. Therefore, an EBM has to show a false-compliant rate below 5 % to be accepted as a screening method for dioxins in food. According to the regulation a "lot is compliant if the result of a single analysis performed by a screening method with a false-compliant rate below 5 % indicates that the level does not exceed the respective maximum level of PCDD/Fs and the sum of PCDD/Fs". If exceeded, the lot cannot be sold on the European market[9].

The use of biomarkers in particular has a long tradition in some MS and Regional Seas Conventions. Within the RSCs (OSPAR, HELCOM, UNEP-MAP and the Bucharest convention) and the International Council for the Exploration of the Sea (ICES), several EBMs have long been included in recommended or agreed monitoring programmes although most are not considered mandatory methods for contracting parties. The OSPAR Coordinated Environmental Monitoring Programme (CEMP) generates data that are used in the Joint Assessment and Monitoring Programme (JAMP), and includes both mandatory (CEMP) components and voluntary (pre-CEMP) components. The division is based on an assessment of whether monitoring guidelines, quality assurance tools and/or assessment tools are available. If at least one is missing, the component is included in the pre-CEMP components. At the moment there is only one mandatory EBM (imposex) and it is combined with a chemical analytical requirement (of sediment and/or biota)[10] whereas PAH- and metal-specific effects as well as general effects are included on a voluntary basis[11] (OSPAR Agreement 2016-01).

---

[8] European Commission. Commission Regulation (EC) No. 589/2014 of 2 June 2014 laying down methods of sampling and analysis for the control of levels of dioxins, dioxin-like PCBs and non-dioxin-like PCBs in certain foodstuffs and repealing Regulation (EU) No. 252/2012. *OJ L* **164** (2014).

[9] However, Finland and Sweden have been granted exemptions to sell certain species of fish and from certain regions (such as Baltic herring) in their territories or to each other regardless of the dioxin content, providing that the consumers are fully informed about the potential health risks.

[10] "H4. Tributyl tin (TBT)-specific biological effects and TBT in sediment or biota. Monitoring of TBT concentrations in the marine environment in either sediments or biota should be carried out in parallel with monitoring of TBT-specific biological effects" (OSPAR decision).

[11] "H10 PAH and metal-specific biological effects"; "H11 general biological effects".

In the context of the Marine Strategy Framework Directive (MSFD)[12], EBMs are included as supplementary criteria for descriptor 8 (voluntary basis)[13] to assess good environmental status (GES) under the second criterion for this descriptor (D8C2). In the absence of harmonised guidelines, the application of D8C2 through collaboration at regional and subregional level should include a list of habitats, species and tissue matrices established by MS according to local conditions. As reported in the last submission according to the EU Water Reporting Obligation (Directive 2008/56/EC), some of the MS evaluated biological effects in compliance with the MSFD by using several biomarkers in different taxa of aquatic organisms and species living in coastal areas, such as birds. In the 2012 MSFD initial assessment, in total 29 different biomarkers (Table 1) and one *in vivo* bioassay were mentioned, although most were only reported by one or a few MS. However, imposex in gastropods was used by 10 MS in this context.

[12] Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy.

[13] Please note that in 2017, the initial COM decision on "criteria and methodological standards on good environmental status of marine waters" was revised. With the new decision, the criterion included on effects of contaminants was changed from being mandatory to being supplementary but also rephrased. The previous wording in COM Decision 2010/477/EU ("Levels of pollution effects on the ecosystem components concerned, having regard to the selected biological processes and taxonomic groups where a cause/effect relationship has been established and needs to be monitored") could suggest that primarily very specific biomarkers analysing effects on higher organisational levels should be considered. In the new decision (COM 2017/848), the EBMs that "fit" in under D8 are "broader".

Table 1. EBMs used by some MS for environmental monitoring in the context of the MSFD (from Niegowska et al. 2018).

Mussels	Metallothionein (MT) content Acetylcholinesterase (AChE) activity Glutathione-S-transferase (GST) activity Micronuclei (MN) formation Lysosomal membrane stability (LMS) Scope for growth (SfG) Glutathione peroxidase (GPx) activity Catalase (CAT) activity Cell damage
Fish	Ethoxyresorufin-O-deethylase (EROD) activity Fish disease index (FDI) Levels of bile metabolite 1-hydroxyprene Intersex Formation of DNA adducts Liver tumours Liver pathologies Blood vitellogenin (Vtg) White blood cells alterations Activities of detoxication enzymes Gonad index % deformed larvae
Birds	Chick mortality Mass mortality Breeding success Egg shell thickness Contamination of eggs (coastal birds)
Other biota	Embryos malformations (amphipods) Imposex (gastropods)

EBMs are also mentioned in relation to the HP14 criterion for the assessment of hazardous waste. The properties which render waste hazardous are laid down in Annex III of Directive 2008/98/EC and are further specified by the Decision 2000/532/EC. Primarily the assessment is based on the chemical composition of the waste. However, if the chemical composition is unknown, EBMs, i.e. ecotoxicological tests, are applied.

EBMs have long been used to assess effluents (WEA, Whole Effluent Assessments) containing complex mixtures. As an example, the German waste water ordinance defines waste-water specific threshold values for EBMs, i.e. mostly *in vivo* biological test systems such as the algae test and fish embryo test (FET) for the discharge of waste water. In the Directive 2010/75/EU on industrial emissions including Best Available Techniques (BAT), some BAT Reference Documents (BREFs)[14] require the monitoring of emissions with EBMs.

EBMs can deliver valuable information about possible pressures caused by chemical contamination that are not captured by chemical monitoring or ecological status assessments. Nevertheless, EBMs are mentioned in several CIS documents, see Table 2 below.

[14] See for example Commission Implementing Decision (EU) 2016/902 of 30 May 2016 establishing best available techniques (BAT) conclusions, under Directive 2010/75/EU of the European Parliament and of the Council, for common waste water and waste gas treatment/ management systems in the chemical sector.

Table 2. CIS and related documents referring to the use of EBMs.

Document	WFD/EQSD relevant articles	EBM-related contents
CIS guidance 19	WFD art. 8 and Annex I on monitoring	EBM mentioned as supplementary methods for surface water quality assessment
CIS guidance 25	WFD art. 8 and Annex I on monitoring. EQSD art. 3 on biota and sediment	EBM mentioned for sediment assessment
CIS guidance 27	WFD art. 16 and Annex V on EQS derivation; EQSD art. 3.3. on option to use sediment and biota for status assessments	EBMs mentioned in Section 6.2. on sediment assessment (tier 2)
Technical report on EBMs (incl. Annex) (European Commission 2014)	See above	EBM considered in detail throughout

In the Technical report of 2014 (Wernersson et al), an overview of the use of EBMs in different MS is included (see Section 2.2. in that report). Bioassays are used in individual MS to provide decision support to prohibit the release of toxic substances into the environment (e.g. WEA Whole Effluent Assessment in the permitting process and evaluation of dredged sediments that are considered for sea disposal). They are also used within a broad screening of different sources (such as sewage treatment plant effluents). Other applications include for example the Dutch alarm system that directly triggers control measures (closing drinking water intakes).

3.2.4. Window of opportunity for EBMs

In 2016 the Water Directors endorsed the need for a new approach to the chemical status assessment explicitly stating that EBMs should be used to elaborate a holistic approach for the evaluation of surface water quality (see discussion document presented under work item 2 during the Water Directors' meeting in Bratislava, 28-29 November 2016[15]).

Four basic principles were also suggested:

1. **Instead of continuing with the list of individual PS, establish EQS at EU level for several critical groups of substances, each group characterised by a specific mode of action (or effect type)**. The EQS would represent the maximum acceptable total presence of substances with that particular mode of action (or effect type). If the EQS were exceeded, MS would have to investigate the reason and tackle the source(s) of the offending substance(s).

[15] https://webcache.googleusercontent.com/search?q=cache:rwEFJRAqU90J:https://circabc.europa.eu/sd/a/ea75fb1b-83fd-4eae-8658-78cf5db1ebc8/Final%2520synthesis%2520Bratislava%2520WD.docx+&cd=1&hl=it&ct=clnk&gl=it&client=safari

2. **Continue to require MS to identify pressures from other substances**, i.e. from those not covered by the group EQS or certain individual EQS. Support this process with the **EU Watch List**, focusing on substances not already captured under the groups. Ensure that MS use **harmonised EQS** for these other substances, developed at EU level. Monitoring would be risk based and proportionate, potentially more cost-effective than the current model.

3. **As regards uPBTs**, all of which are currently priority hazardous substances (PHS), **the emphasis would be on achieving at least a stable level or preferably a downward trend in environmental concentrations** (including in biota and sediment), and in parallel progressively ceasing or phasing out emissions, discharges and losses.

4. **Ensure that MS maintain/revise their inventories of emissions**, covering diffuse as well as point sources, so that they can properly carry out the pressures and impacts analysis and identify appropriate measures. The Commission should be able to use these inventories to assess the trends in emissions to water. A downward trend in emissions could be taken into account in the assessment.

It was also concluded that "applying some aspects of the above principles would require the development of new analytical and risk assessment tools that will need to be mature and reliable enough to be taken up in routine practice".

In the WG Chemicals mandate for 2016-2018 it was decided to continue the activity on Effect-Based Methods (previously Effect-Based Tools). The following activities should be included: "*Effect-based assays; links between chemical and ecological status; mixtures. Possible follow-up of estrogen-screening project. Exchange of information on innovative techniques, approaches and potential application in WFD context*".

"In the WFD review, a more holistic approach, taking into account the presence of mixtures of chemicals acting together (for example through the use of effect-based tools in addition to group EQSs), could be considered, to provide a more accurate assessment of risks and a more appropriate targeting of monitoring and measures"

(*from discussion document endorsed by the Water Directors*)

4. ACTIVITY OF THE EBM SUB-GROUP

4.1. Terms of reference

A specific sub-group was established with representatives from nineMS, Switzerland and several stakeholders. The sub-group elaborated the Terms of Reference (ToR) of the Activity after a long discussion at the WG Chemicals and a consultation with the WG Ecostat Group the Marine Strategy Framework Directive WGs. The ToR were finalised in 2016.

The Main Objective of the activity of the group was to examine and further document the possible implementation of EBMs for monitoring and assessment in the WFD context, alongside traditional chemical analysis, bearing in mind their possible application also under the MSFD. It set out to build on all scientific evidence and practical knowledge available to date, including the conclusions of the CMEP work (European Commission 2014 technical report) and the estrogen monitoring project. The activity presented in the ToR was in line with the Commission Communication on mixtures and with the objectives of the 7th Environment Action Programme.

The ToR were based on a series of specific objectives:

1. Identification of chemical modes of action (MoAs) (e.g. estrogenicity, Ah receptor binding, acetylcholinesterase inhibition, anti-cholinergic activity, photosynthetic inhibition, mutagenicity, immunotoxicity), considered to be of relevance in or via the aquatic environment for the protection of aquatic ecosystems and human health.
2. Perform an inventory of MoAs (if known) for currently regulated and/or monitored compounds (in particular priority and other WFD Annex X substances, watch-list (WL) substances, and RBSP identified to be of concern).
3. Based on 1 and 2, identification and prioritisation of EBMs (*in vivo* and *in vitro*) available for the detection of the relevant MoAs, in the different matrices of the aquatic environment. The prioritisation will consider the level of maturity of the methods, including whether they are available for routine use, and their robustness and reliability.
4. Development, where possible, of *in vivo and in vitro* effect-based trigger values, signaling a risk to or via the aquatic environment (including risks to human health from chronic exposure via consumption of drinking water or fishery products if possible), with the aim of making effect-based methods applicable (alongside chemical tools) in WFD chemical monitoring and assessment.
5. Based on objectives 3 and 4, selection of relevant EBMs (*in vitro* and *in vivo*) that can be used alongside chemical methods for the evaluation of complex mixtures occurring in the different types of aquatic environments (e.g. freshwaters, coastal waters), and aiming at being able to identify significant pressures and water bodies at elevated risk (i.e. support the WFD assessment of pressures and impacts). This will include consideration of the comparability of the results given by the different methods, and as far as possible the definition of quality control criteria for these tools in the context of the WFD, on the lines of the criteria defined by the QA/QC Directive.
6. Evaluation of ecological methods that can be used to address also chemical pollution, including metagenomics approaches.

7. Identification of a list of EBMs to be considered for Marine Strategy Framework Directive application according to D8 criterion 8.2.1 (of Decision 2010/477/EU) and/or considered within the WFD, taking also harmonisation between the WFD and MSFD into account.
8. Assess the availability and suitability of investigative approaches for identifying the underlying causes contributing to the overall risks, to identify sources of emissions and facilitate measures.
9. Assess the practical feasibility and cost effectiveness of implementing at EU-scale possible strategies using EBMs, to better take into account mixture risk assessment and mixture risk management under the WFD for relevant MoAs, as far as possible ensuring consistency with other legislation. In particular, this will include a comparison of the advantages/drawbacks of using effect-based tools alongside chemical tools, compared with using only chemical methods as in the current approach to chemicals under the WFD.

4.2. Meetings of the EBM activity

In total three meetings were organised (Rome, Prague, Ispra[16]) and every step of the activity was reported to WG Chemicals.

[16] WEBLINK to meeting folders

5. DELIVERABLES

Below, the main deliverables for objectives 1-8 are briefly described. For some of the objectives, more details are provided in annexes and cited literature.

The methods described in the report are categorised into three main groups and in line with the categorisation made in the Technical Report (European Commission 2014):

- Bioassays, *in vitro* and *in vivo*, which measure the toxicity of environmental samples under defined laboratory conditions, on cellular or individual (organism) levels, respectively;
- Biomarkers, i.e. biological responses at the cellular or individual (organism) levels, measured in field exposed organisms;
- Ecological indicators, measuring changes observed at higher biological organisation levels, i.e. the population and/or community.

Biomarkers are in turn often divided into those that are to be considered "effect biomarkers"[17] in the sense that the response (endpoint) typically can be linked to negative health effects, whereas some biomarkers are categorised as "exposure biomarkers"[18] in the sense that they are measuring the presence of a compound or its metabolites and interactions with receptors.

Some general pros and cons of these three main categories and subcategories are also described in the Technical Report (European Commission 2014).

Moreover, two publications on conclusions regarding Estrogen Monitoring (see Section 4.1) of European surface and waste water were provided in collaboration with WG Chemicals and as a follow up of CMEP and Science to Policy Initiative (SPI) activity (Kase et al. 2018 and Könemann et al. 2018) showing the feasibility of EBM in comparison with current chemical analytical methods.

[17] Imposex is for example considered to be an effect biomarker of very high ecological relevance since the effects observed are related to reproduction and measured on a high organisational level (tissue/organism). Extensive effects have been observed in the field related to population decline. Another effect biomarker that can be considered to be of very high ecological relevance is reproductive success in eelpout, because it is related to reproduction and measured at a high organisational level, and field effects have been observed in locally impacted areas.

[18] Metallothionein (MT) induction can on the other hand be considered to be an exposure biomarker of low/moderate ecological relevance, because it is involved in the regulation of the intracellular concentrations of essential and non-essential metals, and MTs provide protection against oxidative stress. Thus, if there is a response it is not straightforward to link it to a negative health impact.

5.1. Mode of Action (objectives 1 and 2)

Objectives 1 and 2 of the ToR

Objective 1: Identification of chemical modes of action (MoAs) (e.g. estrogenicity, Ah receptor binding, acetylcholinesterase inhibition, anti-cholinergic activity, photosynthetic inhibition, mutagenicity, immunotoxicity), considered to be of relevance in or via the aquatic environment for the protection of aquatic ecosystems and human health.

Objective 2: Perform an inventory of MoAs (if known) for currently regulated and/or monitored compounds (in particular priority and other WFD Annex X substances, WL substances, and RBSP identified to be of concern).

Objective 1 of the current activity was to identify MoAs that are of highest relevance in or via the aquatic environment, with respect to risk to the environment or human health. The PS were identified because of their relevance from the same perspective. Thus, the MoAs of these substances were to be investigated also through objective 2. A good understanding of the MoAs of the PS and other identified substances of WFD relevance is also crucial in trying to "group them" according to similar MoAs (see principle 1 in the Bratislava document cited above).

The term MoA refers within the Adverse Outcome Pathway (AOP) strategy to the specific mechanism by which the chemical compounds present in water produce their adverse effects on aquatic organisms. The MoA is the process initiated by the interaction of the toxicant with the organisms, for example with a receptor, which progresses through molecular, biochemical, physiological and/or anatomical changes in the organism to result in sub-lethal and lethal effects (Figure 2).

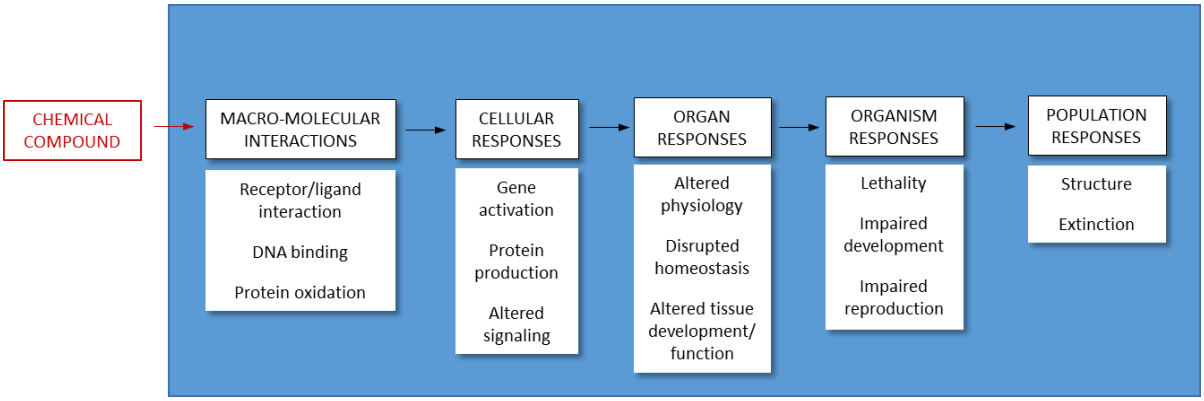

Figure 2. Schematic representation of the MoA, the process through which a chemical compound exerts its adverse effects. Adapted from OECD (http://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm)

In the aquatic environment, many substances from different sources co-occur as chemical mixtures. Even though most of them are present at very low concentrations, their combined action can cause adverse effects on the aquatic organisms (e.g. Carvalho et al 2014). The joint action of chemicals could result in a potentially unlimited number of additive, synergistic or antagonistic combinations. It is impossible to perform ecotoxicity tests to establish EQS for each potential mixture. Therefore, a robust approach for prospective environmental risk

assessment of chemical mixtures is needed. To gain greater insight into the risks posed by environmental contaminants and their mixtures it is beneficial to understand their MoA.

The WFD-specific measures for pollution control are based on the regulation of single substances but do not cover all the substances which are possibly relevant. To assess the chemical status of the water bodies the individual EQS are considered as safety limits, however the combined action of co-occurring compounds (chemical mixtures) is not taken into account. Chemicals can exert independent, additive, synergistic or antagonistic effects (Beyer et al. 2014). Additive and synergistic effects would lead to an increased toxicological effect. A better understanding of the MoA and potential interactions of chemicals is crucial for water quality assessments. According to the three EC Scientific Committees (SCHER, SCENIHR and SCCS)[19], a MoA is a plausible hypothesis about measurable key events by which a chemical exerts its biological effects. The MoA is already applied in computational models for the prediction of the toxicity of mixtures (Raies et al. 2016). Identification of the MoA can lead to an understanding of the molecular target (e.g. biological receptor) of a chemical and extrapolation to anticipated effects or biological responses. In this context, EBMs offer the possibility to monitor the overall response from multiple chemicals in environmental samples and estimate their impact on different levels of biological organisation. For this reason, they have been proposed to complement the chemical analytical methods to provide a more holistic approach to assessing chemical status.

The 2018 JRC technical report on MoA[20] provides an overview of the MoA of the PS in the WFD and other substances of concern (from the first WL and the current exercise to prioritise candidates for the PS list). The purpose of that report was to present an overview of the MoAs reported in ecotoxicological studies. In the report, the substances of interest are grouped into categories based on their chemical structure and common use, e.g. herbicides, PAHs, insecticides; as well as common MoA and toxicological endpoints, e.g. photosynthesis inhibition, endocrine disruption, oxidative stress. Furthermore, the available EBMs linked to the MoA are identified. However, it is not possible to identify single EBMs that account for all the relevant effects (including effects on different organisms) of each PS, alone or in combination. Furthermore, certain factors (e.g. toxicokinetics and toxicodynamics) other than the aqueous concentration may influence the toxicity of the substances, therefore even where an *in vitro* bioassay result might be expected to correlate with the results of field measurements, there may not be an exact correlation (see Section 5.2.3.).

[19] SCHER (Scientific Committee on Health and Environmental Risks), SCENIHR (Scientific Committee on Emerging and Newly Identified Health Risks) and SCCS (Scientific Committee on Consumer Safety). 2012. Toxicity and assessment of chemical mixtures.

[20] Napierska D et al. 2018. Modes of action of the current Priority Substances list under the Water Framework Directive and other substances of interest. JRC Technical Reports JRC110117. Office for official Publications of the European Communities.

Common MoA/effects identified in the JRC technical report on MoA:

- Photosynthesis inhibition
- Endocrine disruption
- Oxidative stress
- Activation of metabolising/detoxifying pathways
- Genotoxicity
- Histopathology
- Stress proteins
- Unique pathway toxicity (e.g. acetylcholiesterase inhibition, imposex, presence of metallothioneins)

To predict the toxicity of a chemical mixture, data on the MoA of each component of the chemical mixture is required. However, for some classes of chemicals, such as the neonicotinoid and pyrethroid insecticides, whose MoA is well-characterised in their target organisms, there is limited information regarding the mechanism that causes toxicity in non-target organisms including aquatic species.

Therefore, the choice of EBMs to detect (specifically) the presence of those substances in the monitored water remains a challenge and further investigation is needed to elucidate the mechanisms behind the toxicity of these compounds.

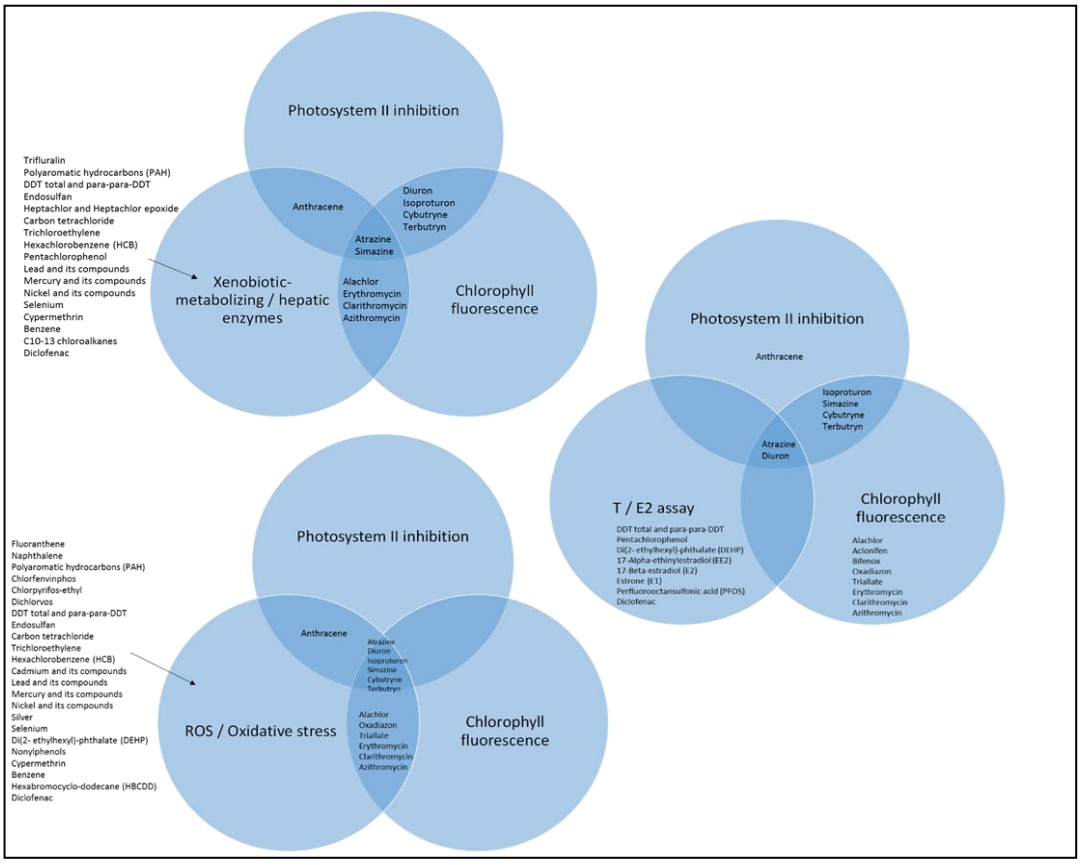

Figure 3. Venn diagrams representing common MoA and endpoints of PS and other substances of interest. Adapted from the JRC technical report on MoA[21].

[21] Napierska D et al. 2018. Modes of action of the current Priority Substances list under the Water Framework Directive and other substances of interest. JRC Technical Reports JRC110117. Office for official Publications of the European Communities.

JRC technical report on MoA - conclusions

- PS grouped by their common MoA/effect
- MoA linked to available EBMs
- Further investigation needed to understand the MoA of some groups of chemicals (e.g. neonicotinoids, pyrethroids)
- Chemicals acting through the same MoA can exert additive effects
- EBMs suitable for monitoring mixture toxicity
- A battery of EBMs is proposed to reduce chemical analysis in water quality assessment
- Standardisation and interlaboratory trial needed before EBM implementation in WFD

A battery of MoA-based assays is proposed in the JRC technical report to assess the chemical status of water environments more holistically (rather than with a limited but ever-growing list of individual EQS), and to try to overcome analytical difficulties and reduce monitoring costs. For this purpose, a more systematic approach should be developed in order to define which panel of assays might be of greatest use for the specific circumstances (e.g. for the combination of substances that might be found). Furthermore, an interlaboratory exercise for harmonisation and validation will be required to ensure comparability among bioassays focused on the same MoA.

5.2. Inventory and selection of EBMs (objective 3, 5 and 7)

To assess which EBMs are now available and sufficiently developed to be applied on a more regular basis, an inventory of such methods was compiled, focusing on available biomarkers, *in vitro* and *in vivo* assays.

Furthermore, for each method, three main aspects were investigated:

1. WFD/MSFD relevance
2. Availability of Standard Operating Procedures (SOPs)
3. Possibilities to evaluate the data (availability of assessment criteria, further described also in Section 5.3.)

Objective 3, 5 and 7 of the ToR

Objective 3: Based on 1 and 2, identification and prioritisation of effect-based methods (*in vivo* and *in vitro*) available for the detection of the relevant MoAs, in the different matrices of the aquatic environment. The prioritisation will consider the level of maturity of the methods, including whether they are available for routine use, and their robustness and reliability.

Objective 5: Based on objectives 3 and 4, selection of relevant effect-based methods (*in vitro* and *in vivo*) that can be used alongside chemical methods for the evaluation of complex mixtures occurring in the different types of aquatic environments (e.g. freshwaters, coastal waters), and aiming at being able to identify significant pressures and water bodies at elevated risk (i.e. support the WFD assessment of pressures and impacts). This will include consideration of the comparability of the results given by the different methods, and as far as possible the definition of quality control criteria for these tools in the context of the WFD, on the lines of the criteria defined by the QA/QC Directive.

Objective 7: Identification of a list of effect-based methods to be considered for Marine Strategy Framework Directive application according to D8 criterion 8.2.1 (of Decision 2010/477/EU) and/or considered within the WFD, taking also harmonisation between the WFD and MSFD into account.

Information on several EBMs was collected when establishing fact sheets in the Technical Report (European Commission 2014). Within this task, additional information on those EBMs was collected and participants were invited to add other methods into an Excel-sheet and also check whether the information already included was correct. EBMs that would be expected to respond to specific compounds or compounds with a common MoA, several MoAs or unknown MoAs and cumulative stress from several stressors, not only toxic substances, were included in the inventory.

Although emphasis has been made to include "as many EBMs as possible" at an initial stage, it should be pointed out that the inventory should not be considered a comprehensive list of available EBMs. Furthermore, not all methods included (in the Annex II to this report) are to be considered "recommended". The inventory should instead be viewed as the "base set" of EBMs that have been considered within the activity. It should also be pointed out that for several of the EBMs in the inventory, the necessary information to fully assess their robustness was not available or found – probably in part due to the time constraints on the activity. In addition, whether a certain EBM can be considered "robust enough", will most likely depend on the intended use (see "WFD applications" below).

MoAs that are covered by the EBMs in the inventory

In total 138 EBMs were finally included, of which 57 could be categorised as *in vitro* assays, 37 as *in vivo* assays and 34 as biomarkers. The inventory collected so far does not claim to be complete and would have to be further developed.

Objective 3 of the ToR is tightly linked with objectives 1 and 2. A summary of EBM availability according to the MoA of each PS is found in Annex I to this report, and a detailed list (inventory) of the EBMs available is presented in Annex II, taking account in particular of their

relevance to the content of Annex I. However, please note that the EBMs in the inventory (numbered list in Annex II) might cover also additional MoAs or be categorised somewhat differently from the MoAs identified in Annex I, and that the examples of substance (groups) the biomarkers can cover, mentioned in Annex II table II.1, is not exhaustive.

For the EBMs in the inventory, the "endpoint" is included in the numbered inventory lists in Annex II to this report. It should be noted that in a toxicological assessment, an endpoint is meant as an observed or measured outcome to indicate or reflect the effect of contaminants on organisms. There is therefore a strong link between the endpoint used and the MoA examined. However different MoAs can result in a common adverse outcome, particularly if it concerns a general endpoint, such as lethality or growth, that could be the result of substances with different MoAs acting together. Furthermore, some of the EBMs included (in particular, *in vivo* bioassays and some general biomarkers) are able to detect general effects from complex mixtures.

Together the EBMs collected so far cover the following MoAs and type of effects:

- Endocrine disruption of sex hormones (of relevance for e.g. reproduction):
  - Activation and antagonistic activity of the estrogen receptor (ER) *in vitro*
  - Neurosteroids *in vivo*
  - Vitellogenin induction (*in vivo* and as biomarker)
  - Spiggin induction (as biomarker)
  - Activation and antagonistic activity of androgen receptor (AR) *in vitro*
  - Activation and antagonistic activity of progestogenic receptor (PR) *in vitro*
  - Imposex (tissue level, as biomarker)
  - Intersex (tissue level, as biomarker)
- Endocrine disruption of glucocorticoids (of relevance to e.g. development, metabolism, immune system):
  - Activation and antagonistic activity of the glucocorticoid receptor (GR)
- Endocrine disruption of thyroid hormones (of relevance to development, growth, and metabolism of all vertebrates, major role in neurogenesis and brain function)
  - Binding assay to thyroid receptor (TR)
  - Activation and antagonistic activity of the thyroid receptor (TR)
- Genotoxicity and mutagenicity
  - DNA strand breaks (*in vitro*)
  - Reporter gene expression (+S9) (*in vitro*)
  - Mutagenicity (point mutation, clastogenic effect)
  - DNA damage (Comet assay) (*in vivo* at early life stage and as biomarker)
  - Gene transcriptions
- Immune response
  - KappaB (*in vitro*)
  - Fish disease (biomarker)
- Activation of metabolic enzymes
  - Activation of the peroxisome proliferator-activated receptor (PPARγ) (*in vitro*)
  - Activation of human pregnane x receptor (PXR) (*in vitro*)
- Oxidative stress
  - Reactive oxygen species (ROS, *in vitro*)
  - Stress proteins (biomarker)

  - Protein carbonylation (biomarker)
  - Gene transcriptions
- Internal regulation
  - Metallothionein (MT) induction (biomarker)
  - Ah receptor activation (of relevance to e.g. detoxification) (as *in vitro* and *in vivo* and biomarker - EROD)
  - PAH metabolites (biomarker)
  - Gene transcriptions (biomarker)
  - P-glycoprotein efflux (P-gp) (biomarker)
- Hemoglobin synthesis
  - Delta-aminolevulinic acid dehydratase (ALA-D) (biomarker)
- Lysosomal membrane stability (biomarker)
- Inhibition of photosynthesis
  - PSII-inhibition (algae, higher plants) (*in vitro*/*in vivo*)
- Neurotoxicity
  - Acetylcholinesterase (AChE) inhibition (overstimulation of neuromuscular junctions) (*in vivo* and as biomarker)
- Cytotoxicity (cell death)
  - In fish cell lines (*in vitro*)
  - In algae (inhibition of photosynthesis and loss in biomass/growth, *in vivo* but single cell organisms)
  - In bacteria (inhibition of bioluminescense, *in vivo* but single cell organisms)
  - Lipid peroxidation (biomarker)
- Embryotoxicity (*in vivo*)
- Spermiotoxicity (*in vivo*)
- Development (*in vivo*)
  - Molting
  - Growth
  - Larval development
- Histopathological changes
  - Fish Liver histopathology (LH) and liver macroscopic neoplasms (MLN) (biomarkers)
  - Mussels (gametogenesis, digestive gland and tube, biomarkers)
- Malformation (*in vivo*)
  - Embryo of amphipods, fish (*in vivo* and biomarkers)
  - Benthic diatoms (biomarker)
  - Mentum deformations in chironomids (biomarker)
- Behaviour (*in vivo*)
  - Immobilisation
  - Swimming behaviour
  - Photomotor response
  - Feeding inhibition
- Reproduction (*in vivo*)

- Invertebrates
- Fish (also in viviparous organism, eelpout, as biomarker)
- Pregnancy rate in marine mammals (biomarker/ecological level)
- Egg shell thinning in predatory birds (biomarker)

- Lethality
  - *In vivo* assays on several trophic levels such as fish (early life stage), invertebrates (also benthic) and aquatic plants
  - Biomarker in mussels (aerial survival)
  - Survival of offspring (mammals and predatory birds, biomarker/ecological level)

WFD and MSFD applications

Besides collecting additional information about the individual methods available, another starting point to the selection process was to identify different WFD- (and MSFD-) relevant "useful applications" of EBMs and explore whether EBMs are available today to fill identified needs. However, to be able to conclude on this, also the WFD relevant "needs" had to be identified.

The most obvious and important use of EBMs, already mentioned in the introduction, would be to cover also other substances that are today not monitored or assessed and to take mixture effects into account (see also introduction). This "coverage" could refer both to the assessment of "(toxicological) status" (further discussed in Chapter 6) but also to identify water bodies that are subject to significant pressures (see Section 5.5.).

The need to 1) assess effects from complex mixtures (of unknown composition) and perhaps even cumulative effects when combined with other stress factors can be distinguished from the need to 2) specifically take mixture effects from substances sharing the same MoA (such as estrogenicity) into account.

EBMs that 3) can be specifically used under the MSFD as indicators for e.g. D8, and those 4) available to assess sediment quality, should also be highlighted.

Applications 1-4 are the main WFD related purposes investigated for individual EBMs. However, it was also assessed whether there would be cases where a particular EBM 5) could be used as a "bioanalytical method", comparable to chemical analytical methods, to assess the status of already regulated compounds, in particular at a screening level. Other potentially useful applications were identified and are discussed below.

One aspect that was investigated was whether EBMs exist that 6) could be used to assess metal bioavailability in cases where BLM modeling is problematic due to e.g. highly deviating water chemistry compared to the validation ranges of the BLM models. EBMs were specifically mentioned in a document developed to facilitate the implementation of BLMs in cases where conditions are outside the applicability range of the BLMs and user-friendly tools (WCA 2014). Please note also that bioavailability models (BLMs) are not available for all chemicals in all environments. In such cases, the use of EBMs may provide a better assessment of environmental status.

Another potential use that was discussed in the activity was whetherEBMs exist that 7) could be used to assess status where metal EQSs cannot be readily used because of high natural background concentrations (background>EQS), situations that could occur in mineralised

areas. To assess whether EBMs exist that could be used for application 6 or 7, a separate investigation was performed by Brix (2018)[22].

Finally, the usefulness of particular methods or groups of methods to 8) assess the quality of drinking water and to 9) assess the quality of effluents or leachates is briefly described.

Below, the identified "available" EBMs are further described and their "fitness for purpose" (for use in any of the applications 1-9) is assessed. For more or less all of the above applications one aspect to consider in the further assessment of whether a particular EBM is fit for purpose is whether it can be used to assess effects relevant in a WFD and/or MSFD context. Their level of maturity (based on an assessment of availability of routine use, robustness, reliability – see ToR objective 3) was also considered. This assessment was performed for the three main EBM groups respectively (see sections on *in vitro* assays, *in vivo* assays and biomarkers below). For ecological indicators, see objective 6.

In **Annex V**, an example of a battery of *in vivo* and *in vitro* bioassays, according to the results of the European Union Framework Programme Project SOLUTIONS, and the Norman Network activity, is described.

Applications investigated for individual EBMs within the three groups (*in vitro*, *in vivo* and biomarkers)

1. Cover complex mixtures (of unknown composition) and perhaps even cumulative effects when combined with other stress factors – to assess status and/or identify significant pressures
2. Cover mixture effects from substances sharing the same MoA– to assess status and/or identify significant pressures
3. Identify relevant MSFD indicators
4. Assess sediment quality
5. Bioanalytical methods to assess status of regulated substances
6. Assess metal bioavailability when water chemistry outside validation range
7. Assess status where high natural metal concentrations (>EQS)
8. Assess quality of drinking water
9. Assess quality of effluents or leachates

Standard Operating Procedures and Performance criteria for EBMs

For the chemical approach, clear quality control mechanisms are in place. The QA/QC Directive (2009/90/EC)[23] states that e.g. all methods, "used for the purposes of chemical monitoring programmes carried out under Directive 2000/60/EC are validated and documented in accordance with EN ISO/IEC-17025 standard or other equivalent standards accepted at international level" (Art 3).

For EBMs the applicability of such a requirement is also justified. Several individual EBMs, in particular the *in vivo* assays (largely stemming from protocols developed in chemicals testing)

[22] The following effects and corresponding EBMs were investigated: ion homeostatis, oxidative stress, lysosomal stability, DNA damage, deformities (in chironomids, diatoms and amphibians), *in vivo* assays (algae, invertebrates, fish), Cytochrome P450, AChE, Urease, bacterial reporter assay, ALAD, MT, eDNA barcoding.

[23] Commission Directive 2009/90/EC of 31 July 2009 laying down, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, technical specifications for chemical analysis and monitoring of water status.

but also some *in vitro* assays, are indeed rapidly advancing in this context. For biomarkers however, such SOPs may be developed in another framework than the regular international standardisation context, see section 5.2.1. below.

Furthermore, the EBM performing laboratories should participate in proficiency testing programmes, see also Art 6 of the QA/QC directive (see text box below). This is further discussed in the following sections.

> Art. 6 of the QA/QC Directive
>
> 1. Member States shall ensure that laboratories or parties contracted by laboratories apply quality management system practices in accordance with EN ISO/IEC-17025 or other equivalent standards accepted at international level.
>
> 2. Member States shall ensure that laboratories or parties contracted by laboratories demonstrate their competences in analysing relevant physico-chemical or chemical measurands by: (a) participation in proficiency testing programmes covering the methods of analysis referred to in Article 3 of this Directive of measurands at levels of concentrations that are representative of chemical monitoring programmes carried out under Directive 2000/60/EC, and (b) analysis of available reference materials that are representative of collected samples which contain appropriate levels of concentrations in relation to relevant environmental quality standards referred to in Article 4(1).
>
> 3. The proficiency testing programmes referred to in paragraph 2(a) shall be organised by accredited organisations or internationally or nationally recognised organisations which meet the requirements of ISO/IEC guide 43-1 or of other equivalent standards accepted at international level. The results of participation in those programmes shall be evaluated on the basis of the scoring systems set out in ISO/IEC guide 43-1 or in the ISO-13528 standard or in other equivalent standards accepted at international level.

The QA/QC directive (Art 4.1.) also states that "Member States shall ensure that the minimum performance criteria for all methods of analysis applied are based on an uncertainty of measurement of 50 % or below (k=2) estimated at the level of relevant environmental quality standards and a limit of quantification equal or below a value of 30 % of the relevant environmental quality standards".

Such a requirement would only be possible to apply and evaluate for those EBMs for which assessment criteria are in place, corresponding to the "environmental quality standards" mentioned (further discussed in Section 5.3.). However, Art 4.2. states also that "In the absence of relevant environmental quality standard for a given parameter, or in the absence of method of analysis meeting the minimum performance criteria set out in paragraph 1, Member States shall ensure that monitoring is carried out using best available techniques not entailing excessive costs." Thus, for those EBMs where assessment criteria are not in place, it would be justified to use at least the best available techniques, taking costs into account.

5.2.1. Biomarker inventory

Applications investigated

Biomarkers have been applied for a long time in regular monitoring programmes, especially within the marine environment. Such programmes have had different purposes but one major reason is for them to act as early warning signals and to detect effects from complex mixtures and non-monitored substances. Thus, for biomarkers the primary applications investigated for individual methods are the first three applications – to assess unknown substances and to take mixtures into account, and to consider relevance in the MSFD context. However, also sediment quality could be relevant, for biomarkers analysing effects occurring in organisms

exposed to sediment. There are also examples of biomarkers that are more or less substance specific. Thus, for the individual biomarkers in the inventory, applications 1-7 were investigated, see box below.

Applications investigated for individual biomarkers in the inventory

1. Cover complex mixtures (of unknown composition) and perhaps even cumulative effects when combined with other stress factors – to assess status and/or identify significant pressure
2. Cover mixture effects from substances sharing the same MoA– to assess status and/or identify significant pressure
3. Identify relevant MSFD indicators
4. Assess sediment quality
5. Bioanalytical methods to assess status of regulated substances
6. Assess metal bioavailability when water chemistry outside validation range
7. Assess status where high natural metal concentrations (>EQS)

Since drinking water investigations (application 8) are exclusively related to the protection of human health, biomarkers are not suitable because they monitor responses in field-collected organisms that are not only exposed differently but which might also have different receptors etc.

Biomarkers, being analysed on field-collected organisms, can also not easily be used to assess effluents or leachates (application 9). Nevertheless, if triggered by a particular biomarker response the corresponding "*in vitro*" or "*in vivo*" method could be used to evaluate effluents (further described in Section 5.4).

Is the biomarker analysig "WFD-relevant effects"?

Overall, most biomarkers included in the inventory can in one way or another be considered to be of WFD relevance but some can be interpreted in a more "stand-alone" manner and some even in absolute terms (if assessment criteria are available today or in the near future). The identification of such biomarkers is the main focus of this section. Other biomarkers will most likely be easier to interpret in a weight–of-evidence manner. This will be further developed in Section 5.3.

To assess the WFD relevance of individual biomarkers, two different approaches were chosen. The first approach was to assess whether the response in itself can be linked to adverse health impacts on the organism. The ecological relevance of each of the biomarkers included in the inventory was therefore assessed based on exposure, type of effect measured and level of biological organisation (subcellular-cellular-tissue-organism levels), see last column of table II.1.

As was previously described, the EQSs for individual substances are based on an assessment of concentrations that cannot be exceeded to achieve different "protection goals". Two of these are the protection of pelagic organisms (fish, algae, aquatic plants) and benthic organisms respectively. These two protection objectives were considered the most relevant to assess for biomarkers.

In the calculation of EQSs, data from laboratory toxicity tests are normally used. In CIS Guidance Document No. 27 there is an indicative list of endpoints that could be considered in the derivation:

- growth (weight, length, growth rate, biomass)
- number (cells, population)
- mortality
- immobilisation
- reproduction
- hatching (rate, time, percentage)
- sex ratio
- development (egg, embryo, life stage)
- malformations (teratogenicity)
- proliferation (cells)
- filtration rate
- carbon uptake (algae)
- reburial (of e.g. certain crustacean species).

This list can serve also as an indication of the type of endpoints that can be considered highly relevant for biomarkers.

Several biomarkers are used to monitor effects that are of high ecological relevance. Such biomarkers could therefore be considered valuable to assess "status". Most of these methods are "general biomarkers". Being "general" means that they cannot be used to directly link the observed effects to a particular substance or sometimes even a group of substances sharing a common MoA. However, this can also be considered their strength, in the sense that they can be used to assess effects from many interacting substances, and cumulative effects, in some cases the result of other types of stress.

Other biomarkers are used to monitor relevant effects but at lower levels of biological organisation (subcellular levels) than those that are normally used as endpoints in the calculation of EQSs. Such endpoints include e.g. blood or plasma protein levels, histopathological endpoints, organ weights (e.g. hepatosomatic index, gonadosomatic index), mRNA induction. For such biomarkers, the link to negative health impacts can sometimes be more difficult to assess. However, if there is a correlation or causal relationship with population sustainability established also these endpoints could be of relevance. Furthermore, exposure biomarkers can often have higher sensitivity and, as a consequence, more subtle effects (early warning levels) can be detected.

In the assessment of relevance of the biomarkers in the inventory, as long as the MoA was considered important and the response likely linked to adverse impacts on health (at least on a tissue level) the biomarker was ranked as being of "moderate ecological relevance".

In a few cases biomarkers that can be used to assess secondary poisoning were also identified. This is clearly also considered relevant from a WFD perspective, since protection against secondary poisoning is also considered in the derivation of EQS for accumulating substances. However, the effects are studied in birds and mammals rather than pelagic organisms (fish) and these types of "biota" are not sampled in the WFD context. Also, the geographical scale of the assessment would be difficult to make at such fine resolution as an individual water body. These biomarkers are therefore most likely of MSFD rather than WFD relevance. Furthermore, one would expect the chemical approach to assess secondary poisoning under the WFD to provide warning of risk at an earlier stage than would biomarkers reflecting secondary poisoning.

Is the biomarker analysing WFD-relevant substances?

The other approach to assess the WFD relevance is whether the biomarker is likely to respond to substances or substance groups that are already considered to be of WFD/MSFD relevance. Annex VIII of the WFD lists several classes of compounds that should in particular be considered in the WFD context. On line four, the following group of substances is specifically mentioned: "Substances and preparations, or the breakdown products of such, which have been proved to possess carcinogenic or mutagenic properties or properties which may affect steroidogenic, thyroid, reproduction or other endocrine- related functions in or via the aquatic environment."

Some biomarkers can be considered "specific" in the sense that they primarily respond to a particular MoA and can thus be used to monitor some of the above substances – but as the combined response to all substances in the mixture. Such biomarkers are often called specific. They can either be linked to one or a few individual substances (such as TBT in the case of imposex) or a particular MoA (such as those of mutagenic or estrogenic substances). The specificity of the biomarkers in the inventory is described in Table II.1.

Regulatory implementation aspects for biomarkers

Table II.3. lists, for the biomarkers included in the inventory, available information on costs for analysis, availability of commercial laboratories performing the tests and whether the biomarker has been already included in regular monitoring programmes, and whether there are established assessment criteria and SOPs (such as international standards but also guidance documents or frequently used scientific publications).

The main robustness check was to investigate whether SOPs are available. One type of SOP available for biomarkers are publications in the "ICES Techniques in Marine Environmental Sciences (TIMES)" series[24]. These documents provide details on methods and procedures relating to chemical and biological measurements in the marine environment. Most of the techniques described have been selected on the basis of performance in ICES or other international intercalibration exercises.

Another important aspect that needs to be considered is whether the results can easily be evaluated. For chemical status assessments, analysed concentrations are compared to EQSs. The procedure to derive the EQSs is described in detail in CIS Guidance Document No. 27. For biomarkers, no such strict procedures, applicable to all biomarkers, have to our knowledge been developed, and for some biomarkers another approach than setting "fixed values" might be more appropriate. This is discussed further in Section 5.3. However, in Table II.3. any known established assessment criteria are included.

Please be aware that the information in Table II.2. and II.3. is not always complete or necessarily relevant to every MS. If the information indicates that commercial laboratories are available, this means that at least one commercial provider is available in at least one MS. Costs are only roughly estimated according to the following categorisation: low <200 Euro; moderate: 200-500; high 500-1000 Euro per sample/assessment.

Costs related to sampling are normally not included but only the costs of the actual analyses. Nevertheless, it should be noted that the costs for analyses can vary between laboratories. In

[24] http://www.ices.dk/publications/our-publications/Pages/-ICES-Techniques-in-Marine-Environmental-Sciences-.aspx

some cases, commercial providers are already available (also indicated), whereas some EBMs have so far been implemented primarily by research institutions.

Information about whether a particular EBM is already included in a regular monitoring programme also indicates the availability of laboratories (also other than commercial) able to perform the analysis and of expertise to aid in further interpretation etc. Information indicating which biomarkers are already used for such regular monitoring is also provided in table II.3.

Practical and strategic aspects

Since biota are being sampled and investigated, a major advantage of using a selected set of biomarkers alongside chemical and biological monitoring is that it is possible to establish an integrated monitoring approach, in the sense that the same samples can be used to assess:

- Concentrations of contaminants in the tissues
  - To assess status
  - To assess trends
- Effects on individual and suborganism levels
- Effects on the population level (such as fish catch)[25]

OSPAR/ICES developed a guidance document on this topic and this is further described in the technical report (European Commission 2014).

An integrated approach thus has many advantages. A large part of the costs involved in analysing biota is related to the sampling. By combining traditional biota monitoring with EBM analyses and monitoring to assess population level effects, a cost-effective monitoring approach can be used in the sense that the sampling frequency can be lower (sampling is done for several purposes at the same time). Also, the data interpretation can be facilitated and based on an integrated approach. Minimising the sampling of organisms (vertebrates) is also positive from an animal welfare perspective.

However, there are some prerequisites and aspects to be particularly aware of in the planning stage of an integrated monitoring approach. First, the amount of sampled material/number of individuals needs to be sufficient for all the analyses to be performed. Whenever this information was available, the amount of sample needed for a particular EBM is included in the summary table on biomarkers in Annex II to the report. Another practical aspect to consider is whether sampling, depending on the scope, should be performed at a particular time of the year. This is related to the variability of the parameter, such as seasonal patterns (related to e.g. reproduction season). Such aspects are also included in the summary table II.2.

For some EBMs, it is possible to store samples for later analysis. This greatly facilitates an integrated approach, in the sense that the necessary expertise and equipment is not required at the point of sampling. As with chemical analyses, some sample preparation might be needed, but the main analysis may still be performed upon arrival at the laboratory or even after longer storage. This information is also included in Table II.2.

Whereas practical aspects, such as amount needed, storage possibilities and seasonal aspects for sampling biota to analyse biomarkers in the inventory are tabulated in Annex II these aspects are not taken into account in the next step – identifying which methods are robust

[25] Please note that we here do not necessarily refer to current BQEs but rather investigations of general species composition and abundance.

enough to be considered being used on a regular basis in a regulatory (WFD and/or MSFD) context. However, such aspects can also have implications for the potential to limit the sampling efforts and indirectly the costs.

Biomarker prioritisation for different WFD/MSFD applications

No strict "evaluation criteria" other than the availability of SOPs and the relevance of the biomarker were used to "prioritise" (ToR objective 3 and 5) biomarkers for further selection.

The individual biomarkers that were considered to fulfill both the "relevance and robustness checks" are further described under Section 5.3. and proposed to be considered under the WFD and/or MSFD umbrellas to also take effects from otherwise non-monitored substances and mixtures into account (applications 1-3). Several if not all biomarkers can be used to identify water bodies that are subject to significant pressures. This will be further discussed in Section 5.5. Some can also be used to assess "toxicological status" (see further discussions in Chapter 7).

From the inventory list it became obvious that not all biomarkers would be readily applicable to both marine and limnic environments and for effect biomarkers, marine biomarkers dominated. A distinction between biomarkers that should be prioritised for marine use (WFD coastal water bodies and MSFD) and those that should be prioritised for limnic use (river and/or lake water bodies within the WFD) is therefore made in Chapter 6.

From these biomarkers, primarily biomarkers studied on gastropods and mussels, but perhaps also some fish species, are anticipated to monitor effects in organisms also being exposed through the sediment (application 4). Thus, the usefulness of a particular biomarker for this application depends on which species is sampled rather than the biomarker analysis itself.

Biomarkers, being analysed on field-collected organisms, are generally not possible to use as bioanalytical methods (application 5) in screening environmental samples. Furthermore, the fact that field-collected organisms are analysed implies that it is not always possible to control for environmental factors (including other substances) influencing the results. They can therefore normally not replace chemical analyses of individual, regulated compounds. However, a few biomarkers are exceptionally specific and – if environmental factors can be excluded as having an impact and/or be taken into account in the evaluation, such data can be considered alongside chemical analytical data of the particular compound. If such biomarkers show unacceptable effects, this should be sufficient evidence to conclude that status is not good, even if chemical concentrations are below the EQS. Only two biomarkers (imposex and egg-shell thinning) could be identified to fulfill these criteria. And could thus be used alongside chemical analysis to assess status (in relation to TBT and DDT) under the MSFD and/or WFD.

As previously mentioned, EBMs could be used to deal with status assessments of metals in water bodies where conditions are outside the applicability range of the BLMs and user-friendly tools (WCA 2014). The types of EBMs mentioned were ecotoxicity tests, bioassays, certain biomarkers and ecological community monitoring. By using these, it was suggested that ecological assemblage specific EQSs or site-specific PNECs from field data might be derived. To assess whether EBMs exist that could either be used for application 6 or 7, a separate investigation was performed by Brix (2018). Most of the EBMs investigated were biomarkers, and although most are already included in Annex II, some additional biomarkers were also being evaluated (ion homeostatis and urease). Three of these biomarkers were considered to analyse effects for which a strong link to ecologically relevant effects had been demonstrated (ion homeostatis – strong link to survival and growth; deformities – clear links between observed deformities and effects on individuals and populations; AChE – strong

correlation between AChE inhibition and acute effects/survival). For Lysosomal Membrane Stability (LMS) and urease the relevance was assessed to be moderate (for LMS because links to organ-level effects have been shown but not yet documented at individual or population levels; for urease because of effects on nitrogen metabolism and inferred effects at individual/population level). For the other biomarkers included, the author concluded that links to ecologically relevant effects are not (yet) demonstrated.

Nevertheless, the author also concluded that none of the biomarkers are BOTH responding specifically to metals (or one particular metal) AND analysing effects for which there is a strong link to effects at high organisational levels. This is in line with the findings above related to application 5 (to use biomarkers as bioanalytical methods). Imposex and egg-shell thinning are the only biomarkers found in the inventory that would fulfill such requirements[26]. Thus, there are at the moment no biomarkers that could be used instead of metal analyses (and bioavailability models) to assess "metal status".

The sensitivity of the different EBMs was also evaluated by Brix (2018), by investigating at which concentrations a biomarker reponse is triggered, and comparing this concentration with the EQSs for individual metals. An EBM that is very sensitive and specific to a certain metal could potentially be valuable in cases where BLMs can't be used or give less reliable results (application 6) due to e.g. water chemistry being far outside the validation range. One would assume that if a sensitive biomarker (responding to concentrations significantly lower than the EQS) doesn't respond, it would indicate that the metal is not sufficiently bioavailable to be cause for concern. However, the report cannot identify any biomarkers sufficiently sensitive to metals to be used in this way. The biomarkers investigated at best respond at the EQSs but not at significantly lower concentrations. ALA-D inhibition occurs near the EQS of lead (Pb) for example and MT responds also at the EQS for some metals but only at concentrations much higher than the EQS for other metals. Therefore, it can be concluded that none of the biomarkers in the inventory would be useful to assess metal bioavailability in cases where e.g. the water chemistry is outside the range of the BLM validation range. For similar reasons, it was concluded that none of the biomarkers investigated could be used to assess metal status in water bodies exposed to high natural background concentrations (above the EQS). These would obviously also be the conclusion for water bodies that are within the validation range. Nevertheless, it should be pointed out that if a response is observed in any of the biomarkers above, metals could indeed be involved, but also other substances or in some cases other stress factors.

For the above reasons, NO biomarkers have been selected (or proposed in Chapter 6) to be used to assess bioavailability of metals and/or toxic stress in areas with high levels of natural background concentrations. However, it could not be ruled out that other, more general and very sensitive variables such as red and white blood cells (biomarkers not included in the inventory or the assessment made by Brix 2018) could be of value on a case-by-case basis in this context.

If sufficiently sensitive biomarkers were available for a particular metal, in situations where the EQS is exceeded, but the bioavailability is uncertain, a "no-response" in such a sensitive biomarker would suggest that status is, after all, "good".

As stated previously, when it comes to biomarkers, focus has been put on identifying EBMs that could be used to assess effects that are related to the protection of pelagic or benthic organisms and to some extent wildlife vulnerable to secondary poisoning. Human health protection cannot be achieved through the use of biomarkers. EROD can e.g. be expected to

[26] Although TBT contains metal (Sn) it is usually not considered in this context, being an "organometal".

respond to dioxins. However, the main driver for the EQS of dioxins and dioxin-like PCBs is human health (and possibly secondary poisoning). EROD as a biomarker cannot be used to assess this risk. Nevertheless, EROD could also respond to PAHs, and for some of the EQSs developed for PAHs the main driver is toxicity to aquatic organisms. Although EROD probably cannot be used as a "stand-alone" biomarker, the biomarker EROD is useful in a weight-of-evidence and trend approach as an early warning signal (see example in Section 5.3.5.).

Thus, biomarkers were found to primarily be useful for the first five applications.

Table 3 below lists identified effect biomarkers that monitor negative health effects at least on tissue level (effects considered of moderate or higher ecological relevance) along with condensed information important to the proposal in Chapter 6. In Chapter 6, the prioritised individual biomarkers that were found to be suitable for a particular WFD/MSFD application are listed, and in Annex II more detailed descriptions of these methods are included.[27]

This is not to say that other biomarkers such as exposure biomarkers could not be of value, but they would be less straightforward to evaluate one by one. In Section 5.3., other approaches to evaluate analytical results using broader batteries of biomarkers are also described.

[27] Please note, for example, that if the table 3 suggests that "assessment criteria" are available, this can refer to both Environmental Assessment Criteria (EACs) and Background Assessment Criteria (BACs), or other values, established only at national level (see also Section 5.3. on assessment criteria). SOPs usually refer to TIMES protocols but for some individual methods only other documents are available – such as common reference documents used when the method is applied or adopted by a Member State in the MSFD context.

Table 3. Criteria for the selection process of biomarkers. The last column lists identified relevant applications. For full explanation about the different applications assessed, see text box in Section 5.2.1. If monitoring is suggested, brackets are added where it has only been performed in campaigns. For some biomarkers, information about this is still missing (marked with a question mark).

Biomarker name	**Ecological Relevance**	**Responds to**	**SOP available?**	**Assessment criteria available?**	**Monitored?**	**Marine?**	**Limnic?**	**Protection goal**	**Applications of relevance**
Imposex	VERY	TBT	Yes	Yes	Yes	Yes	No	Pelagic Benthic	3 (MSFD) 4 (sediment) 5 (regulated substances)
LMS (lysosomal membrane stability)	MODERATE	Complex mixtures and other stressors	Yes	Yes	Yes	Yes	Yes	Pelagic and/or Benthic (depends on species)	1 (complex mixtures) 3 (MSFD) 4 (sediment)
ALA-D (delta-aminolevulinic acid dehydratase)	MODERATE	Lead	Yes	**?**	?	Yes	Yes	Pelagic and/or Benthic (depends on species)	3 (MSFD) 4 (sediment) 5 (regulated substances)
DNA adducts	MODERATE-HIGH	Mixtures of compounds with common MoA (mutagenicity)	Yes	Yes	?	Yes	Yes	Pelagic and/or Benthic (depends on species)	2 (MoA mix) 3 (MSFD) 4 (sediment)
FDI (Fish Disease Index) including LH (liver histopathology) and	HIGH	Complex mixtures and other stressors	Yes	Yes	Yes	Yes	Yes	Pelagic and/or Benthic	1 (complex mixtures) 3 (MSFD)

Biomarker name	Ecological Relevance	Responds to	SOP available?	Assessment criteria available?	Monitored?	Marine?	Limnic?	Protection goal	Applications of relevance
MLN (macroscopic liver neoplasms)								(depends on species)	4 (sediment)
Reproductive success in eelpout	VERY	Complex mixtures and other stressors	Yes	Yes	Yes	Yes	No	Pelagic (but bottom dwellers)	1 (complex mixtures) 3 (MSFD) 4 (sediment)
VTG (vitellogenin) in male fish	MODERATE-HIGH	Mixtures of compounds with common MoA (estrogenicity)	Yes	Yes	yes	yes	yes	Pelagic and/or Benthic (depends on species)	2 (MoA mix) 3 (MSFD) 4 (sediment)
Intersex in male fish	VERY	Mixtures of compounds with common MoA (estrogenicity)	Yes	Yes	?	yes	Yes	Pelagic and/or Benthic (depends on species)	2 (MoA mix) 3 (MSFD) 4 (sediment)
MN (micronucleus)	MODERATE-HIGH	Mixtures of compounds with common MoA (genotoxic/mutagenic)	Yes	Yes	Yes	Yes	Yes	Pelagic and/or Benthic (depends on species)	2 (MoA mix) 3 (MSFD) 4 (sediment)
Amphipod embryo malformation (brackish water)	VERY	Complex mixtures	Yes	Yes	Yes	Yes (Baltic)	Yes	Pelagic and benthic	1 (complex mixtures)

Biomarker name	Ecological Relevance	Responds to	SOP available?	Assessment criteria available?	Monitored?	Marine?	Limnic?	Protection goal	Applications of relevance
									3 (MSFD) 4 (sediment)
AChE (acetylcholinesterase)	HIGH	Mixtures of compounds with common MoA	Yes	Yes	Yes	Yes	Yes	Pelagic and/or Benthic (depends on species)	2 (MoA mix) 3 (MSFD) 4 (sediment)
Comet Assay	MODERATE-HIGH	Mixtures of compounds with common MoA	Yes	Yes	?	Yes	Yes	Pelagic and/or Benthic (depends on species)	2 (MoA mix) 3 (MSFD) 4 (sediment)
Mussel histopathology (gametogenesis)	MODERATE-HIGH	Complex mixtures	?	Yes	?	Yes	No	Benthic (mussels)	1 (complex mixtures) 3 (MSFD) 4 (sediment)
Stress on stress	HIGH/VERY	Complex mixtures and other stressors	?	Yes	Yes	Yes	No	Benthic (mussels)	1 (complex mixtures) 3 (MSFD) 4 (sediment)
SfG (Scope for Growth)	HIGH/VERY	Complex mixtures and other stressors	Yes	YES	Yes	Yes	No	Benthic (mussels)	1 (complex mixtures) 3 (MSFD)

Biomarker name	Ecological Relevance	Responds to	SOP available?	Assessment criteria available?	Monitored?	Marine?	Limnic?	Protection goal	Applications of relevance
									4 (sediment)
Benthic diatom malformation	MODERATE-HIGH	Complex mixtures	Yes	Yes	(Yes)	No	Yes	Benthic (benthic organism)	1 (complex mixtures) 4 (sediment)
Egg-shell thinning	VERY	DDT	Yes	Yes	Yes	Yes	No	Secondary poisoning	3 (MSFD) 5 (regulated substances)
Sea eagle productivity	VERY	Complex mixtures but priority suspects are DDTs	Yes	Yes	Yes	Yes	No	Secondary poisoning	1 (complex mixtures) 3 (MSFD)
Pregnancy rate in seals	VERY	Complex mixtures but priority suspect are PCBs	Yes	Yes	Yes	Yes	No	Secondary poisoning	1 (complex mixtures) 3 (MSFD)
Mentum deformation in chironomids	MODERATE-HIGH	Complex mixtures	?	?	(Yes)	No	Yes	Benthic	1 (complex mixtures) 4 (sediment)

5.2.2. *In vivo* assays

In vivo bioassays are performed using living organisms. They have the capacity to provide an integrated response at organism level to contaminants in a sample. In general, ecologically relevant endpoints are investigated. The advantages of using *in vivo* assays are demonstrated by their broad implementation in pesticide regulation and effluent monitoring, monitoring programmes of Marine Conventions, and in sediment dredging. Many data on the impact of chemicals regulated under REACH, for example, are obtained using bioassays, and their long-term application with standardised protocols (standards, guidelines) offers information on the precision of the procedures.

In vivo bioassays are tests in which whole living organisms (including bacteria and algae) are exposed to environmental samples such as surface water, sediment, waste water, dredged material, or extracts from these samples. Tests are performed in the laboratory or, less frequently, in the field (called "in situ assays").

The "endpoint" is related to the type of effect that is measured, and some examples that are frequently used in this context are:

- Mortality
- Immobilisation
- Effects on reproduction (i.e fertilisation, hatching, embryo development)
- Effects on growth of individuals
- Effects on growth of populations
- Metabolic or physiological changes
- Behavioural changes
- Bioluminescence
- Molecular/Biochemical responses.

In general, *in vivo* bioassays are broad spectrum assays, e.g. an *in vivo* bioassay reacts to a variety of substances and different MoAs. It is important that the evaluation of toxic effects of a sample is based on the response in several species, because they can exhibit intrinsic differences in terms of sensitivity to various chemicals and also depending on the endpoint measured in the test. Both short- and long-term *in vivo* bioassays should preferably be carried out on at least three species from different taxonomic groups and trophic levels (primary producer, decomposer/saprophytic, detritivore/filter feeder, consumer). The battery of ecotoxicological tests should have sufficient sensitivity and an overall discriminatory power responding to as many forms of pollution as possible; consequently, they have little specificity for different MoAs although in some cases (e.g. embryos of fishes) morphological alterations could point to the identification of specific MoAs

Samples often need to be concentrated before using *in vivo* assays in this context, especially if using short-term tests; see also Annex to the technical report (European Commission, 2014).

In the Inventory, a total of 46 *in vivo* bioassays (see Annex II) have been collected; these include the following MoAs: growth biomass (algae), photosyntesis inhibition (PSII), reproduction (*Daphnia magna*, crustaceans, amphipods, snails), lethality, developmental toxicity and behaviour (fish embryos, chironomidae), reproduction-endocrine disruption (*Gammarus*). For the marine environment, *in vivo* bioassays have been collected for rotifera, crustacea, polychaeta, and ostracoda. Generally, most of these bioassays are used in monitoring programmes in the context of wastewater regulation, marine monitoring and sediment dredging.

Applications investigated for *in vivo* bioassays

- Cover complex mixtures (of unknown composition) and perhaps even cumulative effects when combined with other stress factors – to assess status and/or identify significant pressure
- Identify relevant MSFD indicators
- Assess sediment quality
- Assess metal bioavailability when water chemistry outside validation range
- Assess status where high natural metal concentrations (>EQS)
- Assess quality of effluents or leachates

5.2.3. *In vitro* assays

In contrast to *in vivo* assays that capture the effect of chemicals on whole organisms, *in vitro* assays detect unwanted biological effects on a molecular level such as the activation of a cellular receptor or signaling pathway, the induction or inhibition of a specific enzymatic activity or the mutation of a DNA sequence. *In vitro* EBMs are fast and have the potential for automation, and thus allow high-throughput screening of samples. They are widely used for screening purposes in chemical risk assessment because at least in part they can serve as alternatives to animal testing. The ECHA promotes such alternative methods for the assessment of the hazards of substances. As a prominent example, a combination of the Ames test and the micronucleus test was able to detect almost all of the 962 rodent carcinogens and *in vivo* genotoxins tested in a study by Kirkland et al. (2011).

In vitro bioassays which measure the same endpoint and employ the same species as *in vivo* reference models may display different sensitivity for the same substance or chemical mixture.

In Annex VIII, the WFD provides an indicative list of main pollutants in European water bodies. For protection against possible chronic effects caused by environmental contaminants, Annex VIII defines compounds with “carcinogenic or mutagenic properties or properties which may affect steroidogenic, thyroid, reproduction or other endocrine-related functions”.

As discussed above, it is evident that chemical analysis alone cannot cover all potentially harmful compounds present in the water environment, which indicates the need for EBMs. However, due to the high time- and cost efforts and ethical considerations it is not possible to routinely assess effects such as carcinogenicity or reproductive toxicity in water samples with chronic whole-organism *in vivo* bioassays. In contrast, the unwanted biological properties of compounds listed in Annex VIII concern some molecular effects linked to possible chronic effects detectable by various *in vitro* EBMs that are implemented under REACH such as the Ames Fluctuation Test (Reifferscheid et al. 2012) or the micronucleus assay (Reifferscheid et al. 2008) and mammalian cell chromosome aberration assays to detect the mutagenic potential of chemicals.

Against this background it is reasonable to use the same instruments to directly measure these molecular initiating events, such as the activation of a hormone receptor or the mutation of a DNA sequence, in water samples. For *in vitro* EBMs, which are usually based on eukaryotic cell lines or single cell microorganisms (bacteria, yeast), these molecular events are displayed as a quantifiable signal such as fluorescence, light emission or a colour change. Therefore, *in vitro* EBMs are usually less cost-intensive compared to *in vivo* EBMs and have the potential for automation and high-throughput analysis.

For several *in vitro* EBMs it has been demonstrated that they meet the request to capture mixture effects of chemicals acting together. One group of such bioassays detects hormone-like effects based on the activation of nuclear receptors. The assumption is that all agonistic compounds present in a water sample contribute to the activation of a given receptor. By this means, unwanted effects defined in WFD Annex VIII can be addressed in a more holistic and direct way using *in vitro* EBMs.

Effect levels of *in vitro* EBMs are frequently reported as 'biological equivalence concentrations' (BEQ) that express the biological response of the *in vitro* EBM as a concentration of a reference compound resulting in the same response (Brack et al. 2017, Escher et al. 2015, Neale et al. 2015, Altenburger et al. 2015, Wagner et al. 2013, Villeneuve et al. 2000). Thus, the BEQ reflects the overall biological activity with respect to the effect under investigation. A generic guideline to calculate BEQ from experimental data is currently under development by ISO [ISO/NP 23196: Water Quality - Calculation of biological equivalence concentrations (BEQ)]. Results from classic chemical analysis can be translated to biological effects by multiplying measured concentrations by the relative effect potency of the target compound. By this means a measured biological effect can be matched against an expected biological effect based on concentrations of contaminants.

For certain applications, e.g. for screening purposes, threshold values must be defined to assess results from *in vitro* EBMs expressed as a BEQ. Effect-based trigger values (EBT) can be used as such threshold values. Section 5.3.1 and Annex III provide more details on the definition of EBTs and demonstrate this concept for agonists of the ER. An exceedance of the EBT would trigger further actions such as analysis of samples by high-end chemical analysis. Such an approach is of special interest if:

- compounds cannot be detected with routine chemical analysis because of insufficient sensitivity of the method,
- a number of compounds not fully covered by chemical monitoring act in an additive way by the same mode of action and mixture effects have to be considered.

However, the interpretation of *in vitro* test results is more challenging compared to *in vivo* EBMs for two reasons:

1) the manifestation of an adverse outcome at the organism or population level is not only determined by a given molecular initiating event but influenced by a number of biotic and abiotic factors. Furthermore, different molecular initiating events may lead to a common adverse outcome. As a basis for interpretation, the concept of the adverse outcome pathway (AOP) is used to elucidate causal relationships between key molecular initiating events and effects at higher biological levels (Ankley et al. 2010). The possibility that a specific molecular initiating event is the cause of an adverse effect increases with the completeness of the AOP and thus the relevance of a related *in vitro* EBM.

2) different toxicokinetics between cellular *in vitro* EBMs and organismic *in vivo* EBMs can hamper extrapolation from the results of an *in vitro* EBM to a whole organism or population (Brinkmann et al. 2017).

Consequently, an EQS defined for a single compound based on *in vivo* studies cannot be translated directly to a threshold value to assess the results of an *in vitro* EBM. In addition, the contribution of a single compound to a sum effect that is measured by the *in vitro* EBM cannot be quantified without separation of the mixture. In other words, it is not possible to determine the individual contributions to the sum if only the sum is known.

3) *in vitro* EBMs are frequently applied to enriched water samples. In the case of in *vitro* tests addressing estrogenic effects, for example, the sensitivity is sufficient to detect estrogenicity in waste water effluents. However, an enrichment is required for surface water samples. In contrast to chemical analysis, no internal standard can be applied to correct for an incomplete recovery of e.g. estrogenic compounds. This might lead to an underestimation of effects. Possible impacts of the sample matrix on the enrichment can be roughly estimated by spiking a sample aliquot with a reference compound. The assessment of results obtained by *in vitro* EBM with enriched samples is less problematic compared to the testing of enriched samples with *in vivo* EBM. In the latter case, higher concentrations of compounds might trigger unspecific effects that are not related mechanistically to possible chronic effects caused by the same compound at lower concentrations. In the case of *in vitro* assays, results can be matched against EBTs (see below) under consideration of the relative sample enrichment.

Despite these limitations, *in vitro* EBMs are an important tool to feasibly address chronic effects from chemicals in water bodies. The linkage between estrogenic effects and the possible occurrence of chronic effects is well accepted. Besides estrogenicity, further molecular mechanisms – especially with respect to endocrine regulation – are discussed in the context of the AOP (see below) as initiating events for adverse outcomes such as agonistic and antagonistic effects on the androgen receptor. The underlying and discussed uncertainties have to be acknowledged, but without the use of relevant *in vitro* EBMs essential information for the assessment of water quality would be neglected.

For the selection of *in vitro* EBMs to assess the quality of water bodies, three criteria have to be met:

1. **Relevance of the *in vitro* EBM**: as discussed above, positive results from *in vitro* EBMs do not necessarily indicate adverse biological effects at a higher biological level per se. Effects at the molecular level have to be mechanistically linked to apical endpoints. This can be done following the concept of the AOP (Ankley et al. 2010). According to OECD-document ENV/JM/MONO(2016)12. "An AOP describes a sequence of events commencing with initial interaction(s) of a stressor with a biomolecule within an organism that causes a perturbation in its biology (i.e. molecular initiating event, MIE), which can progress through a dependent series of intermediate key events (KEs) and culminate in an adverse outcome (AO) considered relevant to risk assessment or regulatory decision-making". The Adverse Outcome Pathway (AOP) Wiki[28] serves as the primary repository of qualitative information for the international AOP development effort coordinated by the Organisation for Economic Cooperation and Development (OECD). The completeness of a proposed AOP indicates the relevance of an *in vitro* EBM that is able to detect a specific molecular initiating event. A further line of evidence for the relevance of *in vitro* EBM is the availability of field studies that link the occurrence of adverse effects on populations or human health to molecular initiating events or the presence of compounds known to trigger these specific molecular initiating events. Here, prominent examples are studies by Kidd et al. (2007 and 2014) demonstrating effects of 17α-ethinylestradiol on a lake ecosystem. Comparable studies investigating the relevance of *in vitro* effects for the prediction of population status are scarce and are much needed. Finally, the relevance of an *in vitro* EBM is indicated by the inventory of chemical MoAs, as demonstrated in Section 5.1, such as photosynthesis inhibition, endocrine disruption or genotoxicity. Nevertheless, due to improvements in EBT

[28] https://aopwiki.org/

derivation a linkage to EQS with population relevance for many species and to levels of higher biological relevance is possible with high specificity and sensitivity (see Annex III). EBTs can currently be proposed for 21 different MoAs and endpoints covering around 37 EBM (see Table III.12).

2. **Maturity of the *in vitro* EBM**: As for other EBMs the standardisation of an *in vitro* EBM according to ISO/CEN/DIN or validation via OECD is a key element for its use in Europe-wide studies including a number of laboratories. By this means the transferability of an *in vitro* EBM and the comparability of related results can be guaranteed and the method is sufficiently characterised with respect to achievable quantification limits and variabilities. The latter is crucial to determine if results obtained by different laboratories on different samples differ with statistical significance or not. If no standard is available, performance characteristics of the method should be characterised by means of (international) interlaboratory trials. Without this information, EU-wide assessment of data provided by different laboratories is impossible. *In vitro* EBMs that are not validated by interlaboratory trials might be used on a regional scale for e.g. investigative monitoring.
3. **Assessability of results obtained by an *in vitro* EBM**: in general results obtained by an *in vitro* EBM can be assessed relative to other values resulting from measurements with the same *in vitro* EBM or the results can be matched against a defined EBT as outlined in Section 5.3. The relative assessment allows the prioritisation of water bodies, source identification and investigative monitoring. A status assessment would require an accepted EBT for the given *in vitro* EBM. However, in any case it is desirable to use EBM results in terms of a risk assessment for which an EBT has to be proposed (see Table III.12 in annex 3). Therefore, *in vitro* EBMs with defined EBTs are to be favored above EBMs without defined EBTs.

In vitro EBMs capture the presence of known and unknown contaminants (**application 1**) that exhibit the specific MoA detected by this *in vitro* EBM, e.g. all receptor agonists present in a sample contribute to the activation of the receptor. By this means *in vitro* EBMs cover also mixture effects (**application 2**). If receptor antagonists are present as well, the *in vitro* EBM would measure the integral effect of the mixture. However, *in vitro* EBMs do not integrate biological effects on other target molecules or possible mixture effects at higher biological levels. *In vitro* EBMs are applicable to marine samples as well (**application 3**) when working with extracted samples. Due to the low concentrations expected – especially in marine samples - sample enrichment is recommented. *In vitro* EBMs can be used for the characterisation of sediment samples (**application 4**) using pore water and eluates or extracts from sediment. Two *in vitro* EBMs detecting sediment-associated mutagenic and estrogenic effects are used for the assessment of dredged material in Germany (HABAB-WSV 2017, Annex 2). Several *in vitro* EBMs such as assays detecting dioxin-like effects can be used as 'bioanalytical' tools for screening purposes prior to a chemical analysis (**application 5**). *In vitro* EBMs can be used in relation to drinking water production with a special focus on *in vitro* EBMs addressing effects with relevance for human health such as mutagenicity (Richardson et al. 2007). Numerous studies demonstrate that they have the potential to be used to assess effluents from waste water treatment plants and leachates from landfill sites (**application 9**, Escher et al. 2014).

Selected MoAs and respective *in vitro* EBMs are presented in more detail in Chapter 6. The inventory of *in vitro* EBMs shows a number of various methods addressing several MoAs. Based on available information and discussions within the activity, several MoAs

were prioritised that can be addressed by certain *in vitro* EBMs suitable for effect-based assessment of water quality. These are listed in Table I.1. Annex I.

***In vitro* EBMs conclusions**

- *In vitro* EBMs allow the specific detection of relevant MoAs at a molecular level
- *In vitro* EBMs allow for cost-efficient high-throughput measurements
- A number of *in vitro* EBMs are standardised and thus mature for implementation
- Results can be used for a relative assessment, for prioritisation, source identification and investigative monitoring
- *In vitro* EBMs with defined EBTs can be used for screening purposes and possibly even for a status assessment
- The concept of the AOP can be used for the prioritisation of *in vitro* EBMs and should be further developed
- For many EBMs, EBTs are already available, usable and recommended (see annex III)

5.3. EBM Assessment criteria (objective 4)

In general, assessment criteria are needed in order to classify a waterbody and then decide on the measures to be applied. The methods to derive EQSs for chemicals are widely accepted and largely based on procedures already in place within the context of the chemicals legislation (such as REACH and the Biocidal Products Regulation). For EBMs this is not (yet) the case. Depending on which subcategory a particular EBM belongs to, the results need to be interpreted in different ways. For EBMs, there are several assessment criteria that can be applied depending on the type of EBM: *in vitro*, *in vivo*, and biomarkers of various types. A brief description is included below for the three different EBM types and Annex III describes an EBT derivation approach and compilation for *in vitro* and *in vivo* EBM.

Objective 4 of the ToR

Objective 4: Development, where possible, of *in vivo and in vitro* effect-based trigger values, signalling a risk to or via the aquatic environment (including risks to human health from chronic exposure via consumption of drinking water or fishery products if possible), with the aim of making effect-based methods applicable (alongside chemical tools) in WFD chemical monitoring and assessment.

5.3.1. Biomarkers

Biomarker results have so far been evaluated in relative (e.g. time trends or comparisons between reference and impacted sites) and/or in absolute terms (against "fixed assessment criteria" – comparable to the EQSs) but also in an integrated manner (weight of evidence)

(see also European Commission, 2014). In the MSFD context, the employed methods, specific effects and evaluation parameters comprising safety threshold values are to be based on local experience or on knowledge transfer, thus leading to a heterogeneity in the quality of results across Europe.

Fixed assessment criteria for biomarkers

For biomarkers, the ICES Working Group on the Biological Effects of Contaminants (WGBEC) has developed several so-called BAC (Background Assessment Criteria) and EAC (Environmental Assessment Criteria) values (Davies and Vethaak 2012, OSPAR 2013). Although, to our knowledge, there is no strict "guidance document" on which procedures to use (corresponding to the CIS Guidance Document No. 27 on deriving EQSs), and the actual methodology may vary between different biomarkers, the BACs and EACs for biomarkers are generally based on the deviation from reference conditions. The BAC and the EAC could be considered as equivalents of the WFD high/good- and good/moderate-boundaries, respectively. Under the MSFD, EAC is used as the boundary for good environmental status (GES). Available EACs and BACs for the biomarkers in the inventory are included in Table II.3.

As was pointed out in chapter 5.2., the biomarkers in the inventory could be divided into two main subgroups according to their ecological relevance (low relevance vs moderate or higher). It can also be noted that EACs have so far been developed primarily for effect biomarkers, whereas BACs are available also for exposure biomarkers. At least in theory, it would probably be possible to establish EACs for most effect biomarkers of moderate, high or very high ecological relevance (see Table 3 in chapter 5.2.) since they can be related to adverse impacts at least at tissue level. Such biomarkers could therefore be possible to evaluate one by one and using the "one-out-all-out" (OOAO) approach[29]. However, it is important to be aware of the increased risk of false positives when multiple biomarkers are used. In such cases, it may instead be advisable to use a weight–of-evidence approach. For biomarkers with low ecological relevance, it would probably be inappropriate to assess status based only on exceedances of individual assessment criteria. Instead, such biomarkers are more valuable as a source of supportive information in a weight-of-evidence approach.

Options for the use of individual and fixed assessment criteria

For those biomarkers where it would not be appropriate or possible (today) to assess effects in relation to EACs (or similar fixed assessment criteria), an alternative would be to analyse the time trend. A significant trend in observed effects could suggest that although effects might not be severe enough to cause negative health impacts today, effects at higher organisational levels (e.g., population) in the long run cannot be excluded.

The "trend approach" is already included in the WFD context. For priority substances accumulating in sediment and/or biota, concentration trends are to be monitored and MS have to take measures aimed at ensuring "that such concentrations do not significantly increase in sediment and/or relevant biota" (see EQSD art 3.6.). Also, in the MSFD context, some indicators are evaluated using a trend approach instead of or along with fixed thresholds.

[29] The OOAO approach is used under the WFD for example to assess chemical status, which means that it is sufficient that one substance occurs at concentrations above its EQS for the overall chemical status to be considered "not good" (non-compliance).

Another common approach in this context is to use a weight-of-evidence approach or a combination of both trend and weight-of-evidence approaches. The fish biomarkers included in the Swedish monitoring programme are for example evaluated on an annual basis, together with population data and data from chemical monitoring of fish from the same sites. Several long-term trends have been identified at monitoring stations used as reference sites, and the monitored (and regulated) contaminants at those sites are generally decreasing. However, the biomarker results suggest that the perch could be showing signs of toxic stress that could either be related to unknown chemicals, mixtures or a combination of several stress factors including toxic substances (resulting in "cumulative effects"). Biomarker responses include induction of the detoxification enzyme EROD, and increased glutathione reductase (GRed) activity[30] (Figure 4 below). Calcium concentrations are also signalling impacts on ion regulation, red blood cells are reduced in number, gonad size has decreased, etc. (for more information, see Mustamäki et al., 2018). Since similar symptoms are being observed on both the west coast (North Sea) and the east coast (Baltic Sea) and at all three reference stations on the east coast it is now believed that the effects are probably happening on a large scale.

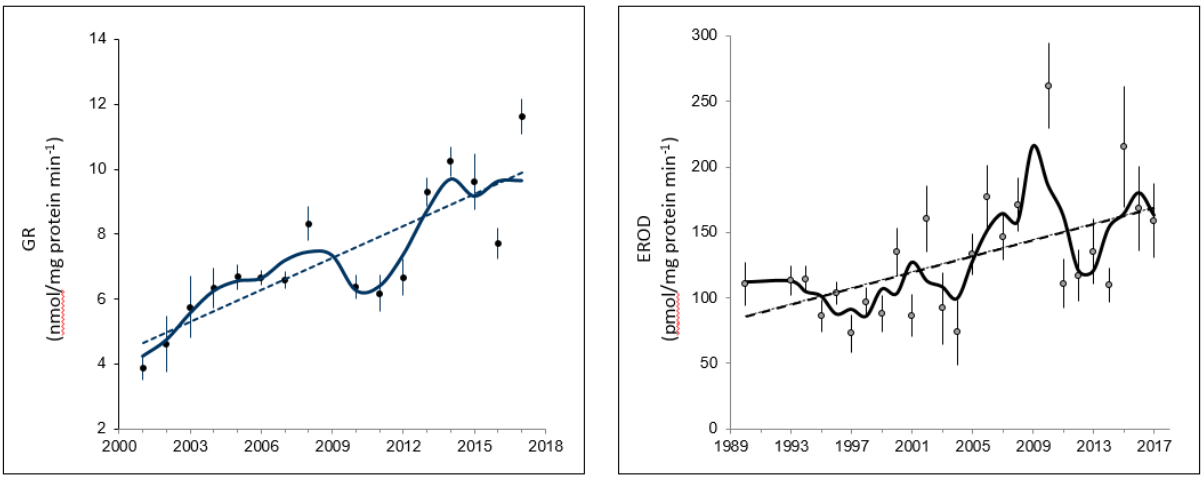

Figure 4. Activity of glutathione reductase (GRed) and detoxication enzyme EROD in the liver of female perch. Sampling is performed on female perch of similar size and at the same time of the year. Mean values with 95% confidence interval. Solid line represents three-year rolling mean values and the dotted line a significant trend. Modified from Mustamäki et al. 2018.

The example above illustrates how the evaluation can be performed using "expert judgment" and focusing on trends for several different variables taken together.

However, in the annex to the technical report (European Commission 2014), there is an example illustrating how a large set of biomarker data, including exposure biomarkers, could be evaluated not only through expert judgement, but also in a formal and transparent way using the scoring scheme below, developed to facilitate the evaluation of biomarker data obtained from a fish biomarker battery and where the individual variables are given different weights.

For each of the individual markers included below, "assessment criteria" have now also been developed, based on responses observed at monitoring stations from reference areas (Hanson et al. 2014). Such values and scoring procedures would likely aid in the

[30] indicating increased oxidative stress

interpretation, make it more transparent and also facilitate taking other data into account as well but without applying the OOAO principle where this would be inappropriate for a particular biomarker. However, please note that for some of the biomarkers included in the scoring, such as a response in dead/malformed embryos, a read-out above the marker score would be sufficient to trigger an exceedence of the limit for impact on function. For more details, refer to the Technical Report (European Commission 2014).

Table 4. Proposed scoring system for an integrated assessment using fish biomarkers (included in the Swedish monitoring programme). The assessment is based on a weight-of-evidence approach where individual biomarkers are grouped based on physiological function. If the score of the biomarkers within each function exceeds the limit, the function is considered impaired. Overall, biomarkers should be regarded as affected if any of the functions "reproduction" or "condition and metabolism" is considered impaired. Overall, biomarkers should also be considered affected if at least two of the other functions are considered impaired.

Function	Score	Limit
Reproduction		
Reduced gonad size	1	1
Increased vitellogenin for male fish	1	
Reduced vitellogenin for female fish	1	
Skewed primary sex ratio (eelpout)	1	
Contition and metabolism		
Reduced condition factor	2	2
Increased condition factor	1	
Change in liver size	1	
Change in glucose	1	
Change in lactate	1	
Liver function		
Change in liver function	3	4
Change in EROD activity	1	
Change in GRed activity	1	
Change in MT	1	
Immune defense		
Change in total white blood cells	2	3
Increase in macrofage centra	2	
Change in lymphocytes	1	
Change in thrombocytes	1	
Change in number of granulocytes	1	
Red blood cells		
Change in hematocrit	2	3
Change in hemoglobin	2	
Change in immature red blood cells	1	
Ion regulation		
Change in potassium	2	3
Change in calcium	2	
Change in chloride- AND sodium	3	
Change in chloride- OR sodium	1	

5.3.2. *In vivo* assays

For *in vivo* assays already used in, e.g. whole-effluent assessments, the assessment principle is similar to that involving an EQS, since the results are expressed as, e.g. EC50s or NOECs, although the "C" (concentration) does not refer to the concentration of a particular substance in this case, but rather to the dilution of the sample, which is tested in a dilution series. Emission limit values for such effluents can then be expressed in these toxicological terms.

In vivo assays are also frequently used in the assessment of contaminated sites, including sediment and surface waters, in which a battery of assays is foreseen. The results can then be evaluated using a weight-of-evidence approach, e.g., the so called "triad approach" (Chapman 1990).

Triad approach

Field observations provide information about possible human impact on ecosystems. However, they do not always show what is causing the impact or which types of management action are needed. For this, multiple lines of evidence may be needed. The sediment quality triad (SQT) is a widely accepted method and conceptual framework to assess sediment quality using three components (Chapman 1990). The three main components in the SQT are: 1) sediment chemistry, 2) sediment toxicity tests, and 3) field observations (Figure 5).

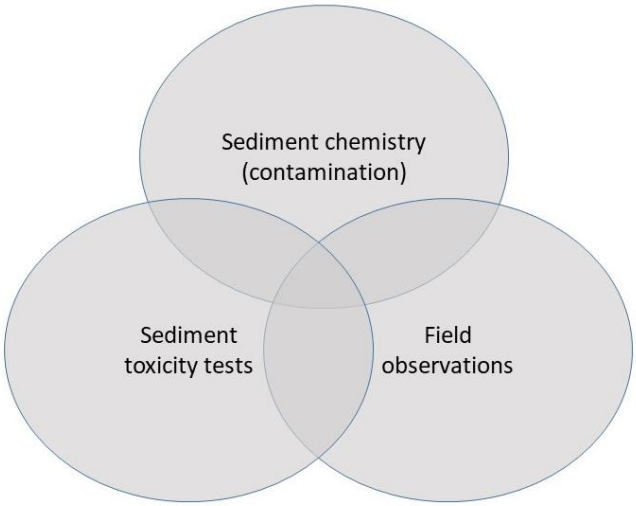

Figure 5. The sediment quality triad is based on the three components sediment chemistry, sediment toxicity tests, and field observations. The figure is modified from Chapman (1995).

The three components provide different pieces of information that can be used to reach the most scientifically justified conclusion. For this, a decision matrix can be used to provide guidance (Table 5). In the decision matrix, each of the three components is given a "Yes" or a "No", depending on the response (impact/no impact). When, for example, all three components are answered with "Yes", there is strong evidence that community effects are (at least partly) caused by toxic chemicals. When all components are answered with "No", there is strong evidence against such effects.

Table 5. Decision matrix for the Sediment Quality Triad. Based on Chapman (1996).

Contamination	Toxicity	Field effects	Possible conclusion
Yes	Yes	Yes	Strong evidence for pollution-induced degradation
No	No	No	Strong evidence against pollution-induced degradation
Yes	No	No	Contaminants are not bioavailable
No	Yes	No	Unmeasured contaminants have the potential to cause degradation
No	No	Yes	Alteration is not caused by contamination

Yes	Yes	No	Toxic contaminants are bioavailable, but *in situ* effects are not demonstrable
No	Yes	Yes	Unmeasured toxic contaminants are causing degradation
Yes	No	Yes	Contaminants are not bioavailable, alterations not due to toxic chemicals

5.3.3. Effect-Based Trigger values (EBT) for *in vitro* assays

For *in vitro* assays, a reference substance is normally used not only to check the performance of the test but also to conduct a positive control for comparison with the observed effects. The results are then expressed as a "biological equivalence concentration".

EBMs are complementary to chemical analysis and can provide relevant information about mixture effects of chemicals in water. Standardised criteria for the application of such methods in a legal framework are needed in order to ensure a robust analysis of results across Europe. Due to the lack of scientific knowledge on the behaviour of single compounds in chemical mixtures, and to the heterogeneity of studies evaluating the efficacy of the numerous EBMs developed over the past decades, many approaches have been proposed to deriving safety threshold values (Tang et al. 2013, Jarosova et al. 2014, Kunz et al. 2015, van der Oost et al. 2017, Escher et al. 2018).

Translating environmental quality standards (EQS) directly into their corresponding biological equivalence concentration (BEQ), which are further expressed as EBTs, is the most widely used approach (Figure 6). BEQ translates the readout of an EBM to the concentration of a reference compound. By analogy to EQS defined for chemical parameters, EBTs serve as a benchmark to differentiate between an acceptable and an unacceptable level of an unwanted biological activity or ecological risk that is elicited by a given water sample.

The risk-quotient based on chemical analysis is given by:

$$R_i = \frac{c_i}{EQS_i}$$

with

R_i	risk-quotient for compound i
c_i	concentration of compound i
EQS_i	environmental quality standard of compound i

The risk-quotient based on EBMs would be given by:

$$R_{EBM} = \frac{BEQ}{EBT}$$

with

R_{EBM}	risk quotient based on *in vitro* EBM
BEQ	biological equivalence concentration determined with an *in vitro* EBM
EBT	effect-based trigger value

A MoA-specific EBT can be used as a guidance value to assess the quality of a water body:

Measurement (BEQ) < x EBT low probability of risk

Measurement (BEQ) ~ EBT hazard risk possible

Measurement (BEQ) > x EBT high probability of risk

This approach can be used for prioritisation in risk characterisation, screening or possibly even status assessments (see Kase et al. 2018).

The definition of EBTs is relevant especially *for in vitro* EBMs because measured effects are not adverse per se as they are for *in vivo* EBMs showing e.g. acute toxicity or growth inhibition etc. As discussed under 5.2.3, *in vitro* EBMs detect molecular initiating events that are related to possible adverse effects.

Similar to the definition of EQS as threshold values for chemical parameters the derivation of EBTs has to deal with inevitable uncertainties. Uncertainties associated with the definition of EQS are caused by a lack of knowledge about possible mixture effects, species extrapolation and, in part, long-term chronic effects if mainly short-term, acute data are available for a certain compound. Mixture effects including known and unknown compounds are better captured by EBTs. However, uncertainties arise based on the limitations of *in vitro* EBTs as discussed in Section 5.2.3.

The general approach for the derivation of EBTs uses the available EQS data specific for every substance in the analysed chemical mixture when the composition is known, or for the most potent reference compound. BEQs are determined through EBMs suitable for the level of biological complexity under assessment and selected reference compounds taking into account their mode of action (MoA). Specific effect concentrations (EC) are compared to an EC of a reference compound. The ratio of both EC-values gives the relative effect potency (REP) of the compound. REPs from different *in vitro* and *in vivo* studies are then considered to derive toxic equivalency factors (TEFs). Multiplication of either REPs or TEFs by the concentration from chemical analysis allows the calculation of toxic equivalency (TEQ) for single toxicants or as a sum of multiple similarly acting substances in a mixture (Figure 6). In the evaluation of adverse outcomes, mixture effects can be assessed for only one MoA in the case of specific methods (e.g. receptor-mediated effects), or for more biological pathways if employing wide-spectrum EBMs.

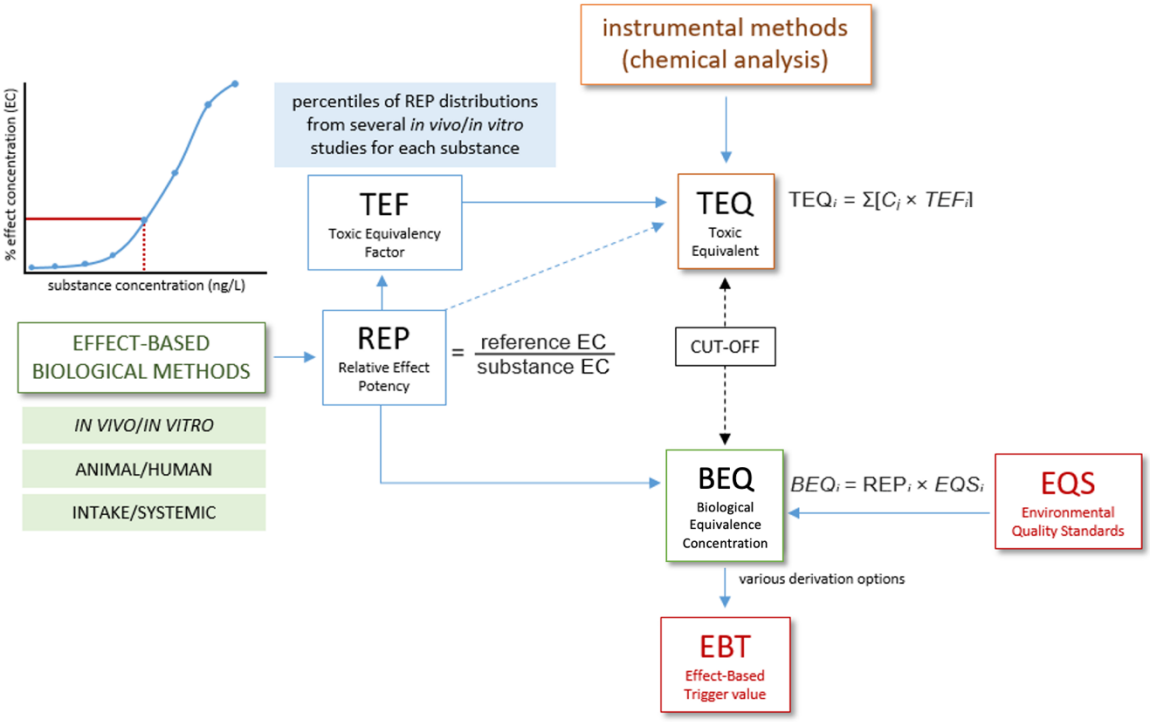

Figure 6. Schematic flowchart of approach commonly used to derive bioanalytical and toxic equivalents (BEQ and TEQ) by combining effect-based methods (EBMs) and confirmatory chemical assessment methods.

The probability of a harmful effect on the environment increases with an increasing risk quotient regardless of whether if this is based on chemical analysis or an EBM. Although the exceedance of a risk quotient of 1 does not necessarily mean that effects in the environment will occur, the risk quotient reflects the likelihood of adverse effects occuring in the aquatic environment.

How to assess the predictive power of a proposed Effect-Based Trigger value (EBT)?

Because *in vitro* EBMs link chemical contamination and adverse effects at higher biological levels by the detection of molecular initiating events, results obtained by *in vitro* EBMs – expressed as BEQ values – can be related to both data from chemical analysis and adverse effects *in vivo*. This can be done based on a specificity and sensitivity analysis using data from the chemical analysis as a reference point or from *in vivo* EBMs. By using such an approach, proposed EBTs can be assessed for their power to predict the presence of chemical contaminants that trigger a given molecular initiating event, or the occurrence of adverse effects at higher biological levels. This is exemplified for the relevant molecular initiating event 'activation of the ER' in Annex III (Section 1) using a data set including data from hr-LC/MS measurements, five *in vitro* EBMs and an *in vivo* transgenic fish model. In Annex III Section 1 it is demonstrated that EBM-specific EBTs can be used to identify samples containing elevated levels of the WL substances E1, E2 and EE2, and the activation of the ER in the brain of a transgenic model fish. In fact, the predictive power for the effect in the fish model was higher than that based on the chemical analytical data. Here, a general outline of the concept is described.

The analysis of sensitivity and specificity is based on the definition of a reference method, e.g. results obtained by chemical methods such as LC/MS and a subsequent benchmarking

of the reference results against the assessments based on a given EBM/EBT combination. By this means, true positive, true negative, false positive and false negative test results are defined as illustrated in Figure 7.

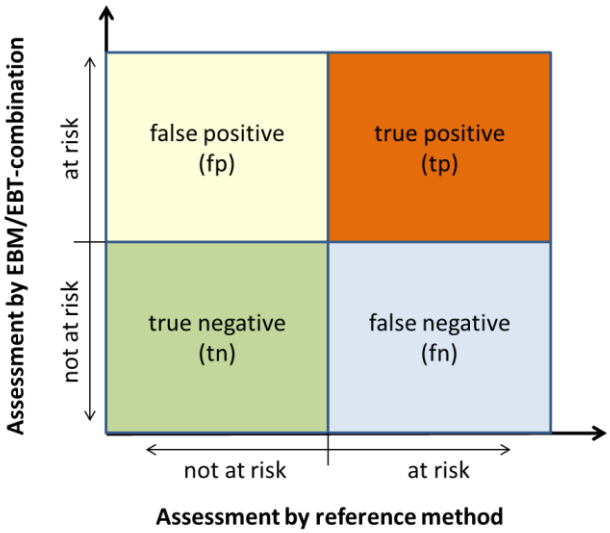

Figure 7. Assignment of true positive, true negative, false positive and false negative test results by comparing an assessment based on an EBM/EBT combination against an assessment based on the reference method.

The analysis of specificity and sensitivity based on a classification of results is frequently done to characterise alternative screening methods in medicine. Kirkland *et al.* (2005) used this approach to evaluate the ability of a battery of *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. The terminology (true positive, true negative, false positive and false negative) is defined with respect to the reference method that provides true results by definition because it is the selected anchor point for this analysis. Based on this classification, the sensitivity and specificity of a given EBM/EBT combination can be calculated as shown in Annex III Section 1.

The sensitivity gives the percentage of true positive assessments against all samples that were identified to be 'at risk' by the reference method. The specificity gives the percentage of true negative assessments against all samples that were identified to be 'not at risk' by the reference method. It is obvious that the parameters sensitivity and specificity have inverse tendencies. A very low EBT would result in 100% sensitivity, i.e. all samples assigned to be at risk by the reference approach would be identified, but in 0% specificity because all samples assigned to be not at risk by the reference approach would be identified as problematic by the *in vitro* EBM/EBT combination. A very high EBT would result in an inverse situation with 0% sensitivity and 100% specificity. Because two categories, i.e. 'at risk' and 'not at risk', have to be distinguished, the sensitivity and specificity of an *in vitro* EBM/EBT combination have to be well above 50% to show any predictive power over flipping a coin. The optimal case would be 100% sensitivity and 100% specificity. An EBM/EBT combination with 90% sensitivity would miss 10% of samples that were assessed to be 'at risk' by the reference method. In the case of 90% specificity, 10% of the samples identified 'at risk' by the EBM/EBT combination would be classified as 'not at risk' by the reference method. Thus, a balanced optimum would be an EBT that maximises sensitivity and specificity together. If different EBM/EBT combinations are available, the optimal option would be the combination showing the highest sensitivity and specificity. However, combinations with lower sensitivity and specificity could be defined e.g. for protected areas resulting in a more conservative quality assessment.

Conclusions on EBTs

- EBTs are used as benchmarks for results obtained by *in vitro or in vivo* EBMs expressed as biological equivalence concentrations
- The predictive power of an EBT-proposal can be assessed based on a sensitivity and specificity analysis
- Specific EBTs for a number of *in vitro* and *in vivo* EBMs are proposed

5.4. Ecological indicators (objective 6)

Under the WFD, population or community-level effects measureable using EBMs might be included as biological quality elements (BQEs) under ecological status. In practice, however, there are few such examples (and so far, only available for benthic assessments). Although a single EBM or even a battery of EBMs measuring effects at lower levels of biological organisation (organism and sub-organism) cannot be seen as measuring the equivalent of a biological quality element, they can deliver valuable information about possible pressures caused by chemical contamination that are not captured by chemical monitoring or current ecological status assessments.

5.4.1. WFD biological quality elements, BQE

Biological indicators are used under the WFD to support impact assessments and to determine ecological status. The biological indicators use different groups of organisms (biological quality elements, BQE), and the intention is that the most sensitive BQE should determine status. In CIS Guidance Document No. 3 (Analysis of Pressures and Impacts), guidance is given in Table 1 on which type of impact the different BQEs respond to. None of the biological index is, however, specifically linked to pollution by hazardous substances.

Moreover, very few biological response variables exist that both respond to toxic chemicals and that can be used under the WFD, given the current requirement in the WFD that effects on BQEs are measured at population- or community-level and consider structural rather than functional aspects (see WFD Annex V. 1.1.1.-1.1.4.):

- Composition, abundance and biomass of phytoplankton
- Composition and abundance of other aquatic flora
- Composition and abundance of benthic invertebrate fauna
- Composition, abundance and age structure of fish fauna.

The Technical Report (European Commission 2014) nevertheless mentions a few methods already used by at least one MS to assess biological effects and that could at least in part respond to toxic substances. Those were the British Infaunal Quality Index (IQI), Danish Quality Index Ver2 (DKIver2), Spanish Multivariate AZTI Marine Biotic Index (M-AMBI) and French "Benthic Opportunistic Annelida Amphipoda Index/Benthic Opportunistic Polychaete Amphipoda Index". Furthermore, four methods under development were described in the 2014 report. Those were the SPEAR index, the NemaSPEAR index, PICT, and a multimetric index based on traits.

In a follow up, it was noted that a multimetric index based on benthic species composition and traits is under development in France. The index is called I2M2 (Indice Invertébrés MultiMétriques). When combined with a "diagnostic toolbox", the most probable pressure acting on the community can be identified (Mondy and Usseglio-Polatera 2013).

To ensure that the assessment of ecological status is similar in all EU MS, the different BQEs are intercalibrated within so-called geographical intercalibration groups (GIGs). For practical reasons, however, most intercalibration has been performed by investigating relationships with the concentration of the limiting factor for primary production (phosphorous or nitrogen). Table 6 gives an overwiew of which pressures the different BQEs respond to. The list is not complete as, for example, acidification is not mentioned.

Table 6. The different BQEs used to assess ecological status, and the anthropogenic pressures to which they are linked (from CIS Guidance Document No. 3, Analysis of pressures and impacts).

Biological Quality Element	**Anthropogenic pressure**
Phytoplankton (ANNEX V, WFD) - Trophic status	Assessment of eutrophication
Macrophytes and Phytobenthos (ANNEX V, WFD)	Assessment of morphology and organic pressures*
Benthic invertebrate Fauna (ANNEX V, WFD): - Saprobic status - AQEM-Evaluation	Assessment of organic pressures*
Fish fauna: Species composition and abundance	Assessment of the river continuity and morphology

*Organic substances that contribute to the oxygen demand of water bodies.

According to Annex V to the WFD, species composition and abundance should be monitored to determine ecological status. In practice, species composition is evaluated using different indices, where species (or taxa) are given different weight based on their sensitivity to anthropogenic disturbance. Although relatively few indices have been developed specifically to detect impacts caused by chemical pollution, indices that describe general species composition could detect changes caused by several types of stress, including chemicals. However, it is often not known which species are sensitive to which chemicals (with the exception of some well-studied substances, including pesticides). Therefore, such indices may respond very differently to different chemicals. Furthermore, chemicals occur in mixtures and most often in combination with other types of anthropogenic pressures (e.g. elevated nutrient load and physical disturbance). This further complicates the interpretation.

Even though EBMs have not been used extensively for the assessment of ecological status (partly due to the limitations discussed above), biomarkers/bioassays have been introduced locally to complement the monitoring of BQEs. By combining EBMs at higher (BQEs) and lower (biomarkers/bioassays) levels of biological organisation it may be possible to identify the (chemical) cause of reduced ecological status. This is something that is necessary to achieve effective programmes of measures (PoMs). This use of EBMs to support BQEs qualifies as "investigative monitoring", and is described in Annex V to the WFD. CIS Guidance Document No. 32 provides a detailed list of species/tissues currently used in European biota monitoring programmes without, however, indicating the

application of specific EBMs, while criteria for the evaluation of BQEs are described in WFD Annex V.

The identification of pollution and anthropogenic pressures on the environment through the measurement of BQEs using ecological methods can be informative but does not in itself prevent effects on aquatic organisms. Indeed, alterations at population and community levels usually result from chronic exposure to chemicals and the initiation of one or more adverse outcome pathways.

5.4.2. Metagenomics

Metagenomics is the study of genetic material recovered directly from environmental samples. As this describes the genetic composition at a high level of biological organisation (communities), it can be considered an ecological indicator. However, metagenomics deviates from traditional ecosystem measurements (e.g. BQEs) in many respects. For assessing the effects of chemicals using metagenomics, microbial communities have been most well studied, e.g. with respect to the effects of antibiotics.

DNA sequencing of microbial communities, ideally in combination with chemical analysis and the evaluation of physico-chemical parameters, can identify links between exposure to (groups of) chemicals and observed effects on microorganisms, including: changed composition of sentinel communities (Kisand et al. 2012); increased abundance of pathogens; changes in metabolic pathways; and the transmission and/or development of antimicrobial resistance (AMR) (Garner et al. 2016, Gupta et al. 2018, Bengtsson-Palme et al, 2017).

Large-scale analyses of microbial DNA in aquatic communities have a particular value for the following reasons:

- They provide very detailed information on anthropogenic effects on the structure and diversity of microbial communities, including, e.g. bacteria, viruses, fungi and to some extent also protists, plants and metazoans (Bengtsson-Palme et al. 2015);
- They provide information about impaired ecosystem functions and services;
- They provide information on the risk for aquatic transmission of a large range of pathogens (bacteria, viruses, etc.);
- They provide information on the risk for antibiotic resistance selection and evolution, ideally in combination with chemical analyses and cultivation data (Bengtsson-Palme et al, 2018);
- They provide information about the type of chemicals and other stressors affecting aquatic communities based on the characteristics of present/lost members and their functional genes.

Waterbodies have been recognised as a transmission route for antibiotic resistant bacteria, but also as a potential arena for the evolution of new forms of resistance (Bengtsson-Palme et al, 2018). Metagenomic profiling of antibiotic resistance genes (ARGs) may provide information on both of these processes. Analyses of antibiotic concentrations could also provide critical input on the risks for selection and hence evolution of resistance in aquatic environments (Bengtsson-Palme and Larsson, 2016). The abundance of specific functional gene categories responsible for certain processes (e.g. detoxification pathways, nitrification etc) may also provide information on impaired ecosystem functions or services of exposed communities.

Analyses are often based on random DNA fragments (shotgun metagenomics). Such a random or untargeted approach is important as it can result in unexpected findings, in contrast to analyses of a limited set of predefined endpoints. Hence, shotgun metagenomics data can also be useful for retrospective datamining when new questions arise. The exceptional diversity of microorganisms present in most samples is still often a challenge for detecting rarer members or genes carried by them, although costs for sequencing have dropped dramatically in recent years. PCR-amplified DNA regions, such as ribosomal sequences partly conserved across bacterial species, can therefore be used for more focused analysis, for example for providing deep taxonomic information. A potential drawback for certain applications is that individual genes can be difficult to link to species. This may be partly overcome by approaches like epicPCR (Spencer et al. 2016). Although more challenging, RNA may also be studied. The quality of databases used for analysing genetic data is critical (Bengtsson-Palme et al, 2017).

5.4.3. EBMs as supportive components for ecological status

EBMs that have established links to BQEs could be used as supportive elements for ecological status. This would mainly be the case for some *in vivo* methods and some biomarkers at higher levels of biological organisation. Supportive quality elements are already used for ecological status within the WFD. Those are physico-chemical elements and hydromorphological elements. Figure 8 shows an example of how ecologically relevant EBMs could be used to assess ecological status by giving the parameters they measure the same weight as physico-chemical elements (including RBSPs). This would need a relatively small change in the legislation (WFD) and in current work flows.

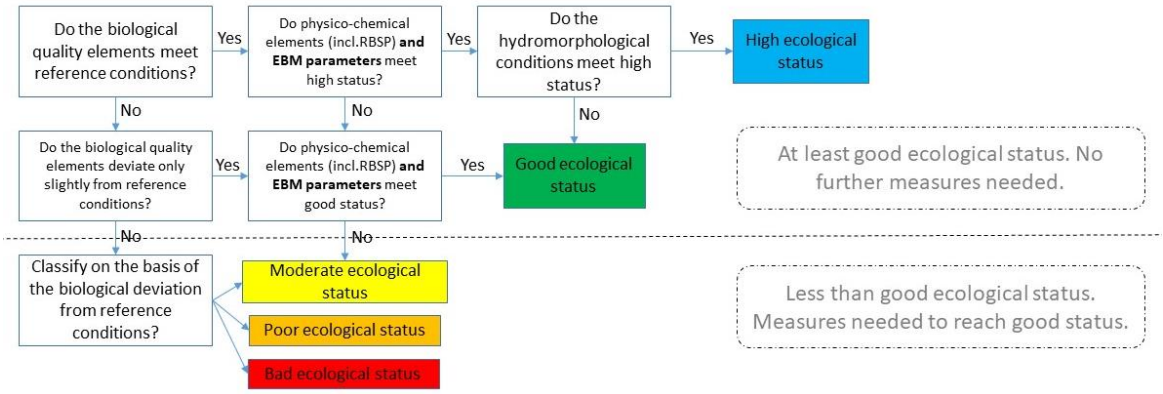

Figure 8. An example of how EBMs could be used as supportive elements to assess ecological status. In the example, EBM parameters are given the same weight as physico-chemical elements (including RBSPs). This means that they could cause a reduction in status to moderate, and thus indicate a need for improvement. The figure is modified from CIS Guidance Document No. 13 (Overall approach to the classification of ecological status and ecological potential).

5.5. EBMs vs pressures and measures (objective 8)

Objective 8 of the ToR

Objective 8: Assess the availability and suitability of investigative approaches for identifying the underlying causes contributing to the overall risks, to identify sources of emissions and facilitate measures.

5.5.1. Identifying water bodies "at risk"

As mentioned in the introduction, the WFD employs the DPSIR (Drivers – Pressures – State – Impact and Response) approach (Pirrone et al. 2005). Therefore, as a starting point, MS need to identify the water bodies that are at risk of failing the WFD objectives ("good status") based on an analysis of pressures including – at least for the PS – an inventory of emissions of individual substances, together with an assessment of impacts. CIS Guidance Document No. 3 was developed to support the MS in performing this analysis of pressures and assessment of impacts. The results from this initial stage are used 1) to optimise monitoring programmes and 2) as the basis for the programmes of measures (PoM). Based on monitoring data, in particular from operational monitoring, the chemical and ecological status of the water bodies is classified. For those water bodies failing any of the WFD objectives (including achieving "good status"), PoMs should be developed with the aim of achieving the objectives (see Figure 9).

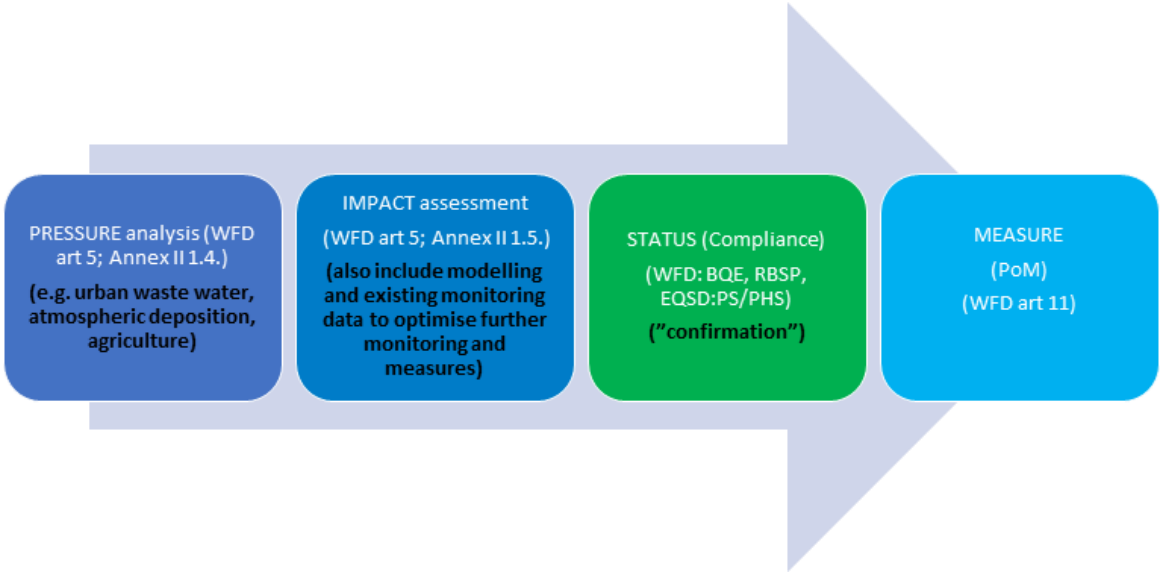

Figure 9. Analysis of pressures and assessment of impacts should guide monitoring efforts and measures. The impact assessment can itself include modelling and monitoring approaches. Although here illustrated as an "arrow", the WFD approach is actually performed in a "6-year-cycle" (iterative process) and monitoring is also performed to assess the effectiveness of the PoM. BQE= biological quality element; RBSP: river basin specific pollutant; PS: priority substance; PHS: priority hazardous substance; PoM: programme of measures.

WFD Annex II 1.5. specifically mentions that also monitoring and modelling can be used to assess the impacts. CIS Guidance Document No. 3 also mentions that for the selection of substances for which EQSs should be developed at national level (RBSPs), the pressures and impacts assessment is an important starting point and, as a safety net to this selection

process, the "presence of pollutants with similar modes of toxic action and hence potentially additive effects" should be taken into account[31].

From the CIS Guidance Document No. 3 it is also apparent that there are numerous ways to identify water bodies at risk and significant pressures, but also to arrive at the list of RBSPs. EBMs could be of value in this context.

5.5.2. Which EBMs?

Most EBMs and all categories thereof (i.e. *in vitro*, *in vivo* and biomarkers) can be used, alongside other methods, to identify water bodies that are subject to significant pressures and thus risk failing the WFD objectives. If effects are observed using EBMs, especially if "severe", impacts are indicated. EBMs are also of interest if there is no obvious reason for an insufficient ecological status of a water body. The use of EBMs can provide insights into the role of chemical contamination. If test results are negative (no observed toxicity) from a battery of sensitive tests, the presense of chemical contaminants cannot be excluded but is less likely to be responsible for the observed ecological effects. The detection of effects by EBMs indicates the likely presence of bioactive uninvestigated compounds.

The selection of EBMs to use in a particular case needs to consider case-specific circumstances. The EBMs need to be sufficiently sensitive and cover the suspected compounds or groups of compounds in a particular case. If the compounds present are largely unknown (not monitored), a battery of EBMs is normally needed. However, also practical aspects need to be taken into account. Costs can be reduced if combining the EBM analysis and sampling with sampling for other purposes. If e.g. biota sampling is planned from the same water body, biomarkers would be a cost-effective approach, whereas if water is analysed, *in vitro* batteries would probably be the first choice. If contaminated sediment is of concern, "in vivo" EBMs could be applied. Also, knowledge about the source/s (type of pressure) is valuable in selecting a suitable EBM or battery of EBMs. If the water body is, e.g. primarily exposed to sewage effluents, the analysis should at least include EBMs that respond to estrogenic substances.

To assess the risk of failing the objectives, less "strict" assessment criteria are necessary than for status classification, although "risk criteria" would be helpful to evaluate EBM results to assess pressures and impacts. However, comparisons between up- and downstream sites, trends, weight-of-evidence approaches and expert judgement on a case-by-case basis (see Section 5.3.) are also possible.

EBMs can also be applied to analyse pressures. *In vivo* and *in vitro* assays are particularly useful here, and can be used also to analyse effluents, leachates etc. that can contain complex mixtures. Whole-effluent analyses are already used routinely for this purpose (see also 3.2.3.).

5.5.3. Identifying measures and assessing their effectiveness

The "chemical approach" so far used has some major advantages. It directly reveals substances that should be targeted. Some "substance-specific" measures include the restriction of particular uses (see e.g. REACH Annex XVII). Some substances can be seen as indicators for a group of compounds, often from the same source or at least type of

[31] See table 3.9. in CIS 3.

source (e.g. dioxins and PAHs from combustion processes and active substances used for plant protection in agriculture). Thus, local measures targeting a particular substance or only a few compounds can also result in a decrease in the load of other chemicals from the same group or same type of application.

Most EBMs do not provide direct information about causative substances and, in fact, several EBMs are used primarily because they respond to many substances with the same or multiple MoAs. They are therefore useful for detecting mixture effects and unknowns, although if effects are observed, further investigation is necessary.

To identify measures where potential pressures are known

An analysis of pressures should normally precede the impacts assessment and status classification (Figure 9). Thus, the main potential contaminant sources, such as sewage treatment plants or industries, to the water body should normally be known already during impact assessment and status classification. By using suitable EBMs (in vivo and in vitro assays) alongside the analysis of "suspect substances" the main problematic pressures can be identified and aid in the identification of cost-effective measures. In cases biomarkers were used in the impact assessment, a corresponding *in vivo* or *in vitro* assay needs to be used. If e.g. intersex has been observed, in vitro assays can be used to test estrogenicity of the effluents from the identified sources (see also Annex 7 to the technical report of 2014).

In many cases it is not necessary to know or regulate emissions of particular substances. In fact, in many cases, the same EBMs that were used to characterize the effluents could probably be used to establish, e.g. toxicity-based emission limit values see e.g. proposal developed within WP 3 of the COHIBA project (Nakari et al 2011). In fact, the United States Environmental Protection Agency (US EPA) started to assess and regulate effluent toxicity for certain installations several decades ago and in the early 90's developed guidance on how to identify the main suspects behind observed effects, using Toxicity Identification Evaluation (TIE) (see e.g. US EPA 1991 a and b). If more detailed information about causative substances is needed to undertake measures, TIE and effects-directed analyses (EDA) can be considered as a second step. EDA and TIE methodologies are further described in the Technical Report of 2014 (European Commission 2014).

To identify measures where pressures are largely unknown

If WFD surveillance monitoring has been conducted using only EBMs, the reasons for the observed effects (i.e., responsible pressures) could still be unknown, leaving it unclear how to respond in terms of operational monitoring and measures. Investigative monitoring, including also "source tracking" (e.g. gradient studies) could in several cases help to identify potential sources and thus suspect substances.

In some cases, if the impact could be suspected to be related to large scale effects, in time and space, checking the same type of impact at other locations as well as the trend would be needed to confirm this (see e.g. Figure 4). If e.g. effect biomarkers were used in a surveillance monitoring program of water bodies that are not (yet) identified to be "at risk" of failing good status, while effects are still observed, additional supportive variables (such as exposure biomarkers, in vitro assays and chemical analyses of substances with relevant MoAs), could provide important clues as to the reasons for the observed response. Other

taxonomic groups could also be investigated, if possible, using the same effect biomarker endpoint to assess the extent of the problem.

6. PROPOSAL – Scenarios that require the application of EBMs in support of the WFD

Objective 9 of the ToR

Objective 9: Assess the practical feasibility and cost effectiveness of implementing at EU-scale possible strategies using effect-based methods, to better take into account mixture risk assessment and mixture risk management under the WFD for relevant MoAs, as far as possible ensuring consistency with other legislation. In particular, this will include a comparison of the advantages/drawbacks of using effect-based tools alongside chemical tools, compared with using only chemical methods as in the current approach to chemicals under the WFD.

From the findings and previous deliverables of the task (Chapter 5) we have elaborated the following proposal on how to support implementation of a more holistic approach to assess toxic substances in a WFD context, related to the first five previously identified applications of EBMs:

1. Cover complex mixtures (of unknown composition) and perhaps even cumulative effects when combined with other stress factors – to assess status and/or identify significant pressure.
2. Cover mixture effects from substances sharing the same MoA – to assess status and/or identify significant pressure.
3. Identify relevant MSFD indicators.
4. Assess sediment quality.
5. Assess status of regulated substances.

This proposal is related to the previous objectives and in line with the last (ninth) objective of the ToR, where also practical feasibility and advantages/drawbacks are to be described.

Cost effectiveness is also to be assessed in the ninth objective of the ToR. However, it is normally not so straightforward to compare prices for an individual EBM to prices for chemical analysis since most EBMs respond to several (types of) substances. Only for those EBMs that respond to a particular substance or small group of substances would a comparison between estimated analytical costs using the EBM approach or the traditional chemical analytical approach be appropriate. However, care should be taken in interpreting even such comparisons. The chemical analysis of "known and well-regulated compounds" is currently performed on a routine basis and this can lower the analytical costs. Routine performance of biomarker analyses is, e.g. generally not (yet) in place, especially if not included in a regular monitoring programme. Nevertheless, information about costs is included for several EBMs in Chapter 5 and Annex II.

In Chapter 6, the focus is on the assessment of feasibility from a technical and scientific point of view. From a more WFD legal perspective, primarily two options are discussed: to assess status (classification) and/or to identify significant pressures and assess impacts (further discussed in Chapter 7).

6.1. EBMs to cover non-monitored substances and mixtures in WFD and MSFD context (applications 1-3)

EBMs are probably the only way to detect the effects of complex mixtures in the environment.

The first two applications - to cover non-monitored substances and mixtures - were considered to be the most important reasons for the use of EBMs in the WFD context and should also be the most important reason to use EBMs in the MSFD context (application 3).

Below, the individual EBMs and EBM batteries so far identified that would be fit for such a purpose are described and their feasibility assessed. First, *in vitro* assays to detect two important MoAs – estrogenicity and genotoxicity – are proposed (6.1.1. and 6.1.2.). These two particular MoAs were chosen because of their biological relevance, implying relevance also to the WFD, and the level of maturity of related EBMs compared to other prioritised *in vitro* assays. The related EBMs can thus be used and evaluated on a routine basis. The biomarkers identified to detect effects from particular MoAs and/or more biological pathways (resulting from complex mixtures and including cumulative effects) for which routine use seems to be possible (today or in the near future) are presented for the marine and freshwater environment, respectively (6.1.3. and 6.1.4.). The methods included were all considered to be mature or relatively mature, based on an assessment of the availability of assessment criteria and/or SOPs. Some of the methods in 6.1.3. are already used as MSFD indicators. Finally, the feasibility of using *in vivo* bioassays to assess mixtures (application 1) is described in Section 6.1.5.

6.1.1. To assess estrogenic activity using *in vitro* assays

Current approach

The risk from estrogenic substances in water bodies is currently assessed using a chemical-analytical, substance-by-substance approach. The presence of three EU WL compounds, 17β-estradiol (E2, natural hormone), 17α-ethinylestradiol (EE2, contraceptivum) and estrone (E1, breakdown product of E2), is quantified by high-resolution mass spectrometry coupled to liquid chromatography (hr-LC/MS) after enrichment by solid phase extraction. The EQS considered for these compounds are 400 pg/l for E2, 35 pg/l for EE2 and 3600 pg/l for E1. Other compounds with estrogenic activity, such as nonyl- and octylphenol, are also included in the EQSD.

This current approach suffers from two limitations:

- the review of the 1st WL under the WFD (Loos et al. 2018) shows that a number of MS are not able to quantify these three compounds at EQS levels due to insufficient LOQ, in particular for EE2. The ability to detect E1 and E2 at levels below EQS was better (16 out of 23 MS for E1 and 16 out of 25 for E2).

- it is well known that further compounds with estrogenic activity are present in the environment and that all these agonists of the ER act in a mixture according to the concept of concentration addition (Kortenkamp 2007 and Kortenkamp et al. 2009). Annex III.1

presents strong evidence that the monitoring of E1, E2 and EE2 alone is insufficient to assess the overall risk of estrogenic endocrine disruption from the presence of ER agonists in water.

Usefulness of EBMs

The activation of the ER by ER agonists is a relevant mode of action that is related to adverse effects at the population level (Kase et al. 2018, Könemann et al. 2018). As outlined above, an EU-wide comprehensive assessment of the WL compounds E1, E2 and especially EE2 is not feasible in the current situation. Based on the results presented in Section 5.3.4, *in vitro* EBM together with respective EBTs would be able to discriminate between a sufficient and insufficient chemical status with respect to E1, E2 and EE2 with sensitivities and specificities near 90%.

In vitro EBMs for the detection of estrogenicity can be readily used for trend monitoring, status assessments, prioritisation of water bodies, identification of sources and investigative monitoring.

Added value of using EBMs

The added value of *in vitro* EBMs is illustrated by their current use, to e.g. screen for estrogenicity in different types of water sample. *In vitro* test batteries, including estrogenicity assays, are frequently used for screening purposes with various types of sample, including effluents from waste water treatment plants (see also Practical feasibility).

Mixture effects from known and unknown compounds with estrogenic potential can be assessed in an integrated manner. The related *in vitro* EBMs measure the overall estrogenic activity present in a mixture of ER agonists. The response is not restricted to a limited number of selected compounds and thus provides a more comprehensive view on the presence of this unwanted effect in surface waters.

The assessment of ER activation by a number of *in vitro* EBMs is well established and there is strong evidence that this molecular initiating event is linked to adverse outcomes at higher biological levels. A study by Arlos et al. (2018) demonstrated the correlation of predicted concentrations of known estrogens expressed as total estrogenicity (E2 equivalent concentrations) with key estrogenic responses such as intersex in the rainbow darter. Therefore, *in vitro* EBMs detecting the activation of the ER would allow a holistic assessment of estrogenic potentials in water samples.

The potential application field of these EBMs is not limited to the WFD context. The use of these EBMs is meaningful also in the context of water reuse and for the assessment of urban waste water. They can be used to identify pressures and potential risks to water bodies and to trace and regulate sources retroactively if effects are observed. Thus, these EBMs can contribute to improved water management in Europe.

Current regulatory context and use of EBMs

Annex VIII to the WFD identifies compounds with “endocrine-related functions” as being among the main pollutants of European water bodies, indicating the relevance of this biological effect. EQS-proposals were developed for E2 and EE2 at EU level in preparation for the 2013 revision of the EQSD, but the substances were not included. Instead, they

were added to the WL. They can also be regulated as RBSPs in individual MS. So far there is no explicit use of *in vitro* EBMs in the WFD, but an EU monitoring project connected to the WL has been performed to assess the applicability of using different types of *in vitro* assays for screening purposes, to identify samples that can be prioritised for further chemical analysis of estrogenic compounds.

Guidance needed?

By analogy with the EQS derivation guidance (CIS Guidance Document No. 27), guidance is needed on how to develop EBTs. Also, a standard (SOP) on suitable pretreatment[32] of surface water samples would be beneficial. Such guidance could be developed based on scientific literature and experiences obtained by the EU estrogen monitoring project. A tiered calculation of EBTs with increasing knowledge could be achieved. For the investigated EBMs in the EU estrogen monitoring project a very high specificity and sensitivity was shown (see Chapter 5 and Annex III). For other EBMs addressing the activation of the ER this could be done too. If EBMs could be used for WFD status assessment, revision of EBTs with each water management cycle of 6 years would allow an update of existing EBTs with increased knowledge and the development of EBTs for *in vitro* EBMs not yet covered by the particular assays presented in Annex III.1.

Practical feasibility

A number of *in vitro* EBMs are available that directly detect the potential of a sample to activate the ER. Three international standards for the determination of the estrogenic potential of water and waste water are published (ISO 19040 parts 1 to 3 - 2018). Three methods (ERα-CALUX, A-YES and YES) successfully passed an international interlaboratory trial.

The detection of estrogenic substances is possible at low E2-levels. Taking a sample enrichment of 10 into account that can be easily performed by solid phase extraction, the sensitivities for the human cell-line-based reporter gene assay and the A-YES are sufficiently below EBTs to facilitate the classification of water bodies with respect to their contamination with estrogenic compounds. Variabilities in all three EBMs were below 50% (see Table III.9).

ER-CALUX and A-YES are available as commercial products. In addition, license-free versions of this type of assay are available, e.g. using the cell line T47D also covered by the international interlaboratory trial. Within ISO 19040-3, validity criteria are defined to cover further cell-line-based reporter gene assays.

Commercial costs for this type of EBM are about 140-200 € per sample ((if performed in-house the costs for personnel and consumables are around 60 Euro/sample). Commercial and non-commercial EBMs for the detection of estrogenicity are available, however establishment of a dedicated cell-culture facility is required.

Conclusions and recommendations

In vitro EBMs for the detection of the ER activation cover a relevant MoA. SOPs for this type of *in vitro* EBMs are available and three assays are even ISO standardised. Further validation and interlaboratory studies for other bioassays evaluating effects by estrogenic compounds would provide a wider choice of methods.

[32] Currently an enrichment factor of 1000 was most appropriate both different European for surface and waste water assessments, the water phase concentration was afterwards back calculated.

Short-term outlook and recommendations for possible further implementation under the WFD

Annex VIII to the WFD identifies the substances that these EBMs respond to[33]. Consideration could be given to allowing the use of *in vitro* EBMs for the assessment of the presence of substances causing effects on endocrine-related functions. However, field studies should be performed to investigate the potential of these EBMs to identify sources of emissions as a basis for subsequent measures for improvement. *In vitro* EBMs combined with suitable EBTs can reliably screen water samples for further chemical analysis.

Medium term outlook (next mandate)

In vitro EBMs for the detection of ER activation might be included in future WL cycles after the development of guidance documents and an interlaboratory comparison of suitable EBMs. A field study should be performed to demonstrate the potential of these EBMs to be used for source identification if elevated levels of estrogenicity are found. This would also demonstrate that an observed effect can be linked to a pressure.

6.1.2. To assess genotoxic activity using *in vitro* assays

Current approach

The risks from genotoxic substances in water bodies are currently assessed using a chemical-analytical, substance-by-substance approach. Some compounds with mutagenic properties, such as PAHs and benzene, are included in the EQSD.

Usefulness of EBMs

The assessment of genotoxicity is a key component of the evaluation of surface water quality. Numerous EBMs permit the evaluation of genotoxicity, i.e. damage to the genetic information within a cell through the interaction of a genotoxic substance with the DNA sequence or structure, potentially leading to mutations (mutagenicity), and further to cancer (carcinogenicity). For the latter reason, the use of EBMs specific for this MoA is fundamental for the protection of human health, considering among other things the exposure of humans to genotoxic substances in drinking water.

Added value of using EBMs

Mutagenicity tests are predictive of integral mutagenic/carcinogenic activity, and can evaluate the combined action of potentially hazardous compounds present, e.g. in drinking water as complex mixtures and not only individual compounds. They are able to take into consideration the synergism, additivity or even antagonism of substances. The extraction method is also very important for this type of assay.

Current regulatory context and use of EBMs

[33] Point 4 in WFD Annex VIII ("INDICATIVE LIST OF THE MAIN POLLUTANTS") reads: "Substances and preparations, or the breakdown products of such, which have been proved to possess carcinogenic or mutagenic properties or properties which may affect steroidogenic, thyroid, reproduction or other endocrine-related functions in or via the aquatic environment."

Annex VIII of the WFD identifies compounds "that possess carcinogenic or mutagenic properties" as among the main pollutants for European water bodies, indicating the relevance of this biological effect. So far, there is no explicit use of *in vitro* EBMs in the WFD context.

Guidance needed?

A brief guidance document to define the most suitable tests for a particular application and the extraction method to be applied would be useful.

Practical feasibility

In vitro EBMs for the detection of mutagenic and clastogenic potentials are used under REACH (Council Regulation (EC) No 440/2008).

Mutagenicity tests are rapid, relatively cheap and have the potential for automation and thus high-throughput screening.

Several EBMs can be used to assess genotoxicity in the presence or absence of an external metabolic activation, e.g. by the use of S9-mix in the Ames or micronucleus (MN) tests, Comet assay, P53 assay, SOS-umu test, SOS-chromo test and others.

Below are described some examples.

The *Salmonella typhimurium*/mammalian microsome assay (Ames) is the most widely used short-term test to identify genetic damage and to assess the mutagenic potential of compounds and mixtures. The Ames test employs several histidine-dependent auxotrophic mutant strains of bacteria (*Salmonella* spp., *E. coli*) to detect several types of mutations that occur upon exposure to toxicants, e.g. substitutions, additions, or deletions of one or several DNA nucleotides. This test is based on the principle of reverse mutation or back mutation, so it is also known as the bacterial reverse mutation assay.

The Ames test has many advantages, it is a very versatile assay, its different modifications have been developed to determine mutagenic potencies, and it is recommended by several agencies, e.g. the German Institute for Structural Engineering (Deutsches Institut für Bautechnik – DIBt) for the assessment of construction material. A positive result from the test will indicate that the chemical is mutagenic and therefore may act as a carcinogen, as cancer is often linked to mutation. The response in the Ames test is the result of the effect of the whole mixture of (geno)toxic compounds potentially present in drinking water samples or in other environmental samples. The Ames test and the Ames Fluctuation Test are standardised according to ISO (ISO 16240:2005, ISO 11350:2012)

The MN technique is a useful tool to investigate the ability of substances to interfere with chromosome structure and function, and has been included in the OECD guidelines for chemical testing[34]. Performed on actively dividing cells (bone marrow or erythrocytes), it allows for a rapid and reliable assessment of chromosomal aberrations through the measurement of the frequency with which stained micronucleated cells are observed under a microscope. Micronuclei are formed after exposure to genotoxic compounds or stressful conditions whenever a chromosome or its fragment resulting from incorrectly repaired/unrepaired DNA break is not incorporated into a daughter nucleus during cell division. The MN assay is standardised according to ISO (ISO 21427-1:2006, parts 1 and 2).

[34] OECD/OCDE Guideline for the testing of chemicals 474, adopted 29 July 2016.

The Comet assay, also known as the single-cell gel electrophoresis (SCGE) assay, is a highly sensitive EBM applied to detect DNA damage at the level of single cells, and is widely used in biomonitoring. The effects of genotoxic events in cells exposed to chemicals are evaluated based on the pattern formed by DNA fragments. Undamaged DNA maintains its compact structure which does not allow it to migrate on the agarose gel; in contrast, genotoxic substances may disrupt this organisation enabling DNA fragments to move faster towards the positively charged anode. The assay has been successfully introduced to assess sediments and surface water in The Netherlands. The MN and Comet assays are also frequently used as biomarkers (see e.g. 6.1.3.).

Conclusions and recommendations

Short term outlook and recommendations for possible further implementation under the WFD

Genotoxicity/mutagenicity tests could be included in the WFD for screening purposes and for compliance checking. EBTs could be unnecessary because the qualitative detection of mutagenicity in a waterbody is a signal (Yes/No) of status. Therefore, the establishment of a threshold is of limited practical utility after definition of a suitable enrichment procedure.

6.1.3. To assess ecologically relevant effects from chemical mixtures in the marine environment using biomarkers

Current approach

Mixture effects from complex toxic mixtures and/or substances sharing the same MoA are not taken into account with the current WFD approach.

In the MSFD context, EBMs can be included but on an optional basis.

Usefulness of EBMs

Several EBMs are available to take effects from both complex mixtures and the combined effects from substances having the same MoA into account (see table in Section 5.2). Based on this assessment, a set of biomarkers is proposed for the assessment of ecologically relevant effects from complex chemical mixtures (**application 1**) and for mixtures of substances having the same MoA (*application 2*) in the marine environment, in table 7.

The maturity of each method is briefly described, based on the availability of SOPs, assessment criteria and whether it is already included in regular national monitoring. All can be considered usable for MSFD assessments (**application 3**), whereas two would be less applicable to WFD assessments.

Table 7. Biomarkers proposed for the assessment of ecologically relevant effects from complex chemical mixtures and effects from mixtures of substances having the same MoA in the marine environment.

Biomarker name	Remark
LMS, lysosomal membrane stability	Applicable under both WFD and MSFD to take complex mixtures into account Mature method: EACs and SOP established and already included in monitoring
DNA adducts	Applicable under both WFD and MSFD to assess genotoxicity Relatively mature method: EACs and SOP established, monitored but could not be confirmed to be performed on a regular basis at national level
FDI, Fish Disease Index (including LH and MLN)	Applicable under both WFD and MSFD to take complex mixtures into account Mature method: EACs and SOP established and already included in monitoring
Reproductive success in eelpout	Applicable under both WFD and MSFD to take complex mixtures into account Mature method: EACs and SOP established and already included in monitoring
VTG in male fish	Applicable under both WFD and MSFD to assess estrogenicity. Mature method: EACs and SOP established and already included in monitoring
Intersex in male fish	Applicable under both WFD and MSFD to assess estrogenicity Relatively mature method: assessment criteria proposed but not yet formally adopted (?), scientific publication refered to for method description, monitored but could not be confirmed to be performed on a regular basis at national level.
Micronucleus	Applicable under both WFD and MSFD to take genotoxic effects into account Mature method: EACs and SOP established and already included in monitoring
Amphipod embryo malformation (brackish water)	Applicable under both WFD and MSFD (but primarily Baltic Sea) to take complex mixtures into account Mature method: EACs and SOP established and already included in monitoring
AChE	Applicable under both WFD and MSFDto take AChE induction into account. Mature method: EACs and SOP established and already included in monitoring
Comet Assay	Applicable under both WFD and MSFD to assess genotoxicity. Relatively mature method: EACs and SOP established, monitored but could not be confirmed to be performed on a regular basis at national level.
Mussel histopathology (gametogenesis)	Applicable under both WFD and MSFD to take complex mixtures into account Relatively mature method: EACs established but a SOP may not formally be established, monitored but could not be confirmed to be performed on a regular basis at national level.
Stress on stress	Applicable under both WFD and MSFD to take complex mixtures into account Relatively mature method: EACs established but a SOP may not formally be established, monitored.
SfG, Scope for Growth	Applicable under both WFD and MSFD to take complex mixtures into account Mature method: EACs and SOP established and already included in monitoring

Biomarker name	Remark
Sea eagle productivity	Primarily applicable under MSFD to assess secondary poisoning and to take complex mixtures into account Mature method: EACs and SOP established and already included in monitoring
Pregnancy rate in seals	Primarily applicable under MSFD to assess secondary poisoning and to take complex mixtures into account Mature method: EACs and SOP established and already included in monitoring

Please note that imposex and egg-shell thinning are not included in the table above. This is because their added value is not related to the possibility of detecting mixture effects. Instead, they are included in Section 6.3. Also, ALA-D is not included, being too specific to be useful in this context. Benthic diatom malformation and mentum malformation in chironomids are also not included because they are not applicable to the marine environment. Nevertheless, marine diatoms have been considered to detect effects from toxic substances[35].

Added value of using EBMs

Using the biomarker battery identified above, effects from mixtures can be detected in fish and mussels as well as at higher trophic levels, e.g. in top predators (seals and sea eagles). Most can respond to very complex mixtures, while some respond to substances having a particular MoA (e.g. estrogenicity or genotoxicity, both of high relevance to the health of the organism); see also 6.1.1. and 6.1.2. for a corresponding *in vitro* approach to assess these pathways.

Most of the biomarkers can probably be used under the MSFD and WFD, although the last two (sea eagle productivity and pregnancy rate in seals) would be less appropriate to assess status at water-body scale, since they reflect larger geographical-scale exposure.

Current regulatory context and use of EBMs

The biomarkers identified above could not be used for WFD status classification at present because WFD Biological Indices (according to WFD Annex V) should reflect effects at population level.

Guidance needed?

Guidance would be helpful; see below.

Practical feasibility

For most of the biomarkers in Table 7, assessment criteria are available or under development and although these criteria may need to be checked in more detail, some are already included in national legislation in relation to the MSFD[36] and used in assessments under RSCs.

[35] The marine diatom *Thalassiosira pseudonana* has e.g. been used as a model organism to assess the effects of PAH exposure at the molecular level, either to single compounds or to mixtures (Bopp and Lettieri 2007). Specific pathways are affected after PAH exposure and one of the major pathways is the silicification process. Genes encoding for proteins involved in silica uptake and metabolism, such as the silicon transporter and silaffin 3, could be suitable molecular biomarkers of exposure to the PAHs (Carvalho et al. 2011).

[36] See e.g. Swedish national regulation HVMFS 2012:18.

Most of the biomarkers in Table 7 are already included in national monitoring programmes but for some it was not possible to confirm this.

Monitoring costs can be reduced if combined with chemical analysis of biota (see also Section 5.2. and the table in Annex II for more detailed information).

Conclusions and recommendations

EBMs are usable under the MSFD on a voluntary basis. The above biomarkers could be considered for use in MSFD assessments and monitoring programmes. For those MS that decide to use them for MSFD assessments, it should be possible, for the sake of harmonisation, to take those assessments into account in the WFD status assessments, as long as they are appropriate/applicable also to the WFD context. In this case, the same assessment criteria/threshold values should probably be used in the two assessments. However, in MS also having access to data for additional biomarkers providing more information on status, the assessment scheme in Section 5.3 based on scoring and weight-of-evidence could also be considered.

An option could be to allow MSFD assessment to trigger WFD monitoring. If "non-compliance" is observed and the reasons are not known but land-based sources are suspected, this should trigger an investigative monitoring approach, in line with the WFD monitoring requirements, but also continued (surveillance) monitoring (see also section 5.5.3.).

Short-term recommendations for possible further implementation under the WFD

Consider, if possible, adding a link between the "MSFD initial assessments" and WFD surveillance and investigative monitoring needs by opening the possibility to take results from the MSFD assessment into account in the WFD assessment of the marine environment. This would likely require slight revision of WFD Annex V. A somewhat different definition of the BQEs would also be possible to, e.g. allow ecologically relevant biomarkers to be used, which could be considered also for the definition of marine BQEs. See also the more general suggestion in section 5.4.3. on the use of EBMs as supportive elements.

Medium term (next mandate)

Develop a guidance document on how to proceed if "non-compliance" is observed and the reasons are not known. A proper first step would likely be to assess the geographical scale of the response observed by using additional monitoring sites (at WFD water-body level within these areas but also targeting identified potential local sources) (see also Section 5.5.).

A guidance document would also be useful to support and promote the use of all available information (such as data for other biomarkers and trends) in the assessment of status (in both the MSFD and WFD context). The values and scoring procedures suggested in Sections 5.2. and 5.3 would likely aid in the interpretation of the data, make it more transparent and also facilitate taking other data into account, without applying the OOAO principle in relation to biomarkers for which this would be inappropriate.

For some of the methods above, especially those considered to be "relatively mature", a revised assessment could be done to take the specific aspects identified for each method into consideration. For some methods, intercalibration exercises could be necessary, as well as more formal adoption of SOPs.

Finally, the development of formal guidance on how to establish assessment criteria for biomarkers, possibly by using and/or adapting the current framework used for BQEs, could be considered.

6.1.4. To assess ecologically relevant effects from chemical mixtures in the freshwater environment using biomarkers

Current approach

Mixture effects from complex toxic mixtures and/or substances sharing the same MoA are not taken into account in the current WFD approach.

Usefulness of EBMs

Several EBMs are available to quantify effects from both complex mixtures and the combined effects from substances having the same MoA (see table in Section 5.2). Based on this assessment, a set of biomarkers is proposed for the assessment of ecologically relevant effects from complex chemical mixtures (application 1) and from mixtures of substances having the same MoA (application 2) in the freshwater environment (Table 8).

The maturity of each method is briefly described, based on the availability of SOP, assessment criteria and whether it is already included in regular national monitoring.

Table 8. Biomarkers proposed for the assessment of ecologically relevant effects from complex chemical mixtures and from mixtures of substances having the same MoA in the freshwater environment.

Biomarker name	**Remark**
Benthic diatom malformation	Applicable to take complex mixtures into account, in particular from metals Mature method: SOP established, assessment criteria available to asses "risk" (significant pressures) and monitored at least in campaigns.
Mentum deformation in chironomids	Applicable to take complex mixtures into account Relatively mature method although it is uncertain whether any formal SOPs or assessment criteria have been established. Background levels of mentum malformations have been described in the scientific literature and informal "criteria" are frequently used to assess effects at contaminated sites.
LMS, lysosomal stability	Applicable to take complex mixtures into account Relatively mature method: SOP established and the EACs established should be possible to implement also for freshwaters, but unclear whether included in any freshwater monitoring programmes.
DNA adducts	Applicable to assess genotoxicity. Relatively mature method: SOP established, but unclear whether EACs could be implemented for freshwaters, and unclear whether included in any freshwater monitoring programmes.
FDI, Fish Disease Index (including LH and MLN)	Applicable to take complex mixtures into account Relatively mature method: SOP established, but unclear whether EACs could be developed or implemented for freshwaters, and unclear whether included in any freshwater monitoring programmes.

Biomarker name	Remark
VTG in male fish	Applicable to assess estrogenicity. Relatively mature method: SOP established, but unclear whether assessment criteria could be implemented for freshwaters, and unclear whether included in any freshwater monitoring programmes.
Intersex in male fish	Applicable to assess estrogenicity Relatively mature method: unclear whether assessment criteria proposed (not yet formally adopted) could be implemented for freshwaters, and unclear whether included in any freshwater monitoring programmes. Scientific publication referred to for method description.
Micronucleus	Applicable to take genotoxic effects into account Relatively mature method: SOP established, but unclear whether assessment criteria could be implemented for fresh waters, and unclear whether included in any freshwater monitoring programmes.
Amphipod embryo malformation	Applicable to take complex mixtures into account Relatively mature method: SOP established, but unclear whether EACs could be implemented for fres waters; monitored on occasion in freshwaters but with uncertain regularity.
AChE	Applicable to take AChE induction into account. Relatively mature method: SOP established, but unclear whether EACs could be implemented for freshwaters, and unclear whether included in any freshwater monitoring programmes.
Comet Assay	Applicable to assess genotoxicity. Relatively mature method: SOP established, but unclear whether assessment criteria could be implemented for freshwaters, and unclear whether included in any freshwater monitoring programmes.

Please note that ALA-D is not included in the table above, being too specific to be useful in this context.

Added value of using EBMs

Using the biomarker battery identified above, effects from mixtures occurring in fish, invertebrates and diatoms can be detected. Some can respond to very complex mixtures while some respond to substances having the same MoA (e.g. estrogenicity and genotoxicity, both of high relevance to the health of the organism).

Most of the biomarkers in Table 8 could probably be useful in the WFD context, but at the moment the level of maturity is uncertain or lower for freshwater biomarkers than for marine biomarkers. Nevertheless, the most mature freshwater biomarker EBM included in the table is the assessment of benthic diatom malformations, and several fish biomarkers could be used in both freshwater and marine contexts.

Current regulatory context and use of EBMs

The biomarkers identified above could not currently be used for WFD status classification because WFD biological indices (according to WFD Annex V) should reflect effects at population level.

Guidance needed?

Guidance would be useful; see below.

Practical feasibility

For several of the biomarkers above, the applicability of available assessment criteria to freshwater organisms needs to be checked and discussed in more detail.

Monitoring costs can be reduced if combined with chemical analysis of biota (see also Section 5.2. and Table II.3. on practical implementation aspects).

The diatom malformation method could be considered especially promising to use in combination with current BQE assessments (while assessing ecological effects from eutrophication) to minimise costs.

Conclusions and recommendations

Short-term recommendations for possible further implementation under the WFD

Consider, as a first step, promoting the use of any of the above methods, as appropriate, in the analysis of pressures and assessment of impacts under the WFD, and for investigative monitoring, to take complex mixtures and combined effects from substances having the same MoA (e.g. estrogenicity, genotoxicity) into account. Although it would be possible to use such an approach already today, a minor clarification in WFD Annex II would be appropriate to further promote the use of EBMs in that context.

Medium term (next mandate)

A guidance document would be useful to help MS identify a suitable battery of methods depending on the types of pressure, and to facilitate the assessment of the results. In the longer term, additional "risk indicator" values could be developed (already in place for diatom malformation, see Annex II). To use the biomarkers also in status classification, additional guidance would be needed on how to assess the results (see Chapter 5).

6.1.5. To assess relevant effects from chemical mixtures in the freshwater environment using a battery of *in vivo* assays

Current approach

There is currently no assessment of cumulative effects from complex mixtures under the WFD.

Usefulness of EBMs

The use of *in vivo* assays provides the possibility to detect effects in the environment caused by mixtures of pollutants. The limitation is that it is difficult to determine which substances have caused the effects, but EBMs can be used in this context as screening tools. The use of "*in vivo*" assays should include at least three trophic levels (for example algae, crustaceans, fish embryos) in order to cover the main exposure routes in an aquatic ecosystem.

Some *in vivo* EBMs such as the FET (Fish embryo toxicity test) allow different types of acute and chronic effects to be detected, including possible mixture effects. The fish embryo toxicity (FET) test is the alternative approach to classical acute fish toxicity testing since the latter is not compatible with most current animal welfare legislation. Furthermore, through the FET, different endpoints (lethal and sublethal) can be measured simultaneously, giving a more detailed description of the MoAs of contaminants such as those underlying embryotoxicity or teratogenicity. The effects detected are all relevant to human health.

Alternative approaches have also been developed to use fish cell lines (e.g. gills) instead of fish. Predictability might be equally good. FET should be restricted to 96 h. The Fish Egg Test DIN EN ISO 15088 (48 hours) provides results that correlate well with fish toxicity, and is already used in e.g. Germany for testing of waste water.

Usually, different *in vivo* tests at different trophic levels, coupled with *in vitro* methods, can be used to detect complex mixtures. An example of a test battery was applied in the SOLUTIONS Project (see Annex V).

Added value of using EBMs

In vivo assays could be used as screening tools to identify waterbodies at risk, and for waste water assessment.

Current regulatory context and use of EBMs

In vivo assays could not be used for WFD status classification at present because the WFD BQEs (according to WFD Annex V) should reflect effects at population level.

Guidance needed?

Guidance would be useful.

Practical feasibility

One example in this context is the FET test using zebrafish. It has several advantages: short life cycle of zebrafish, good reproduction rate, high egg fertility, external fecundation, transparency of embryos, rapid embryo development, low costs and reduced spaces for the reproduction and analysis.

Conclusions and recommendations

Short term outlook and recommendations for possible further implementation under the WFD

To perform at least one *in vivo* assay per trophic level (e.g. zebrafish embryos, algae and crustaceans) once a year at representative sampling points in each river basin in order to detect effects that may have not been identified by chemical analysis or other EBMs.

6.2. Assess sediment quality (application 4)

6.2.1. *In vivo* assays to assess sediment quality

Current approach

The EQSD states that MS may derive EQS at national level for PS in sediment, but that the level of protection must be at least the same as provided by the EQS in the Directive[37].

Bioassays are mentioned in CIS Guidance Document No. 27 as a potential tier 2 step in the evaluation of sediment quality (see also "Current regulatory context").

Usefulness of EBMs

The monitoring and assessment of sediment quality using EBMs is important to detect toxicity from, e.g. the complex mixtures frequently occurring in contaminated sediments, and/or to confirm toxicity when the bioavailability of contaminants in sediment is uncertain or the chemical assessment criteria are uncertain.

Added value of using EBMs

Different *in vivo* assays are already frequently used in the assessment of contaminated sites and of dredged material. Since different species could be sensitive to different substances or groups of substances, it is important to use test batteries to cover a large number of compounds and take mixture effects into account. Test batteries typically include bioassays using algae, crustaceans and other invertebrates (see also the table in Annex II, which contains several *in vivo* bioassays applicable to sediments or extracts thereof). Also, some biomarkers are available to assess sediment quality.

At contaminated sites, EBMs are frequently used in combination with chemical and biological methods in the TRIAD approach described in Section 5.3.2.

Current regulatory context and use of EBMs

CIS Guidance Document No. 27 foresees the use of EBMs when sediment EQS are exceeded, in particular in situations where the EQS is considered "uncertain", e.g. when high assessment factors have been used or when the EQS is based on recalculation of a water EQS to sediment using the equilibrium partitioning approach. In the latter case, the sediment EQS is usually based not on toxicity to benthic organisms but on toxicity to pelagic organisms. and there might be differences in sensitivity.

Guidance needed?

Guidance should propose different test batteries depending on, e.g. the type of pressure (suspected group of compounds) and environment (freshwater, brackish or marine) and purposes (to confirm EQS exceedance of a particular substance or to detect effects from mixtures).

Practical feasibility

[37] The first two paragraphs of Article 3 of the Directive read: "Member States may opt, in relation to one or more categories of surface water, to apply an EQS for a matrix other than that specified in paragraph 2, or, where relevant, for a biota taxon other than those specified in Part A of Annex I.

Member States that make use of the option referred to in the first subparagraph shall apply the relevant EQS laid down in Part A of Annex I or, if none is included for the matrix or biota taxon, establish an EQS that offers at least the same level of protection as the EQS laid down in Part A of Annex I.

Only a few standardised *in vivo* bioassays are directly applicable to whole sediment testing (see tables in Annex II). However, several assays/toxicity tests can be applied to liquid matrices obtained from sediment, such as the elutriate and/or the pore water (SETAC, 1993). A sediment elutriate is an environmental matrix that enables the replication of sediment mobilisation phenomena and the prediction of the release of contaminants from the sediment to the water column. The use of in vivo EBMs can be included to evaluate the potential effects of disposing of dredged material (see e.g. the Italian Ministerial Decree 173/2016) in open water and is nowadays also applied to the quality evaluation of *in situ* sediment (contaminated sites).

Conclusion

Short-term outlook and recommendations for possible further implementation under the WFD

Consider, as a first step, promoting the use of *in vivo* bioassays in the analysis of pressures and assessment of impacts under the WFD, and for investigative monitoring, including to take complex mixtures into account. Although it would be possible to use such an approach already today, a minor clarification in WFD Annex II would be appropriate to further promote the use of EBMs in that context.

A battery of *in vivo* bioassays to evaluate sediment quality could also be considered. For marine testing, coordination with methods already used or recommended within RSC assessments would facilitate the harmonisation of assessments.

Medium term and next mandate

A guidance document would be useful to help MS identify a suitable battery of tests depending on the types of pressure, and to facilitate the assessment of the results.

6.3. EBMs to assess status of regulated substances (application 5)

6.3.1. To assess dioxin activity

Current approach

Dioxins and dioxin-like PCBs are listed as PS in Annex X to the WFD. Annex X defines seven polychlorinated dibenzo-para-dioxins (PCDD), ten polychlorinated dibenzofurans (PCDF) and twelve dioxin-like PCBs (dl-PCBs) to be quantified as a sum parameter. The concentrations of individual congeners are multiplied by the respective toxic equivalence factor (TEF) defined by the WHO (Van den Berg et al. 2005) and summed to a total toxic equivalent (TEQ) of the sample that must not exceed 0,0065 µg/kg in biota. The chemical analysis of this group of chemicals is complex because of the necessary sample clean-up and instrumental requirements.

Usefulness of EBMs

Regulation (EC) No 1881/2006 sets maximum levels for certain contaminants in foodstuffs including dioxins and dl-PCBs, and the current EQS in the EQSD is based on such a value. Analytical methods and requirements for the control of these compound classes are specified in Commission Regulation (EU) 2017/644, including the option to use specific EBM to screen samples for the presence of PCDDs, PCDFs and dl-PCBs. The screening assays aim to detect samples exceeding a defined action level triggering a

chemical analysis of the sample. Due to the physico-chemical properties of this compound class they are adsorped to particulate matter and accumulate in sediments and biota.

In the 2011 EQS dossier for dioxins and dioxin-like PCBs, the relevance of biomarkers and other EBMs is specifically mentioned. Eichbaum and coworkers (Eichbaum et al. 2014) summarised different *in vitro* bioassay applications for detection of dioxin-like compounds and considered the comparability of TEQs derived from bioassay results and chemical analysesobtained using various approaches for various matrices and samples, from single reference materials and compound mixtures to more complex samples such as sediments.

The Micro EROD protocol was recently described in Nature Protocols (Schiwy et al. 2015). Eichbaum et al. (2018) reported on an intra- and inter-laboratory comparison study between four independent laboratories. A bioassay battery consisting of RTL-W1 (7-ethoxy-resorufin-O-deethylase; EROD), H4IIE (micro-EROD), and H4IIE-luc cells was used to assess aryl hydrocarbon receptor-mediated effects of sediments from two major European rivers, differently contaminated with dioxin-like compounds. Each assay was validated by characterisation of its limit of detection (LOD) and quantification (LOQ), z-factor, reproducibility, and repeatability. Dioxin-like compound concentrations were measured using high-resolution gas chromatography/high-resolution mass spectrometry (hr-GC/hr-MS) and compared to bioassay-specific responses via TEQs at intra- and inter-laboratory levels. The micro-EROD assay exhibited the best overall performance among the bioassays.

Added value of using EBMs

As described, dioxins and dl-PCBs consist of a complex mixture of different congeners. The overall toxic potential of this mixture is assessed via the TEQ as described above. *In vitro* EBMs that determine the activation of the aryl-hydrocarbon receptor (AhR) express the overall contamination of a given sample in terms of a BEQ using 2,3,7,8-TCDD as a reference compound. By this means the BEQ directly reflects the overall toxic potential of the sample and captures mixture effects.

Current regulatory context and use of EBMs

Concepts described in Regulations (EC) No 1881/2006 and (EU) 2017/644 could be used in the context of the WFD.

Guidance needed?

Action values for the WFD context have to be defined and EBM calibrations (BEQ against TEQ) have to be performed. In addition, the use of passive sampling should be elaborated as an alternative to biota sampling. If an EQS is exceeded, chemical analysis could be considered in cases where, e.g. information about individual congeners or sources is needed to identify suitable measures through “fingerprinting”.

Practical feasibility

EBMs for the analysis of dioxins and dl-PCBs in foodstuff are already implemented in EU regulations and used efficiently in practice. Costs for using the *in vitro* assay approach are generally lower compared to chemical analysis of dioxins, furans and dl-PCBs. The results are expressed in TEQ and the underlying concept can be transferred to the WFD because the WFD defines EQS for these compound classes in biota.

Standardisation activities for EBMs to detect dioxin-like effects are under discussion in ISO TC147 / SC5.

Conclusions and recommendations

Short-term outlook and recommendations for possible further implementation under the WFD

By analogy to Commission Regulation (EU) 2017/644, the WFD might allow the use of specific *in vitro* EBMs for screening purposes to analyse levels of dioxins and PCBs in biota.

Medium term (next mandate)

Development of guidance, if needed.

6.3.2. To assess TBT effects (imposex) in the marine environment

Current approach

The TBT EQS (expressed for water) in the EQSD was developed to protect gastropods against adverse effects, and the most critical endpoint is imposex. In the MSFD context, ten MS have also adopted the biomarker imposex to assess TBT-related effects in the marine environment, but under the WFD only a chemical approach is to be used. TBT accumulates in sediment and whereas the EQS in the Directive is expressed for water some MS have adopted a sediment EQS. OSPAR contracting parties are also obliged to monitor imposex as well as TBT in sediment or biota (but not in water).

Usefulness of EBMs

The biomarker imposex is clearly WFD-relevant because the effects are of high ecological relevance and they can be linked to TBT exposure.

Added value of using EBMs

Since imposex is frequently monitored in the marine environment and taken into account in the MSFD assessment, for harmonisation purposes and cost effectiveness it would be useful to also take imposex into account in the marine WFD context when assessing the impact of TBT on status.

Imposex analyses provide a time-integrated reponse. Effects are irreversible and therefore monitoring frequency can be lower than for water samples. Furthermore, the gastropods studied in the imposex analysis are exposed to TBT in both the aquatic and sediment phase and an imposex response obviously indicates that the substance is sufficiently bioavailable to cause severe effects.

Furthermore, assessing the impact of TBT on status based only on water sampling is likely to underestimate the effects, because the substance accumulates in sediment.

Current regulatory context and use of EBMs

If imposex is observed (Vas Deference Sequence Index (VDSI) above the EAC, see Section 5.3. and Annex II table) chemical status (with respect to TBT) is clearly not good, but under the WFD this cannot currently be taken into account.

Guidance needed?

Guidance would not be necessary to be able to decide on the monitoring approach but see below.

Practical feasibility

A disadvantage is that imposex cannot be analysed in areas with. e.g. severe contamination, due to the disappearance of the organisms, nor in areas where appropriate organisms to monitor do not exist for other reasons (including in limnic or very brackish water environments). An EBM approach might therefore not be implementable everywhere, and we therefore suggest that it should be optional for MS to use this approach in a WFD context.

Conclusions and recommendations

Short-term outlook and reccomendations for possible further implementation under the WFD

Since imposex is frequently monitored in the marine environment and taken into account in the MSFD assessment, for harmonisation purposes and cost effectiveness, it would be useful to also take imposex into account in the marine WFD context when assessing the impact of TBT on waterbody status.

A OOAO approach should be used, meaning that status is not good if the VDSI is above the EAC. Conversely, it cannot be said that status is good simply when the imposex VDSI is below the EAC because imposex is a severe effect and the EQS was developed to take into account also effects that could occur in other, non-monitored species. Thus, imposex biomarkers should be used along with the chemical approach.

Medium term (next mandate)

A guidance document could be useful to support MS in the identification of suitable measures at local level. The substance has been banned for use as an antifouling agent but high concentrations are still found in marinas, also in surface sediment, soil and storm water, and linked to activites such as boating uptake.

6.3.3. To assess secondary poisoning from DDT

Current approach

The DDT EQS in the EQSD are expressed for water. However, the protection objective is somewhat unclear, since no EQS dossier has been located.

Usefulness of EBMs

The biomarker egg-shell thinning is clearly relevant to both the WFD and MSFD because the effects are of high ecological relevance and they can be linked to DDT exposure.

Therefore, if this biomarker responds (effects observed above the EACs), chemical status with respect to DDT is probably not good. However, it may be difficult to link effects observed in this biomarker to pressures on a particular water body. The effects occur due to ingestion of feed such as fish, but perhaps also terrestrial animals, and the species investigated (sea eagle) can catch its prey from a large geographical area.

Added value of using EBMs

The biomarker provides a direct measurement of DDT-related effects (secondary poisoning).

Current regulatory context and use of EBMs

The biomarker egg-shell thinning should already fit into the current MSFD context

Guidance needed?

See below.

Practical feasibility

See above and the aspects included in the Annex II tables.

Conclusions and recommendations

Short-term outlook and recommendations for possible further implementation under the WFD

Consider, if possible, adding a link between the "MSFD initial assessments" and WFD surveillance and investigative monitoring by making it possible to take results from the MSFD assessment into account in the WFD assessment of the marine environment.

Medium term (next mandate)

It could be worth developing guidance on how to proceed if "non-compliance" is observed in the MSFD context but the WFD EQS is not exceeded and the reasons are not known. Although the substance behind the effects is known in this case, the sources would need to be identified (source tracking at both large and small geographical scale).

7. CONCLUSIONS AND PROPOSED ACTIONS

7.1. Final Considerations

Current chemical monitoring under the WFD takes into account several compounds whose toxicity to humans and the environment is well known. However, the presence of "unknown" chemicals with potentially harmful effects emitted by human activities is not captured by a targeted chemical analysis. Possible mixture effects are not fully covered either. This situation creates the need to develop a new holistic way to address effects of known and unknown compounds in the environment. This need was addressed at the EU Water Directors meeting in October 2016 (Bratislava, 2016). As discussed in this Report, EBMs contribute to a more holistic way to assess surface water quality with respect to chemical contamination. Based on the evidence presented in this report, EBMs relevant to environmental and human health were collected and selected, taking particular account of their maturity, assessability and the extent of their use. As described in the JRC technical report on MoA, the PS and other substances of interest can be grouped according to their common MoA (**Annex I**). In some cases, the MoA is very specific to a certain group of substances or even a single chemical. For example, some herbicides inhibit photosynthesis. Therefore, EBM related to this MoA, e.g. which measure photosystem II inhibition and chlorophyll fluorescence, would detect the biological effects of herbicides displaying this MoA.

An integrated platform linking EBMs to currently employed chemical and ecological assessment methods has been proposed in the JRC report about the integrated assessment of the current PS list under the WFD and other substances of interest (see **Annex IV**). To monitor a number of selected MoAs using their respective EBMs would be complementary to the assessment of chemical status and ecological status using current chemical methods and biological indices, and add a line of evidence to the question of whether chemical contamination contributes to a finding of poor ecological status.

Prioritised and recommended MoAs are presented in **Annex I** Table 2**.** These MoAs, in particular estrogenicity and genotoxicity, can be mainly addressed by well developed or even standardised *in vitro* EBMs with specific EBTs (see Annex I and III) and/or biomarkers. General agreement among participants of the activity exists that MoAs for Estrogenicity, Mutagenicity/Genotoxicity, Dioxin-like effects and Herbicidal effects can already be detected by EBMs, and that in future also Neurotoxicity will be measurable.

Whereas *in vitro* assays could be used to monitor more or less any compartment, the biomarker approach would likely fit the best in combination with biota monitoring for chemical analysis. Several biomarkers have reached a high level of maturity, especially in the marine context (see Section 5.2.). Biomarkers can also be used in freshwater environments, although the level of maturity is generally lower, largely due to their limited use so far. Chapters 5 and 6, and **Annex II** of this report provide more details about these methods.

At the moment, very few biological indices exist that respond to toxic chemicals (see Section 5.5.). Those identified so far have primarily been developed to assess impacts on benthic communities. On the other hand, several *in vivo* EBM and biomarkers with a high level of "maturity" are available that can be used to cover complex mixtures and even cumulative effects from several stressors and at ecologically relevant levels. Chapter 5, 6

and **Annex II** of this report provide more details about these methods. Furthermore, the assessment criteria for biomarkers could generally also be developed in a similar way as for the BQEs, using reference conditions and the likelihood of negative effects at population level (see Section 5.3.). *In vivo* EBMs are widely used in Europe and could also be relevant for the detection of complex mixtures, for sediment assessment and for the evaluation of effluent quality (see **Chapter 6**), as well as to link chemical and ecological status.

7.2. Possible use of EBMs in the WFD monitoring and assessment Programmes

The use of EBMs provides a straightforward approach to evaluate the risks from chemical mixtures, meaning mixtures of substances with the same MoAs and complex mixtures. Selected *in vitro*, *in vivo* and biomarker EBMs can be useful to evaluate the mixture risks depending on local situations and pressures.

EBMs could also be considered for the assessment of chemical status linked to a specific MoA (e.g. dioxin-like effects, estrogenicity) and/or included as a supportive component for BQEs in determining ecological status. Furthermore, EBMs are suitable to evaluate the chemical quality of sediments.

The use of EBMs under the legislative framework could have the following advantages

- early warning of effects before adverse outcomes at population level occur (precautionary principle); this point is particularly relevant to the effects of climate change (e.g. through flooding) on chemical contamination;
- evaluation of effects from chemical mixtures even with unknown composition
- inclusion of realistic behaviour of interacting chemicals (bioactivation, metabolites, generation of new compounds through spontaneous reactions, additive and/or synergetic/antagonistic effects);
- evaluation of desired endpoints based on groups of similarly acting substances (shared mode of action);
- use as a screening method to rationalise monitoring programmes.

Two possible ways to implement EBMs under the WFD have been identified. The options could be used together. In both cases, specific guidance would be needed.

Option 1: <u>To include EBMs as a supportive component in the assessment of chemical and ecological status.</u>

The aim would be to better take the effects from chemicals and chemical mixtures into account in the chemical and ecological status classification. EBMs could provide evidence on whether pollutants are causing or contributing to impaired ecological status. EBMs are of particular interest if causes of insufficient ecological status are unknown and the chemical status assessment gives good results. EBMs showing a response would indicate the possible presence of detrimental contamination caused by pollution and could thus trigger further activities such as source identification, including through the use of EBMs. *In vitro* bioassays could contribute particularly to the assessment of chemical status, while

biomarkers at higher levels of biological organisation could contribute particularly to the assessment of ecological status, and *in vivo* bioassays and biomarkers at lower levels of biological organisation could be used to assess chemical or ecological status, depending on the specific bioassay/biomarker.

Option 2: To include EBMs in monitoring and screening for the identification of pressures.

The selected EBMs could be useful for pressures and impacts assessments, for which the regulatory requirements are more flexible than in relation to status assessment (see Section 5.5.). It would be possible to use EBMs to prioritise or de-prioritise water bodies for further (operational or investigative) monitoring. The use of EBMs in this context is already possible, but the addition of a clarification in, e.g. WFD Annex II, would probably promote such an approach. EBMs would certainly be useful for investigative monitoring programmes under the WFD, and at sites where ecological and chemical status give different evaluations, e.g. to identify the cause of a reduction in ecological status or the source of pollution (for effective measures).

7.3. Next steps

Whatever approach is adopted, the benefit of employing prioritised EBMs for the assessment of chemical mixtures should be evaluated through an EU-wide action based on an interlaboratory exercise involving MS to communicate about standardised operating procedures (SOPs) and data interpretation in order to avoid large gaps in methodological coherence between countries. Based on these results, guidelines for the use of selected EBMs should be developed, including an in-depth assessment of associated costs and required facilities and expertise for the use of these EBMs.

Implementation of an EBM "watch list" at selected European sites with the use of selected *in vitro*, *in vivo* and biomarkers would be highly recommended. If elevated effect levels were to be identified at certain sampling sites, a case study should be initiated to investigate the suitability of EBMs as a means to identify pollutant sources (to inform the subsequent development of effective measures).

At the same time, further activity on the derivation of assessment criteria for EBMs is needed. In particular, EBTs for the different MoAs should be developed further in the context of the activity of WG Chemicals. The EBTs should protect both the environment and human health and be linked to the EQS concept. A specific simplified Technical Guidance Document (TGD) on EBT derivation and assessment criteria should be elaborated (a good example is reported in Annex III). For certain MoAs (for example neurotoxicity) further research is needed prior to any recommendations for policy, but a first round of discussion in WG Chemicals is highly recommended due to the implications of this aspect also for human health protection (see Annex VI). For EBMs that are intended as supportive elements for ecological status (see section 5.4.3.), assessment criteria could be based on the same principles as the background assessment criteria (BAC) and environmental assessment criteria (EAC) used under the MSFD (see Section 5.3). Finally, an EU-wide interlaboratory exercise with the use of reference materials composed of substances representing the main toxicity drivers is proposed in Annex IV as a frontline approach in support of an integrated monitoring platform aimed at taking into consideration the effects from chemical mixtures.

8. References

Adeel M, Song X, Wang Y, Francis D, Yanga Y. 2017. Environmental impact of estrogens on human, animal and plant life: A critical review. Environment International, Vol 99, 107-119.

Altenburger R, Ait-Aissa S, Antczak P, Backhaus T, Barcelo D, Seiler TB, Brion F, Busch W, Chipman K, de Alda ML, de Aragao Umbuzeiro G, Escher BI, Falciani F, Faust M, Focks A, Hilscherova K, Hollender J, Hollert H, Jager F, Jahnke A, Kortenkamp A, Krauss M, Lemkine GF, Munthe J, Neumann S, Schymanski EL, Scrimshaw M, Segner H, Slobodnik J, Smedes F, Kughathas S, Teodorovic I, Tindall AJ, Tollefsen KE, Walz KH, Williams TD, Van den Brink PJ, van Gils J, Vrana B, Zhang X, Brack W. 2015. Future water quality monitoring--adapting tools to deal with mixtures of pollutants in water resource management. Sci Total Environ 512-513, 540-551.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL. 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environmental Toxicology and Chemistry 29 (3):730-741.

Arlos MJ, Parker WJ, Bicudo JR, Law P, Hicks KA, Fuzzen MLM, Andrews SA, Servos MR. 2018. Modeling the exposure of wild fish to endocrine active chemicals: Potential linkages of total estrogenicity to field-observed intersex. Water research 139:187-197. doi:10.1016/j.watres.2018.04.005.

Bengtsson-Palme J, Larsson DGJ. 2016. Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation. Environ Int. 86:140

Bengtsson-Palme J., Larsson DGJ., Kristiansson E. 2017. Using metagenomics to investigate human and environmental resistomes. J Antimicrob Chemother. https://doi.org/10.1093/jac/dkx199

Bengtsson-Palme J, Kristiansson E, Larsson DGJ. 2018. Environmental factors influencing the development and spread of antibiotic resistance. FEMS Micro Rev. 1;42. https://doi.org/10.1093/femsre/fux053

Blaber SJM. 1970. The occurrence of a penis-like outgrowth behind the right tentacle in spent females of *Nucella lapillus* (L.). Proceedings of the Malacological Society of London, 39: 231–233.

Brack W, Altenburger R, Schuurmann G, Krauss M, Lopez Herraez D, van Gils J, Slobodnik J, Munthe J, Gawlik BM, van Wezel A, Schriks M, Hollender J, Tollefsen KE, Mekenyan O, Dimitrov S, Bunke D, Cousins I, Posthuma L, van den Brink PJ, Lopez de Alda M, Barcelo D, Faust M, Kortenkamp A, Scrimshaw M, Ignatova S, Engelen G, Massmann G, Lemkine G, Teodorovic I, Walz KH, Dulio V, Jonker MT, Jager F, Chipman K, Falciani F, Liska I, Rooke D, Zhang X, Hollert H, Vrana B, Hilscherova K, Kramer K, Neumann S, Hammerbacher R, Backhaus T, Mack J, Segner H, Escher B, de Aragao Umbuzeiro G. 2015. The SOLUTIONS project: challenges and responses for present and future emerging pollutants in land and water resources management. Sci Total Environ 503-504, 22-31.

Brion F, BriDe Gussem V, Buchinger S, Hollert H, Carere M, Porcher JM, Piccini B, Feray C, Dulio V, Konemann S, Simon E, Werner I, Kase R, Ait-Aissa S. 2019. Monitoring estrogenic activities of waste and surface waters using a novel *in vivo* zebrafish embryonic (EASZY) assay: Comparison with in vitro cell-based assays and determination of effect-based trigger values. Environment International 130: 104896.

Brix K. 2018. Evaluation of Effects Based Methods for Regulating Metals in Aquatic Ecosystems. White paper on behalf of the Metals Environmental Research Associations (MERA). Report from EcoTox consultant, Florida, US. Available at https://copperalliance.eu/resources/effects-based-methods-metals-water-framework-directive/

Bopp SK, Lettieri T. 2007. Gene regulation in the marine diatom *Thalassiosira pseudonana* upon exposure to polycyclic aromatic hydrocarbons (PAHs). Gene. 396(2):293-302.

Carvalho RN, Burchardt AD, Sena F, Mariani G, Mueller A, Bopp SK, Umlauf G, Lettieri T. 2011. Gene biomarkers in diatom *Thalassiosira pseudonana* exposed to polycyclic aromatic hydrocarbons from contaminated marine surface sediments. Aquat Toxicol. 101(1):244-53.

Carvalho RN, Arukwe A, Ait-Aissa S, Bado-Nilles A, Balzamo S, Baun A, Belkin S, Blaha L, Brion F, Conti D, Creusot N, et al. 2014. Mixtures of chemical pollutants at European legislation safety concentrations: how safe are they? Toxicological Sciences, 141(1), pp.218-233.

Chapman P. 1990. The sediment quality triad approach to determining pollution-induced degradation. The Science of the Total Environment 97/98: 815–825.

Chapman P. 1995. Extrapolating laboratory toxicity results to the field. Environmental toxicology and chemistry 14:927-930.

Chapman P. 1996. Presentation and interpretation of sediment quality triad data. Ecotoxicology 5:327-339.

Davies IM and Vethaak AD. 2012. Integrated marine environmental monitoring of chemicals and their effects. ICES Cooperative Research Report No. 315. 277 pp.

European Commission. 2014. Technical report on aquatic effect-based monitoring tools. Technical Report: 2014-077. ISBN 978-92-79-35787-9.

Escher BI, Ait-Aissa S, Behnisch PA, Brack W, Brion F, Brouwer A, Buchinger S, Crawford SE, Du Pasquier D, Hamers T, Hettewer K, Hilscherova K, Hollert H, Kase R, Kienle C, Tindall AJ, Turek J, van der Oost R, Vermeirssen E, Neale PA. 2018. Effect-based trigger values for *in vitro* and *in vivo* bioassays performed on surface water extracts supporting the environmental quality standards (EQS) of the European Water Framework Directive. Science of the Total Environment, 628-629, pp. 748-765.

Féral C. 1980. Variations dans l'évolution du tractus génital mâle externe des femelles de trois Gastéropodes Prosobranches gonochoriques de stations atlantiques. Cahiers Biologie Marine, 21: 479–491.

Ferrero VE, Pedotti M, Chiadò A, Simonelli L, Calzolai L, Varani L, Lettieri T, 2014. Rational modification of estrogen receptor by combination of computational and experimental analysis. PLoS One. 30;9(7):e102658. doi: 10.1371/journal.pone.0102658. Garner E, Wallace JS, Arango-Argoty G, Wilkinson C, Fahrenfeld N, Heath LS, Zhang L, Arabi M, Aga DS, Pruden A. 2016. Metagenomic profiling of historic Colorado Front Range flood impact on distribution of riverine antibiotic resistance genes. Scientific Reports 6, 38432.

Le Fol V, Aït-Aïssa S, Sonavane M, Porcher JM, Balaguer P, Cravedi JP, Zalko D, Brion F. 2017. In vitro and in vivo estrogenic activity of BPA, BPF and BPS in zebrafish-specific assays. Ecotoxicol Environ Saf., 142: 150-156.

Gupta SK, Shin H, Han D, Hur H-G, Unno T. 2018. Metagenomic analysis reveals the prevalence and persistence of antibiotic- and heavy metal-resistance genes in wastewater treatment plant. Journal of Microbiology 56, 408-415.

Hanson N, Larsson Å, Förlin L. 2014. Gränsvärden för biomarkörer och dess tillämpning i bedömningsgrunder för fiskhälsa38.

Harries JE, Sheahan DA, Jobling S, Matthiessen P, Neall M, Sumpter JP, Taylor T, Zaman N, 1997. Estrogenic activity in five United Kingdom rivers detected by measurement of vitellogenesis in caged male trout. Environmental Toxicology and Chemistry 16 (3):534-542. doi:10.1897/1551-5028(1997)016<0534:eaifuk>2.3.co;2

Helander B, Olsson A, Bignert A, Asplund L, Litzén K. 2002. The Role of DDE, PCB, Coplanar PCB and Eggshell Parameters for Reproduction in the White-tailed Sea Eagle (*Haliaeetus albicilla*) in Sweden. Ambio Vol. 31 No. 5, August 2002.ICCVAM 2011 National Institute of Environmental Health, Sciences National Institutes of Health, U.S. Public Health Service, Department of Health and Human Services, "Test Method Evaluation Report The LUMI-CELL® ER (BG1Luc ER TA) Test Method: An In Vitro Assay for Identifying Human Estrogen Receptor Agonist and Antagonist Activity of Chemicals" NIH Publication Number 11-7850

ISO 19040-1:2018 Water quality — Determination of the estrogenic potential of water and waste water — Part 1: Yeast estrogen screen (*Saccharomyces cerevisiae*).

ISO 19040-2:2018 Water quality — Determination of the estrogenic potential of water and waste water — Part 2: Yeast estrogen screen (A-YES, *Arxula adeninivorans*).

ISO 19040-3:2018 Water quality — Determination of the estrogenic potential of water and waste water — Part 3: In vitro human cell-based reporter gene assay.

Jarosova B, Blaha L, Giesy JP, Hilscherova K. 2014. What level of estrogenic activity determined by *in vitro* assays in municipal waste waters can be considered as safe?. Environment international, 64, pp. 98-109.

Jobling S, Reynolds T, White R, Parker MG, Sumpter JP. 1995. A variety of environmentally persistent chemicals, including some phthalate plasticizers, are weakly estrogenic. Environmental Health Perspectives 103 (6):582-587. doi:10.2307/3432434.

Kahlert M. 2012. Utveckling av en miljögiftsindikator – kiselalger i rinnande vatten. Länsstyrelsen Blekinge län, Karlskrona, Report 2012:12, 40 pp.

Kase R, Javurkova B, Simon E, Swart K, Buchinger S, Koenemann S, Escher BI, Carere M, Dulio V, Ait-Aissa S, Hollert H, Valsecchi S, Polesello S, Behnisch P, di Paolo C, Olbrich D, Sychrova E, Gundlach M, Schlichting R, Leborgne L, Clara M, Scheffknecht C, Marneffe Y, Chalon C, Tusil P, Soldan P, von Danwitz B, Schwaiger J, Moran Palao A, Bersani F, Perceval O, Kienle C, Vermeirssen E, Hilscherova K, Reifferscheid G, Werner I. 2018. Screening and risk management solutions for steroidal estrogens in surface and wastewater. Trac-Trends Anal Chem 102:343-358. doi:10.1016/j.trac.2018.02.013.

Kidd KA, Paterson MJ, Rennie MD, Podemski CL, Findlay DL, Blanchfield PJ, Liber K. 2014. Direct and indirect responses of a freshwater food web to a potent synthetic oestrogen. Philosophical Transactions of the Royal Society B-Biological Sciences 369 (1656). doi:10.1098/rstb.2013.0578.

Kidd KA, Blanchfield PJ, Mills KH, Palace VP, Evans RE, Lazorchak JM, Flick RW. 2007. Collapse of a fish population after exposure to a synthetic estrogen. Proceedings of the National Academy of Sciences USA 104 (21):8897-8901.

Kirkland D, Aardema M, Henderson L, Muller L. 2005. Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-

[38] Available at http://naturvardsverket.diva-portal.org/smash/resultList.jsf?dswid=-8813&language=sv&searchType=SIMPLE&query=Gr%C3%A4nsv%C3%A4rden+f%C3%B6r+biomark%C3%B6rer+och+dess+till%C3%A4mpning+i+bed%C3%B6mningsgrunder+f%C3%B6r+fiskh%C3%A4lsa&af=%5B%5D&aq=%5B%5B%5D%5D&aq2=%5B%5B%5D%5D&aqe=%5B%5D&noOfRows=50&sortOrder=author_sort_asc&sortOrder2=title_sort_asc&onlyFullText=false

carcinogens - I. Sensitivity, specificity and relative predictivity. Mutation Research-Genetic Toxicology and Environmental Mutagenesis 584 (1-2):1-256.

Kirkland D, Reeve L, Gatehouse D, Vanparys P. 2011. A core *in vitro* genotoxicity battery comprising the Ames test plus the *in vitro* micronucleus test is sufficient to detect rodent carcinogens and *in vivo* genotoxins. Mutation Research 721 (1):27-73.

Kisand V, Valente A, Lahm A, Tanet G, Lettieri T, 2012. Phylogenetic and functional metagenomic profiling for assessing microbial biodiversity in environmental monitoring. PLoS ONE 7(8): e43630.

Könemann S, Kase R, Swart K, Simon E, Buchinger S, Schlüsener M, Ternes T, Hollert H, Werner I, Ait-Aissa S, Escher B, Dulio V, Valsecchi S, Polesello S, Behnisch P, Jarvukova B, Perceval O, di Paolo C, Olbrich D, Sychrova E, Gundlach M, Schlichting R, Leborgne L, Clara M, Scheffknecht C, Marneffe Y Chalon C, Tusil P, Soldan P, von Danwitz B, Schwaiger J, Moran Palao A, Bersani F, Vermeirssen E, Hilscherova K, Reifferscheid G, Carere M. 2018. Effect-based and chemical analytical methods to monitor estrogens under the Water Framework Directive. Trends in Analytical Chemistry 102:225-235.

Kunz PY, Kienle C, Carere M, Homazava N, Kase R. 2015. *In vitro* bioassays to screen for endocrine active pharmaceuticals in surface and waste waters. Journal of Pharmaceutical and Biomedical Analysis, 106, pp. 107-115.

Kunz, P. Y.; Simon, E.; Creusot, N.; Jayasinghe, B. S.; Kienle, C.; Maletz, S.; Schifferli, A.; Schonlau, C.; Ait-Aissa, S.; Denslow, N. D.; Hollert, H.; Werner, I.; Vermeirssen, E. L. M. 2017 Effect-based tools for monitoring estrogenic mixtures: Evaluation of five in vitro bioassays. Water research 110, 378-388

Lyons BP, Bignell JP, Stentiford GD, Bolam TPC, Rumney HS, Bersuder P, Barber JL, Askem CE, Nicolaus MEE, Maes T. 2017. Determining good environmental status under the marine strategy framework directive: case study for descriptor 8 (chemical contaminants). Marine Environmental Research (124):118-129.

Matthiessen P, Sumpter JP. 1998. Effects of estrogenic substances in the aquatic environment. In: Braunbeck T., Hinton D.E., Streit B. (eds) Fish Ecotoxicology. EXS, vol 86. Birkhäuser, Basel.

Matthiessen, P.; Ankley, G. T.; Biever, R. C.; Bjerregaard, P.; Borgert, C.; Brugger, K.; Blankinship, A.; Chambers, J.; Coady, K. K.; Constantine, L. 2017. Recommended approaches to the scientific evaluation of ecotoxicological hazards and risks of endocrine-active substances. J Integrated environmental assessment management, 13, (2), 267-279.

Mondy C, Usseglio-Polatera P. 2013. Using conditional tree forests and life history traits to assess specific risks of stream degradation under multiple pressure scenario. Science of the total environment 461-462:750-760.

Mustamäki N, Olsson J, Franzén F, Förlin L, Larsson Å, Parkkonen J, Faxneld S, Danielsson S, Nyberg E. 2018. Faktablad från Integrerad kustfiskövervakning 2018:2. Holmöarna, Bottniska viken, 1989-2017 (Fact Sheet in Swedish), pp. 7. https://pub.epsilon.slu.se/15605.

Napierska D, Sanseverino I, Loos R, Gomez Cortes L, Niegowska M, Lettieri T. 2018. Modes of action of the current Priority Substances list under the Water Framework Directive and other substances of interest. JRC Technical Reports JRC110117. Office for official Publications of the European Communities.

Niegowska M, Napierska D, Sanseverino I, Gomez Cortes L, Marinov D, Loos R, Lettieri T. 2018. JRC Technical Report on the Integrated assessment of the current Priority Substances list under the Water Framework Directive and other substances of interest. In preparation.

Pirrone N, Trombino G, Cinnirella S, Algieri A, Bendoricchio G, Palmeri L. 2005. The Driver-Pressure-State-Impact-Response (DPSIR) approach for integrated catchment-

coastal zone management: preliminary application to the Po catchment-Adriatic Sea coastal zone system. Regional Environmental Change 5(2–3), pp. 111–137.

OECD 2009, Test No. 455: The Stably Transfected Human Estrogen Receptor-alpha Transcriptional Activation Assay for Detection of Estrogenic Agonist-Activity of Chemicals.

OECD. 2016. OECD No. 233 Second Edition Users' Handbook Supplement to the Guidance Document for Developing and Assessing Adverse Outcome Pathways. ENV/JM/MONO(2016)12.

OSPAR. 2013. Background documents and technical annexes for biological effects monitoring (Update 2013). OSPAR Commission, London. Publication 589, 238 pp.

SETAC. 1993. Guidance document on sediment toxicity tests and bioassays for freshwater and marine environments. From the Workshop on Sediment Toxicity Assessment", held at Slot Moermond Congrescentrum Renesse, The Netherlands. 8-10 November 2993. Editors: Hill, IR, Matthiessen, P and Heimbach F.

Smith BS. 1971. Sexuality in the American mud snail, Nassarius obsoletus SAY. Proceedings of the Malacological Society of London, 39: 377–378.

Smith BS. 1981. Tributyltin compounds induce male characteristics on female mud snails Nassarius obsoletus = Ilyanassa obsoleta. Journal of Applied Toxicology, 1: 141–144.

Spencer SJ, Tamminen MV, Preheim SP, Guo MT, Briggs AW, Brito IL, Weitz DA, Pitkänen LK, Vigneault F, Virta MPJ, Alm EJ. 2016. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. The ISME *Journal vol 10,* pp. 427–436.

Urbaniak C, Sielaff AC, Frey KG, Allen JE, Singh N, Jaing C, Wheeler K, Venkateswaran K. Detection of antimicrobial resistance genes associated with the International Space Station environmental surfaces. 2018. Sci Rep. 16;8(1):814. doi: 10.1038/s41598-017-18506-4.US EPA. 1992. Toxicity Identification Evaluation: Characterization of Chronically Toxic Effluents, Phase I. United States Environmental Protection Agency. EPA/600/6-91/005F.

US EPA. 1991 a. Methods for Aquatic Toxicity Identification Evaluations: Phase I Toxicity Characterization Procedures. Second Edition. EPA/600/6-91/003. Environmental Research Laboratory, Duluth, MN.

US EPA. 1991. Toxicity Identification Evaluation: Characterization of Chronically Toxic Effluents, Phase I. EPA/600/6-911005. Environmental Research Laboratory, Duluth, MN.

Van der Oost R, Sileno G, Janse T, Nguyen MT, Besselink H, Brouwer A. 2017. SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment; PART II. Field feasibility survey. Environmental Toxicology and Chemistry, 36(9), pp. 2400-2416.

Van der Oost R, Beyer J, Vermeulen NPE. 2003. Fish bioaccumulation and biomarkers in environmental risk assessment: a review. Environmental Toxicology and Pharmacology 13(2):57-149.

Vos JG, Dybing E, Greim HA, Ladefoged O, Lambre C, Tarazona JV, Brandt I, Vethaak AD. 2000. Health effects of endocrine-disrupting chemicals on wildlife, with special reference to the European situation. Critical reviews in toxicology 30 (1):71-133. doi:10.1080/10408440091159176.

Walter H, Consolaro F, Gramatica P, Scholze M, Altenburger R. 2002. Mixture Toxicity of Priority Pollutants at No Observed Effect Concentrations (NOECs). Ecotoxicology 11(5):299–310.

Wernersson AS et al. 2015. The European technical report on aquatic effect-based monitoring tools under the water framework directive. Environ Sci Eur 27:7.

Wittwehr, C.; Aladjov, H.; Ankley, G.; Byrne, H. J.; de Knecht, J.; Heinzle, E.; Klambauer, G.; Landesmann, B.; Luijten, M.; MacKay, C.; Maxwell, G.; Meek, M. E.; Paini, A.; Perkins, E.; Sobanski, T.; Villeneuve, D.; Waters, K. M.; Whelan, M. 2017. How Adverse Outcome Pathways Can Aid the Development and Use of Computational Prediction Models for Regulatory Toxicology. Toxicological Sciences, 155, (2), 326-336.

ANNEX I. MoA examples

Summary of already existing effect-based methods (EBMs), which can be used to monitor the mode of action (MoA) reported in the literature for the priority substances (PS), Watch List (WL) and emerging substances (From the JRC technical report on MoA[39]).

[39] Napierska D et al. 2018. Modes of action of the current Priority Substances list under the Water Framework Directive and other substances of interest. JRC Technical Reports JRC110117. Office for official Publications of the European Communities.

Table I.1: EBMs classified based on the MoA.

Substance	Photosystem II inhibition/algal growth inhibition	Chlorophyll fluorescence	AChE inhibition	ALA-D activity	Cytochrome P4501a/EROD activity	CYP19A1/ aromatase induction	Vtg induction	Intersex in male fish	T/E2 assay	MT induction	Genotoxicity (DNA adducts)	Genotoxicity (MN frequency, Comet assay)	Fish Embryotoxicity (FET)	Amphibian Embryotoxicity (FETAX)	SfG	LMS	HSP	Thyroid hormone(s) assay	Imposex in gastropoda (VDSI, RPSI)	ROS/oxidative stress	Xenobiotic-metabolising/ hepatic enzymes	Liver histopathology	Histopathology of organs other than liver	PAH metabolites
Herbicides																								
Alachlor		■												■						■	■			
Atrazine	■	■				■			■											■	■			
Diuron	■	■			■	■			■											■				
Isoproturon	■	■																		■				
Simazine	■	■												■						■	■			
Trifluralin												■									■			
Aclonifen		■																						
Bifenox		■																						
Cybutryne	■	■																		■				
Terbutryn	■	■																		■				
Oxadiazon		■																		■				
Triallate		■																		■				
Polyaromatic hydrocarbons (PAHs)																								
Anthracene	■																			■	■		■	
Fluoranthene					■															■			■	
Naphthalene																				■				
Polyaromatic hydrocarbons (PAH)					■						■									■	■	■	■	■
Organophosphorus insecticides																								
Chlorfenvinphos			■													■				■				
Chlorpyrifos-ethyl			■																	■				
Dichlorvos			■									■								■				
Malathion			■																					
Omethoate			■																					
Organochlorine insecticides																								
Cyclodiene pesticides					■		■	■																
DDT total and para-para-DDT							■	■	■			■								■	■			
Endosulfan						■	■											■		■	■	■	■	
Hexachloro-cyclohexane																	■					■	■	
Dicofol						■																		

Substance	Photosystem II inhibition/algal growth inhibition	Chlorophyll fluorescence	AChE inhibition	ALA-D activity	Cytochrome P4501a/EROD activity	CYP19A1/ aromatase induction	Vtg induction	Intersex in male fish	T/E2 assay	MT induction	Genotoxicity (DNA adducts)	Genotoxicity (MN frequency, Comet assay)	Fish Embryotoxicity (FET)	Amphibian Embryotoxicity (FETAX)	SfG	LMS	HSP	Thyroid hormone(s) assay	Imposex in gastropoda (VDSI, RPSI)	ROS/oxidative stress	Xenobiotic-metabolising/ hepatic enzymes	Liver histopathology	Histopathology of organs other than liver	PAH metabolites
Heptachlor and Heptachlor epoxide														■							■			
Chlorinated solvents																								
Carbon tetrachloride																	■			■	■			
Tetrachloroethylene													■											
Trichloroethylene													■		■	■				■	■			
1,2-Dichloroethane											■													
Dichloromethane																								
Hexachlorobutadiene																							■	
Trichloromethane (Chloroform)																								
Aromatic organochlorine compounds																								
Hexachlorobenzene (HCB)			■		■															■	■		■	
Pentachlorobenzene					■															■				
Pentachlorophenol						■			■						■						■		■	
Trichlorobenzenes																								
Dioxins, PCBs, BDEs																								
Brominated Diphenyl Ethers (BDEs)						■												■						
Dioxins and coplanar PCBs					■																			
Metals																								
Cadmium and its compounds										■						■				■			■	
Lead and its compounds				■												■				■	■			
Mercury and its compounds										■						■				■	■	■	■	
Nickel and its compounds																■				■	■	■	■	
Silver														■						■				
Uranium												■												
Selenium													■							■	■			
Endocrine disrupters																								
Di(2- ethylhexyl)-phthalate (DEHP)							■	■	■						■					■				
Nonylphenols						■	■	■					■		■	■				■			■	

Substance	Photosystem II inhibition/algal growth inhibition	Chlorophyll fluorescence	AChE inhibition	AIA-D activity	Cytochrome P4501a/EROD activity	CYP19A1/ aromatase induction	Vtg induction	Intersex in male fish	T/E2 assay	MT induction	Genotoxicity (DNA adducts)	Genotoxicity (MN frequency, Comet assay)	Fish Embryotoxicity (FET)	Amphibian Embryotoxicity (FETAX)	SfG	LMS	HSP	Thyroid hormone(s) assay	Imposex in gastropoda (VDSI, RPSI)	ROS/oxidative stress	Xenobiotic-metabolising/ hepatic enzymes	Liver histopathology	Histopathology of organs other than liver	PAH metabolites
Octylphenols							■	■															■	
Tributyltin compounds										■						■			■					
17-Alpha-ethinylestradiol (EE2)							■	■	■															
17-Beta-estradiol (E2)							■	■	■															
Estrone (E1)							■	■	■															
Pyrethroid insecticides																								
Cypermethrin					■							■	■							■	■		■	
Bifenthrin													■							■	■			
Deltamethrin												■	■							■	■			
Esfenvalerate													■							■	■			
Permethrin													■							■	■			
Perfluorinated surfactant																								
Perfluorooctan-sulfonic acid (PFOS)							■		■				■					■						
Benzene											■									■	■			
Quinoline fungicide																								
Quinoxyfen																								
C10-13 chloroalkanes														■				■			■	■		
Hexabromocyclo-dodecane (HBCDD)													■			■				■				
Antibiotics																								
Erythromycin		■																		■	■			
Clarithromycin		■																		■	■			
Azithromycin		■																		■	■			
Neonicotinoid insecticides																								
Imidacloprid			■									■	■							■				
Thiacloprid												■	■							■				
Thiamethoxam												■	■							■				
Clothianidin												■	■							■				
Acetamiprid												■	■							■				
Anti inflammatory drug																								

Substance	Photosystem II inhibition/algal growth inhibition	Chlorophyll fluorescence	AChE inhibition	ALA-D activity	Cytochrome P4501a/EROD activity	CYP19A1/ aromatase induction	Vtg induction	Intersex in male fish	T/E2 assay	MT induction	Genotoxicity (DNA adducts)	Genotoxicity (MN frequency, Comet assay)	Fish Embryotoxicity (FET)	Amphibian Embryotoxicity (FETAX)	SfG	LMS	HSP	Thyroid hormone(s) assay	Imposex in gastropoda (VDSI, RPSI)	ROS/oxidative stress	Xenobiotic-metabolising/ hepatic enzymes	Liver histopathology	Histopathology of organs other than liver	PAH metabolites
Diclofenac																								
Antioxidant																								
2,6-Di-tert-butyl-4-methylphenol																								
Sunscreen agent / UV filter																								
2-Ethylhexyl 4-methoxycinnamate																								
Carbamate insecticide and herbicide																								
Methiocarb																								
Sulfonylurea herbicide																								
Nicosulfuron																								

As discussed in this report, EBMs facilitate a more holistic way to assess water quality with respect to chemical contamination. Based on evidence presented in this report, EBMs with relevance for environmental and human health were selected, taking also the maturity and the assessability of the EBMs into account. The EBMs for the selected MoAs could be used to complement chemical analysis and the assessment of ecological status using BQEs. Prioritised MoAs are presented in Table 2. These MoAs can mostly be addressed using well developed or standardised EBMs

Table I.2: Recommended modes of action (MoA) for inclusion in WFD monitoring.
EBM=Effect-Based-Method, SOP=Standard Operating Procedure, EBT=Effect-Based-Trigger-value, SW= Surface Water, WW=Waste Water, DW=Drinking Water

MoA with proven relevance	**Protection aim/ reasoning**	**Effect based method (EBM)**	**Reference compound**	**Standardised SOP**	**Defined effect based trigger value (EBT) to reference compound[40]**	**Known applicability**
Relevant MoAs with developed EBMs for potential implementation in the WFD						
Activation of estrogen receptor (ER)	Aquatic wildlife (fish) Is the most investigated MoA of endocrine disruption relevant for aquatic and human health; currently, mixture effects are not assessed. Well-developed *in vitro* EBMs capturing additive effects of ER-agonists are available.	ERα-CALUX	17-beta-estradiol	ISO 19040-3	0.283 ng/l E2-equivalence	SW, WW, DW, sediments
		T47D				
		A-YES		ISO 19040-2	0.400 ng/L	
		ER GeneBLAzer		Validity for ISO 19040-3 to be demonstrated	0.242 ng/l E2-equivalence	
		Hela 9903		Validity for ISO 19040-3 to be demonstrated	0.182 ng/l E2-equivalence	
		MELN		Validity for ISO 19040-3 to be demonstrated	0.557 ng/l E2-equivalence	
		p-YES		No standard	0.500 ng/l	

[40] Values reported from literature studies, their relevance has to be discussed within the WG Chemicals.

					E2-equivalence	
Phytotoxicity/ PSII-inhibition	Aquatic wildlife (algae, plants) Herbicides and antibiotics (including as mixtures) cause toxic effects in algae. Available EBMs capture mixture effects and can be extended to detect specific PSII-inhibition.	Green algae assay	Diuron	DIN EN ISO 8692		SW, WW, sediments
Mutagenicity	Human health Potentially to be applied for the assessment of water bodies used for drinking water production.	Ames-Fluctuation assay	2-amino-anthracene (2-AA) Nitrofurantoin (NF) 4-nitro-*o*-phenylenediamine (4-NOPD)	ISO 11350	assessment based on yes/no EBT to be developed	SW, WW, DW, sediments
Dioxin-like effects	Wildlife, secondary poisoning, human health Implemented in Commission Regulation (EU) 2017/644	DR CALUX Assay	2,3,7,8-tetrachloro-dibenzo-*p*-dioxin (TCDD)	validation available	Needed definition of an EBM-specific EBT (action value) based on a biological equivalence to screen for an exceedance of the TEQ-value (0,0065 µg/kg) defined as EQS in the WFD.	Biota, sediments
		Micro-EROD Assay using H4IIe-Zellen	different options	No standard		
		EROD Assay using RTL-W1	different options	No standard		
		H4IIe-Luc Assay	different options	No standard		
Relevant MoAs with need for further research and method development						

Neurotoxicity (for future)	Aquatic wildlife, human health Emerging MoA which is unwanted from human and ecotoxicological perspective	Acetylcholinesterase (AChE) inhibition Need for the development of further EBMs to address neurotoxicity more comprehensively, see outlook neurotoxicity	different options see DIN	DIN 38415-1	assessment based on yes/no EBT to be developed	SW, WW, DW, sediments
Binding to human estrogen receptor (ER)	Aquatic wildlife Potentially to be applied for the detection of agonists and antagonists of the ER as well as mixture effects	*In vitro* human ERα Competition assay	17-beta-estradiol	Validated (Ferrero VE et al. 2013)		SW, WW, DW, sediments
Activation of estrogen receptor (ER)	Aquatic wildlife	m-YES	17-beta-estradiol	Under validation		SW, WW, DW, sediments
Binding to aryl hydrocarbon receptor (AhR)	Aquatic wildlife, human health Potentially to be applied for detecting PAH, dioxins, pesticides and other toxic unknown ligands	*In vitro* AhR Competition assay	2,3,7,8-tetrachloro-dibenzo-*p*-dioxin (TCDD)	Under validation		SW, WW, DW, sediments
Antimicrobial resistance (AMR)	Aquatic wildlife, human health Potentially to be applied for detecting antimicrobial resistance genes	Ion AmpliSeq™ Antimicrobial Resistance (AMR) Research Panel		Validated (Urbaniak et al. 2018)		SW, WW, DW, sediments

ANNEX II. Inventory results

List of EBMs in the inventory

List of *in vitro* assays and the respective endpoints included in the inventory:

1. DR CALUX/DR
2. PAH CALUX
3. ERalfa CALUX/ER-Luc (agonistic/antagonistic)
4. ER-CALUX
5. T47D-Kbluc
6. BG1Luc4E2
7. ERa_Luc_BG1
8. AR CALUX (agonistic/antagonistic)
9. YES (Yeast estrogen screen)
10. YAS (Yeast androgen screen)
11. micronucleus assay
12. TTR binding assay
13. umu-Test
14. PPARg-GeneBLAzer
15. PPARy-CALUX
16. HG5LN-hPXR
17. PXR-CALUX
18. MELN
19. ER-GeneBLAzer
20. SSTA ERα-HeLa-9903
21. A-YES
22. 3d YES
23. ISO-LYES (Sumpter)
24. ISO-LYES (McDonnell)
25. Anti-ER-GeneBLAzer
26. Anti-ERa_Luc_BG1
27. Anti-A-YES
28. AR-GenBLAzer
29. MDA-kb2

30. A-YAS
31. anti AR-GenBLAzer
32. anti MDA-kb2
33. anti AR-CALUX
34. anti PR-CALUX
35. ZELH-zfERbeta2 and ZELH-zfERalpha
36. HELN-PRB
37. GR-GeneBLAzer
38. antiGR-GeneBLAzer
39. TTR RLBA
40. TTR FITC_T4
41. XETA
42. Anti-TR-LUC-TRE
43. Comet assay
44. SOS Chromotest
45. Ames Fluctuation Test (TA98)
46. Ames Fluctuation Test (TA100)
47. RT gill-W1
48. RTG2
49. SAF1
50. AREc32
51. anti HELN-PRB
52. PLHC-1 / EROD
53. AREGeneBLAzer
54. Nrf2-CALUX
55. P53 CALUX
56. kappaB CALUX
57. PSII-inhibition (algae and higher plants via Imaging-Pulse-Amplitude-Modulation)

List of *in vivo* assays and the respective endpoints included in the inventory:

58. EASZY (Cyp19a1b-GFP)
59. REACTIV (unspiked)
60. RADAR (unspiked)
61. anti-AR RADAR (spiked)
62. Vibrio Fischeri (Bacteria) bioluminescence

63. Lumistox
64. 72h Algal growth inhibition
65. 24h Synchronous algae reproduction
66. 24h Combined algae assay (growth)
67. 2h Combined algae assay (PSII)
68. 48h *Daphnia magna* immobilisation
69. *Daphnia magna* reproduction test
70. *Ceriodaphnia dubia,* survival/ reproduction test
71. FET (Danio rerio) Fish Embryo Acute Toxicity test – mortality and sublethal effects
72. *Oryzias latipes* (fish)
73. *Oryzias melastigma* (fish)
74. *Oryzias mykiss* (fish)
75. 14d (fish) *Danio rerio*, mortality
76. *Crassostrea gigas (Bivalvia)* embryo-larval development
77. *Mytilus* sp. (mollusca) embryo larval development
78. 7d *Gammarus* sp. feeding (in situ assay)
79. 7d *Gammarus* sp. acetylcholinesterase (in situ assay)
80. *Gammarus* sp. reprotoxicity (in situ assay)
81. *Gammarus* sp. endocrine disrupting (in situ assay)
82. *Ceramium tenuicorne* (red macroalga) growth rate
83. *Nitocra spinipes* (harpactoid copepod) survival
84. *Potamopyros antipodarum* (snail) survival rate and reproductive output
85. *Nassarius reticulata* (snail)
86. *Hyalella azteca* (amphipod)
87. *Gmelinoides fasciatus* (amphipod)
88. *Corophium volutator* (amphipod)
89. *Brachionus* (rotifera)
90. *Artemia franciscana* (crustacea) mortality
91. 48h/7d *Acartia tonsa* (crustacea) mortality, larval development
92. *Tigriopus fulvus* (crustacea)
93. *Hediste diversicolor* (Polychaeta)
94. *Paracentrotus lividus (*echinodermata) fecundity, larval development
95. *Heterocypris incongruens* (Ostracoda) growth inhibition, mortality
96. *Chironomus* assay
97. Mussel larvae

98. *Lumbriculus* assay
99. *Nitocra spinipes* LDR test (larval development rate)
100. *Amphibalanus Amphitrite* (crustacea) mortality
101. Fetax (amphibian embryos)

List of biomarkers and the respective endpoints included in the inventory:

102. Imposex, VDSI index - Penis and Vas Deference development
103. Imposex, RPSI index - Relative Penis Size Index
104. LMS (Lysosomal Membrane Stability) - minutes destabilisation period
105. MT (metallothionein) induction - concentration of MT (common unit: ug/mg cytosolic protein)
106. ALA-D (delta-amino-leuvulinic acid dehydratase) - porphobilinogen (PBG) formed per unit time and protein (nmol/l PBG/mg protein/min)
107. Cytochrome P450 1A activity /EROD (resorufin production; pmol/min/mg protein)
108. DNA adducts - number of adducted nucleotides per number of undamaged nucleotides, but also analysed as diagonal radioactive zones, DRZs (composite of multiple overlapping DNA adducts)
109. PAH metabolites - e.g. 1-hydroxypyrense or 1-hydroxyphentanthrene (ng/mg)
110. LH (Liver Histopathology) - occurrence of changes
111. MLN (Macroscopic Liver Neoplasm) - visible tumors on the surface of fish livers
112. Externally visible fish diseases - different types; FDI (Fish Disease Index) is calculated based on EVD (externally visible diseases), MLN, LH.
113. Reproductive success in eelpout - mean prevalence malformed fry, late dead fry, early dead fry and total abnormal fry. Different malformation classes.
114. VTG (vitellogenin) - concentration in blood plasma (ng/ml), of different types; in male
115. VTG (vitellogenin) - concentration in blood plasma (ng/ml), in female
116. Intersex in male fish - intersex prevalence (presence/absence)
117. Spiggin
118. Micronucleus assay - permanent and hereditary double DNA strand breaks (frequency of MN (FMN%) and frequency Nucleus abnormalities (FNA) - need to compare samples with a blank)
119. Amphipod embryo malformation - number (ratio) of malformed embryos
120. Stress proteins (Hsp) - amount of protein (semi quantitative), relative density units
121. Acetylcholinesterase (AChE) assay - AChE inhibition (nmol/min and mg protein)
122. Comet assay - tail moment, % DNA tail, length
123. Mussel histopathology (gametogenesis) - cell type composition (digestive gland epithelium), digestive tube epithelial atrophy and thinning, lysosomal alterations and inflammation
124. Stress on stress - anoxic/aerial survival (LT50 and TMM, time to maximum mortality)
125. SfG, Scope for Growth - alterations in energy available for growth and reproduction
126. Benthic diatom malformation - number (frequency) of malformed valves
127. Egg shell thinning of bird eggs

128. Sea eagle productivity
129. Pregnancy rate in seal
130. Genes involved in xenobiotic biotransformation and regulation (e.g. cytochrome 1A, AhR, ugt, metallothioneins)
131. Genes involved in oxidative stress (e.g. gpx, cat, HSPs), apoptotic response (e.g bax, p53, caspase), DNA repair (e.g nucleotide-excision repair xpa and xpc genes)
132. Mentum deformation in chironomids
133. Lipid peroxidation (LPO)
134. Protein carbonylation
135. P-glycoprotein efflux (P-gp)

Linking biomarkers to MoA and WFD relevance

The relationship between the above identified biomarkers and the type of MoA and substances they respond to, as well as the ecological and/or human toxicology relevance anticipated from effect observations, are illustrated in Table II.1.

Table II.1: Categorisation of the biomarkers in the inventory of effect and exposure biomarkers, information about the type of MoA or effect they respond to, an assessment of the level of relevance of the response and the specificity of the biomarker.

No	Biomarker	Effect or exposure?	Type of MoA	MoA – specified (if relevant)	Biological organisational level of the response	General or specific biomarker	Ecological (or human) relevance of observed response using the particular EBM
1	Imposex VDSI	Effect	tissue changes - reproduction	imposex - imposition of male sex characteristics on females	tissue and individual	Specific (TBT)	very high - related to reproduction and measured at high organisational level, extensive field effects observed and that were related to population decline
2	Imposex RPSI	Effect	tissue changes - reproduction	imposex - imposition of male sex characteristics on females	tissue and individual	Specific (TBT)	very high - related to reproduction and measured at high organisational level, extensive field effects observed and that were related to population decline
3	LMS	Effect	detoxification/ internal regulation/immune function	lysosomal stability - destabilitsation of lysososomes means disturbed degradation (of material taken up into the cell by endocytosis) and regulation of the catabolic rate of cellular macromolecules, proteins in particular	cellular and lower	General (can respond to e.g. PAHs, metals, OC, redox cycling compounds)	moderate - related to many functions at cellular level and can cause different types of effects at individual level, including lethality
4	MT	Exposure	detoxification/ internal regulation	MT induction: regulation of the intracellular concentrations of essential and non-essential metals. MTs provide protection against oxidative stress	cellular and lower	Moderately specific (metals, in particular Cu, Zn, Cd but also other chemicals that can induce oxidative stress - they are also metallothionein inducers)	low-moderate
5	ALA-D	Effect	Hb synthesis	Hb synthesis	cellular and lower	Specific (Pb)	moderate - related to cellular function and can in the long run lead to effects at individual level
6	Cytochrome P450 1A activity /EROD	Exposure	detoxification/	Ah-receptor activation/detoxification	cellular and lower	Specific (Ah receptor agonists such as dioxins,	low-moderate

No	Biomarker	Effect or exposure?	Type of MoA	MoA – specified (if relevant)	Biological organisational level of the response	General or specific biomarker	Ecological (or human) relevance of observed response using the particular EBM
			internal regulation			planar PCBs, PAHs. At high levels - can be inhibited too; as well as from metals and oestrogens)	
7	DNA adducts	Effect	genotoxicity/ mutagenicity	formation of DNA adducts; includes uptake, metabolism and repair	cellular and lower	Moderately specific (genotoxic compounds such as PAHs, known to form adducts)	moderate/high - related to DNA damage and thus cellular function, can in the long run lead to different types of effects at individual level
8	PAH metabolites	Exposure	detoxification/ internal regulation	analysis of metabolites; final stage of biotransformation	cellular and lower	Specific (PAH)	low-moderate
9	LH	Effect	tissue changes (histopathology)	five different classes - non specific, early non neoplastic, foci of cellular alteration, benign neoplasms, malign neoplasms; preferably also type of lesions (fibrosis, granuloma, apoptosis…). Part of FDI	tissue and individual	General (but can be diagnostic depending on lesion; included in "OSPAR PAH EBMs")	high - tissue alterations possibly indicating malfunction
10	MLN	Effect	tissue changes (histopathology)	histologically confirmed cases of macroscopic liver neoplasms (malign and benign) but not including pre-neoplastic stages. Part of FDI	tissue and individual	Moderately specific (cancer-inducing chemicals; included in "OSPAR PAH EBMs")	high - tissue alterations possibly indicating malfunction
11	Externally visible fish diseases	Effect	tissue changes (histopathology) and disease (immunological)	different categories of EVD (Externally visible Diseases) - an index is based on MLN, LH as well and also impact on host and size, sex and season - for MLN and LH also age	tissue and individual	General (most frequently viruses/bacteria but other factors (such as chemicals) can influence immune system)	high - tissue alterations possibly indicating malfunction

No	Biomarker	Effect or exposure?	Type of MoA	MoA – specified (if relevant)	Biological organisational level of the response	General or specific biomarker	Ecological (or human) relevance of observed response using the particular EBM
12	Reproductive success in eelpout	Effect	reproduction - lethality, malformation, sex ratio	malformation, survival and sex ratio of fry (viviparous organism)	off spring	General	very high - related to reproduction and measured at high organisational level, field effects observed in locally impacted areas
13	VTG males	Effect	endocrine disruption - sex hormones (reproduction)	sex hormone disruption - expression in males: estrogenicity	cellular and lower	Specific (xenooestrogens, such as EE2 (more potent than E2), weak activity from alkylphenols, some phthalates, parabens, phytosterols).	moderate/high - related to cellular function and can in the long run lead to effects at individual level and cause impaired reproduction
14	VTG females	Effect	endocrine disruption - sex hormones (reproduction)	sex hormone disruption - inhibition in females: anti-estrogenicity	cellular and lower	Specific (see above?)	moderate/high - related to cellular function and can in the long run lead to effects at individual level
15	intersex in male fish	Effect	tissue changes - reproduction	intersex - imposition of female sex characteristics on males	tissue and individual	Specific (oestrogenic substances, such as estrogenic steroids (estrone, estradiol, ethinyl estradiol) and/or phenolic compounds (alkylphenols and their ethoxylates). If observed in marine top predator fish: could be the result from biomagnification of weak estrogen PBTs such as OCs and brominated flame retardants)	very high - related to reproduction and measured at high organisational level, field effects observed in locally impacted areas
16	Spiggin	Effect	endocrine disruption - sex hormones (reproduction)	sex hormone disruption - androgenic effects in females	cellular and lower	Specific (androgens such as the pharmaceuticals levonorgestrel and noretisteron)	moderate/high - related to cellular function and can in the long run lead to effects at individual level and cause impaired reproduction
17	MN	Effect	genotoxicity/ mutagenicity	clastogenicity - permanent and hereditary double DNA strand breaks;	cellular and lower	Moderately specific (substances that cause double DNA strand breaks; cytogenetic damage	moderate/high - responds to clastogenic substances that in the long term can give rise to negative effects

No	Biomarker	Effect or exposure?	Type of MoA	MoA – specified (if relevant)	Biological organisational level of the response	General or specific biomarker	Ecological (or human) relevance of observed response using the particular EBM
						detects the activity of clastogenic and aneugenic test substances)	
18	amphipod embryo malformation	Effect	tissue changes - reproduction	embryo malformation	offspring	General (strong correlation observed between effects and conc of metals and org compounds in field collected sediments)	very high - related to reproduction and measured at high organisational level, field effects observed in locally impacted areas
19	stress proteins	Effect	oxidative stress	includes oxidative stress - cell ability to handle oxygen radicals etc that could damage DNA and proteins	cellular and lower	General	low-moderate
20	AChE	Effect	neurotoxicity - behaviour, fitness, survival, reproduction	AChE inhibition - Ach will accumulate - causes overstimulation of neuromuscular junctions	cellular and lower	Specific (OP and carbamate pesticides but can also respond to heavy metals and detergents)	high - related to cellular function and can lead to effects at individual level and cause lethality
21	Comet	Effect	genotoxicity/mutagenicity	DNA single and double strand breaks, depending on if neutral or alkaline protocol is applied	cellular and lower	Moderately specific (substances that cause double DNA strand breaks)	moderate/high - related to DNA damage and thus cellular function, can in the long run lead to different types of effects at individual level
22	mussel histopathology	Effect	histopathology (cellular level) - reproduction	gametogenesis	cellular and lower	General (has been related to PAH, PCB, heavy metals)	moderate/high - related to cellular function and can in the long run lead to effects at individual level and cause impaired reproduction
23	stress on stress	Effect	lethality	aerial survival	tissue and individual	General	high/very high - related to survival of individual
24	SfG	Effect	energy - reproduction, growth	alterations in energy available for growth and reproduction (J/h*g)	tissue and individual	General (including DEHP, aromatics, PCP, Cu, TBT, dichlorvos)	high/very high - related to growth of individual
25	benthic diatom malformation	Effect	malformation	malformation of valves	tissue and individual	General but effect was so far primarily found to correlate with metals	moderate/high - high organisational level endpoint and serious effect; malformations can be suspected to lead to later effects at population or

No	Biomarker	Effect or exposure?	Type of MoA	MoA – specified (if relevant)	Biological organisational level of the response	General or specific biomarker	Ecological (or human) relevance of observed response using the particular EBM
							community level, but population sensitivity to deformations not known.
26	Egg-shell thinning of bird eggs	Effect	reproduction	thinning of egg shell	offspring	Specific (DDT)	very high - related to reproduction and measured at high organisational level, extensive field effects observed and related to population decline, related to secondary poisoning
27	sea eagle productivity	Effect	reproduction	survival of off spring	offspring	General	very high - related to reproduction and measured at high organisational level, extensive field effects observed and related to population decline, related to secondary poisoning
28	pregnancy rate in seal	Effect	reproduction	pregnancy rate	tissue and individual	General (PCB)	very high - related to reproduction and measured at high organisational level, extensive field effects observed and related to population decline, related to secondary poisoning
29	Genes involved in xenobiotic biotransformatio n and regulation	Exposure	gene transcription-exposure	detoxification of contaminants	cellular and lower	General	low-moderate
30	Genes involved in oxidative stress, apoptotic response, Dna repair	Effect	gene transcription-effect	oxidative damage, DNA damage, apoptosis	cellular and lower	Moderately specific	low-moderate
31	Mentum deformation in chironomids	Effect	malformation	malformation of chironomids (mentum/mouth parts)	tissue and individual	General (including metals)	moderate/high - high organisational level endpoint and probably serious effect; malformations can be suspected to lead to later effects at population or community level,

No	Biomarker	Effect or exposure?	Type of MoA	MoA – specified (if relevant)	Biological organisational level of the response	General or specific biomarker	Ecological (or human) relevance of observed response using the particular EBM
							but population sensitivity to deformations not known.
32	Lipid peroxidation	Effect	cytotoxicity	oxidative damage - excess of ROS that generates oxidative degradation of lipids, resulting in cell damage	cellular and lower	General	low-moderate
33	Protein carbonylation	Effect	protein alteration	oxidative damage - excess of ROS that yields a reactive carbonyl moiety in proteins	cellular and lower	General (metals, organic pollutants)	low-moderate
34	P-glycoprotein efflux	Exposure	detoxification	cellular detoxification activity	cellular and lower	General (including pharmaceuticals, PAHs, metals, OCs, PFCs, algal toxins)	low-moderate

Practical aspects of biomarkers

Practical aspects, such as amount needed, storage possibilities and seasonal aspects for sampling biota to analyse biomarkers in the inventory are tabulated below. Whether effects are irreversible could also be important to know during planning (sampling frequency needed) and data interpretation[41].

(If information was not available to the activity, a "?" appears.)

[41] If effects, that are irreversible in nature, are observed, it cannot be ruled out that the exposure occured at a previous point in time. If exposure has seized but was previously larger, this could explain situations where chemical and toxicological data point in different directions.

Table II.2: Practical aspects of biomarker application in sampled biota.

No	Biomarker	Biota sampled	Sample needed	Pre-treatment and storage	Irreversible effect?	Sampled during particular season?
1	Imposex VDSI	soft tissues of female gastropods, typically Nucella, Nassarius, Buccinum, Neptunea, Hydrobia.	40 individuals (but also depends on size in case chemical analysis to be performed)	preferably not frozen and not stored longer than a week before analysis	irreversible - can occur during early life stages and persist years after - depends on life cycle of species.	not influenced by season; organisms could be difficult to find during cold winter conditions
2	Imposex RPSI	soft tissues of female gastropods, typically Nucella, Nassarius, Buccinum, Neptunea, Hydrobia.	40 individuals (but also depends on size in case chemical analysis to be performed)	preferably not frozen and not stored longer than a week before analysis	irreversible - can occur during early life stages and persist years after - depends on life cycle of species	could be influenced by season; organisms could be difficult to find during cold winter conditions
3	LMS	mussels, fish etc. Various tissues depending on organism: hemocytes (blood), liver, hepatopancreas, midgut, head kidney	20 individuals recommended (? - in more recent TIMES publication only 10; for practical reasons no more than 12 seems to be possible, if using NRR)	can be stored but depends on method (if using histochemical); if NRR no - analysis required within 24h	?	avoid reproductive season
4	MT	fish; liver cells	100 mg liver per analysis	can be stored	reversible - disappears rapidly after exposure has seized	one month outside spawning season
5	ALA-D	fish - blood (lysed red blood cells)	20-50ul	blood sample to be taken on live fish or within 5 min after killing, but can then be stored (liquid nitrogen)	reversible (?)	one month outside spawning season
6	Cytochrome P450 1A activity/ EROD	fish - liver (or gills)	100 mg liver per analysis (?, but could be higher according to standard)	can be stored but liver needs to be separated, store in liquid nitrogen	reversible	needs to be standardised (if to be compared between years)
7	DNA adducts	fish liver cells	100 mg liver per analysis	can be stored in liquid nitrogen or at -70	irreversible, but can be removed by cell death and repair processes	no
8	PAH metabolites	fish bile fluid	100 ul bile	can be stored frozen (-20)	reversible - the metabolites represent exposure during last hours-few days, at most 2 weeks	seasonal trend
9	LH	fish liver tissues	30-50 individuals	can be stored after fixation in 70% alcohol	?	outside spawning season

No	Biomarker	Biota sampled	Sample needed	Pre-treatment and storage	Irreversible effect?	Sampled during particular season?
10	MLN	fish liver tissues	50 individuals	can be stored after fixation - e g formalin	irreversible (?)	outside spawning season
11	Externally visible fish diseases	fish whole organism	large number of individuals needed - normally a low impact percentage	?	Depends on disease	Any season. However, natural occurance may differ between seasons.
12	Reproductive success in eelpout	fish whole organism	40 individuals (females)	cannot be stored	irreversible	autumn (depends on reproduction), in Sweden November
13	VTG males	blood plasma from male fish	12 individuals or more, constant size	centrifugation performed within 30 min, can then be stored at -20 C but not too long	reversible but t½ is days-weeks	outside breeding season, flounder before off shore migration
14	VTG females	blood plasma from female fish	20	Centrifugation performed within 30 min, can then be stored at -20 C but not too long	reversible but t½ is days-weeks	outside breeding season, flounder before off shore migration
15	intersex in male fish	gonads from male fish	50 individuals (higher numbers needed in marine than limnic, because lower frequency)	can be stored in Bouins solution or buffered formal saline	irreversible - can occur in younger life stages when sexual differentiation takes place; check also VTG (more recent exposure)	outside breeding season
16	Spiggin	female sticklebacks	?	?		
17	MN	fish, newt embryo, molluscs – tissue depends on species: gill cells, haemocytes, erythrocytes	1000-2000 for molluscs (20 individuals), higher for fish (10-20 individuals)	should be analysed within 3d	irreversible	any season
18	amphipod embryo malformation	amphipods	50 individuals	cannot be stored, need to be analysed alive	irreversible	late stage of embryo development (in SE: End of January)
19	stress proteins	any?	mg	can be stored if snap frozen on dry ice and stored at - 80C or liquid N	?	? (can be species dependent)
20	AChE	fish and ? – tissue depends on species (gills, muscle tissue, brain tissue)	low µL	can be stored if quick frozen and stored at - 80C	irreversible for organophosphate and carbamate	?

No	Biomarker	Biota sampled	Sample needed	Pre-treatment and storage	Irreversible effect?	Sampled during particular season?
21	Comet	fish - applicable to most eukaryotic cell types	a small number of cells (at least 100 cells per slide, each slide represents an individual; 10-20 individuals)	protocol recently made available for conserving fish erythrocytes	reversible	applicable alle the seasons but reproductive status may affect results
22	mussel histopathology	mussels	50 individuals	process asap but analysis is performed on frozen tissue sections	?	?
23	stress on stress	mussels	40 individuals	cannot be stored	?	?
24	SfG	mussels	?	?? (probably not possible to store)	? (perhaps reversible but substance depending?)	avoid spawning season, preferably during period of maximum growth (early summer-early autumn).
25	benthic diatom malformation	diatoms collected from biofilm	at least 400 valves; see EU standard sampling protocol	yes, and non-destructive analysis	irreversible	Autumn
26	egg shell thinning of bird eggs	bird eggs	?	?		Spring/summer
27	sea eagle productivity	sea eagle	?	?		Spring/summer
28	pregnancy rate in seal	seal	?	?		
29	Genes involved in xenobiotic biotransformatio n and regulation	fish (liver), invertebrates, bivalves (hepatopancreas)	mg	can be stored in liquid nitrogen or at -70	reversible	
30	Genes involved in oxidative stress, apoptotic response, Dna repair	fish (liver), invertebrates, bivalves (hepatopancreas)	mg	can be stored in liquid nitrogen or at -70	irreversible	avoid reproductive season
31	Mentum deformation in chironomids	chironomids	?	?		
32	Lipid peroxidation	several tissues depending on species	mg	can be stored in liquid nitrogen or at -70	Irreversible	avoid reproductive season
33	Protein carbonylation	several tissues depending on species	alt 1: mg alt 2: about 0.1 g of tissue. For field studies at least 20 individuals	alt 1: yes, if stored in liquid nitrogen or at -70 alt 2: yes	Irreversible	no but seasonal differences may appear

No	Biomarker	Biota sampled	Sample needed	Pre-treatment and storage	Irreversible effect?	Sampled during particular season?
				but sample should be stored at - 80°		
34	P-glycoprotein efflux	mussels (gills)	biopsy (4-6 mm)	cannot be stored, need to be alive	reversible	

Regulatory implementation aspects for biomarkers

The table below lists, for the biomarkers included in the inventory, available information on costs for analysis, availability of commercial providers, whether the biomarker has been included in regular monitoring programmes, and whether there are established assessment criteria and SOPs (such as international standards but also guidance documents or scientific publications frequently used).

Please be aware of this information not being complete or necessarily relevant to every MS. If commercial labs are shown to be available, this means that there is at least one commercial provider in at least one MS. Costs are only roughly estimated according to the following categorisation: low <200 Euro; moderate: 200-500; high 500-1000 Euro. Regarding assessment criteria, please refer to special section for explanation of how these were developed.

One type of SOP available for biomarkers are publications in the "ICES Techniques in Marine Environmental Sciences (TIMES)" series[42]. These documents provide details of methods and procedures relating to chemical and biological measurements in the marine environment. Most of the techniques described have been selected on the basis of performance in ICES or other international intercalibration exercises.

[42] http://ices.dk/publications/library/Pages/default.aspx#Default=%7B%22k%22%3A%22%22%2C%22r%22%3A%5B%7B%22n%22%3A%22owstaxIdPublicationType%22%2C%22t%22%3A%5B%22string(%5C%22%2300764ee64-19f6-4fb5-84a6-224dbed54d0b%5C%22)%22%2C%22%5C%22%C7%82%C7%824c307c233030373634656536342d313966362d346662352d383461362d323234646265643534643062 7c54696d6573%5C%22%22%5D%2C%220%22%3A%22and%22%2C%22k%22%3Afalse%2C%22m%22%3Anull%7D%5D%7D

Table II.3: Information on the biomarkers important in implementation contexts, used in this report to assess their "maturity".

No	Biomarker	Commercial laboratory available?	Analytical costs	Included in national monitoring programme or surveys?	Marine, limnic or both?	Assessment criteria established?	SOP
1	Imposex VDSI	Yes	Moderate	Included in JAMP under OSPAR SE, FR and ES: Included in regular national monitoring of marine environment Probably also other MSs (applied in 10 MSs in the MSFD initial assessment from 2012)	So far primarily studied on marine and brakish organisms, different species may be necessary for coastal vs off shore	OSPAR assessment criteria available. For e.g. *Neptunea antiqua* (red whelk) and Nucella lapillus (dog whelk) the EAC is 2.0 and the BAC is 0.3. For *Nassarius reticulatus* (netted dog whelk) the EAC is 0.3. There are also imposex values for Nucella, Nassarius, Buccinum and Hydrobia.	ICES TIMES 24[43] and 37[44]; JAMP guidelines.
2	Imposex RPSI	Yes	Moderate	SE: Included in regular national monitoring of marine environment ES: Imposex is included in MSFD monitoring and assessment Probably also other MSs (applied in 10 MSs in the MSFD initial assessment from 2012)	So far primarily studied on marine and brakish organisms, different species may be necessary for coastal vs off shore	No OSPAR values (but see above for VDSI and "imposex")	ICES TIMES 24[45]; JAMP guidelines.
3	LMS	Yes	Low-moderate (about 300 Euro?)	FR: considered for inclusion in D8 assessment under MSFD (fish and mussels)	All	OSPAR assessment criteria are available. Criteria are	ICES TIMES 56[46]

[43] Gibbs PE. 1999. Biological effects of contaminants: Use ofimposex in the dogwhelk (*Nucella lapillus*).
as a bioindicator of tributyltin pollution. . ICES Techniques in Marine Environmental Sciences No.24 37pp.
http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times24/TIMES24.pdf
[44] http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times37/TIMES37.pdf on intersex in perwinkle (imposex).
[45] Gibbs PE. 1999. Biological effects of contaminants: Use ofimposex in the dogwhelk (*Nucella lapillus*).
as a bioindicator of tributyltin pollution. . ICES Techniques in Marine Environmental Sciences No.24 37pp.
http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times24/TIMES24.pdf
[46] Martínez-Gómez, C., Bignell, J. and Lowe, D. 2015. Lysosomal membrane stability in mussels. ICES Techniques in Marine Environmental Sciences No. 56. 41 pp.http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times56/TIMES%2056.pdf Please note that

No	Biomarker	Commercial laboratory available?	Analytical costs	Included in national monitoring programme or surveys?	Marine, limnic or both?	Assessment criteria established?	SOP
				SE: National screening campaign (on blue mussels) performed recently; not included in regular national monitoring (yet); FI: included in D8 MSFD monitoring programme since 2014 ES: Included in MSFD monitoring and assessment (applied in 2 MSs in the MSFD initial assessment; from 2012)		independent of species but depends on method. For neutral red method: BAC 120 minutes and EAC 50 minutes. For histochemical method: BAC 20 minutes and EAC 10 minutes.	Intercalibration performed in 2013 [47]
4	MT	?	Low (?)	SE: Included previously in regular national monitoring of the marine environment but now samples are stored in the Swedish national specimen bank ES: Included in MSFD monitoring and assessment (applied in mussels, used in 3 MSs in the MSFD initial assessment; from 2012)	All	OSPAR BACs	ICES TIMES 26[48] OSPAR JAMP recommends any of three methods, (one without Hg?)
5	ALA-D	?	Low	? (no reported use in the MSFD initial assessment from 2012)	All	? No OSPAR values	ICES TIMES 34[49]

there is also a previous report (no 36) on the same technique (Moore & Lowe 2004): http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times36/TIMES36.pdf

[47] http://extra.lansstyrelsen.se/havmoterland/SiteCollectionDocuments/Publikationer/2013-70.pdf

[48] Hylland K, 1999. Biological effects of contaminants: Quantification of metallothionein (MT) in fish liver tissue. . ICES Techniques in Marine Environmental Sciences No.26; 25pp. http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times26/TIMES26.pdf

[49] Hylland K, 2004. Biological effects of contaminants: Quantification of δ-aminolevulinic acid dehydratase (ALA-D) activity in fish blood. ICES Techniques in Marine Environmental Sciences No 34, 18 pp.
http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times34/TIMES34.pdf

No	Biomarker	Commercial laboratory available?	Analytical costs	Included in national monitoring programme or surveys?	Marine, limnic or both?	Assessment criteria established?	SOP
6	Cytochrome P450 1A activity /EROD	Yes (?)	Low	SE: Included in regular national monitoring of marine environment ES: Included in MSFD monitoring and assessment (applied in 2 MSs in the MSFD initial assessment; from 2012)	All	BAC values established within OSPAR/ICES. Differs between species (available for dab, flounder, cod, plaice, megrim, dragonet, red mullet at different seasons). assessment criteria have been proposed (through HELCOM) when discussed as a pre core indicator	ICES TIMES 23[50] and 13[51] (microplate method)
7	DNA adducts	?	Moderate-high?	SE: Included previously in regular national monitoring of the marine environment but now samples are stored in the Swedish national specimen bank (One MS reported use in the MSFD initial assessment from 2012)	All	BAC and EAC values established, species dependent	ICES TIMES 25[52]
8	PAH metabolites	?	Low	SE: Tested in SE. Samples (gall bladder) are stored in the Swedish national specimen bank. ES: Included in MSFD monitoring and assessment	All	BAC and EAC values established, species dependent but all values not available for all species	TIMES, and scientific paper?

[50] Stabb and McIntosh, 1998. Biological effects of contaminants: Determination of CYP1A-dependent mono-oxygenase activity in dab by fluorimetric measurement of EROD activity. ICES Techniques in Marine Environmental Sciences No.23. 23 pp.
http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times23/TIMES23.pdf

[51] Galgani and Payne, 1991.Biological effects of contaminants: Microplate method for measurement ofethoxyresoruim-O-deethylase (EROD) in fish. ICES Techniques in Marine Environmental Sciences No. 13. 15pp.
http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times13/TIMES13.pdf

[52] Reichert WL, French BL, Stein JE, 1999. Biological effects of contaminants: Measurement of DNA adducts in fish by 32p-postlabelling. ICES Techniques in Marine Environmental Sciences No 25. 52pp.
http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times25/TIMES25.pdf

No	Biomarker	Commercial laboratory available?	Analytical costs	Included in national monitoring programme or surveys?	Marine, limnic or both?	Assessment criteria established?	SOP
				FR: considered for inclusion in D8 assessment under MSFD (One MS reported use of 1-hydroxypyrene in fish in the MSFD initial assessment from 2012)			
9	LH	?	?	SE: liver, gut, gonads and spleen are stored (for future analysis if necessary) (Two MSs reported use of "fish liver pathologies" in the MSFD initial assessment from 2012)	All	EAC available but related to FDI (> or = 2). For nonspecific: related to trend in FDI	ICES TIMES 38[53]
10	MLN	?	?	SE: liver, gut, gonads and spleen are stored (for future analysis if necessary) (Two MSs reported use of "fish liver tumors" in the MSFD initial assessment from 2012)	All	EAC available but related to FDI (> or = 2)	?, scientific paper?
11	Externally visible fish diseases	?	?	SE: not included but pilot ongoing. FR: considered for inclusion in D8 assessment under MSFD (Two MSs reported use of FDI in the MSFD initial assessment from 2012)	All (but in practice used on flounder etc.)	EACs available but varies between different types and sex	paper?
12	Reproductive success in eelpout	Yes (?)	High	SE: Included in regular national monitoring of marine environment (One MS reported use of "% deformed fish larvae" in the 2012 MSFD initial assessment)	Marine	BAC and EAC established within OSPAR/ICES (WGBEC), för malformed fry, late dead fry, early dead fry and total abnormal fry (Jakob Strand, DK)	scientific publication? (https://www.slu.se/globalassets/ew/org/inst/aqua/externwebb/k-

[53] Feist SW, Lang T, Stentiford GD, Köhler A. 2004. Biological effects of contaminants: Use of liver pathology of the European flatfish dab (*Limanda limanda* L.) and flounder (*Platichthys flesus* L.) for monitoring. ICES Techniques in Marine Environmental Sciences No 38. 50 pp. http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times38/TIMES38.pdf

No	Biomarker	Commercial laboratory available?	Analytical costs	Included in national monitoring programme or surveys?	Marine, limnic or both?	Assessment criteria established?	SOP
							lab/provfiske-vid-kusten/undersoknin gstyp-tanglake-20141216.pdf)
13	VTG males	?	Low	SE: Included in regular national monitoring of marine environment (one MS applied results in the 2012 MSFD initial assessment)	All	BAC established within OSPAR, for flounder and cod; see also HELCOM candidate indicator report	ICES TIMES no 31 (RIA and ELISA)[54] (possible to measure either protein (VTG) or gene expression; JAMP recommends protein; can detect conc below 10 ng/ml.)
14	VTG females	?	?	? SE: pilot study performed	?	No OSPAR values.	?
15	Intersex in male fish	?	moderate? (combined analysis of gonads and liver gives higher cost effectiveness)	FR: considered for inclusion in D8 assessment under MSFD (one MS applied results in the 2012 MSFD initial assessment)	All	BAC for dab, flounder, cod, red mullet, eepout. According to Davies & Vethak (2012) a >5% prevalence would be considered the cut-off point for definition of an affected population	Bateman et al 2004 [55](this is the paper refered to from Davies & Vethak, 2012)
16	Spiggin	?	?	?	?	No	
17	MN	?	Low	Buschini et al., 2004; Klobucar et al., 2010; Gutierrez et al., 2015. SE: Tested in a 3-year pilot study on fish, results not yet evaluated	All	OSPAR BAC values established. BACs differ between species and refers to certain tissue:	OECD 474, APAT ISO 21427-1:2006

[54] Scott AP and Hylland K, 2002. Biological effects of contaminants: Radioimmunoassay (RIA) and enzyme-linked immunosorbent assay (ELISA) techniques for the measurement of marine fish vitellogenins. ICES Techniques in Marine Environmental Sciences No.31. http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times31/TIMES31.pdf

[55] Bateman KS, Stentiford GD, and Feist SW, 2004. A ranking system for the evaluation of intersex condition in European flounder (*Platichthys flesus*). Environmental Toxicology and Chemistry, 23: 2831–2836.

No	Biomarker	Commercial laboratory available?	Analytical costs	Included in national monitoring programme or surveys?	Marine, limnic or both?	Assessment criteria established?	SOP
				ES: Included in MSFD monitoring and assessment (in mussels) FR: considered for inclusion as part of D8 assessment under MSFD (Two MSs applied results from MN in mussels in the 2012 MSFD initial assessment)		*Mytilus edulis* (blue mussel), for gill 2.5% and for blood 2.5% *Mytilus galloprovincialis* (Mediteranean mussel), for blood 3.9 % *Mytilus trossulus* (bay mussel), for blood 4.5% *Platichthys flesus* (flounder), blood 0.3% *Limanda limanda* (dab), blood 0.5% *Zoarces viviparus* (eelpout) blood 0.4% *Gadus morhua* (cod) blood 0.4 % *Mullus barbatus* (red mullet) blood 0.3%	(ISO 21427-1 Nov 2006[56]
18	Amphipod embryo malformation	?	Low	SE: Included in regular national monitoring of marine environment (Two MSs applied results in the 2012 MSFD initial assessment)	All but depends on species (in SE: Monoporeia affinis - available in Baltic Sea, but also lakes below highest coastline)	Yes, BAC and EAC values established. But method not included in ICES integrated strategy	TIMES, paper 41[57]
19	Stress proteins	?	Low	?	All	No	Western blot or ELISA (?)

[56] Qualité de l'eau - Évaluation de la génotoxicité par le mesurage de l'induction de micronoyaux - Partie 1 : évaluation de la génotoxicité à l'aide de larves d'amphibiens.
[57] Sundelin B et al., 2008. Biological effects of contaminants: the use of embryo aberrations in amphipod crustaceans for measuring effects of environmental stressors. ICES Techniques in Marine Environmental Sciences No. 41. 27 pp.
http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times41/TIMES41.pdf

No	Biomarker	Commercial laboratory available?	Analytical costs	Included in national monitoring programme or surveys?	Marine, limnic or both?	Assessment criteria established?	SOP
20	AChE	?	Low	Fulton and Key, 2001; Dellali et al., 2001; Fernando et al., 2005; Monteiro et al., 2007. Tested in SE ES: Included in MSFD monitoring and assessment (in fish and mussels) FR: considered for inclusion in D8 assessment under MSFD SE: Included in regular national monitoring of marine environment (applied in mussels in 3 MSs in the MSFD initial assessment; from 2012)	All, present in most Animals	Yes, both BAC and EAC, differs between species.	ICES TIMES 22[58]
21	Comet	?	Low/moderate	Pavlica et al., 2001; Buschini et al., 2004; Boettcher et al., 2010; Scalon et al., 2010; Klobucar et al., 2010; Parolini et al., 2013 ?	All, limnic fish more frequently so far	Yes, BAC and depends on species: *Mytilus edulis* (blue mussel): 10%, *Gadus morhua* (cod): 5% and *Limanda limanda* (dab): 5%.	ICES TIMES 58[59]
22	Mussel histopathology	?	?	?	? (also limnic mussels?)	BAC and EAC, varies between type of effect	No formal SOP established but a common reference in this context is Peters, 1988[60]
23	Stress on stress	?	Low	ES: Included in MSFD monitoring and assessment	? (also limnic mussels?)	Yes, BAC is 10 days and EAC is 5 days for Mytilus (blue mussels).	No formal SOP established but the method is

[58] Bocquene G, Galgani F, 1998. Biological effects of contaminants: Cholinesterase inhibition by organophosphate and carbamate compounds. ICES Techniques in Marine Environmental Sciences No. 22
http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times22/TIMES22.pdf
[59] Bean and Akcha, 2016. Biological effects of contaminants: Assessing DNA damage in marine species through single-cell alkaline gel electrophoresis (comet) assay. ICES Techniques in Marine Environmental Sciences No. 58, 21 pp.
http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times58/TIMES%2058.pdf
[60] Peters EC, 1988. "Recent investigations of the disseminated sarcomas of marine bivalve molluscs. Amer. Fish. Soc.Spec. Publ.18: 74-92.

No	Biomarker	Commercial laboratory available?	Analytical costs	Included in national monitoring programme or surveys?	Marine, limnic or both?	Assessment criteria established?	SOP
				(no MSs used in 2012 MSFD initial assessment but one MS had established targets)			considered very simple. A common reference in this context is Veldhuizen-Tsoerkan et al., 1990[61]
24	SfG	?	?	ES: Included in MSFD monitoring and assessment (one MS used results in 2012 MSFD initial assessment)	? (also limnic mussels?)	Yes, BAC and EAC for Mytilus is 25 and 15 J h\(^{-1}\) g\(^{-1}\) respectively	ICES TIMES 40[62]
25	Benthic diatom malformation	Yes	low	SE: No national survey done yet but regional campaigns performed in SE (see Kahlert 2012)	Limnic (both streams and lakes) but ture marine	Assessment criteria in SE, but to be used as risk indication (>2% malformations: risk)	Sampling and storage is standardised (EN 13946:2014). Method to identify malformations is included in "Undersökningstyp Påväxt i sjöar och vattendrag – kiselalgsanalys"[63] & Kahlert M, 2012. See also Lavoie et al., 2017.

[61] Veldhuizen-Tsoerkan MBDA et al., , 1991. A field study on stress indices in the sea mussel Mytilus edulis. Application of the "stress approach" in biomonitoring. Arch. Environ. Contam. Toxicol. 21: 497-504.
[62] Widdows and Staff, 2006. BIOLOGICAL EFFECTS OF CONTAMINANTS: MEASUREMENT OF SCOPE FOR GROWTH IN MUSSELS. ICES Techniques in Marine Environmental Sciences No. 40, 34pp.
http://ices.dk/sites/pub/Publication%20Reports/Techniques%20in%20Marine%20Environmental%20Sciences%20(TIMES)/times40/TIMES40.pdf
[63] https://www.havochvatten.se/download/18.6d9c45e9158fa37fe9f8d1a2/1482318545797/undersokningstyp-pavaxt-i-vatten-kiselalgsanalys-version-3-2.pdf

No	Biomarker	Commercial laboratory available?	Analytical costs	Included in national monitoring programme or surveys?	Marine, limnic or both?	Assessment criteria established?	SOP
26	Egg shell thinning of bird eggs	No, not known	Very high (but difficult to separate costs from national monitoring of population productive parameters)	SE: Included in regular national monitoring of marine environment (One MS used results in the MSFD initial assessment from 2012)	Marine	Assessment criteria in SE (for MSFD use): 0.59 mm (based on eggs sampled in 1856-1935)	Helander et al 2002
27	Sea eagle productivity	No, not known	Very high (but difficult to separate costs from national monitoring of population productive parameters)	SE: Included in regular national monitoring of marine environment (One MS used results from "bird breeding success" in the 2012 MSFD initial assessment)	Marine	Assessment criteria established in HELCOM. Productivity: The threshold value is 0.97 nestlings. Brood size: The threshold value is 1.64 nestlings. Breeding success: The threshold value is 0.59 (59%).	Naturvårdsverket, 2004 HELCOM 2012[64] In Sweden based on the assessment of nests 15 km or less from coast line
28	Pregnancy rate in seal	No, not known	Very high (but difficult to separate costs from national monitoring of population productive parameters)	SE: Included in regular national monitoring of marine environment (No MSs used results in the D8 MSFD initial assessment from 2012 but one MS defined GES and environmental targets for "reproductive health of marine mammals")	Marine	Assessment criteria established in Sweden (HVMFS 2012:18) for MSFD use (for grey seal in the Baltic Sea): good environmental status when pregnancy rate is above 80%.	Naturvårdsverket 2004 HELCOM 2012[65]

[64] Naturvårdsverket 2004. Handledning för miljöövervakning. Undersökningstyp: Havsörn, bestånd. Programområde Kust och hav. Version 1:0: 2004-05-26. 2 HELCOM 2012. Baltic Sea Environmental Proceedings No. 129B. The development of a set of core indicators: Interim report of the HELCOM CORESET project. Part B. Descriptions of the indicators. Helsinki Commission. See also http://www.helcom.fi/Core%20Indicators/White-tailed%20sea%20eagle%20productivity%20HELCOM%20core%20indicator%202018.pdf

[65] Naturvårdsverket 2004. Naturvårdsverket 2004b. Handledning för miljöövervakning; Undersökningstyp: Patologi hos gråsäl, vikaresäl och knubbsäl. Programområde Kust och hav. Version 1:0: 2004-01-23. HELCOM 2012. Baltic Sea Environmental Proceedings No. 129B. The development of a set of core indicators: Interim report of the HELCOM CORESET project. Part B. Descriptions of the indicators. Helsinki Commission.

No	Biomarker	Commercial laboratory available?	Analytical costs	Included in national monitoring programme or surveys?	Marine, limnic or both?	Assessment criteria established?	SOP
29	Genes involved in xenobiotic biotransformation and regulation	?	Low	?	All	?	Scientific literature
30	Genes involved in oxidative stress, apoptotic response, Dna repair	?	Low	?	All	?	Scientific literature
31	Mentum deformation in chironomids	Yes (?)	?	SE: used occasionly in the assessment of contaminated sites (sediments)	Limnic	No?	Scientific literature
32	Lipid peroxidation	?	Low	?	All	No?	Scientific literature
33	Protein carbonylation	?	Low	Prevodnik et al., 2007; Almroth et al., 2008; Parolini et al., 2013; Toni et al., 2011; Cattaneo et al., 2012	All	No?	Scientific literature
34	P-glycoprotein efflux	?	Low	?	All	?	Scientific literature

ANNEX III. Trigger value procedures

Sensitivity and specificity analysis of effect-based trigger-values (EBT) regarding the screening of known chemical and *in vivo* mixture risks

Background

This part presents a specificity and sensitivity analysis of *in vitro* EBMs for the detection of ER-agonists in combination with EBTs. The activation of the ER by ER-agonists is a relevant MoA that is related to adverse effects at the population level. The three WL compounds estrone (E1), 17β-estradiol (E2) and 17α-ethinylestradiol (EE2) activate the ER in an additive way. The respective EBT used for the assessment of the results obtained by *in vitro* EBMs has to be defined in a way to maximise sensitivity and specificity for known mixture risks based on chemical analysis for the WL compounds.

The results obtained using *in vitro* EBMs are benchmarked not only against the results of chemical analysis, but also against those from a transgenic fish model (*D. rerio*, EASZY assay Brion et al. 2019, in order to characterise their predictive power for effects at higher biological levels, and their potential to serve as an 'early warning' signal for *in vivo* effects. Although the stimulation of the ER in brain tissue that is detected by the transgenic fish model is not an adverse effect per se, it clearly demonstrates that ER agonists present in a sample are bioavailable, taken up by the organism, and distributed within the organism and across the blood-brain barrier, resulting in concentrations that are high enough to trigger the activation of the ER in the brain above control levels, possibly causing further effects in the fish.

The sensitivity and specificity analysis in this Annex is based on published data for 33 surface- and waste water samples analysed in the EU estrogen monitoring project (see Kase et al. 2018, Könnemann et al. 2018) using five different *in vitro* EBMs (ERα-CALUX, MELN; p-YES, Hela 9903 and ER GeneBlazer) and three chemical analytical methods based on hr-LC/MS for the quantification of E1, E2 and EE2. Furthermore, all samples were also tested in the abovementioned transgenic fish model (*D. rerio*, EASZY assay Le Fol et al. 2017 and Brion et al. 2019). In previous studies it was demonstrated that the expression of the green fluorescence protein (gfp) fused to the cyp19a1b-gene reflects the behaviour of the endogenous brain aromatase gene in zebra fish (D. rerio, EASZY tg cyp19a1b-GFP transgenic fish line) and thus its brain-specific response to hormonal regulation. By this means, this transgenic fish line allows the detection of ER-agonists in environmental samples including the toxicokinetics of compounds present in the sample. The induction of the brain aromatase gene is not yet an adverse apical endpoint per se but it clearly indicates the impact of ER-agonists on a key molecular initiating event in the context of a whole organism.

Methodology:

Step 1: The data from chemical analysis was used to calculate a chemical analytical cumulative risk quotient for each sample as follows:

$$RQ_{chem} = \frac{c_{E1}}{EQS_{E1}} + \frac{c_{E2}}{EQS_{E2}} + \frac{c_{EE2}}{EQS_{EE2}}$$

with

RQ_{chem}	cumulative risk quotient based on chemical analysis
c_i	concentration of the analytes E1, E2 and EE2 determined by hr-MS
EQS_i	proposed environmental quality standards for E1, E2 and EE2 (3600, 400 and 35 pg/L, respectively)

The rationale to calculate a cumulative risk quotient is the known additive behaviour of these three ER-agonists. The calculated cumulative risk quotients for the 33 samples are published in Kase et al. 2018. A cumulative RQ above 1 indicates a population relevant risk for aquatic species based on data from chemical analysis. The assessments based on *in vitro* results with different EBT scenarios were benchmarked against these cumulative RQs as described in 'step 3' (see below).

Step 2: The data from EASZY *in vivo* was assessed as follows:

If the EASZY-assay was stimulated significantly above the negative control (dimethyl sulfoxide, DMSO) in response to exposure of the sample, the sample was defined as active, i.e. the risk quotient (*in vivo*) was >1. The concentration – response curves were modelled according to a Hill equation using the Regtox 7.0.6 Microsoft Excel TM macro[66], and EC20 values were calculated. For active environmental samples, the estrogenic activity is expressed as an E2-equivalent concentration (EEQ) using the ratio EC20 of E2/EC20 active sample.

The limit of quantification (LOQ) that defines the threshold above which samples were assessed as positive was calculated as follows: LOQ = mean GFP expression in DMSO controls + 3 x S.D. This was done by taking into account all the individual responses from all the DMSO controls (mean of the mean). The value was then expressed in ng E2/l by extrapolation to a mean E2 standard curve (obtained from all the E2 standard curves generated). The LOQ in terms of an E2 equivalence concentration and taking account of an enrichment factor of 10 was determined as 6.3 ng/L E2 equivalents.

$$RQ_{in\ vivo} = \frac{BEQ}{AL}$$

with

$RQ_{in\ vivo}$	Risk quotient derived by *in vivo* analysis
BEQ	Biological equivalence concentration resulting from sample measurements
AL	Activation Limit for EASZY (6.3 ng/L E2 equivalents)

Step 3: Risk calculations from the selected *in vitro* EBM based on EBT

The results of *in vitro* EBM are also expressed in terms of a biological equivalence concentration (BEQ). For the selected *in vitro* EBM, the results are provided as E2-equivalence concentrations (EEQ) in ng/l. The EEQ value represents the combined effect

[66] http://www.normalesup.org/~vindimian/fr_index.html

of all ER-agonists present in the sample. The EEQ value is compared to the EBT by analogy with the chemical risk assessment.

$$RQ_{EBM} = \frac{EEQ}{EBT}$$

with

RQ_{EBM}	risk quotient based on *in vitro* EBM
EEQ	E2-equivalence concentration determined with an *in vitro* EBM
EBT	effect-based trigger value

In recent publications, EBTs for the assessment of estrogenic potentials in water samples were proposed (Jarosova et al. 2014, Kunz et al. 2015, van der Oost et al. 2017, Escher et al. 2018). These proposed EBTs do not differentiate between various *in vitro* EBMs that can be used for the detection of ER-agonists, i.e. all assay results are assessed against the same EBT. The use of one EBT for different EBMs detecting the same MoA might be problematic because of EBM-specific differences in relative potencies for bioactive compounds. Therefore, a given EBT might be suitable for the assessment of one *in vitro* EBM but over-protective or under-protective in combination with another *in vitro* EBM. If possible, EBM-specific EBTs should be derived and tested for their performance against proposed generic EBTs.

In the case of ER activation, alternative approaches are available to derive EBTs that are specific for different *in vitro* EBM. One method is presented by Escher *et al.* (2018). The definition of EBTs is specific for a single *in vitro* EBM taking into account its performance characteristics such as limit of detection for model compounds and relative potencies of model compounds in relation to the reference compound E2. The specific EBTs were determined by read-across from published data. EBTs for the following *in vitro* EBMs were given by Escher *et al.* (2018): ER GeneBLAzer, Hela 9903, MELN and ERα-CALUX.

The second method proposed to derive EBM-specific EBTs for ER activation is based on the mean value of the above cited generic EBTs, i.e 400 pg/l EEQ. This mean EBT is modified based on the sensitivity of the *in vitro* EBM, its variability and relative potencies of prominent reference compounds. The details of this approach, termed 'sensitivity factor approach' (SFA), are described in Annex III.3.

The selected proposals for EBTs to assess estrogenicity in water samples are summarised below.

Table III.1. Proposed EBTs in ng/l E2-equivalence concentration for the assessment of estrogenic potentials. na: not available.

In vitro EBM	**Low generic**	**Median generic**	**High generic**	**Read across (RA)[4] specific**	**Sensitivity factor approach (SFA)[5] specific**
ER Gene BLAzer	0.3[1]	0.4[2]	0.5[3]	0.340	0.400
Hela 9903				1.01	0.266
p-YES				na	0.266
MELN				0.370	0.266
ERα-CALUX				0.100	0.400

Step 4: Assessment of sensitivity and specificity

The assessment of the results obtained using the *in vitro* EBM in comparison with the different suggested EBTs is benchmarked against the calculated cumulative risk quotient based on the chemical analysis (step 1) and *in vivo* results (step 2), and expressed in terms of true positive (tp), false positive (fp), true negative (tn) and false negative (fn)[67] test results as shown in Table 4 and 5.

Table III.2: Definition of true negative (tn), true positive (tp), false positive (fp) and false negative (fn) results with data from chemical analysis as reference point.

	$RQ_{chem} < 1$	**$RQ_{chem} \geq 1$**
$EEQ < EBT \rightarrow RQ_{EBM} < 1$	true negative (t_n)	false negative (f_n)
$EEQ \geq EBT \rightarrow RQ_{EBM} \geq 1$	false positive (f_p)	true positive (t_p)

An example of this benchmarking is shown in Figure AIII.1 for EEQ-values obtained by the *in vitro* EBM ERα-CALUX with a generic EBT of 0.4 ng/l EEQ (= 400 pg/l). The R_{chem} values are given in log-space to achieve a symmetric representation of the values.

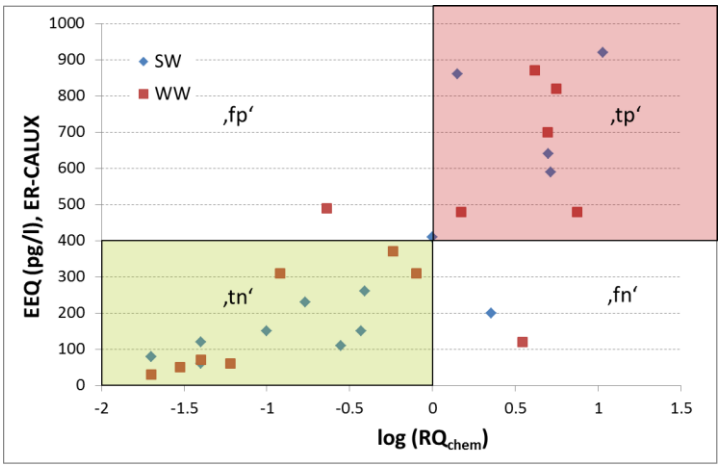

Figure AIII.1: Benchmarking of EEQ-values measured with the ERα-CALUX against RQ_{Chem}. An EBT of 0.4 ng/l EEQ (= 400 pg/l) was used. True negative results are located in the green box, indicating no risk based on chemical analysis and the *in vitro* EBM. True positive results are located in the red box, indicating risk based on chemical analysis and the *in vitro* EBM. False positive results are located in the upper left part and false negative results are located in the lower right part of the diagram. SW:= surface water sample, WW:= waste water sample, tp: true positive, fp: false positive, fn: false negative, tn: true negative.

A higher EBT would result in a lower number of 'false positive' results but in a higher number of 'false negative' results. The other way around: a lower EBT would result in a higher number of 'false positive' results but in a lower number of 'false negative' results.

[67] It has to be pointed out that the categories true/false positive and true/false negative are defined based on the chemical analysis restricted to the target compounds E1, E2 and EE2. This assessment does not necessarily reflect the real risk associated with a water sample since further ER-agonists may be present that are not detected by chemical analysis. Thus, the assessment 'false positive' results from the comparison with R_{Chem} that is an estimate of the real risk associated with a sample. The 'false negative'-results might be caused either by specific antagonistic compounds in the sample or by unspecific interferences with the *in vitro* EBM. In the first case the *in vitro* EBM would reflect the true estrogenic potential of the sample by taking agonistic and antagonistic mixture effects into account and the actual risk would be overestimated by the chemical analysis. The latter case would represent a real false negative test result and an existing risk would have gone undetected by the *in vitro* EBM. In this respect, sufficient control experiments and the definition of validity criteria are important to demonstrate the functionality of the *in vitro* EBM for a given sample. If validity criteria are not met, the sample cannot be assessed by the *in vitro* EBM. This situation is comparable to the presence of compounds interfering with a chemical analysis, e.g. due to ion suppression in mass spectrometry.

The sensitivity and specificity for the various combinations of *in vitro* EBM with EBT are calculated as follows:

$$Y_{specificity}(\%) = \frac{t_n}{t_n + f_p} \cdot 100$$

$$Z_{sensitivity}(\%) = \frac{t_p}{t_p + f_n} \cdot 100$$

with

$Y_{specificity}(\%)$	specificity in %
$Z_{sensitivity}(\%)$	sensitivity in %
t_n	true negative, i.e. no risk indicated by chemical analysis and *in vitro* EBM
t_p	true positive, i.e. risk indicated by chemical analysis and *in vitro* EBM
f_n	false negative, i.e. risk indicated by chemical analysis but not by *in vitro* EBM
f_p	false positive, i.e. no risk indicated by chemical analysis but by *in vitro* EBM

The same approach as described above can be used to assess the sensitivity and specificity of the proposed EBTs in combination with the selected *in vitro* EMBs to predict effects in the transgenic *in vivo* model. The definition of negative (tn), true positive (tp), false positive (fp) and false negative (fn) is done by analogy with the benchmarking against chemical analysis as shown in Table III.3.

Table III.3. Definition of true negative (tn), true positive (tp), false positive (fp) and false negative (fn) results with data from *in vivo* analysis with EASZY assay as reference point.

	$RQ_{in\ vivo} < 1$	**$RQ_{in\ vivo} \geq 1$**
EEQ < EBT → $RQ_{EBM} < 1$	true negative (t_n)	false negative (f_n)
EEQ ≥ EBT → $RQ_{EBM} \geq 1$	false positive (f_p)	true positive (t_p)

Results:

The raw data for this analysis are available in a supplementary Excel file, this annex is focused on the presentation and discussion of the main findings of the sensitivity and specificity analysis.

The performance of the assessment based on *in vitro* EBM is based on the three parameters

- Sensitivity
- Specificity
- Variability of sensitivity and specificity between different in vitro EBM

using data from:

- chemical analysis and a
- transgenic fish model

as reference for benchmarking the predictive power of a given *in vitro* EBM / EBT-combination.

Sensitivity: A risk indicated either by the cumulative risk quotient using concentration data for E1, E2 and EE2 or by the activation of the transgenic fish model should be captured also by the *in vitro* EBM. Otherwise, the *in vitro* EBM would fail to detect samples that are defined as problematic by the reference approach.

Specificity: The *in vitro* EBM in combination with the EBT should only flag samples that were identified as problematic by the reference approach. Otherwise, the *in vitro* EBM would overestimate the risk associated with a given sample.

Variability of sensitivity and specificity between different *in vitro* EBM: As described above, some generic EBTs are proposed in different publications that are claimed to be applicable to all *in vitro* EBM for the same MoA. An EBT might fit well, i.e. high sensitivity and specificity, for a given *in vitro* EBM but be insufficient for another EBM. The variability reflects the applicability of an EBT to a range of *in vitro* EBMs.

It is obvious that the parameters sensitivity and specificity have inverse tendencies. A very low EBT would result in 100% sensitivity, i.e. all samples assigned to be at risk by the reference approach were identified, but at 0% specificity because all samples assigned to be not at risk by the reference approach were identified as problematic by the *in vitro* EBM/EBT – combination. A very high EBT would result in an inverse situation with 0% sensitivity and 100% specificity. Because two categories have to be distinguished, the sensitivity and specificity of an *in vitro* EBM/EBT – combination has to be well above 50% to have any predictive power over flipping a coin. The optimal case would be a 100% sensitivity and 100% specificity. A balanced optimum would be an EBT that maximises sensitivity and specificity together.

A low variability of a generic EBT indicates a broad applicability of the proposed EBT for the *in vitro* EBM that were investigated. If the variability of proposed specific EBTs is lower, specific EBTs should be used to increase the predictive power of the *in vitro* EBM.

Figure AIII.2 summarises the results of the sensitivity and specificity analysis benchmarked against risk assessments based on chemical analysis (RQ(chem), top) and the use of the transgenic fish model (RQ(*in vivo*), bottom). The values for the proposed generic EBTs are distinguished from those of the specific EBT proposals.

Table III.4. Sensitivities and specificities in % for five *in vitro* EBM detecting the presence of ER agonists assessed by a proposed EBT of 0.4 ng/l E2-equivalence concentration (Kunz et al. 2015).

in vitro EBM	**EBT [ng EEQ/l]**	**RQ(chem)**		**RQ(*in vivo*)**	
		Sensitivity %	Specificity %	Sensitivity %	Specificity %
ER GeneBLAzer	0.4	81.3	82.4	**88.9**	**100**
Hela 9903		75	94.1	72.2	100
p-YES		87.5	70.6	83.3	73.3
MELN		93.8	64.7	100	80
ERα-CALUX		**87.5**	**94.1**	83.3	100
Mean		85.0	81.2	85.5	90.1
%Cv		7.5	14.8	10.6	12.8

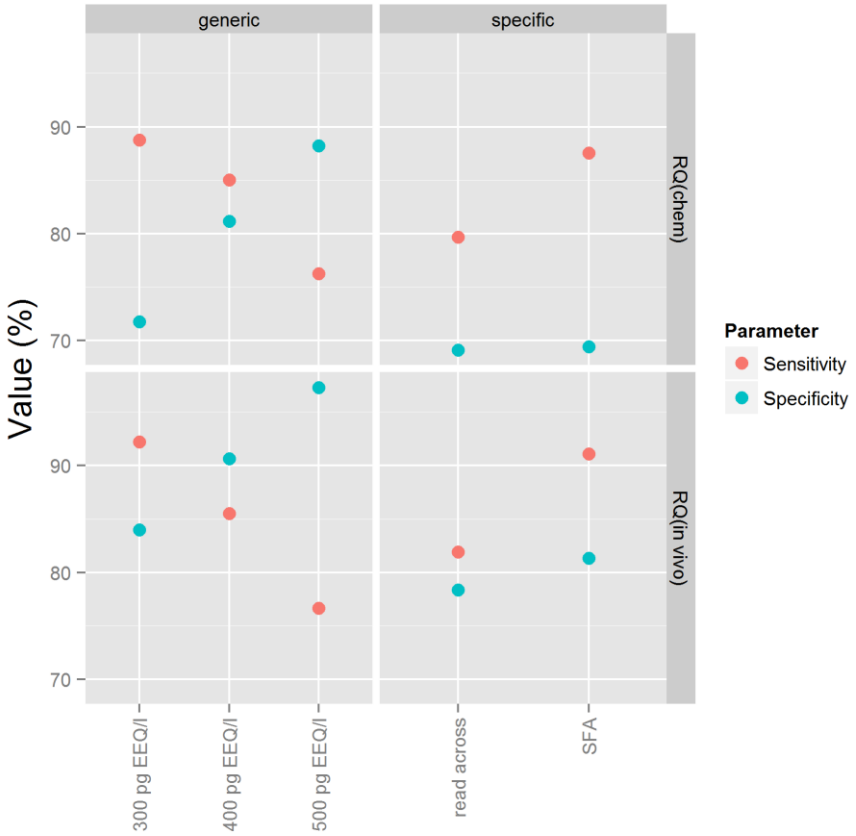

Figure AIII.2: Sensitivity and specificity analysis of *in vitro* EBM / EBT-combinations. Mean percentages for sensitivity (red dots) and specificity (blue dots) across all investigated *in vitro* EBMs are presented. Both parameters were calculated based on a comparison either with a risk assessment based on chemical analysis (RQ(chem), top) or results from a transgenic fish model (RQ(*in vivo*), bottom).

All EBT proposals proved to have a predictive power for the risk assessment based on chemical analysis of E1, E2 and EE2 as well as for the activation of the transgenic fish model (Figure AIII.2). However, specific differences in the performance can be observed. The mean sensitivity for RQ(chem) drops from about 89% for an EBT-proposal of 300 pg EEQ/l (Jarosova et al. 2014) to 76% for an EBT-proposal of 500 pg EEQ/l (van der Oost et al. 2017) whereas the mean specificity increases from 72% to 88%. Similar tendencies are to be observed for the benchmarking against RQ(*in vivo*). In this case, the mean sensitivity drops from 92% to 77% whereas the mean specificity increases from 84% to 97%. The best balance between sensitivity and specificity is reached for an EBT-proposal of 400 pg EEQ/l (Kunz et al. 2015). The generic EBT proposal of 400 pg EEQ/l showed a higher concordance compared to the specific EBT-proposals. The mean sensitivity and specificity were higher for the generic EBT-proposal of 400 pg EEQ/l than for the EBM-specific EBT-proposals based on the read-across approach. It has to be pointed out that in this case the calculated mean value was impacted strongly by one individual *in vitro* EBM, namely Hela 9903 with a proposed EBT of 1010 pg EEQ/l. In this case the sensitivity for RQ(chem) was only 38% and for RQ(*in vivo*) 33%. In contrast, specificities were high with values of 100% each. This indicates that the proposed EBT for this specific *in vitro* EBM was probably too high. Compared to the sensitivity factor approach (SFA) described in Annex III.3 the generic EBT-proposal of 400 pg EEQ/l had a lower sensitivity but a higher

specificity. This is most pronounced for RQ(chem) where the mean specificity for the SFA-approach is 70% and for the generic EBT-proposal 81%.

Figure AIII.3 shows the variability of sensitivity and specificity between different *in vitro* EBM assessed by the respective EBT-poposals.

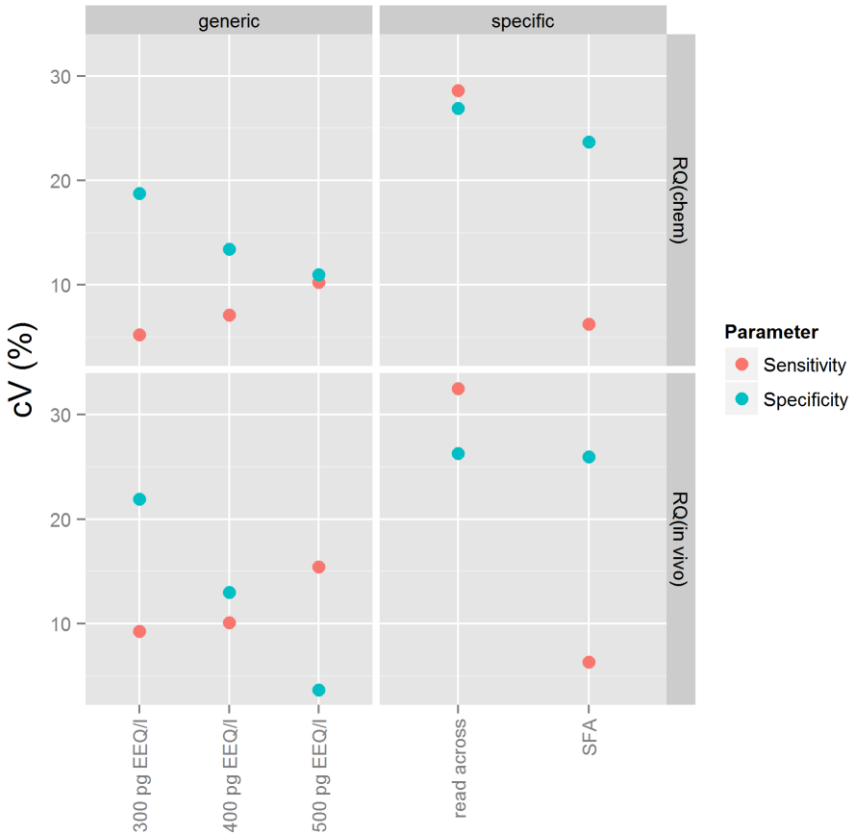

Figure AIII.3: Variability of sensitivity and specificity analysis of *in vitro* EBM / EBT-combinations. Variabilities for sensitivity (red dots) and specificity (blue dots) across all investigated *in vitro* EBMs are presented. Both parameters were calculated based on a comparison either with a risk assessment based on chemical analysis (RQ(chem), top) or results from a transgenic fish model (RQ(*in vivo*), bottom).

The lowest overall variabilities are observed for the EBT-proposals of 400 pg EEQ/l and 500 pg EEQ/l. In the case of the sensitivity factor approach, the variability for the determination of sensitivity was lower but for specificity higher. It has to be pointed out that the variability is not completely independent from the determination of sensitivity and specificity. For extreme EBTs resulting in e.g. 100% sensitivity for all *in vitro* EBMs, the variability for the determination of the sensitivity will be 0%. Thus, the assessment of variability has to include both sensitivity and specificity, and has a meaningful outcome only for EBT proposals resulting in a balanced sensitivity and specificity. As discussed above, generic EBT-proposals suffer from the inherent possibility that they might be not applicable to a selected *in vitro* EBM despite performing well with another *in vitro* EBM. In the example presented here, the generic EBT-proposal of 400 pg EEQ/l performed best with respect to a balanced sensitivity and specificity performance and a low variability over a range of *in vitro* EBMs. Based on previous discussions with water experts, this EBT was suggested in an international estrogen monitoring recommendation as a moderate and balanced option as well (Dulio and Kase 2017). Nevertheless, it is important to have tools to derive specific EBTs as proposed by Escher et al. 2018, and to use the SFA described in

Annex III.3 to derive EBM-specific EBTs in cases where a generic EBT-proposal results in high variabilities.

Discussion:

As presented a sensitivity and specificity analysis can be done to assess the performance of proposed EBTs in combination with *in vitro* EBMs. This approach is able to elucidate the power of *in vitro* EBMs to assess in combination with EBTs the likelihood that a sample is at risk according to its chemical composition and/or the likelihood of the occurrence of an unwanted effect at a higher biological level. This type of analysis is easy to perform, is not based on any assumptions, and is independent of expert judgement. However, it requires the respective data obtained by *in vitro* EBMs, chemical analysis and/or *in vivo* EBMs. Such data sets are not yet available for most *in vitro* EBMs but if an *in vitro* EBM is discussed as a possible candidate to be used as an element in water quality assessment it is recommended to perform a sensitivity and specificity analysis as outlined in this Annex.

In fact, such data sets can be used as training sets to define optimal EBT-proposals. This is done by maximising sensitivity and specificity for the chemical risk, and using the possibility to observe effects at higher biological levels, or both, as illustrated in Figure AIII.4 and described in detail by Brion et al. 2018 (in preparation). As an example, EEQ values in pg EEQ/l obtained by the *in vitro* EBM 'ER CALUX" were used. The cumulative positive assessments by RQ(chem) and RQ(*in vivo*) were plotted against the log(EEQ). The first positive assessment based on RQ(chem) occurs at an EEQ of 120 pg EEQ/l, the second at an EEQ of 200 pg EEQ/l. Up to these EEQ-levels, no positive *in vivo* result was observed. The highest EEQ at which no effect in the transgenic fish model was observed was 260 pg EEQ/l. The first positive result obtained by the transgenic fish model was observed at 310 pg EEQ/l. Above 370 pg EEQ/l, the cumulative positive assessments for RQ(chem) and RQ(*in vivo*) increase. Thus, an EBT that differentiates best between positive and negative assessments by the reference methods must lie between 260 and 310 pg EEQ/l. Based on this approach, the EBT-proposal for the ER CALUX was set to an average value between these two EEQ-values.

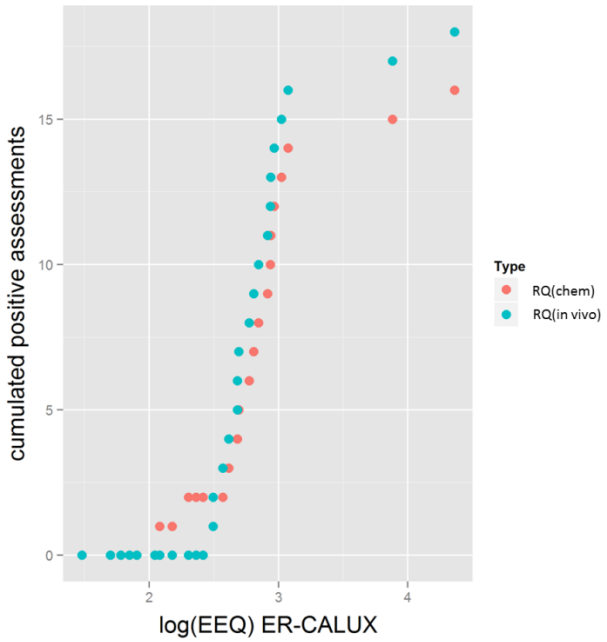

Figure AIII.4: Cumulative positive surface and waste water assessments vs. log (EEQ). The cumulative positive assessments of RQ(chem) (red dots) and RQ(*in vivo*) (blue dots) are plotted against the log(EEQ) measured by the ER CALUX.

Table III.5 summarises the EBT-proposals for all investigated *in vitro* EBM based on this approach together with the respective values for sensitivity and specificity.

Table III.5: Proposed EBTs in ng/l E2-equivalence concentration for the assessment of estrogenic potentials by Brion *et al.* 2018 and resulting sensitivity and specificity in %.

in vitro EBM	**Brion et al. 2018**	**RQ(chem)**		**RQ(*in vivo*)**	
		Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
ER GeneBLAzer	0.242	87.5	76.5	100	100
Hela 9903	0.182	93.8	82.4	93.3	94.4
p-YES	0.500	87.5	88.2	93.3	83.3
MELN	0.557	87.5	70.6	93.3	100
ERα-CALUX	0.283	87.5	76.5	100	100
	Mean	88.8	78.8	95.5	96.0
	%Cv	2.8	6.7	7.3	3.7

The EBTs shown in Table 6 result in the highest mean sensitivity and specificity. The variability between various *in vitro* EBMs is comparably low. In sum, the values indicate that it is possible to classify samples by means of *in vitro* EBM in good accordance with chemical analysis and results obtained by an organismic EBM. These proposed EBTs showed the highest predictive power and are recommended for the assessment of the respective *in vitro* EBM for the detection of ER agonists in water samples. However, these proposals have to be validated using an independent data set following the approach described above.

All EBTs resulting from the different approaches lie within in a small range of EBT-proposals from 0.1 ng EEQ/l up to 1.01 ng EEQ/l. The majority of proposals are in the range from 0.18 ng EEQ/l up to 0.56 ng EEQ/l. The majority of proposed EBTs were able to differentiate between the exceedance of EQS-values for E1, E2 and EE2 and effect induction at a higher biological level. This finding indicates a good overall consistency of the EBT-proposals.

Interestingly, the results from the in vitro EBMs show a higher agreement with the results obtained by the transgenic fish model in comparison to the results from the chemical analysis, i.e. the observed sensitivities and specificities for RQ (*in vivo*) are higher compared to RQ(chem) independent of the individual EBT-proposal. In fact, four samples showed an RQ(chem) < 1 but were assessed as positive by the in vitro EBMs in most cases. According to the definition these assignments were 'false positive' results. A further sample showed a risk based on chemical analysis but in most cases this sample was identified as a negative result. However, the assessment of these samples by the *in vitro* EBMs was in good agreement with the classification based on the transgenic fish model.

This finding indicates that there is a possibility of underestimating a risk based on the chemical assessment. This reinforces the need for a more holistic assessment of water quality because the chemical analysis of only three agonists of the ER might not capture significant other agonists present in the environment. This leads inevitably to lower specificities if the results from the chemical analysis are defined as the 'true' reference point. This example demonstrates the potential of *in vitro* EBMs for a more holistic way to assess water quality as acknowledged by the EU Water Directors.

Proposal for a tiered approach as a general framework to define EBTs

The presented methods and concepts used for the definition of EBT-proposals and the evaluation of these proposals can be used to build up a framework for the definition of EBTs based on available information to facilitate their use for e.g. prioritisation, screening or status assessment.

Similar to the definition of EQS as threshold values for chemical status assessments, the derivation of EBTs has to deal with inevitable uncertainties. As already discussed, uncertainties associated with the definition of EQS are caused by a lack of knowledge about the total composition of an environmental sample and possible mixture effects by the compounds present in the sample.

The EBT derivation approach used in this proposed concept is linked to EQS derivation which is protective for eco- and human toxicological risks according to the current knowledge level with the main difference that it also addresses known and unknown mixture risk and not only single-substance-based risks.

EBM have the advantage that they cover mixture effects and effects of unknown contaminants in an environmental sample as they measure the integrated effect caused by all compounds present in a sample. They can be used to address known and unknown mixture effects for population-relevant effects (Kase et al. 2018). In the case of biomarkers and many *in vitro* assays, specific molecular events are used as a marker for apical effects such as mortality, developmental or reproductive toxic effects. This can result in

uncertainties about the translation from a molecular effect to an adverse outcome in the organism. Depending on the knowledge about the investigated MoA, the level of uncertainty varies. For some MoA a link between *in vitro* results and adverse population relevant effects and risks can be established (Ankley et al. 2010, Matthiesen et al. 2017, Wittwehr et al. 2017, Kase et al. 2018).

A tiered approach for the derivation of EBTs is proposed that is driven by the availability of data for the given MoA. This allows, on the one hand, the initial definition of EBTs for a broad range of EBMs to be used for prioritisation and screening purposes and, on the other hand, the subsequent refinement of EBTs for prioritised EBMs to reduce uncertainties in water-body classifications. In general, uncertainties for both EQS and EBTs are reduced by increased quality of the underlying data. The following flow chart (Figure AIII.5) outlines the suggested approaches for the derivation of EBTs based on existing data.

EBTs derived from the highest tier available are based on a broader data basis resulting in reduced uncertainties. Therefore, it is recommended to check data availability in advance and follow the flow chart from Tier 4 to Tier 1.

The decision-making process for EBT derivation starts with testing the highest knowledge level Tier 4 before moving downwards to Tier 1 as follows:

Tier 4: The most powerful data basis for the derivation of EBTs is given by parallel *in vitro* and *in vivo* and chemical EQS compliance measurements. In other words, the *in vitro* effect quantified by an EBM is calibrated against mechanistically linked *in vivo* effects and quantified chemical-mixture effects and risks (Brion et al. 2018 in prep.). A transgenic fish line is used for the detection of ER-agonists in environmental samples including the toxicokinetics of compounds present in the sample. This approach combines the established population relevance according to the chemical assessments of single compounds and direct *in vivo* results covering further unknown compounds with the same MoA. By this means, the most direct link from *in vitro* results to unwanted endpoints of higher relevance and EQS compliance can be established. In principle, this approach can be transferred to other apical and adverse *in vivo* or *in vitro* effects of other MoA, e.g. PSII inhibition for herbicidal activity.

Advantages:

- Combines data from chemical monitoring and *in vivo* studies to define EBTs with the highest discriminative power based on real environmental samples including mixture effects of known and unknown compounds.

Limitations:

- Comparatively high efforts and labour costs for the generation of the required data
- Transferrable to other MoAs if a suitable *in vivo* model is available
- Calibration was performed only against one *in vivo* method with its own strengths and weaknesses.

Conclusion: This approach links cell-based EBMs to organismic EBMs and data from chemical monitoring resulting in a robust EBT to differentiate between samples ‘at risk’ and ‘not at risk’. Each EBM requires its own calibration with comparatively high efforts.

Tier 3: If both data from chemical monitoring of compounds with an associated EQS-based mixture risk and results from an EBM are available for the same samples, the results from the EBM can be calibrated against the combined risk quotient calculated for the detected

compounds in the sample (Kase et al. 2018 and Könemann et al. 2018). Moreover, if EBM-specific knowledge of sensitivity, variability and relative potencies is available, the EBT can be adjusted to the uncertainty of used methods by the application of a sensitivity factor. This approach was discussed and prioritised by participants of an EBT workshop in Switzerland in June 2017, in which some experts from the EBM activity participated (http://www.ecotoxcentre.ch/projects/aquatic-ecotoxicology/monitoring-of-steroidal-estrogens/). For this approach, it is recommended to use the maximal sensitivity factor range of the respective EBM-specific EBT according to Escher et al. 2018. The method is described in Annex III.4 in more detail.

Advantages:

- Based on EU EQS which indicate a population-relevant risk level for many species establishing a relevant point of departure (POD)
- Only four EBM-specific parameters are necessary and can be transferred to other MoAs where information about EQS and EBM is available, e.g. photosynthesis II inhibition and dioxin-like effects.
- Simple to implement in regulation as the use of one screening EBT for each endpoint plus sensitivity factor will result in a low number of EBTs which need to be implemented.
- Applicable with test-specific knowledge, such as Limit of Quantification (LOQ), Coefficient of Variation (CV), and Relative Effect Potencies (REP) for all new and existing methods possible.

Limitations:

- Depends on the availability of high-quality data, which is given only for selected, well characterised EBM
- Needs other approaches to set a first sensitivity range, but can be then applied independently.

Conclusion: This approach is recommended for all MoA for which no *in vivo* data are available, but for which monitoring for comparison with EQS is successfully applied and sufficient knowledge about performance characteristics of the respective EBM is available.

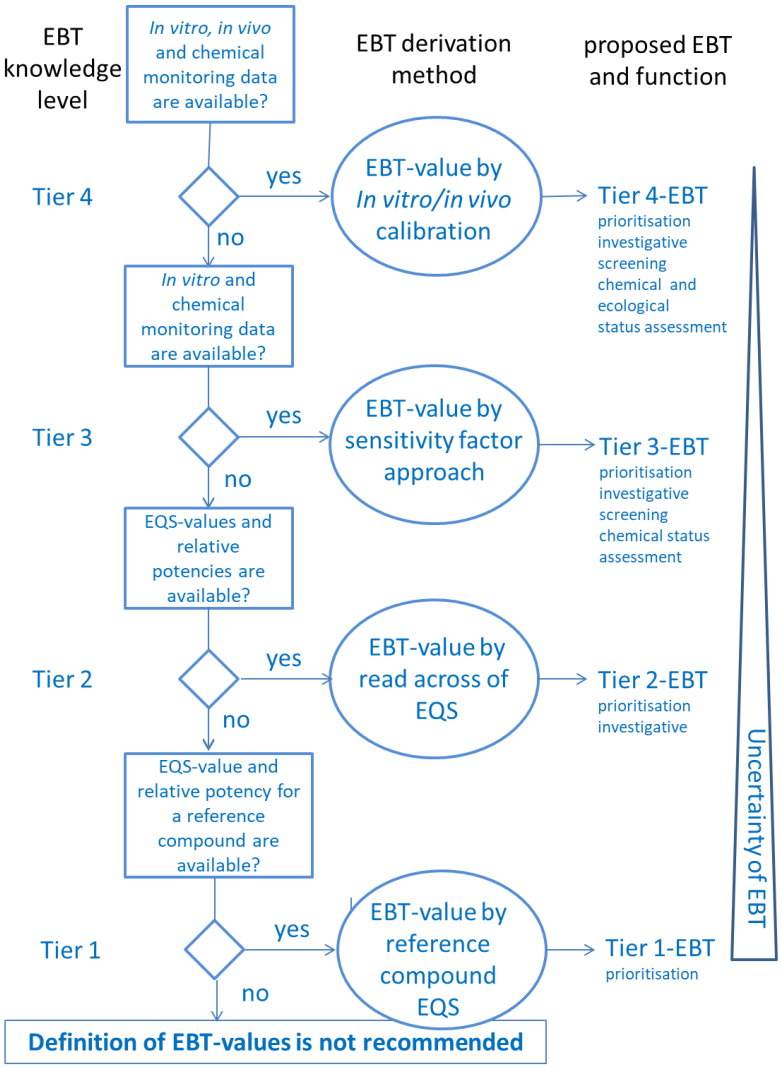

Figure AIII.5: Proposed concept for tiered Effect-Based Trigger value (EBT) derivation. EQS: Environmental Quality Standard.

Tier 2: If no experimental data from monitoring campaigns are available, EBTs can be derived by a read-across approach based on EQS-values of single compounds and the respective relative potencies of the compounds for the given EBM. (Escher et al. 2018). The proposed EQS-read across to define EBTs was applied to a large number of EBMs using more than 100 individual EQS-values (See Annex III.5).

Advantages:

- Based on multiple EQS indicating population-relevant risk levels for many species establishing a relevant point of departure (POD)
- Can be applied for MoAs for which EQS and REPs of EBM are available
- Based on existing data resulting in efficient and fast implementation.

Limitations:

- The approach depends on the quality and availability of data and possibly leads to higher uncertainties if only a limited number of compounds with associated EQS-values can be used for the EBT-derivation.

- Derived EBT depends on the selection of compounds to be included in the calculations. Stronger guidance is needed for the decision to select or de-select a compound for the EBT-derivation.
- Approach does not take into account EBM-specific inter-test CV and LOQs.

Conclusion: Recommended for all MoA for which no *in vivo* or chemical-analytical monitoring data are available.

Tier 1: If no read across approach is possible, the EQS of the reference compound in a certain mode of action (MoA), can be used as described above for an initial estimation of an EBT based on the respective EQS and the relative potency of the reference compound for the selected EBM. The reference compound should be either the most potent compound for the EBM or should be characterised as the main driver of the given biological effect in the environment. If no EQS for the reference compound is available, a certain BEQ level could be used instead of an EBT, but the interpretation of the results could be weakened so BEQs are not recommended for EBT derivation.

Advantages:

- Very simple, the same concentration of the reference compound can elicit an adverse effect at EQS level
- Can be applied for many MoAs for which EQS are available.

Limitations:

- The method does not take into account test-specific differences.
- The choice of the reference compound can significantly influence results.

Conclusion: Only recommended for prioritisation of effect levels if no other EBT derivation method is applicable.

Safety and screening value of tiered EBT for surface water assessments – MoA 'estrogen receptor activation'

The choice of EBT influences the safety and screening value of the EBM in surface water, illustrated as follows. The safety and screening value was calculated based on 80 surface-water measurements performed in the estrogen monitoring project using 5 different EBMs and compared with 48 high resolution LC/MS analytical measurements. For the calculations, the EBTs derived for the four tiers (see above) were used. The EBT-dependent risk indication for chemical analytical risks and the percentage of additional samples are summarised in Tables III.6 and III.7.

Table III.6: Different Effect-Based Trigger value (EBT) approaches applied to results from measuring surface water (SW) samples with different EBM regarding risk indication and screening value (adapted from Kase et al. 2018).

EBM and condition	Tier 2: EBT Escher et al 2018 [ng/L]	Tier 2: Positive chemical analytical risk indication in SW	Tier 2: Additional positive samples without chemical analytical risk in SW	Tier 3: sensitivity factor approach [ng/L]	Tier 3 Positive chemical analytical risk indication in SW	Tier 3: Additional positive samples without chemical analytical risk in SW	Tier 4 : EBT Brion et al. 2018 [ng/L]	Tier 4: Positive chemical analytical risk indication in SW	Tier 4: Additional positive samples without chemical analytical risk in SW
ER GeneBlazer	0.340	7/7=100%	0/7=0%	0.400	5/7=71%	0/7=0%	0.242	7/7=100%	0/7=0%
Hela 9903	1.01	1/7=14%	0/7=0%	0.266	5/7=71%	0/7=0%	0.182	7/7=100%	0/7=0%
pYES	Na	na	na	0.266	6/7=86%	4/7=57%	0.500	6/7=86%	0/7=0%
MELN	0.370	6/7=86%	4/7=57%	0.266	6/7=86%	5/7=71%	0.557	6/7=86%	1/7=14%
ER Calux	0.100	7/7=100%	5/7=71%	0.400	6/7=86%	0/7=0%	0.283	6/7=86%	0/7=0%
Mean		75%	32%		80%	26%		91%	3%

Table III.7. Summary risk indication and screening properties of different Effect-Based Trigger value (EBT) approaches.

EBT option	Percentage of positive chemical risk indication of steroidal estrogens mixture risk for 16 surface water samples (cumulative RQ>1) in estrogen monitoring project	Percentage screening for other xenoestrogens: additional samples to analyse without known mixture risk of steroidal estrogens
Tier 1: Generic EBT = 400 pg/L*	77%	11%
Tier 2: EBT according to Escher et al. 2018**	75%	32%
Tier 3: sensitivity factor approach**	80%	26%
Tier 4: EASZY EBT approach**	91%	3%

*tested and published in Kase et al. 2018, ** calculated in annex 2

EBM data are validated against the risk identification based on hr-LC/MS chemical analysis (risk identification and the low additional screening percentage for other xenoestrogens). The EBTs derived from tier 4 resulted in the highest percentage of positive risk assessments and the lowest percentage of false positive risk assessments. The average percentage of positive surface water assessments decreased with decreasing tier used for the EBT derivation and the average percentage of false positive assessments increased with decreasing tier. In terms of safety, the Tier 4 EBT are most appropriate, followed by tier 3 EBT. This result supports the tiered uncertainty approach in Fig. 1.

The situation at the moment is that the assessment using chemical monitoring data is accepted and implemented. The respective data are a kind of an anchor for 'alternative' methods – such as EBM. It is likely that an assessment based on an EBM will be compared to an assessment based on a chemical measurement. With respect to the possible application of EBMs for screening (comparable to the use of EBMs for "dioxins in food") it is especially necessary to "validate" the EBM-readout against the assessment based on chemical analysis (as assumed to be true). There would be no added value for the EBM (with respect to screening) if there is a high number of false negative and/or false positive assessments ("true" or "false" defined based on the outcome of the accepted chemical assessment and not necessarily "true" or "false" as an "absolute" assessment).

An application of the read-across Tier 2 EBT shows that the very sensitive EBMs have low EBTs and the EBMs with low sensitivity have high EBTs. This leads – e.g. in the case of the ER-CALUX – to a situation in which a high percentage of samples would pose an inacceptable risk although no risk is indicated by chemical analysis. This might reduce the acceptabilty of EBMs. This is partially compensated for by the Tier 3 proposed in the document that takes into account low variability, sensitivity and relative potency and the proof of concept that the EBM have shown population-relevant effects with high specificity and sensitivity.

Further data showing preliminary results from the effect-based WL project are presented.

Preliminary results from the effect-based watch list project

In the course of the effect-based WL project (presented at the last EBM activity meeting in Rome) further data was generated that support the findings described above. The following figure shows the BEQ for around 40 representative WL water bodies investigated by the ERα-CALUX assay following ISO 19040-3.

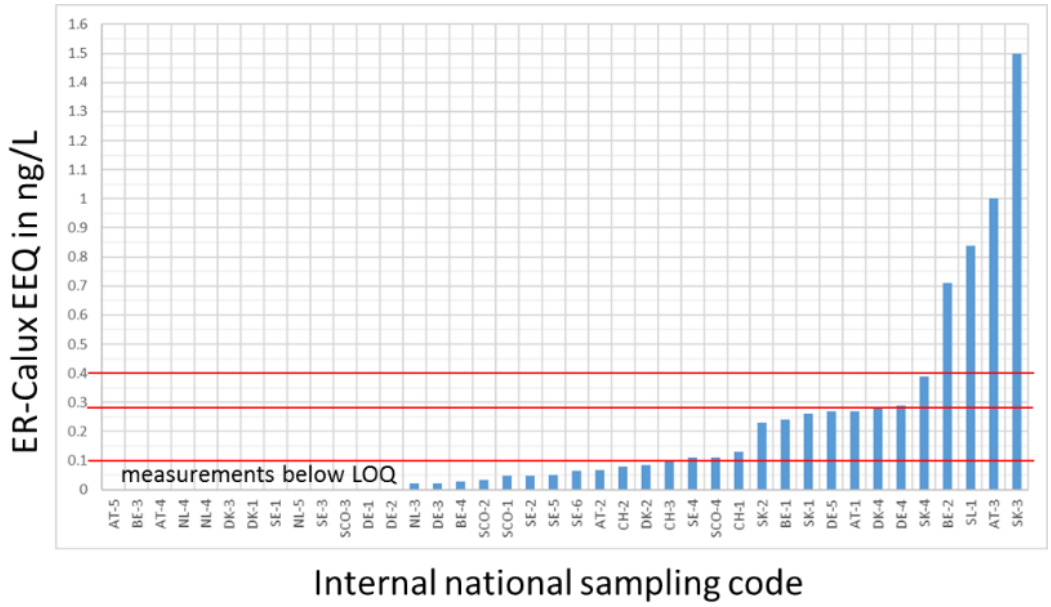

Figure AIII.6: Preliminary effect-based measurement data of around 40 EU WL samples, measured with ERα-CALUX (LOQ were between 15 to 48 pg/L EEQ); investigated EBTs are included as red lines.

Table III.8: Effect-based trigger value (EBT) exceedances for 40 WL samples assessed using the EBTs derived from the tiered approach.

	Tier 1	Tier 2	Tier 3	Tier 4
ERα-CALUX EBT	400 pg/L	100 pg/L	400 pg/L	283 pg/L
Percentage of positive watch list samples	10%	38%	10%	15%

Similar to the results of the estrogen monitoring project it can be also shown on representative WL samples that application of tier 2 EBTs according to Escher et al. 2018 leads to the highest percentage of positive samples with 38%, which would mean additional chemical analysis.

Application of tier 3 or tier 4 EBTs would lead to 10-15 % of positive samples. In conclusion, the chemical analytical monitoring burden could be lowered remarkably by using higher-tier EBTs on representative samples. Higher-tier EBTs also have a good screening value in combination with standardised methods, e.g. ERα-CALUX method. Therefore, Tier 3 and 4 EBT are recommended additionally for screening.

Moreover, in a compliance check of five *in vitro* methods with three chemical analytical methods, it was shown on 33 water samples that an effect-based status assessment would be very useful because all samples can be classified as compliant or non-compliant with high specificity and sensitivity (Kase et al. 2018).

Both findings indicate a good screening and status assessment potential for different EBMs for the MoA of ER-receptor-mediated estrogenicity using higher-tier EBTs.

Description of sensitivity factor EBT approach (Tier 3)

EBTs are needed to assess whether a sample poses an acceptable or an in-acceptable risk to the aquatic environment. EBT can be derived for certain endpoints or be test-specific. Without EBT, any applied inclusion of EBMs (e.g. for screening, prioritisation or status assessments) will be difficult to achieve in frame of the review of the EU Water Framework Directive (WFD).

Test-specific EBT have the advantage that the specificity and sensitivity can be increased compared to endpoint-specific EBT, which have to cover a broad set of different EBM. On the other hand, it is not feasible or meaningful in a regulatory context to provide a separate EBT for each EBM as they can never keep up with the fast development of EBM and are very difficult to implement due to a large variability in available methods and potential new method developments.

The most preferred solution out of three options to derive EBTs was discussed at an EBT workshop in Dübendorf (CH) in 2017 and further discussed at the last EBM- activity meeting in Rome (IT).

This section describes the combination of an endpoint-specific EBT with a test-specific classification, based on its respective sensitivity, variability and specificity, for the MoA 'ER receptor activation'. This approach intends to combine the advantage of an easy-to-implement EBT derivation with a test-specific adjustment regarding specificity and sensitivity.

The EBM-dependent parameters are: the LOQ for the reference substance E2, the inter-test coefficient of variation (CV%) and the relative effect potencies (REP) for a less potent reference substance such as E1 and for a more potent reference substance such as EE2.

In a first step, data for eight *in vitro* EBMs (see Table A1) were compiled, five of which were already characterised in the Estrogen Monitoring project (Kase et al. 2018). Three of the selected EBMs (A-YES, L-YES and ER-Calux) are standardised according to ISO (ISO 19040 parts 1 to 3).

Methods

Table III.9: Effect-based method (EBM)-specific characteristics for eight estrogen receptor (ER) activation assays*

E1 = estrone, E2 = 17β-estradiol, EE2 = 17α-ethinylestradiol, CV = coefficient of variation, REP = relative effect potency

No	Name of EBM	EBM characterised in Estrogen Monitoring project (EM), OECD or ISO	LOQ for E2 [pg/L] at 1000-fold enrichment**	Inter-test CV [%] (intra-lab or inter-lab)	REP E1	REP EE2
1	ER GeneBlazer	EM	32	29	0.08	1.67
2	Hela 9903	EM, OECD	41	20	0.02	1.18
3	pYES	EM	7.5	33	0.11	1.00
4	MELN	EM	17	64	0.29	0.79
5	ERα-CALUX	EM, OECD, ISO	8	9	0.01	1.30
6	L-YES (Mc Donnell)	ISO	318	28	0.11	1
7	A-YES	ISO	13	14	0.22	1.2
8	VM7Luc4E2	OECD	25.5	17	0.033	1.15
		Mean:	**58**	**27**	**0.11**	**1.16**
		STDEV:	**106**	**17**	**0.10**	**0.26**

*corresponding data are in Kunz et al. 2017, Kase et al. 2018, ICCVAM 2011, OECD 2009 or were provided by ISO contact points who are co-authors of this proposal. VM7Luc4E2 data were kindly provided by Timo Hamers from University of Amsterdam, NL.

**LOQs were calculated from sample concentrations. 3 x STDEV from the negative control with n=3 was the minimum LOQ requirement. The final LOQ was then divided by the relative enrichment factor (REF). REF = (solid-phase extraction (SPE) concentration factor (1000) / test specific dilution factor (x))

Remark: The VM7Luc4E2 does not normally work with 1000-fold enriched samples (as indicated in the table) and uses a maximal 250-fold enrichment, normally lower depending on the activity of the samples. Moreover, VM7Luc4E2 has an additional enrichment step of 50 before 200-fold dilution. For other methods the REF might also be adapted depending upon the activity of the samples.

Starting from a generic screening EBT of 0.4 ng/L EEQ (see Kase et al. 2018), a maximum sensitivity factor of four can be estimated to address test-specific differences. This factor was because the maximal ratio between the lowest and highest EBT for the MoA 'ER-activation' published in Escher et al. 2018 and the generic screening EBT of 0.4 ng/L (the BEQ) was 4. The following classification scheme of sensitivity factors (see Table III.10) was presented in June 2017 at the EBT workshop in Dübendorf (CH). This approach aims at simplifying regulatory use, and can be adapted with test-specific EBTs according to Escher et al. 2018 and with a test-specific sensitivity classification (see Table III.11).

Table III.10: Proposal for a classification scheme of sensitivity factors for estrogen receptor (ER) activation.

LOQ = limit of quantification, E1 = estrone, E2 = 17β-estradiol, EE2 = 17α-ethinylestradiol, CV = coefficient of variation, REP = relative effect potency

Sensitivity factor classification*	LOQ for E2 [pg/L] at 1000-fold enrichment*	Inter-test CV [%] for E2	Relative potencies compared to mean REP for E1 and EE2 across investigated bioassays	Rounded mean sensitivity factor classification defines the mean sensitivity factor
Very high (I)	LOQ < 1/10 of EBT	CV < 10	REP > mean + 2x STDEV	**0.5**
High (II)	1/10 of EBT < LOQ < 1/3 of EBT	10 < CV < 20	mean + 2x STDEV > REP > mean + 1x STDEV	**1**
Moderate (III)	1/3 EBT < LOQ < 1xEBT	20 < CV < 35	REP = mean ± 1x STDEV	**1.5**
Low (IV)	1 EBT < LOQ < 1.5xEBT	35 < CV < 50	mean -1 x STDEV > REP > mean - 2x STDEV	**2.5**
Very low (V)	1.5 EBT < LOQ < 2.5xEBT	50 < CV < 65	mean - 2x STDEV > REP > mean - 3x STDEV	**4**
out of range	LOQ > 2.5xEBT	CV > 65	REP < mean - 3x STDEV	**not possible**

*If the rounded mean classification is exactly between 2 classes, e.g. between high (II) and moderate (III) it will be rounded to the lower mean (in this case moderate) sensitivity classification in order to increase the protectiveness. If only one parameter for one EBM is not available or out of range no sensitivity classification can be performed. This approach intends to stimulate minimum data availability and data quality for each EBM before using them for screening purposes and in combination with EBTs. The sensitivity factor needs to be adapted for each relevant MoA according to available test specific EBTs calculated according to Escher et al. 2018

Results

The sensitivity categorisation scheme was applied for all eight EBMs to calculate a sensitivity factor. The results are shown in table A3. Four EBMs (ERα-CALUX, A-YES, VM7Luc4E2 and ER-GeneBlazer) were ranked in the category 'high sensitivity', resulting in a sensitivity factor of 1. The other four EBMs (Hela 9903, MELN, p-YES, L-YES Mc Donnell) were ranked in the category 'moderate sensitivity' resulting in a sensitivity factor of 1.5. The sceening EBT of 0.4 ng/L EEQ was modified by the test-specific sensitivity factor to allow a comparison of different EBT approaches as shown in Table III.11. Five of these EBMs (No.s 1-5) were applied in the EU estrogen monitoring project and showed a good risk indication of steroidal estrogens compared to analytical results obtained by hr-LC/MS.

Table III.11: Sensitivity factor categorisation for the eight selected EBM for the MoA 'estrogen receptor (ER) activation'.

LOQ = limit of quantification, E1 = estrone, E2 = 17β-estradiol, EE2 = 17α-ethinylestradiol, CV = coefficient of variation, REP = relative effect potency, sensitivity classification: 1 (very high), 2 (high), 3 (moderate), 4 (low), 5 (very low)

No	Name of EBM	LOQ for E2 [pg/L] at 1000 fold upconcentration*	Sensitivity classification based on LOQ	Intertest CV [%]	Sensitivity classification based on CV	REP E1	Sensitivity classification based on E1 REP	REP EE2	Sensitivity classification based on EE2 REP	Mean sensitivity classification	Resulting sensitivity factor
1	ER GeneBlazer	32	1	29	3	0.08	3	1.67	2	2.25	1.0
2	Hela 9903	41	2	20	3	0.02	3	1.18	3	2.75	1.5
3	pYES	7.5	1	33	3	0.11	3	1.00	3	2.50	1.5
4	MELN	17	1	64	5	0.29	3	0.79	4	3.25	1.5
5	ERα-CALUX	8	1	9	1	0.01	4	1.30	3	2.25	1.0
6	YES (Mc Donnell)	318	3	28	3	0.11	3	1.00	3	3.00	1.5
7	A-YES	13	1	14	2	0.22	3	1.20	3	2.25	1.0
8	VM7Luc4E2	25.5	1	17	2	0.033	3	1.15	3	2.25	1.0
	Mean:	**58**		**27**		**0.11**		**1.16**			
	STDEV:	**106**		**17**		**0.10**		**0.26**			

Effect-Based Trigger value (EBT) compilation using the tiered EBT approach

Table III.12: EBT compilation using the tiered EBT approach; EBTs in bold are proposed for use based on current knowledge.

* UBA/JRC Dossier BPA 2016; ** Ecotox Centre Dossier Chlorpyrifos 2016; *** Ecotox Centre Dossier Diuron 2017; na: not available; dossiers available upon request. EBTs are not rounded and are shown as calculated. Data for Tier 2 are based on Escher et al. 2018.

No	Measured endpoint or molecular target	Effect-Based Method /Assay name	Role in Adverse Outcome Pathway AOP	Reference compound	Tier 1 EBT [ng/L]	Tier 2 EBT [ng/L]	Tier 3 EBT [ng/L]	Tier 4 EBT [ng/L]	Comment
1	Activation of aryl hydrocarbon receptor (AhR)	H4L1.1c4 AhR assay	Toxicokinetics	Benzo[a]pyrene	50.000	**6.358**			
2	Activation of aryl hydrocarbon receptor (AhR)	PAH-CALUX	Toxicokinetics	Benzo[a]pyrene	50.000	**6.205**			
3	Activation of peroxisome proliferator-activated receptor (PPARγ)	PPARg-GeneBLAzer	Toxicokinetics	Rosiglitazone	na	**36.000**			
4	Activation of peroxisome proliferator-activated receptor (PPARγ)	PPARy-CALUX	Toxicokinetics	Rosiglitazone	na	data too preliminary to derive final effect threshold			
5	Activation of pregnane x receptor (PXR)	HG5LN-hPXR	Toxicokinetics	Di(2-ethylhexyl)-phthalate	1300.000	**16273.280**			
6	Activation of pregnane x receptor (PXR)	PXR-CALUX	Toxicokinetics	Di(2-ethylhexyl)-phthalate	1300.000	**272494.999**			
7	Activation of estrogen receptor (ER)	MELN	Hormone receptor regulation	17β-Estradiol	0.400	0.368	0.266	**0.557**	

8	Activation of estrogen receptor (ER)	ER-GeneBLAzer	Hormone receptor regulation	17β-Estradiol	0.400	0.337	0.400	**0.242**	
9	Activation of estrogen receptor (ER)	ERa_Luc_BG1	Hormone receptor regulation	17β-Estradiol	0.400	**0.625**			
10	Activation of estrogen receptor (ER)	SSTA ERα-HeLa-9903	Hormone receptor regulation	17β-Estradiol	0.400	1.008	0.266	**0.182**	
11	Activation of estrogen receptor (ER)	ER-CALUX	Hormone receptor regulation	17β-Estradiol	0.400	0.104	0.400	**0.283**	
12	Activation of estrogen receptor (ER)	A-YES	Hormone receptor regulation	17β-Estradiol	0.400	0.558	**0.400**		
13	Activation of estrogen receptor (ER)	3d YES	Hormone receptor regulation	17β-Estradiol	0.400	**0.882**			
14	Activation of estrogen receptor (ER)	ISO-LYES (Sumpter)	Hormone receptor regulation	17β-Estradiol	0.400	**0.968**			
15	Activation of estrogen receptor (ER)	VM7Luc4E2	Hormone receptor regulation	17β-Estradiol	0.400	na	**0.400**		
16	Activation of estrogen receptor (ER)	p-YES	Hormone receptor regulation	17β-Estradiol	0.400	na	0.266	**0.500**	
17	Activation of estrogen receptor (ER)	ISO-LYES ((McDonnell))	Hormone receptor regulation	17β-Estradiol	0.400	1.068	**0.266**		
18	Estrogenic signalling	REACTIV (unspiked)	Hormone receptor regulation	17β-Estradiol	0.400	**0.797**			
19	Antagonistic activity on the estrogen receptor (ER)	anti ER-GeneBLAzer	Hormone receptor regulation	Tamoxifen	na	currently not applicable because regulated chemicals are of low potency -> no read-across possible			

20	Antagonistic activity on the estrogen receptor (ER)	anti ERa_Luc_BG1	Hormone receptor regulation	Tamoxifen	na	currently not applicable because regulated chemicals are of low potency -> no read-across possible			
21	Antagonistic activity on the estrogen receptor (ER)	anti A-YES	Hormone receptor regulation	Tamoxifen	na	currently not applicable because regulated chemicals are of low potency -> no read-across possible			
22	Activation of androgen receptor (AR)	AR-GeneBLAzer	Hormone receptor regulation	Methyltrienolone (R1881)	na	currently not applicable because all regulated chemicals are of low potency (REP 1.10-3 to 1.2.10-5 compared to the hormone agonist R1881)-> no read-across possible			

23	Activation of androgen receptor (AR)	MDA-kb2	Hormone receptor regulation	5α-Dihydrotestosterone (DHT)	na	currently not applicable because all regulated chemicals are of low potency (REP 1.10-3 to 1.2.10-5 compared to the hormone agonist DHT)-> no read-across possible			
24	Activation of androgen receptor (AR)	A-YAS	Hormone receptor regulation	5α-Dihydrotestosterone (DHT)	na	currently not applicable because only two chemicals were active, which are also estrogenic at lower concentration			
25	Androgenic activity	RADAR (unspiked)	Hormone receptor regulation	17α-methyl testosterone (17MT)	na	currently not applicable because none of the tested chemicals were active			

26	Antagonistic activity on the androgen receptor (AR)	anti AR-GeneBLAzer	Hormone receptor regulation	Flutamide	na	**3284.262**			
27	Antagonistic activity on the androgen receptor (AR)	anti MDA-kb2	Hormone receptor regulation	Flutamide	na	**3458.463**			
28	Antagonistic activity on the androgen receptor (AR)	anti AR-CALUX	Hormone receptor regulation	Flutamide	na	**14431.888**			
29	Anti-androgenic activity	anti AR RADAR (spiked)	Hormone receptor regulation	Flutamide	na	**3631.287**			
30	Antagonistic activity on the progestogenic receptor (PR)	anti PR-CALUX	Hormone receptor regulation	Endosulfan	5.000	1967.111			
31	Activation of glucocorticoid receptor (GR)	GR-GeneBLAzer	Hormone receptor regulation	Dexamethasone	na	currently not applicable because all regulated chemicals are of low potency (REP 2.10-4 to 4.10-6 compared to the potent agonist dexmethasone) -> no read-across possible			

32	Antagonistic activity of glucocorticoid receptor (GR)	anti GR-GeneBLAzer	Hormone receptor regulation	Mifepristone	na	currently not applicable because all regulated chemicals are of low potency (REP 3.10-4 to 7.10-6 compared to the potent antagonist Mifepristone) -> no read-across possible			
33	Competition with T4 for binding to transthyretin (TTR)	TTR RLBA	Hormone receptor regulation	Thyroxine (T4)	na	**58.432**			
34	Competition with T4 for binding to transthyretin (TTR)	TTR FITC-T4	Hormone receptor regulation	Thyroxine	na	**49.153**			
35	Modulation of thyroid hormone signaling	XETA (unspiked)	Hormone receptor regulation	Triiodothyronine (T3)	na	**0.621**			
36	Antagonistic activity on the thyroid receptor (TR)	Anti-TR-LUC-GH3	Hormone receptor regulation	Bisphenol A	240.000*	**603.416**			
37	Induction of oxidative stress response	AREc32	Adaptive Stress responses	Dichlorvos	0.600	**155834.865**			
38	Induction of oxidative stress response	AREGeneBLAzer	Adaptive Stress responses	Dichlorvos	0.600	**392090.410**			
39	Induction of oxidative stress response	Nrf2-CALUX	Adaptive Stress responses	Dichlorvos	0.600	**25579.901**			

40	Growth inhibition	72h Algal growth inhibition	Population and organism response	Diuron	70.000***	**116.460**			
41	Growth inhibition	24h Synchronous algae reproduction	Population and organism response	Diuron	70.000***	**109.362**			
42	Growth inhibition	24h Combined algae assay (growth)	Population and organism response	Diuron	70.000***	**129.676**			
43	Photosynthesis inhibition	2h Combined algae assay (PSII)	Population and organism response	Diuron	70.000***	**73.740**			
44	Immobilization	48h Daphnia immobilization test	Population and organism response	Chlorpyrifos	0.460**	**14.993**			
45	Mortality after 48h	Fish embryo toxicity	Population and organism response	Bisphenol A	240.000*	**275568.416**			
46	Mortality after 96/120h	Fish embryo toxicity	Population and organism response	Bisphenol A	240.000*	**182805.837**			
47	Steroidgenesis modulation assay	H295 R	Steroidgenesis	Atrazine	**600.000**	na			
48	Steroidgenesis modulation assay	H295 R	Steroidgenesis	Forskolin	na	na			Forskolin is the most potent inducing compound

ANNEX IV. Integrated platform for EBMs and application of Reference Materials

An integrated platform linking EBMs to currently employed chemical and ecological assessment methods has been proposed in the JRC report on the integrated assessment of the current PS list under the WFD and other substances of interest (Niegowska et al. 2018; Figure IV.1). EBMs selected based on endpoints most widely targeted by chemicals present in water bodies (e.g. oxidative stress, photosynthesis inhibition, endocrine disruption, carcinogenicity) would provide effect concentrations that, on the basis of comparison with reference materials, could be related to EQS for a range of model organisms (Carvalho et al. 2014). Few endpoints would be sufficient to cover several mechanisms of toxic action and thus allow mixture effects in biota to be prevented in a timely manner in line with the precautionary principle.

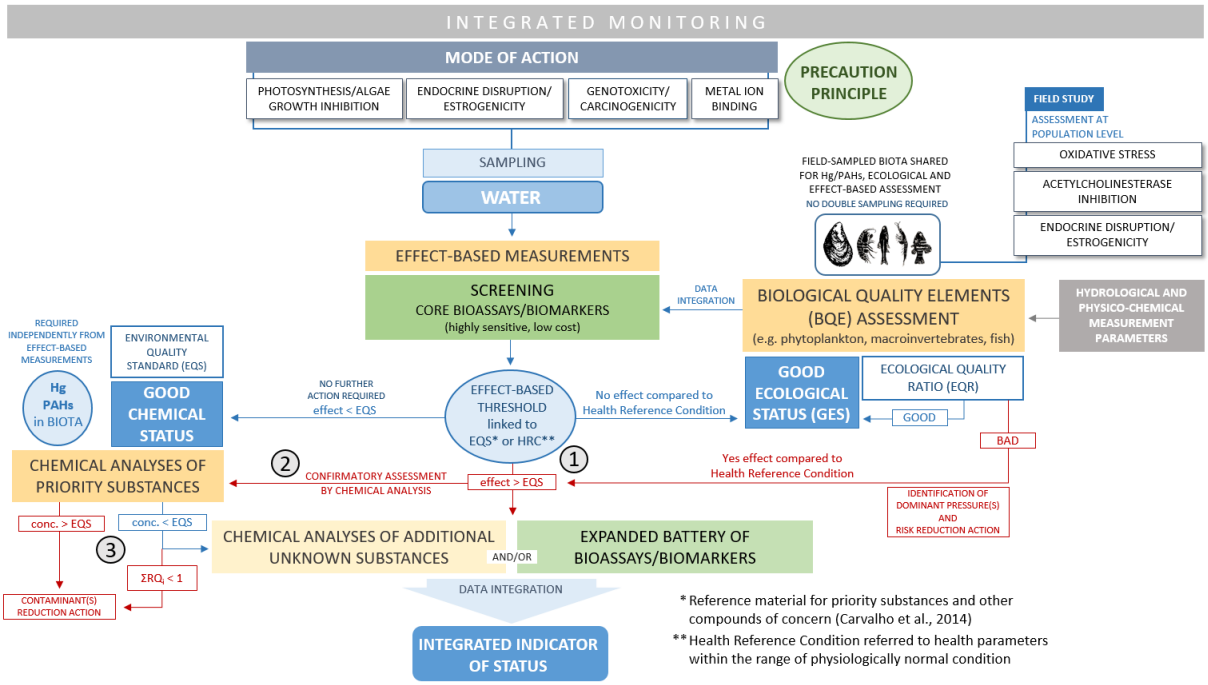

Figure IV.1: Framework for surveillance/operational monitoring linking EBMs with chemical and ecological methods (Niegowska et al. 2018).

The purpose of this platform would be to reduce the chemical assessment taking into account cost effectiveness without significantly impacting the entire workflow in terms of biological sampling which could be performed once for effect-based measurements, ecological assessment and detection of mercury (Hg) and polycyclic aromatic hydrocarbons (PAHs). Instrumental analysis would be executed for PS only if EBM results indicated effect concentrations above the safety threshold, in order to confirm the presence of specific compounds and take action to reduce contamination or perform additional analysis of unknown chemicals when measured substances do not exceed their EQS.

At population level, biomarkers relevant to the most ecologically relevant endpoints (e.g. oxidative stress, acetylcholinesterase (AChE) inhibition, endocrine disruption) could be

employed to inform about effects in biota compared to health reference conditions (HRC) corresponding to physiologically optimal parameters already defined to a large extent (e.g. normal AChE activity in flounder). Altogether, EBMs, ecological and chemical methods applied in a complementary manner according to the proposed platform would generate an integrated indicator of status as a holistic assessment of water quality and health conditions of biota exposed to realistically occurring chemical mixtures.

Example of a possible EU-wide exercise with Reference Materials

An approach evaluated recently in an EU-wide exercise proposed the use of a known chemical mixture (with EQS available for each component) as a reference material (RM) for EBMs (Carvalho et al. 2014). The RM compounds were selected based on their chemical structure and MoA to represent main pollutant groups found in surface waters which enabled the expression of results with reference to EQS even for unknown substances. Calibration curves generated from the RM for a range of EBMs were used to extrapolate the obtained effect concentration (EC) as EQS multipliers (xEQS) in a straightforward manner without the need to derive correction factors (Figure IV.2).

The availability of a standardised RM is crucial to assess the performance of an EBM in a laboratory and should be used for quality control alongside routine measurements, and especially if a laboratory starts using an established EBM. If further EBM are developed in future, their performance could be benchmarked against this reference material.

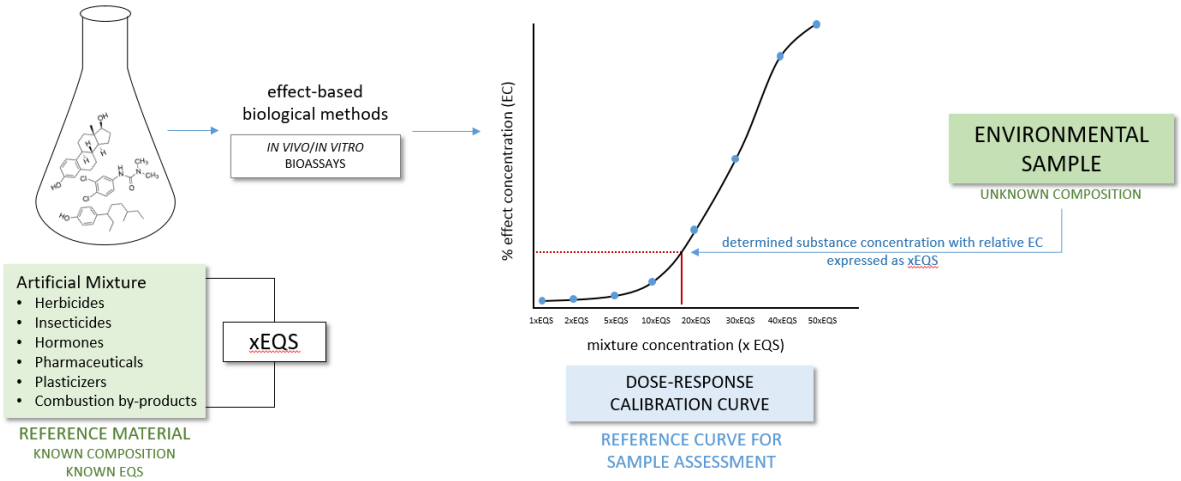

Figure IV.2: Workflow representing the approach based on the reference mixture proposed by Carvalho et al. 2014. The calibration curve generated from the RM at different concentrations is used to extrapolate the concentration of substances present in sampled water at which effects are induced.

The selection of chemicals for RM should be further investigated in order to identify differences in terms of assay performance based on the compounds included and to establish the most appropriate reference mixture composition for the effects to be assessed. Specific RM mixtures could be created for environmental sites where contamination by particular substances is expected or suspected, thus providing RM for surface water pollution profiling.

ANNEX V. Example of a Battery of EBMs

The purpose of this annex is to (i) provide an overview of the recent bioanalytical test batteries typically used for water monitoring and assessment, and (ii) provide a recommendation for the use of a "standardised" bioassay battery for the evaluation of water quality.

Within the NORMAN Working Group (WG) 2 on Bioassays and Biomarkers in Water Quality Monitoring, in partnership with the SOLUTIONS project, a comprehensive review on the integration of bioassays and biomarkers in water quality monitoring and the selection of bioassays for a coherent battery of EBMs was conducted [1-3]. The bioassay batteries of different projects have been reviewed and compared in order to identify and to suggest a common battery of bioassays.

A recent NORMAN network interlaboratory study (ILS) verified whether a battery of miniaturised bioassays, conducted in 11 different laboratories following their own protocols, would produce comparable results when applied to evaluate blinded samples consisting of pristine water extracts spiked with four emerging pollutants as single chemicals or mixtures [3]. Assays evaluated effects on aquatic organisms from three different trophic levels (algae, daphnids, zebrafish embryos) and mechanism-specific effects using *in vitro* estrogenicity (ER-Luc, YES) and mutagenicity (Ames, Ames Fluctuation) assays [3]. Within the SOLUTIONS project, Busch and co-workers [2] systematically compiled organic contaminants detected in freshwater monitoring studies, provided an overview of the current knowledge available about the MoA for the detected compounds, performed a hazard ranking to identify priority mixtures, and reflected on the challenges in selecting appropriate bioassays for effect-based monitoring. Furthermore, they suggested a list of organic compounds that could serve as a reference list for EBM validation studies [2].

In the SOLUTIONS project a broad battery of *in vitro* bioassays based on human and fish cell lines as well as whole-organism assays using bacteria, algae, daphnids and fish embryos were assembled for use in water quality monitoring [4]. The selection of bioassays was guided by the principles of AOPs in order to cover relevant steps in toxicity pathways known to be triggered by environmental water samples. In a proof-of-concept study the effects of 34 water pollutants, which were selected based on hazard quotients, available EQS and MoA information, were fingerprinted in the bioassay test battery. The proof-of-concept study not only demonstrated the utility of fingerprinting single chemicals for an improved understanding of the biological effect of pollutants, but also highlighted the need to apply bioassays for water quality monitoring in order to prevent underestimation of the overall biological effect [4].

Based on the discussions within NORMAN and SOLUTIONS, a common battery of bioassays has been suggested that covers major toxicological endpoints. The recommended bioassay battery is also detailed in a Policy brief from the SOLUTIONS project (Figure V.1). It is suggested to complement *in vitro* assays by apical bioassays representing at least fish (fish embryo testing), invertebrates (Daphnia) and algae (cell multiplication inhibition) considered also as BQEs for pelagic communities under the WFD. Of the MoA-specific *in vitro* assays, priority should be given to endocrine disruption and mutagenicity. Dioxin-like effects should be analysed particularly in sediments, biota and equilibrium passive samplers since typical drivers of these effects are very hydrophobic and accumulate in these matrices.

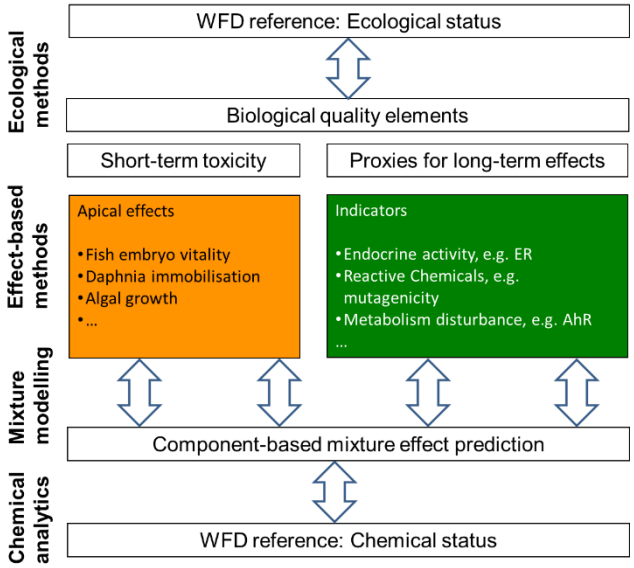

Figure V.1: Recommended test battery in the context of chemical and ecological status monitoring (redrawn from the SOLUTIONS Policy Brief Effect-based monitoring, Brack et al. 2018).

References:

1. Neale, P.A., et al., *Integrating chemical analysis and bioanalysis to evaluate the contribution of wastewater effluent on the micropollutant burden in small streams.* Science of the Total Environment, 2017. **576**: p. 785-795.
2. Busch, W., et al., *Micropollutants in European rivers: A mode of action survey to support the development of effect-based tools for water monitoring.* Environ. Toxicol. Chem., 2016. **DOI:10.1002/etc.3460**.
3. Di Paolo, C., et al., *Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts - Towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring.* Water Res, 2016. **104**: p. 473-484.
4. Neale, P.A., et al., *Development of a bioanalytical test battery for water quality monitoring: Fingerprinting identified micropollutants and their Contribution to effects in surface water.* Water Research, 2017. **123**: p. 734-750.
5. Brack, W., et al., *Towards a holistic and solution-oriented monitoring of chemical status of European water bodies: how to support the EU strategy for a non-toxic environment?* Environmental Sciences Europe, 2018. **30**(1): p. 33.
6. Brack, W., et al., *Towards the review of the European Union Water Framework Directive: Recommendations for more efficient assessment and management of chemical contamination in European surface water resources.* Sci Total Environ, 2017. **576**: p. 720-737.

ANNEX VI Neurotoxicity Outlook

Neurotoxicity was identified within the EU project SOLUTIONS as one of the most important emerging MoAs in the environment. The numbers of potential neurotoxicants in the environment is rising and can pose a risk for humans and the environment. Considering the increasing numbers of environmental contaminants with potential neurotoxic potential, eco-neurotoxicity should be also considered in future risk assessments. In order to do this, novel test systems are needed that can cope with species differences within ecosystems. The selection of *in vitro* assays could be guided by AOPs relevant for eco-neurotoxicity. The German Federal Ministry of Education and Research (BMBF) founded the project NeuroBox and the EU NORMAN network is performing a ringtest with neurotoxic substances considering behavioural changes in *Danio rerio*. Moreover, EURL ECVAM of the Joint Research Centre (JRC) is working on *in vitro* approaches to detect developmental neurotoxicity (DNT) triggered by a single chemical or mixtures of chemicals.

For example, the JRC has developed human stem cell-based *in vitro* assays for evaluation of neurite outgrowth, synaptogenesis and neuronal electrical activity. This battery of assays in also included in an EFSA/OECD DNT project which aims to develop a guidance document on the use of DNT *in vitro* methods. The perturbation of these key neurodevelopmental processes (e.g. synaptogenesis, neuronal network formation and function) were identified as key events in several AOPs.

An evaluation of neurotoxicity (including developmental stage) is also being performed using non-mammalian species since the mechanisms underlying the development and function of the nervous system are well conserved across the phylogenic tree. Many of the basic molecular processes are identical in mammals and in non-mammalian species. Therefore, several alternative species including *Danio rerio*, *Oryzias latipes* or *Xenopus laevis* are used as vertebrate non-mammalian models and complementary to *in vitro* approaches. The small size, transparency during embryogenesis and speed of development make these species suitable for chemical testing. The gathering of data from these multiple information sources, could be used to develop Integrated Approaches to Testing and Assessment (IATA) designed in a fit-for-purpose manner for different regulatory purposes, including aquatic and human health protection. In the light of these developments, a relevant selection of neurotoxicity assays for environmental assessments can be discussed in more detail in the medium term to advance the reliability and scope of assessments for neurotoxicity.

Acknowledgement: Dr. Anna Bal-Price from EURL ECVAM at JRC for her support and knowledge sharing regarding ongoing neurotoxicity activities.