Contribution à la validation de la mesure de l'activité oestrogénique *in vitro* dans les eaux de surface et les eaux usées

S. Aït-Aïssa

Décembre 2018

Rapport final

Avec le soutien de

Contexte de programmation et de réalisation

Ce rapport a été réalisé dans le cadre du programme scientifique et technique AQUAREF pour l'année 2017, au titre de l'action « Approche bioanalytique : validation de la mesure de l'activité estrogénique dans les matrices environnementales », thème « Nouveaux Outils de surveillance».

Auteur :

Selim Aït-Aïssa
INERIS
selim.ait-aissa@ineris.fr

Vérification du document :

Eric Thybaud
INERIS
Eric.thybaud@ineris.fr

Olivier Geffard
IRSTEA
olivier.geffard@irstea.fr

Les correspondants

AFB : Pierre-François STAUB, pierre-francois.staub@onema.fr
Olivier PERCEVAL, olivier.perceval@onema.fr

INERIS : Selim AIT-AISSA

Référence du document : Selim Aït-Aïssa – Contribution à la validation de la mesure de l'activité oestrogénique *in vitro* dans les eaux de surface et les eaux usées – Rapport AQUAREF 2018 – DRC-18-159227-11202B - 16 p. + annexes.

Droits d'usage :	*Accès libre*
Couverture géographique :	*International*
Niveau géographique :	*National*
Niveau de lecture :	*Professionnels, experts*
Nature de la ressource :	*Document*

SOMMAIRE

Liste des annexes :

Annexe A : Kunz et al (2017) Effect-based tools for monitoring estrogenic mixtures: Evaluation of five in vitro bioassays. *Water Research*.

Annexe B : Könneman et al (2018) Effect-based and chemical analytical methods to monitor estrogens under the European Water Framework Directive. *Trends in Analytical Chemistry*.

Annexe C : Kase et al. (2018) Screening and risk management solutions for steroidal estrogens in surface and wastewater. *Trends in Analytical chemistry*.

Annexe D : Carvalho et al. (2014) Mixtures of Chemical Pollutants at European Legislation Safety Concentrations: How Safe are They? *Toxicological Sciences*.

Annexe E : DiPaolo et al. (2016) Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts - Towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring. *Water Research*.

CONTRIBUTION A LA VALIDATION DE LA MESURE DE L'ACTIVITE OESTROGENIQUE IN VITRO DANS LES EAUX DE SURFACE ET LES EAUX USEES

S. Aït-Aïssa

RESUME

L'utilisation de bioessais *in vitro* permettant la quantification d'estradiol-équivalents (EEQ) est une approche pertinente pour surveiller les substances à activité œstrogène-mimétiques dans les eaux. L'objectif de cette étude est d'évaluer la validité de cet outil pour une surveillance en routine et son application future dans un contexte de surveillance réglementaire. Dans une première partie, un état l'art bibliographique non exhaustif répertorie les différentes méthodes et protocoles utilisés et identifie les étapes clés (échantillonnage, extraction, modèles cellulaires, traitement des données) qui peuvent être sources de variabilité inter-essais. La seconde partie rapporte les résultats d'études inter-laboratoires européennes auxquelles l'INERIS a contribué. Ces études visaient à évaluer les performances (intra- et inter-essais) de 5 bioessais différents, en comparaison avec des analyses chimiques des hormones (E1, E2, EE2), pour l'évaluation de mélanges de substances modèles et d'échantillons d'eaux usées et de rejets. Ces travaux démontrent les bonnes performances (bonne répétabilité, reproductibilité, sensibilité et spécificité) des outils évalués pour répondre aux exigences analytiques de surveillance des hormones oestrogéniques dans ces matrices environnementales. Des premières recommandations sur des valeurs de concentrations *in vitro* en EEQ au-delà desquelles un risque est prédit (valeurs seuils) ont été émises pour permettre leur application à l'évaluation et la gestion des masses d'eaux dans un contexte réglementaire.

Mots clés :

Bioessais, estradiol-équivalents, comparaisons inter-laboratoires, surveillance DCE, eaux de surface, rejets.

CONTRIBUTION TO THE VALIDATION OF IN VITRO ESTROGENIC ACTIVITY ASSESSMENT IN WASTE- AND SURFACE WATERS
S. Aït-Aïssa

ABSTRACTS

The use of receptor-based in vitro assays assessing estradiol-equivalents (EEQs) in environmental matrices is a relevant tool to monitor estrogenic substances in water bodies. The aim of this study is to assess its potential for application in routine monitoring in a regulatory context. First, we report a non-exhaustive state of the art on the principal techniques and methods available and identify critical steps in the protocol that could be the source of variability, i.e. sampling, extraction, in vitro biological models and data analysis. Then, inter-laboratory studies were conducted at the European level to assess the method performances of 5 different in vitro bioassays and compare it with the best available analytical methods for hormone quantification in both reconstituted mixtures and real samples of waste- and surface waters. Overall, these studies demonstrated the good repeatability and reproducibility of the bioassays, as well as very good sensitivity to detected low levels of estradiol-equivalents, as required by WFD criteria. First recommendations are provided regarding the threshold values as EEQs to be used in operational monitoring in regulatory surveillance of water bodies.

Key words:

Bioassays, estradiol-equivalents, interlaboratory comparison, WFD surveillance, surface waters, wastewaters

1. CONTEXTE & OBJECTIFS

L'évaluation de la qualité chimique des milieux fait face à des défis scientifiques et techniques importants pour prendre en compte la complexité des contaminations et la toxicité de ces mélanges. Les milieux aquatiques sont contaminés par une multitude de micropolluants, de classes chimiques et d'origines très diverses, qui peuvent subir des transformations biotiques ou abiotiques et générer des métabolites et produits de dégradation parfois plus toxiques que les composés parents. Face à cette complexité chimique, une évaluation pertinente de la qualité chimique implique de disposer de méthodes de détection qui soient suffisamment représentatives et/ou exhaustives, i.e. intégratrices de l'ensemble des contaminants en présence. Dans le contexte de la directive cadre européenne sur l'eau (DCE 2000/60/CE), les analyses chimiques ciblant une liste finie de substances prioritaires sont utilisées pour l'évaluation de la qualité chimique des masses d'eau. Si ces approches permettent détecter et quantifier les polluants prioritaires, elles ne fournissent qu'une vision très partielle de la contamination environnementale (i.e. ciblées uniquement sur les composés actifs connus) et donc du danger (éco)toxique associé à ces matrices puisqu'elles ne prennent pas en compte des effets de mélange.

Dans cette optique, l'utilisation des méthodes de détection basée sur une réponse biologique est envisagée : elles fournissent une information qui intègre l'ensemble des substances actives au sein d'un échantillon et sont donc complémentaires aux analyses chimiques de substances individuelles. L'utilisation de tels outils dans la surveillance de l'état chimique s'inscrit dans les perspectives de la directive fille, avec la "future application d'outils pour la surveillance autre que substance par substance" (récital 18, 2013/39/EU).

Si des bioessais ont été développés et validés depuis de nombreuses années pour caractériser l'écotoxicité globale d'eaux de surface et d'effluents, l'utilisation de bioessais ciblant des modes d'action particuliers est plus récente. Ces bioessais sont de plus en plus appliqués à la caractérisation de la contamination des milieux aquatiques dans le cadre de programmes de recherche (Wernersson et al 2015, Neale et al 2017). Il existe un besoin de validation de ces nouvelles méthodes dans un contexte réglementaire. Le laboratoire AQUAREF, dont l'une des missions est de proposer et valider de nouvelles méthodes pour la surveillance de la qualité chimique des milieux aquatiques en appui à la DCE, s'inscrit dans cette démarche.

Dans le cadre de la présente étude, nous nous intéressons spécifiquement aux bioessais de détection d'une classe de perturbateurs endocriniens (PE), les substances à activités oestrogéniques. Parmi les multiples polluants aquatiques, nombreux sont des PE du fait de leur capacité à altérer le fonctionnement normal du système endocrinien et engendrer des effets sur la reproduction, le développement ou l'homéostasie d'un organisme et/ou de sa descendance. Au niveau moléculaire et cellulaire, la perturbation endocrinienne implique divers mécanismes d'action, parmi lesquels, l'interaction des substances chimiques avec les récepteurs nucléaires joue un rôle central dans la médiation des effets néfastes pour les organismes. Dans l'environnement aquatique, la présence de substances capables d'interagir avec le récepteur des œstrogènes (ER), a ainsi été associée à la féminisation de poissons (Desbrow et al. 1998) avec des répercutions à l'échelle de la population (Kidd et al. 2007). Aussi, en raison des risques encourus par les populations exposées, une attention croissante a été portée sur ces composés durant la dernière décennie. Au niveau réglementaire, cela s'est traduit en 2013 par la volonté d'inclure dans la liste de vigilance (*watch-list*) de la directive fille (2013/39/EU) de la directive cadre sur l'eau (DCE), 3 substances stéroïdiennes à activité oestrogénique, l'œstradiol (E2), l'estrone (E1) et l'éthinyl-œstradiol (EE2).

L'inclusion de l'E2 et l'EE2 dans la liste de vigilance se confronte à des difficultés analytiques importantes du fait de leurs effets à très faibles doses qui ont conduit à la dérivation de normes de qualité environnementale (NQE) très basses (i.e. 0.4 ng/L pour l'E2 et 0.035 ng/L pour l'EE2). Atteindre ces faibles niveaux de détection, notamment pour l'EE2, pose un vrai challenge analytique. Actuellement, les laboratoires en charge des analyses chimiques pour la surveillance en routine de l'état chimique ne disposent pas de méthodes suffisamment sensibles pour analyser ces composés à ces niveaux de concentrations. Si de telles méthodes sont actuellement disponibles au sein de certains laboratoires académiques ou privés, leur transfert vers les laboratoires de routine implique un investissement financier conséquent. Par conséquent, pour répondre aux besoins d'analyse de l'E2 et l'EE2, le recours à d'autres méthodes, comme les bioessais *in vitro* basés sur le mode d'action de ces substances (i.e. activation du récepteur des œstrogènes ou ER), trouve aujourd'hui un grand intérêt dans ce cadre réglementaire. Dans le cas spécifique des composés œstrogéno-mimétiques, différents bioessais *in vitro* sont actuellement disponibles et candidats pour répondre aux enjeux de la surveillance des hormones oestrogéniques (Jarosova et al 2014, Kunz et al. 2015). Si ces différents bioessais existants sont *a priori* pertinents en termes de spécificité et de sensibilité pour répondre à cet objectif, il existe aujourd'hui un **besoin d'inter-comparaison** et **d'harmonisation** des méthodes et pratiques pour valider leur application en surveillance des milieux aquatiques dans un **contexte réglementaire**.

Dans ce contexte, la présente étude menée dans le cadre d'AQUAREF visait à contribuer à la standardisation de la méthodologie d'évaluation de l'activité oestrogénique dans des eaux environnementales à l'aide de bioessais *in vitro* et *in vivo*, pour le diagnostic sur la présence de substances à activité oestrogénique. Le travail réalisé, initié en 2013, a consisté à :

- dresser un état de l'art non exhaustif recensant les méthodes biologiques actuellement utilisées pour détecter des substances à activité œstrogénique dans des matrices environnementales de type sédiment, eaux de surface et eaux usées. Ce travail a fait l'objet d'un rapport d'étape spécifique produit en 2015, et dont les principales conclusions sont rappelées dans le présent document.
- contribuer aux travaux européens sur les bioessais d'oestrogénicité menés en interface avec les instances réglementaires en charge de la DCE. Ces travaux incluent une participation aux groupes d'experts sur les bioessais et aux exercices de comparaison inter-laboratoires de bioessais d'évaluation de l'activité oestrogénique.

2. METHODOLOGIE DE LA MESURE DE L'ACTIVITE ESTROGENIQUE DANS LES EAUX

Un rapport bibliographique a été produit en décembre 2015 (Aït-Aïssa et Creusot, 2015) dressant un inventaire non exhaustif des méthodes et des pratiques existantes sur la base d'une revue de la littérature scientifique et des documents techniques existants. L'objectif était de présenter les grands principes de la méthodologie et d'identifier les étapes techniques clés qui peuvent être sources potentielles de variabilité inter-essais ou inter-laboratoires, dans une optique de standardisation.

La Figure 1 présente un protocole générique de l'évaluation de l'activité oestrogénique par les bioessais et déclinant les grandes étapes méthodologiques : la préparation de l'échantillon (incluant l'échantillonnage et la méthode d'extraction), le test cellulaire employé, ou encore l'interprétation et le traitement statistique des résultats.

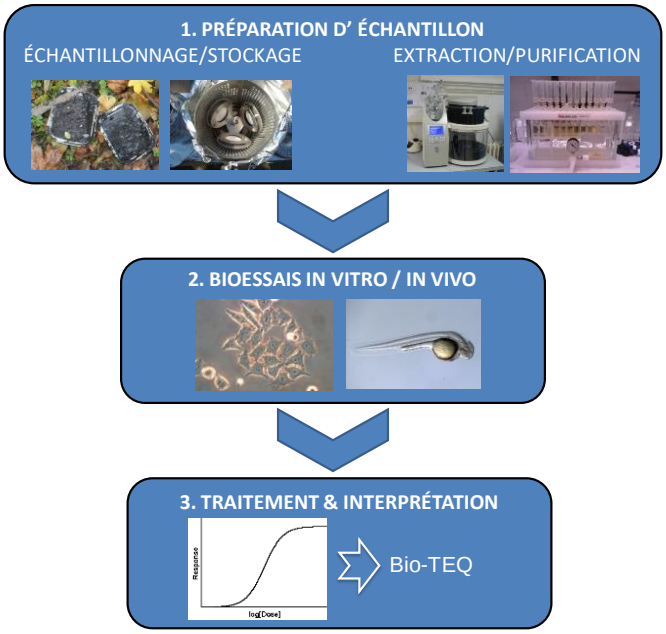


Figure 1. Démarche générique de l'évaluation de l'activité oestrogénique d'échantillons du milieu aquatique (eaux, sédiments) à l'aide de bioessais.

Ce bilan a mis en évidence la nécessité d'utiliser une méthode d'extraction pour concentrer les molécules actives généralement associées à la phase organique dissoute. Le choix de la méthode d'extraction dépendra de l'objectif visé, toutefois les méthodes à large spectre étant favorisées du fait de la diversité des molécules chimiques potentiellement impliquées. La validation des méthodes d'extraction (e.g. calculs de rendement d'extraction) doit se faire sur la base des bioessais *in vitro* en complément aux analyses chimiques (e.g. Creusot et al 2016, Neale et al 2018).

Concernant les modèles cellulaires *in vitro* utilisés, une diversité d'outils est aujourd'hui disponible pour la mesure du potentiel œstrogénique de matrices environnementales. Si certains de ces modèles ont fait l'objet d'études de (pré)validation pour l'étude de substances chimiques seules (i.e. HeLa-9903, ER-CALUX), il subsiste un besoin de les valider pour l'analyse d'échantillons environnementaux. Les récents travaux normatifs au niveau international concernant l'évaluation du potentiel œstrogénique des eaux et eaux usées pour les modèles basés sur les levures et les modèles basés sur des lignées cellulaires humaines vont dans ce sens (normes ISO-19040-1, ISO-19040-2, ISO-19040-3). Toutefois ces normes n'incluent pas l'utilisation d'une étape d'extraction. Il est également noté que les outils aujourd'hui disponibles sont très majoritairement basés sur le récepteur des œstrogènes (ER) humain alors que certaines études mettent en évidence des différences inter-espèces dans l'activation de ce récepteur. Par conséquent, de manière à appréhender le danger pour les organismes aquatiques, le développement de modèles basés sur les récepteurs de poisson semble pertinent.

Un intérêt majeur de ce type de bioessais est la possibilité de fournir une mesure quantitative de la concentration en substances actives, sous forme d'équivalents-estradiol (EEQ). Cette quantification passe par l'établissement de courbes concentration-réponse pour un échantillon donné et la comparaison de concentrations effectrices en échantillon (e.g. concentration induisant 20% d'effet ou EC20) à celle de la substance de référence, l'estradiol. Dans la littérature différentes méthodes sont utilisées pour déterminer ces EEQ (Wagner et al 2014). Si elles sont *a priori* équivalentes, l'utilisation de méthodes différentes peut être source de variabilité inter-essai et influer sur la valeur donnée par le bioessai.

S'il est aujourd'hui admis que les bioessais *in vitro* sont des outils pertinents pour répondre aux besoins actuels de surveillance et de gestion des substances à activité œstrogénique dans les milieux aquatiques, l'harmonisation des méthodes existantes constitue un point incontournable et nécessaire pour permettre leur utilisation effective dans les réseaux de surveillance.

3. INTER-COMPARAISON DE BIOESSAIS POUR LA SURVEILLANCE DES SUBSTANCES ŒSTROGENIQUES DANS LES EAUX

3.1 INTER-COMPARAISON DE BIOESSAIS IN VITRO POUR L'EVALUATION DE MELANGES DE SUBSTANCES

Une première étude inter-laboratoires a été menée sur 5 bioessais *in vitro* (YES, ERα-CALUX, MELN, T47D-KBluc and GeneBLAzer-ERα). L'objectif était de caractériser et comparer leurs performances intra- et inter-essais à quantifier l'activité estrogénique d'eaux dopées par un mélange de substances oestrogéniques (E1, E2, EE2 et bisphénol A). Les concentrations utilisées étaient représentatives d'échantillons fortement (10 fois les valeurs de NQE) et faiblement contaminés (proches des valeurs de NQE).

Cette étude était coordonnée par le centre suisse d'écotoxicologie (Ekotox Zentrum de l'Eawag). Ces travaux ont fait l'objet d'une publication présentée en Annexe A (Kunz et al 2017).

Dans l'ensemble, les résultats montrent que les 5 bioessais évalués présentent des niveaux de performance comparables pour l'analyse de ces mélanges. Le coefficient de variation (CV) moyen des concentrations en équivalents-estradiol (EEQ) pour tous les essais et tous les échantillons est de 32%. Pour les deux concentrations testées, une plus faible variabilité est observée lorsque les essais sont réalisés le jour-même (CV=30%) qu'en inter-jour (CV=37%). L'étape d'extraction en phase solide (SPE) de la phase organique dissoute n'a pas induit de variabilité additionnelle par rapport aux essais sur extraits dopés directement. Dans cette étude, parmi les 5 bioessais évalués, le bioessai sur ER-CALUX est celui qui a présenté la meilleure répétabilité et précision (CV global de 13%).

Cette étude a également permis de soulever certains points de vigilance/recommandations. Parmi elles :

- pour la comparaison avec les concentrations individuelles en substances, il est recommandé de définir et d'utiliser le facteur d'équivalence relative à E2 (EEF) des substances individuelles propre à chaque bioessai. L'EEF d'une substance peut varier d'un bioessai à un autre et ainsi influer sur la valeur d'EEQ prédite par les analyses chimiques, et donc sur la précision de la valeur d'EEQ donnée par le bioessai.

- l'analyse des données concentration-réponse se base sur la dérivation de la PC10 (concentration induisant 10 % de l'effet maximal du témoin positif) qui peut être sujette à une certaine variabilité. De ce fait, il est recommandé d'ajuster les gammes de dilution de sorte à obtenir une bonne résolution de la réponse autour des concentrations induisant ces faibles effets (bas de la courbe).

Au bilan, cette étude, basée sur un mélange simple de 4 molécules à activité oestrogénique, a fait la démonstration de la pertinence des bioessais *in vitro*, en combinaison avec une méthode d'extraction/concentration, pour déterminer des EEQ à des concentrations inférieures ou égales aux NQE.

3.2 COMPARAISON DE BIOESSAIS ET ANALYSES CHIMIQUES POUR L'ANALYSE DES HORMONES ESTROGENIQUES DANS LES EAUX DE SURFACE ET LES REJETS

Suite à la proposition d'inclure l'E_1, l'E_2 et l'EE_2 dans la *watch-list* de la DCE, un groupe de travail a été initié en 2013, sous l'impulsion de l'institut fédéral allemand d'Hydrologie (BfG), du centre suisse d'Ecotoxicologie appliqué de l'eawag-EPFL et de l'institut national supérieur de la Santé (ISS, Italie), afin de mener une réflexion sur l'utilisation de bioessais *in vitro* pour la surveillance des œstrogènes dans les milieux aquatiques. Ce groupe, qui incluait plus de 24 partenaires académiques et privés européens dont l'INERIS, a abouti au lancement d'un projet de type *Science-Policy Interface* soutenu par la Direction Générale de l'Environnement de la commission Européenne[1].

L'objectif général de ce projet visait à démontrer la faisabilité d'utiliser des bioessais *in vitro* pour la surveillance des substances à activité oestrogénique dans les masses d'eaux. Plus spécifiquement, il a contribué à :

- évaluer les performances relatives de diagnostic des composés oestrogéniques par un panel de 5 bioessais, en comparaison inter-bioessais et vis-à-vis de méthodes analytiques performantes (HR-MS/MS) ;
- faire la démonstration du lien surveillance chimique-surveillance biologique à large échelle ;
- proposer des critères d'évaluation DCE pour la surveillance des hormones dans les eaux.

Le projet a porté sur 33 échantillons d'eaux de surface (16) et eaux usées (17) collectés dans 7 pays européens. L'extraction des eaux a été centralisée par le même laboratoire et les extraits ensuite envoyés aux différents laboratoires en charges des bioessais (ER-CALUX, MELN, GeneBLAzer-ERα, HeLa-9903, pYES) et des analyses chimiques (3 méthodes par LC/MS/MS). L'ensemble des résultats a fait l'objet de deux publications scientifiques publiées en 2018 dans la revue *Trends in Analytical Chemistry*, présentées en Annexes B (Könneman et al 2018) et C (Käse et al 2018) de ce rapport.

D'une manière générale, les travaux menés dans le cadre de ce projet ont permis plusieurs avancées significatives sur la validation des bioessais d'estrogénicité dans un contexte réglementaire de surveillance.

Il a démontré une excellente performance analytique des outils *in vitro* pour détecter et quantifier des EEQ dans les matrices eaux usées et eaux de surface, avec des limites de quantification (LQ) inférieures à la NQE (Figure 2). Tous les bioessais ont permis de quantifier des EEQ dans tous les échantillons analysés (Figure 3). Des corrélations significatives, sur la base des EEQ, sont rapportées entre les résultats des bioessais (tous tests) et ceux des analyses chimiques des hormones E1, E2 et EE2 (Annexe C, Figure 6). Ce constat, qui corrobore ceux d'études antérieures, démontre la pertinence du concept de l'utilisation des méthodes biologiques (en phase de screening) en remplacement des analyses chimiques pour la surveillance des hormones stéroïdiennes de la watch list.

Dans l'ensemble, cette étude montre la validité des bioessais *in vitro* pour quantifier des EEQ dans des eaux de surface et des rejets.

[1] http://www.ecotoxcentre.ch/projects/aquatic-ecotoxicology/monitoring-of-steroidal-estrogens/

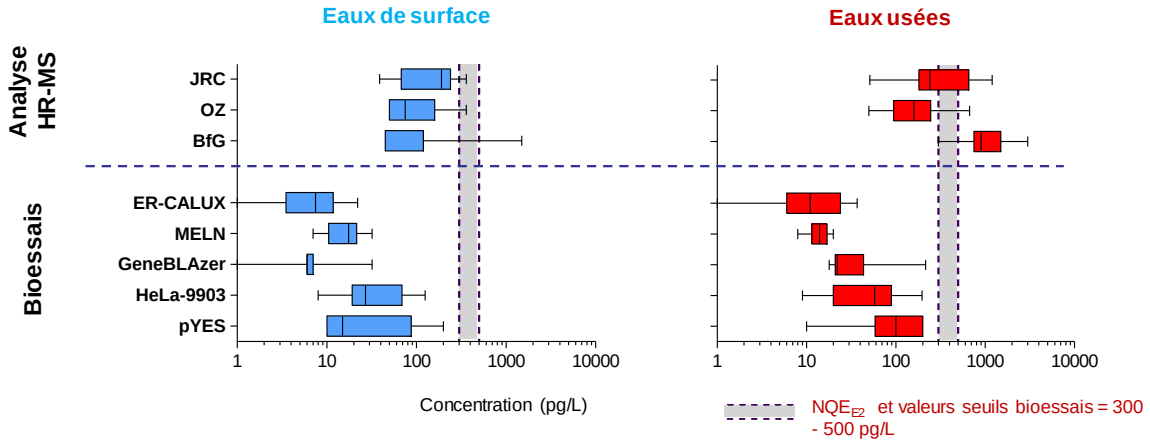


Figure 2. Limites de quantification (LQ) des bioessais in vitro *(en EEQ) et des analyses chimiques (pour l'E2) dans les eaux de surface et eaux usées. Comparaison aux gammes de valeurs seuils 0,3-0,5 ngEEQ/L) éprouvées dans ce projet.*

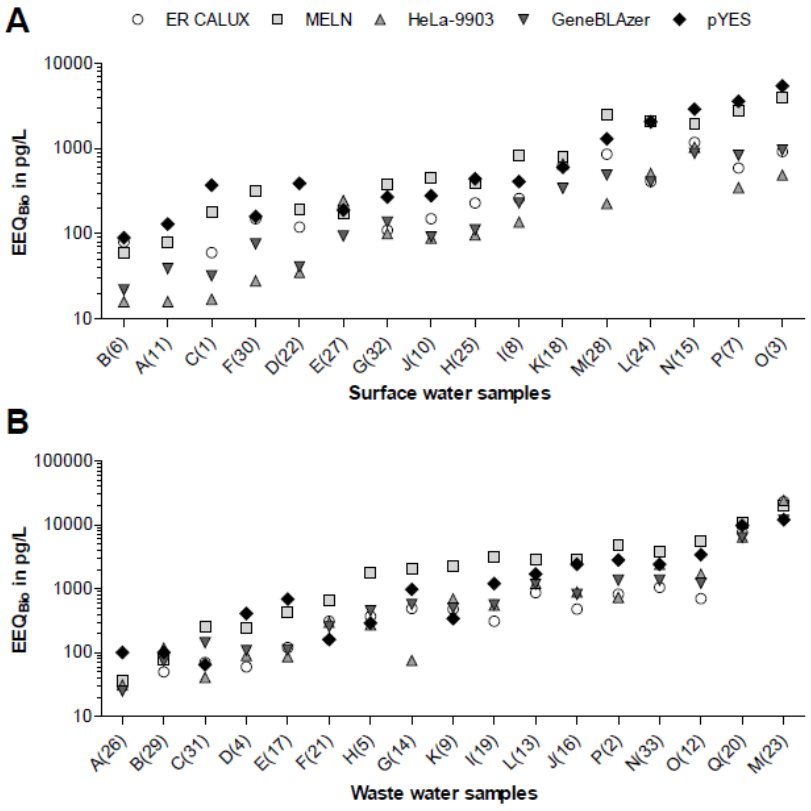


Figure 3. Concentrations en EEQ dans tous les échantillons d'eaux de surface (A) et d'eaux usées (B) mesurées avec les différents bioessais in vitro.

Ce projet a mis en avant la bonne sensibilité des bioessais *in vitro* pour répondre aux critères de détection requis dans la DCE (LQ inférieure à la NQE de l'E2). A contrario, pour 15% des eaux de surfaces et 40 % des rejets, les LQ des analyses chimiques étaient trop fortes pour pouvoir statuer sur l'état chimique de l'échantillon au regard des concentrations en hormones.

La comparaison des EEQ données par les différents bioessais (valeurs moyennes des différents bioessais) avec une valeur seuil générique de 0,4 ng EEQ/L (correspond à une concentration en estradiol-équivalent dérivée la NQE de l'E2, Kunz et al 2015) montrent que 61±11% des eaux de surface et 38±10% des eaux usées analysées seraient en deçà de cette valeur seuil et présentent donc un statut conforme avec le critère DCE (Annexe C, Tableau 2). Ce pourcentage varie cependant de 50 à 75% pour les eaux de surface et de 24 à 47% pour les eaux de rejet selon le bioessai utilisé. Cette variabilité inter-bioessai, inhérente aux modèles cellulaires utilisés, montre que pour un certain nombre d'échantillons ayant une valeur d'EEQ proche de 0,4 ng/L, l'utilisation d'un bioessai ou un autre peut conduire à un faux négatif ou à un faux positif. L'amélioration de cette valeur seuil pourrait donc passer par la définition de valeur seuils spécifiques à chaque bioessai, ce qui permettrait une évaluation plus juste du risque associé à la valeur mesurée *in vitro* pour un bioessai donné.

Au bilan, ce projet a fait la démonstration des performances des bioessais, en termes de sensibilité et de spécificité, pour l'analyse des hormones œstrogènes stéroïdiens dans les masses d'eaux en répondant aux critères de surveillance réglementaire (i.e. seuils de détection). Des études complémentaires dans un contexte de surveillance opérationnelle seront nécessaires pour éprouver et affiner les valeurs seuils propres à chaque bioessai.

4. EVALUATIONS INTER-LABORATOIRES DE BATTERIE DE BIOESSAIS POUR EVALUER DES MELANGES DE SUBSTANCES PRIORITAIRES

Dans une démarche européenne d'utilisation des bioessais pour la surveillance des substances prioritaires et de polluants émergents, deux études inter-laboratoires visant à intercomparer des batteries de bioessais pour évaluer la toxicité de mélanges reconstitués de substances, ont été menées respectivement sous l'impulsion du JRC (Joint Research Center) de la commission européenne, d'une part, et du réseau NORMAN, d'autre part. Les panels de bioessais évalués incluaient des bioessais d'écotoxicité (e.g. poisson, algues, daphnies) et différents bioessais basés sur les mécanismes d'action des toxiques, y compris les bioessais d'oestrogénicité portés par l'INERIS.

Les résultats de ces études ont fait l'objet de deux publications scientifiques présentées en Annexes E (Carvalho et al 2014) et F (DiPaolo et al 2016) de ce rapport.

L'étude JRC a évalué des eaux dopées par deux mélanges de 14 ou 18 substances prioritaires de la DCE à des concentrations équivalentes à leur NQE (Carvalho et al 2014). Les résultats ont montré la capacité des mélanges de substances prioritaires étudiés à induire une réponse dans différents bioessais, à des concentrations parfois proches des seuils NQE préconisés. Concernant les bioessais d'activité oestrogénique, ils ont permis de détecter la présence des œstrogènes aux sein de ces mélanges et de quantifier une concentration EEQ à des niveaux des seuils NQE.

Ces études 1) démontrent la pertinence d'une batterie de bioessais pour mettre en évidence des effets de substances avec des mode d'actions variés, 2) suggèrent la prise en compte des effets de substances lorsqu'elles sont en mélanges, y compris à des concentrations susceptibles d'être sans effet (i.e. NQE). Par exemple, le test ER-CALUX montre un effet pour des substances présentes en mélange à des niveaux équivalents aux valeurs de leurs NQE (Figure 5 de l'article présenté en Annexe D). Ces études ont confirmé la bonne performance des bioessais d'oestrogénicité pour détecter des molécules actives (E2, EE2, bisphénol A) au sein de mélanges incluant d'autres molécules sans effet (anti)oestrogénique connu.

5. CONCLUSIONS

Ces travaux ont contribué à montrer la pertinence et la validité des bioessais *in vitro* pour la mesure de l'activité estrogénique dans des échantillons d'eaux de surface et de rejets dans un contexte réglementaire. La démonstration a été faite, au niveau européen, de la performance des outils proposés pour répondre aux exigences analytiques de surveillance des hormones oestrogéniques dans ces matrices environnementales. Des premières recommandations sur les valeurs seuils réglementaires ont été émises pour permettre leur application à l'évaluation et la gestion des masses d'eaux dans un contexte réglementaire.

Dans la continuité de ces études, des travaux complémentaires visent à éprouver la pertinence des valeurs d'EEQ mesurées *in vitro* vis-à-vis d'un effet mesuré à l'échelle de l'organisme, à l'aide du bioessai *in vivo* EASZY développé chez le poisson zèbre. Ces travaux sont réalisés dans le cadre du réseau de surveillance prospective (action RSP 5 : « Application de nouveaux outils pour préparer la surveillance future », supportée par AQUAREF). La mise en œuvre d'une démarche intégrant la mesure d'effet *in vivo* aidera à la définition/amélioration de valeurs seuils spécifiques des différents bioessais *in vitro*.

Enfin, l'ensemble de ces données alimentent les travaux en cours au sein de groupes d'experts au niveau national (groupe de travail national sur les Bioessais) et à la commission européenne (DG-Environment, sous-groupe « Effect-based methods » du groupe de travail « Chemicals ») visant à proposer une approche plus holistique de l'évaluation de l'état chimique des eaux pour les prochains cycles de la DCE. Un des objectifs étant d'introduire les outils biologiques dans la stratégie globale de gestion du risque.

6. VALORISATION

Rapports

Aït-Aïssa S et Creusot N (2015) Approche bio-analytique : validation de la mesure de l'activité œstrogénique dans les matrices environnementales – Revue bibliographique. Rapport AQUAREF-INERIS DRC-15-136927-12327A, 2015 – 32 p.

Carvalho R *et al* (2013) Testing comparability of existing and innovative bioassays for water quality assessment: an European-wide exercise. JRC Scientific and Policy Report.

Publications scientifiques

Carvalho RN, Arukwe A, Aït-Aïssa S, Bado-Nilles A, Balzamo S, Baun A, Belkin S, Blaha L, Brion F, Conti D, Creusot N, Essig Y, Ferrero VE, Flander-Putrle V, Fürhacker M, Grillari-Voglauer R, Hogstrand C, Jonáš A, Kharlyngdoh JB, Loos R, Lundebye AK, Modig C, Olsson PE, Pillai S, Polak N, Potalivo M, Sanchez W, Schifferli A, Schirmer K, Sforzini S, Stürzenbaum SR, Søfteland L, Turk V, Viarengo A, Werner I, Yagur-Kroll S, Zounková R, Lettieri T. (2014) Mixtures of Chemical Pollutants at European Legislation Safety Concentrations: How Safe are They? Toxicological Sciences, 141(1):218-33

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

Kase R, Javurkova B, Simon E, Swart K, Schlüsener M, Buchinger S, Könemann S, Escher BI, Carere M, Dulio V, Aït-Aïssa S, Hollert H, Valsecchi S, Polesello S, Behnisch P, di Paolo C, Olbrich D, Tavazzi S, 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 oestrogens in surface and wastewater. Trends in Analytical Chemistry, 102, 343-358.

Köneman S, Kase R, Simon E, Stwart K, Buchinger S, Schlüsener M, Hollert H, Escher BI, Werner I, Aït-Aïssa S, Dulio V, Valsecchi S, Polesello S, Schaan S, Behnisch P, Javurkova B, Perceval O, Di Paolo C, Clayton H, Olbrich D, Tavazzi S, Sychrova E, Gundlach M, Schlichting R, Leborgne L, Clara M, Scheffknecht C, Marneffe Y, Chalon C, Tušil P, Soldan P, von Danwitz B, Schwaiger J, Moran Palao A, San Martín Becares I, Bersani F, Vermeirssen E, Hilscherova K, Reifferscheid G, Carere M (2018) Effect-based and chemical analytical methods to monitor oestrogens under the EuropeanWater Framework Directive. Trends in Analytical Chemistry, 102, 225-235.

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

Communications

Aït-Aïssa S, Brion F, Perceval O. Performances relatives des outils bio-analytiques pour la détection d'activité oestrogénique dans les eaux de surface et les rejets. *Comité de pilotage du plan micropolluants, MTES, La Défense, Novembre 2017*.

Carvalho R, Arukwe A, Aït-Aïssa S, Bado-Nilles A, Balzamo S, Baun A, Belkin S, Blaha L, Brion F, Conti D, Creusot N, Essig Y, Ferrero VE, Flander-Putrle V, Fürhacker M, Grillari-Voglauer R, Hogstrand C, Jonáš A, Kharlyngdoh JB, Loos R, Lundebye AK, Modig C, Olsson PE, Pillai S, Polak N, Potalivo M, Sanchez W, Schifferli A, Schirmer K, Sforzini S, Stürzenbaum SR, Søfteland L, Turk V, Viarengo A, Werner I, Yagur-Kroll S, Zounková R, Lettieri T « Toxicological effects of chemical mixtures of pollutants at European environmental quality standards » *24th annual meeting of the SETAC Europe, May 11-15, 2014, Basel, Suisse*.

Di Paolo C, Seiler TB, Seidensticker S, Ottermann R, Hollert H, Keiter S, Dulio V, Schulze T, Brack W, Buchinger S, Cousin X, Hamers T, Zwart N, Hilscherová K, Zounková R, Jonas A, van der Oost R, Nguyen MT, Aït-Aïssa S, Maillot-Maréchal E, Escher B, Tang J, Vermeirssen E, Marneffe Y, Chalon C, Breitholtz M. The NORMAN interlaboratory study on biotesting of spiked water extracts. *25th annual meeting of the SETAC Europe, May 2015, Barcelona, Spain.*

Käse R, Werner I, Hollert H, Vermeirssen E, Buchinger S, Behnisch P, Jarosova B, Lettieri T, Carvalho RN, Loos R, Clayton H, Perceval O, Ait-Aissa S, Creusot N, Reifferscheid G, Ternes T, Heiss C, Seiler T, Kunz PY, Kienle C, Dulio V, Valsecchi S, Carere M. Effect-based and chemical analytical monitoring for the steroidal estrogens: An international project to cope with a monitoring challenge. *25th annual meeting of the SETAC Europe, May 2015, Barcelona, Spain*.

Kunz P., Aït-Aïssa S, Creusot N, Homazava, Jayasinghe, Maletz, Schifferli, Schönlau, Denslow N, Hollert H, Werner I. « Effect-based tools for monitoring (xeno)estrogens in surface waters: Variability and reproducibility of sample preparation and 5 different in vitro assays » *24th annual meeting of the SETAC Europe, May 11-15, 2014, Basel, Suisse*.

Lettieri T, Arukwe A, Aït-Aïssa S, Bado-Nilles A, Balzamo S, Baun A, Belkin S, Blaha L, Brion F, Conti D, Creusot N, Essig Y, Ferrero VE, Flander-Putrle V, Fürhacker M, Grillari-Voglauer R, Hogstrand C, Jonáš A, Kharlyngdoh JB, Loos R, Lundebye AK, Modig C, Olsson PE, Pillai S, Polak N, Potalivo M, Sanchez W, Schifferli A, Schirmer K, Sforzini S, Stürzenbaum SR, Søfteland L, Turk V, Viarengo A, Werner I, Yagur-Kroll S, Zounková R,. Carvalho RN (2014) Mixtures of Chemical Pollutants at European Legislation Safety Concentrations: How Safe are They? *24th annual meeting of the SETAC Europe, May 11-15, 2014, Basel, Suisse*.

7. REFERENCES BIBLIOGRAPHIQUES

Aït-Aïssa S, Creusot N (2015) Approche bio-analytique : validation de la mesure de l'activité œstrogénique dans les matrices environnementales – Revue bibliographique. Rapport AQUAREF-INERIS DRC-15-136927-12327A, 2015 – 32 p.

Creusot N, Devier MH, Budzinski H, Aït-Aïssa (2016) Evaluation of an extraction method for a mixture of endocrine disrupters in sediment using chemical and in vitro biological analyses. Environmental Science and Pollution Research 23(11): 10349-10360.

Desbrow C, Routledge EJ, Brighty GC, Sumpter JP, Waldock M (1998) Identification of estrogenic chemicals in STW effluent. 1. Chemical fractionation and in vitro biological screening. Environmental Science and Technology 32(11): 1549-1558.

Jarosova B, Ersekova A, Hilscherova K, Loos R, Gawlik BM, Giesy JP, Blaha L (2014) Europe-wide survey of estrogenicity in wastewater treatment plant effluents: the need for the effect-based monitoring. Environmental Science and Pollution Research 21(18): 10970-10982.

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 of the United States of America 104(21): 8897-8901.

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: 107-115.

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

Neale PA, Brack W, Ait-Aissa S, Busch W, Hollender J, Krauss M, Maillot-Marechal E, Munz NA, R. Schlichting R, Schulze T, Vogler B, Escher BI (2018) Solid-phase extraction as sample preparation of water samples for cell-based and other in vitro bioassays. Environmental Science-Processes & Impacts 20(3): 493-504.

Wernersson AS, Carere M, Maggi C, Tusil P, Soldan P, James A, Sanchez W, Dulio V, Broeg K, Reifferscheid G, Buchinger S, Maas H, Van Der Grinten E, O'Toole S, Ausili A, Manfra L, Marziali L, Polesello S, Lacchetti I, Mancini L, Lilja K, Linderoth M, Lundeberg T, Fjallborg B, Porsbring T, Larsson DGL, Bengtsson-Palme J, Forlin L, Kienle C, Kunz P, Vermeirssen E, Werner I, Robinson CD, Lyons B, Katsiadaki I, Whalley C, den Haan K, Messiaen M, Clayton H, Lettieri T, Carvalho RN, Gawlik BM, Hollert H, Di Paolo C, Brack W, Kammann U, Kase R (2015) The European technical report on aquatic effect-based monitoring tools under the water framework directive. Environmental Sciences Europe 27: 1-11.

ANNEXE A: KUNZ ET AL (2017) EFFECT-BASED TOOLS FOR MONITORING ESTROGENIC MIXTURES: EVALUATION OF FIVE IN VITRO BIOASSAYS, *WATER RESEARCH* (11 PAGES)

Water Research 110 (2017) 378–388

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Effect-based tools for monitoring estrogenic mixtures: Evaluation of five *in vitro* bioassays

Petra Y. Kunz [a, 1, 2], Eszter Simon [a, 1], Nicolas Creusot [b], B. Sumith Jayasinghe [c], Cornelia Kienle [a], Sibylle Maletz [d], Andrea Schifferli [a], Christine Schönlau [d, 3], Selim Aït-Aïssa [b], Nancy D. Denslow [c], Henner Hollert [d], Inge Werner [a], Etiënne L.M. Vermeirssen [a, *]

[a] Swiss Centre for Applied Ecotoxicology Eawag-EPFL, 8600 Dübendorf, Switzerland
[b] INERIS, Institut National de l'Environnement Industriel et des Risques, Unité ECOT, Verneuil en Halatte, France
[c] University of Florida, Center for Environmental and Human Toxicology, Gainesville, FL, USA
[d] RWTH Aachen University, Institute for Environmental Research, Aachen, Germany

ARTICLE INFO

Article history:
Received 4 August 2016
Received in revised form
12 October 2016
Accepted 23 October 2016
Available online 24 October 2016

Keywords:
ERTA
(Xeno)estrogens
Water Framework Directive
Precision
Repeatability
EEQ derivation

ABSTRACT

In vitro estrogen receptor transactivation assays (ERTAs) are increasingly used to measure the overall estrogenic activity of environmental water samples, which may serve as an indicator of exposure of fish or other aquatic organisms to (xeno)estrogens. Another potential area of application of ERTAs is to assist the monitoring of the potent steroids 17β-estradiol (E2) and 17α-ethinylestradiol (EE2) under the Water Framework Directive (WFD) watch-list mechanism. Chemical analysis of E2 and EE2 is currently hampered by limits of quantification being mostly above the proposed annual average Environmental Quality Standards (AA-EQS) of 0.4 and 0.035 ng/L, respectively. Sensitive ERTAs could circumvent current detection challenges by measuring total estrogenic activity expressed as E2-equivalent (EEQ) concentrations.

However, the use of different ERTAs results in different EEQ concentrations for the same sample. Reasons for these differences are known, but it remains unclear how to use and interpret bioassay results in a harmonised way. The aim of this study was to compare the intra- and inter-day variability of EEQ measurements using five different ERTAs (YES, ERα-CALUX, MELN, T47D-KBluc and GeneBLAzer-ERα) with regard to their applicability as effect-based tools in environmental monitoring.

Environmentally relevant artificial mixtures of (xeno)estrogens were prepared to represent samples with higher (i.e. multiple times the AA-EQS for E2) or lower pollution levels (i.e. around the AA-EQS for E2). Mixtures were tested either directly or following solid phase extraction (SPE). The SPE step was included, as environmental samples typically require enrichment before analysis. Samples were analysed repeatedly to test intra-day and inter-day variability. Estrogenicity was quantified using the 10% effect level (PC10) of the positive control (E2) and expressed as EEQ concentrations.

The average coefficient of variation (CV) of EEQ concentrations for the five ERTAs and all samples was 32%. CV was lower for intra-day experiments (30%) compared to inter-day experiments (37%). Sample extraction using SPE did not lead to additional variability; the intra-day CV for SPE extracted samples was 28%. Of the five ERTAs, ERα-CALUX had the best precision and repeatability (overall CV of 13%).

1. Introduction

Concerns about natural and xenoestrogens in the aquatic environment arose as early as the 1970s (see Aherne et al., 1985). Following observations on hermaphroditic fish and subsequently observations on vitellogenin induction in male rainbow trout in the UK, caused by treated sewage effluents (Purdom et al., 1994),

* Corresponding author.
E-mail address: etienne.vermeirssen@oekotoxzentrum.ch (E.L.M. Vermeirssen).
[1] Authors contributed equally to this manuscript.
[2] Present address: Federal Office for the Environment FOEN, Air Pollution Control and Chemicals Division, Biocides and Plant Protection Products Section, Berne, Switzerland.
[3] Present address: Örebro University, MTM Research Centre, Örebro, Sweden.

http://dx.doi.org/10.1016/j.watres.2016.10.062
0043-1354/

estrogens in the aquatic environment became a research focus (Sumpter and Johnson, 2008). It became apparent, that the issue of estrogens affecting the aquatic ecosystem was not limited to the UK (Tyler et al., 1998) but of concern world-wide (e.g. Folmar et al., 1996; Vethaak et al., 2002; Ying et al., 2002). Consequently, annual average environmental quality standards (AA-EQS) – values specifying the permissible concentration of potentially hazardous chemicals in an environmental sample – were developed to protect aquatic wildlife from excessive exposure. However, so far, these EQS-values have not been adopted into regulatory frameworks for lack of monitoring options (European Commission, 2013; Johnson et al., 2013).

First developed some 25 years ago, *in vitro* bioassays, such as estrogen-receptor transactivation assays (ERTA, e.g. McDonnell et al., 1991), are increasingly used to assess estrogenic activity in the aquatic environment. To this end, the assay response of a sample is compared to that of a reference compound, typically 17β-estradiol (E2), which is the natural ligand of the estrogen receptor. In this way, the estrogenic activity of a sample can be quantified and expressed as an E2 equivalent (EEQ) concentration. The EEQ concentration signifies the E2 concentration that elicits the same effect as the combined activity of all compounds with agonistic and antagonistic estrogenic properties present in the sample (see e.g. Escher and Leusch, 2012).

Such ERTAs have proven highly valuable for ecotoxicological research and monitoring purposes (Wernersson et al., 2015). Our current knowledge on the extent of estrogenic contamination in surface waters can partly be attributed to the application of these tools in many countries (e.g. Vethaak et al., 2002; ARCEM, 2003). In addition, ERTAs have been pivotal in identifying the various compounds that cause estrogenicity in surface waters and effluents. Typically, the main contributors are estrone (E1), E2 and 17α-ethinylestradiol (EE2, Desbrow et al., 1998). To a lesser extent alkylphenols, phtalates and bisphenol A (BPA), may also contribute to the estrogenic activity in the aquatic environment (Vethaak et al., 2005; Fernandez et al., 2009). ERTAs have been suggested to be used as effect-based tools to monitor the efficiency of wastewater treatment processes (Maletz et al., 2013; Schindler Wildhaber et al., 2015) and assess water quality of surface waters affected by effluent (Kienle et al., 2015). Finally, ERTAs offer the possibility of monitoring compounds for which AA-EQS are very low (i.e. 35 pg/L for EE2, Johnson et al., 2013) and current chemical analysis techniques are not yet sensitive enough (Loos, 2012; Kunz et al., 2015).

Over the years, many different ERTAs have been developed, which differ in their sensitivity to estrogenic compounds. The most common ERTAs use either modified yeast cells (e.g. McDonnell et al., 1991; Routledge and Sumpter, 1996) or human cell lines (e.g. Soto et al., 1995; Balaguer et al., 1999; Leusch et al., 2010). Obviously, in a yeast cell, the human estrogen receptor is behaving in an "artificial environment" when compared to human cell lines. For example, uptake and metabolism of estrogenic compounds differ between yeast and human cell lines due to differences in cell membrane structure and the presence of enzymes (Zacharewski, 1997, but see also Bovee et al., 2008 and Bovee and Pikkemaat, 2009). Generally, human cell lines are more sensitive to the natural ligand E2, typically by a factor of ca. 10 (e.g. Leusch et al., 2010). Furthermore, different ERTAs do not respond the same way to other natural estrogens or xenoestrogens such as EE2, nonylphenol or BPA. Consequently, relative sensitivities to specific estrogenic compounds vary across assays. This complicates the evaluation of assay responses to mixtures of chemicals or environmental samples containing multiple receptor agonists and antagonists (Ihara et al., 2014). Depending on its chemical makeup, analysis of a specific environmental sample with different ERTAs will therefore result in different EEQ concentrations (Leusch et al., 2010). Besides the type of ERTA, another factor affecting EEQ concentrations are the different methods applied for data analysis. Often the dose-response curves of ERTA obtained for the reference compound and sample analysed are not parallel and may have different maxima. In this case, an interpolation of EEQ concentrations at either different sample concentrations or different effect levels will directly affect the EEQ concentration (Wagner et al., 2013), see also (Villeneuve et al., 2000). Generally, bioassay data analysis strategies are not described in sufficient detail (for review see: Wagner et al., 2013), which further complicates the comparison of published data.

Given the low contamination levels at which the analytes of interest have to be detected (i.e. EEQ concentrations <1 ng/L), an enrichment step prior to the bioassay is necessary. Several concentration methods are available: e.g. evaporation, liquid-liquid extraction or solid phase extraction (SPE; López de Alda and Barceló, 2001). When such methods are used it is important to consider whether there is sufficient recovery of the most important ER-agonists (as well as antagonists) that can be present in an environmental sample.

Our overall objective is to validate and then apply a set of robust ERTAs, for monitoring surface water quality as well as treated effluents in a regulatory context (Escher et al., 2013; Kunz et al., 2015; Wernersson et al., 2015; Kienle et al., 2015; Schindler Wildhaber et al., 2015), see also Mehinto et al. (2015). The aim of this study was to compare intra- and inter-day variability and precision of five frequently applied ERTAs for mixtures of estrogenic compounds commonly detected in treated wastewater (E1, E2, EE2, and BPA). Concentrations were chosen to be representative of treated sewage effluent (after biological treatment EEQ concentrations are typically between 0.05 and 8 ng/L, (Vethaak et al., 2005; Margot et al., 2013) and effluent-receiving surface waters (EEQ concentrations in surface waters are typically between LOQ and 2 ng/L, (Vermeirssen et al., 2008; Williams et al., 2012). In addition, we examined recovery of estrogens after SPE extraction.

2. Materials and methods

2.1. Chemicals and solvents

The hormones E2 (≥98% pure) and E1 (>99% pure); the pharmaceutical EE2 (≥98% pure) and the industrial chemical BPA (>99% pure) were purchased form Sigma-Aldrich (Buchs, Switzerland). Stock solutions were made in ethanol and stored in the dark at −20 °C. Analytical grade ethanol, n-hexane, acetone (all EMSURE®) and 30% hydrochloric acid (HCl; Suprapur®) were purchased from Merck (Zug, Switzerland). Analytical grade methanol (OPTIMA) was obtained from Fisher Scientific (Reinach, Switzerland). Ultrapure water was produced using a Barnstead Nanopure system (18.2 MΩ cm, Thermo Scientific, Allschwil, Switzerland).

2.2. Sample preparation

Ethanol was spiked with E1, E2, EE2 and BPA (Mixture). Ultrapure water was spiked with a mixture of ethanolic stock solutions of E1, E2, EE2 and BPA prior to SPE (Mixture-SPE). An overview of the compounds and concentrations in the mixtures and the spiked water samples is provided in Table 1. Mixtures were prepared at two concentration levels. $Mixture_{low}$ and $Mixture\text{-}SPE_{low}$ represented surface water and their respective SPE extracts with an EEQ concentration around the AA-EQS proposal for E2 (0.4 ng E2/L). $Mixture_{high}$ and $Mixture\text{-}SPE_{high}$ were prepared to resemble effluent receiving surface waters and their respective SPE extracts (1.3–5.6 ng EEQ/L, Table 3). As positive controls (PC), ethanol (1 mL) and ultrapure water (1 L) were spiked with 6 ng E2 (PC_{E2} and PC-

Table 1
Overview of concentrations in mixtures, the methods used for sample preparation (solid phase extraction, SPE), concentration factors (CF), and parameters evaluated.

	Concentration (ng/mL)			Concentration (ng/L)		
	PC_{E2}	$Mixture_{low}$	$Mixture_{high}$	$PC\text{-}SPE_{E2}$	$Mixture\text{-}SPE_{low}$	$Mixture\text{-}SPE_{high}$
17β-estradiol (E2)	6	0.1	1	6	0.1	1
Estrone (E1)	–	0.8	8	–	0.8	8
17α-ethinylestradiol (EE2)	–	0.01	0.1	–	0.01	0.1
Bisphenol-A (BPA)	–	50	500	–	50	500
Sample preparation	–	–	–	SPE	SPE	SPE
Analysis	ERTA	ERTA	ERTA	ERTA	ERTA	ERTA
CF				500	1000	500
Parameters	intra- and inter-day variability			$ERTA_{SPE}$ variability		

$Mixture_{low}$ and $Mixture\text{-}SPE_{low}$: represent surface water with an EEQ concentration around the proposed AA-EQS for E2 (0.4 ng/L).
$Mixture_{high}$ and $Mixture\text{-}SPE_{high}$: represent effluent receiving surface water (1.3–5.6 ng EEQ/L, Table 3).
PC_{E2} and $PC\text{-}SPE_{E2:}$ positive controls with E2 only.

Table 2
Characteristics of estrogen receptor transactivation assays.

	YES	ERα-CALUX	T47D-KBluc	MELN	GeneBLAzer-ERα
Organism/cell-line	Genetically modified yeast (*S. cervisiae*)	Human U2OS osteosarcoma	Human T47D breast adeno-carcinoma	Human mammary adeno-carcinoma (MCF7)	Human embryonic kidney HEK293
Receptor	stably transfected hERα[a]	stably transfected hERα	endogenous hERα (hERβ)	endogenous hERα	stably transfected hERα
Reference compound[b]	17β-estradiol (E2) (2.6–341 ng/L)	E2 (0.03–27 ng/L)	E2 (0.03–27 ng/L)	E2 (0.03–27 ng/L)	E2 (1.25–2724 ng/L)
Endpoint for estrogenicity	Colour-change	Luciferase activity	Luciferase activity	Luciferase activity	Fluorescence
EC_{50}[c] of the reference compound	(–)	3.2 ng E2/L	2.7 ng E2/L	1.4 ng E2/L	27 ng E2/L
Reference	(Routledge and Sumpter, 1996)	(van der Linden et al., 2008)	(Wilson et al., 2004)	(Balaguer et al., 1999)	(Huang et al., 2011)

[a] hERα: human estrogen receptor alpha.
[b] Concentration ranges of the reference dose-response curves are given in brackets.
[c] EC_{50}: the concentration of the reference compound inducing half-maximal response.

Table 3
Concentrations of mixture components in the mixtures (top panel; 17β-estradiol, E2; estrone, E1; 17α-ethinylestradiol, EE2; and bisphenol A, BPA) as well as E2 equivalence factors (EEF) and nominal E2 equivalent (EEQ) concentrations (lower panel).

	Concentration (ng/mL for Mixtures and ng/L for $Mixtures_{SPE}$)			
	E2	E1	EE2	BPA
$Mixture_{low}$ and $Mixture\text{-}SPE_{low}$	0.1	0.8	0.01	50
$Mixture_{high}$ and $Mixture\text{-}SPE_{high}$	1	8	0.1	500

	EEF				Nominal EEQ (ng/mL or ng/L)	
	E2	E1	EE2	BPA	$Mixture_{low}$ and $Mixture\text{-}SPE_{low}$	$Mixture_{high}$ and $Mixture\text{-}SPE_{high}$
YES	1	0.26	1.2	6.5E-05	0.32	3.2
ERα-CALUX	1	0.02	1.3	2.7E-05	0.13	1.3
MELN	1	0.02	0.9	4.5E-05	0.13	1.3
GeneBLAzer-ERα®	1	0.27	6.1	1.4E-04	0.38	3.8
T47D-KBluc[a]	1	0.53	2.9	4.9E-06	0.56	5.6

[a] EEFs were provided by the laboratories performing the assays except for T47D-KBluc (for source of EEFs see Supplementary data).

Table 4
Contribution of each compound in the mixtures as a percentage of the total 17β-estradiol equivalent (EEQ) concentrations. The individual compound contribution in $Mixture_{low}$ and $Mixture_{high}$ was concurrent due to the same mixture composition. Dominating compounds in the mixture are highlighted.

Contribution to EEQ (%)	YES	ERα-CALUX	MELN	GeneBLAzer-ERα	T47D-KBluc
Estrone (E1)	**66%**	12%	13%	**56%**	**77%**
17β-estradiol (E2)	31%	**77%**	**78%**	26%	18%
17α-ethinylestradiol (EE2)	4%	10%	7%	16%	5%
Bisphenol-A (BPA)	1%	1%	2%	2%	0%

SPE_{E2}, respectively, see Table 1). All samples were prepared from the same stock solution and aliquots were generally shipped to testing labs at ambient temperature in 2 μL of DMSO after solvent evaporation. For the intra-day and inter-day variability experiments with ERα-CALUX and T47D-KBluc ethanolic solutions were sent to the lab on dry ice. Testing labs redissolved the samples in the appropriate solvent before analysis. Detailed sample handling procedures are provided in the Supplementary data.

2.3. Solid phase extraction

Five replicate 1 L aliquots of the Mixture-SPE_{low} and blank samples (unspiked ultrapure water) and five replicate 0.5 L aliquots of the Mixture-SPE_{high} and PC-SPE_{E2} samples were adjusted to pH 3 (±0.1) with 0.5 M HCl and extracted using LiChrolut® EN-RP18 cartridges (Merck, Darmstadt, Germany; Escher et al., 2008a). Briefly, cartridges were conditioned with 2 mL hexane, 2 mL acetone, 6 mL methanol and 6 mL ultrapure water. After passing the samples, the cartridges were dried under nitrogen. The cartridges were eluted with 4 mL acetone and 1 mL methanol. Eluates were evaporated under nitrogen to near dryness and reconstituted in 1 mL ethanol, resulting in concentration factors of 1000 (Mixture-SPE_{low}, Blank) or 500 (Mixture-SPE_{high}, PC-SPE_{E2}). Extracts of the same sample type were pooled (five times 1 mL) and stored at −20 °C in amber glass vials. This procedure was repeated on different days to produce five sets of three samples per day. Pools from the five SPE days were split into five aliquots (1 mL per ERTA) and evaporated till dryness under a gentle nitrogen flow. Subsequently, 2 μL of DMSO was added to each vial.

2.4. In vitro estrogen receptor transactivation assays

The principle of different ERTAs is very similar. Estrogenic chemicals present in a sample diffuse into the ERTA cell and bind to the estrogen receptor resulting in the dimerization of two estrogen-bound receptors. This dimer complex then interacts with other transactivation factors present in cells and binds to and activates specific DNA sequences called estrogen responsive elements, which regulate the transcription of estrogen responsive genes. Upon activation, the reporter gene produces a gene product (e.g. luciferase or β-galactosidase) that can be measured and quantified (e.g. as a light emitting reaction or a colour change). Five ERTAs were compared, one yeast-based assay and four assays using human cell lines (Table 2).

The reference compound (E2) and samples (PCs, Mixtures and $Mixtures_{SPE}$; Table 1) were tested in triplicate dilution series. Each participating laboratory performed the analysis according to their own standard operating procedures with regard to e.g. 96-well plate layout, number of solvent controls, number of dilution steps etc. Detailed descriptions of the assays are provided in the Supplementary data.

2.5. Data analysis

Data were accepted when the coefficient of variation (CV) of triplicates was ≤20%, if not, outliers were excluded from further analysis. Reference dose response data were fitted using the four-parameter Hill function (Equation (1); GraphPad Prism®, GraphPad Software, La Jolla, USA); the acceptance criterium for the fit was: $R^2 \geq 0.98$).

$$Induction = Bottom + \frac{(Top - Bottom)}{1 + 10^{(LogEC50 - X) * Hill\ Slope}} \quad (1)$$

X = Log of dose or concentration
Top = Maximum response (fitted for the reference curve)
Bottom = Minimum response (fixed to the measurement of the solvent control)
$LogEC_{50}$ = Log of concentration at which 50% of the maximum response is observed
HillSlope = Slope factor

Induction data of the reference and test sample were then normalised using Equation (2), where response refers to the pertinent measured activity in the assay.

$$Induction[\%] = \frac{Response - Bottom}{Top - Bottom} \quad (2)$$

Next, dose response curves of the normalised data were fitted from 0 to 100%, where 0% referred to the response in the solvent control and 100% was the response maximum fitted for the reference E2. Limits of quantification (LOQ) were calculated as 10-fold the standard deviation (SD) of the averaged induction of the solvent control for each assay plate. MELN plate layout involved a reference on one plate and solvent controls split over two or three assay plates within a single series of analyses. For MELN, therefore, the average of all (six or nine) solvent controls within one series were used for the LOQ determination. The 10% effect level was then interpolated from the normalised reference dose-response curve to determine the PC concentration needed for 10% effect (PC_{10}). The 10% effect level was also interpolated from the normalised sample dose-response curve to determine the relative enrichment factor (REF, Equation (3)) needed to produce 10% effect (REF_{10}; OECD, 2015; see also Figure S1, Supplementary data). In case of the dose-response curve of the sample did not reach 10%, the REF required to reach the 10% effect level was extrapolated (Inglese et al., 2006; see also Figure S3b in Supplementary data).

$$REF = Concentration\ Factor_{SPE} \times Dilution\ Factor_{bioassay} \quad (3)$$

The estrogenic activity of the sample (EEQ_{sample}) was determined by dividing the PC_{10} by the REF_{10} (Equation (4)).

$$EEQ_{sample} = PC_{10}/REF_{10} \quad (4)$$

The EEQ concentrations determined in the ERTAs were then reported as either ng/mL (ethanol solutions) or ng/L (SPE extracts).

2.6. Tests for intra- and inter-day repeatability and apparent recovery

To test how each method performs within one laboratory we determined: (1) intra-day variability of ERTAs, (2) inter-day variability of ERTAs and (3) overall variability of ERTAs after SPE.

To assess “recoveries” (the yield of a preconcentration or extraction stage of an analytical process) and “apparent recoveries” (the quantity observed value/reference value, obtained using an analytical procedure that involves a calibration graph) (Burns et al., 2002) for the five ERTAs and the various sample types (see Table 1) we calculated ratios between measured and nominal EEQ concentrations. Nominal concentrations of the PC-samples were 6 ng E2 per mL ethanol. Expected nominal estrogenic activities were calculated for the Mixtures and $Mixtures_{SPE}$ using assay-specific E2-equivalence factors (EEF). These factors are compound specific and were derived by dividing the 50% effect concentration of the reference compound (E2) by the 50% effect concentration of $compound_i$ (Table 3). The expected nominal estrogenic activities for a mixture were calculated by adding up the nominal concentrations of each compound in the mixture (Table 1) multiplied with its EEF ($EEQ_{nominal} = \sum (Concentration_{compound} \times EEF_{compound})$). EEF values in Table 3 were provided by the participating laboratories performing the pertinent bioassay, except for the T47D-KBluc assay, where literature data were used (see Supplementary data).

3. Results

3.1. Intra- and inter-day variability for ERTAs

Intra-day variability of EEQ concentrations ranged between 4 and 30% for YES and ERα-CALUX and between 12 and 124% for MELN, GeneBLAzer-ERα and T47D-KBluc (Table 5, Figs. 1 and 2). Inter-day variability was below 50% for all assays except for MELN. An outlier test (Grubbs Test; GraphPad Software) pointed to two MELN data points and one GeneBLAzer-ERα data point as outliers (see Fig. 1). When these three results were removed, intra- and inter-day variability improved notably from 118 to 23% and from 136 to 86% for the MELN, and from 124 to 44% for the GeneBLAzer-ERα (Table 5, outliers are excluded in Fig. 2).

3.2. $ERTA_{SPE}$ variability

Across the three samples (PC-SPE_{E2}, Mixture-SPE_{low} and Mixture-SPE_{high}) tested, ERα-CALUX and YES had the lowest CVs (<30%), followed by MELN and GeneBLAzer-ERα (<50%) and T47D-KBluc (up to 153%; Table 5). The Grubbs Test pointed to one outlier in one T47D-KBluc dataset (see Fig. 1). With the outlier removed, all T47D-KBluc results had a CV below 60% (Table 6, Fig. 2). None of the blank SPE samples had EEQ concentrations above LOQ in any of the ERTAs.

3.3. Nominal versus measured EEQ concentrations

For the PC_{E2} samples in the intra-day variability experiment, mean EEQ results across the ERTAs were between 71 and 135% (Table 5) of the nominal value (i.e. 6 ng/mL). For the mixtures ($Mixture_{high}$ and $Mixture_{low}$) the situation is more complex. EEQ concentrations calculated for ERα-CALUX, MELN, T47D-KBluc were between 77 and 136% of nominal values (with one outlier removed; Table 5). However, YES and GeneBLAzer-ERα (without outlier) tended to underestimate the estrogenic activity of both mixtures (as low as ca. 40% of nominal values).

Results for the inter-day variability experiment showed a similar pattern across test systems. The average EEQ concentration for the PC_{E2} samples in four ERTAs was close to the nominal value (90%; range 82–101%), MELN had a much lower result of 42%. For the mixtures, ERα-CALUX, MELN and T47D-KBluc had average EEQ concentrations close to nominal values (106%, with one outlier removed). YES and GeneBLAzer-ERα again tended to underestimate the estrogenic activity of both mixtures (as low as 30% of nominal values).

For the variability of the ERTAs after SPE (i.e. $ERTA_{SPE}$), the average recovery of PC_{E2} across the five ERTAs was acceptable (86%), but the spread across the assays was very large (Table 6). Recovery for YES was too high (178%) and the T47D-KBluc result was much too low (merely 21%). Also results for mixtures diverged substantially, recoveries between 27 and 138%, with one extreme T47D-KBluc result (959%).

Fig. 3 shows that only ERα-CALUX had an excellent overall performance (recoveries between 96 and 101%). YES and GeneBLAzer-ERα performed reasonably well, with an overall average of 131% and 68% respectively.

4. Discussion

4.1. Repeatability of ERTAs suits regulatory monitoring of water quality status

Of the five ERTAs tested, ERα-CALUX performed best overall. The average CV of the nine samples tested for intra- and inter-day as well as $ERTA_{SPE}$ variability was 13%. YES was the second best performing ERTA, with an overall average CV of 22%. Both offer sound results, for example, when comparing this to a CV of 15–30% that is typically feasible and required for chemical analysis (Horwitz, 1982; European Commission, 2014). Furthermore, the overall average CV for YES is also much lower than requirements for chemical analysis and monitoring of water quality status outlined within the EU WFD (i.e. uncertainty of measurement of 50%; European Commission, 2009). T47D-KBluc and GeneBLAzer-ERα had moderate overall CVs (31 and 39%; with outliers excluded). MELN had the highest overall CV of tested ERTAs (54%; with outliers excluded). The fact

Table 5
Intra- and inter-day variability overview of five estrogen-receptor transactivation assays (ERTAs). Measured 17β-estradiol (E2) equivalent concentrations (EEQ; mean, SD and coefficient of variation (CV); n = 5) of the positive control (PC) as well as mixtures with high or low concentrations of (xeno)estrogens ($Mixture_{high}$ and $Mixture_{low}$) are compared with nominal EEQ concentrations.

	PC_{E2}					$Mixture_{high}$					$Mixture_{low}$				
	EEQ (ng/mL)	SD	CV (%)	$EEQ_{nominal}$ (ng/mL)	EEQ as % of $EEQ_{nominal}$	EEQ (ng/mL)	SD	CV (%)	$EEQ_{nominal}$ (ng/mL)	EEQ as % of $EEQ_{nominal}$	EEQ (ng/mL)	SD	CV (%)	$EEQ_{nominal}$ (ng/mL)	EEQ as % of $EEQ_{nominal}$
Intra-day															
YES	8.1	2.46	30%	6.0	135%	2.2	0.52	24%	3.2	67%	0.14	0.02	13%	0.32	44%
ERα-CALUX	6.1	0.46	8%	6.0	101%	1.8	0.07	4%	1.3	136%	0.10	0.02	20%	0.13	77%
MELN	4.4 [a,b]	2.59	59%	6.0	73%	1.5 [a,b]	0.34	23%	1.3	114%	0.11 [a,b]	0.06	53%	0.13	88%
including outlier						*3.1* [a,b]	*3.60*	*118%*	*1.3*	*240%*					
GeneBLAzer-ERα	4.3 [b]	1.07	25%	6.0	71%	1.5 [b]	0.99	66%	3.8	39%	0.37	0.16	44%	0.38	96%
including outlier											*0.81*	*1.00*	*124%*	*0.38*	*212%*
T47D-KBluc	5.4	1.67	31%	6.0	90%	7.1	0.82	12%	5.6	127% [c]	0.48	0.21	44%	0.56	86% [c]
Inter-day															
YES (n = 3)	6.1	0.49	8%	6.0	101%	3.4	0.41	12%	3.2	106%	0.15	0.07	49%	0.32	46%
ERα-CALUX	5.7	0.54	9%	6.0	95%	1.5	0.07	4%	1.3	119%	0.13	0.02	17%	0.13	99%
MELN	2.5 [a,b]	1.62	64%	6.0	42%	1.3 [a,b]	1.30	103%	1.3	98%	0.12 [a,b]	0.10	86%	0.13	93%
including outlier											*0.29* [a,b]	*0.40*	*136%*	*0.13*	*230%*
GeneBLAzer-ERα	4.9 [a,b]	1.43	29%	6.0	82%	1.3 [a]	0.45	33%	3.8	35%	0.11 [a,b]	0.04	36%	0.38	30%
T47D-KBluc	4.9	1.87	38%	6.0	82%	6.9	2.76	40%	5.6	123% [c]	0.59	0.14	24%	0.56	106% [c]

Outliers were identified and results are presented without these outliers or *with outliers included* (see text and Fig. 1 for further details).

[a] LOQ was larger than 10% effect for the majority of the replicates (see Table S1 and Figure S3 for full details).

[b] For the majority of the replicates the 10% effect level (PC10) had to be extrapolated either for the sample or the reference dose-response curve (see Table S1 and Figure S3 for full details).

[c] For T47D-KBluc, the relative potencies used to calculate the nominal EEQ concentration were based on literature data and not determined in the testing lab. Consequently, $EEQ_{nominal}$ is associated with high uncertainty and thus also the measured EEQ concentration as a percentage of $EEQ_{nominal}$.

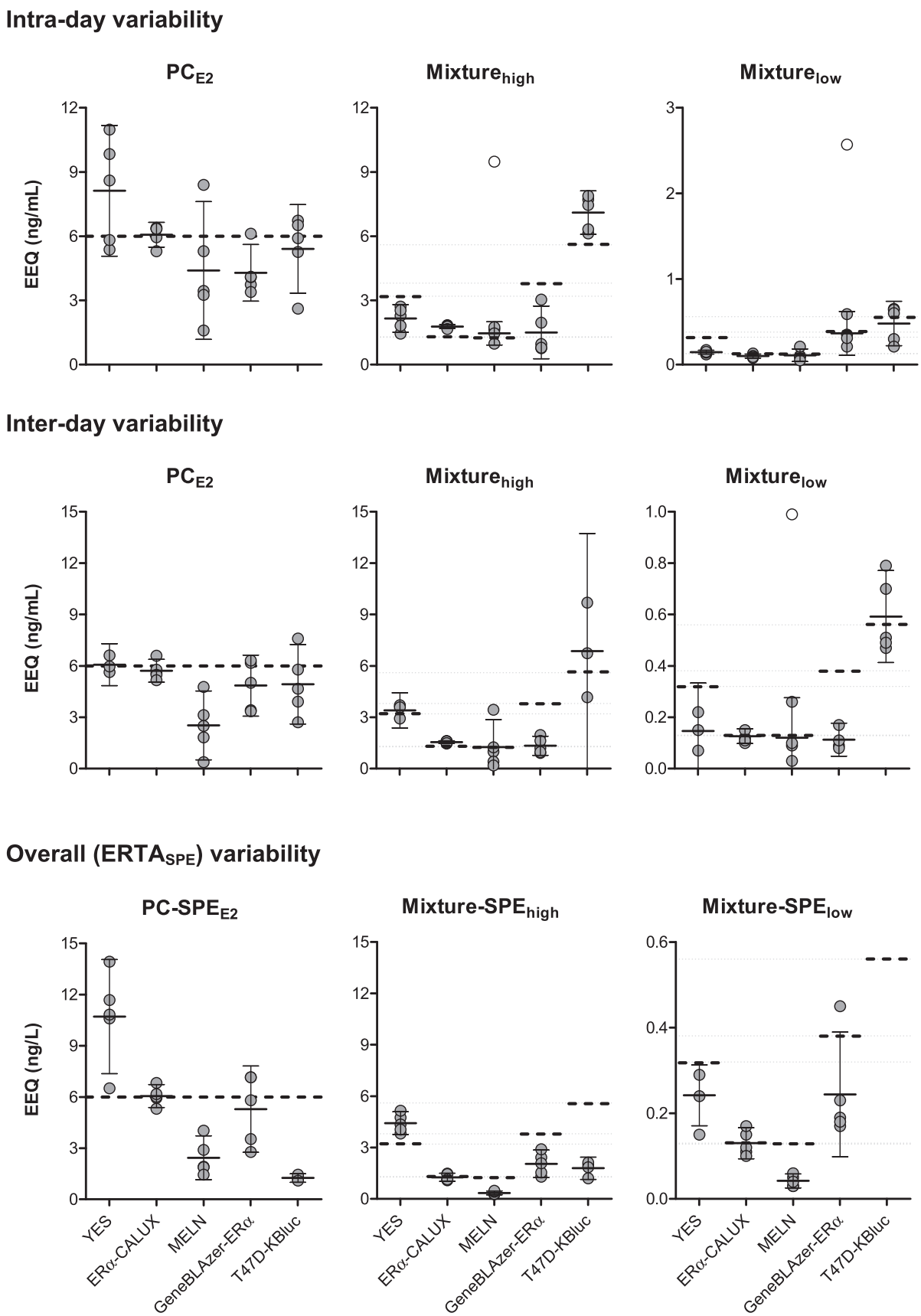


Fig. 1. Intra-day (top row), inter-day (middle row) and overall variability (bottom row) of 17β-estradiol (E2) equivalent (EEQ) concentrations (mean and 95% confidence interval) derived from repeated ERTA measurements (n = 5). Data are shown for the E2 positive control (PC_{E2}) and mixtures with high and low (xeno)estrogen concentrations that were tested directly (intra- and inter-day; in mL of ethanol) or after solid phase extraction ($ERTA_{SPE}$, overall; in L of water). Dashed lines indicate the expected EEQ concentrations for PC_{E2} and two mixtures. Outliers are indicated as open circles and are excluded for means and confidence intervals. The following T47D-KBluc data are outside y-axis ranges: the outlier for Mixture-SPE_{high} (21 ng/L) and the data for Mixture-SPE_{low} (between 2.3 and 9.9 ng/L).

that outliers occurred that could not be tied to obvious methodological errors is a clear indication that ERTAs require multiple biological replicates to produce a robust result. Taken together, although results across the ERTAs varied appreciably, results go on to show that these assays perform within limits that are acceptable for the regulatory monitoring of water quality status (e.g. uncertainty of measurement <50%; European Commission, 2009).

4.2. Robust EEF values are required to calculate nominal EEQ concentrations

When averaging all data (i.e. nine samples analysed with five ERTAs) measured EEQ concentrations matched nominal concentrations fairly well (mean = 83%, SD = 19%; n = 5 ERTAs, excluding the extreme T47D-KBluc result of 959%), meaning that results were both relatively accurate as well as fairly precise. However, some assays clearly performed better than others (Fig. 3). ERα-CALUX yielded an excellent match between measured and nominal data across all nine samples (i.e. data are close to the 100% line). YES also performed well on average, but YES data scattered more than ERα-CALUX data. MELN and GeneBLAzer-ERα tended to underestimate nominal EEQ concentrations. T47D-KBluc performed well for the mixture samples but very poorly with respect to the Mixture-SPE samples (see Table 6).

When breaking down the total data set, the pairs of measured and nominal EEQ concentrations that are easiest to interpret are those of PC_{E2} samples tested for intra- and inter-day comparisons. This because the compound tested in PC_{E2} sample corresponds 1-to-1 with the E2 reference used in all ERTAs and results can be evaluated in a straightforward manner in terms of accuracy (i.e. the match between measured and nominal concentrations). Accuracy was excellent for four assays (average 95%), only MELN showed an appreciable deviation from 100% (i.e. 58%). A larger mismatch between PC_{E2} $EEQ_{measured}$ and PC_{E2} $EEQ_{nominal}$ may be tied to an inconsistency in the concentration of the reference compound used in the testing lab or an inconsistency in redissolving the PC_{E2} sample at the testing lab (samples were shipped dry). To exclude this possible source of error, an aliquot of the same reference solution can be sent to all participating labs and samples can be sent in solution rather than dry. In fact, this approach was adopted for a recent ISO ring test of the YES (conducted at the end of 2015).

Pairs of measured and nominal concentrations that are more difficult to interpret involve the mixture samples. The main reason for this is that the nominal EEQ concentrations for the mixtures are calculated using EEFs and these EEFs are associated with uncertainty. Thus it is not possible to calculate an accurate nominal EEQ concentration. A good example for this is the EEF for E1 in the T47D-KBluc. We found disparate published T47D-KBluc EEF values for E1, (i.e. 0.02; 0.1; 0.61; 1.4 see Supplementary data for full details). These differences between studies may be caused by metabolism of E1 (Bovee and Pikkemaat, 2009; Hoogenboom et al., 2001) that may vary between labs. Anyway, it is clear that the inclusion or rejection of one or more “disparate” EEF values to calculate an average EEF (we used 0.53) will dramatically affect the calculation of nominal EEQ concentrations. In turn, this affects the match (or mismatch) between measured and nominal EEQ concentrations when we evaluate our results in terms of accuracy.

EEF uncertainty, or inconsistencies with concentrations in the redissolved samples, are not relevant for all compounds and all assays. EEFs of BPA are so low – for all assays – that BPA constitutes less than 2% of nominal EEQ concentrations. Consequently, a two-fold error in BPA concentration would only cause a 2% effect on $EEQ_{nominal}$ and even a five-fold change in the EEF of BPA would only produce less than 10% effect on $EEQ_{nominal}$. Estrone EEF values are low for ERα-CALUX and MELN, thus the mixture is dominated by E2 (see Table 4). Consequently, ERα-CALUX and MELN results are expected to be less influenced by uncertainty related to either the accuracy of the E1 concentration in the sample or the accuracy of the assay specific E1 EEF. This “E2 dominance” may have contributed to the good match that we observed between measured and nominal concentrations in ERα-CALUX (108%, average of intra- and inter-day mixture samples) and MELN (98%). The situation is different for YES (66%) and GeneBLAzer-ERα (50%) where E1 appears to dominate the mixture. For T47D-KBluc (111%), E1 and E2 have an almost even contribution in the mixture. Consequently, for three ERTAs, slight errors in E1 concentration or E1 EEF will directly and appreciably impact the match between measured and nominal EEQ concentrations.

As discussed above, E1 EEFs are uncertain for T47D-KBluc but this is also the case for YES. A review by (Jarošová et al., 2014) lists 10 YES E1 EEF values between 0.1 and 0.68. We have observed similar variability in E1 EEF values in the YES (between 0.05 and 0.41 n = 9 tests performed in one lab over a two year period). Consequently, to reduce uncertainty with respect to EEFs and the calculation of nominal EEQ concentrations of mixtures, it is advisable to determine the EEFs within the experiment. Keeping in mind that the sensitivity to different compounds in a bioassay is not necessarily stable over time, the OECD guideline for testing estrogen receptor agonists and antagonists advises to always test a weak or partial agonist along with E2: “to help ensure proper functioning of the assay” (OECD, 2015).

4.3. Sample enrichment – no obvious source for additional variability

Besides investigating ERTA intra- and inter-day variability, another aim was to address possible additional variability caused by SPE. Sample enrichment is required when applying ERTAs to measure estrogenic activity in surface waters at environmentally relevant levels (i.e. <1 ng/L). Overall intra-day variability for PC_{E2} and Mixture samples was 30% (n = 15; five ERTAs and three samples), overall intra-day variability for the three SPE samples was 28%. One reason for the absence of additional variability caused by SPE samples may be that the SPE samples were already pools of five individual SPEs. Consequently, our experimental design will have evened out some of the variability that is inherent to SPE.

Again, ERα-CALUX results showed good accuracy for the three tested samples (i.e. results around 100%, Fig. 3). This indicates a robust recovery of steroidal estrogens with this particular SPE and supports earlier observations with the same method (Escher et al., 2008b; Kienle et al., 2015). However, with a maximum recovery for PC-SPE_{E2} and YES of 178% and a minimum recovery of 21% in T47D-KBluc, data from the other four ERTAs were less accurate than ERα-CALUX data (one extreme T47D-KBluc result of 959% could not be resolved, i.e. no obvious method errors could be uncovered). Although we expected recovery bias to be introduced by the SPE step, this excessive recovery variability seen for some ERTAs does not match the ERα-CALUX results (very close to 100%), previous published results on SPE extraction efficiency (Escher et al., 2008b; Kienle et al., 2015) and unpublished SPE validation studies. We conclude that the high recovery variability seen in SPE samples must be linked to the handling of the samples in the testing labs. For example, although good accuracy was achieved for T47D-KBluc and intra- and inter-day PC_{E2} samples (respective measured concentrations at 90 and 82% of nominal concentrations; Table 5), low accuracy was observed for the PC-SPE_{E2} sample (21%; Table 6). As the PC-SPE_{E2} sample for ERα-CALUX (accurate at 101%) and T47D-KBluc (not accurate at 21%) came from the same pool and were evaluated in the same testing lab, sample handling (e.g. redissolving or storage) in the testing lab seems a plausible source for error.

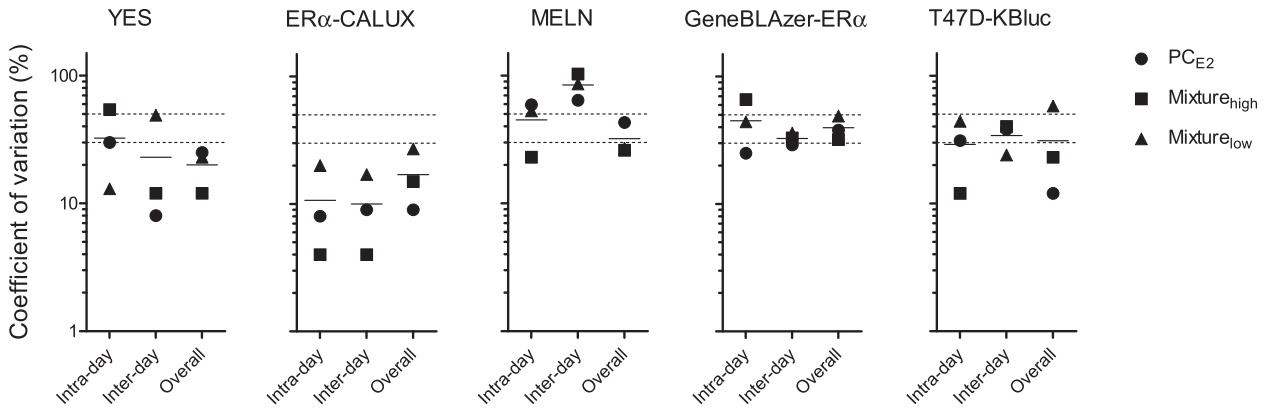


Fig. 2. Coefficient of variation (CV) of 17β-estradiol (E2) equivalent (EEQ) concentrations in the three experiments (intra-day; inter-day; overall, ERTA$_{SPE}$). Data are shown for three samples types (E2 positive control, PC$_{E2}$, and mixtures with high and low (xeno)estrogen concentrations, Mixture$_{high}$ and Mixture$_{low}$) tested in five ERTAs. CV was calculated without outliers (see Table 5). Dashed lines indicate CV levels of 30 and 50%.

Table 6
Overall ERTA$_{SPE}$ variability and recovery. Measured 17β-estradiol (E2) equivalent concentrations (EEQ; mean, SD and coefficient of variation (CV); n = 5) of the positive control (PC) as well as mixtures with high or low concentrations of (xeno)estrogens (Mixture$_{high}$ and Mixture$_{low}$) are compared with calculated nominal EEQ concentrations.

	PC-SPE$_{E2}$					Mixture-SPE$_{high}$					Mixture-SPE$_{low}$				
	EEQ (ng/L)	SD	CV (%)	EEQ$_{nominal}$ (ng/L)	EEQ as % of EEQ$_{nominal}$	EEQ (ng/L)	SD	CV (%)	EEQ$_{nominal}$ (ng/L)	EEQ as % of EEQ$_{nominal}$	EEQ (ng/L)	SD	CV (%)	EEQ$_{nominal}$ (ng/L)	EEQ as % of EEQ$_{nominal}$
Intra-day															
YES	10.7	2.69	25%	6.0	178%	4.4	0.53	12%	3.2	138%	0.24	0.06	23%	0.32	76%
ERα-CALUX	6.1	0.54	9%	6.0	101%	1.3	0.19	15%	1.3	96%	0.13	0.03	27%	0.13	98%
MELN	2.4 [a]	1.04	43%	6.0	41%	0.4 [a]	0.09	26%	1.3	27%	0.04 [a]	0.01	27%	0.13	34%
GeneBLAzer-ERα	5.3 [a,b]	2.03	38%	6.0	88%	2.1 [a,b]	0.65	32%	3.8	54%	0.24 [a]	0.12	49%	0.38	64%
T47D-KBluc	1.2	0.15	12%	6.0	21%	1.8	0.41	23%	5.6	32% [c]	5.37	3.11	58%	0.56	959% [c]
including outlier						*5.6*	*8.61*	*153%*	*5.6*	*101%*					

Outliers were identified and results are presented without these outliers or *with outliers included* (see text and Fig. 1 for further details).
[a] LOQ was larger than 10% effect for the majority of the replicates (see Table S1 and Figure S3 for full details).
[b] For the majority of the replicates the 10% effect level (PC10) had to be extrapolated either for the sample or the reference dose-response curve (see Table S1 and Figure S3 for full details).
[c] For T47D-KBluc, the relative potencies used to calculate the nominal EEQ concentrations were based on literature data and not determined in the testing lab. Consequently, EEQ$_{nominal}$ is associated with high uncertainty and thus also the measured EEQ concentration as a percentage of EEQ$_{nominal}$. This extreme T47D-KBluc result of 959% could not be resolved, no errors were found in calculations.

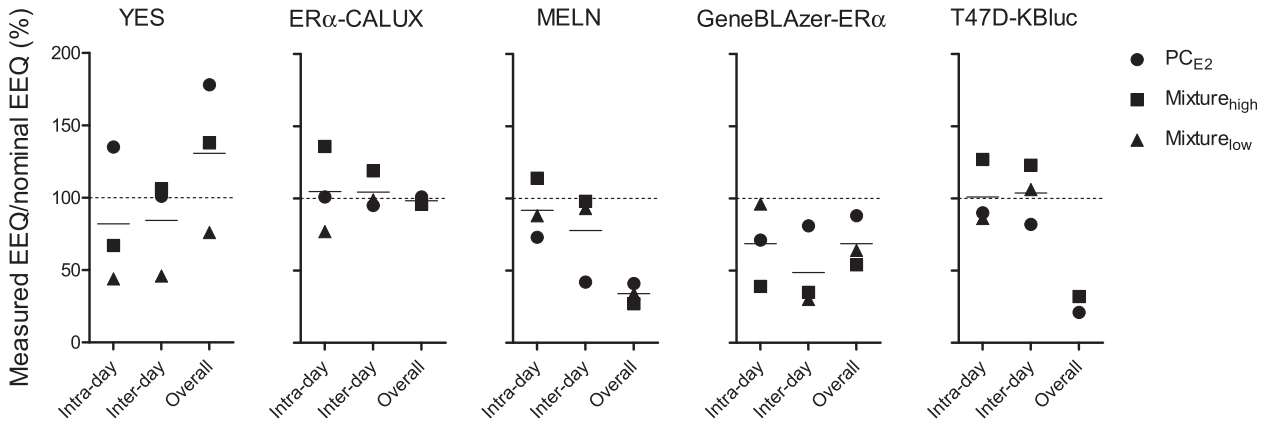


Fig. 3. Ratios of the average measured 17β-estradiol (E2) equivalent (EEQ) concentration and the nominal EEQ concentrations for three samples types (E2 positive control, PC$_{E2}$, and mixtures with high and low (xeno)estrogen concentrations, Mixture$_{high}$ and Mixture$_{low}$; solid lines indicate the average for three samples) in five ERTAs. Data are shown for all three experiments (intra-day; inter-day; overall, SPE and ERTA).

4.4. PC10 EEQ concentration derivation – a uniform method across samples and ERTAs

ERTA results will inherently vary due to the many laboratory steps that are involved to reach a result. One of these steps that will contribute to the variability of the EEQ concentration is the way EEQ concentrations are derived. We selected a PC10 approach (OECD, 2015; see also Escher et al. (2013) for a non-logistic PC10

approach) to determine EEQ concentrations mainly because the method allows for an unbiased comparison between different samples and different bioassays. The method is particularly advantageous when dealing with environmental samples where: 1) samples do not induce full effects (due to low contamination [e.g. effect levels in surface water samples often do not reach 50% even using high concentration factors] or cytotoxicity at high concentrations or the presence of ER-antagonists); 2) effects observed in samples may be associated and influenced by effects on growth parameters in the assay (e.g. growth stimulation or cytotoxicity); and 3) slopes between the reference and samples are often not parallel, possibly due to toxicokinetic and/or toxicodynamic effects (i.e. how different compounds are being taken up by cells and how compounds interact at the receptor). The impact of the last two aspects (cytotoxicity and non-parallelism) is minimised when testing lower sample doses that are associated with lower effect levels (e.g. a 10% effect level).

Generally, instead of a PC10 evaluation, the observed effect can be interpolated at the tested sample doses from the E2 reference curve. To calculate a mean interpolated EEQ concentration, one dilution per tested sample or several dilutions from a sample dose-response curve are interpolated. For all of the mentioned approaches there is a risk to create bias in the derived EEQ concentrations if data selection and analysis are not bound to specific criteria and are not harmonised (e.g. by arbitrary selection of dilutions to be tested or used for EEQ mean derivation). With regard to weakly active samples that only allow for a single or a few interpolations at low effect levels, there is not much difference between the PC10 or EEQ mean approaches. For samples with higher activity, more interpolated data can be produced, including data from higher effect levels, possibly covering sections of dose-response curves where non-parallelism may occur (see examples in Supplementary data).

The five ERTAs were performed according to standard procedures of the labs involved. Consequently, this caused certain constraints in terms of a harmonised analysis, as the PC10 approach requires specific conditions to be met: i.e. LOQ below 10% effect and dose-response data that cover the 10% effect range. These conditions were not always met for all samples and all ERTAs (see labels in Tables 5 and 6 and Table S1 and Figure S3 for full details). Particularly the issue of having to extrapolate to the 10% effect level (mainly because samples were not always tested at high enough dilutions) can cause uncertainty. To evaluate possible bias caused by the PC10 evaluation method we also processed all our raw data to calculate mean interpolated EEQ concentrations. With the exception of one sample, Mixture-SPE$_{low}$ in the T47D-KBluc, only minor differences were observed between the two approaches (see Supplementary data, Figure S4), supporting the use of the PC10 approach as a straightforward and faster method for EEQ derivation.

The two conditions that a PC10 approach implies (LOQ<10% and effect data that cover the 10% range) were not always met. Both issues do not constitute a prohibitive hurdle for data analysis however. The LOQ determined as 10-fold the standard deviation is very conservative (e.g. two-fold the standard deviation was used in Mehinto et al. (2015)). LOQs of the MELN data were typically between 15 and 30%. As for MELN, a single reference dose-response curve was used for the data evaluation of two to three plates, we averaged control data from those two or three assay plates (three controls per plate) to calculate standard deviations and LOQs. Such pooling of control data across plates leads to higher control standard deviations (and thus LOQs) when compared to the calculation of LOQs per individual plate. When MELN LOQs were calculated based on an individual plate basis, LOQs ranged between 6 and 15% (instead of 15–30%). In the GeneBLAzer-ERα and T47D-KBluc, samples were often not diluted to levels below 10% effect. Also this issue is easy to resolve by testing more diluted samples. Alternatively, the conditions can be satisfied by deriving EEQ concentrations at higher effect levels, e.g. PC20 (Creusot et al., 2013).

4.5. Further standardisation as a route to apply ERTA in regulatory monitoring

Although the overall variability was good and measured EEQ concentrations generally matched nominal EEQ concentrations, several ERTA elements would benefit from further standardisation. First, the use of the same (e.g. certified) standard by all participating labs will contribute to a more reproducible result. Second, beside the reference E2 it is advisable to also test a standard of the most dominant compound(s) in the environmental sample. For typical surface waters affected by domestic effluent this would be E1 and possibly EE2. In this way, there is an internal experimental control on the EEF of the most important (known) compounds in the mixture. Third, a better characterisation of EEFs of estrogenic compounds that can contribute to the effect is advisable. For as yet unknown reasons, the spread in published EEF data for the same ERTA can be very large (i.e. > factor 10; see also Alvarez et al. (2013)). Fourth, the ERTA should include sufficient elements that allow for a robust set of validity criteria. ERTA data should include a full reference dose-response curve (below 10% and above 90%) and have sufficient values at the lower effect levels (e.g. <10% effect) for both reference and samples. Sufficient solvent controls should be available per plate to allow for a robust LOQ determination that is below the PC10 (or e.g. PC20) level. Fifth, a harmonised EEQ concentration derivation method will contribute to more uniform results generated with the same set of raw data. A harmonised evaluation method will also be critical when considering linking the outcome of an ERTA to EEQ-based trigger values in regulatory monitoring. For example, this will be important when applying ERTAs to support WFD monitoring for watch list chemicals such as E2 or EE2.

One important aspect that needs consideration is that the different bioassays will produce different EEQ concentrations from the same environmental mixture. The reason being assay specific differences in sensitivity to the individual compounds (as reflected by the different EEFs). For example, an E1 dominated mixture tested with T47D-KBluc will lead to higher EEQ concentrations than the same mixture tested with YES (Alvarez et al., 2013). This needs to be considered when applying ERTAs in regulatory monitoring. When deciding regulatory binding trigger values for ERTAs, these values need to be assay specific (Jarošová et al., 2014).

5. Conclusions

- ERTA repeatability is good to excellent for ERα-CALUX and acceptable for YES
- ERTA accuracy is generally good to excellent for ERα-CALUX
- When ERTAs are coupled with sample enrichment the ability to determine EEQ concentrations at or below proposed water quality standards is straightforward

Consequently, ERTAs are suitable to satisfy criteria for regulatory monitoring, especially ERα-CALUX performs well.

- Accuracy may be improved by sending E2 reference (e.g. certified standards) and samples in solutions (labs can weigh vials and report the weight to monitor loss through evaporation)
- Better EEFs are needed to permit an accurate calculation of EEQ concentrations of mixtures, EEFs have to be established within the experiment itself

- To facilitate the PC10 analysis and thus improve repeatability, accuracy and precision, the low end of the reference and sample dose response curves require sufficient resolution

Acknowledgements

This research was primarily sponsored by the Swiss Centre for Applied Ecotoxicology (Dübendorf/Lausanne, Switzerland) with partial funding provided by the European Union's Seventh Framework Programme (grant agreement no. 308339, project DEMEAU "Demonstration of promising technologies to address emerging pollutants in water and wastewater"). The analysis performed at INERIS was partly supported by the ONEMA (AQUAREF programme). Nadzeya Homazava assisted with the preparation of spike solutions and SPE samples.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.watres.2016.10.062.

References

Aherne, G.W., English, J., Marks, V., 1985. The role of immunoassay in the analysis of microcontaminants in water samples. Ecotoxicol. Environ. Saf. 9 (1), 79–83.
Alvarez, D.A., Shappell, N.W., Billey, L.O., Bermudez, D.S., Wilson, V.S., Kolpin, D.W., Perkins, S.D., Evans, N., Foreman, W.T., Gray, J.L., Shipitalo, M.J., Meyer, M.T., 2013. Bioassay of estrogenicity and chemical analyses of estrogens in streams across the United States associated with livestock operations. Water Res. 47 (10), 3347–3363.
ARCEM, 2003. Hormonwirksame Stoffe in Österreichs Gewässer – ein Risiko? Umweltbundesamt GmbH, Wien, p. 462.
Balaguer, P., Francois, F., Comunale, F., Fenet, H., Boussioux, A.M., Pons, M., Nicolas, J.C., Casellas, C., 1999. Reporter cell lines to study the estrogenic effects of xenoestrogens. Sci. Total Environ. 233 (1–3), 47–56.
Bovee, T.F.H., Schoonen, W.G.E.J., Hamers, A.R.M., Jorge Bento, M., Peijnenburg, A.A.C.M., 2008. Screening of synthetic and plant-derived compounds for (anti)estrogenic and (anti)androgenic activities. Anal. Bioanal. Chem. 390, 1111–1119.
Bovee, T.F.H., Pikkemaat, M.G., 2009. Bioactivity-based screening of antibiotics and hormones. J. Chrom. A 1216, 8035–8050.
Burns, D.T., Danzer, K., Townshend, A., 2002. Use of the terms "recovery" and "apperent recovery" in analytical procedures. Pure Appl. Chem. 74 (11), 2201–2205.
Creusot, N., Budzinski, H., Balaguer, P., Kinani, S., Porcher, J.-M., Aït-Aïssa, S., 2013. Effect-directed analysis of endocrine-disrupting compounds in multi-contaminated sediment: identification of novel ligands of estrogen and pregnane X receptors. Anal. Bioanal. Chem. 405 (8), 2553–2566.
Desbrow, C., Routledge, E.J., Brighty, G.C., Sumpter, J.P., Waldock, M., 1998. Identification of estrogenic chemicals in STW effluent. 1. Chemical fractionation and in vitro biological screening. Environ. Sci. Technol. 32 (11), 1549–1558.
Escher, B.I., Allinson, M., Altenburger, R., Bain, P.A., Balaguer, P., Busch, W., Crago, J., Denslow, N.D., Dopp, E., Hilscherova, K., Humpage, A.R., Kumar, A., Grimaldi, M., Jayasinghe, B.S., Jarosova, B., Jia, A., Makarov, S., Maruya, K.A., Medvedev, A., Mehinto, A.C., Mendez, J.E., Poulsen, A., Prochazka, E., Richard, J., Schifferli, A., Schlenk, D., Scholz, S., Shiraishi, F., Snyder, S., Su, G., Tang, J.Y.M., Burg, B.v.d., Linden, S.C.v.d., Werner, I., Westerheide, S.D., Wong, C.K.C., Yang, M., Yeung, B.H.Y., Zhang, X., Leusch, F.D.L., 2013. Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays. Environ. Sci. Technol. 48 (3), 1940–1956.
Escher, B.I., Bramaz, N., Mueller, J.F., Quayle, P., Rutishauser, S., Vermeirssen, E.L.M., 2008a. Toxic equivalent concentrations (TEQs) for baseline toxicity and specific modes of action as a tool to improve interpretation of ecotoxicity testing of environmental samples. J. Environ. Monit. 10 (5), 612–621.
Escher, B.I., Bramaz, N., Quayle, P., Rutishauser, S., Vermeirssen, E.L.M., 2008b. Monitoring of the ecotoxicological hazard potential by polar organic micropollutants in sewage treatment plants and surface waters using a mode-of-action based test battery. J. Environ. Monit. 10, 622–631.
Escher, B.I., Leusch, F., 2012. Bioanalytical Tools in Water Quality Assessment. IWA Publishing, London.
European Commission, 2009. 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.
European Commission, 2013. Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 Amending Directives 2000/60/EC and 2008/105/EC as Regards Priority Substances in the Field of Water Policy.
European Commission, 2014. COMMISSION REGULATION (EU) 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. Official Journal of the European Union L 164/18.
Fernandez, M.P., Noguerol, T.-N., Lacorte, S., Buchanan, I., Piña, B., 2009. Toxicity identification fractionation of environmental estrogens in waste water and sludge using gas and liquid chromatography coupled to mass spectrometry and recombinant yeast assay. Anal. Bioanal. Chem. 393, 957–968.
Folmar, L.C., Denslow, N.D., Rao, V., Chow, M., Crain, D.A., Enblom, J., Marcino, J., Guillette Jr., L.J., 1996. Vitellogenin induction and reduced serum testosterone concentrations in feral male carp (Cyprinus carpio) captured near a major metropolitan sewage treatment plant. Environ. Health Perspect. 104 (10), 1096–1101.
Hoogenboom, L.A.P., De Haan, L., Hooijerink, D., Bor, G., Murk, A.J., Brouwer, A., 2001. Estrogenic activity of estradiol and its metabolites in the ER-CALUX assay with human T47D breast cells. APMIS 109, 101–107.
Horwitz, W., 1982. Evaluation of analytical methods used for regulation of foods and drugs. Anal. Chem. 54 (1).
Huang, R., Xia, M., Cho, M.-H., Sakamuru, S., Shinn, P., Houck, K.A., Dix, D.J., Judson, R.S., Witt, K.L., Kavlock, R.J., Tice, R.R., Austin, C.P., 2011. Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. Environ. Health Perspect. 119 (8), 1142–1148.
Ihara, M., Ihara, M.O., Kumar, V., Narumiya, M., Hanamoto, S., Nakada, N., Yamashita, N., Miyagawa, S., Iguchi, T., Tanaka, H., 2014. Co-occurrence of estrogenic and antiestrogenic activities in wastewater: quantitative evaluation of balance by in vitro ERα reporter gene assay and chemical analysis. Environ. Sci. Technol. 48 (11), 6366–6373.
Inglese, J., Auld, D.S., Jadhav, A., Johnson, R.L., Simeonov, A., Yasgar, A., Zheng, W., Austin, C.P., 2006. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc. Acad. Nat. Sci. 103, 11473–11478.
Jarošová, B., Bláha, L., Giesy, J.P., Hilscherová, K., 2014. What level of estrogenic activity determined by in vitro assays in municipal waste waters can be considered as safe? Environ. Int. 64 (0), 98–109.
Johnson, A.C., Dumont, E., Williams, R.J., Oldenkamp, R., Cisowska, I., Sumpter, J.P., 2013. Do concentrations of ethinylestradiol, estradiol, and diclofenac in European rivers exceed proposed EU environmental quality standards? Environ. Sci. Technol. 47 (21), 12297–12304.
Kienle, C., Vermeirssen, E., Kunz, P., Werner, I., 2015. Grobbeurteilung der Wasserqualität von abwasserbelasteten Gewässern anhand von ökotoxikologischen Biotests. Studie im Auftrag des BAFU. Schweizerisches Zentrum für angewandte Ökotoxikologie Eawag-EPFL, Dübendorf, p. 61.
Kunz, P.Y., Kienle, C., Carere, M., Homazava, N., Kase, R., 2015. In vitro bioassays to screen for endocrine active pharmaceuticals in surface and waste waters. J. Pharm. Biomed. Analysis 106, 107–115.
Leusch, F.D.L., de Jager, C., Levi, Y., Lim, R., Puijker, L., Sacher, F., Tremblay, L.A., Wilson, V.S., Chapman, H.F., 2010. Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters. Environ. Sci. Technol. 44 (10), 3853–3860.
Loos, R., 2012. Analytical Methods for the New Proposed Priority Substances of the European Water Framework Directive (WFD).
López de Alda, M.J., Barceló, D., 2001. Use of solid-phase extraction in various of its modalities for sample preparation in the determination of estrogens and progestogens in sediment and water. J. Chromatogr. A 938 (1–2), 145–153.
Maletz, S., Floehr, T., Beier, S., Klümper, C., Brouwer, A., Behnisch, P., Higley, E., Giesy, J.P., Hecker, M., Gebhardt, W., Linnemann, V., Pinnekamp, J., Hollert, H., 2013. In vitro characterization of the effectiveness of enhanced sewage treatment processes to eliminate endocrine activity of hospital effluents. Water Res. 47 (4), 1545–1557.
Margot, J., Kienle, C., Magnet, A., Weil, M., Rossi, L., de Alencastro, L.F., Abegglen, C., Thonney, D., Chèvre, N., Schärer, M., Barry, D.A., 2013. Treatment of micropollutants in municipal wastewater: ozone or powdered activated carbon? Sci. Total Environ. 1 (461–462), 480–498.
McDonnell, D.P., Nawaz, Z., Densmore, C., Weigel, N.L., Pham, T.A., Clark, J.H., O'Malley, B.W., 1991. High level expression of biologically active estrogen receptor in Saccharomyces cerevisiae. J. Steroid Biochem. Mol. Biol. 39 (3), 291–297.
Mehinto, A.C., Jia, A., Snyder, S.A., Jayasinghe, B.S., Denslow, N.D., Crago, J., Schlenk, D., Menzie, C., Westerheide, S.D., Leusch, F.D.L., Maruya, K.A., 2015. Interlaboratory comparison of in vitro bioassays for screening of endocrine active chemicals in recycled water. Water Res. 83, 303–309.
OECD, 2015. OECD Guideline for the Testing of Chemicals Draft Proposal for an Updated TG 455 Performance-based Test Guideline for Stably Transfected Transcriptional Activation in Vitro Assays to Detect Estrogenic Agonist Activity of Chemicals.
Purdom, C.E., Hardiman, P.A., Bye, V.V.J., Eno, N.C., Tyler, C.R., Sumpter, J.P., 1994. Estrogenic effects of effluents from sewage treatment works. Chem. Ecol. 8 (4), 275–285.
Routledge, E.J., Sumpter, J.P., 1996. Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. Environ. Toxicol. Chem. 15 (3), 241–248.
Schindler Wildhaber, Y., Mestankova, H., Schärer, M., Schirmer, K., Salhi, E., von Gunten, U., 2015. Novel test procedure to evaluate the treatability of wastewater with ozone. Water Res. 75 (0), 324–335.
Soto, A.M., Sonnenschein, C., Chung, K.L., Fernandez, M.F., Olea, N., Serrano, F.O.,

1995. The E-Screen assay as a tool to identify estrogens - an update on estrogenic environmental-pollutants. Environ. Health Perspect. 103, 113–122.
Sumpter, J.P., Johnson, A.C., 2008. 10th Anniversary Perspective: reflections on endocrine disruption in the aquatic environment: from known knowns to unknown unknowns (and many things in between). J. Environ. Monit. 10 (12), 1476–1485.
Tyler, C.R., Jobling, S., Sumpter, J.P., 1998. Endocrine disruption in wildlife: a critical review of the evidence. Crit. Rev. Toxicol. 28 (4), 319–361.
van der Linden, S.C., Heringa, M.B., Man, H.-Y., Sonneveld, E., Puijker, L.M., Brouwer, A., van der Burg, B., 2008. Detection of multiple hormonal activities in wastewater effluents and surface water, using a panel of steroid receptor CALUX bioassays. Environ. Sci. Technol. 42 (15), 5814–5820.
Vermeirssen, E.L.M., Eggen, R.I.L., Escher, B.I., Suter, M.J.-F., 2008. Estrogens in Swiss rivers and effluents - sampling matters. Chimia 62 (5), 389–394.
Vethaak, A.D., Rijs, G.B.J., Schrap, S.M., Ruiter, H., Gerritsen, A., Lahr, J., 2002. Estrogens and Xeno-estrogens in the Aquatic Environment of the Netherlands: Occurrence, Potency and Biological Effects. RIZA/RIKZ, Lelystad, p. 292.
Vethaak, A.D., Lahr, J., Schrap, S.M., Belfroid, A.C., Rijs, G.B.J., Gerritsen, A., de Boer, J., Bulder, A.S.G., Grinwis, G.C.M., Kuiper, R.V., Legler, J., Murk, T.A.J., Peijnenburg, W., Verhaar, H.J.M., de Voogt, P., 2005. An integrated assessment of estrogenic contamination and biological effects in the aquatic environment of The Netherlands. Chemosphere 59, 511–524.
Villeneuve, D.L., Blankenship, A.L., Giesy, J.P., 2000. Derivation and application of relative potency estimates based on in vitro bioassay results. Environ. Toxicol. Chem. 19 (11), 2835–2843.
Wagner, M., Vermeirssen, E.L.M., Buchinger, S., Behr, M., Magdeburg, A., Oehlmann, J., 2013. Deriving bio-equivalents from in vitro bioassays: assessment of existing uncertainties and strategies to improve accuracy and reporting. Environ. Toxicol. Chem. 32 (8), 1906–1917.
Wernersson, A.-S., Carere, M., Maggi, C., Tusil, P., Soldan, P., James, A., Sanchez, W., Dulio, V., Broeg, K., Reifferscheid, G., Buchinger, S., Maas, H., Van Der Grinten, E., O'Toole, S., Ausili, A., Manfra, L., Marziali, L., Polesello, S., Lacchetti, I., Mancini, L., Lilja, K., Linderoth, M., Lundeberg, T., Fjallborg, B., Porsbring, T., Larsson, D., Bengtsson-Palme, J., Forlin, L., Kienle, C., Kunz, P., Vermeirssen, E., Werner, I., Robinson, C.D., Lyons, B., Katsiadaki, I., Whalley, C., den Haan, K., Messiaen, M., Clayton, H., Lettieri, T., Negrão Carvalho, R., Gawlik, B.M., Hollert, H., Di Paolo, C., Brack, W., Kammann, U., Kase, R., 2015. The European technical report on aquatic effect-based monitoring tools under the Water Framework Directive. Environ. Sci. Eur. 27 (1), 7.
Williams, R.J., Churchley, J.H., Kanda, R., Johnson, A.C., 2012. Comparing predicted against measured steroid estrogen concentrations and the associated risk in two United Kingdom river catchments. Environ. Toxicol. Chem. 31 (4), 892–898.
Wilson, V.S., Bobseine, K., Gray, L.E., 2004. Development and characterization of a cell line that stably expresses an estrogen-responsive luciferase reporter for the detection of estrogen receptor agonist and antagonists. Toxicol. Sci. 81 (1), 69–77.
Ying, G.G., Kookana, R.S., Ru, Y.J., 2002. Occurrence and fate of hormone steroids in the environment. Environ. Int. 28 (6), 545–551.
Zacharewski, T.R., 1997. In vitro bioassays for assessing estrogenic substances. Environ. Sci. Technol. 31 (3), 613–623.

ANNEXE B : KÖNNEMAN ET AL (2018) EFFECT-BASED AND CHEMICAL ANALYTICAL METHODS TO MONITOR ESTROGENS UNDER THE EUROPEAN WATER FRAMEWORK DIRECTIVE. *TRENDS IN ANALYTICAL CHEMISTRY* (11 PAGES)

Trends in Analytical Chemistry 102 (2018) 225–235

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Effect-based and chemical analytical methods to monitor estrogens under the European Water Framework Directive

Sarah Könemann [a, b, *, 1], Robert Kase [b, 1], Eszter Simon [b], Kees Swart [c], Sebastian Buchinger [d], Michael Schlüsener [d], Henner Hollert [a], Beate I. Escher [e, f], Inge Werner [b], Selim Aït-Aïssa [g], Etienne Vermeirssen [b], Valeria Dulio [g], Sara Valsecchi [h], Stefano Polesello [h], Peter Behnisch [c], Barbora Javurkova [i], Olivier Perceval [k], Carolina Di Paolo [a], Daniel Olbrich [b], Eliska Sychrova [i], Rita Schlichting [e], Lomig Leborgne [l], Manfred Clara [m], Christoph Scheffknecht [n], Yves Marneffe [o], Carole Chalon [o], Petr Tušil [p], Přemysl Soldàn [p], Brigitte von Danwitz [q], Julia Schwaiger [r], Maria Isabel San Martín Becares [s], Francesca Bersani [t], Klara Hilscherová [i], Georg Reifferscheid [d], Thomas Ternes [d], Mario Carere [u]

[a] Institute for Environmental Research, RWTH Aachen University, Worringerweg 1, 52074 Aachen, DE, Germany
[b] Swiss Centre for Applied Ecotoxicology Eawag-EPFL, Überlandstrasse 131, 8600 Dübendorf, CH, Switzerland
[c] BioDetection Systems b.v., Science Park 406, 1098 XH Amsterdam, The Netherlands
[d] Bundesanstalt für Gewässerkunde, Am Mainzer Tor 1, 56068 Koblenz, DE, Germany
[e] Helmholtz Centre for Environmental Research - UFZ, Permoserstrasse 15, 04318 Leipzig, DE, Germany
[f] Eberhard Karls University Tübingen, Environmental Toxicology, Center for Applied Geosciences, 72074 Tübingen, DE, Germany
[g] INERIS, Rue Jaques Taffanel, Parc Technologique ALATA, 60550 Verneuil-en-Halatte, France
[h] Istituto di Ricerca Sulle Acque, Via del Mulino 19, 20861 Brugherio (MB), Italy
[i] Masaryk University, Research Centre for Toxic Compounds in the Environment (RECETOX), Kamenice 753/5, 625 00 Brno, Czech Republic
[k] French National Agency for Water and Aquatic Environments, 5 Square Felix Nadar, 94300 Vincennes, France
[l] Agence de l'eau Adour-Garonne, 90 rue de Fétéra, CS 87801, 31078 Toulouse Cedex 4, France
[m] Umweltbundesamt, Spittelauer Lände 5, 1090 Wien, Austria
[n] Umweltinstitut, Institut für Umwelt und Lebensmittelsicherheit des Landes Vorarlberg, Montfortstraße 4, 6901 Bregenz, Austria
[o] Institut Scientifique de Service Public (ISSeP), Rue Chéra 200, 4000 Liège, Belgium
[p] T.G. Masaryk Water Research Institute, Podbabská 2582/30, Praha 6, 16000, Czech Republic
[q] Landesamt für Natur, Umwelt und Verbraucherschutz NRW (LANUV), Auf dem Draap 25, 40221 Düsseldorf, DE, Germany
[r] Bayerisches Landesamt für Umwelt, Demollstrasse 32, 82407 Wielenbach, DE, Germany
[s] Instituto de Recursos Naturales, Universidad de León, Avenida de Portugal 42, 24071 León, ES, Spain
[t] Centro Ricerche (SMAT), Società Metropolitana Acque Torino S.p.A.C. So Unità d'Italia 235/3, 10127 Torino, Italy
[u] National Institute of Health, Department Environment and Health, Roma, Italy

ARTICLE INFO

Article history:
Available online 1 March 2018

Keywords:
Science-policy interface
Estrogen screening
Endocrine disruption
Surface and waste water assessment
Emerging pollutants
EU watch-list
Steroid analyses
In vitro bioassays
Integrated effects of mixtures

ABSTRACT

The European Decision EU 2015/495 included three steroidal estrogens, estrone, 17β-estradiol and 17α-ethinyl estradiol, in the "watch-list" of the Water Framework Directive (WFD). As consequence, these substances have to be chemically monitored at the level of their environmental quality standards, which can be challenging. This project aimed to identify reliable effect-based methods (EBMs) for screening of endocrine disrupting compounds, to harmonise monitoring and data interpretation methods, and to contribute to the current WFD review process. Water and wastewater samples were collected across Europe and analysed using chemical analyses and EBMs. The results showed that 17β-estradiol equivalents were comparable among methods, while results can vary between methods based on the relative potencies for individual substances. Further, derived 17β-estradiol equivalents were highly correlated with LC-MS/MS analyses. This study shows that the inclusion of effect-based screening methods into

* Corresponding author. Sarah Könemann, Institute for Environmental Research, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany.
E-mail address: sarah.koenemann@rwth-aachen.de (S. Könemann).
[1] Authors contributed equally to this manuscript.

https://doi.org/10.1016/j.trac.2018.02.008

monitoring programmes for estrogens in surface waterbodies would be a valuable complement to chemical analysis.

1. State of the art

Over the past two decades, numerous scientific studies have demonstrated that endocrine disrupting chemicals (EDCs) elicit adverse effects on sensitive aquatic species, such as fish [1–7]. Steroidal estrogens, like the natural hormones estrone (E1) and 17β-estradiol (E2), as well as the synthetic hormone 17α-ethinyl estradiol (EE2), are of particular environmental concern [8–11]. Due to their steady release via waste water effluents into surface waters [12,13] and their high biological activity, even very low concentrations of E2 and EE2 have been shown to cause reproductive toxicity with negative effects at the population level [14–16]. As a consequence, E1, E2, and EE2 were included in a European Union (EU) Water Framework Directive (WFD) "watch-list" [17–20]. The WFD watch-list mechanism aims to collect high-quality monitoring data on concentrations of emerging pollutants and potentially hazardous substances, whose currently available monitoring information shows either quantitative or qualitative deficiencies [21]. To collect more high-quality data, listed substances have to be monitored at representative EU sampling sites for a period of at least 12 and up to 48 months. The watch-list mechanism is expected to support future substance prioritisation processes, enable the implementation of measures, and facilitate environmental risk assessment across the EU.

Chemical monitoring of estrogens for the watch-list mechanism is challenging, because the European Commission set maximum acceptable method detection limits (MDLs) at EQS levels of 400 pg/L for E1 and E2, and 35 pg/L for EE2 [18,22]. Most routine analytical methods used by the Member States cannot meet these requirements, especially for EE2, based on [23,24]. Hence, the quality assessment of water bodies based on current methods is a challenge for the detection/quantification limits that are too high to detect if EQS are being exceeded or not. Effect-based methods are able to detect estrogenic substances at sub-ng or even pg levels and have the potential to be used as a complementary screening tool [12,25–27]. In addition, they do not require *a priori* knowledge of the substances to be monitored, as they are able to determine the biological response caused by complex mixtures of unknown compounds. Thus, effect-based methods may be suitable to serve as a valuable link between chemical analytical and ecological quality assessments, since the effects can rarely be linked to individual compounds.

As described in an EU technical report, which was elaborated in the context of the Chemical Monitoring and Emerging Pollutants (CMEP) expert group under the Common Implementation Strategy (CIS) of the WFD, effect-based tools can be categorised into three main groups: Bioassays (*in vitro*, *in vivo*), biomarkers, and ecological methods [28]. With regard to steroidal estrogens and other EDCs, *in vitro* reporter gene assays have been used predominantly to determine the total estrogen receptor (ER) mediated estrogenicity of an environmental sample [29]. Among the most commonly applied assays are *in vitro* methods such as estrogen receptor transactivation assays (ER-TAs), which use various cell types including yeast, human and other mammalian cell lines that were transfected with a human estrogen receptor coupled to a reporter gene [30]. Activation of the ER leads to the expression of the reporter gene product, usually an enzyme that modifies another chemical, causing a quantifiable response. The resulting estrogenic potential of a sample is expressed as an E2 equivalent concentration (EEQ), indicating the estrogenic activity of the sample or sample dilution in terms of equivalency to the estrogenic activity of the corresponding E2 reference concentration [31].

Although ER-TAs are highly advantageous methods for the detection of ER activation and quantification of very low estrogen concentrations in surface waters [23], these methods are not included within current WFD monitoring programmes [20]. One reason for this is the lack of data that demonstrate their applicability as a monitoring and screening tool in combination with chemical analytical methods (see *e.g.* Ref. [14]). Such information would greatly increase their regulatory acceptance. As a response to this need, an EU-wide project involving 24 research organisations and environmental agencies from 12 countries was carried out to evaluate the usefulness of specific *in vitro* methods for identifying the presence of the watch-list substances, E1, E2, and EE2, in surface and waste waters. The project aimed to compare the chemical and effect-based data resulting from the analysis of 16 surface and 17 waste water treatment plant effluent samples. Analyses were conducted in seven participating laboratories using different LC/MS- (three laboratories) and effect-based methods (five laboratories). The objectives of the study were (i) the demonstration of reliable effect-based screening methods for the monitoring of estrogenic EDCs in waste water and surface water, (ii) the harmonisation of data interpretation methods, and (iii) providing recommendations for the implementation of cost-effective and reliable effect-based methods in WFD monitoring programmes.

2. The project

2.1. Sampling

A total number of 16 surface water (SW) and 17 waste water (WW) samples were collected according to a protocol developed by the participants (SI, Part A). Selected sampling sites were located in seven European countries in Central and Southern Europe (Fig. 1): Austria (1 SW/3 WW), Belgium (2/2), Czech Republic (2/2), France (1/1), Germany (4/4), Italy (5/3), and Spain (1/2). Sample collection was carried out from September to November 2015 by ten participating institutions. The samples were taken based on prior knowledge on their contamination with estrogens and represented a gradient of contamination from high to moderate.

2.2. Sample preparation

The sample preparation included the filtering of a part of the SW (see SI, Part A) and all WW samples over glass fibre filters (Millipore, type 4, retention 2.7 μm, circle size 4.7 cm). Since a filtration step can have an impact on the composition of a sample and its estrogenic activity [32], the filtration step was investigated during a feasibility study prior to the main study presented here. The results of the pre study did neither show a significant reduction in estrogenicity in the control nor in tested environmental samples (data not shown). Subsequently, all samples were enriched by means of solid-phase extraction (SPE; 11 L sample to 11 mL extract) and extracts were passed over silica gel (SiOH) columns (methods focussing on E1, E2 and EE2). While for surface water each extract was split into eleven

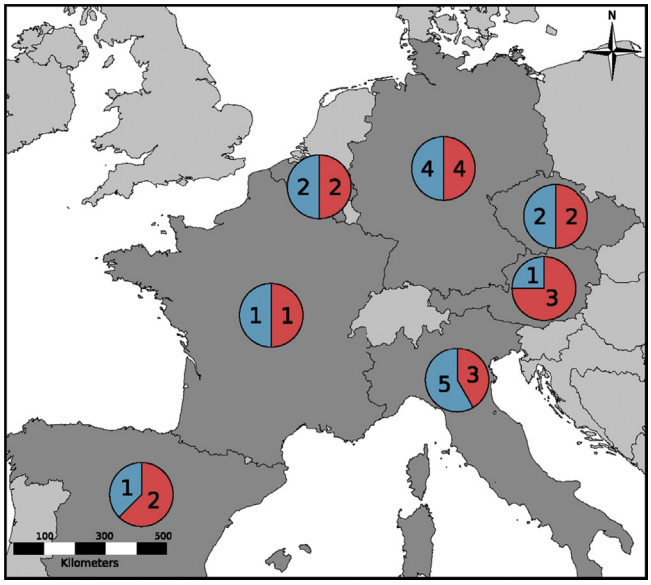


Fig. 1. Samples taken in various European States (dark grey). The circles indicate the number of surface water (blue) and waste water samples (red) taken in each country.

1 mL aliquots that were each passed over a single SiOH column, for waste water a single column was inadvertently used to treat the whole extract (11 mL). For LC-MS/MS analysis this means that matrix was less efficiently removed from WW extracts (relative to SW extracts) and higher matrix loads would have impeded low LOQs in WW LC-MS/MS analysis. For bioassay analysis this means that, should additional ER-agonists (*i.e.* other than E1, E2 and EE2) have been present in the extracts, a reduced clean-up efficiency would have reduced ER-agonist removal which in turn would have caused enhanced effects in bioassays. Full details of sample preparation are provided in SI, Part A.

2.3. Chemical and effect-based analyses

Participating laboratories received spiked reference samples, blanks and encoded water extracts. The chemical analyses were conducted in three different labs, which applied an LC-MS/MS with negative electron spray ionisation (detailed information in SI, Part D Table S2). The effect-based methods were conducted in five different labs: Estrogen Receptor Chemical Activated LUciferase gene eXpression (ER-CALUX) at Biodetection Systems (BDS), luciferase-transfected human breast cancer cell line (MELN) gene-reporter assay at INERIS [33], ER-GeneBLAzer assay at the Helmholtz Centre for Environmental Research (UFZ) [34], the stably transfected human estrogen receptor-alpha transcriptional activation Assay using hERα-HeLa-9903 cells (HeLa-9903 assay) at RECETOX [35], and planar Yeast Estrogen Screen (pYES) at the German Federal Institute of Hydrology (BfG) [36,37]. The pYES is a method, which combines a chromatographic separation of the sample by thin layer chromatography (TLC) with a subsequent performance of the YES on the planar surface of the TLC-plate [38–40]. Like the common assays which are performed in micro-well-plates, this approach allows the quantification of the overall estrogenic activity present in the sample by means of E2-equivalence concentrations. Furthermore, like methods based on LC/MS, it also allows the estimation of concentrations of individual estrogenic compounds, *e.g.* E1, E2 and EE2, due to the chromatographic separation of the sample. For this purpose the respective standard compounds are used for a calibration on the same TLC plate – in the present study E1, E2, EE2, and estriol (E3) were applied in a mixture at three different levels. Due to the limited separation power of the thin layer chromatography compared to HPLC and GC in particular, a co-migration of estrogenic compounds cannot be excluded. Therefore, under the assumption of effect addition, the estimated individual concentrations represent the possible maximal concentration of the respective compound. This approach can be used to identify and quantify substance groups causing ER-activation.

2.4. Blanks and positive controls

Ultrapure water (11 L) was used as extraction blank. An extraction blank was included with each extraction run of 10 samples, subjected to clean-up and distributed the same as the sample extracts. Further, each analysis using effect-based methods included a negative control. To avoid solvent effects on cell viability, its concentrations did not exceed a defined value (see SI, Part D Table S3). As positive controls for ensuring the validity and enabling a comparison of the methods, surface water samples (11 L each) from the Netherlands were spiked with E2 and EE2 at two concentrations by the central lab (BDS). The “low spike” (600 pg/L) represented a concentration slightly above the proposed EQS for E2 (400 pg/L). The “high spike” (6000 pg/L) represented a concentration that is quantifiable with high certainty by both effect-based and chemical methods.

2.5. Data evaluation – effect-based methods

Raw data and information on relative enrichment factors (REF) of the extracts were collected from participating laboratories. The REF expresses the combination of: 1) sample enrichment using SPE and 2) extract dilution steps in each of the applied effect-based methods. Estrogenic activity of the extracts was expressed as E2-equivalence concentration (pg EEQ/L water) (described in detail in SI, Part B). Briefly, dose-response curves of the reference compound, E2, and the dilution series of the water extracts and blanks were fitted using a five-parametric non-linear regression with normalised data. The concentration of the positive control (E2) needed to induce 10% effect of the maximum E2-induction (PC_{10}), was calculated. Subsequently, the relative REF of the sample, that stimulates the assay at PC_{10} level was determined by interpolation. The PC_{10} reference concentration was divided by the corresponding sample dilution (REF) to obtain the EEQ of the sample. EEQs derived by the PC_{10} method are presented in the results section.

2.6. Data evaluation – chemical analysis

Internal standard calibration and interpolation using a linear regression model were performed to determine concentrations (pg/L) of the individual steroidal estrogens in sample extracts. Identification of selected analytes was performed based on two to three Multiple Reaction Monitoring (MRM) transitions between the precursor ion and two or three most abundant product ions, depending on the laboratory where analyses were done. The first transition was used for quantification purposes whereas the second and third transitions were used to confirm the presence of the target compound in the sample. Quantified analytes were identified by comparing the retention time (RT) of the corresponding standard and the ratio between two ion transitions recorded (±20%) in the standard and water samples.

2.7. Calculation of sample-dependent LOD and LOQ

The Limits of quantification (LOQ) for effect-based methods the LOQs were calculated as 3-fold the standard deviation (SD) of the

averaged response of the negative control on each assay plate. The effect level of 3-fold the SD was interpolated from the E2 reference curve and divided by the REF of the sample to derive the LOQ. The actual reporting for effect-based methods occurred at the 10% effect level which was always above LOQ (typically at 2–5% effect levels).

In case of the chemical analysis the limits of detection (LOD) were determined for each compound in each sample based on the signal intensity of the internal standards or the analyte peak by a signal-to-noise (S/N) ratio of 3:1 and LOQ by a S/N ratio of 10:1.

When comparing LOQs of effect-based methods with those of chemical analyses the various key differences between the two approaches need to be taken into account (for further background see SI, Part C).

2.8. Comparison of chemical and biological analysis

The EEQ_{bio} is the ratio of the effect concentration of the reference compound estradiol EC_{50}(E2) (pg/L) and the sample EC_{50}(sample) (Equation (1)) and was derived in this study using the PC_{10} approach (see above). The EEQ_{chem} was calculated from the sum of the relative effect potencies REP_i times the detected concentration of estrogenic chemical i, c_i [41]. The REP, in turn, is the ratio of the effect concentration of the reference compound estradiol EC_{50}(E2) and the chemical i's EC_{50}(i) (Equation (2)).

$$EEQ_{bio} = \frac{EC_{50}(E2)}{EC_{50}(sample)} \quad (1)$$

$$EEQ_{chem} = \sum_{i=1}^{n} REP_i \cdot c_i = \sum_{i=1}^{n} \frac{EC50(E2)}{EC50(i)} \cdot c_i \quad (2)$$

Due to the analytical method detection limits of E2 and EE2, we evaluated the potential contribution of non-detected estrogens to the overall $EEQ_{chem,LOD/2}$ using Equation (3), where values below the LOD ("non-detects") were included as LOD/2. If the analytical lab reported data as <LOQ, we used LOQ/2 in Equation (3) instead of LOD/2. In Equation (3), n refers to the total number of chemicals included in the analysis, m refers to the number of chemicals below LOD. Ci is the average value of three analytical measurements,

$$EEQ_{chem,LOD/2} = \sum_{i=1}^{n-m} REP_i \cdot c_i + \sum_{j=1}^{m} REP_j \cdot LOD_j/2 \quad (3)$$

2.9. Correlation analysis

The correlation analysis among effect-based methods (EEQ_{bio}) was performed with GraphPad Prism, using the Pearson correlation (r) [42].

3. Results and discussion

3.1. Reference chemicals and validation

All essential criteria for method performance were fulfilled in this study (described in more detail in the SI, Part E). As shown in Table S4 (SI, Part E), the chemical analytical as well as effect-based methods showed good recovery in the spiked samples. No estrogenic activity or quantifiable concentrations of E1, E2, and EE2 were measured in the blank samples (*i.e.* procedure-, extraction- and solvent blanks). As the derived effect concentrations in the effect-based methods and chemically measured EE2 concentrations matched with the nominal concentrations of the spiked samples, the observed effects can be ascribed to the samples themselves.

3.2. Results of chemical analysis

Measured concentrations of the three estrogens E1, E2 and EE2 differed widely between sampling sites as well as between surface and waste water samples. Differences among SW samples can be explained by varying river characteristics, *e.g.* flow (dilution factor), or temperature, as well as differences in estrogenicity of treated WW, that are released into the SW. The results of the analyses, which are summarised in Fig. 2, show a 3.2 to 3.6 times higher mean concentration for E1 and E2 in WW (Fig. 2B) compared to SW (Fig. 2A). Due to the highly contaminated WW sample M(23), possibly influenced by an industrial discharge of EE2, the mean concentration of EE2 across all WW samples was approximately 20 times higher compared to SW (Fig. 2). Estrone (E1) was quantified in all samples. For E1 maximum concentrations of 5.6 ng/L (sample P(7)) and 20.5 ng/L (sample Q(20)) in SW and WW were measured, respectively. E2 was the second most frequently quantified estrogen and measured above LOQ in nine of 16 SW and six of 17 WW samples. Measured concentrations ranged from 0.4 ng/L (sample N(33)) to 1.1 ng/L (sample Q(20)) in WW, and from 0.06 ng/L (sample J(10)) to 0.5 ng/L (sample N(15)) in SW. The synthetic EE2 was least frequently quantified and measured above LOQ in four of 16 SW and four of 17 WW samples with a maximum concentration of 0.3 ng/L in SW sample O(3) and 7.5 ng/L in WW sample M(23). These concentration ranges and patterns are in accordance with recent review studies [43,44].

Our results underline the analytical difficulties that have recently been highlighted for E2 and EE2 by several studies and workshops [16,45], stressing the challenges that emerge for routine methods used in national monitoring programmes. Despite the use of quite advanced chemical analytical techniques (status 2015), the detection and quantification of E2 and EE2 in SW and WW samples was problematic in some cases. While it was possible to quantify E1 in almost all samples, the percentage of quantifications was significantly reduced for E2 and even more for EE2 (Fig. 3). This was partially due to the fact that insufficient silica gel was used to reduce the matrix effects in WW. WW is considered as worst-case regarding matrix effects [46,47].

However, the quantification of substances itself is not the only challenge faced by those routinely applying analytical methods for watch-list monitoring. According to the EU Commission Decision 2015/495, which established the first watch-list, the indicative methods applied by Member States have to meet the minimum requirement for method detection limits (MDL) equal to the proposed EQSs of E1 at 3.6 ng/L, E2 at 0.4 ng/L and EE2 at 0.035 ng/L [18]. To take into consideration the matrix effects of different waters, LODs and LOQs had to be calculated for each sample (SI Part F, Table S7). The three techniques used in the current study were able to meet MDL requirements for E1 in all SW and WW samples. Also for E2, in 96% of surface water samples and 94% of waste water samples detection was possible at the level of the proposed EQS. In the case of EE2, the minimum criteria were not met, since only 56% and 16% of SW and WW samples, respectively, could be monitored at the EQS level. These findings are in accordance with a recent report from 2015, which showed that the lowest LOQ found in literature at that time was sufficient for compliance monitoring of E1 and E2 in inland surface waters, while the criteria were not met for EE2 by several Member States [24]. It has to be pointed out that, in this project, the silica clean-up step for the sample extracts differed between WW and SW samples (see methods section) favouring the presence of polar compounds in extracts of WW samples. This difference likely reduced the sensitivity of the analytical method for the target compounds in WW samples. Furthermore, sample extraction was performed at pH 3 possibly increasing concentrations of humic acids and thus lowering sensitivity of LC/MS-based methods applied. Under ideal

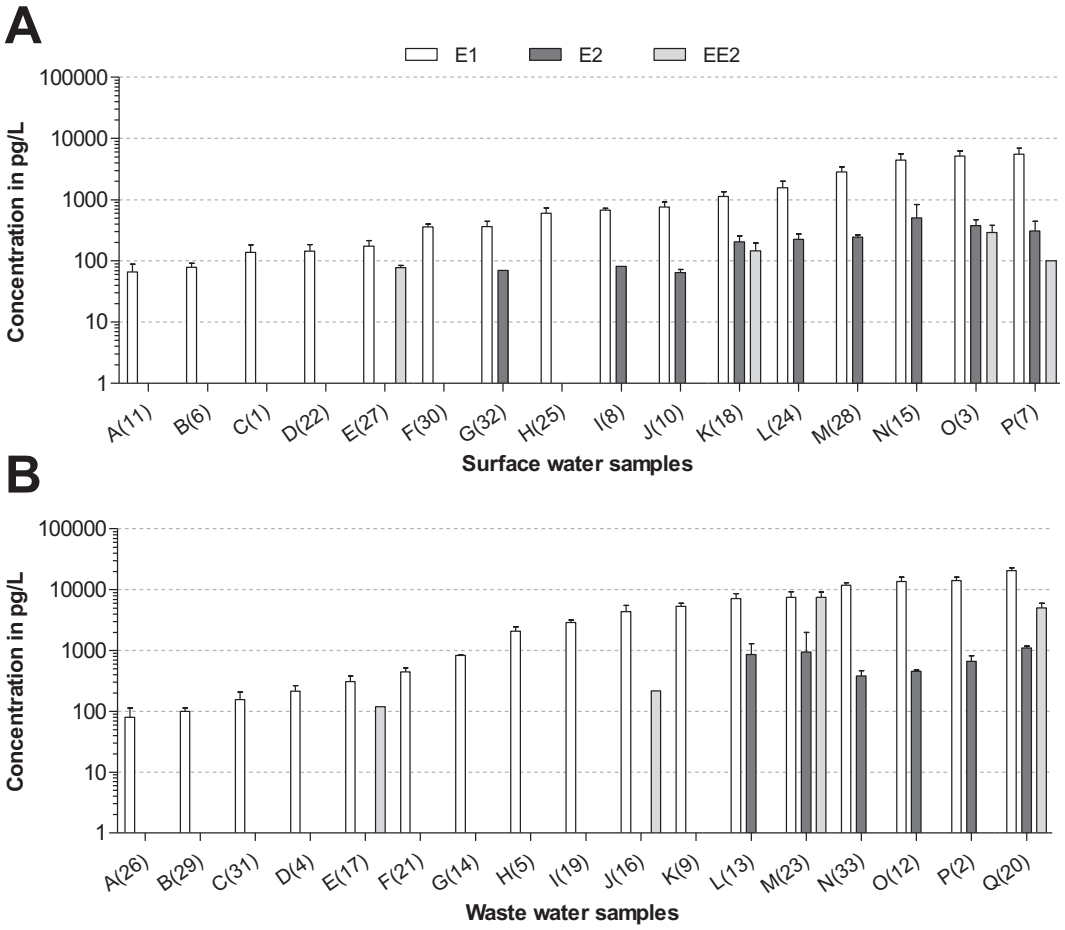


Fig. 2. Chemical analytically measured concentrations for SW (A) and WW extracts (B) above LOQ for E1, E2 and EE2. The bars show the mean concentration of all three applied methods for each analyte showing results > LOQ, the standard deviation is shown when two or three methods reported results. The sample-dependent LOQs are listed in the supplementary information together with the measurement data of analytical methods (SI, Part F, Tables S6 and S7).

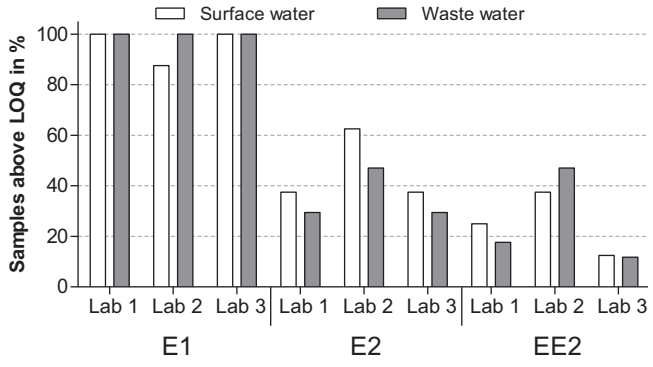


Fig. 3. Mean percentage of quantified (>LOQ) samples for each substance in SW and WW. The sample-dependent LOQs are listed in the supplementary information together with the measurement data of the analytical methods (SI Part F, Table S7).

conditions, we estimate that analytical methods can achieve LODs and LOQs of a factor 2 to 3 lower in WW samples. It has to be recognised that the LODs of chemical analytical methods used exclusively for steroidal estrogens already significantly decreased from 2013 (LOD E2 and EE2 of 100 pg/L) to 2015 (E2: 60 pg/L, EE2: 85 pg/L) and will certainly decrease further [16,23].

Nevertheless, if steroidal estrogens were to be included in the EU priority list for monitoring, very strict minimum performance criteria would apply. As stated in the Commission Directive 2009/90/EC, an analytical method used for monitoring of priority substances needs a LOQ equal or below a value of 30% of the EQS [48]. These requirements can presently be met only for E1, but not for E2 or EE2 in all SW. Regarding the quantification of E2, and EE2, existent routine analytical techniques still lag behind the requirements. This result is supported by two recent reviews on the performance of current analytical methods that have shown that 35% of reviewed methods complied with the EQS for E2, while only one method complied with the EQS for EE2 [49,50]. In order to not only detect but also quantify at such low concentrations as required for regulatory monitoring application, a further decrease of LOQs is necessary, which is difficult to achieve for routinely used non-tailored analytical methods in the short-term.

3.3. Quantification limits of chemical-analytical and in vitro effect-based methods

The LOQs for all methods applied in this study are summarised in Fig. 4. Since E2 is used as the reference compound for all effect-based methods, the LOQ of E2 is shown for the chemical-analytical methods as an example. When comparing LOQs across the different methods it has to be taken into account that LOQs were derived along different approaches (see method section and SI, Part C for further details). The effect-based *in vitro* methods were generally able to quantify effects at one to two orders of magnitude lower concentrations than the analytical methods used. For effect-based

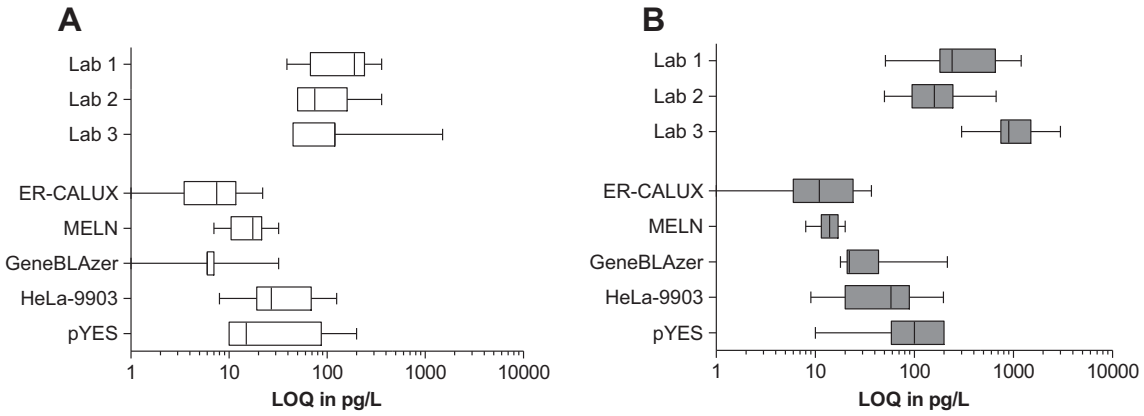


Fig. 4. Sample-dependent LOQs in surface water (A) and waste water (B) extracts. For the chemical analytical method the LOQ of E2 is shown as an example and for the effect-based methods the LOQ of the integrated effects is represented. Plots indicate the distribution of data, thereby the bottom and the top of the box are the first and third quartiles, while the line inside the box is the median. The whiskers show the minimum and maximum of all data.

methods, LOQs ranged between 0.002 ng/L and 0.2 ng/L for SW as well as WW, while for chemical-analytical methods LOQs for E2 were 0.04 ng/L to 1.5 ng/L in SW and 0.05 ng/L to 3 ng/L in WW. This increase in LOQs for chemical-analytical methods in WW samples (Fig. 4B) compared to surface water (Fig. 4A) can be ascribed to the higher complexity of the waste water matrix [46,47] as well as the less efficient clean-up used for WW samples.

3.4. Measured estrogenic effects

As a result of these low effect-based quantification limits, estrogenic activities were detected in all tested samples. As expected, highest EEQs were measured in WW samples (Fig. 5A and B) . In SW, EEQ_{bio} ranged from 0.16 ng/L measured with HeLa-9903 in sample B(6) to up to 5.4 ng/L measured with pYES in sample O(3).

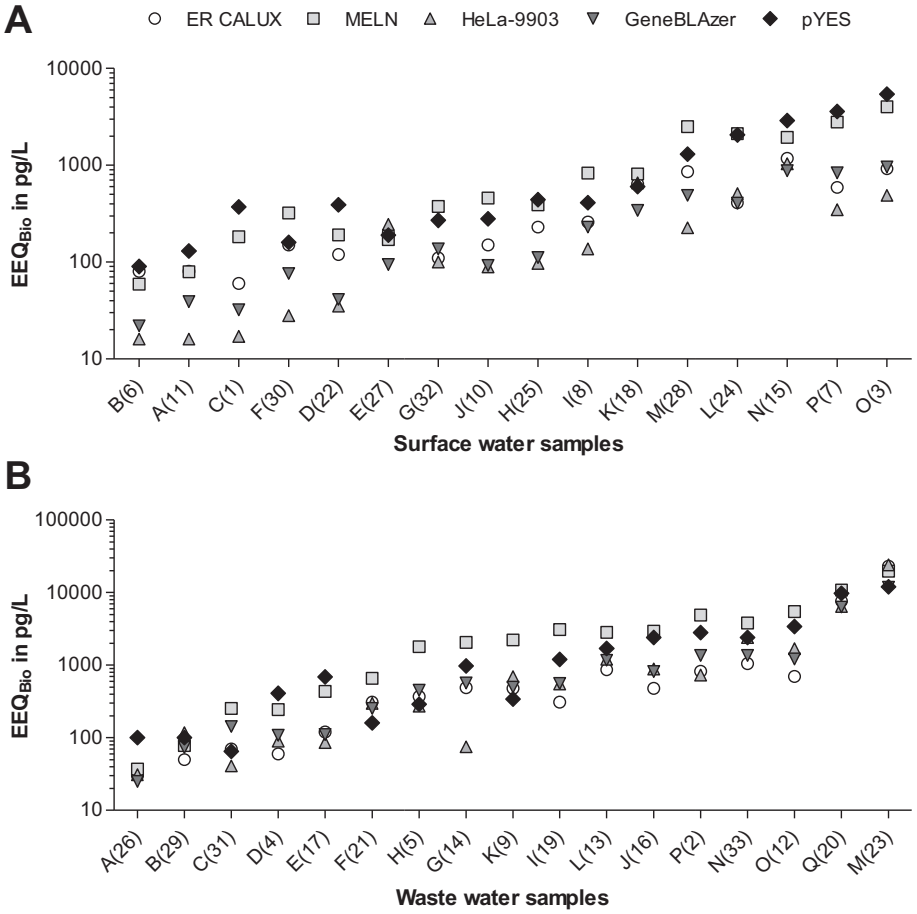


Fig. 5. Measured E2-equivalents for all SW (A) and WW (B) extracts. The symbols show the EEQs for each bioassay, which were calculated according to the method described in Section 2.5. The sample-dependent LOQs are mentioned in the supplementary information, together with the measurement data of effect-based methods (SI Part F, Tables S8 and S9).

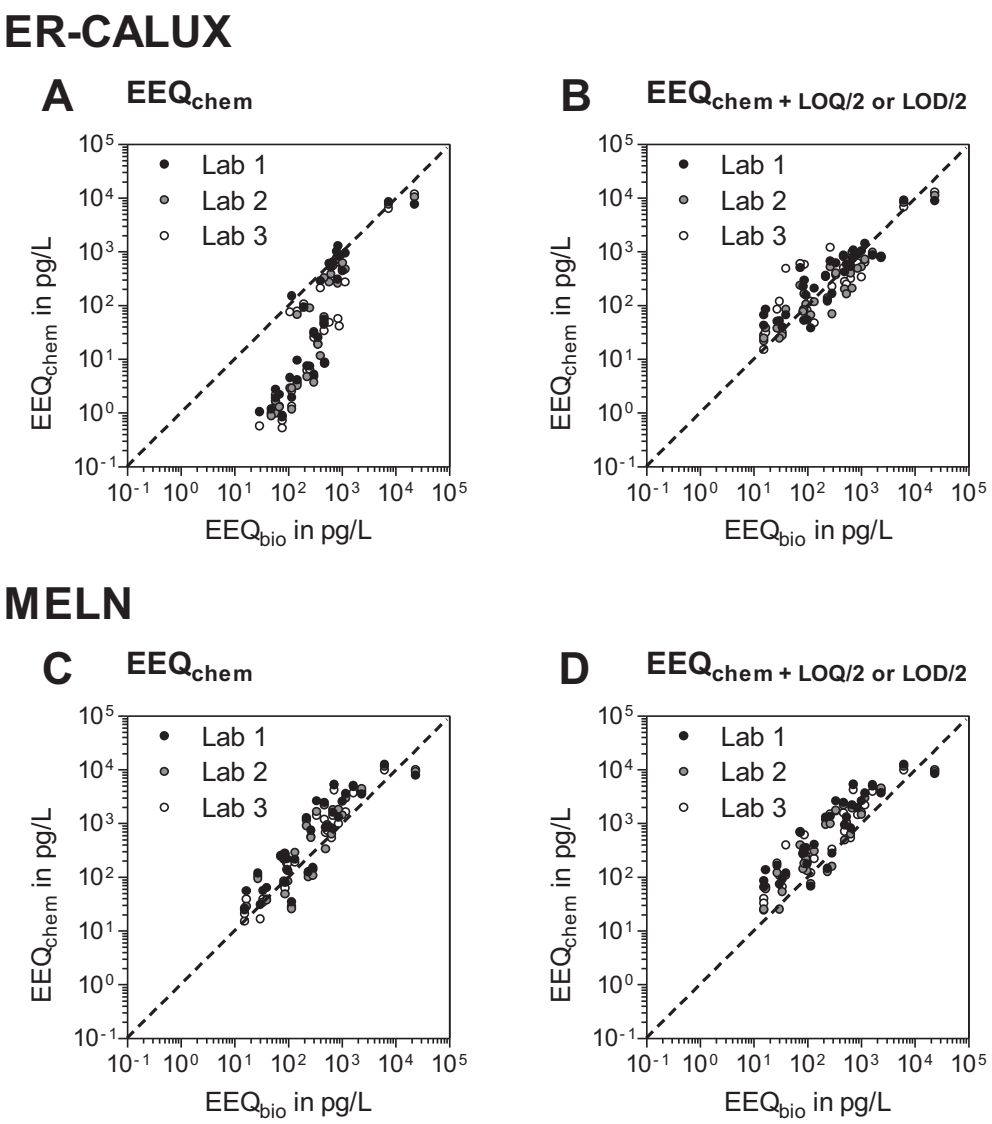


Fig. 6. Comparison of EEQ_{chem} with EEQ_{bio}. Exemplary graphs are shown for the ER-CALUX (A, B) and MELN assay (C, D) (further figures in the SI, Part G). Graphs on the left show the EEQ_{chem} derived from values > LOQ, while the graphs on the right show the $EEQ_{chem+LOD/2 \text{ or } LOQ/2}$ calculated by including LOD/2 or LOQ/2. The dashed line indicates perfect agreement of EEQ_{chem} with EEQ_{bio}.

In WW, the lowest EEQ_{bio} of 0.03 ng/L was measured in sample A(26) with ER-GeneBLAzer, while the highest EEQ_{bio} of 24 ng/L was measured in sample M(23) with HeLa-9903. Further, it is evident that EEQ_{bio} for SW samples determined with the MELN, as well as the pYES, were higher (>50%) than the EEQ_{bio} measured with the other effect-based methods. A possible reason for this pattern, which was less pronounced in WW, could be a higher sensitivity of the MELN and pYES towards E1 (see SI Part F, Table S8), combined with a larger proportion of E1 in surface water. Additionally, alterations in the method's performance occur due to differences between the test systems, which was already mentioned in previous studies [23,44,51] and is further discussed for this project in an associated publication [52].

3.5. Comparison of chemical analysis and in vitro effect-based methods

We cannot *a priori* expect consistency between EEQ_{chem} calculated from E1, E2, and EE2 concentrations and EEQ_{bio}. Although the extraction and clean-up method focused on E1, E2, and EE2, other natural estrogens and xenoestrogens (both agonists and antagonists) might still be present in the extracts and contribute to the mixture effects detected by effect-based methods. Thus, there can be situations where EEQ_{chem} is lower than EEQ_{bio} because: 1) agonists other than E1, E2, and EE2 were present in the sample but not quantified by LC-MS/MS analyses or 2) some target compounds were present but below LOQ or LOD, thus they were not included in EEQ_{chem} but still contributed to EEQ_{bio}. Alternatively, EEQ_{chem} can be higher than EEQ_{bio} when antagonists supress the response of the assay.

For ER-CALUX, the comparison of EEQ_{bio} with EEQ_{chem} (Fig. 6A) indicated an underestimation of EEQ_{bio} by EEQ_{chem} at low concentrations of steroidal estrogens. When E1 concentrations are low, typically E2 and EE2 concentrations are below LOQ (Fig. 2). However, as stated above, also below their LOD/LOQ, these chemicals may be present and contribute to the biological mixture effect (*i.e.* EEQ_{bio}). We therefore also calculated the $EEQ_{chem,LOD/2}$ that uses the LOD/2 or LOQ/2 for those E2 and EE2 concentrations below the LOD

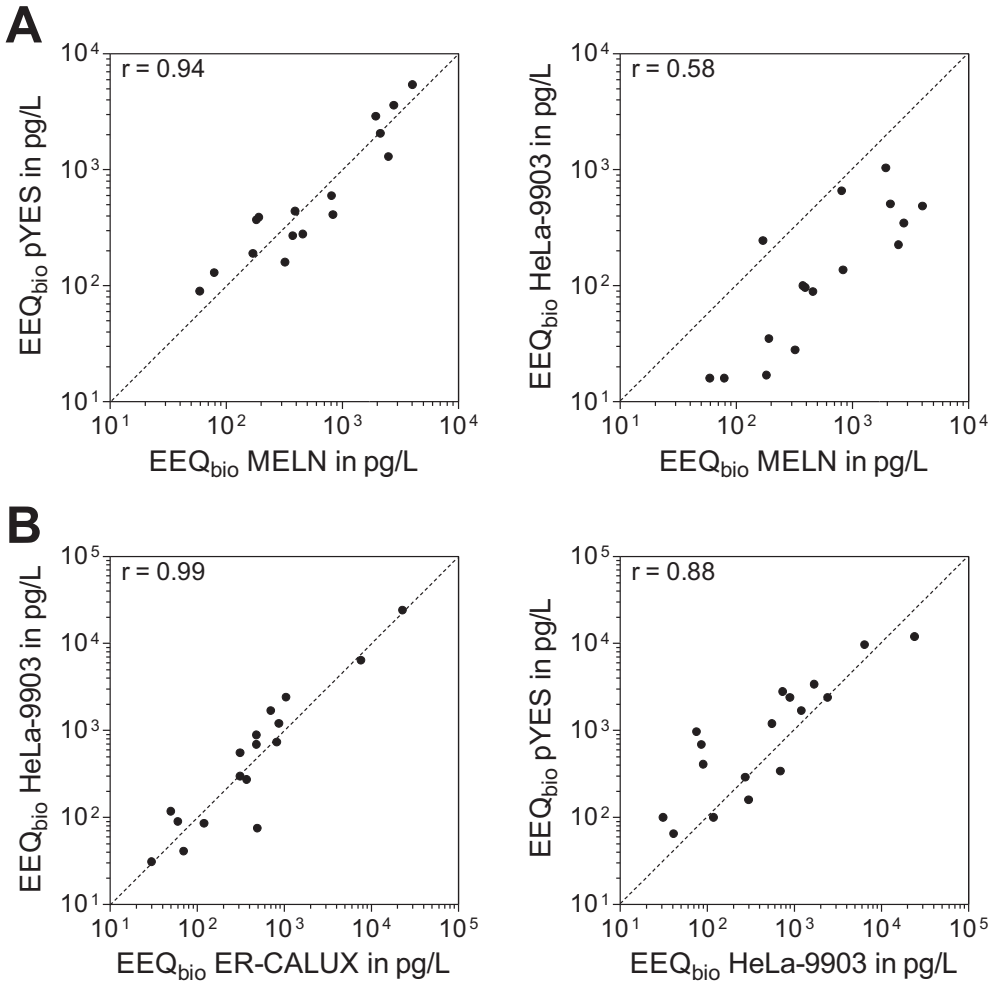


Fig. 7. Exemplary graphs of correlation analysis of effect-based methods for SW (A) and WW (B) showing the strongest and weakest correlations. The correlation analysis was based on the method described in Section 2.9. The dashed line indicates perfect agreement of the compared effect-based methods. All correlations were significant with a p value < 0.0001 except for MELN and HeLa-9903 (top right panel) which had a p value ≈ 0.01. Further graphs are shown in SI, Part H, Figs. S2 and S3.

or LOQ. The increase in EEQ_{chem}, due to the inclusion of LOQ/2 and LOD/2 data (SI, Part F, Tables S10–14), shifts the EEQ_{chem} - EEQ_{bio} data cluster towards the one-to-one line (Fig. 6B). In fact, there is now a slight overestimation of the biological effect in the range where EEQ concentrations are low (up to ca.100 pg/L). The fact that the agreement between EEQ_{chem} and EEQ_{bio} has become much better (going from Fig. 6A and B) is a good indication that E2 and EE2 are indeed present and were captured by effect-based methods.

The situation for MELN is markedly different from that of ER-CALUX. For MELN the direct comparison between EEQ_{chem} and EEQ_{bio} is already very good (Fig. 6C). In fact, EEQ_{chem} tends to be above EEQ_{bio} already before adding the additional EEQ_{chem} component using LOD/2 or LOQ/2 for E2 and EE2. The inclusion of LOD/2 or LOQ/2 in the EEQ_{chem} calculation caused a notable overestimation of EEQ_{chem} for almost all samples (>90% of data above the 1 to 1 line in Fig. 6C). The other three bioassays show results that are intermediate between ER-CALUX and MELN, with a general trend towards a slight underestimation of EEQ_{chem} for samples with low EEQ_{bio} and an overestimation after adding LOD/2 or LOQ/2 (see Fig. S1).

The marked differences between ER-CALUX and MELN are not unexpected. MELN has the highest relative E1 effect potency of all tested bioassays (0.29 compared to 0.01 for ER-CALUX; Table S5). Thus, EEQ_{chem} results for MELN are strongly based on E1 concentrations – a compound that was always measured (except for a few samples by Lab 2, Fig. 3). Consequently, for MELN the relative contribution of E2 and EE2 at LOD/2 or LOQ/2 on top of measured E1 concentrations is relatively small though still noticeable for samples with low EEQ concentrations (compare Fig. 6C and D).

3.6. Comparison of effect-based methods

To compare the five effect-based methods amongst each other, a correlation analysis was conducted by plotting the EEQs of one method against the EEQs of all other methods for SW samples and WW samples, respectively (Fig. 7).

The results of this analysis are summarised in Tables 1 and 2 and show a strong correlation and thus good comparability of pYES, MELN and ER-CALUX. For SW samples, the strongest correlations were seen for pYES/MELN ($r^\circ = 0.94$) and pYES/ER-GeneBLAzer ($r^\circ = 0.94$), while the weakest correlation was determined for MELN/HeLa-9903 ($r^\circ = 0.58$). For WW samples, test results correlated strongly among all methods (Table 2), and the strongest correlation ($r^\circ = 0.99$) was observed for ER-CALUX/HeLa-9903. It is known that effect-based methods differ in their REPs for individual ER-agonists [53–55] which can explain that results obtained by the HeLa-9903 assay correlated less strongly with other test results.

Table 1
Pearson correlation coefficients of all bioassays for SW. The values were calculated according to the method mentioned in Section 2.9. All correlations were significant with a p value < 0.0001 (***) and a p value ≈ 0.01 (*).

	MELN	ER-GeneBLAzer	HeLa-9903	pYES
ER-CALUX	0.81***	0.91***	0.86***	0.76***
MELN		0.93***	0.58*	0.94***
ER-GeneBLAzer			0.77***	0.94***
HeLa-9903				0.61*

Table 2
Pearson correlation coefficients of all bioassays for WW. The values were calculated according to the method mentioned in Section 2.9. All correlations were significant with a p value < 0.0001 (***).

	MELN	ER-GeneBLAzer	HeLA-9903	pYES
ER-CALUX	0.94***	0.98***	0.99***	0.89***
MELN		0.98***	0.94***	0.97***
ER-GeneBLAzer			0.97***	0.96***
HeLa-9903				0.88***

Based on these differences effect-based methods can be split into two groups: pYES and MELN with high E1 REP and ER-CALUX, HeLa-9903 and ER-GeneBLAzer with lower E1 REP.

4. Conclusions and trends

By including E1, E2, and EE2 in the watch-list of the WFD, the European Commission recognised the need to assess environmental occurrence and impact of these endocrine disrupting substances. However, the current WFD monitoring approach, which is based on chemical analytical measurements and compliance with specific EQSs, has been shown to be limited with regard to the ability to detect these substances at required concentrations [18,51]. As demonstrated in this study, chemical analytical methods (status 2015) were unable to quantify the steroidal estrogens E2 and EE2 at EQS concentrations in all samples although E1 was measured effectively. Using effect-based methods, EEQ concentrations could be determined in all samples. As these EEQ concentrations are the responses to mixtures of known as well as unknown substances, effect-based methods have the potential to be highly valuable tools complementing routine monitoring and water quality assessment for estrogenic compounds. Effect-based methods are of particular regulatory interest as tools to screen and prioritise samples for further analysis by chemical analytical methods. Furthermore, DIN/EN/ISO standards to determine the estrogenic potential of water samples – covering human cell lines (*e.g.* ER-CALUX) and yeast based assays – will be available in early 2018 under ISO/DIS19040. The availability of such standards will facilitate the integration of effect-based methods into regulatory schemes.

Our study showed that EEQ results obtained from all effect-based methods applied were comparable – especially at higher concentrations found in WW – but results can vary between methods based on the relative effect potencies for individual substances. This has to be considered for the interpretation of data and determination of threshold values. As stated above: 1) *in vitro* effect-based methods cannot deliver single substance based measurements, but are suitable to assess overall estrogenicity in water samples and 2) results of these methods need to be confirmed by advanced chemical analysis. Along these lines, the inclusion of effect-based methods into monitoring programmes as a screening tool (detailed description in Kase *et al.*, [52]) for estrogenic substances in surface water bodies would be a valuable complement to chemical analysis currently foreseen by the Directive 2013/39/EU and WFD [28, 56, 57].

Conflict of interests

The Federal Institute of Hydrology did not receive any kind of financial support from the Pharmaceutical Associations. Other authors declare no conflict of interest.

Acknowledgements

The major part of the project was funded by in-kind contribution from numerous project partners showing high interest in this project. Essential parts of the project, e.g. extraction and reporting were externally co-funded by the Swiss Centre of Applied Ecotoxicology EAWAG-EPFL, NORMAN Network, INERIS and Pharmaceutical Associations (Pfizer, Johnson & Johnson, TEVA which joined the project later and only supported the extraction at central Lab BDS). In total, 70 colleagues associated to 24 institutes and agencies in 12 nations contributed to this project. The authors would like to thank: Joint Research Centre (EC), The French Agency for Biodiversity (previously ONEMA) (FR), INERIS (FR), Bio Detection Systems (NL), Swiss Centre for Applied Ecotoxicology (CH), Federal Institute of Hydrology (DE), Federal Environment Agency (DE), RWTH Aachen (DE), RECETOX (CZ), NORMAN-Network, Helmholtz Centre for Environmental Research-UFZ (DE), IRSA-CNR (IT), Italian Institute of Health (IT), University of Leon (ES), Water Research Institute T.G.Masaryk (CZ), Bavarian State Office for Environment (DE), LANUV (DE), Environment Agency Austria (AT), ISSeP (Scientific Institute of Public Service) Wallonia (BE), SMAT (IT), BrianzAcque (IT), Vettabbia scarl, MilanoDepur (IT), Agence de l'Eau Adour-Garonne (FR), Ontario Ministry of the Environment and Climate Change (CAN), McGill University (CAN), Environmental Institute (SK) and DG Environment of the European Commission (EC). The project partners from UFZ, RWTH Aachen, INERIS and RECETOX were supported by the SOLUTIONS project, that is funded by the European Union Seventh Framework Programme (FP7-ENV-2013-two-stage Collaborative project) under grant agreement number 603437. Additionally, we would like to thank Christin Kuehnert (UFZ), Maria König (UFZ) and Michael Gundlach (RWTH) for experimental assistance. Special thanks for stewardship of this study we would like to address to Helen Clayton and Stéphanie Schaan from European Commission, Environment DG.

The Federal Institute of Hydrology got financial support by the Federal Ministry for the Environment, Nature Conservation, and Building (BMUB).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.trac.2018.02.008.

References

[1] John P. Sumpter, Andrew C. Johnson, 10th anniversary perspective: reflections on endocrine disruption in the aquatic environment: from known knowns to unknown unknowns (and many things in between), J. Environ. Monit. 10 (12) (2008) 1476–1485.
[2] Georg Streck, Chemical and biological analysis of estrogenic, progestagenic and androgenic steroids in the environment, Trac. Trends Anal. Chem. 28 (6) (2009) 635–652.
[3] Daniel J. Caldwell, Frank Mastrocco, Paul D. Anderson, Reinhard Länge, John P. Sumpter, Predicted-no-effect concentrations for the steroid estrogens estrone, 17β-estradiol, estriol, and 17α-ethinylestradiol, Environ. Toxicol. Chem. 31 (6) (2012) 1396–1406.
[4] Helmut Segner, Ayako Casanova-Nakayama, Robert Kase, Charles R. Tyler, Impact of environmental estrogens on yfish considering the diversity of estrogen signaling, Gen. Comp. Endocrinol. 191 (2013) 190–201.

[5] Jinmiao Zha, Liwei Sun, Yiqi Zhou, Philip A. Spear, Mei Ma, Zijian Wang, Assessment of 17α-ethinylestradiol effects and underlying mechanisms in a continuous, multigeneration exposure of the Chinese rare minnow (gobiocypris rarus), Toxicol. Appl. Pharmacol. 226 (3) (2008) 298–308.

[6] Karen A. Kidd, Paul J. Blanchfield, Kenneth H. Mills, Vince P. Palace, Robert E. Evans, James M. Lazorchak, Robert W. Flick, Collapse of a fish population after exposure to a synthetic estrogen, in: Proceedings of the National Academy of Sciences, 2007, pp. 8897–8901.

[7] Stefanie Grund, Eric Higley, René Schönenberger, Marc JF. Suter, P.John Giesy, Thomas Braunbeck, Markus Hecker, Henner Hollert, The endocrine disrupting potential of sediments from the upper danube river (Germany) as revealed by in vitro bioassays and chemical analysis, Environ. Sci. Pollut. Control Ser. 18 (3) (2011) 446–460.

[8] Barbara V. Rutishauser, Maija Pesonen, Beate I. Escher, Gabriele E. Ackermann, Hans-Rudolf Aerni, Marc J-F. Suter, Rik IL. Eggen, Comparative analysis of estrogenic activity in sewage treatment plant effluents involving three in vitro assays and chemical analysis of steroids, Environ. Toxicol. Chem. 23 (4) (2004) 857–864.

[9] Andrew C. Johnson, Egon Dumont, Richard J. Williams, Rik Oldenkamp, Iwona Cisowska, John P. Sumpter, Do concentrations of ethinylestradiol, estradiol, and diclofenac in european rivers exceed proposed eu environmental quality standards? Environ. Sci. Technol. 47 (21) (2013) 12297–12304.

[10] Adam R. Schwindt, Dana L. Winkelman, Kristen Keteles, Mark Murphy, Alan M. Vajda, An environmental oestrogen disrupts fish population dynamics through direct and transgenerational effects on survival and fecundity, J. Appl. Ecol. 51 (3) (2014) 582–591.

[11] Eric Higley, Stefanie Grund, Paul D. Jones, Tobias Schulze, Thomas-B. Seiler, Urte Lübcke-von Varel, Werner Brack, Jan Wölz, Hanno Zielke, John P. Giesy, Endocrine disrupting, mutagenic, and teratogenic effects of upper danube river sediments using effect-directed analysis, Environ. Toxicol. Chem. 31 (5) (2012) 1053–1062.

[12] C.E.J.R. Desbrow, E.J. Routledge, G.C. Brighty, J.P. Sumpter, M. Waldock, Identification of estrogenic chemicals in stw effluent. 1. Chemical fractionation and in vitro biological screening, Environ. Sci. Technol. 32 (11) (1998) 1549–1558.

[13] Patricia Burkhardt-Holm, Walter Giger, Herbert GUttinger, Ueli Ochsenbein, Armin Peter, Karin Scheurer, Helmut Segner, Erich Staub, Marc J-F. Suter, Where have all the fish gone? Environ. Sci. Technol. 39 (21) (2005) 441A–447A.

[14] Frederic DL. Leusch, Christiaan De Jager, Yves Levi, Richard Lim, Leo Puijker, Frank Sacher, Louis A. Tremblay, Vickie S. Wilson, Heather F. Chapman, Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters, Environ. Sci. Technol. 44 (10) (2010) 3853–3860.

[15] Jon P. Nash, David E. Kime, Leo TM. Van der Ven, Piet W. Wester, François Brion, Gerd Maack, Petra Stahlschmidt-Allner, Charles R. Tyler, Long-term exposure to environmental concentrations of the pharmaceutical ethynylestradiol causes reproductive failure in fish, Environ. Health Perspect. (2004) 1725–1733.

[16] Christiane Heiss, Recommendation for a Monitoring Strategy for Estrogens in Coastal and Continental Surface Waters, 2013 [cited 2018 February 5th]; Available from: http://www.bafg.de/DE/05_Wissen/02_Veranst/2013/2013_02_27_votum_en.pdf?__blob=publicationFile.

[17] European Union, Directive 2013/39/eu of the european parliament and of the council of 12 august 2013 amending directives 2000/60/ec and 2008/105/ec as regards priority substances in the field of water policy, Off. J. Eur. Union (2013).

[18] EU Decision, 495/2015, commission implementing decision (eu) 2015/495 of 20 march 2015 establishing a watch list of substances for union-wide monitoring in the field of water policy pursuant to directive 2008/105/ec of the european parliament and of the council, Off. J. Eur. Union L 78 (2015) 40–42.

[19] Mario Carere, Stefano Polesello, Robert Kase, Bernd Manfred Gawlik, The emerging contaminants in the context of the eu water framework directive, in: Emerging Contaminants in River Ecosystems, Springer, 2015, pp. 197–215.

[20] EU WFD, Directive 2000/60/ec of the European Parliament and of the Council Establishing a Framework for the Community Action in the Field of Water Policy, The European Parliament and the Council of the European Union, Brussels, Belgium, 2000.

[21] R.N. Carvalho, L. Ceriani, A. Ippolito, Development of the First Watch List under the Environmental Quality Standards Directive Water Policy, 2015.

[22] A.S. Wernersson, C. Maggi, M. Carere, Technical Report on Aquatic Effect-based Monitoring Tools, Office for Official Publications of the European Communities, 2014.

[23] Petra Y. Kunz, Cornelia Kienle, Mario Carere, Nadzeya Homazava, Robert Kase, In vitro bioassays to screen for endocrine active pharmaceuticals in surface and waste waters, J. Pharmaceut. Biomed. Anal. 106 (2015) 107–115.

[24] Robert Loos, Analytical Methods for Possible Wfd 1st Watch List Substances, 2015 [cited 2018 February 5th], Report EUR 27046 EN, Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.829.1423&rep=rep1&type=pdf.

[25] Mario Carere, Stefano Polesello, Caterina Sollazzo, Bernd Manfred Gawlik, Chemical monitoring and emerging pollutants in the common implementation strategy of the water framework directive, Trac. Trends Anal. Chem. 36 (2012) 12–14.

[26] Markus Hecker, Henner Hollert, Endocrine disruptor screening: regulatory perspectives and needs, Environ. Sci. Eur. 23 (1) (2011) 1.

[27] A.D. Vethaak, A.C. Belfroid, Estrogens and Xeno-estrogens in the Aquatic Environment of The Netherlands: Occurence, Potency and Biological Effects, Rijkswaterstaat, RIKZ, 2002.

[28] Ann-Sofie Wernersson, Mario Carere, Chiara Maggi, Petr Tusil, Premysl Soldan, Alice James, Wilfried Sanchez, Valeria Dulio, Katja Broeg, Georg Reifferscheid, The european technical report on aquatic effect-based monitoring tools under the water framework directive, Environ. Sci. Eur. 27 (1) (2015) 1–11.

[29] Karin Kinnberg, Evaluation of in Vitro Assays for Determination of Estrogenic Activity in the Environment, Danish Ministry of the Environment, Danish Environmental Protection Agency, 2003.

[30] Diane M. Klotz, Barbara S. Beckman, Steven M. Hill, John A. McLachlan, Marian R. Walters, Steven F. Arnold, Identification of environmental chemicals with estrogenic activity using a combination of in vitro assays, Environ. Health Perspect. 104 (10) (1996) 1084.

[31] Beate I. Escher, Nadine Bramaz, Jochen F. Mueller, Pamela Quayle, Sibylle Rutishauser, Etiënne LM. Vermeirssen, Toxic equivalent concentrations (teqs) for baseline toxicity and specific modes of action as a tool to improve interpretation of ecotoxicity testing of environmental samples, J. Environ. Monit. 10 (5) (2008) 612–621.

[32] Charles W. Walker, John E. Watson, Adsorption of estrogens on laboratory materials and filters during sample preparation all rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher, J. Environ. Qual. 39 (2) (2010) 744–748.

[33] Patrick Balaguer, Fabienne François, Franck Comunale, Hélène Fenet, Anne-Marie Boussioux, Michel Pons, Jean-Claude Nicolas, Claude Casellas, Reporter cell lines to study the estrogenic effects of xenoestrogens, Sci. Total Environ. 233 (1) (1999) 47–56.

[34] Beate I. Escher, Mayumi Allinson, Rolf Altenburger, Peter A. Bain, Patrick Balaguer, Wibke Busch, Jordan Crago, Nancy D. Denslow, Elke Dopp, Klara Hilscherova, Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays, Environ. Sci. Technol. 48 (3) (2013) 1940–1956.

[35] OECD, Test No. 455: Performance-based Test Guideline for Stably Transfected Transactivation in Vitro Assays to Detect Estrogen Receptor Agonists. OECD Publishing.

[36] Denise Spira, Georg Reifferscheid, Sebastian Buchinger, Combination of high-performance thin-layer chromatography with a specific bioassay-a tool for effect-directed analysis, JPC J. Planar Chromatogr.-Mod TLC 26 (5) (2013) 395–401.

[37] M.B. Müller, C. Dausend, C. Weins, F.H. Frimmel, A new bioautographic screening method for the detection of estrogenic compounds, Chromatographia 60 (3) (2004) 207–211.

[38] Sebastian Buchinger, Denise Spira, Kathrin Bröder, Michael Schlüsener, Thomas Ternes, Georg Reifferscheid, Direct coupling of thin-layer chromatography with a bioassay for the detection of estrogenic compounds: applications for effect-directed analysis, Anal. Chem. 85 (15) (2013) 7248–7256.

[39] Ines Klingelhöfer, Gertrud E. Morlock, Bioprofiling of surface/wastewater and bioquantitation of discovered endocrine-active compounds by streamlined direct bioautography, Anal. Chem. 87 (21) (2015) 11098–11104.

[40] Andreas Schönborn, Andrea Grimmer, Coupling sample preparation with effect-directed analysis of estrogenic activity-proposal for a new rapid screening concept for water samples, JPC J. Planar Chromatogr.-Mod TLC 26 (5) (2013) 402–408.

[41] Maria König, Beate I. Escher, Peta A. Neale, Martin Krauss, Klára Hilscherová, Jiří Novák, Ivana Teodorović, Tobias Schulze, Sven Seidensticker, Muhammad Arslan Kamal Hashmi, Impact of untreated wastewater on a major european river evaluated with a combination of in vitro bioassays and chemical analysis, Environ. Pollut. 220 (2017) 1220–1230.

[42] Lothar Sachs, Angewandte Statistik: Anwendung Statistischer Methoden, Springer-Verlag, 2013.

[43] Ze-hua Liu, Yoshinori Kanjo, Satoshi Mizutani, Removal mechanisms for endocrine disrupting compounds (edcs) in wastewater treatment—physical means, biodegradation, and chemical advanced oxidation: a review, Sci. Total Environ. 407 (2) (2009) 731–748.

[44] Barbora Jarošová, Luděk Bláha, John P. Giesy, Klára Hilscherová, What level of estrogenic activity determined by in vitro assays in municipal waste waters can be considered as safe? Environ. Int. 64 (2014) 98–109.

[45] R. Loos, Analytical Methods Relevant to the European Commission's 2012 Proposal on Priority Substances under the Water Framework Directive, JRC scientific and policy report, 2012.

[46] Tom Benijts, Riet Dams, Willy Lambert, André De Leenheer, Countering matrix effects in environmental liquid chromatography–electrospray ionization tandem mass spectrometry water analysis for endocrine disrupting chemicals, J. Chromatogr. A 1029 (1) (2004) 153–159.

[47] Y.K.K. Koh, T.Y. Chiu, A. Boobis, E. Cartmell, J.N. Lester, M.D. Scrimshaw, Determination of steroid estrogens in wastewater by high performance liquid chromatography–tandem mass spectrometry, J. Chromatogr. A 1173 (1) (2007) 81–87.

[48] EC Directive, 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, Off. J. Eur. Union L 201 (2009) 36.

[49] Sándor Görög, Advances in the analysis of steroid hormone drugs in pharmaceuticals and environmental samples (2004–2010), J. Pharmaceut. Biomed. Anal. 55 (4) (2011) 728–743.
[50] Helena Tomšíková, Jana Aufartová, Petr Solich, Lucie Nováková, Zoraida Sosa-Ferrera, José Juan Santana-Rodríguez, High-sensitivity analysis of female-steroid hormones in environmental samples, Trac. Trends Anal. Chem. 34 (2012) 35–58.
[51] Carolina Di Paolo, Richard Ottermanns, Steffen Keiter, Selim Ait-Aissa, Kerstin Bluhm, Werner Brack, Magnus Breitholtz, Sebastian Buchinger, Mario Carere, Carole Chalon, 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. 104 (2016) 473–484.
[52] Robert Kase, Barbora Javurkova, Eszter Simon, Kees Swart, Sebastian Buchinger, Sarah Könemann, Beate Escher, Mario Carere, Valeria Dulio, Selim Ait-Aissa, Henner Hollert, Sara Valsecchi, Stefano Polesello, Peter Behnisch, Carolina Di Paolo, Daniel Olbrich, Eliska Sychrova, Michael Gundlach, Rita Schlichting, Lomig Leborgne, Manfred Clara, Christoph Scheffknecht, Yves Marneffe, Carole Chalon, Petr Tusil, Premysl Soldan, Brigitte von Danwitz, Julia Schwaiger, Antonio Moran, Francesca Bersani, Olivier Perceval, Cornelia Kienle, Etienne Vermeirssen, Klara Hilscherova, Georg Reifferscheid, Inge Werner, Screening and risk management solutions for steroidal oestrogens in surface and wastewater, Trends Anal. Chem. (2018) (accepted for publication).
[53] Juliette Legler, Martine Dennekamp, A. Dick Vethaak, Abraham Brouwer, Jan H. Koeman, Bart van der Burg, Albertinka J. Murk, Detection of estrogenic activity in sediment-associated compounds using in vitro reporter gene assays, Sci. Total Environ. 293 (1) (2002) 69–83.
[54] Albertinka J. Murk, Juliette Legler, Marola MH. Van Lipzig, John HN. Meerman, Angelique C. Belfroid, Albertus Spenkelink, Bart Van Der Burg, Gerard BJ. Rijs, Dick Vethaak, Detection of estrogenic potency in wastewater and surface water with three in vitro bioassays, Environ. Toxicol. Chem. 21 (1) (2002) 16–23.
[55] Frederic DL. Leusch, Tools to Detect Estrogenic Activity in Environmental Waters, 2008 [cited 2018 February 5th]; Available from: http://www.waterrf.org/resources/Lists/SpecialReports/Attachments/2/GWRC_EDC_ToolsToDetect.pdf.
[56] Werner Brack, Valeria Dulio, Marlene Ågerstrand, Ian Allan, Rolf Altenburger, Markus Brinkmann, Dirk Bunke, Robert M. Burgess, Ian Cousins, Beate I. Escher, 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. 576 (2017) 720–737.
[57] Ian J. Allan, Branislav Vrana, Richard Greenwood, Graham A. Mills, Benoit Roig, Catherine Gonzalez, A “toolbox” for biological and chemical monitoring requirements for the european union's water framework directive, Talanta 69 (2) (2006) 302–322.

ANNEXE C. KASE ET AL. (2018) SCREENING AND RISK MANAGEMENT SOLUTIONS FOR STEROIDAL ESTROGENS IN SURFACE AND WASTEWATER. *TRENDS IN ANALYTICAL CHEMISTRY* (16 PAGES)

Trends in Analytical Chemistry 102 (2018) 343–358

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Screening and risk management solutions for steroidal estrogens in surface and wastewater

Robert Kase [a, *], Barbora Javurkova [b], Eszter Simon [a], Kees Swart [c], Sebastian Buchinger [d], Sarah Könemann [a, e], Beate I. Escher [f, g], Mario Carere [h], Valeria Dulio [i], Selim Ait-Aissa [i], Henner Hollert [e], Sara Valsecchi [j], Stefano Polesello [j], Peter Behnisch [c], Carolina di Paolo [e], Daniel Olbrich [a], Eliska Sychrova [b], Michael Gundlach [e], Rita Schlichting [f, g], Lomig Leborgne [k], Manfred Clara [l], Christoph Scheffknecht [m], Yves Marneffe [n], Carole Chalon [n], Petr Tusil [o], Premysl Soldan [o], Brigitte von Danwitz [p], Julia Schwaiger [q], Antonio Moran Palao [r], Francesca Bersani [s], Olivier Perceval [t], Cornelia Kienle [a], Etienne Vermeirssen [a], Klara Hilscherova [b], Georg Reifferscheid [d], Inge Werner [a]

[a] Swiss Centre for Applied Ecotoxicology, Eawag-EPFL, Ueberlandstrasse 131, 8600 Dübendorf, Switzerland
[b] Research Centre for Toxic Compounds in the Environment (RECETOX), Kamenice 753/5, Pavillion A29, 62500 Brno, Czech Republic
[c] BioDetection Systems b.v., Science Park 406, 1098 XH Amsterdam, Netherlands
[d] Bundesanstalt für Gewässerkunde, Am Mainzer Tor 1, 56068 Koblenz, Germany
[e] Institute for Environmental Research, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany
[f] Helmholtz Centre for Environmental Research (UFZ), Permoserstrasse 15, 04318 Leipzig, Germany
[g] Eberhard Karls University Tuebingen, Hoelderlinstr. 12, 72074 Tuebingen, Germany
[h] Italian Institute of Health, Department Environment and Health, Roma, Italy
[i] INERIS, Rue Jaques Taffanel, Parc Technologique ALATA, 60550 Verneuil-en-Halatte, France
[j] Water Research Institute - National Research Council of Italy, Istituto di Ricerca sulle Acque, Via del Mulino 19, 20861 Brugherio (MB), Italy
[k] Agence de l'eau Adour-Garonne, 90 rue de Fétéra, CS 87801, 31078 Toulouse Cedex 4, France
[l] Umweltbundesamt, Spittelauer Lände 5, 1090 Wien, Austria
[m] Umweltinstitut, Institut für Umwelt und Lebensmittelsicherheit des Landes Vorarlberg, Montfortstraße 4, 6901 Bregenz, Austria
[n] Institut Scientifique de Service Public (ISSeP), Rue Chéra 200, 4000 Liège, Belgium
[o] T.G. Masaryk Water Research Institute, Podbabská 2582/30, Praha 6, 16000, Czech Republic
[p] Landesamt für Natur, Umwelt und Verbraucherschutz NRW (LANUV), Auf dem Draap 25, 40221 Düsseldorf, Germany
[q] Bayerisches Landesamt für Umwelt, Demollstr. 32, 82407 Wielenbach, Germany
[r] Instituto de Recursos Naturales, Universidad de León, Avenida de Portugal 42, 24071 León, Spain
[s] Centro Ricerche SMAT, Società Metropolitana Acque Torino S.p.A., C.so Unità d'Italia 235/3, 10127 Torino, Italy
[t] The French Agency for Biodiversity, 94300 Vincennes, Paris, France

ARTICLE INFO

Article history:
Available online 8 March 2018

Keywords:
Effect-based methods
Science-policy interface
Oestrogen screening
Endocrine disruption
Surface and wastewater assessment
Risk-based trigger values
EU watch-list
Water framework directive
Chemical monitoring of Emerging pollutants

ABSTRACT

Background: The European Commission Implementing Decision EU 2015/495 included three steroidal estrogens, namely 17α-ethinyl estradiol, 17β-estradiol, and estrone, in the so-called "watch list" of the EU Water Framework Directive (WFD). The monitoring of these compounds is difficult because the detection limits of the majority of the available analytical methods cannot achieve the very low target concentrations required to meet proposed environmental quality criteria. In 2014, a combined Science-Policy Interface/Chemical Monitoring of Emerging Pollutants project was launched to meet this monitoring challenge. The project involved 24 research organizations and environmental agencies from 12 different countries.

Methods: Sixteen surface water (SW) and 17 wastewater (WW) samples were collected across Europe and analysed using five *in vitro* effect-based and three chemical analytical methods. A general description of the project and data evaluation is provided by Könemann and colleagues in the companion publication 2018. In our study, we compared bioanalytical and chemical analytical results with regard to their application for aquatic status assessment. Therefore we considered the potential to predict population-relevant risks for aquatic organisms and the specificity and sensitivity of the various methods used in

* Corresponding author. Swiss Centre for Applied Ecotoxicology Eawag-EPFL, Überlandstrasse 133, 8600 Dübendorf, Switzerland.
E-mail addresses: Robert.Kase@oekotoxzentrum.ch, kase.robert@googlemail.com (R. Kase).

https://doi.org/10.1016/j.trac.2018.02.013

both approaches. Finally, we tested and discussed the applicability and relevance of previously suggested effect-based trigger values (EBT).
Results and discussion: Results of the risk assessment based on chemical analytical data correlated highly with estrogenic activities (expressed as 17β-estradiol equivalents (EEQ) determined using effect-based methods), demonstrating the ability of the bioassays to predict the mixture risk caused by steroidal estrogens. For about 15% of SW and 40% of WW samples detection limits of chemical-analytical methods were too high to allow a status assessment, while detection limits of all effect-based methods were below proposed EBT. This demonstrates that effect-based methods are suitable for status assessment of surface waters. The *in vitro* effect-based methods were quite specific for steroidal estrogens and highly sensitive, meaning they have a low probability to detect false positive or negative results. After testing of three alternative EBT proposals, we concluded to use preliminary 400 pg/L EEQ as screening EBT corresponding to the AA-EQS of E2. Further test specific refinements in the application of this value are not excluded.
Conclusions: We conclude that water quality assessment can progress from a purely analytical approach to effect-based monitoring, from single substance to known and unknown mixture assessment and from *in vitro* screening to population-relevant risk assessment. Despite a few limitations, effect-based *in vitro* methods are recommendable for monitoring steroidal estrogens under the WFD because they, a) are able to sensitively quantify the activity of steroidal estrogens in all surface and wastewater samples, b) are able to detect the combined effect of estrogen mixtures including unknown chemicals with estrogen receptor activating properties, c) allow an ecotoxicological status assessment using EBT to calculate risk quotients. This approach is similar to the risk ratio used in regulatory environmental risk assessments, but allows for an integrated mixture assessment.
Outlook: The results of this study support the introduction of a holistic approach for the regulation of chemicals in the aquatic environment under the EU WFD, as proposed recently by EU Water Directors. An ecotoxicological status assessment for one of the most relevant modes of action of endocrine disruption will allow authorities responsible for water quality assessment to focus available monitoring resources and to improve the ecological status of EU waterbodies.

1. Introduction

1.1. State of the art and current regulation

Steroidal estrogens, which are commonly present in wastewaters (WW) and surface waters (SW) (e.g. Ref. [1]), can cause reproductive toxicity to aquatic biota, especially to fish [2–6]. The most potent steroidal estrogens: the synthetic hormone 17α-ethinyl estradiol (EE2), the natural hormones 17β-estradiol (E2), and estrone (E1) the main transformation product of both, were recently included in the European Union's watch list [7–9] of substances for monitoring in SW. This watch-list mechanism is designed to allow target EU-wide monitoring of substances of possible concern to support the prioritization process in future reviews of the priority substances list under the Water Framework Directive (WFD). According to the WFD, the good chemical status of a waterbody is reached when all substances included in the priority list are detected below their defined Environmental Quality Standards (EQS) that are based on annual average and maximum allowable concentrations. The previously proposed and not regulatory binding annual average Environmental Quality Standards (AA-EQS) for E1, E2 and EE2 are as low as 3600 pg/L, 400 pg/L and 35 pg/L, respectively [10,11]. So far, only a very limited number of institutes from both the public and the private sector, and major environmental agencies in Europe have developed capacities to quantify the steroidal estrogens at their EQS levels. Such low EQS, together with complexity of matrices such as SW and WW, and the instability of some of the analytes, make the monitoring of these compounds under the watch-list mechanism of the WFD and national surveillance schemes difficult.

Estrogenic adverse effects in environmental watersare generally caused by mixtures of different estrogenic chemicals including metabolites, which increases their risk to wildlife [12]. A recent study demonstrated that mixtures of priority pollutants present at and below their individual EQS concentrations can cause relevant biological effects and may pose significant risks to wild species and ecosystems in spite of the fact that individual chemicals were at concentrations in compliance with regulations [13]. Mixture toxicity has also been highlighted in the context of the European strategy on endocrine disrupting chemicals [14]. The European Commission recently acknowledged the need to develop and implement methodologies for the identification of chemical mixtures of potential concern and for the assessment of their impacts on both environmental and human health [15]. Such methodologies should help to link the knowledge of chemical contamination and the observation of adverse effects on and via the aquatic environment. For all described reasons (risks at low concentrations, difficulty and costs of high-end analytics, mixture toxicity, and linking chemical contamination to ecological status) alternative monitoring and risk assessment methods are urgently needed.

Effect-based methods detect cumulative effects and are useful to bridge the gap between chemical contamination and ecological status [16,17]. Complementary, mechanism-specific bioassays can provide information on modes of action (MoA) that are intrinsically of concern for ecosystems and human health [18]. Focusing on estrogenic effects, *in vitro* bioassays, can detect an activation of estrogenic receptor(s) (ER) by mixtures of estrogens and xenoestrogens: In parallel, they can detect single analytes at sufficiently low concentration levels [19]. The response of the assays is expressed as E2-equivalent (EEQ) values. The applicability of this approach has been demonstrated in different projects during the past decade e.g. Refs. [20–23]. Also the combination of effect-based and chemical analytical monitoring to identify and assess the risks from steroidal estrogens has been discussed and proposed as a potential tool for WFD monitoring [24–29].

1.2. Current trends

The WFD also involves evaluation of the ecological status of waterbodies, e.g. monitoring of biological diversity. Consequently, severe adverse effects at the population level of aquatic organisms should be captured, but there are currently no tools under WFD implemented to monitor Endocrine Disruption efficiently. The current proposal for the WFD [26–29a,b] is to screen water samples by *in vitro* assays for estrogenic activity and subsequently target the more demanding chemical monitoring on a reduced number of samples that show up as positive in the bioanalytical screening, as was previously elaborated by two international workshops in 2013 and 2017 [29a,b]. The water status evaluated with *in vitro* bioassays would be newly called an ecotoxicological status for risk from ER-mediated effects and is intended to improve chemical and ecological status assessment for one of the most relevant modes of action of endocrine disruption and fish reproduction toxicity [4]. As described, numerous studies recommend the use of these effect-based methods. However, clear recommendations for water managers regarding the use of *in vitro* methods, especially as regards harmonized data evaluation and effect-based trigger values (EBT) distinguishing “acceptable” and “not acceptable” water quality are still missing.

Moreover, these methods with their clearly defined EBT are also necessary for wastewater management. Wastewaters or Wastewater Treatment Plants (WWTPs) effluents are the dominant sources of steroidal estrogens in waterbodies, e.g. Refs. [30,31]. Therefore, although levels of estrogens in WW and WWTPs effluents are not directly regulated by the WFD, their monitoring is necessary to quantify the loads emitted into waterbodies and to show reduction of risks after having applied proper mitigation actions. Various options for the removal of pharmaceuticals and hormones by WWTPs were recently reviewed [24,32]. Although elimination rates of estrogenicity (caused mainly by steroidal estrogens) at conventional WWTPs with tertiary treatment are high (usually from about 80% to more than 90%, [33,34]) the residual activity in discharges can still represent a risk for aquatic biota and remains, besides untreated WW, their major known source [35]. Advanced treatment steps such as e.g. ozonation, UV treatment etc. can further eliminate well above 90% of steroidal estrogens [32]. Under such circumstances, concentrations in surface waters cannot be measured by most currently available chemical analytical methods due to matrix effects [36].

To test if *in vitro* assays may be suited for regulatory monitoring and risk assessment of low levels of estrogens in both WW and SW, 24 research organizations and environmental agencies from 12 different European countries joined forces in a project, which also supports the activities of the Working Group “Chemicals” under the Common Implementation Strategy (CIS) for the WFD and as a follow-up to their “Science-Policy Interface” activity [17,37]. The project results are summarized in two publications: the general description of the project, experimental details and recommendations on harmonized *in vitro* data evaluation, are described in a companion paper by Könemann et al. [36]. The current publication focuses on testing the most useful EEQ EBT for screening and discussion about environmental risk for SW and WW. A demonstration of applicability of the proposed EBT value to 16 SW and 17 WW samples collected across Europe is also presented here. The samples represented a gradient from low polluted to highly polluted samples and were analysed by sensitive HPLC-MS/MS methods at three different institutes and by five different *in vitro* assays at five different institutes in order to:

- Evaluate the applicability and relevance of the *in vitro* methods for the monitoring of steroidal estrogens with reference to the classification of the chemical status of waterbodies
- Discuss and recommend the extent to which the bioassays can be used for screening and prioritizing environmental samples, while considering risks for aquatic organisms
- Propose suitable effect-based trigger values (EBT) for screening and discriminating between unpolluted and polluted samples with the aim of classifying waterbodies.
- Contribute towards the review process of WFD and to integrated effect-based-methods (EBM) into regulation

Within this project, we aimed to bridge the gap between conventional analytical and effect-based monitoring and risk assessment for steroidal estrogens.

2. Methods

Sampling, sample preparation, positive and negative controls, and chemical and biological analyses methods, are described in detail in Ref. [36] and are briefly summarized here.

2.1. Samples, sample preparation, in vitro and chemical analyses

A total of 16 SW and 17 WW samples with 11 L sampling volume were collected by 10 participating institutes at sites expected to be polluted from Austria, Belgium, Czech Republic, Germany, France, Italy and Spain. The samples were frozen within 48 h and sent to Bio Detection Systems (BDS), Amsterdam, The Netherlands. Samples were filtered (see Könemann et al., 2018) and subsequently extracted by using solid phase extraction with C18 cartridges (Phenomenex Strata C18-E, 55 μm, 70 Å, 500 mg/6 mL). Additional silica gel clean-up was applied to the extracts to reduce matrix effects and reach detection limits in the sub ng/L range in the chemical analysis. For some wastewater samples, a single silica gel (SiOH) column was inadvertently used to treat the entire sample extract (11 mL), while for each surface water sample, extract was split into eleven 1 mL aliquots for clean-up. Extracts were then homogenized, divided into 1 mL aliquots, and sent to Federal Institute of Hydrology (BfG), Institut National de l'Environnement Industriel et des Risques (INERIS), Research Centre for Toxic Compounds in the Environment (RECETOX), and Helmholtz Centre for Environmental Research (UFZ), where analyses by five *in vitro* effect-based methods were performed: ER-Calux (at BDS), pYES, MELN, HeLa 9903, and ER-GeneBlazer and to Joint Research Centre (JRC), BfG, and Swiss Centre for Applied Ecotoxicology (Ecotox Centre, EC) performing HPLC MS/MS analysis. EC used an additional silica gel clean-up for 3 of 17 WW samples prior to the chemical analysis. More detailed information on methods is available in the companion publication Könemann et al., 2018 [36]. All data were analysed centrally in a harmonized way.

2.2. Chemical analytical data evaluation and compliance assessment

Measured concentrations for E1, E2, and EE2 were expressed in pg/L. Measurements below LOQ, but above LOD were indicated as < LOQ. Measurements below LOD were indicated as < LOD (SI Tables 1–3).

The measurements were compared with EQS proposals (AA-EQS EE2 = 35 pg/L, AA-EQS E2 = 400 pg/L, AA-EQS E1 = 3600 pg/L) in order to assess potential compliance. Because the European Commission did not propose an EQS for E1 [10] the Swiss EQS proposal for E1 of 3600 pg/L [11] was used. The compliance of samples was set to 0 if the measurement result exceeded an EQS proposal. Compliance was designated as “not assessable” if the results were below LOD, with an LOD (LOQ/3) above the EQS proposal. If EQS proposals were not exceeded and LODs were below the proposed

EQS, the compliance assessment was set to 1 (compliant). For comparison reasons, the results for WW samples were treated the same way as the SW samples, even though WW discharges are usually diluted by the receiving waters. Only if concentrations of estrogens in the flow of the receiving waterbody and the effluent are known, dilution factors can be applied to calculate a final estrogenicity (Equation (8) in chapter 3.4.7.2) and the receiving water can be assessed as compliant even though the WW could be "non-compliant". On the other hand in our project the concentrations of estrogens in the receiving waters were not measured and the precautionary principle (to set the same quality requirement for WW and SW) was applied in order to avoid the risk of a false negative assessment (extrapolation of compliance in case of non-compliance) for the receiving waterbody. Our approach intends to stimulate the consideration and measurements of background concentrations without only focusing on the estrogenicity of the WW discharges. The overall compliance corresponds to the rounded average compliance to 0 or 1 of all three analytical assessments.

2.3. Effect-based trigger values (EBT)

In recently published studies, a narrow concentration range of published EBT values was proposed, although different approaches were applied to derive them. The following shortly described EBT were tested and discussed in our study in section 3.4.4:

Jarošová et al. [33] derived "safe environmental concentrations" of EEQ in municipal wastewater effluents, based on a simplified assumption that mainly (>90%) steroidal estrogens are causing ER-mediated estrogenicity. These potentially safe concentrations were derived using the estrogenic relative potencies in bioassays, the *in vivo* predicted no-effect concentrations of the compounds, and their relative contributions to the measured EEQ of WW effluents. The predicted safe concentrations ranged from 100 to 400 pg/L EEQ with a median EBT of 300 pg/L EEQ. We used in our study the median EBT because we worked with five different bioassays which we intended to characterize with alternative EBT proposals.

Van der Oost at al [38]. used bioanalytical equivalents (BEQ) of selected substances that trigger the bioassay and a background BEQ to derive an EBT of 500 pg/L EEQ. The background BEQ was calculated with 60 pg EEQ/L, and a safe BEQ (based on lowest NOEC of triggering substances) was 7 pg/L EEQ. The finally proposed EBT (500 pg/L EEQ) was mainly based on a BEQ-converted species sensitivity distribution (SSD) that provided the concentration that is a potential hazard for 5% of aquatic species (HC_5 BEQ = 500 pg/L EEQ).

Kase et al. and Kunz et al. [25,26] proposed to use the proposed AA-EQS of E2 as an EBT for estrogen receptor mediated estrogenic activity, thus proposed 400 pg/L EEQ as EBT. This was done primarily for different reasons: a) The EBT is compatible with the EU AA-EQS proposal for E2 which is based on fish toxicity SSD for population relevant effects 400 pg/L EEQ, b) E2 is a natural steroid hormone and has an *in vitro* and *in vivo* potency between E1 and EE2, therefore it is likely better suited than E1 or EE2 for assessing mixture effects, c) E2-equivalents are commonly used in bioanalysis and biomonitoring, thus data are easily comparable with previous studies, d) EE2 has a slightly higher potency *in vitro* than E2, but *in vivo* it is 10–20 times more potent. If EE2 equivalents were to be used, there is a high probability for risk overestimation and obtaining false positive results, due to the possibility of E2 and E1 playing a more prominent role. e) Steroidal estrogens normally occur as a mixture in WW and in receiving waterbodies. Jarosova and colleagues [33] compiled data of 353 wastewater measurements from three studies with a median concentration of 7–12 ng/L E1, 1.3–1.7 ng/L E2 and 0.47–0.6 ng/L EE2. Based on relative potencies of used bioassays this means that the *in vitro* ER mediated mixture effect is likely dominated by E2. Therefore as a simplified approach the use of 400 pg/L EEQ as EBT seem to be arbitrary, however this can be justified and was successfully tested in this study.

2.4. Effect-based compliance assessment and validation status of methods

The measured EEQs were compared with EBTs to assess the potential compliance of samples. The compliance of samples was set to 0 ("non-compliant") if the measured EEQ exceeded an EBT. The compliance assessment was set to "not assessable" if LOQ or LOD were above the EBT, but this never occurred during the measurements. If EBT were not exceeded and LOD were below EBT the compliance was set to 1 ("compliant"). The overall compliance corresponds to the rounded average compliance to 0 or 1 of all five effect-based assessments. After EBT discussion in section 3.4.4 we used the 400 pg/L EEQ as preliminary EBT for compliance assessment.

Validation activities: Two of our five *in vitro* assays used in this study are currently being OECD validated (HeLa 9903 and ER-Calux [39]) and the ER-Calux, A-YES and L-YES are DIN/EN/ISO standardized in 2018. The ER-ER-GeneBlazer is used in the US within the Tox21 program of the National Institute of Health and US Environmental Protection Agency [40].

2.5. Sensitivity and specificity analysis for methods in compliance assessments

Sensitivity and specificity were determined for both chemical and effect-based techniques. This was done to evaluate their suitability for European WFD monitoring programs. High sensitivity means that the method is less prone to detect false negatives, in other words the method has a low risk of erroneous compliant assessments. While high specificity means that the method is less prone to detect false positives, in other words the method has a low risk of erroneous non-compliant assessments.

In order to assess the sensitivity and specificity of three chemical analytical and five *in vitro* effect-based methods, each method was compared with the overall compliance derived from the chemical or effect-based analyses (see section 2.2 and 2.4)

Four options regarding conformity in the results of the comparison of the specific method with the overall compliance were possible:

- Overall compliance is 1 and the compliance of the specific method is 1, this means "conformity in compliance ":= CC
- Overall compliance is 1 and the compliance of the specific method is 0, this means "non-conformity in compliance":= NCC
- Overall compliance is 0 and the compliance of the specific method are 0, this means "conformity in non-compliance":=CNC
- Overall compliance is 0 and the compliance of the specific method is 1, this means "non-conformity in non-compliance":=NCNC

Based on the conformity rating of each method sensitivity and specificity was calculated according to Equations (1) and (2).

$$\text{Sensitivity } [\%] = \frac{\sum \text{CNC}}{(\sum \text{CNC} + \sum \text{NCC})} * 100 \qquad (1)$$

$$\text{Specificity } [\%] = \frac{\sum CC}{(\sum CC + \sum NCNC)} * 100 \quad (2)$$

2.6. Risk quotient and -scenario calculations for chemical and ecotoxicological status assessments

The chemical analytical Risk Quotient (RQ_{chem}) was calculated by dividing the Measured Environmental Concentration (MEC) by the proposed EQS (Equation (3)). The biological RQ_{bio} was calculated by dividing the measured Bioanalytical Equivalent Concentration (BEQ), in our case Estradiol Equivalent Concentration (EEQ) by the effect-based trigger value (EBT) (Equation (4)). RQ > 1 signifies an unacceptable risk for aquatic organisms.

$$RQ_{chem} = \frac{\text{Measured Environmental Concentration (MEC)}}{\text{Proposed EQS}} \quad (3)$$

$$RQ_{effect-based} = \frac{EEQ}{\text{Proposed EBT}} \quad (4)$$

The chemical analytical RQ is used for chemical status assessment and the effect-based RQ for an eco-toxicological status assessment. After EBT discussion in section 3.4.4 we used the 0.4 ng/L EEQ as preliminary EBT. Equation (3) for single analytes, can be adapted to mixture effects of multiple measured substances with the same MoA, via calculation of cumulative RQs, according to Kortenkamp 2007 [41]. Kortenkamp proposed the concentration addition concept as an accurate approach for regulatory use if EDC have the same MoA (Equation (5)).

$$\sum RQ_{E1,\,E2,\,EE2} = \frac{\text{MEC E1}}{3600\ \text{pg/L}} + \frac{\text{MEC E2}}{400\ \text{pg/L}} + \frac{\text{MEC EE2}}{35\ \text{pg/L}} \quad (5)$$

The application of this mixture concept is well supported by additional evidence for endocrine disruptors and other relevant mixtures [42,43]. Moreover, the equation can be further improved to consider unknown unquantified risks by taking into account the specific LODs and LOQs of chemical analytical methods, i.e. by setting concentrations of samples with non-detectable analytes either to 0, LOD/2 of LOD. Three cumulative risk scenarios were calculated in this way to derive the minimum known, the likely, and the maximal risks of steroidal estrogens in the samples:

Mixture risk scenarios for chemical analytical methods:

1) Minimal cumulative risk scenario: $\sum RQ_{EE2,\,E2,\,E1} = \sum (MEC_{EE2,\,E2,\,E1}/AA\text{-}EQS_{EE2,\,E2,\,E1})$
2) Medium cumulative risk scenario: $\sum RQ_{EE2,\,E2,\,E1} = \sum (MEC_{EE2,\,E2,\,E1}$ or $LOD/2_{EE2,\,E2,\,E1}/AA\text{-}EQS_{EE2,\,E2,\,E1})$
3) Maximal cumulative risk scenario: $\sum RQ_{EE2,\,E2,\,E1} = \sum (MEC_{EE2,\,E2,\,E1}$ or $LOD_{EE2,\,E2,\,E1}/AA\text{-}EQS_{EE2,\,E2,\,E1})$

The cumulative RQs for these three risk scenarios are based on the chemical measurements (SI Tables 1–3).

2.7. Correlation analysis of risk quotients and EEQ measurements

We compared the chemical analytical mixture risk scenarios with effect-based biological responses. Cumulative RQs of the minimum and maximum risk scenarios were plotted on a logarithmic scale against the biological EEQ (BEQ) responses and a log-linear regression line ($y = ax^b$) was calculated. Data were tested for log-linearity (scatter-plot), constant variance (TA-plot) and normality (Q-Q plot). Moreover, a double-sided correlation analysis significance test was performed with the Pearson correlation coefficient at $p \leq 0.0001$; $p \leq 0.001$, $p \leq 0.01$ and $p \leq 0.05$ [44]. More specific p-values and confidence intervals were calculated with graph pad 5 using a two-tailed column Pearson normality test after normality check.

2.8. Calculation of a risk indication score (RIS) and screening score for effect-based methods

To measure how precisely a biological response indicates a population relevant mixture risk, we calculated a Risk Indication Score (RIS). A chemical analytical cumulative $RQ_{mix} > 1$ (Equation (5) above) for estrogens indicates an "unacceptable" risk for aquatic organisms and their populations. This is mainly the case for fish species as, based on current knowledge, they include the most sensitive species for estrogenic effects and were used to derive the EQS. RQmix was compared to exceedances of EEQ measured by effect based methods of different EBT (see chapter 2.3) ranging from 300 to 500 pg/L.

There were two possible outcomes:

- If the cumulative chemical $RQ_{E1E2EE2}$ was >1 and the respective EBT was exceeded by the biological response, it was counted as successful risk indication.
- If the cumulative chemical $RQ_{E1E2EE2}$ was >1 and the respective EBT was not exceeded by the biological response, it was counted as failed risk indication.

The number of successful risk indications was scored and normalized to the maximal number of possible risk indications by calculation of the RIS (Equation (6)).

$$RIS\ [\%] = \frac{(\sum \#RQ_{E1,\,E2,\,EE2} > 1 \text{ and BEQ} > \text{EBT}) x100}{\sum \#RQ_{E1,\,E2,\,EE2} > 1} \quad (6)$$

In a few cases, an EBT exceedance was observed where cumulative chemical $RQ_{E1E2EE2}$ was <1. This was scored as "ncr (no chemical risk indication, but positive biological response) and an "ncr* screening score" was calculated (Equation (7)).

$$ncr^*\ [\%] = \frac{(\sum \#RQ_{E1,\,E2,\,EE2} < 1 \text{ and BEQ} > \text{EBT}) x100}{\sum \#RQ_{E1,\,E2,\,EE2} > 1} \quad (7)$$

RIS and ncr* score were calculated using three proposed EBTs. The two parameters identify the specificity of effect-based methods to predict risks caused by steroidal estrogens and their potential as screening methods. This screening allows to detect additional risks caused by estrogen receptor activating substances other than steroidal estrogens or where chemical analysis was not able to quantify steroidal estrogens due to high LOQs.

3. Results and discussion

3.1. Compliance assessment with chemical analytical methods

Here we describe the results of our compliance assessment based on chemical analytical data. Results of the chemical analysis of surface water and effluent samples tested in this study including LOD and LOQ values are provided in SI Tables 1–3 and by Könemann et al. [36].

On average, SW water samples had a much higher percentage of compliant samples compared to WW samples (54% SW vs. 12% WW, Table 1). Due to matrix effects and associated higher LOD/

Table 1
Chemical analytical compliance assessments of 16 surface water (SW) samples and 17 wastewater (WW) samples, which were analysed by HPLC MS/MS methods. Compliance frequency is shown as percentage, arithmetic mean and coefficient of variation (CV). Analytical data is provided in SI Tables 1–3.

	Lab1 [%]	Lab2 [%]	Lab3 [%]	Mean ± CV [%]
SW compliant samples	44	63	56	54.2 ± 9.5
SW non-compliant samples	31	25	38	31.3 ± 6.3
SW not assessable samples	25	12	6	14.6 ± 9.5
WW compliant samples	18	18	0	11.8 ± 10.2
WW non-compliant samples	53	47	41	47.1 ± 5.9
WW not assessable samples	29	35	59	41.2 ± 15.6

LOQs, a much higher number of WW samples (41%) could not be assessed compared to SW samples (15%). For SW and WW sample assessments, the CV is in the range of 6–10%. Due to insufficient silica gel clean-up of WW samples, matrix effects were not reduced in an optimal way. Under ideal conditions, analytical methods can achieve LODs and LOQs of a factor 2 to 3 lower in WW samples. It has to be recognized that LODs of chemical analytical methods for steroidal estrogens have been lowered significantly since 2013 and are likely to decrease further [25,29]. The differences in sample assessment ("compliant" vs. "non-compliant" vs. "not assessable") among the three analytical methods were far more frequently caused by differences in LOQs rather than by differences in detected concentrations (SI Tables 1–3). The percentage of "not assessable" samples (up to 25% in SW and 59% in WW) confirmed the existence of challenges for the chemical monitoring of estrogens E2 and especially EE2. Moreover, if the strict requirements of compliance assessment currently applied to monitoring of priority substances under the WFD (i.e. the LOQ should be 1/3 of the EQS [45]), none of the results obtained in this study could be considered acceptable for compliance assessment of the chemical status of surface waters.

3.2. Ecotoxicological status assessment with in vitro effect-based methods

Bioanalytical equivalent concentrations (BEQ), or EEQ for estrogenicity, are a measure of the effect caused by mixtures of unknown and potentially unidentified chemicals expressed as the equivalent concentration of a known reference compound that would elicit the same effect as the sample [46]. Effect-based trigger values (EBT) for *in vitro* bioassays can be derived by combining calculated or measured BEQs of selected substances that trigger a specific effect, with a benchmark approach using known chemical, toxicological and biological data [38]. Similar to conventional risk assessment using chemical concentration and EQS, exceedance of effect-specific EBTs indicates an elevated, unacceptable risk (hazard & exposure) for the aquatic ecosystem due to chemicals with a particular MoA such as estrogenicity. Measured BEQs below an EBT indicate a low and acceptable ecological risk.

Similar to the results of chemical methods, a greater percentage of SW samples (61%) were "compliant" compared to WW samples (39%) (Table 2). The matrix effects did not lead to any "not assessable" categorization and ER mediated estrogenicity was quantified in all samples. Results obtained using different effect-based methods showed good agreement in average status assessments with CVs of 10–11% in SW and WW samples.

3.3. Comparison of status assessment by chemical analytical and in vitro effect-based methods

The *in vitro* effect-based methods were less matrix-dependent and provided generally lower LOQs than chemical analytical methods (Table 3). As it was mentioned before not all matrix effects were optimally removed during WW samples extraction. All results obtained with *in vitro* effect-based methods allowed a risk assessment because the LOQ was below the target EBT values for all samples. Status assessment with chemical analytical and *in vitro* effect-based methods showed overlapping values and means. The main difference in chemical analytical detection methods is the single substance based approach which can be matrix dependent. *In vitro* effect-based methods allow an integrative activity measurement of all ER activating substances and are not matrix independent as long no cytotoxic and ant-estrogenic effects occur, which was not the case in our samples. Therefore the different responses and LOQs of both method can be compared, a detailed discussion is available in Könemann et al. [36].

3.4. Comparison of chemical analytical and in vitro effect-based methods

3.4.1. Mixture risk scenarios of chemical analytical measurements

Since the main MoA of E1, E2, and EE2 is activation of ER, their interactions in environmental mixtures are most probably concentration additive [36]. The risk quotient addition model, which is derived from the mixture model of concentration addition, was shown to be sufficiently accurate for regulatory use [41]. Therefore, the sum of individual RQs of E1, E2, and EE2, represents the combined risk for the mixture of E1, E2, and EE2 and is called cumulative RQ. To show full ranges of potential RQs, the minimal, the medium, and maximal cumulative RQs were thus determined for each sample.

Depending on the mixture risk scenario, the sum of RQ and mean percentage of samples at unacceptable risk increased (Table 4). The minimal cumulative mixture risk scenario showed in 7 SW and 9 WW samples RQs above 1, whereas 9 SW and all WW samples presented an unacceptable risk in the medium mixture risk scenario. The maximal possible mixture risk scenario showed 10 out of 16 SW samples and all WW samples RQs above 1. A dilution factor was not taken into account for the WW samples (see chapter 3.5.2).

Table 2
In vitro effect-based (ecotoxicological) compliance assessments of 16 surface water (SW) and 17 wastewater (WW) samples using an EBT of 400 pg/L EEQ. Results of compliance assessments are shown as percentage, arithmetic mean, and coefficient of variation (CV). EEQ concentrations are provided in SI Table 4,5.

	ER-Calux [%]	p-YES [%]	MELN [%]	HeLa 9903 [%]	ER-Gene-Blazer [%]	Mean ± CV [%]
SW compliant samples	62	50	50	75	69	61.3 ± 11.2
SW non-compliant samples	38	50	50	25	31	38.8 ± 11.2
SW not assessable samples	0	0	0	0	0	0.0
WW compliant samples	47	35	24	47	35	37.6 ± 9.8
WW non-compliant samples	53	65	77	53	65	62.4 ± 9.8
WW not assessable samples	0	0	0	0	0	0.0

Table 3
LOQ comparison of chemical analytical (LOQ for E2) and *in vitro* effect-based (EEQ corresponding to E2 equivalents) measurements of 16 surface water (SW) and 17 wastewater (WW) samples. Data are shown as arithmetic mean, range, and coefficient of variation (CV). The measurements are provided in SI Tables 1–5.

Methods	LOQ_{SW}			LOQ_{WW}		
	Mean ± SD [pg/L]	Min-max range [pg/L]	CV%	Mean ± SD [pg/L]	Min-max range [pg/L]	CV%
3 chemical analytical methods [E2]	181 ± 291	39–1500	161	627 ± 726	50–3000	116
5 *in vitro* effect-based methods [EEQ]	28 ± 33	2–200	116	60 ± 62	1–216	104

Table 4
Average cumulative risk quotients (RQ) for EE2, E2, and E1 based on mean concentrations of three chemical analytical measurements for 16 surface water (SW) and 17 wastewater (WW) samples calculated for minimal, medium and maximal risk scenarios.

Cumulative RQ for SW samples				Cumulative RQ for WW samples			
Sample code	Minimal	Medium	Maximal	Sample code	Minimal	Medium	Maximal
A (11)	0.02	0.41	0.77	A (26)	0.02	1.78	3.53
B (6)	0.02	0.38	0.72	B (29)	0.03	2.01	3.78
C (1)	0.04	0.45	0.8	C (31)	0.04	7.78	15.48
D (22)	0.04	0.41	0.76	D (4)	0.06	3.45	6.29
E (27)	2.27	2.34	2.38	E (17)	3.49	3.25	5.06
F (30)	0.1	1.77	3.37	F (21)	0.12	2.74	5.32
G (32)	0.28	0.58	0.97	G (14)	0.23	11.2	21.21
H (25)	0.17	1.6	2.62	H (5)	0.58	18.41	34.99
I (8)	0.39	1.21	2.06	I (19)	0.8	5.5	9.75
J (10)	0.37	0.66	0.98	J (16)	7.45	14.54	25.5
K (18)	4.99	4.99	4.99	K (9)	1.49	18.3	34.52
L (24)	1	4.18	7.06	L (13)	4.15	6.93	10.07
M(28)	1.41	4.11	6.36	M(23)	219.18	219.1	219.4
N (15)	2.5	4.99	7.26	N (33)	4.26	15.37	25.69
O (3)	10.71	12.11	16.6	O (12)	4.94	17.01	27.51
P (7)	5.17	5.92	8.19	P (2)	5.6	7.85	9.88
				Q (20)	151.6	151.6	151.6
Mean cumulative RQ	1.84	2.88	4.12	Mean cumulative RQ	23.76	29.81	35.85
Percentage of samples presenting an unacceptable risk	44%	63%	63%		53%	100%	100%

3.4.2. Mixture risk scenarios of chemical analytical measurements compared with in vitro effect based methods

We compared the risk derived for the minimum and the maximum mixture risk scenarios based on chemical analytical data with risk indicated by the effect based methods. Figs. 1 and 2 show that the sum of the risk quotients ("mixture risk of steroidal estrogens") derived from chemical measurements were highly correlated with the measured EEQs in the respective samples. Depending on the choice of LOD, LOD/2 or 0 to replace non-detects, the respective cumulative mixture risk estimation (minimal to maximal) can change considerably (by several orders of magnitude) if non-detects occur.

Overall, EEQ measured with the five *in vitro* effect-based methods correctly assigned the "chemical status" of wastewater samples as determined by the sum of the risk quotients for E1, E2, and EE2, with highest EEQ signals detected at sites where EE2 was present at concentrations above the LOQ and thus quantified. In most cases, the fit for maximal RQ was worse than for the minimal scenario indicating that undetected compounds were likely not present in the mixtures and non-detected compounds did not play a significant role. The significant correlations indicate that for our selection of samples (33 samples from seven countries with variable pollution levels) estrogenicity was mainly caused by the steroidal estrogens E1, E2, and EE2 and anti-estrogenicity or other xenoestrogens played a minor role. This is supported by the iceberg modelling presented by Könemann et al. [36]. However, this result can be partially influenced by the choice of more or less specific extraction methods.

Even though our results confirmed that the steroidal estrogens were dominantly responsible for triggering the estrogenic activities measured by the bioassays, an important question is if the assumptions are applicable to all surface and wastewaters. Although most studies comparing steroidal estrogen concentrations with biological activity support the applicability of effect-based methods [13,18,36,38,46,47], there can be some exceptions. In our dataset, the risk for one SW sample was evaluated as unacceptable in all three chemical analyses, whereas not by any effect-based method (Table 5, sample E (27)). In this sample, EE2 was detected in the range of 73–85 pg/L, which indicated an elevated risk (RQ = 2.27), while concentrations of E2 and E1 were very low (Table SI 1–3, sample E (27)). We can explain this by comparing the potential of E1, E2 and EE2 to induce responses *in vitro* and *in vivo*. While E1 and E2 usually trigger *in vitro* responses at similar concentrations as *in vivo* responses, EE2 is 10–20 times more potent *in vivo* than *in vitro* [33,48]. This is reflected in very low EQS of EE2, but in contrast to the relatively high EBT, which integrates the risk of chemical mixtures. The precautionary principle (to work with EE2-equivalents and an EBT of 35 pg/L) cannot be applied here, because then all samples including background samples [26,38,49] would be assessed as presenting an elevated risk. Since the main source of EE2 is its excretion after use of contraceptive pills, however, EE2 is usually present together with other natural steroidal estrogens, and risk is therefore correctly indicated by results of effect based methods and the proposed EBT. Sample E (27) appears to be an exception as other WW samples (Q (20), M(23)) contained relatively high concentrations of EE2 (which contributed to more than 93% of cumulative RQ) and high EEQs were measured in all five *in vitro* effect-based methods due to the mixture effects in the samples. However, in the aquatic environment EE2 degrades less rapidly than the natural hormones [50] and may still be present

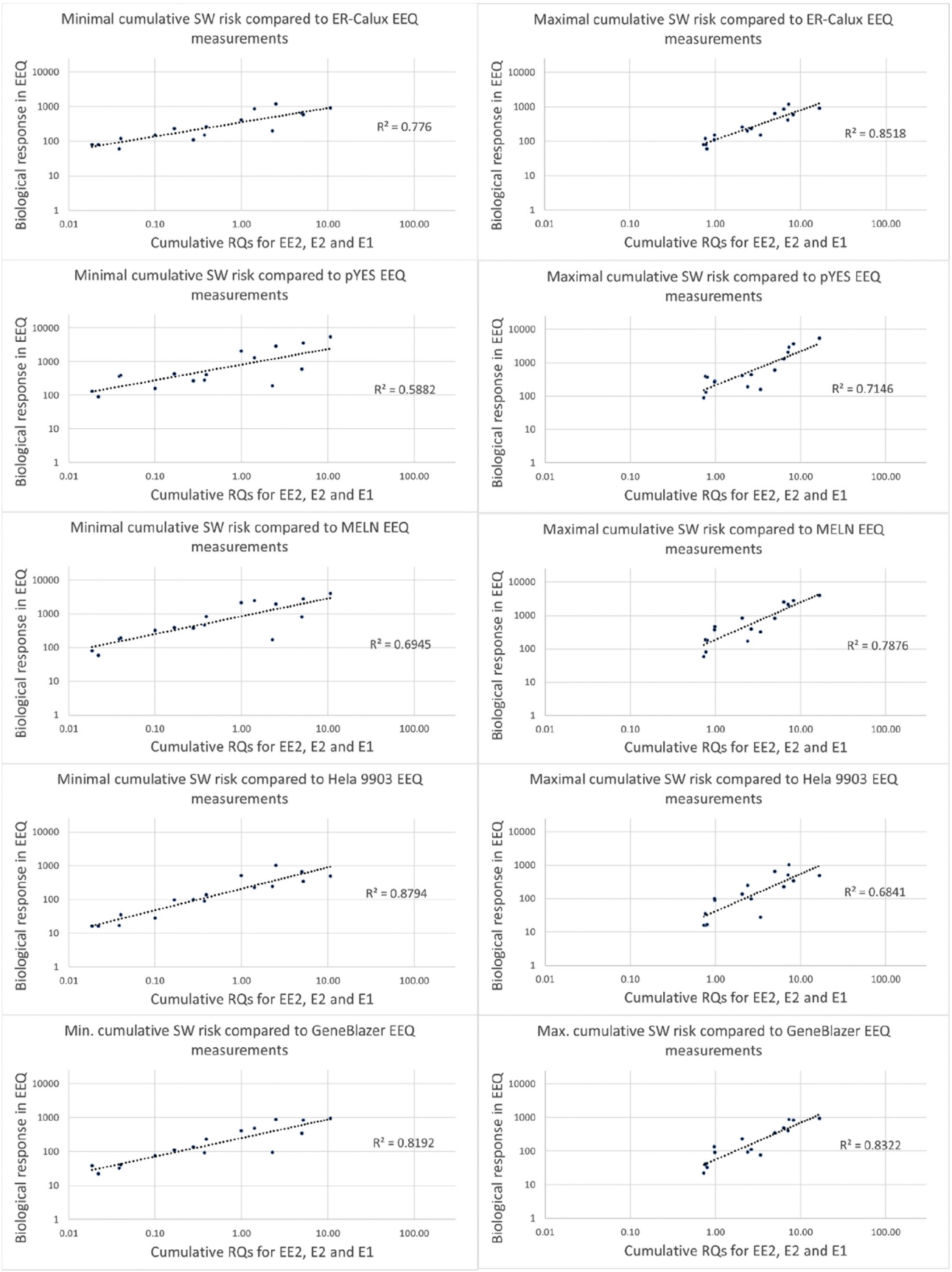


Fig. 1. Minimal and maximal cumulative risk quotients (RQ) compared with measured ER-Calux, pYES, MELN, Hela 9903, ER-GeneBlazer biological responses of ER activation [EEQ in pg/L] in 16 surface water (SW) samples. The *in vitro* effect-based methods are shown vertically from row 1 to 5. All correlations shown in the figures are highly significant with $p < 0.0001$ (SI Table 13).

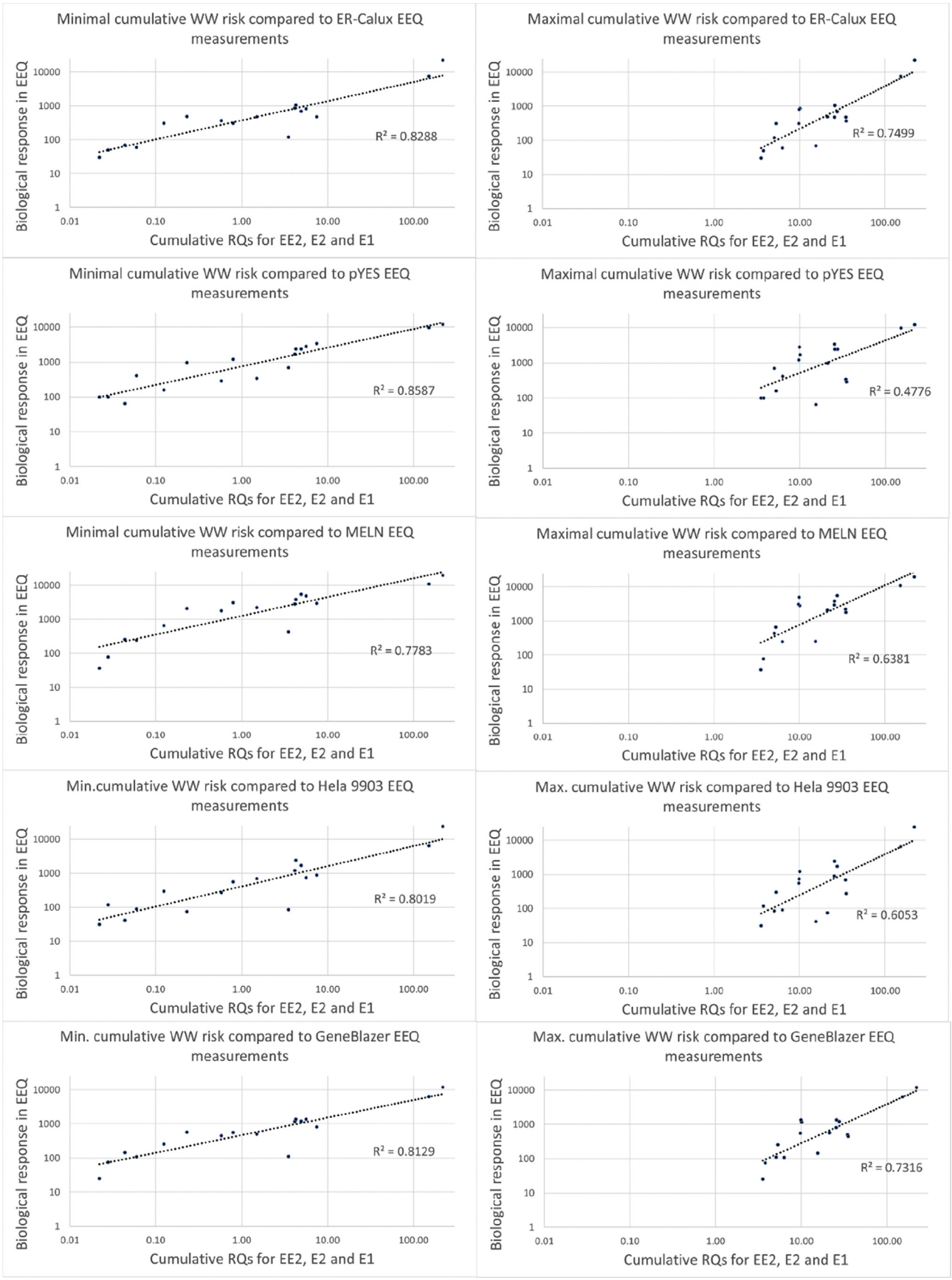


Fig. 2. Minimal and maximal cumulative risk quotients (RQ) compared with measured ER-Calux, pYES, MELN, Hela 9903, ER-GeneBlazer biological responses of ER activation [EEQ in pg/L] in 17 wastewater (WW) samples. The *in vitro* effect-based methods are shown vertically from row 1 to 5. All correlations shown in the figures are highly significant with $p < 0.0001$., only pYES graph at maximal risk had $p < 0.01$ (SI Table 14).

further downstream from the source. Where this is the case, EE2 could be missed both by chemical analysis and effect-based methods, but for effect-based methods the test specific relative potency for EE2 can be considered and further refinements in EBT are possible. EE2 alone can also be discharged by industrial sources, but such sources can be expected to discharge concentrations far above EBT which could be detected by *in vitro* effect-based methods.

In theory, anti-estrogenic compounds could modulate ER activity and, therefore, interfere with EEQ measurements. While most literature focuses on agonistic (estrogenic) modes of action, substantially less information is available on anti-estrogenic MoA in environmental waters [51]. Overall, anti-estrogens are not considered to be a major issue in common municipal WW [33]. So far, anti-estrogenic activity has not been commonly detected in environmental waters by effect-based methods and many limitations exist in measuring anti-estrogenic effects (lack of standardization, and potential artefact problems due to DOC) [48]. The highly significant correlation between measured estrogens and bioassay results found in this study supports the hypothesis that anti-estrogenic substances or other ER-receptor activating substances play a minor role in WW and SW samples containing WW (Figs. 1 and 2).

Given that the EQS values of the steroidal estrogens are based on population-relevant long-term effect data (the EQS were derived from Species Sensitivity Distribution based on data from 9 to 11 fish species), the mixture risk can be considered as directly indicative for population-relevant effects in fish species. As a consequence, the (receptor activation-based) biological response measured with the bioassays, which was highly correlated with cumulative RQs (Figs. 1 and 2), can also be considered to estimate the risk for aquatic species. Therefore, we investigated whether the biological EEQ response exceeded the EBT of 400 pg/L in those cases where a cumulative population relevant mixture-risk was identified.

3.4.3. Risk indication of in vitro effect-based methods for cumulative population relevant effects

The risk indication for all three EBT scenarios was calculated in the SI Tables 10–12 and showed a good agreement of chemical and biological risk indicators. For the moderate EBT of 400 pg/L the relative RIS (score of biological responses which indicates quantifiable chemical mixture risk) of all five *in vitro* assays was 77% ± 13% CV for SW samples and 91% ± 5% CV for WW samples. There are two reasons for the higher percentage of RIS in WW. First, there is naturally higher variability in the evaluation of samples with lower activities, such as SW compared to WW (EEQs close to the EBT can result in different category). Second, the composition of municipal wastewaters in Europe have been shown to typically contain steroidal estrogens with EE2 contributing ≤40% of total estrogenic activity (EEQ) [33]. As discussed above, the bioassays indicate the risks most precisely when EE2 is not the predominant estrogenicity driver but occurs in combination with other steroids (chapter 3.4.2).

Our results also demonstrate the potential of the effect-based methods to screen samples for other estrogens than the three target compounds. For example, two WW samples were evaluated as “compliant” or “not assessable” when their RQ were calculated based on chemical data (SI Table 11, samples G (14) and I (19)), however, most bioassays indicated elevated risk. The most probable reason is that the screening function of bioassays is not limited to steroidal estrogens and confirms findings of recently published approaches for screening endocrine active pharmaceuticals and other receptor activating substances [20,21,25]. In this study, the screening for other receptor activating compounds was measured by an ncr* score (the relative positive risk indication without chemical analytical verification ratio normalized to the number of chemical positive findings) and resulted in 11.4% “biological positives” for SW samples and 26.7% for WW samples. In other words, with the selected EBT, the effect-based methods were able to screen 11–27% more positive samples for SW and WW.

Finally, one of the *in vitro* effect-based methods (HeLa 9903) occasionally showed EEQs below the EBT where all other bioassays and chemical methods showed a risk (SI Table 11, SW samples P (7), M(28)). Variability can generally account for some negative risk indications, which occur when the detected EEQs are close to the EBT. These samples contained very high concentrations of E1. E1 is typically a less potent ER ligand than E2, but it is particularly less potent in HeLa 9903 with an estrogenic potency relative to E2 of 0.018 (the relative potencies of other used *in vitro* effect-based methods are listed in the SI Table 5 of the companion publication [36]), and thus the contribution of E1 to EEQ was lower than for most other bioassays with exception ER-Calux. Low potencies and higher variabilities, which are indicated by LOQs can lead to reduced detectability by some bioassays, and test specific refinements should be considered. This is also a matter of identifying criteria for benchmarking of bioassays suitable for these application purpose and to add test specific “sensitivity factors” which can be multiplied with EEQs to meet a screening EBT. This was identified as a future need and included as one of the aims of a subsequent project (see Conclusions and Outlook).

3.4.4. Comparison of trigger value (EBT) scenarios to assess the risk indication of in vitro effect-based methods

Different trigger values were applied to assess risk indication and screening function of used methods. Results provided in chapter 3.4.3 show that an EBT of 400 pg/L can distinguish with high precision (77% ± 13%–91% ± 5%) between more and less polluted SW and WW sites, indicated by a quantifiable population relevant mixture risk. To investigate the impact of the choice of EBT on results all three proposed EBT, 300 pg/L EEQ [33], 400 pg/L EEQ [25,26] and 500 pg/L EEQ [38] were compared.

Application of the lowest EBT of 300 pg/L resulted in the highest RIS of 83 ± 6% for SW samples, and 93 ± 6% for WW samples, as well as in the highest ncr* score with 26% for SW and 33% for WW. Use of the moderate EBT of 400 pg/L led to slightly (2–6%) lower RIS for WW and SW samples, compared to the strictest EBT scenario of 300 pg/L. On the other hand, the moderate EBT scenario reduced the ncr*to half for SW (11%) and to two thirds for WW (27%), compared to the strictest EBT scenario. The least stringent EBT of 500 pg/L lowered the RIS for WW and SW samples to 8% and 16%, respectively, compared to the strictest EBT scenario. The ncr* score decreased to 2% in SW and 20% in WW. The ncr* score can vary

Table 5
Mean positive risk indication scores (RIS) and coefficients of variations (CV) of the 5 *in vitro* effect-based methods for the identification of population relevant risks (RQs>1) applying different trigger values (EBT) 300, 400 and 500 pg/L. Additionally the percentage of positive biological responses without chemical verification (ncr*) was calculated. Data were used from SI Tables 10–12.

EBT approach	Risk indication of steroidalestrogens RIS [%]				Screening of other xenestrogens and unquantifiable steroidal oestrogens Mean percentage of ncr* related to chemical positives [%]			
	SW	CV	WW	CV	SW	CV	WW	CV
EBT = 300 pg/L	82.9	6.4	93.3	6.1	25.7	35.6	33.3	13.6
EBT = 400 pg/L	77.1	12.8	91.1	5.0	11.4	15.7	26.7	14.9
EBT = 500 pg/L	65.7	21.7	84.4	9.9	2.2	4.9	20.0	16.4

depending on the chemical composition and activity of estrogen mixtures.

Generally the ncr* should allow screening for non-target ER receptor-activating substances. A high ncr* as shown in our strictest EBT scenario would mean that there was a need for additional analyses by costly high end chemical analytical methods (26–33%). Application of a higher EBT of 500 pg/L results in a low ncr*. (2% in SW) indicating that samples containing unknown estrogens will not be selected. The moderate EBT scenario means that fewer (half to one third) samples have to be analysed further by chemical analysis. This moderate EBT scenario is still quite protective and specific with 77–91% of positive RIS based on quantified chemical analytical mixture risks. The moderate EBT scenario of 400 pg/L EEQ has an additional screening function for other non-target ER activating substances, combined with a high specificity for the risks of steroidal estrogens.

3.4.5. Specificity and sensitivity of chemical analytical and in vitro effect-based methods in compliance assessments

A suitable method should be specific and sensitive. A specificity and sensitivity analysis was therefore performed with each methods applied in this study in order to characterize and compare their suitability for monitoring (Fig. 3AB).

For SW samples the three chemical analytical methods performed with moderate specificity >73% and moderate sensitivity >69%. The Lab3 method achieved higher sensitivity of 75% and the Lab2 method achieved the highest specificity level of close to 85%. The five *in vitro* effect-based methods had, in most cases, high specificity and high sensitivity >90% in SW. Only the sensitivity of one method HeLa 9903 was lower (66.7%). ER-Calux, p-YES and MELN were the most sensitive assays (100% sensitivity), and ER-Calux, HeLa 9903 and ER-ER-GeneBlazer had the highest specificity (100%).

For WW samples, the chemical analytical methods performed with low specificity (in a range of 52–56%) and with low to moderate sensitivity (59–67%), likely due to matrix effects which were not removed efficiently by the silica gel cleaning step (see methods section 2.1). Most *in vitro* effect-based methods performed well in WW and showed both high specificity and high sensitivity >85%. Only the p-YES and MELN were less specific with 71% and 57% specificity, respectively. This can be explained by the higher sensitivity for E1 of both methods. For WW, MELN and ER-ER-GeneBlazer were the most sensitive assays (100% sensitivity), and ER-Calux and HeLa 9903 were the most specific. Finally, most (exception of MELN and pYES in SW which performed similarly) (Fig. 3 AB). This can be explained by the quantification problems of HPLC MS/MS that often occurred in WW and in some SW samples (SI Tables 1–3 and 6,7).

3.4.6. Comparability of chemical analytical and in vitro effect-based methods

A recent literature review [52] highlighted the need for sufficiently sensitive analytical methods for E2 and EE2 in order to be able to comply with the WFD reporting requirements [8,9]. Our study applied advanced analytical methods and confirmed this finding. The main advantages of chemical analytical methods are the quantification of single analytes, however, for both E2 and EE2 LOQs were often >EQS. Chemical analytical methods were able to detect steroidal estrogens above their EQS in 56% of SW samples and only in 16% of WW samples [36], demonstrating that the chemical analytical detection of E2 and EE2 is currently at the limit of feasibility with advanced methods. First monitoring results of the EU watch-list substances in 2017 confirm these results. In 21 EU member states the LOQs for EE2 E2 were above their EQS in >95% and approx. 50% of unquantified water samples analysed. In our study, if steroidal estrogens were detected, the average coefficients of variation CV % of quantifiable concentration measurements in SW (for E1 = 22.2%, E2 = 28.3%, EE2 = 15.8%) and WW (for E1 = 18.9%, E2 = 36.2%, EE2 = 14.6%) showed good agreement between the three chemical analytical analyses (SI Tables 1–3), but the methods showed also significant absolute variability in LOQs (Table 3) making a comparable risk-assessment difficult.

Overall, i*n vitro* effect-based methods were highly sensitive (90–94%) and specific (83–92%) in both SW and WW assessments (Fig. 3AB). The main advantage of the *in vitro* effect-based methods is their ability to account for the mixture toxicity and integrate the effects of unknown chemicals with the same MoA (e.g. metabolites) as well as synergistic or antagonistic mixture effects. In our study the CV of SW and WW sample assessments was in the range of 10–11%, showing good comparability of all five *in vitro* methods regarding the status assessment (Table 2) without any not-assessable samples.

Variability of *in vitro* effect based methods is similar to that of chemical analytical methods. To quantify intra- and inter-test variability, five *in vitro* effect-based methods were recently compared [47]. In this comparison the CV of EEQ concentrations measured in the five in *vitro* assays and for all samples was around 32% for comparing artificial mixtures. CV was lower for intra-day experiments (30%) compared to inter-day experiments (37%). ER-Calux had the best precision and repeatability with an overall CV of 13%. Further validation, inter-laboratory comparison studies and standardization of these effect-based methods may still improve their suitability for monitoring (Mehinto et al. [53]). In our study the five used *in vitro* effect-based methods correlated well among each other as it was shown in the companion publication by Könemann et al. [36]. In line with the results of our study, other studies have also confirmed that *in vitro* effect-based methods are

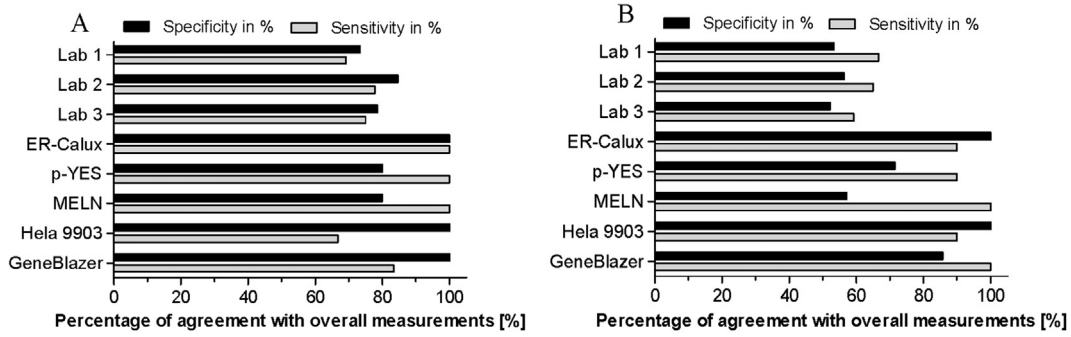


Fig. 3. AB: Specificity and sensitivity assessment of three analytical LC MS/MS methods (Lab 1, Lab 2, Lab 3) and five *in vitro* effect-based methods for specificity and sensitivity applied in this study. Results are given as relative values in % for 16 surface water samples (A) and 17 wastewater samples (B) and were compared to the overall compliance assessment. A high specificity indicates if a method is less prone to false positive assessments. A high sensitivity indicates if a method is less prone to false negative assessments.

able to benchmark contamination by estrogenic compounds correctly [23,54].

The results of our study support the application of *in vitro* effect-based methods for surface and especially for WW monitoring. Results of individual bioassays and chemical analyses correlated highly (Figs. 1 and 2) demonstrating the suitability of all effect-based methods to indicate steroidal estrogen pollution with population relevant mixture risks. Our proposed approach for an ecotoxicological status assessment is in line with the latest results of the EU project SOLUTIONS from König et al. [21], which measured pollution patterns in the river Danube using a large set of effect-based as well as chemical analytical methods. In this study an integrated analytical-bioanalytical approach was well suited to detect the impact of untreated wastewater on Danube River water quality. Both chemical and biological effect patterns were affected in a consistent way.

Limitations and recommendations: To prevent overestimation of these *in vitro* effect-based methods to protect the environment, their limitations are summarized as follows:

a) Cell culture based assays cannot account for organism-level toxicokinetic changes e.g. metabolism. Differences in toxicokinetics are likely in less frequently investigated species and modes of actions might in specific cases result in over- or underestimations of *in vivo* effects when evaluated purely by *in vitro* effect-based methods with one trigger value. Despite this our results showed that also the highly potent EE2 was correctly identified in mixtures by *in vitro* methods, see chapter 3.4.3. But generally the chemical analytical approach also does not consider the toxicokinetics and bioavailability, and EQS are always limited to available effect-data sets.
b) Although not confirmed in any study including this one, the interaction of the generally much less potent (compared to the steroidal estrogens) anti-estrogenic compounds could lead to underestimation of risks of estrogenicity cannot be excluded for very specific sample compositions. In case if high concentrations of anti-estrogenic substances would bind to ER receptor the detection of ER-agonists can be lowered.

Considering these limitations and options, we recommend direct use of *in vitro* effect-based methods with a preliminary screening EBT of 400 pg/L EEQ under the WFD for the following pragmatic reasons: a) there are currently no better available tools to monitor this type of endocrine disruptor pollution providing important link between MoA and adverse effects, b) they will circumvent current monitoring problems of steroidal estrogens, c) the methods are cost-efficient and can decrease the financial burden of monitoring, d) they are readily available, e) they address mixture effects.

3.4.7. Risk management

3.4.7.1. Surface water risk management options. The development of analytical techniques to detect EDCs in environmental matrices still remains one of the main challenges for environmental chemists [54,55]. Due to analytical difficulties in the last decades no representative EU-wide monitoring dataset is available and risk characterizations are mainly known from modelling. For example Johnson and colleagues [1] estimated at median flow conditions an average an EQS exceedance of EE2 in 12% by length of Europe's rivers, which can in some countries also be higher than 30%. This single substance related population relevant risk is certainly increased by other ER activating substances.

One of the main recommendations (in view of the future review of the WFD) is to integrate effect-based methods into monitoring of water quality and to adopt them as a key approach for addressing chemical mixtures interactions with aquatic organisms [16]. Our study supports this recommendation: specific effect-based methods proved to be suitable tools to indicate the risks associated to the quantified and unquantified fractions of EE2, E2, and E1 for various water samples and should therefore be applied as screening tools to identify polluted waterbodies. Especially because of the low LOQs and low absolute variability in LOQs they are reliable and suitable regarding risk assessments and prioritizations.

Besides, effect-based methods are the only currently available tools to address unknown mixture risks and circumvent the monitoring limitations of current chemical analytical methods, as mentioned above. But another question needs to be solved for risk management: How to proceed when an EBT is exceeded?

First of all, an EBT exceedance can identify waterbodies at risk for receptor-mediated estrogenicity. This can allow focusing of monitoring resources on priority sites. For example, if 100 waterbodies are screened and only 10% are at risk, for 90% of remaining waterbodies no costly high-end chemical analysis needs to be performed. Taking into account our chemical analytical findings for SW assessment: 54% ± 10% of the samples were rated as "compliant", 31% ± 6% "non-compliant" and 15% ± 10% "not-assessable" due to too high LOD/LOQ (Table 1). With effect-based methods, the results overlapped with those obtained with analytical methods, (61% ± 11% of the samples assessed as "compliant" and 39% ± 11 "non-compliant" at an EBT of 400 pg/L EEQ) (Table 2), but thanks to the shift from chemical analytical assessments to an ecotoxicological effect-based assessment, the percentage of not assessable samples can be reduced to zero with an obvious benefit in terms of assessment feasibility and costs.

Secondly, the choice is given to use directly the EBT RQ (Equation (4) in chapter 2.6) for an ecotoxicological status assessment or if additionally an identification of substances for an investigative purpose is needed, e.g. via application of mini Effect-Directed Analysis (EDA), but this would definitely increase the costs. A cost estimation for a mini EDA for ER activation is in the range of 5 k Euro per sample (pers. communication Timo Hamers VU University of Amsterdam). Based on our study, if a ecotoxicological risk is identified, we suggest to mitigate if possible the risk or to identify the cause substance. Considering the current costs of mini-EDAs for substance identification and the high probability that most of the effects of concern are caused by mixtures this favours a direct risk reduction. An ecotoxicological effect-based assessment can be established if in the EU context the EBT is harmonized and highly validated and comparable effect-based methods are used for screening. In this study, we characterized and discussed the screening value of *in vitro* effect-based methods. Most of the *in vitro* effect-based methods are less expensive compared to high-end chemical analytical methods considering installation costs and analysis costs per sample. A short cost discussion subchapter is provided in the SI.

3.4.7.2. Wastewater risk management options. Although no legal discharge limits for micropollutants exist at the EU level [31], WW are often monitored as the main known sources of these compounds to waterbodies. Jarosova and colleagues [33] compiled data of 353 wastewater measurements from three studies with a median concentration of 7–12 ng/L E1, 1.3–1.7 ng/L E2 and 0.47–0.6 ng/L EE2, so it will depend on the dilution factor and the background concentration of the receiving water if the EQS can be met and population relevant risks can be excluded. Also, several activities aim at limiting unnecessary risks of pharmaceuticals, such as the EU Strategy on pharmaceuticals [8] which aims at reducing discharges, emissions, and losses, or the Eco-Pharmaco-Stewardship Initiative for industrial wastewaters [48]. Our study identified a high RIS (mean 91%) of 5 *in vitro* effect-based methods for WW and

notable chemical analytical limitations for the detection of steroidal estrogens in these samples because of their complex matrix composition. This offers a direct use of effect-based methods to WW risk regulation at local, national, and EU-wide level. This is of special importance, as the main entrance pathway of the synthetic EE2 as well as the overall ER-mediated estrogenicity into our waterbodies is municipal WW. ER-mediated estrogenicity in WW can be reduced by around a factor of 10 with additional wastewater treatment techniques and the *in vitro* effect-based methods can be also used to monitor these technical options to reduce the pharmaceutical and anthropogenic mixture risks before entering the aquatic environment [56–58].

Without knowledge of the ER mediated estrogenicity risk in the receiving waterbody, we suggest using the same ecotoxicological status assessment for WW as for SW to ensure an assessment compliance, because a compliant WW cannot lead to a change in SW compliance assessment to a non-compliant assessment. With further knowledge about the ER mediated estrogenicity in the receiving waterbody the EEQ of WW can be combined with a dilution factor to estimate a more appropriate overall ER mediated estrogenicity RQ in SW (Equation (8)). Similarly, for chemical analytical cumulative risk assessment, the combined risk can be calculated with a dilution factor and Measured Environmental Concentrations MEC (Equation (9)).

$$\text{RQ effect} - \text{based} = \frac{\text{WW EEQ}}{\text{EBT}} * \frac{1}{\text{dilution factor}} + \frac{\text{SW EEQ}}{\text{EBT}} * (1 - 1/\text{dilution factor}) \quad (8)$$

$$\text{RQ chem} = \left(\frac{\text{WW MEC EE2}}{\text{EQS EE2}} + \frac{\text{WW MEC E2}}{\text{EQS E2}} + \frac{\text{WW MEC E1}}{\text{EQS E1}}\right) * \frac{1}{\text{dilution factor}} + \left(\frac{\text{SW MEC EE2}}{\text{EQS EE2}} + \frac{\text{SW MEC E2}}{\text{EQS E2}} + \frac{\text{SW MEC E1}}{\text{EQS E1}}\right) * (1 - 1/\text{dilution factor}) \quad (9)$$

Because we intended a 1:1 comparability in SW vs. WW assessments, our sampling locations had varying dilution factors, and we had, in most cases, too limited knowledge about the EEQ in the receiving water, we used the simplified Equation (4) (Methods 2.6) for our calculations. For further studies, another practical solution would be to measure EEQ directly in the mixing zone of the receiving waterbody, meaning one EEQ for one ecotoxicological assessment.

4. Conclusions and Outlook

Considering their relevance, applicability for screening as well as limitations, we propose the application of effect-based methods under the WFD, in particular the use of *in vitro* effect based methods for identifying ER-mediated risk in WFD monitoring programs. The methods are: I) capable of addressing relevant combined mixture effects, II) able to overcome detection problems encountered with analytical techniques for the EU watch list substances (EE2 and E2), III) suitable as screening tools for the identification and prioritization of waterbodies requiring further examination, and IV) suitable for measuring ecotoxicological status in relation to receptor-mediated estrogenicity, one of the most relevant MoA of EDCs.

A recent study [16] emphasized the need for the harmonization and standardization of EBT. The derivation of EBT is one of the tasks identified for the activity on effect-based methods started by the EU Working Group Chemicals under the Common Implementation Strategy for the WFD [59]. The results of our study confirm that a preliminary screening EBT of 400 pg/L EEQ is suitable for the identification of population relevant analytical (and mixture) risks from steroidal estrogens, at the same time as achieving the screening of other ER-activating substances. This EBT is recommended as a suitable threshold or cut-off value to discriminate samples of greater level of estrogenic pollution with the aim of classifying waterbodies. The application of this value can be further refined, taking into account differences in sensitivity of the used methods via sensitivity factors and risk classifications.

Our study demonstrated SW and WW risk management options by using risk indication scores (RIS) based on tested EBT covering population relevant effects for aquatic organisms. The tested concepts proved to be applicable for WW and for most SW (91% of RIS vs. 77% of RIS). Effect-based methods were highly sensitive (90–94%) and specific (83–92%) in both SW and WW assessments. In special situations where EE2 occurs at low, yet EQS-exceeding, concentrations and alone mainly contributes to ER-mediated estrogenicity, a false negative assessment might occur. This issue was only discovered in one of 16 SW or in total 33 samples. Based on the highly significant correlations between all measured estrogens risks and bioassay results found in this study (Fig. 1 + 2) it was possible to identify that anti-estrogenicity and matrix effects played a minor role in most of our samples. This presented approach allows us to screen, prioritize, and manage environmental samples using the ER-EBT concept very similar and compatible to the current chemical status assessment of the WFD. Furthermore, ER-Calux, A-YES, and L-YES will be standardized at DIN/EN/ISO level by early 2018 supporting their availability for regulatory use. Our study showed the use of very specific *in vitro* effect-based methods with tested EBT is able to bridge the gap between conventional analytical and effect-based monitoring and risk assessment for steroidal estrogens.

The combination of the results of this study demonstrates that water quality assessment can progress from a purely analytical approach to effect-based monitoring, from single substance to known and unknown mixture assessment, and from *in vitro* screening to population-relevant risk assessment. This approach can support the introduction of the proposed new holistic approach to the regulation of chemicals in the aquatic environment under the EU Water Framework Directive, an objective which EU water directors agreed in November 2016 to investigate [60] and which has also been recommended by international platforms such as, the NORMAN network and the EU-funded SOLUTIONS project [16]. A follow-up study regarding the use of different effect-based methods under the EU watch list mechanism in 2017 and 2018 is intended. This follow-up study aims to characterize the screening function for ER-mediated effects with regulatory relevant EU watch list samples. Moreover, it intends to apply an integrative effect-based approach for other relevant pharmaceutical MoA such as COX inhibition [61,62].

Acknowledgements

The major part of the project was funded by in-kind contribution from numerous project partners showing high interest in this project. Essential parts of the project, e.g. extraction and reporting were externally co-funded by the Swiss Centre of Applied Ecotoxicology EAWAG-EPFL, NORMAN Network, INERIS and Pharmaceutical Associations (Pfizer, Johnson & Johnson-Johnson & Johnson, TEVA which joined the project later and only supported the extraction at central Lab BDS). In total, 70

colleagues associated to 24 institutes and agencies in 12 nations contributed to this project. The authors would like to thank: Joint Research Centre (EC), The French Agency for Biodiversity (previously ONEMA) (FR), INERIS (FR), Bio Detection Systems (NL), Swiss Centre for Applied Ecotoxicology (CH), Federal Institute of Hydrology (DE), Federal Environment Agency (DE), RWTH Aachen (DE), RECETOX (CZ), NORMAN-Network, Helmholtz Centre for Environmental Research-UFZ (DE), IRSA-CNR (IT), Italian Institute of Health (IT), University of Leon (ES), Water Research Institute T.G.Masaryk (CZ), Bavarian State Office for Environment (DE), LANUV (DE), Environment Agency Austria (AT), ISSeP (Scientific Institute of Public Service) Wallonia (BE), SMAT (IT), BrianzAcque (IT), Vettabbia scarl, MilanoDepur (IT), Agence de l'Eau Adour-Garonne (FR), Ontario Ministry of the Environment and Climate Change (CAN), McGill University (CAN), Environmental Institute (SK) and DG Environment of the European Commission (EC). The project partners from UFZ, RWTH Aachen, INERIS and RECETOX were supported by the SOLUTIONS project, that is funded by the European Union Seventh Framework Programme (FP7-ENV-2013-two-stage Collaborative project) under grant agreement number 603437. Additionally, we would like to thank Christin Kuehnert (UFZ) and Maria König (UFZ) for experimental assistance. Special thanks for stewardship of this study we would like to address to Helen Clayton and Stéphanie Schaan from European Commission, Environment DG and Joint Research Centre (EC) for chemical analysis of samples.

The Federal Institute of Hydrology got financial support by the Federal Ministry for the Environment, Nature Conservation, and Building (BMUB).

List of abbreviations

AA-EQS	Annual-Average Environmental Quality Standard
EDA	Effect directed Analysis
EDC	Endocrine Disrupting Compounds
EQS	Environmental Quality Standard
BEQ	Bioanalytical equivalents
COX	cyclooxygenase
CV	Coefficient of variation
EE2	17-alpha-ethinylestradiol
E2	17-beta-estradiol
E1	Estrone
EBT	Effect-based trigger values
EEQ	17-beta-estradiol-equivalents
ER	Estrogen Receptor
EU	European Union
HPLC MS	High Pressure liquid chromatography–mass spectrometry
LOD	Limit of Detection
LOQ	Limit of Quantification
MoA	Mode of Action
MEC	Measured Environmental Concentration
ncr	No Chemcical Risk indication but positive biological response
RIS	Risk Indication Score
RQ	Risk Quotient
SW	Surface waters
WFD	Water Framework Directive
WW	Wastewaters

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.trac.2018.02.013.

References

[1] A.C. Johnson, E. Dumont, R.J. Williams, R. Oldenkamp, I. Cisowska, J.P. Sumpter, Do concentrations of ethinylestradiol, estradiol, and diclofenac in European rivers exceed proposed EU environmental quality Standards? Environ. Sci. Technol. 47 (2013) 12297–12304. https://doi.org/10.1021/es4030035.

[2] J. Zha, L. Sun, Y. Zhou, P.A. Spear, M. Ma, Z. Wang, Assessment of 17α-ethinylestradiol effects and underlying mechanisms in a continuous, multi-generation exposure of the Chinese rare minnow (Gobiocypris rarus), Toxicol. Appl. Pharmacol. 226 (2008) 298–308. https://doi.org/10.1016/j.taap.2007.10.006.

[3] A.R. Schwindt, D.L. Winkelman, K. Keteles, M. Murphy, A.M. Vajda, An environmental oestrogen disrupts fish population dynamics through direct and transgenerational effects on survival and fecundity, J. Appl. Ecol. 51 (2014) 582–591. https://doi.org/10.1111/1365-2664.12237.

[4] H. Segner, A. Casanova-Nakayama, R. Kase, C.R. Tyler, Impact of environmental estrogens on Yfish considering the diversity of estrogen signaling, Gen. Comp. Endocrinol. 191 (2013) 190–201. https://doi.org/10.1016/j.ygcen.2013.05.015.

[5] L. Viganò, J.L. Loizeau, A. Mandich, G. Mascolo, Medium- and Long-Term Effects Of estrogenic contaminants on the middle river Po fish community as reconstructed from a sediment core, arch, Environ. Contam. Toxicol. 71 (2016) 454–472. https://doi.org/10.1007/s00244-016-0315-3.

[6] K.A. Kidd, P.J. Blanchfield, K.H. Mills, V.P. Palace, R.E. Evans, J.M. Lazorchak, R.W. Flick, Collapse of a fish population after exposure to a synthetic estrogen, Proc. Natl. Acad. Sci. 104 (2007) 8897–8901. https://doi.org/10.1073/pnas.0609568104.

[7] M. Carere, S. Polesello, R. Kase, B.M. Gawlik, The emerging contaminants in the context of the EU water framework directive, in: M. Petrovic, S. Sabater, A. Elosegi, D. Barceló (Editors), Emerg. Contam. River Ecosyst. Occur. Eff. Under Mult. Stress Cond, Springer International Publishing, Cham, 2016, pp. 197–215. https://doi.org/10.1007/698_2015_5011.

[8] Off.J. Eur. Union L 226 56, Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as Regards Priority Sin the Field of Water Policy, 2013, pp. 1–17. n.d.

[9] Off. J. Eur. Union L 78, Commission Implementing Decision (EU) 2015/495 of 20 March 2015 Establishing a Watch list of Substances for Union-wide Monitoring in the Field of Water Policy Pursuant to Directive 2008/105/EC of the European Parliament and of the Council, 2015, p. 40. n.d.

[10] European Commission, Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL amending Directives 2000/60/EC and 2008/105/EC as Regards Priority Substances in the Field of Water Policy, COM, 2011, p. 876 final 2011/0429 (COD), Brussels, 2012.

[11] R. Kase, R.I.L. Eggen, M. Junghans, C. Götz, J. Hollender, Assessment of micropollutants from municipal wastewater- combination of exposure and ecotoxicological effect data for Switzerland, ISBN: 978-953-307-233-3, in: Wastewater - Evaluation and Management,Fernando Sebastián García Einschlag, InTech, 2011, pp. 32–55 (chapter 2), Available at: http://cdn.intechopen.com/pdfs/14570.pdf. (Accessed 26 May 2017).

[12] WHO/UNEP, State of the science of endocrine disrupting chemicals-2012, An Assessment of the State of the Science of Endocrine Disruptors Prepared by a Group of Experts for the United Nations Environment Programme (UNEP) and World Health Organization (WHO), 2013. http://www.who.int/ceh/publications/endocrine/en/. (Accessed 13 April 2017).

[13] R.N. Carvalho, A. Arukwe, S. Ait-Aissa, A. Bado-Nilles, S. Balzamo, A. Baun, S. Belkin, L. Blaha, F. Brion, D. Conti, N. Creusot, Y. Essig, V.E. V Ferrero, V.V. Flander-Putrle, M. Fürhacker, R. Grillari-Voglauer, C. Hogstrand, A. Jonáš, J.B. Kharlyngdoh, R. Loos, A.K. Lundebye, C. Modig, P.E. Olsson, S. Pillai, N. Polak, M. Potalivo, W. Sanchez, A. Schifferli, K. Schirmer, S. Sforzini, S.R. Stürzenbaum, L. Søfteland, V. Turk, A. Viarengo, I. Werner, S.S. Yagur-Kroll, R. Zounkova, T. Lettieri, Mixtures of chemical pollutants at European legislation safety concentrations: how safe are they? Toxicol. Sci. 141 (2014) 218–233. https://doi.org/10.1093/toxsci/kfu118.

[14] European Commission, 4th Report on the Implementation of the "Community Strategy for Endocrine Disrupters" a Range of Substances Suspected of Interfering with the Hormone Systems of Humans and Wildlife (COM, 1999, p. 706. SEC(2011) 1001 final, Brussels, 2011, http://ec.europa.eu/environment/chemicals/endocrine/pdf/sec_2011_1001.pdf. (Accessed 13 April 2017).

[15] The combination effects of chemicals. Chemical mixtures. COM, European CommissAssessment of Micropollutants from Municipal Wastewater- Combination of Exposure and Ecotoxicological Effect Data for Switzerlandion, Communication from the Commission to the Council, 2012, p. 252. Brussels, 2012, http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52012DC0252. (Accessed 13 April 2017).

[16] W. Brack, V. Dulio, M. Ågerstrand, I. Allan, R. Altenburger, M. Brinkmann, D. Bunke, R.M. Burgess, I. Cousins, B.I. Escher, F.J. Hernández, L.M. Hewitt, K. Hilscherová, J. Hollender, H. Hollert, R. Kase, B. Klauer, C. Lindim, D.L. Herráez, C. Miège, J. Munthe, S. O'Toole, L. Posthuma, H. Rüdel, R.B. Schäfer, M. Sengl, F. Smedes, D. van de Meent, P.J. van den Brink, J. van Gils, A.P. van Wezel, A.D. Vethaak, E. Vermeirssen, P.C. von der Ohe, B. Vrana, Towards the review of the European Union Water Framework management of chemical contamination in European surface water resources, Sci. Total Environ. 576 (2017) 720–737. https://doi.org/10.1016/j.scitotenv.2016.10.104.

[17] R. Kase, H. Clayton, F. Martini, Science-Policy Interface (SPI) Activity on Prioritization of Research Needs, Knowledge Availability and Dissemination for the Working Group E (Chemical Aspects) 2010-2012, 2012. Available at: https://circabc.europa.eu/w/browse/5bf63ff3-b24b-4365-8a57-38e4d56b941c. (Accessed 13 April 2017).
[18] C. Di Paolo, R. Ottermanns, S. Keiter, S. Ait-Aissa, K. Bluhm, W. Brack, M. Breitholtz, S. Buchinger, M. Carere, C. Chalon, X. Cousin, V. Dulio, B.I. Escher, T. Hamers, K. Hilscherová, S. Jarque, A. Jonas, E. Maillot-Marechal, Y. Marneffe, M.T. Nguyen, P. Pandard, A. Schifferli, T. Schulze, S. Seidensticker, T.-B. Seiler, J. Tang, R. van der Oost, E. Vermeirssen, R. Zounková, N. Zwart, H. Hollert, 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. 104 (2016) 473–484. https://doi.org/10.1016/j.watres.2016.08.018.
[19] M. Hecker, H. Hollert, Endocrine disruptor screening: regulatory perspectives and needs, Environ. Sci. Eur. 23 (2011) 15. https://doi.org/10.1186/2190-4715-23-15.
[20] RIWA, T. Slootweg, C. J Houtman, Evaluatie van hormonale activiteit gemeten in de Rijn bij Lobith (2010-2011), 2012. http://www.riwa-rijn.org/evaluatie-van-hormonale-activiteit-gemeten-in-de-rijn-bij-lobith-2010-2011/.
[21] M. König, B.I. Escher, P.A. Neale, M. Krauss, K. Hilscherová, J. Novák, I. Teodorović, T. Schulze, S. Seidensticker, M.A. Kamal Hashmi, J. Ahlheim, W. Brack, Impact of untreated wastewater on a major European river evaluated with a combination of in vitro bioassays and chemical analysis, Environ. Pollut. 220 (2017) 1220–1230. https://doi.org/10.1016/j.envpol.2016.11.011.
[22] B. Jarošová, A. Erseková, K. Hilscherová, R. Loos, B.M. Gawlik, J.P. Giesy, L. Bláha, Europe-wide survey of estrogenicity in wastewater treatment plant effluents: the need for the effect-based monitoring, Environ. Sci. Pollut. Res. Int. 21 (2014) 10970–10982. https://doi.org/10.1007/s11356-014-3056-8.
[23] F.D.L. Leusch, C. De Jager, Y. Levi, R. Lim, L. Puijker, F. Sacher, L.A. Tremblay, V.S. Wilson, H.F. Chapman, Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters, Environ. Sci. Technol. 44 (2010) 3853–3860.
[24] P. Schröder, B. Helmreich, B. Škrbić, M. Carballa, M. Papa, C. Pastore, Z. Emre, A. Oehmen, A. Langenhoff, M. Molinos, J. Dvarioniene, C. Huber, K.P. Tsagarakis, E. Martinez-Lopez, S.M. Pagano, C. Vogelsang, G. Mascolo, Status of hormones and painkillers in wastewater effluents across several European states—considerations for the EU watch list concerning estradiols and diclofenac, Environ. Sci. Pollut. Res. 23 (2016) 12835–12866. https://doi.org/10.1007/s11356-016-6503-x.
[25] P.Y. Kunz, C. Kienle, M. Carere, N. Homazava, R. Kase, In vitro bioassays to screen for endocrine active pharmaceuticals in surface and wastewaters, J. Pharm. Biomed. Anal 106 (2015) 107–115. https://doi.org/10.1016/j.jpba.2014.11.018.
[26] R. Kase, P.Y. Kunz, H. Hollert, I. Werner, Contribution on Bioanalytical Assays for Steroidal Estrogens, in EU JRC report by R. Loos "Analytical Methods Relevant to the European Commission's 2012 Proposal on Priority Substances Under the Water Framework Directive, 2012. https://doi.org/10.2788/51497.
[27] A.-S. Wernersson, M. Carere, C. Maggi, P. Tusil, P. Soldan, A. James, W. Sanchez, V. Dulio, K. Broeg, G. Reifferscheid, S. Buchinger, H. Maas, E. Van Der Grinten, S. O'Toole, A. Ausili, L. Manfra, L. Marziali, S. Polesello, I. Lacchetti, L. Mancini, K. Lilja, M. Linderoth, T. Lundeberg, B. Fjällborg, T. Porsbring, D.J. Larsson, J. Bengtsson-Palme, L. Förlin, C. Kienle, P. Kunz, E. Vermeirssen, I. Werner, C.D. Robinson, B. Lyons, I. Katsiadaki, C. Whalley, K. den Haan, M. Messiaen, H. Clayton, T. Lettieri, R.N. Carvalho, B.M. Gawlik, H. Hollert, C. Di Paolo, W. Brack, U. Kammann, R. Kase, The European technical report on aquatic effect-based monitoring tools under the water framework directive,, Environ. Sci. Eur. 27 (2015) 7. https://doi.org/10.1186/s12302-015-0039-4.
[28] A.-S. Wernersson, et al., Technical Report on Aquatic Effect-Based Monitoring Tools, Technical Report 2014–077, 2014. https://doi.org/10.2779/7260.
[29a] C. Heiss, B. Gawlik, B. von Danwitz, D. Wilke, G. Reifferscheid, G. Maack, H. Hollert, L. Portugall, M. Fuerhacker, M. Carere, N. Homazava, O. Perceval, P. Kunz, R. Kase, S. Buchinger, Selim, V. Dulio, Recommendation for a Monitoring Strategy for Estrogens in Coastal and Continental Surface Waters, 2013. http://www.bafg.de/DE/05_Wissen/02_Veranst/2013/2013_02_27_votum_en.pdf?__blob=publicationFile. (Accessed 30 May 2017);
[b] Valeria Dulio, Robert Kase, Recommendations for the use of effect-based methods for monitoring of estrogens in surface waters – Outcomes of the Dübendorf Workshop (21 June 2017), 2017. Available at, http://www.ecotoxcentre.ch/media/178820/estrogen-monitoring-recommendations-2017_final.pdf. (Accessed 23 March 2018).
[30] S. Jobling, R. Williams, A. Johnson, A. Taylor, M. Gross-Sorokin, M. Nolan, C.R. Tyler, R. van Aerle, E. Santos, G. Brighty, Predicted exposures to steroid estrogens in U.K. Rivers correlate with widespread sexual disruption in wild fish populations, Environ. Health Perspect. 114 (2005) 32–39. https://doi.org/10.1289/ehp.8050.
[31] N. Bolong, A.F. Ismail, M.R. Salim, T. Matsuura, A review of the effects of emerging contaminants in wastewater and options for their removal,, Desalination 239 (2009) 229–246.
[32] M.O. Barbosa, N.F.F. Moreira, A.R. Ribeiro, M.F.R. Pereira, A.M.T. Silva, Occurrence and removal of organic micropollutants: an overview of the watch list of EU Decision 2015/495, Water Res. 94 (2016) 257–279. https://doi.org/10.1016/j.watres.2016.02.047.
[33] B. Jarosova, L. Blaha, J.P. Giesy, K. Hilscherova, What level of estrogenic activity determined by in vitro assays in municipal wastewaters can be considered as safe? Environ. Int. 64 (2014) 98–109. https://doi.org/10.1016/j.envint.2013.12.009.
[34] F.D.L. Leusch, H.F. Chapman, M.R. van den Heuvel, B.L.L. Tan, S.R. Gooneratne, L.A. Tremblay, Bioassay-derived androgenic and estrogenic activity in municipal sewage in Australia and New Zealand, Ecotoxicol. Environ. Saf. 65 (2006) 403–411. https://doi.org/10.1016/j.ecoenv.2005.07.020.
[35] A.C. Johnson, R.J. Williams, A model to estimate influent and effluent concentrations of estradiol, estrone, and ethinylestradiol at sewage treatment works, Environ. Sci. Technol. 38 (2004) 3649–3658. http://www.ncbi.nlm.nih.gov/pubmed/15296317.
[36] S. Könemann, R. Kase, E. Simon b, K. Swart, S. Buchinger, M. Schlüsener, H. Hollert, B.I. Escher, I. Werner, S. Aît-Aïssa, V. Dulio, S. Valsecchi, S. Polesello, P. Behnisch, B. Javurkova, O. Perceval, C. Di Paolo, D. Olbrich, S. Tavazzi, E. Sychrova, M. Gundlach, R. Schlichting, L. Leborgne, M. Clara, C. Scheffknecht, Y. Marneffe, C. Chalon, P. Tušil, P. Soldàn, B. von Danwitz, J. Schwaiger, A. Moran Palao, I. San Martín Becares, F. Bersani, E. Vermeirssen, K. Hilscherová, G. Reifferscheid, M. Carere, Effect-Based and Chemical Analytical Methods to Monitor Estrogens Under the European Water Framework Directive, TrAC Trends Anal. Chem 102 (May 2018) 225–235. https://doi.org/10.1016/j.trac.2018.02.008 (publication with shared first authorship).
[37] F. Martini, C. Fragakis, S. Michel, A. Natacha, G. Kinga, R. Kase, G. Monacelli, Y. Reyjol, R. Ward, Science–Policy Interface in Support of the Water Framework Directive , 2013. https://doi.org/10.2777/35629. CIS-SPI activity report 2010-12, Luxembourg.
[38] R. Van der Oost, G. Sileno, M. Suárez-Muñoz, M.T. Nguyen, H. Besselink, B. Bram, SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment; Part I: model design and effect-based trigger values, Environ. Toxicol. Chem. 36 (9) (2017 Sep) 2385–2399. https://doi.org/10.1002/etc.3836.
[39] OECD, Test Guideline 455, Stably Transfected Human Estrogen Receptor-α Transcriptional Activation Assay for Detection of Estrogenic Agonist-Activity of Chemicals, 2009. https://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oecd/oecd-tg455.pdf. (Accessed 13 April 2017).
[40] R. Huang, S. Sakamuru, M.T. Martin, D.M. Reif, R.S. Judson, K.A. Houck, W. Casey, J.H. Hsieh, K.R. Shockley, P. Ceger, J. Fostel, K.L. Witt, W. Tong, D.M. Rotroff, T. Zhao, P. Shinn, A. Simeonov, D.J. Dix, C.P. Austin, R.J. Kavlock, R.R. Tice, M. Xia, Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway, Sci. Rep. 4 (2014). https://doi.org/10.1038/srep05664.
[41] A. Kortenkamp, Ten years of mixing cocktails: a review of combination effects of endocrine-disrupting chemicals, Environ. Health Perspect. 115 (2007) 98–105. https://doi.org/10.1289/ehp.9357.
[42] A. Kortenkamp, Low dose mixture effects of endocrine disrupters and their implications for regulatory thresholds in chemical risk assessment, Curr. Opin. Pharmacol. 19 (2014) 105–111. https://doi.org/10.1016/j.coph.2014.08.006.
[43] A. Kortenkamp, T. Backhaus, M. Faust, State of the art report on mixture toxicity. http://ec.europa.eu/environment/chemicals/effects/pdf/report_mixture_toxicity.pdf, 2009.
[44] L. Sachs, Angewandte Statistik: Anwendung Statistischer Methoden, ninth ed., Springer-Verlag, 1999.
[45] Off.J. Eur. Union L 201, 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, 2009, p. 36. n.d.
[46] B.I. Escher, F.D.L. Leusch, Bioanalytical Tools in Water Quality Assessment, IWA publishing, London, 2012.
[47] P.Y. Kunz, E. Simon, N. Creusot, B.S. Jayasinghe, C. Kienle, S. Maletz, A. Schifferli, C. Schönlau, S. Aït-Aïssa, N.D. Denslow, H. Hollert, I. Werner, E.L.M. Vermeirssen, Effect-based tools for monitoring estrogenic mixtures: evaluation of five in vitro bioassays, Water Res. 110 (2017) 378–388. https://doi.org/10.1016/j.watres.2016.10.062.
[48] D.J. Caldwell, B. Mertens, K. Kappler, T. Senac, R. Journel, P. Wilson, R.D. Meyerhoff, N.J. Parke, F. Mastrocco, B. Mattson, R. Murray-Smith, D.G. Dolan, J.O. Straub, M. Wiedemann, A. Hartmann, D.S. Finan, A risk-based approach to managing active pharmaceutical ingredients in manufacturing effluent, Environ. Toxicol. Chem. 35 (2016) 813–822. https://doi.org/10.1002/etc.3163.
[49] B. Jarosova, L. Blaha, B. Vrana, T. Randak, R. Grabic, J.P. Giesy, K. Hilscherova, Changes in concentrations of hydrophilic organic contaminants and of endocrine-disrupting potential downstream of small communities located adjacent to headwaters, Environ. Int. 45 (2012) 22–31. https://doi.org/10.1016/j.envint.2012.04.001.
[50] M.D. Jurgens, K.I.E. Holthaus, A.C. Johnson, J.L. Smith, M. Hetheridge, R.J. Williams, The potential for estradiol and ethinylestradiol degradation in English rivers, Environ. Toxicol. Chem. 21 (2002) 480–488. http://www.ncbi.nlm.nih.gov/pubmed/11883412. (Accessed 13 April 2017).
[51] F.D.L. Leusch, P.A. Neale, A. Hebert, M. Scheurer, M.C.M. Schriks, Analysis of the sensitivity of in vitro bioassays for androgenic, progestagenic, glucocorticoid, thyroid and estrogenic activity: suitability for drinking and environmental waters, Environ. Int. 99 (2017) 120–130. https://doi.org/10.1016/j.envint.2016.12.014.
[52] E.J. Tiedeken, A. Tahar, B. McHugh, N.J. Rowan, Monitoring, sources, receptors, and control measures for three European Union watch list substances of emerging

concern in receiving waters – a 20year systematic review, Sci. Total Environ. 574 (2017) 1140–1163. https://doi.org/10.1016/j.scitotenv.2016.09.084.
[53] A.C. Mehinto, A. Jia, S.A. Snyder, B.S. Jayasinghe, N.D. Denslow, J. Crago, D. Schlenk, C. Menzie, S.D. Westerheide, F.D.L. Leusch, K.A. Maruya, Inter-laboratory comparison of in vitro bioassays for screening of endocrine active chemicals in recycled water, Water Res. 83 (2015) 303–309. https://doi.org/10.1016/j.watres.2015.06.050.
[54] B.I. Escher, M. Allinson, R. Altenburger, P.A. Bain, P. Balaguer, W. Busch, J. Crago, N.D. Denslow, E. Dopp, K. Hilscherova, A.R. Humpage, A. Kumar, M. Grimaldi, B.S. Jayasinghe, B. Jarosova, A. Jia, S. Makarov, K.A. Maruya, A. Medvedev, A.C. Mehinto, J.E. Mendez, A. Poulsen, E. Prochazka, J. Richard, A. Schifferli, D. Schlenk, S. Scholz, F. Shiraishi, S. Snyder, G. Su, J.Y.M. Tang, B. van der Burg, S.C. van der Linden, I. Werner, S.D. Westerheide, C.K.C. Wong, M. Yang, B.H.Y. Yeung, X. Zhang, F.D.L. Leusch, Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays, Environ. Sci. Technol. 48 (2014) 1940–1956. https://doi.org/10.1021/es403899t.
[55] T.F.T. Omar, A. Ahmad, A.Z. Aris, F.M. Yusoff, Endocrine disrupting compounds (EDCs) in environmental matrices: review of analytical strategies for pharmaceuticals, estrogenic hormones, and alkylphenol compounds, TrAC Trends Anal. Chem. 85 (2016) 241–259. https://doi.org/10.1016/j.trac.2016.08.004.
[56] C. Kienle, R. Kase, I. Werner, Evaluation of Bioassays and Wastewater Quality: in vitro and in vivo Bioassays for the Performance Review in the Project “Strategy MicroPoll,”, 2011. http://www.oekotoxzentrum.ch/media/2229/2011_kienle_bioassays_micropoll.pdf. (Accessed 13 April 2017).
[57] A.R. Ribeiro, O.C. Nunes, M.F.R. Pereira, A.M.T. Silva, An overview on the advanced oxidation processes applied for the treatment of water pollutants defined in the recently launched Directive 2013/39/EU, Environ. Int. 75 (2015) 33–51. https://doi.org/10.1016/j.envint.2014.10.027.
[58] J. Rivera-Utrilla, M. Sánchez-Polo, M.Á. Ferro-García, G. Prados-Joya, R. Ocampo-Pérez, Pharmaceuticals as emerging contaminants and their removal from water. A review, Chemosphere 93 (2013) 1268–1287. https://doi.org/10.1016/j.chemosphere.2013.07.059.
[59] EU WG Chemicals, ACTIVITY: EFFECT-BASED TOOLS/METHODS for WG Chemicals as Part of the Water Framework Directive CIS Work Programme (2016-2018) Endorsed by the Water Directors, 2016. Terms of References for the Effect-Based Activity - 08-07-2016.docx, https://circabc.europa.eu/sd/a/e39a14f1-ae0a-4ec7-9dec-2a8c6c98645a/Draft. (Accessed 13 April 2017).
[60] EU WG Chemicals, Priority Substances – Outcome of Water Directors Meeting, 2016. Chem 2016–12 (10) WD2016 outcome_Priority Substances.pdf, https://circabc.europa.eu/sd/a/096c7e10-1be3-4ffa-b5d3-e8cab1b86c7f/WG. (Accessed 13 April 2017).
[61] K. Bernhard, C. Stahl, R. Martens, H.-R. Köhler, R. Triebskorn, M. Scheurer, M. Frey, Two novel real time cell-based assays quantify beta-blocker and NSAID specific effects in effluents of municipal wastewater treatment plants, Water Res. 115 (2017) 74–83. https://doi.org/10.1016/j.watres.2017.02.036.
[62] R. Kase, A. Duffek, Applied Effect-Based Monitoring of EU Watch List Samples, 2017. https://circabc.europa.eu/w/browse/209ca179-1ab3-4dc1-81f9-0b3e7fdb8885. (Accessed 13 April 2017).

ANNEXE D. CARVALHO ET AL. (2014) MIXTURES OF CHEMICAL POLLUTANTS AT EUROPEAN LEGISLATION SAFETY CONCENTRATIONS: HOW SAFE ARE THEY? *TOXICOLOGICAL SCIENCES* (16 PAGES)

TOXICOLOGICAL SCIENCES 2014
doi: 10.1093/toxsci/kfu118
Advance Access publication June 23, 2014

Mixtures of Chemical Pollutants at European Legislation Safety Concentrations: How Safe are They?

Raquel N. Carvalho,[a] Augustine Arukwe,[b] Selim Ait-Aissa,[c] Anne Bado-Nilles,[c,d] Stefania Balzamo,[e] Anders Baun,[f] Shimshon Belkin,[g] Ludek Blaha,[h] François Brion,[c] Daniela Conti,[e] Nicolas Creusot,[c] Yona Essig,[i] Valentina E. V. Ferrero,[a] Vesna Flander-Putrle,[j] Maria Fürhacker,[k] Regina Grillari-Voglauer,[k] Christer Hogstrand,[l] Adam Jonáš,[h] Joubert B. Kharlyngdoh,[m] Robert Loos,[a] Anne-Katrine Lundebye,[n] Carina Modig,[m] Per-Erik Olsson,[m] Smitha Pillai,[k] Natasa Polak,[i] Monica Potalivo,[e] Wilfried Sanchez,[c] Andrea Schifferli,[p] Kristin Schirmer,[o,q,r] Susanna Sforzini,[s] Stephen R. Stürzenbaum,[i] Liv Søfteland,[n] Valentina Turk,[j] Aldo Viarengo,[s] Inge Werner,[p] Sharon Yagur-Kroll,[g] Radka Zounková,[h] and Teresa Lettieri[a,1]

a*European Commission—DG Joint Research Centre, Institute for Environment and Sustainability, Via E. Fermi 2749, 21027 Ispra (VA), Italy;* b*Norwegian University of Science & Technology, Trondheim, Norway;* c*National Institute for Industrial Environment and Risks, Verneuil en Halatte, France;* d*Reims University, France;* e*Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Rome, Italy;* f*Department of Environmental Engineering,Technical University of Denmark, Kgs Lyngby, Denmark;* g*Institute of Life Sciences, The Hebrew University of Jerusalem, Israel;* h*Faculty of Science, Masaryk University, RECETOX, Brno, Czech Republic;* i*Analytical and Environmental Sciences Division, King's College London, UK;* j*Marine Biology Station Piran, National Institute of Biology, Slovenia;* k*University of Natural Resources and Life Sciences, Vienna, Austria;* l*Diabetes and Nutritional Sciences Division, King's College London, UK;* m*Life Science Center, Örebro University, Sweden;* n*National Institute of Nutrition and Seafood Research, Bergen, Norway;* o*Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland;* p*Swiss Centre for Applied Ecotoxicology, Eawag-EPFL, Dübendorf, Switzerland;* q*ETH Zürich, Department of Environmental Systems Science, Swiss Federal Institute of Technology, Zürich, Switzerland;* r*EPF Lausanne, School of Architecture, Civil and Environmental Engineering, Lausanne, Switzerland; and* s*Department of Environmental and Life Sciences, Università del Piemonte Orientale Vercelli Novara Alessandria, Alessandria, Italy*

[1]To whom correspondence should be addressed. Fax: +39-0332786645. E-mail: teresa.lettieri@jrc.ec.europa.eu.

Received April 8, 2014; accepted June 6, 2014

The risk posed by complex chemical mixtures in the environment to wildlife and humans is increasingly debated, but has been rarely tested under environmentally relevant scenarios. To address this issue, two mixtures of 14 or 19 substances of concern (pesticides, pharmaceuticals, heavy metals, polyaromatic hydrocarbons, a surfactant, and a plasticizer), each present at its safety limit concentration imposed by the European legislation, were prepared and tested for their toxic effects. The effects of the mixtures were assessed in 35 bioassays, based on 11 organisms representing different trophic levels. A consortium of 16 laboratories was involved in performing the bioassays. The mixtures elicited quantifiable toxic effects on some of the test systems employed, including i) changes in marine microbial composition, ii) microalgae toxicity, iii) immobilization in the crustacean *Daphnia magna*, iv) fish embryo toxicity, v) impaired frog embryo development, and vi) increased expression on oxidative stress-linked reporter genes. Estrogenic activity close to regulatory safety limit concentrations was uncovered by receptor-binding assays. The results highlight the need of precautionary actions on the assessment of chemical mixtures even in cases where individual toxicants are present at seemingly harmless concentrations.

Key words: **bioassays; effects; mixtures; ecotoxicology; biomarkers.**

ABBREVIATIONS

DEQ	Diuron equivalent
E2	17β-estradiol
EE2	Ethinylestradiol
EEQ	Estradiol equivalent
EQS	Environmental Quality Standard
AA-EQS	Annual average EQS
MAC-EQS	Maximum allowed concentration EQS
WFD	Water Framework Directive
ER	Estrogen receptor

In Europe, as in most other industrialized regions of the world, diverse classes of chemical pollutants are released into the aquatic environment, mainly from agriculture, industry, medical facilities, and household waste. The European Union (EU) Directive 2000/60/EC (Water Framework Directive, WFD) has established a strategy for water protection that includes specific measures for pollution control to achieve good ecological and chemical status at the European level. Good chemical status is defined in terms of compliance with the safety limit concentration for substances of concern (Environmental Quality Standards, EQS) which are aimed to ensure that they do not cause any harmful effects to or via the aquatic environment. For technical and economic reasons, there is a tendency to limit chemical analysis to already regulated substances that are known to pose a threat to humans or aquatic organisms. How-

ever, environmental samples are usually very complex and can contain numerous natural and anthropogenic chemicals, even though most are present at very low concentrations. When assessing the chemical status of an aquatic environment, the individual substance EQS values are considered as safety limits, disregarding the very likely scenario of a combined action of co-occurring pollutants. Although it has been assumed that safety factors applied to the derivation of EQS values protect against the combined action of pollutants, there has been a growing concern from both scientists and the public regarding this issue. In response, the European Commission has issued a communication on combination effects of chemicals (COM 2012-252) asking for a stronger effort to ensure that the risks associated with chemical mixtures are properly understood and assessed. Biological based assays (bioassays) offer the possibility to monitor the overall response from multiple chemicals in an environmental sample and assess the impact on different levels of biological organization, such as community, population, individual and/or sub-organism levels. However, different bioassays are rarely tested on identical samples and therefore available information on the comparability, complementarity, and potential uses of the different bioassays is severely lacking.

To address the challenges posed by mixtures of pollutants to the water quality monitoring, artificial mixtures were created and effects measured using diverse bioassays, including non-OECD standards, to investigate the response to identical samples.

Two mixtures were prepared, Mix14 and Mix19, with 14 and 19 substances of concern, respectively, at concentrations equivalent to the Annual Average Environmental Quality Standard (AA-EQS). The substances were selected to include a wide range of chemical groups with known toxicological effects. Mix14 contained priority substances (PSs) whose quality standards were taken from European legislation (COM 2011-876, 2008/105/EC, 2013/39/EU3), whereas Mix19 contained five additional emerging pollutants that may become PSs in the future, selected by taking into account their prevalence in European surface waters (Loos *et al.*, 2009, 2013) and their known effects.

Thirty five *in vitro* and *in vivo* bioassays routinely used by the participating laboratories were performed. The selection of bioassays took into account the endpoints and trophic levels commonly used for the risk assessment of chemicals under European legislation (EC 1907/2006), whereas other bioassays measured endpoints associated with the expected mode of action of substances present in the mixtures. The assessed endpoints included acute toxicity (in microalgae, bacteria, yeast, amoeba, nematode, and cell lines), immunotoxicity in fish, fish embryo toxicity (FET), frog teratogenicity, estrogenic activity, the response of several molecular biomarkers in transgenic bacteria, yeast and nematode, and gene expression analysis of molecular biomarkers in cell lines. The tests were carried out using 11 organisms from different trophic levels, microcosm, several cell lines, and biomarker reporter systems.

To our knowledge, this is the first time that such a complex mixture, harboring different classes of chemicals at regulatory safety concentrations, has been tested using such a broad range of bioassays and test organisms. This paper describes the outcome of this exercise, focusing specifically on the results of the bioassays that exhibited a significant quantifiable effect of the mixtures at concentrations considered safe for each compound.

MATERIALS AND METHODS

Preparation of Reference Mixtures

Mixtures Mix14 and Mix19 contained the chemicals listed in Table 1 at concentrations equivalent to the AA-EQS, which for simplification is designated from now on as EQS. For each mixture, 1000-fold concentrated reference materials were prepared, with organic compounds in methanol and inorganic chemicals in 2% nitric acid. Additional 10,000-fold concentrated reference materials were prepared for Mix14 to allow the assessment of effects at a wider range of concentrations. The chemicals used for the preparation of the reference mixtures were of ≥98% purity, whereas for BaP and DEET the purity was ≥96 and ≥97%, respectively.

The short-term stability of the organic reference materials was assessed according to an isochronous study (ISO Guide 35, 2006) in order to simulate problematic transport or storage conditions with a reference temperature of −20°C and a test temperature of 24°C for up to 8 weeks. During the isochronous study, no significant degradation was observed in all the reference materials produced and dispatched, as checked by applying a two-tailed t-test with 99% as confidence level (for details, see Supplementary Materials and Methods). The organic and inorganic reference materials were transported in dry ice and stored in all laboratories under the reference temperatures of −20°C and 4°C, respectively. It was therefore assumed that the reference mixtures used by the different laboratories were identical, at least until reconstitution. Mixtures or solvent control (SC) (methanol and 2% nitric acid) was directly diluted into bioassay media following a common protocol and tested at final concentrations of 1× and 10×EQS for Mix14 and 1×EQS for Mix19, unless stated otherwise.

Marine Microcosm

Seawater (SW) was collected at the middle of the Gulf of Trieste (45° 32’ 55, 68” N, 13° 33’ 1, 89”E) at depth of chlorophyll maximum on 18 July 2013. Sampling was performed using a Niskin sampler and the SW was immediately pre-filtered through a 53-μm acid-washed Nitex filter to remove larger phytoplankton grazers. All samples were kept at environmental temperature, protected from light, and brought to the Marine Biology Station, Piran within 1 h after sampling. The time zero sample was taken before distributing the water into acid-washed and sterilized 1-l bottles. Each exposure mixture was added directly to 1 l of SW and triplicates were generated for

TABLE 1
Composition of Chemicals in the Reference Mixtures

Substances	CAS[b]	Use	Mode of action/reported effects	AA-EQS (μg/l)
Atrazine	1912-24-9	Herbicide	Photosystem II inhibitor	0.6[c]
Benzo[a]pyrene (BaP)	50-32-8	By-product of incomplete combustion of organic material	Intercalation of BaP metabolites in DNA causing mutagenesis, carcinogenesis	0.00017[c]
Cadmium	7440-43-9	Industrial by-product; used in metal plating and to make pigments, batteries, and plastics.	Indirect formation of reactive oxygen species, depletion of glutathione, lipid peroxidation	0.08[c]
Chlorfenvinphos	470-90-6	Insecticide	Inhibition of cholinesterase activity	0.1[c]
Chlorpyrifos	2921-88-2	Insecticide	Inhibition of cholinesterase activity	0.03[c]
DEHP	117-81-7	Plasticizer	DNA damage, carcinogenicity	1.3[d]
Diclofenac	15307-79-6	Pharmaceutical pain killer; non-steroidal anti-inflammatory drug (NSAID)	Can cause adverse hepatic effects in certain organisms	0.1[d]
Diuron	330-54-1	Herbicide	Photosystem II inhibitor	0.2[c]
17β-estradiol	50-28-2	Natural estrogen	Natural estrogen	0.0004[d]
Fluoranthene	206-44-0	Product of incomplete combustion	Causes mutagenesis, carcinogenesis	0.0063[c]
Isoproturon	34123-59-6	Herbicide	Photosystem II inhibitor	0.3[c]
Ni	7440-02-0	Industry, preparation of alloys	Depletion of glutathione levels, binds to sulfhydryl groups of proteins, carcinogenicity	4[c]
4-nonylphenol	25154-52-3	Mostly used for the production of surfactants (nonylphenolethoxylates)	Endocrine disruptor	0.3[c]
Simazine	122-34-9	Herbicide	Photosystem II inhibitor	1[c]
Carbamazepine[a]	298-46-4	Pharmaceutical (anti-epileptic, mood-stabilizing drug)	Teratogenicity	0.5[e]
Sulfamethoxazole[a]	723-46-6	Pharmaceutical (antibiotic)	Interferes with folic acid synthesis	0.6[e]
Triclosan[a] (Irgasan)	3380-34-5	Anti-bacterial and antifungal agent used in cosmetics and detergents	Inhibition of cellular efflux pumps	0.02[e]
N,N-diethyl-m-toluamide (DEET)[a]	134-62-3	Insect repellent	Affects insect odorant receptors, inhibits cholinesterase activity (nervous system)	41[e]
Bisphenol A[a]	80-05-7	Plasticizer	ER agonist	1.5[e]

[a]Used only in Mix19 (in addition to the other chemicals also present in Mix14).
[b]Chemical Abstracts Service.
[c]According to European Directive 2013/39/EU.
[d]Taken from COM 2011-876.
[e]Proposal from Ecotox Centre, Switzerland.

each treatment. At the same time, two sets of controls were prepared in triplicate: SC (0.1% methanol (v/v) and 0.002% nitric acid in 1-l SW) and SW without any addition. All bottles were incubated in a thermostatic room at constant temperature (15°C) and day/night light conditions. The pH was adjusted to standard SW pH (8.3) with 0.1-M NaOH. After 6, 12, 24, and 48 h of exposure, equal volumes were taken from each of the triplicate bottles for bacterial production and phytoplankton pigment analyses.

Bacterial production was measured as protein synthesis rates of plankton bacteria population using the ^{3}H-leucine incorporation method (Smith and Azam, 1992) and expressed as the number of cells/l/h, using 20-fg C bacterium^{-1} as the conversion factor.

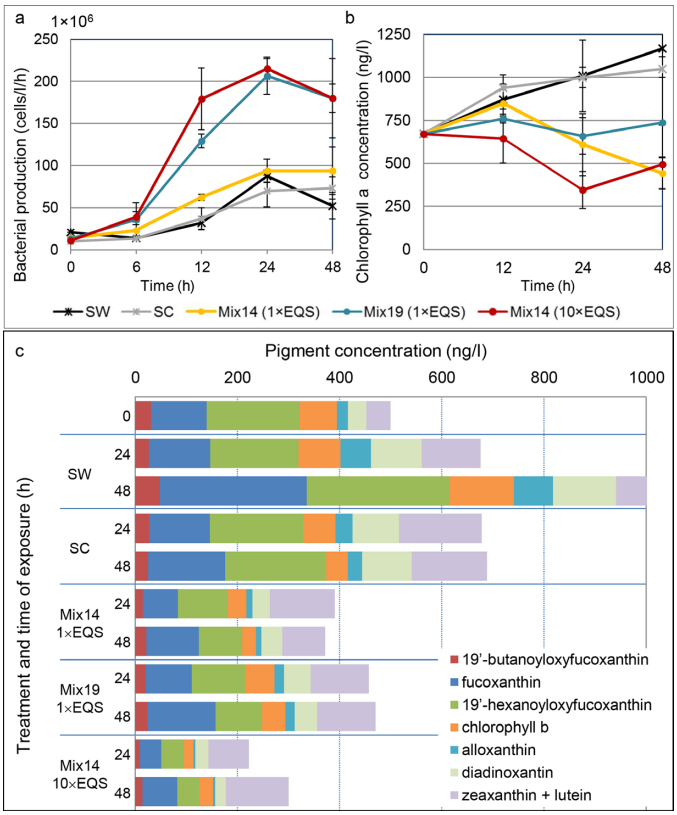


FIG. 1. Marine microcosm. Effect of the chemical mixtures on the natural phytoplankton and bacterioplankton community. Endpoints measured were bacterial production (a), chlorophyll *a* concentration (b), and other phytoplankton pigments (c). For comparison, identical SW samples have been left untreated (SW) or were exposed to SC. Error bars represent the standard deviation ($n = 3$).

The qualitative and quantitative analyses of phytoplankton pigments in the water samples were determined using a reverse-phase HPLC (high performance liquid chromatography) method (Barlow *et al.*, 1993). Water samples were filtered through Whatman GF/F filters, extracted in 90% acetone, sonicated and centrifuged for 10 min at 4000 rpm to remove particles. The supernatant was mixed with 1-M ammonium acetate (1:1), the pigments were separated by RP-HPLC using a 3-μm C18 column (Pecosphere, 35 × 4.5 mm, Perkin Elmer) and detected by absorbance at 440 nm using a diode array detector. The data were statistically evaluated using two-way ANOVA.

Freshwater Microalgae

Cultures of three microalgal species in exponential growth phase were exposed to the test mixtures and the effects on growth rate and photosynthesis (for freshwater algae only) were assessed. SC at equivalent dilutions as the reference mixtures was tested in parallel. The tests were conducted with three replicates for each treatment. Sigmoidal curves were fitted to the data with GraphPad Prism 5 Software (La Jolla, CA, USA). The EC_{50} and EC_{10} values were calculated from the fit.

Pseudokirchneriella subcapitata cultures with a cell density of 2×10^5 cells/ml were exposed to samples in 96-well plates according to Escher *et al.* (2008). The two mixtures were tested at concentrations ranging from 0.03× to 100×EQS for Mix14 and from 0.8× to 100×EQS for Mix19. Diuron was used as a reference compound and the data expressed as diuron-equivalent concentration (DEQ), by multiplying the relative potencies of the photosystem II (PSII) inhibitors diuron, atrazine, isoproturon, and simazine with their known concentration in the mixture (Vermeirssen *et al.*, 2010).

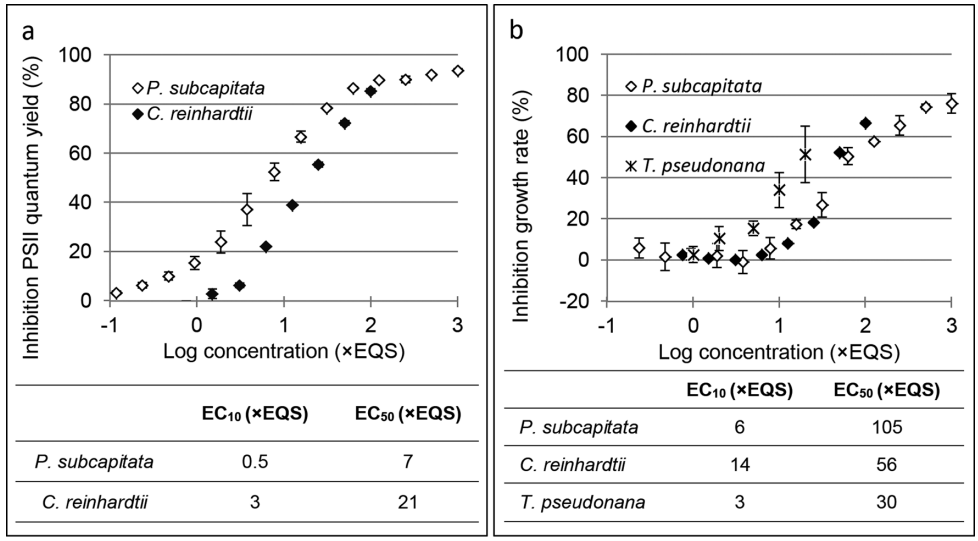
PSII inhibition was measured via the effective quantum yield method using a Maxi-Imaging PAM (pulse amplitude modulation, IPAM) (Walz, Effeltrich, Germany) as described previously (Escher *et al.*, 2008) after 2- and 24-h of exposure. Algae growth was measured by absorbance (685 nm) in a microtiter plate photometer (Synergy 4, Biotek, Winooski, VT) after 2-, 20-, and 24-h exposure. Freshwater algal growth inhibition measurements with *P. subcapitata* were performed by three laboratories for longer exposure times (72 h and 96 h) with Mix14 (1× and 10×EQS) and Mix19 (1×EQS).

Chlamydomonas reinhardtii (CC-125, wild-type mt+137c) was cultured in Talaquil medium, as reported previously (Pillai *et al.*, 2014). The growth conditions were 25°C with constant agitation and illumination of 100 μmol photon m^{-2} s^{-1}. *C. reinhardtii* (2.5×10^5 cells/ml) were exposed to Mix14 for 24 h in a total volume of 20 ml. A dose-dependent response of Mix14 ranging from 0.7× to 100×EQS was investigated. The growth rate was estimated by measuring the cell numbers by CASY counter (Roche Innovatis AG, Switzerland). The photosynthetic yield was determined after 2 h and 24 h with PhytoPAM (Heinz Wald Gmbh, Germany).

Thalassiosira pseudonana (strain CCMP 1335) was obtained as axenic culture from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP, West Boothbay Harbour, Maine, USA) and cultured in artificial seawater (ASW-f/2) at 16°C and photoperiod 13/11-h light/dark. *T. pseudonana* cultures were synchronized according to Hildebrand *et al.* (2007) and exposed to the mixtures at cell density of 1×10^6 cells/ml in a total volume of 20 ml. A dose-dependent response of Mix14 ranging from 1× and 20×EQS and Mix19 at 1×EQS were investigated after 24, 48, and 72 h. Cell densities were determined by measuring the absorption at 450 nm using a microplate spectrophotometer (Biorad, Hercules, CA) and used to calculate growth rates and growth inhibition, as previously described (Bopp and Lettieri, 2007).

Daphnia Magna *Acute Immobilization test*

The test followed the ISO 6341 (2012) standard method. Five newly hatched neonates (age <24 h) were placed in glass beakers (100 ml) and exposed to the mixtures in the dark at 18–22°C. Four replicates were made per treatment (i.e., 20 animals per treatment and 20–40 animals in the control group). The number of immobile animals was counted after 24 and 48 h. Potassium dichromate was used as a reference compound, with an



	EC_{10} (×EQS)	EC_{50} (×EQS)
P. subcapitata	0.5	7
C. reinhardtii	3	21

	EC_{10} (×EQS)	EC_{50} (×EQS)
P. subcapitata	6	105
C. reinhardtii	14	56
T. pseudonana	3	30

FIG. 2. Cytotoxicity to microalgae. Dose response curves of Mix14 were generated for the inhibition of photosynthesis after 2-h exposure (a) and inhibition of growth after 24-h exposure (b) of the freshwater microalgae *P. subcapitata* and *C. reinhardtii* and the growth of marine diatom *T. pseudonana*. The x-axis is displayed as concentration of Mix14, in terms of EQS. The EC_{10} and EC_{50} values obtained from the fit of the data are shown for each of the endpoints. No effect from exposure to the solvent was observed for any of the organisms. Error bars represent the standard deviation, $n = 3$.

EC_{50} of 1.8 mg/l (95% CI, 1.7–1.9 mg/l), fulfilling the validity criteria in the ISO standard of an EC_{50} between 0.9 and 2.4 mg/L22.

The concentration-response relationships were calculated with the ToxCalc software (Ver 5.0) (Tidepool) with maximum likelihood logit regression.

D. Magna *Reproduction Test*

The test followed the OECD Test No. 211 (2012) and the ISO 10706 (2001) guidelines, with newly hatched daphnids placed separately in glass beakers. Exposure to the mixtures, control, and solvent occurred at 21 ± 1°C and photoperiod 16/8-h light/dark (10 animals per condition). During 21 days of exposure, the survival and the reproduction were monitored. Exposure mixtures were changed three times a week and daphnids were fed with green algae (Pseudokirchneriella, Chlorella, and Scenedesmus spp.). Offspring produced by parent animals were counted and removed. Survival of parent animals and the number of live offspring were evaluated and expressed as a percentage of control.

Mean, standard deviation, and the number of replicates were used for statistical evaluation using GraphPad QuickCalc online software, and statistical significance of differences between control and exposure mixtures was tested by unpaired t-test.

FET Test

The FET test was conducted according to the OECD TG. 236 (2013) and the ISO 15088 (2008) guidelines with zebrafish (*Danio rerio*) embryos. Fertilized eggs were exposed to the mixtures under static conditions for 5 days: 10 embryos per 40-ml media and three replicates per treatment in two independent experiments. Embryos were monitored daily for mortality, the number of hatched embryos, type of deformities (head, tail deformities, absence of gas bladder) and the number of defected embryos, underdeveloped embryos and length.

Statistical evaluation of the data was done by ANOVA followed by Dunnett and Fisher LSD *post hoc* test (for data in individual experimental runs). Homogeneity of variance and normality were tested by Levene and Shapiro-Wilk tests, respectively. Nonparametric Kruskal-Wallis test was used for data without normal distribution and a Chi-square test was used for testing differences in frequencies. Statistica for Windows (StatSoft) and Microsoft Excel were used for calculations.

Frog Embryo Teratogenesis Assay Xenopus

The test followed the ASTM E 1439-98 (1998) guideline and was performed under constant temperature (20°C) and low light. *Xenopus laevis* adults were maintained in 20-l plastic tanks in dechlorinated tap water (males and females together, four animals per tank) and were fed with a mixture of ground beef liver, lung, and heart with gelatin and reptile multivitamin mix. Room and water temperature was 19°C, 12-h day/night rhythm.

Two breeding pairs were placed in separated plastic tanks equipped with bottom plastic nets, thermostats set to 23°C, and bubblers. Both males and females were stimulated with human chorionic gonadotropin (females 500 IU and males 300 IU) in the form of Pregnyl 5000 (N.V. Organon, Holland) injected into the dorsal lymph sac. Eggs were staged according to Nieuwkoop and Faber (1994). After reaching stage 46, normally cleaving embryos were manually collected from the tank with a plastic

dropper and placed in sterile plastic Petri dishes for the exposure to the mixtures or SC, in five replicates, each containing 30 embryos in 10 ml of solution. Solutions were changed every 24 h, and dead embryos were removed. After 96 h, embryos from each dish were moved to test tubes and anesthetized with 5 ml of 100-mg/l tricainemethanesulfonate, and then fixed with 5 ml of 3% formaldehyde. The embryos were observed with a light microscope, digitally photographed, and measured with QuickPhoto MICRO software. The parameters evaluated in this test included mortality, embryo length, and the number and type of malformations and were assessed according to the Atlas of Abnormalities (Bantle, 1991).

Differences from controls were analyzed by ANOVA followed by Dunnett and Fisher Least Significant Difference *post hoc* test and the results controlled by nonparametric Kruskal-Wallis test.

In vitro *Human Estrogen Receptor Transactivation Assays*

The detection of (anti)estrogenic activity by the ER-CALUX, the MELN, and the Yeast Estrogen Screen (YES) assays is based on stably transfected transcriptional activation of responsive elements (luciferase for the two former assays and β-galactosidase for the last). The results in these tests were expressed as EC_{50} (the concentration causing 50% of the maximum effect) as well as estradiol equivalent (EEQ) concentration, which were derived from chemical and bioassay data (Vindimian *et al.*, 1983).

ER-CALUX. The reference mixtures were reconstituted in MQ water, subjected to solid phase extraction, and diluted in dimethyl sulfoxide (DMSO) prior to the exposure. Human U2-OS osteosarcoma cells stably transfected with estrogen receptor alpha (ERα) were seeded into 96-well plates in DMEM/F12 medium (without phenol red) and supplemented with stripped serum. After 24 h of incubation (37°C, 5% CO_2), the medium was replaced by medium containing the water extracts (1% DMSO). A dose-dependent response ranging from 1× to 1000×EQS was investigated for Mix14 and from 1× to 100×EQS for Mix19. After 24 h of incubation, the medium was removed and the cells were lysed in 30 µl of Triton-lysis buffer. The amount of luciferase activity was quantified using a luminometer (MicroLumat Plus, Berthold Technologies, Switzerland). All plates included a dose-response curve of the reference compound 17β-estradiol. All mixtures and estradiol were analyzed in triplicates. Only test concentrations where no cytotoxicity was observed using a microscope were used for quantification of the response (Van der Linden *et al.*, 2008). The data were evaluated by fitting a dose-response using GraphPad Prism 5 Software (La Jolla, California, USA).

MELN assay. The MELN cell line was obtained by stable transfection of MCF-7 human breast cancer cells with ERα (Balaguer *et al.*, 2001). Cells were seeded into 96-well plates at a density of 50,000 cells/well in phenol red free DMEM supplemented with 3% stripped serum. After 24 h of incubation (37°C, 5% CO_2), the mixtures, the reference compound 17β-estradiol, and SC were added in fresh medium. A dose-dependent response ranging from 0.12× to 475×EQS was investigated for Mix14 and from 0.08× to 26×EQS for Mix19. After overnight exposure (18 h), 0.3mM of D-luciferin was added to the wells. After 5 min, the luminescence signal was measured in living cells for 2 s/well using a luminometer (µBeta, Wallac). All mixtures, estradiol, and SC were analyzed in triplicates. Modelling of dose-response curves was done using the Regtox Microsoft Excel macro based on the Hill equation model.

YES assay. The YES was performed according to Routledge and Sumpter (1996) with recombinant yeast *Saccharomyces cerevisiae* provided by John Sumpter (Brunel University, Uxbridge, UK). At test initiation, 1:2 dilution series of the reference substance 17β-estradiol, the mixtures, and SC (ethanol) were pipetted into triplicate wells on 96-well plates and the solvent was evaporated completely under sterile conditions. Suspension with 4×10^7 yeast cells was seeded on the test plate (200 µl/well) and incubated at 30°C. After 72 h, cell density ($OD_{620\,nm}$) and color change ($OD_{540\,nm}$) were measured using a plate reader (Synergy 2, Biotek). A dose-dependent response ranging from 0.8× to 1000×EQS was investigated for Mix14 and from 0.8× to 100×EQS for Mix19. The data were fitted to a sigmoidal curve with GraphPad Prism 5 Software (La Jolla, CA, USA). The fit provided the EC_{10} and EC_{50} as well as EEQ.

In vitro *Human ERα Competition Assay*

To test the binding ability of the recombinant receptor we used the PolarScreen ERα competitor green assay developed by Life Technologies, with a recombinant wild-type ERα ligand binding domain ($wtER\alpha^{LBD}$) (Ferrero *et al.*, 2014). The assay is based on the displacement of the Fluormone ES2 from the ER receptor by competitor molecules and a consequent decrease in the maximum fluorescence signal. The intensity of the fluorescence polarization (P) signal was measured with an Infinite 200 Pro multimode plate reader (Tecan).

A dose-dependent response ranging from 0.01× to 200×EQS was investigated for Mix14 and from 0.001× to 20×EQS for Mix19. 17β-estradiol was used as a reference compound. The data were fitted to a sigmoidal one site competition four parameters logistic curve with OriginPro Software. The fit provided the IC_{50} (concentration of test compound required to reduce the maximum polarization value at 50%) as well as EEQ. IC_{50} values were obtained by the average of at least four different experiments.

Zebrafish Embryo Estrogenic Activity Assay

The estrogenic potency of the mixtures was assessed by the *in vivo* test EASZY (Detection of *E*ndocrine *A*ctive *S*ubstances acting through human ER, using transgenic cyp19a1b-GFP zebrafish embryos) (Brion *et al.*, 2012). Newly fertilized ze-

brafish eggs were exposed to the mixtures for 96 h under static condition. A range of three dilutions was tested, from 0.04× to 4×EQS for Mix14 and from 0.04× to 0.4×EQS for Mix19, with 17α-ethinylestradiol (EE2) (0.05nM) as a reference compound. Three independent experiments were performed. At the end of each experiment, the fluorescence of each living zebrafish embryo was acquired using a fluorescence microscope and quantified using ImageJ. The data (expressed as mean fold induction above control) were analyzed to determine the estrogenic activity of each mixture using a parametric two-way ANOVA and post-hoc test using R statistical software.

Escherichia Coli Bioluminescent Reporter Strains

A panel of 12 engineered bioluminescent microbial reporters was studied, each harboring a plasmid-born fusion of a stress responsive gene promoter (*recA*, *katG*, *micF*, *zntA*, *arsR*, *fabA*, *grpE*, *marR*, *cydA*, *sodA*, *yqjF*, and *soxS*; see Supplementary table 2) to a bioluminescence gene cassette (*Photorhabdus luminescens luxCDABE*) (van der Meer and Belkin, 2010).

The reporter strains were grown overnight in 170-μl lysogeny broth (LB) medium supplemented with 100-μg/ml ampicillin. The cultures were diluted 100-fold in M9 medium and regrown with shaking at 37°C for 3 h. Culture aliquots were transferred into an opaque white 96-well microtiter plate (Greiner Bio-One) and diluted 1:1 with the mixture or the individual model chemical as a positive control (see Supplementary table 2). Each mixture was tested in a concentration series ranging from 0.08× to 5×EQS; additional concentrations up to 50×EQS were tested for Mix14.

Luminescence was measured at 37°C for 10-min intervals using a VICTOR^2 plate reader (Wallac, Turku, Finland) and displayed as arbitrary relative luminescence units (RLUs). Activity was calculated as the difference in the intensity of the signal in the presence and absence of the inducer (ΔRLU) (Belkin *et al.*, 1997). All experiments were carried out in duplicate and repeated at least three times. The lowest concentration detected was determined as the concentration at which the ΔRLU was >2, and was validated by the use of a paired t-test.

Caenorhabditis Elegans *Bioluminescent Reporter Strains*

Five *Caenorhabditis elegans* transgenic strains were used: *cyp-35A2* (58cop (25.3.47)), *mtl-2* (62cop (6.15.47)), *ugt-1* (59cop (8.13.47)), *gst-38* (54cop (7.7.47)), and *gcs-1* (23cop (5.23.47)). Each strain was dual-labeled, by linking the promoter of the biomarker to the coding region of a Red Fluorescent Protein (mCherry) and an invariant transmembrane vesicular GABA transporter, *unc-47,* to the coding region of a green fluorescent protein (GFP). All strains were maintained at 20°C on nematode growth medium (NGM) agar plates that were seeded with *Escherichia coli* (OP50).

The exposure mixtures and SC were prepared in OP50 and tested in parallel and BaP (100 μg/ml) and $CdCl_2$ (100μM) were used as positive controls for *cyp35A2* and *mtl-2*, respectively. NGM agar plates (20-ml volume) were inoculated with 200 μl of the spiked OP50 and the seeded plates were incubated at room temperature for 24 h (to allow for bacterial growth). All strains were aged synchronized, placed (as L1 larvae) on the NGM plates and exposed to the respective conditions for 48 h at 20°C. Single worms were picked onto a glass slide with a drop of M9 and immobilized with sodium azide (2%). Images were captured with a Nikon DS-2Mv digital camera and NIS-Elements F 2.20 software linked to a Nikon ECLIPSE TE2000-S-inverted microscope, using the filters G-2A (Ex 510nm–560nm) for mCherry and FITC (Ex 465nm–495nm) for GFP. The fluorescence intensities from 10 worms per condition were analyzed using ImageJ.

For the growth size assay, wild-type nematodes ($N = 10$ per condition) were plated on NGM plates (containing the OP50 medium with the mixtures) and maintained up to 120 h. Adult nematodes were transferred to new plates between 72 h and 96 h to remove hatched offspring. Images of worms were obtained using an inverted microscope and the flat volumetric surface area and length determined by tracing the nematodes using the Image-Pro Express software (Media Cybernetics, Inc.). Data obtained from the fluorescence experiments were analyzed using the one-way ANOVA followed by the Tukey's multiple comparison test for significant differences between the treatments. The phenotypic assays were assessed by means of the two-way ANOVA. All tests were executed with GraphPad Prism.

Gene Expression Analysis with Quantitative Real-Time PCR

Cell lines were from and maintained according to ATCC. Human epithelial cervix cells (HeLa) were cultured in Dulbecco's Modified Eagle Medium (DMEM) + 10% Fetal Bovine Serum (FBS). Chicken epithelial hepatocellular (LMH) cells were cultured in Waymouth's MB + 10% in 0.1% gelatin-coated flasks. Both cell lines were kept at 37°C, 5% CO_2. Zebrafish epithelial liver (ZFL) cells were cultured in 50% L-15/ 35% DMEM High glucose/ 15% Ham's F12 supplemented with 5% FBS, 15-mM HEPES, 0.15-g/l sodium bicarbonate, 1X Insulin-Transferrin-Selenium at 28°C and 3% CO_2. The exposure mixtures or solvent was reconstituted in MQ water and immediately before use mixed with cell culturing medium (1:4) to get the desired exposure concentration, with no effect on the pH of the cell culturing media. Cells were plated in 6- or 12-well plates, and after 18–20 h exposed to the mixtures. HeLa and LMH cells were treated for 24 h and ZFL for 40 h, $n = 4$.

Following exposure the cells were lysed and total RNA was isolated using the NucleoSpin RNA II kit (Macherey-Nagel, Germany) and quantified by Nano-Vue (GE Healthcare). cDNA synthesis followed the qScript cDNA synthesis kit (Quanta Biosciences) and real-time qRT-PCR of each sample was performed in triplicate using the KAPA SYBR FAST qPCR kit (Kapa Biosystems) on an Mx 3000P qPCR machine (Stratagene). The thermocycling conditions were as follows; denaturation 5 min at 95°C followed by 40 cycles of 95°C for 2 s and 60°C for 30 s. The obtained Ct values were normalized using

elongation factor 1 alpha 1 (*eef1a/1*) and relative gene expression was determined using the ΔΔCt method (Schmittgen and Livak, 2008). The primers used and the genes they are directed against are listed in Supplementary table 3. These included androgen receptor (AR), ERα, ER beta (ERβ), metallothionein (MT2A), cytochrome P450, family 1 subfamily A, polypeptide 1 (CYP1A1), glutathione S-transferase, cyclooxygenase-2 (COX2), interleukin-6 (IL6), interleukin-8 (IL8), and tumor suppressor protein (p53). Data variance were analyzed using the GraphPad Prism 5 software by one-way (ANOVA) followed by Dunnet post-test for multiple group comparison.

RESULTS

The effects of two chemical mixtures were assessed for a wide range of biological endpoints and organisms from different trophic levels (for a complete overview see Table 2).

Effect on a Marine Microcosm Composition

Natural bacterioplankton and phytoplankton communities were altered by both Mix14 and Mix19 mixtures. Bacterioplankton population exposed to Mix14 and Mix19 was able to grow at rates significantly higher ($p < 0.0001$) than SC and untreated SW (Fig. 1a). Conversely, after 24 h of incubation the phytoplankton biomass, expressed as chlorophyll *a* concentration, decreased significantly compared with both controls, where an increase (up to 900 ng/l) was recorded (Mix14 at 10×EQS $p < 0.0001$; Mix19 at 1×EQS $p < 0.003$; Mix14 at 1×EQS $p < 0.02$) (Fig. 1b). At the same time, the phytoplankton composition, assessed in terms of chemotaxonomic pigments, changed in Mix14 10×EQS, Mix14 1×EQS, and less in Mix19 10×EQS. A major decrease in pigment concentration was recorded for silicoflagellates (19′-butanoyloxyfucoxanthin), diatoms (fucoxanthin), prymnesiophytes (19′-hexanoyloxyfucoxanthin), but much less for cryptophytes (alloxanthin) and green algae (chlorophyll *b)* (Fig. 1c). A significant increase was observed only for cyanophytes (zeaxanthin + lutein) in all treatments.

Effects on Microalgae

The chemical mixtures induced acute toxicity in the three microalgae tested. The limit of detection of toxic compounds in the mixture was lower for PSII inhibition than growth (Fig. 2). PSII was significantly inhibited in the freshwater algae exposed for 2 h to Mix14, with EC_{50} at 7×EQS for *P. subcapitata* and 21×EQS for *C. reinhardtii* (Fig. 2a). A similar response was obtained for the exposure of *P. subcapitata* to Mix19, with EC_{50} at 13×EQS.

The growth rate of all three species was reduced in a dose-dependent manner (Fig. 2b) after 24-h exposure to Mix14, with an EC_{50} of 30 (*T. pseudonana*) < 56 (*C. reinhardtii*) < 105 (*P. subcapitata*) ×EQS. The growth inhibition assays with *P. subcapitata* performed for 72 h and 96 h of exposure by other three laboratories measured no significant effect at 1×EQS for either Mix14 or Mix19, similar to the results obtained at 24-h exposure. Exposure to Mix14 at a higher concentration (10×EQS) in the three laboratories led to inhibition of *P. subcapitata* growth by 31, 13, and 14%, respectively.

Effects on D. Magna

The calculated EC_{50} values for acute immobilization at 24-h and 48-h exposure to Mix14 was 8× and 2.8×EQS, respectively (Fig. 3a). Additionally, the results with Mix14 at 10×EQS were comparable among the three laboratories performing the bioassay (Fig. 3b). Both mixtures at 1×EQS did not induce any significant effect on the acute immobilization of *D. magna* neither in the chronic reproduction test. However, exposure to Mix14 at 10×EQS proved to be highly toxic with longer exposure times leading to 100% mortality after 3 days.

Embryo Toxicity and Development

After exposure for 5 days, effects in several endpoints related to FET were observed at 1×EQS for Mix19 and 10×EQS for Mix14, as detailed in Table 3. Effects specifically observed included mortality, a change in the number of hatched embryos, head deformations, tail deformations, absence of gas bladder, generally underdeveloped embryos, and embryo length (examples shown in Fig. 4). On shorter times of exposure, only higher concentrations of the mixture triggered significant effects in FET, particularly in terms of the number of defective embryos after 72 h and in the number of hatched embryos after 96 h (Table 3).

The studied mixtures also impaired the development of frog embryo. Using the Frog Embryo Teratogenesis Assay Xenopus (FETAX), 43 ± 12% and 34 ± 14% malformed frog embryos were observed for exposure to 1×EQS of Mix14 and Mix19, respectively, whereas exposure to 10×EQS of Mix14 caused 62 ± 10% malformed embryos. The effects were significantly different from SC (ANOVA, Dunnett post-test, $p < 0.05$), which proved to be moderately toxic (15 ± 12% malformed embryos). The most commonly observed malformations in FETAX included incomplete gut coiling and skeletal malformations such as flexed and waivy tail (see Fig. 4 and Table 3). Eye deformities or thoracic edema were also recorded in lower frequency.

In the bioassays using the nematode *C. elegans*, growth was uniform among the different treatments with the mixtures or solvent during the first 72 h (namely the larval stages L1–L4), but started to deviate after worms had reached adulthood. Nematodes chronically exposed (from L1 stage) to Mix19 at 1×EQS were marked by a statistically significant reduction in final length after 120 h (see Supplementary fig. 2). Though smaller in final size, these worms nevertheless reached adulthood and were able to reproduce, suggesting that the observed phenotype did not affect developmental or reproductive indices.

TABLE 2
Summary of Bioassays, Results, and Partner Laboratories in the EU-Wide Exercise

Organism/test	Biological endpoint	Exposure	Effects	EC_{50} (×EQS)	Comments
Microcosmos in marine water	Bacteria production and pigment concentration	6, 12, 24, 48 h	Increase in bacterioplankton decrease in phytoplankton	-	pH adjusted
Vibrio fischeri, Microtox EN ISO 11348-3	Inhibition bioluminescence	15, 30 min	No toxicity effect, stimulation of luminescence	-	pH adjusted
Escherichia coli (luminescent transgenic organisms)[a]	Induction of biomarkers	up to 800 min	Mix14: *zntA*, *arsR* induction Mix19:*cydA*, *micF* induction	-	-
Pseudokirchneriella subcapitata ISO 8692	Growth inhibition	24 h	Effect observed <10×EQS	105 (Mix14) 116 (Mix19)	72, 96 h tested in some labs
Pseudokirchneriella subcapitata	Inhibition of photosynthesis (PSII)	2 h	Effect observed <10×EQS	7.3 (Mix14) 12.6 (Mix19)	-
Chlamydomonas reinhardtii	Growth inhibition	24, 48, 72 h	Effect observed <10×EQS	56 (Mix14)	Mix19 tested only at 1×EQS
	Inhibition of photosynthesis (PSII)	2, 24 h	Effect observed <10×EQS	19.2 (Mix14)	
Thalassiosira pseudonana	Growth inhibition	24, 48, 72 h	Effect observed <10×EQS	28 (Mix14)	Mix19 tested only up to 2×EQS
Saccharomyces cerevisiae	Growth	8 h	No effect	-	-
	Genotoxicity	8 h	No effect	-	-
	Acute toxicity	4 h	Acute toxicity significant ($p < 0.05$) only >25×EQS	-	-
(Transgenic fluorescent) *Daphnia magna* EN ISO 6341	Acute immobilization	24, 48 h	Effect observed <10×EQS (Mix14)	7 (24 h) 3.4 (48 h)	Mix19 tested up to 2×EQS
Daphnia magna CSN ISO 10706	Reproduction test	21 days	100% mortality after 3 days at 10×EQS (Mix14)	-	No effect at 1×EQS with respect to SC
Caenorhabditis elegans	Growth	120 h	Effect in development for Mix19 (1×EQS)	-	Growth uniform between exposures until 72 h, deviating after 96 h
	Lipid accumulation	48 h	Increased accumulation of lipids in storage compartments (Mix14 10×EQS)	-	Mix19 tested only at 1×EQS
	Pharyngeal pumping	48, 72 h	No effect on food intake (pharyngeal pumping)	-	-
	Movement	48, 72, 96 h	No significant trends in movement	-	-
Caenorhabditis elegans Dual-fluorescent transgenic organisms	Induction of several stress response proteins	48 h	Mix19 (1×EQS) induced expression of gst-38, involved in phase II detoxification	-	No effect on mtl-2, ugt-, gcs-1, and Cyp-35a2-
Danio rerio FET (EN ISO 15088)	FET	120 h	Malformations observed for Mix14 (10×EQS) and Mix19 (1×EQS)	-	Mix14 (1×EQS) no effect
Xenopus laevis	Frog embryo teratogenicity, embryo malformation	96 h	Mix14 (10×EQS): 62 ± 10%;	-	15 ± 12% malformed embryos in SC

TABLE 2
Continued

Organism/test	Biological endpoint	Exposure	Effects	EC_{50} (×EQS)	Comments
FETAX ASTM E 1439-98			Mix14 (1×EQS): 43 ± 12%; Mix19 (1×EQS): 34 ± 14% No effect on embryo length		
Dictyostelium discoideum (soil-living amoeba)	Lysosomal membrane stability	3 h	Effects statistically not different from the solvent	-	-
	Replication	24 h	No effect	-	-
Gasterosteus aculeatus (Three-spined stickleback) *Ex vivo* splenic leucocyte immune activities	Leucocyte distribution Cellular mortality Respiratory burst Lysosomal membrane integrity Phagocytosis activity	18 h	No effect on any of the endpoints tested	-	-
MTT assay, cell lines: RTG-2	*In vitro* cytotoxicity	72 h	No effect	-	-
		20 h	No effect	-	-
RPTEC/TERT1, HepG2, MCF7 HUVEC/TERT			No effect	-	-
Neutral red test H4IIE-luc cells	Acute cytotoxicity		No effect	-	-
xCELLigence Primary hepatocytes cultures, juvenile Atlantic salmon (*Salmo salar* L.)	Cytotoxicity system	up to 120 min	No effect	-	-
Atlantic salmon (*Salmo salar* L.)	ELISA (Vtg, Zrp regulation)	5 days	No effect	-	Maximum concentration tested was 0.16 EQS
Regulation biomarkers	qRT-PCR (Vtg, ERα, Zrp)	5 days	No effect	-	
HeLa, LMH, ZFL cells[b] Regulation biomarkers qRT-PCR	Gene expression	24, 40 h	HeLa: regulation of MT2A, AR, p53, GSTK1, IL6, IL8 LMH: regulation of IL8	-	No effect ZFL cells
YES	ER-binding activity	72 h	Activity measured for Mix14 and Mix19	92.3 (Mix14) 90.5 (Mix19)	-
ER-CALUX		24 h	Activity measured for Mix14 and Mix19	4.9 (Mix14) 4.7 (Mix19)	-
ER-activated luciferase induction MELN cells		18 h	Activity measured for Mix14 and Mix19	34.2 (Mix14) 13.3 (Mix19)	-
$wtER\alpha^{LBD}$ binding assay		2 h	Binding measured for Mix14 and Mix19	IC_{50} 74.9 (Mix14) IC_{50} 7.8 (Mix19)	-
EASZY, *in vivo* transgenic zebrafish larvae		96 h	Activity measured for Mix14 above 4×EQS	-	-
PLHC-1 cells EROD induction	Dioxin-like activity	24 h	No effect	-	-
AR-CALUX	AR-binding activity	24 h	No effect	-	-
AR-activated luciferase induction MDA-kb2 cells		18 h	No effect	-	-

TABLE 2
Continued

Organism/test	Biological endpoint	Exposure	Effects	EC_{50} (×EQS)	Comments
PPAR-CALUX	PPAR γ2-like activity	24 h	No effect	-	-
PXR-activated luciferase induction, HG5LN-PXR cells	PXR-binding activity	18 h	Effect >10×EQS	-	-

[a] All tested reporter genes are detailed in Supplementary table 2.
[b] All tested reporter genes are detailed in Supplementary table 3.

TABLE 3
Effect of Mixtures on the FET Test with Zebrafish and the FETAX

	Time	Endpoint	Chemical mixture		
			Mix14 10×EQS	Mix14 1×EQS	Mix19 1×EQS
FET	72 h	Number of defected embryos	a	-	-
	96 h	Number of hatched embryos	a	-	-
	120 h	Number of defected embryos	a, c	-	a
		Head deformities	a	-	-
		Absence of gas bladder	a	-	-
		Underdeveloped embryos	a	-	a, c
FETAX	96 h	Total number of malformed embryos	a	a	a
		Incomplete gut coiling	a	-	a
		Tail malformation	a	-	a

a: endpoint significantly different from SC (chi-square test, $p < 0.05$); c: endpoint significantly different from SC (ANOVA followed by Fisher LSD *post hoc* test).

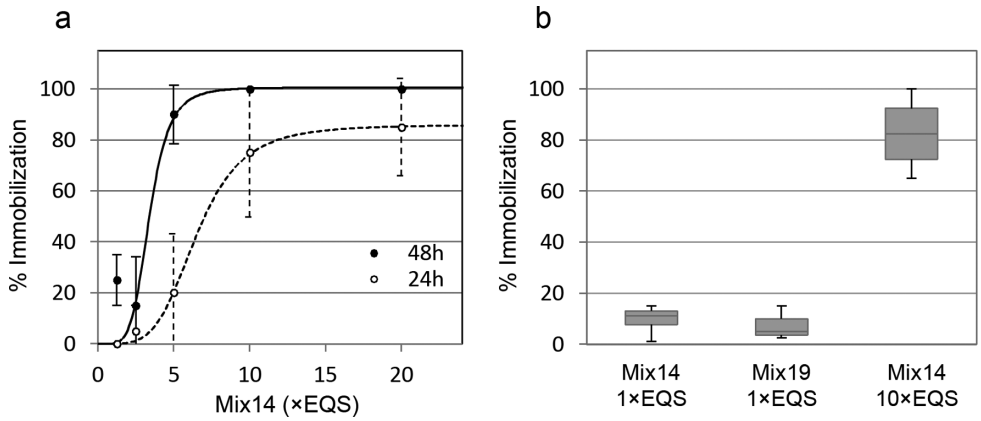


FIG. 3. Acute immobilization in *D. magna*. (a) Dose response of Mix14 in EQS equivalent concentrations, for immobilization at 24-h exposure (open symbols) and 48-h exposure (closed symbols). The lines represent the fit of non-linear regression model to the data for the calculation of the EC_{50}. Error bars represent the standard deviation, $n = 4$. (b) Combined immobilization data from three different laboratories for Mix14 (at 1× and 10×EQS) and Mix19 (at 10×EQS).

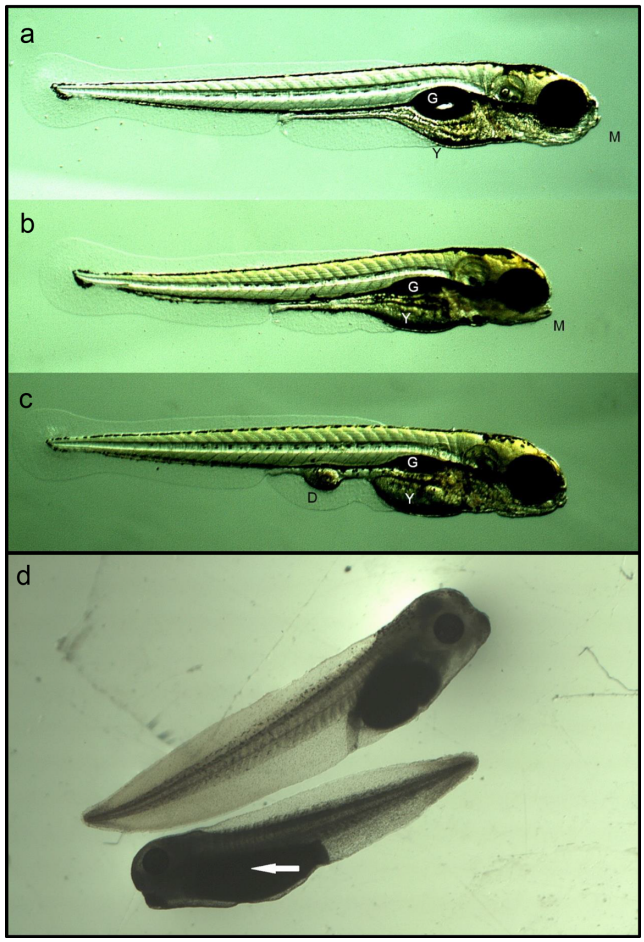
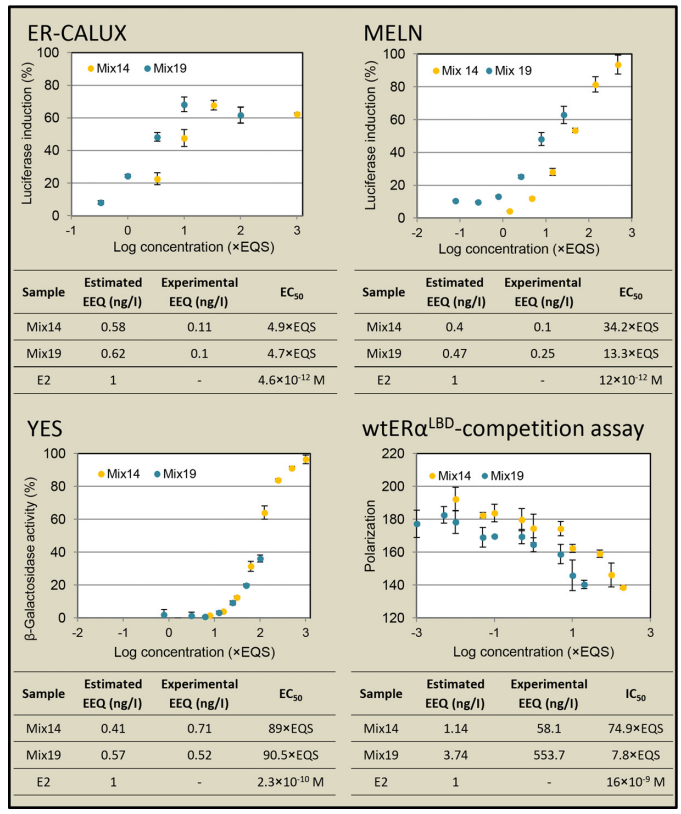


FIG. 4. Embryos of *Danio rerio* from the FET (a)–(c) and *Xenopus laevis* from FETAX (d). (a) Control fish embryo 120-h post fertilization. (b) Embryo exposed to Mix14 at 10×EQS for 120 h - typical underdeveloped (smaller) embryo with non-inflated gas (swimming) bladder (G), deformed head especially at the mouth region (M), and not fully consumed yolk (Y). (c) Embryo from the same exposure as in panel (b) with highlighted deformation nearby the anal region (D), non-inflated gas bladder (G), and not fully consumed yolk (Y). (d) Control 96-h embryo of *X. laevis* (upper individual) compared with underdeveloped and malformed embryo exposed for 96 h to MIX19 1xEQS (the arrow shows the incomplete intestine coiling, which was the most frequent malformation observed).



Sample	Estimated EEQ (ng/l)	Experimental EEQ (ng/l)	EC_{50}
Mix14	0.58	0.11	4.9×EQS
Mix19	0.62	0.1	4.7×EQS
E2	1	-	4.6×10^{-12} M

Sample	Estimated EEQ (ng/l)	Experimental EEQ (ng/l)	EC_{50}
Mix14	0.4	0.1	34.2×EQS
Mix19	0.47	0.25	13.3×EQS
E2	1	-	12×10^{-12} M

Sample	Estimated EEQ (ng/l)	Experimental EEQ (ng/l)	EC_{50}
Mix14	0.41	0.71	89×EQS
Mix19	0.57	0.52	90.5×EQS
E2	1	-	2.3×10^{-10} M

Sample	Estimated EEQ (ng/l)	Experimental EEQ (ng/l)	IC_{50}
Mix14	1.14	58.1	74.9×EQS
Mix19	3.74	553.7	7.8×EQS
E2	1	-	16×10^{-9} M

FIG. 5. Estrogenic activity measurement using *in vitro* bioassays. Dose-dependent estrogenic activity of Mix14 and Mix19 was measured via ER-activated luminescence induction using the ER-CALUX and the MELN system, the β-galactosidase activity using the YES test, and the competition assay using the recombinant wtERα^{LBD}. The EC_{50} values are shown, calculated from the fit to the data measured with the two mixtures and of E2 in the test, as well as the estimated and experimental EEQ concentrations. The error bars represent the standard deviation, $n = 3$.

Nuclear Receptors Binding Activity

The activity of four different human receptors was assessed in this study with respect to the tested mixtures, i.e., ER, AR, peroxisome proliferator-activated receptor (PPAR), and pregnane X receptor (PXR). No activity was measured associated with the binding to the AR, PPAR in all concentrations tested, whereas PXR-mediated activity was measured only at concentrations of the mixture >50×EQS (Table 2).

Four *in vitro* methods, ER-CALUX, MELN, YES, and a competition assay with recombinant wtERα^{LBD} detected estrogenic activity of the mixtures close to the EQS concentration (Fig. 5). The model compound 17β-estradiol was used as a reference compound (EC_{50} values shown in Fig. 5) with the three ER-mediated transactivation assays yielding EC_{50} values that were similar to those previously reported (Leusch *et al.*, 2010). Estrogenic activity was detected at lower concentrations of the mixtures for the ER-CALUX, followed by the MELN assay, the recombinant ERα competition assay, and finally the YES assay (Fig. 5).

In addition, the *in vivo* EASZY test was performed using transgenic zebrafish larvae. In this test, Mix14 induced GFP expression in a dose-dependent manner, which was significant at 4×EQS (Fig. 6), whereas for Mix19, tested only up to 0.4×EQS, no effect was observed.

Molecular Biomarkers

Among the bioluminescent *E. coli* reporters, the sensor elements exhibiting the lowest detection thresholds for Mix14 were the *zntA* and *arsR* gene promoters, indicating the presence of heavy metals at concentrations higher than 6.2×EQS (Supplementary fig. 1). In Mix19, the *micF* gene promoter (indicator of

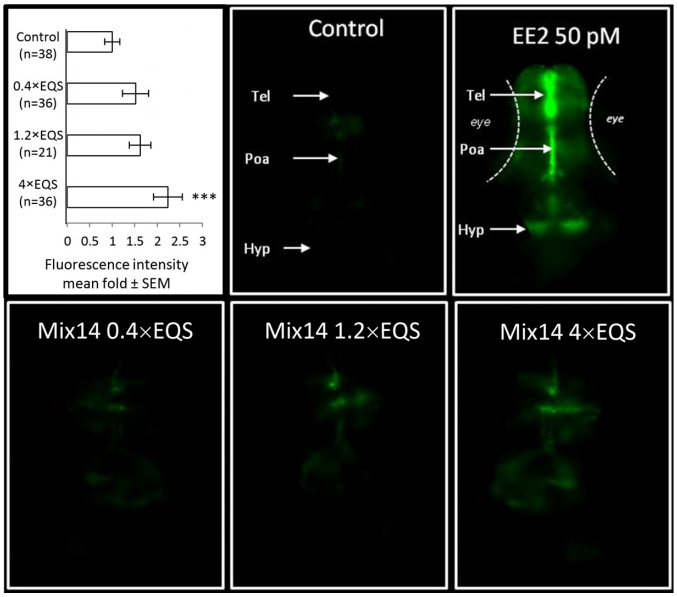


FIG. 6. *In vivo* estrogenic activity of Mix14 as shown by induction of GFP in 96-hpf-old transgenic cyp191ab-GFP zebrafish larvae. Exposure was done at different concentrations of Mix14, during 96 h from fertilization, under static condition, after which fluorescence imaging on living zebrafish was performed. GFP was expressed in various brain regions in radial glial cells. Dorsal view, magnification X10, Tel: telencephal; Poa: preoptic area; Hyp: inferior lobe of hypothalamus. EE2 50pM was used as positive control. The mean fluorescent intensity is shown in the graph, indicating the number of larvae imaged for each condition (n), ***$p < 0.001$. EE2 led to a 26-fold induction.

chemical-induced oxidative stress) and *cydA* (indicator of respiratory inhibition) were induced above 0.16× and 5×EQS, respectively (Supplementary fig. 1).

In addition, a transgenic *C. elegans* strain, carrying the red fluorescent protein reporter gene under the promoter of the glutathione-S-transferase *gst-38*, was responsive to Mix19. GST is a protein involved in phase II detoxification and its induction was significant ($p < 0.05$) in Mix19 at 1×EQS, but not in Mix14, even at 10×EQS (Supplementary fig. 2).

Finally, the expression of several genes was modified in HeLa cells following exposure to the mixtures (Supplementary fig. 3). The highest regulation was found for the *IL6* gene with an increase by 4-fold in Mix19 and by 2.5-fold in Mix14 at 1×EQS. The other regulated genes showed a decreased expression (<2-fold decrease) in Mix14 (at 1× and 10×EQS), but not in Mix19, and included the *AR*, *mt2A, GSTK1, IL8*, and *p53* genes (Supplementary fig. 3). None of the tested genes responded to the mixtures in the ZFL cells. In LMH cells, only *IL8* showed a small downregulation following exposure to Mix14 at 10×EQS.

Additional bioassays performed in the exercise either displayed no effect with the mixtures or measured an effect only at concentrations higher than 10×EQS (Table 2). Some widely used bioassays did not detect an effect of the mixtures at low concentrations. This was the case of the acute toxicity bioassay with *Vibrio fischeri*, which was tested in four different laboratories, with a measured EC_{50} around 400× and 200×EQS for Mix14 and Mix19, respectively.

DISCUSSION

In the last few years, concern over the impact of chemical mixtures on human and ecosystem health has been highlighted by the scientific community and brought to the attention of the European Commission (SCHER, SCENIHR and SCCS, 2012).

The exercise described here employed chemical mixtures at concentrations of the individual compounds believed to be safe and studied the hazard to wildlife organisms of different trophic levels. Artificial mixtures were produced as reference solutions to ensure that the chemical composition and concentrations were known, and in this way facilitate a direct association between chemical and biological effect. Such cause-effect relationships would likely be harder to reach with complex environmental samples, although this is definitely an important matter to address in the future.

By using a battery of ecotoxicity bioassays, ranging from gene-expression tests to whole organism bioassays, we demonstrate biologically relevant effects of chemical mixtures where each contaminant exists at or in some cases considerably below the EQS concentration. Effects of the mixtures at 1×EQS were observed across a wide range of taxa that included bacteria, algae, nematodes, fish and amphibians. These results seriously question the present paradigm for assessing the safety of chemicals to the environment and demonstrate that regulatory safety concentrations (EQS) may not provide sufficient protection when multiple chemicals are present.

The interpretation of the toxicity results measured in our artificial mixtures with respect to environmental samples could be a matter of discussion. Most of the chemical pollutants in environmental samples are usually found at concentrations considerably below the safety limits for toxicological effects, and concentrations exceeding the EQS values of priority pollutants are reported for only a minority of the monitored samples. A summary of a literature search on EQS exceedances from surface water monitoring data in Europe in recent years can be found in Supplementary table 1. WFD EQS exceedances (in some countries) concern usually only a small number of "ubiquitous" substances [e.g., mercury, cadmium, tributyltin, brominated diphenylethers, some polyaromatic hydrocarbons (PAHs), nickel, and Di(2-ethylhexyl)phthalate (DEHP)].

On the other hand, the number of chemicals present in environmental samples likely exceeds the 14 or 19 included in the artificial mixtures of this exercise. When multiple components in a sample, even at low concentrations, affect the same pathway, their combined toxicity can usually be described by the concentration addition concept and may induce significant toxicity to aquatic organisms (Broderius, 1990). This was confirmed in this study for the algae toxicity elicited by the four herbicides in the mixture (diuron, atrazine, isoproturon, and simazine), acting as

PSII inhibitors, and the endocrine disruptor compounds (E2, 4-nonylphenol and bisphenol A) binding to the ER and activating the expression of reporter genes.

A less predictable hazard may arise from combinations of chemicals from different classes and with different modes of action. This is the case for the well-known heavy metal modulation of cytochrome P450 1A1 (CYP1A1) expression and activity, responsible for xenobiotic metabolism and activation (Anwar-Mohamed *et al.*, 2009). Another example is the inhibition by several contaminants of cellular efflux pumps, which are multi-xenobiotic resistance transporters, thus potentiating the cellular accumulation and toxicity of other chemicals. This mechanism has been reported in echinoid larvae (Anselmo *et al.*, 2012) as well as in zebrafish embryo (Fischer *et al.*, 2013).

The fact that the correlation between ecological and chemical indicators has not been straightforward in the implementation of the WFD, further substantiates the need for complementary indicators. The assessment of biological effects in key trophic organisms could play this part in linking ecological and chemical assessment by providing the combined toxicity from all chemicals present.

This study shows that co-occurring chemicals can elicit an effect in some ecologically relevant and surrogate organisms in a manner that may imbalance the entire ecosystem. The concentrations selected for each chemical in the mixtures were that of the AA-EQS, a safety threshold under European legislation aiming to protect the environment from chronic toxicity effects. However, the mixture at AA-EQS in this study was able to induce effects in both chronic and acute toxicity tests. Even stronger toxicological effects were visible when the mixtures were tested at concentrations corresponding to the maximum allowed concentration (MAC-EQS), as indicated by the responses in several of the bioassays.

At the lower trophic level, the study showed that the mixtures at EQS equivalent concentrations affected the bacteria-phytoplankton composition in a marine microcosm, with a significant reduction in the phytoplankton community and an increase in the bacteria population. The increase in bacterial growth rate might be due to fast selection of bacteria that are capable of utilizing selected pollutants or dissolved organic carbon released by decaying phytoplankton. Unfortunately, no measurements of dissolved organic compounds were performed simultaneous with the treatments to assess this possibility. An imbalanced composition of bacteria/plankton population would likely influence the ecosystem functioning (food wed, biodiversity, ecosystem services) (Naeem *et al.*, 2000).

No effect was observed at the AA-EQS equivalent concentration of the mixtures at the single species level for the three microalgae (*P. subcapitata*, *C. reinhardtii*, and *T. pseudonana*), indicating that this value is sufficiently protective when considering only four herbicides with a similar mode of action. However, at concentrations of the mixture corresponding to the MAC-EQS, an effect was measured for the PSII inhibition endpoint.

Going up in the trophic levels, other endpoints for which an effect was observed close to EQS concentrations included the acute immobilization of *D. magna* and effects on toxicity and development of fish and frog embryos. Several of the substances in the mixtures have been described as embryotoxic or teratogenic. These include the pharmaceuticals sulfamethoxazole and carbamazepine (Richards and Cole, 2006), chlorpyrifos (Bonfanti *et al.*, 2004), atrazine (Fort *et al.*, 2004), the polyaromatic hydrocarbons BaP (Fort *et al.*, 1989), and fluoranthene (Hatch and Burton, 1998), E2 and bisphenol A (Saili *et al.*, 2013). However, the effects of these substances have been reported only at concentrations exceeding those currently detected in surface waters and the ways they interfere with developmental processes is poorly understood. Their combined action cannot directly explain the observed toxicity of the mixtures to fish and frog embryos in this study. Developmental effects and daphnia immobilization are general endpoints that may be triggered by a multitude of substances, molecular targets, and intercalating events. They represent a bigger challenge in linking the observed effect from the mixture to specific substances.

Diverse and unpredictable combinatorial effects of mixtures have been well documented, when the individual substances appear safe when tested alone, including for endocrine disrupting chemicals (EDCs) with other compounds (Fagin, 2012).

Additional responses of the mixtures at concentrations close to EQS values were measured in this exercise by estrogen-receptor mediated *in vitro* and *in vivo* bioassays. Several chemical substances released into the environment are able to mimic the action of natural hormones by binding to the ER and may show estrogenic activity, thereby influencing the sexual function and differentiation in aquatic organisms. Some of the substances included in the mixtures are known ligands for the ER, including the natural estrogen 17β-estradiol, 4-nonylphenol, bisphenol A, and possibly triclosan although with lower potency (Svobodová *et al.*, 2009; Torres-Duarte *et al.*, 2012). It is possible that also other substances in the mixtures may bind to the ER in an agonist or antagonist way. Binding of different compounds in the mixtures to the ER without activation of the downstream pathway could explain the highest experimental EEQ in the wtERα^{LBD} competition assay, with respect to the estimated EEQ. A difference between estimated and experimental EEQ was also observed in the ER-CALUX and MELN assays and may be the result of a mixture antagonistic effect, although this requires further investigation. Binding of several molecules to hERα is well known and proven also by co-crystallization of the receptor (Baker, 2011). The binding can occur in an agonist or antagonist way. This suggests a wide flexibility of the ligand binding domain to accommodate chemically different structures into its active site.

The *in vitro* tests used in this study are suitable assays for monitoring of estrogenic activity in water samples, and interestingly, the estrogenic activity was further confirmed in intact fish embryos as measured by the brain-specific upregulation of the ER-mediated cyp191ab expression during early and criti-

cal developmental stages. The rising interest in bioassays as alternative tools for the detection of estrogens in water close to the European regulatory limits lies in the fact that EQS values of estrogenic compounds of concern (E2 and EE2) are below the analytical limits of quantification of most routine chemical methods (Loos, 2012).

We could show that exposure to mixtures of dissimilarly acting substances at concentrations considered environmentally acceptable can exert significant effects on the biota. In this exercise, the bioassays showed i) general comparability among the laboratories for the same assay, ii) complementarity covering several trophic levels of the ecosystem, and iii) potential for the future implementation in water management as holistic approaches for the ecological risk assessment of chemicals under realistic conditions.

Chemical monitoring alone cannot assess the quality status of water impacted by anthropogenic mixtures. Bioassays can be included in the workflow, and their selection should be based on the outcome of a risk assessment of the specific water body, taking into account the known sources of pollutants (e.g., agriculture, industry, household, hospital, etc.), expected concentrations but also considering the methods cost, technical time, and concentration range applicability. In any case, there is no "one size fits all" bioassay that could provide the toxicological potency of every mixture toward all aquatic organisms in all water bodies, but rather a battery of bioassays that should be selected as "fit for purpose". Whether the focus is on low concentration of pollutants such as those found in most fresh and marine waters, or higher concentration of pollutants, e.g., in wastewater treatment plant effluents, different batteries of bioassay can be selected to provide a snapshot of the ecosystem health.

Furthermore, the use of tailor-made reference mixtures with rather-characterized modes of acting chemicals, as described in this study, could i) aid the "quantification" of the observed effects in terms of toxicity units, ii) allow intercalibration among laboratories using the same bioassay, and iii) help establishing a threshold for "no observed mixture effect" in future regulatory applications.

In conclusion, the present study highlights an urgent need to revise tools and paradigms used to assess the safety of chemicals to the environment. Bioassays as part of a multi-tier approach to water quality monitoring can fill the gap between chemical and ecological assessments for a more holistic characterization of water quality.

SUPPLEMENTARY DATA

Supplementary data are available online at http://toxsci.oxfordjournals.org/.

FUNDING

RADAR (265721—Collaborative Project FP7-KBBE-2010-4); French Office Water and Aquatic Environment; French Ministry for Ecology and Sustainable Development (P181 DRC50).

ACKNOWLEDGMENTS

Authors from JRC thank Ramon Lavado, Joaquin Pinto-Grande, and Simona Tavazzi for technical assistance and Giovanni Bidoglio and Bernd Manfred Gawlik for their encouragement. Authors from RECETOX acknowledge technical assistance of Alena Nekvapilová and Mirco Dalla Bona. King's College London thanks Dr Chris Hopkins (Knudra Transgenics) for the generous provision of transgenic *C. elegans* stains. Daniela Baumgartner and Dr Marion Junghans at the Swiss Centre for Applied Ecotoxicology are thanked for technical assistance and data analysis. *Competing financial interests declaration*: R.G.V. is cofounder of Evercyte GmbH and declares conflict of interest. All remaining authors declare they have no competing financial interests.

REFERENCES

Anselmo, H. M., van den Berg, J. H., Rietjens, I. M., and Murk, A. J. (2012) Inhibition of cellular efflux pumps involved in multi xenobiotic resistance (MXR) in echinoid larvae as a possible mode of action for increased ecotoxicological risk of mixtures. *Ecotoxicology* **21**, 2276–2287.

Anwar-Mohamed, A., Elbekai, R. H., and El-Kadi, A. O. (2009) Regulation of CYP1A1 by heavy metals and consequences for drug metabolism. *Expert Opin. Drug Metab. Toxicol.* **5**, 501–521.

ASTM E 1439-98 (1998) *Standard Guide for Conducting the Frog Embryo Teratogenesis Assay—Xenopus*.

Baker, M. E. (2011) Insights from the structure of estrogen receptor into the evolution of estrogens: Implications for endocrine disruption. *Biochem. Pharmacol.* **82**, 1–8.

Balaguer, P., Boussioux, A. M., Demirpence, E., and Nicolas, J. C. (2001) Reporter cell lines are useful tools for monitoring biological activity of nuclear receptor ligands. *Luminescence* **16**, 153–158.

Bantle, J. A. (1991) *Atlas of Abnormalities: A Guide for the Performance of FETAX*. Oklahoma State University, Stillwater.

Barlow, R. G., Mantoura, R. F. C., Gough, M. A., and Fileman, T. W. (1993) Pigment signatures of the phytoplankton composition in the northeastern Atlantic during the 1990 spring bloom. *Deep-Sea Res.* **40**, 459–477.

Belkin, S., Smulski, D. R., Dadon, S., Vollmer, A. C., Van Dyk, T. K., and Larossa, R. A. (1997) A panel of stress-responsive luminous bacteria for the detection of selected classes of toxicants. *Water Res.* **31**, 3009–3016.

Bonfanti, P., Colombo, A., Orsi, F., Nizzetto, I., Andrioletti, M., Bacchetta, R., Mantecca, P., Fascio, U., Vailati, G., and Vismara, C. (2004) Comparative teratogenicity of chlorpyrifos and malathion on Xenopus laevis development. *Aquat. Toxicol.* **70**, 189–200.

Bopp, S. K., and Lettieri, T. (2007) Gene regulation in the marine diatom Thalassiosira pseudonana upon exposure to polycyclic aromatic hydrocarbons (PAHs). *Gene* **396**, 293–302.

Brion, F., Le Page, Y., Piccini, B., Cardoso, O., Tong, S. K., Chung, B. C., and Kah, O. (2012) Screening estrogenic activities of chemicals or mixtures

in vivo using transgenic (cyp19a1b-GFP) zebrafish embryos. *PloS ONE* **7**, e36069.

Broderius, S. J. (1990) Modeling the joint toxicity of xenobiotics to aquatic organisms: Basic concepts and approaches. In *Aquatic Toxicology and Risk Assessment* (M. A. Mayes and M. G. Barron Eds.), Vol. **14**, pp. 107–130. American Society for Tenting and Materials, Philadelphia, PA.

COM 2011-876 final—European Commission proposal for a Directive amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy.

COM 2012-252—Communication from the commission to the council—The combination effects of chemicals.

Escher, B. I., Bramaz, N., Mueller, J. F., Quayle, P., Rutishauser, S., and Vermeirssen, E. L. (2008)Toxic equivalent concentrations (TEQs) for baseline toxicity and specific modes of action as a tool to improve interpretation of ecotoxicity testing of environmental samples. *J. Environ. Monit.* **10**, 612–621.

European Directive 2008/105/EC—Environmental quality standards in the field of water policy. Official Journal of the European Union, L384/84.

European Directive 2013/39/EU—European Directive amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy. Official Journal of the European Union, L226/1.

European Regulation (EC) No 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). Official Journal of the European Union, L396/1.

Fagin, D. (2012) Toxicology: The learning curve. *Nature* **490**, 462–465.

Ferrero, V. E. V., Pedotti, M., Chiadò, A., Simonelli, L., Calzolai, L., Varanib, L., and Lettieri, T. (2014) Rational modification of estrogen receptor by combination of computational and experimental analysis. PLoS ONE, in press.

Fischer, S., Kluver, N., Burkhardt-Medicke, K., Pietsch, M., Schmidt, A. M., Wellner, P., Schirmer, K., and Luckenbach, T. (2013) Abcb4 acts as multi-xenobiotic transporter and active barrier against chemical uptake in zebrafish (Danio rerio) embryos. *BMC Biol.* **11**, 69–84.

Fort, D.J., James, B.L., and Bantle, J.A. (1989) Evaluation of the developmental toxicity of five compounds with the frog embryo teratogenesis assay: Xenopus (FETAX) and a metabolic activation system. *J. Appl. Toxicol.* **9**, 377–388.

Fort, D.J., Rogers, R.L., Thomas, J.H., Buzzard, B.O., Noll, A.M., and Spaulding, C.D. (2004) Comparative sensitivity of Xenopus tropicalis and Xenopus laevis as test species for the FETAX model. *J. Appl. Toxicol.* **224**, 443–457.

Hatch, A. C., and Burton, G. A., Jr (1998) Effects of photoinduced toxicity of fluoranthene on amphibian embryos and larvae. *Environ. Toxicol. Chem*. **17**, 1777–1785.

Hildebrand, M., Frigeri, L. G., and Davis, A. K. (2007) Synchronized growth of Thalassiosira pseudonana (Bacillariophyceae) provides novel insights into cell-wall synthesis processes in relation to the cell cycle. *J. Phycol.* **43**, 730–740.

ISO 10706 (2001) *Water quality—Determination of long term toxicity of substances to Daphnia magna Straus (Cladocera, Crustacea)*.

ISO 15088 (2008) *Water quality—Determination of the acute toxicity of waste water to zebrafish eggs (Danio rerio)*.

ISO 6341 (2012) *Water quality—Determination of the inhibition of the mobility of Daphnia magna Straus (Cladocera, Crustacea)—Acute toxicity test*.

ISO Guide 35 (2006) *Reference materials—General and statistical principles for certification*.

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

Van der Linden, S. C., Heringa, M. B., Man, H. Y., Sonneveld, E., Puijker, L. M., Brouwer, A., and Van der Burg, B. (2008) Detection of multiple hormonal activities in wastewater effluents and surface water, using a panel of steroid receptor CALUX bioassays. *Environ. Sci. Technol.* **42**, 5814–5820.

Loos, R. (2012). Analytical methods relevant to the European Commission's 2012 proposal on priority substances under the Water Framework Directive. JRC scientific and policy report.

Loos, R., Carvalho, R., António, D. C., Comero, S., Locoro, G., Tavazzi, S., Paracchini, B., Ghiani, M., Lettieri, T., Blaha, L., *et al.* (2013) EU-wide monitoring survey on emerging polar organic contaminants in wastewater treatment plant effluents. *Water Res.* **47**, 6475–6487.

Loos, R., Gawlik, B. M., Locoro, G., Rimaviciute, E., Contini, S., and Bidoglio, G. (2009) EU-wide survey of polar organic persistent pollutants in European river waters. *Environ. Pollut.* **157**, 561–568.

Van der Meer, J. R., and Belkin, S. (2010) Where microbiology meets microengineering: Design and applications of reporter bacteria. *Nat. Rev. Microbiol.* **8**, 511–522.

Naeem, S., Hahn, D. R., and Schuurman, G. (2000). Producer-decomposer codependency influences biodiversity effects. *Nature* **403**, 762–764.

Nieuwkoop, P. D., and Faber, J. (1994) *Normal Table of Xenopus laevis (Daudin)*. Garland Publishing Inc., New York.

OECD Guidelines for Testing of Chemicals—Test No. 211 (2012) *Daphnia Magna Reproduction Test*. OECD, Paris, France.

OECD Guidelines for the Testing of Chemicals—Test No. 236 (2013) *Fish Embryo Toxicity (FET) Test*. OECD, Paris, France.

Pillai, S., Behra, R., Nestler, H., Suter, M. J., Sigg, L., and Schirmer, K. (2014) Linking toxicity and adaptive responses across the transcriptome, proteome, and phenotype of Chlamydomonas reinhardtii exposed to silver. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3490–3495.

Richards, S. M., and Cole, S. E. (2006) A toxicity and hazard assessment of fourteen pharmaceuticals to Xenopus laevis larvae. *Ecotoxicology* **15**, 647–656.

Routledge, E. J., and Sumpter, J. P. (1996) Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. *Environ. Toxicol. Chem.* **15**, 241–248 .

Saili, K.S., Tilton, S.C., Waters, K.M., and Tanguay, R.L. (2013) Global gene expression analysis reveals pathway differences between teratogenic and non-teratogenic exposure concentrations of bisphenol A and 17β-estradiol in embryonic zebrafish. *Reprod. Toxicol.* **38**, 89–101.

SCHER, SCENIHR, SCCS (2012) *Toxicity and Assessment of Chemical Mixtures*.

Schmittgen, T. D., and Livak, K. J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108.

Smith, D. C., and Azam, F. (1992) A simple, economical method for measuring bacterial protein synthesis rates in seawater using 3H-leucin. *Mar. Microb. Food Webs* **6**, 107–114.

Svobodová, K., Plačková, M., Novotná, V., and Cajthaml, T. (2009) Estrogenic and androgenic activity of PCBs, their chlorinated metabolites and other endocrine disruptors estimated with two in vitro yeast assays. *Sci. Total Environ.* **407**, 5921–5925.

Torres-Duarte, C., Viana, M., and Vazquez-Duhalt, R. (2012) Laccase-mediated transformations of endocrine disrupting chemicals abolish binding affinities to estrogen receptors and their estrogenic activity in zebrafish. *Appl. Biochem. Biotech.* **168**, 864–876.

Vermeirssen, E. L., Hollender, J., Bramaz, N., van der Voet, J., and Escher, B. I. (2010) Linking toxicity in algal and bacterial assays with chemical analysis in passive samplers deployed in 21 treated sewage effluents. *Environ. Toxicol. Chem.* **29**, 2575–2582.

Vindimian, E., Robaut, C., and Fillion, G. (1983) A method for cooperative or noncooperative binding studies using nonlinear regression analysis on a microcomputer. *J. Appl. Biochem.* **5**, 261–268.

ANNEXE E. DIPAOLO ET AL. (2016) BIOASSAY BATTERY INTERLABORATORY INVESTIGATION OF EMERGING CONTAMINANTS IN SPIKED WATER EXTRACTS - TOWARDS THE IMPLEMENTATION OF BIOANALYTICAL MONITORING TOOLS IN WATER QUALITY ASSESSMENT AND MONITORING. *WATER RESEARCH* (12 PAGES)

Water Research 104 (2016) 473–484

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts – Towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring

Carolina Di Paolo [a], Richard Ottermanns [a], Steffen Keiter [a, b], Selim Ait-Aissa [c], Kerstin Bluhm [a], Werner Brack [d], Magnus Breitholtz [e], Sebastian Buchinger [f], Mario Carere [g], Carole Chalon [h], Xavier Cousin [i, j], Valeria Dulio [c], Beate I. Escher [d, k, l], Timo Hamers [m], Klára Hilscherová [n], Sergio Jarque [n], Adam Jonas [n], Emmanuelle Maillot-Marechal [c], Yves Marneffe [h], Mai Thao Nguyen [o], Pascal Pandard [c], Andrea Schifferli [p], Tobias Schulze [d], Sven Seidensticker [a, l], Thomas-Benjamin Seiler [a], Janet Tang [k], Ron van der Oost [q], Etienne Vermeirssen [p], Radka Zounková [n], Nick Zwart [m], Henner Hollert [a, *]

[a] Institute for Environmental Research, RWTH Aachen University, Aachen, Germany
[b] Man-Technology-Environment Research Centre, School of Science and Technology, Örebro University, Örebro, Sweden
[c] INERIS, Verneuil-en-Halatte, France
[d] UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany
[e] Department of Applied Environmental Science - ITM, Stockholm University, Stockholm, Sweden
[f] Department Biochemistry and Ecotoxicology, Federal Institute of Hydrology, Koblenz, Germany
[g] Italian Institute of Health, Rome, Italy
[h] ISSeP (Scientific Institute of Public Service), Liège, Wallonia, Belgium
[i] Laboratoire d'Ecotoxicologie, Ifremer, L'Houmeau, France
[j] Laboratoire de Physiologie et Génétique des Poissons, Inra, Rennes, France
[k] National Research Centre for Environmental Toxicology - Entox, The University of Queensland, Brisbane, Australia
[l] Centre for Applied Geosciences, Eberhard Karls University Tübingen, Germany
[m] Institute for Environmental Studies -IVM, VU University Amsterdam, The Netherlands
[n] Research Centre for Toxic Compounds in the Environment - RECETOX, Faculty of Science, Masaryk University, Brno, Czech Republic
[o] Waterproef Laboratory, Edam, The Netherlands
[p] Swiss Centre for Applied Ecotoxicology Eawag-EPFL, Dübendorf, Switzerland
[q] WATERNET Institute for the Urban Water Cycle, Division of Technology Research & Engineering, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 12 May 2016
Received in revised form
30 July 2016
Accepted 9 August 2016
Available online 12 August 2016

Keywords:
Triclosan
Acridine

ABSTRACT

Bioassays are particularly useful tools to link the chemical and ecological assessments in water quality monitoring. Different methods cover a broad range of toxicity mechanisms in diverse organisms, and account for risks posed by non-target compounds and mixtures. Many tests are already applied in chemical and waste assessments, and stakeholders from the science-police interface have recommended their integration in regulatory water quality monitoring. Still, there is a need to address bioassay suitability to evaluate water samples containing emerging pollutants, which are a current priority in water quality monitoring. The presented interlaboratory study (ILS) verified whether a battery of miniaturized bioassays, conducted in 11 different laboratories following their own protocols, would produce comparable results when applied to evaluate blinded samples consisting of a pristine water extract spiked

* Corresponding author. Department of Ecosystem Analysis, Institute for Environmental Research (Biology 5), RWTH Aachen University, Worringerweg 1, D-52074, Aachen, Germany.
E-mail address: henner.hollert@bio5.rwth-aachen.de (H. Hollert).

http://dx.doi.org/10.1016/j.watres.2016.08.018

17α-ethinylestradiol
3-Nitrobenzanthrone
Organism-level toxicity
Mechanism-specific toxicity

with four emerging pollutants as single chemicals or mixtures, i.e. triclosan, acridine, 17α-ethinylestradiol (EE2) and 3-nitrobenzanthrone (3-NBA). 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 fluctuation) assays. The test battery presented complementary sensitivity and specificity to evaluate the different blinded water extract spikes. Aquatic organisms differed in terms of sensitivity to triclosan (algae > daphnids > fish) and acridine (fish > daphnids > algae) spikes, confirming the complementary role of the three taxa for water quality assessment. Estrogenicity and mutagenicity assays identified with high precision the respective mechanism-specific effects of spikes even when non-specific toxicity occurred in mixture. For estrogenicity, although differences were observed between assays and models, EE2 spike relative induction EC_{50} values were comparable to the literature, and E2/EE2 equivalency factors reliably reflected the sample content. In the Ames, strong revertant induction occurred following 3-NBA spike incubation with the TA98 strain, which was of lower magnitude after metabolic transformation and when compared to TA100. Differences in experimental protocols, model organisms, and data analysis can be sources of variation, indicating that respective harmonized standard procedures should be followed when implementing bioassays in water monitoring. Together with other ongoing activities for the validation of a basic bioassay battery, the present study is an important step towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring.

1. Introduction

Water quality investigation and monitoring in Europe and worldwide is facing a challenge. There is societal, regulatory and scientific consensus on the urgent need to achieve good water quality in national and transboundary river basins. Meanwhile, an immense variety of contaminants is constantly reaching aquatic systems, which complicates the identification of drivers of chemical toxicity to be routinely monitored (von der Ohe et al., 2011). Further, there is a lack of direct indicators on the regulatory level to verify the biological relevance of chemical monitoring in different water bodies. While the ecological status assessment is certainly of high environmental relevance, it is based primarily on biodiversity indices that often do not present consistency with respective chemical monitoring (Wernersson et al., 2015). Therefore, complementary monitoring strategies are required to achieve the Water Framework Directive (WFD) aim to maintain and improve water quality in Europe (EC, 2000).

Effect-based tools such as bioassays and biomarkers are particularly useful to bridge the gap between chemical contamination and ecological status, since they can cover a broad range of toxicity mechanisms in diverse organisms, and account for additional risks posed by non-target compounds and mixtures. Bioassays already provide the regulatory basis to derive environmental quality standards (EQS) (EC, 2011) and to evaluate pelagic toxicity under the REACH authorization process (ECHA 2014). They are also applied to assess effluents from domestic wastewater treatment plants and industrial sectors (OSPAR, 2007; Gartiser et al., 2009). Moreover, the recommendation to integrate bioassays in regulatory water quality monitoring (Hecker and Hollert, 2011; Hamers et al., 2013; Wernersson et al., 2015) is supported by many tests being available as standardized methods (OECD guidelines, ISO standards). However, there are still open questions that prevent their application in effect-based monitoring of water bodies. A major issue is whether reliable results can be achieved when evaluating effects of samples containing diverse aquatic pollutants and chemical mixtures. Particularly, the evaluation of emerging contaminants, such as pharmaceuticals, personal care and disinfection products, is a current priority in regulatory water quality monitoring (Loos et al., 2009; von der Ohe et al., 2012).

In response to that, the present interlaboratory study (ILS) was developed as a collaborative exercise to investigate whether a battery of miniaturized bioassays would produce consistent results for the evaluation of blinded samples containing pristine water extract spiked with representative emerging pollutants as single-chemicals or mixtures. These included:

(i) Triclosan, a chlorinated phenoxy phenol used as biocide in personal care and household products, already suggested as candidate priority substance (von der Ohe et al., 2012);
(ii) 17α-ethinylestradiol (EE2), a synthetic estrogenic human and veterinary pharmaceutical recently included in the European chemical watch list for water quality monitoring (EC, 2013; Kunz et al., 2015);
(iii) Acridine, a heterocyclic aromatic hydrocarbon of industrial origin and a carbamazepine transformation product found in aquatic sediments and groundwater (Hartnik et al., 2007; de Voogt and Laane, 2009);
(iv) 3-Nitrobenzanthrone (3-NBA), a potent mutagenic diesel exhaust component that occurs in aquatic sediments and rainwater (Murahashi et al., 2003; Lübcke-von Varel et al., 2012).

The water extract included a realistic environmental matrix as a sample component, increasing the relevance of the study for water quality assessment. Methods evaluated effects on organisms from three trophic levels (algae, daphnids, fish) and mechanism-specific effects using *in vitro* estrogenicity and mutagenicity assays. The resulting interlaboratory trial brings a novel approach since, with very few exceptions (Carvalho et al., 2014; Escher et al., 2014), previous bioassay ILS focused on only one or few methods, a single mode of action, or single chemical or sample (Hoss et al., 2012; Reifferscheid et al., 2012; Feiler et al., 2014). Finally, a unique aspect of this study that is reflected in the discussion is the clear aim to promote the regulatory use of bioassays for water quality monitoring at the European policy-makers level.

2. Material and methods

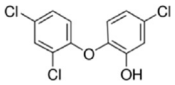
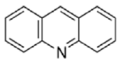
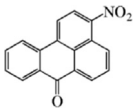
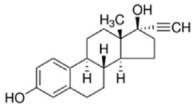
2.1. Chemicals

Information on the test chemicals is provided in Table 1.

2.2. Participant institutes and design of study

The study was coordinated by the Department of Ecosystem

Table 1
Chemical properties of the compounds used for water extract spiking.

Chemical	CAS number	Formula	Supplier	Purity	Structure	Molecular weight (g mol^{-1})	logK_{ow}	Solubility in water (mgL^{-1})
Triclosan	3380-34-5	$C_{12}H_7Cl_3O_2$	Sigma-Aldrich (Germany)	≥97%		289.6	4.76 [a]	10 (20 °C) [a]
Acridine	260-94-6	$C_{13}H_9N$	Merck (Germany)	>98%		179.2	3.40 [a]	38.4 mg/L (24 °C) [a]
3-Nitrobenzanthrone (3-NBA)	17117-34-9	$C_{17}H_9NO_3$	Chiron AS (Norway)	>98%		275.3	4.5 [b]	0.025 [b]
17α-Ethinylestradiol (EE2)	57-63-6	$C_{20}H_{24}O_2$	Sigma-Aldrich (Germany)	≥98%		296.4	3.67 [a]	11.3 (27 °C) [a]

[a] National Center for Biotechnology Information. PubChem Compound Database (September 2015).
[b] Predicted data, US Environmental Protection Agency's EPISuite™, KOWWIN v1.67 estimate.

Analysis, Institute for Environmental Research, RWTH Aachen University, Germany. The 11 participant laboratories (Table S1, S.I.) are associates of the NORMAN working group (WG) on bioassays and biomarkers. The battery composition was defined during a WG meeting in agreement with the different participants, considering the relevance of different bioassays for water quality assessment. After, the WG participants responded to a query regarding their interest in performing the different tests. Finally, three to four laboratories were selected to perform each bioassay, with inclusion of all interested.

2.3. Battery of bioassays

The bioassay battery (Table 2, Table S2) evaluated effects on organisms from different trophic levels: unicellular green algae growth inhibition (Algae), daphnid immobilization (*Daphnia*), and zebrafish embryo lethality and morphological effects (FET). Mechanism-specific assays evaluated estrogenicity (ER-Luc and YES) and mutagenicity (Ames). Experiments were performed in miniaturized format following static exposure without vessel pre-incubation with test solutions.

2.4. Water sample extract spiking

A 180 L water sample was collected at the pristine creek Wormsgraben (Harz Mountains, Germany), transported to the laboratory in stainless steel drums, extracted using large-volume solid phase extraction (Schulze et al., in preparation), and concentrated in 18 mL methanol. The method is described in the supplementary material. The water extract was evaluated in some bioassays (Table S3) by the coordinator.

Chemicals for spiking (Table 1) were selected due to relevance as emerging pollutants and bioactivity. Effect-data from previous studies and own preliminary tests (Table S3) provided the basis for spiking composition decision. Two or three spikes were designed per assay (Table 3) having either the most active toxicant(s) for each method or a final chemical mixture containing a fixed ratio of respective single chemical(s). Concentrations aimed to produce full dose-response curves considering as maximum test concentration 1 $\mu L_{extract}/mL_{medium}$, corresponding to an enrichment factor of 10 (10 $mL_{water\text{-}equivalent}/mL_{medium}$). Spikes for *Daphnia*, FET, ER-Luc and Ames were prepared by water extract evaporation to dryness, addition of DMSO as carrier, and spiking of chemicals using stock solutions in DMSO followed by separation in aliquots for each participant. For algae and YES, the water extract was spiked with the chemicals in methanol, divided in aliquots, and evaporated to dryness. Aliquots were coded and shipped at room temperature to the laboratories, who were not informed on sample composition during the testing period. DMSO was also provided for solvent control conditions. Samples were then stored at 4 °C.

2.5. Exposure setup and tested concentration ranges

Experiments were repeated mostly three times per bioassay, in each test with 3–4 replicate wells/vessels for each test condition following exposure setups described in Table 3.

2.6. Integrated data and statistical analysis

Bioassay results (expressed as described in Table 2) were evaluated following the same data preparation and statistical analysis methods. Results from experimental replicates were pooled and EC_{50} values were calculated for grouped experiments either by 2-parameter Weibull function using R language package (*Daphnia*), two parameter log-logistic curve from 0 to 100% with the two adjustable parameters being slope and EC_{50} by GraphPad Prism 6 (Algae, FET, Ames), or four-parameter log-logistic function with GraphPad (ER-Luc, YES). Differences between logEC_{50} values from different laboratories were compared by *t*-test or one-way ANOVA followed by Tukey's multiple comparisons test. EC_{50} values obtained in $\mu L_{extract}/mL_{bioassay}$ were converted to nominal concentrations of individual chemicals contained in each sample. For Algae, *Daphnia* and FET, ratios between EC_{50} ($\mu L_{extract}/mL_{bioassay}$) values of single-chemical and mixture spikes ($EC_{50\text{-}single}$:$EC_{50\text{-}mixture}$) were calculated. That allowed comparing single- and mixture-spike effects, since the mixture contained a fixed ratio of triclosan and acridine. For ER-Luc and YES, toxic-equivalent factors to respective standard chemical, 17β-estradiol (E2) or EE2, were obtained. Relative estrogenic potencies are expressed as E2 or EE2 equivalents (EEQ), calculated as a ratio between the EC_{50} of the reference compound and the EC_{50} of the spiked sample: $EEQ = EC_{50\text{-}E2 \text{ or } EE2}/EC_{50\text{-}sample}$. The only exception was the water extract, for which the EEQ was obtained with the PC_{10} approach (Besselink, 2015).

Table 2
Bioassays performed in the ILS, with indication of respective method title, endpoints, model organisms, exposure duration and protocol.

Bioassay	Method title	Endpoints/expressed results	Model organism	Exposure duration (h)	Exposure vessels	Medium per vessel or well (mL)	Protocols followed by laboratories (identified by codes) [a]
Algae test	Freshwater algal growth inhibition test	Growth inhibition/Growth inhibition normalized to solvent control	*Pseudokirchneriella subcapitata*	72	96-well plates	0.2	10, 9, 11: OECD Test No. 201 (OECD, 2011) or ISO 8692:2012 (ISO, 2012b) modified to 96-well plate
	Combined algae assay	Inhibition of microalgae growth and photosynthesis/Growth and photosynthesis inhibition normalized to solvent control	*P. subcapitata*	24	96-well plates	0.3	2, 3: Combined algae assay (Escher et al., 2008)
Daphnia test	*Daphnia* sp. acute immobilization test	Immobilization of daphnids/Immobilization occurrence	*D. magna*	48	96-well plates, glass tubes, glass beakers	0.2 10 20	5, 6, 7, 10 and 11: OECD Test No. 202 (OECD, 2004) or ISO 6341:2012 (ISO, 2012a)
FET test	Fish embryo acute toxicity test	Fish embryo lethality and occurrence of morphological sublethal endpoints/Occurrence of lethality and cumulative occurrence of lethal and sublethal morphological endpoints	*Danio rerio*	96	96-well plates	0.2	4, 9 and 10: OECD Test No. 236 (OECD, 2013b) with observation of sublethal morphological endpoints modified to 96-well plate
YES assay	Yeast estrogen screening assay	Estrogen receptor binding activity/Induction values converted to % of standard maximum response (after subtracting the solvent response from both sample and standard)	Recombinant yeast cells	18–72	96-well plates	0.2	1: β-galactosidase recombinant yeast following ISO/TC 147/SC 5 N 804 (ISO, 2013); 6: β-galactosidase recombinant yeast (Routledge and Sumpter, 1996)
				2.5	96-well plates	0.2	9: Luciferase recombinant yeast (Leskinen et al., 2003, Leskinen et al. 2005)
ER-Luc assay	Estrogen receptor luciferase reporter-gene assays with permanent cell lines	Estrogen receptor binding activity/Induction values converted to % of standard maximum response (after subtracting the solvent response from both sample and standard)	Luciferase reporter gene permanent human cell lines	19–24	96-well plates	0.2	5: T47D-kbLuc breast cancer cells (Wilson et al., 2004); 8: BG1Luc4E2 ovarian cancer cells (Rogers and Denison, 2000; OECD, 2012); 10: osteosarcoma cells (Maletz et al., 2013; Besselink, 2015)
Ames assay	Ames fluctuation assay	Induction of reverse mutations/Revertant numbers converted to % of positive control maximum response (after subtracting solvent revertants from both sample and positive control)	*Salmonella* strains TA100 and TA98	48 h	24-/384-well plates	0.5 (+2.5)/0.05	1, 8, 10: ISO 11350 (ISO, 2012c);3: (Reifferscheid et al., 2012; Escher et al., 2014)

[a] Laboratory code numbers are described in Table S1.

Table 3
Composition of the spiked water samples for each bioassay, consisting of one or two single-chemical spiking and a chemical mixture for each bioassay.

Bioassay	Sample	Composition of spiking of 10,000 times concentrated water extract				Exposure setup		
		Triclosan (mg/mL extract)	Acridine (mg/mL extract)	EE2 (μg/mL extract)	3-NBA (μg/mL extract)	Maximal test concentration (mL extract/ L medium)	Serial dilution steps	Number of tested dilutions
Algae test	Triclosan	0.1	–	–	–	1–3 [a]	1: 2	5–7 [a]
	Acridine	–	10	–	–	50–33 [b]	(2-fold)	16 [b]
	Mixture	0.1	10	100	–			
Daphnia test	Triclosan	1	–	–	–	1	1: 2	4–5
	Acridine	–	15	–	–		(2-fold)	
	Mixture	1	15	100	2			
FET test	Triclosan	3	–	–	–	0.77	1: 1.3	5
	Acridine	–	2	–	–	1	(1.3-fold)	
	Mixture	3	2	100	2	0.58		
YES assay	EE2	–	–	100	–	0.1–2	3: 10 and 1: 3	9–16
	Mixture	1	2	100	–		(3.3 and 3-fold)	
ER-luc assay	EE2	–	–	1	–	0.5–1	1: 10	7
	Mixture	1	2	1	–		(10-fold)	
Ames assay	3-NBA	–	–	–	2	1	1: 2	6
	Mixture	0.1	2	100	2		(2-fold)	

[a] Freshwater algal growth inhibition test with unicellular green algae.
[b] Combined algae assay.

3. Results and discussion

Differences between assay results are indicated either as not significant (n.s.) or according to p values. Effect-concentration values for different tests and laboratories are detailed in S.I.

3.1. Toxic effects on aquatic organisms

Aquatic organisms differed in terms of sensitivity to triclosan (algae > daphnids > fish) and acridine (fish > daphnids > algae) spikes. Present EC_{50} nominal ($EC_{50\text{-nom}}$) for single-chemical spikes (Fig. 1) were in same range as literature data for tests performed in microtiter plates (Table S4) but tended to be higher than literature values based on measured concentrations or for experiments in higher medium volume.

3.1.1. Algae test

The OECD/ISO Algae test was the most sensitive aquatic organism assay to triclosan, in agreement with freshwater algal growth being more sensitive than endpoints in bacteria, protozoa, macrophytes, daphnids, amphibians and fish (Orvos et al., 2002; Tatarazako et al., 2004; Harada et al., 2008; Tamura et al., 2013). Detected 72 h growth-inhibition $EC_{50\text{-nom}}$ (14.7 and 25.7 μg/L, n.s.) are in the same range as previous 72 and 96 h $EC_{50\text{-nom}}$ for *P. subcapitata* determined also in 96-well plates (Harada et al., 2008; Rosal et al., 2010). However, our values are 3–50 times higher than results obtained by incubation in 20–100 mL of medium (i.e. 100–500 times the present volume) (Orvos et al., 2002; Tatarazako et al., 2004; Yang et al., 2008; Tamura et al., 2013). Since triclosan is relatively hydrophobic, adsorption to the plate material could have occurred (Rojíčková et al., 1998). Triclosan is also prone to phototransformation (Tixier et al., 2002), which could be another source of variability. The OECD TG (2011) already discusses the interference of these aspects with single-chemicals, which can provide a basis for investigating the stability of water extracts components during exposure. Finally, the water extract matrix could have decreased triclosan bioavailability due to its high sorption capacity to organic matter (Reiss et al., 2002).

For acridine, even if our $EC_{50\text{-nom}}$ differed (5.9 and 4.1 mg/L, $p < 0.01$), values were in good agreement with previous 72 h $EC_{50\text{-nom}}$ for *Desmodesmus subspicatus* following exposure in 24-well plates (Eisentraeger et al., 2008). However, values were circa one order of magnitude higher than 96 h $EC_{50\text{-measured}}$ for *Selenastrum capricornutum* (current *P. subcapitata*) exposed in 100–250 mL

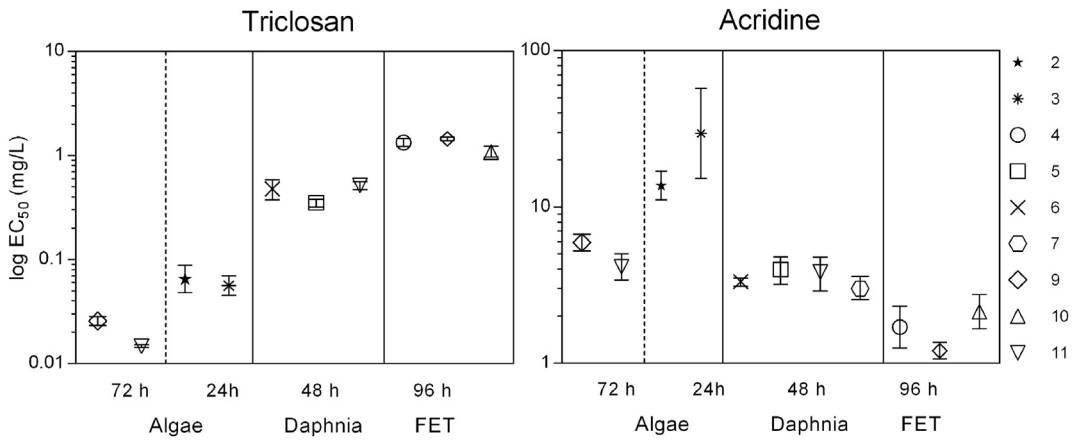


Fig. 1. Effect-concentration values (log EC_{50} and 95% C.I., mg/L) obtained for pooled data from one to three experiments for each assay for the triclosan (left) and acridine (right) spikes in the algae (72 h or 24 h growth inhibition), *Daphnia* (48 h immobilization) and FET (96 h cumulative effects) tests. Results are presented according to laboratory code numbers (Table S1).

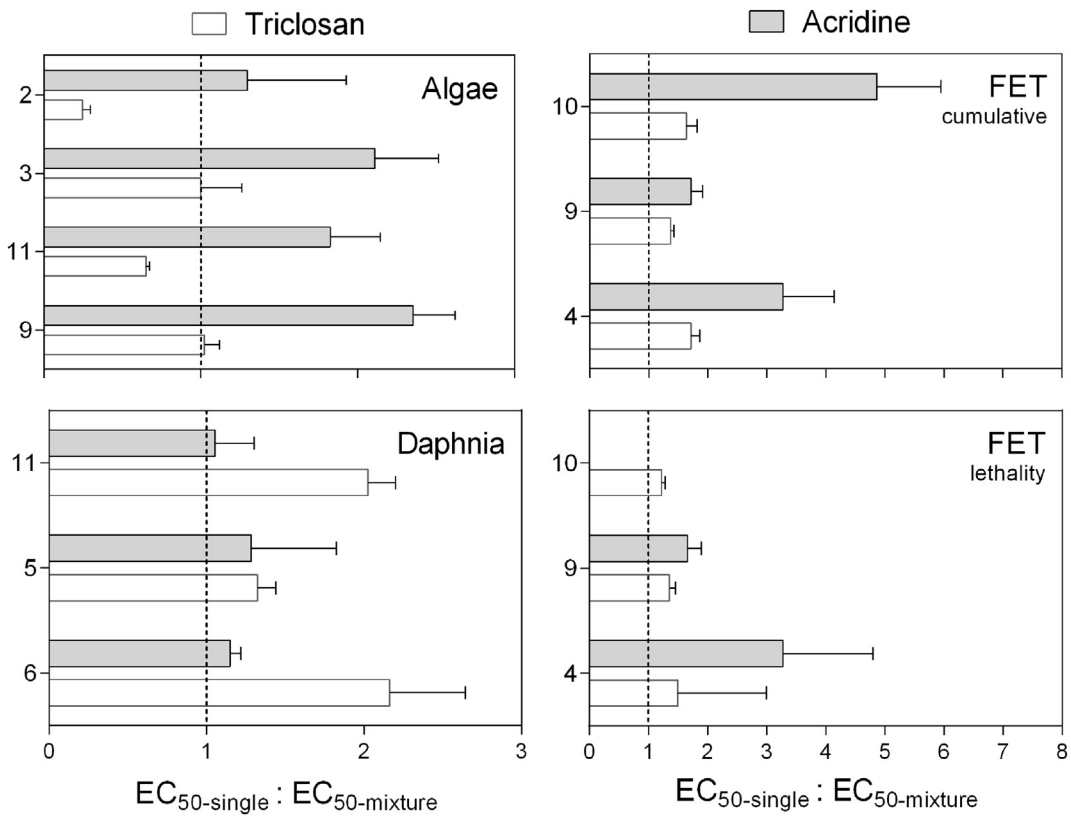


Fig. 2. Ratios between EC_{50} values (μL/mL) for the single-chemical and mixture spikes containing a fixed ratio of respective single compounds ($EC_{50\text{-single}}$: $EC_{50\text{-mixture}}$) for the triclosan (white bars) and acridine (grey bars) spikes in the algae, *Daphnia* and FET (cumulative effects and lethality) tests. Error bars correspond to the ratios between 95% C.I. for single chemicals and the $EC_{50\text{-mixture}}$ value. Y-axis correspond to laboratory code numbers (Table S1).

medium (Blaylock et al., 1985; Dijkman et al., 1997). Sensitivity differences are not known for acridine due to non-specific toxicity mechanism (Dijkman et al., 1997). Decrease in exposure concentration instead may be relevant, since 40–60% losses occurred already prior to exposure start, followed by additional circa 10% decrease during 72 h incubation in 24-well plates (Eisentraeger et al., 2008). Therefore for acridine chemical losses during sample shipping, handling and experiments could have interfered with effective test concentrations.

In the combined algae assay, 24 h growth inhibition $EC_{50\text{-nom}}$ values for triclosan (65.0 and 56.2 μg/L, n.s.) and acridine (13.7 and 29.6, $p < 0.001$) spikes were 2–3 and 2–7 times higher than for the OECD tests, respectively. That indicates time-dependency of effects for both chemicals on algae growth. No tendency for specific photosynthesis inhibition was observed since the photosynthesis endpoint was equally or less sensitive than growth inhibition (results not shown) (Escher et al., 2008; Tang and Escher, 2014). Still, this is a very relevant endpoint since many current WFD priority and emerging compounds present this mode of action.

$EC_{50\text{-single}}$:$EC_{50\text{-mixture}}$ ratios for triclosan (Fig. 2) reached values near or less than 1 and were lower than those for acridine, suggesting its effects were prevalent in the mixture. EE2 is not considered to have caused substantial growth inhibition, since the higher exposure concentration (0.1 mg/L) was seven to ten-fold lower than previous NOEC (0.71 mg/L) or LOEC (1.2 mg/L) (Maes et al., 2014).

3.1.2. *Daphnia* test

The OECD/ISO *Daphnia* immobilization test presented intermediate sensitivity to both triclosan and acridine spikes. Present triclosan 48 h immobilization $EC_{50\text{-nom}}$ (351–516 μg/L, n.s.) are in similar range as previous studies (Orvos et al., 2002; Harada et al., 2008; Peng et al., 2013). The compound was also found to cause effects in *D. magna* reproduction test lasting 21 days, with LOEC values for reduced number of neonates being circa half of respective 48 h immobilization EC_{50} (Orvos et al., 2002; Peng et al., 2013).

Also for acridine the obtained $EC_{50\text{-nom}}$ (3.0–5.1 mg/L, n.s.) agree with previous results (Blaylock et al., 1985; Feldmannová et al., 2006; Eisentraeger et al., 2008). Acridine caused also reduction in offspring number produced per brood in semi-static exposure during 14 d, with the LOEC being less than half of respective acute EC_{50} (Blaylock et al., 1985).

Considering $EC_{50\text{-single}}$:$EC_{50\text{-mixture}}$ ratios (Fig. 2), acridine values were near 1 and lower than for triclosan, indicating that its effects were prevalent in the mixture. EE2 effects are considered to be negligible, since its highest exposure concentration (0.1 mg/L) was 50 times lower than previous NOEC (Goto and Hiromi, 2003). Although no information for 3-NBA was found in the literature, acute effects are not considered relevant due to low concentrations.

3.1.3. FET test

The OECD FET test presented the lowest sensitivity to triclosan and the highest sensitivity to acridine among aquatic organism tests.

Triclosan 96 h $LC_{50\text{-nom}}$ (1.3–1.9 mg/L, n.s.) and $EC_{50\text{-nom}}$ (Table S5) are circa three times higher than previous 96 h $LC_{50\text{-nom}}$ for zebrafish embryos exposed in 24-well plates (Oliveira et al., 2009) or medaka in petri dishes under semi-static conditions (Ishibashi et al., 2004). This discrepancy could be related to differences in medium volumes and ratios surface area to volume of

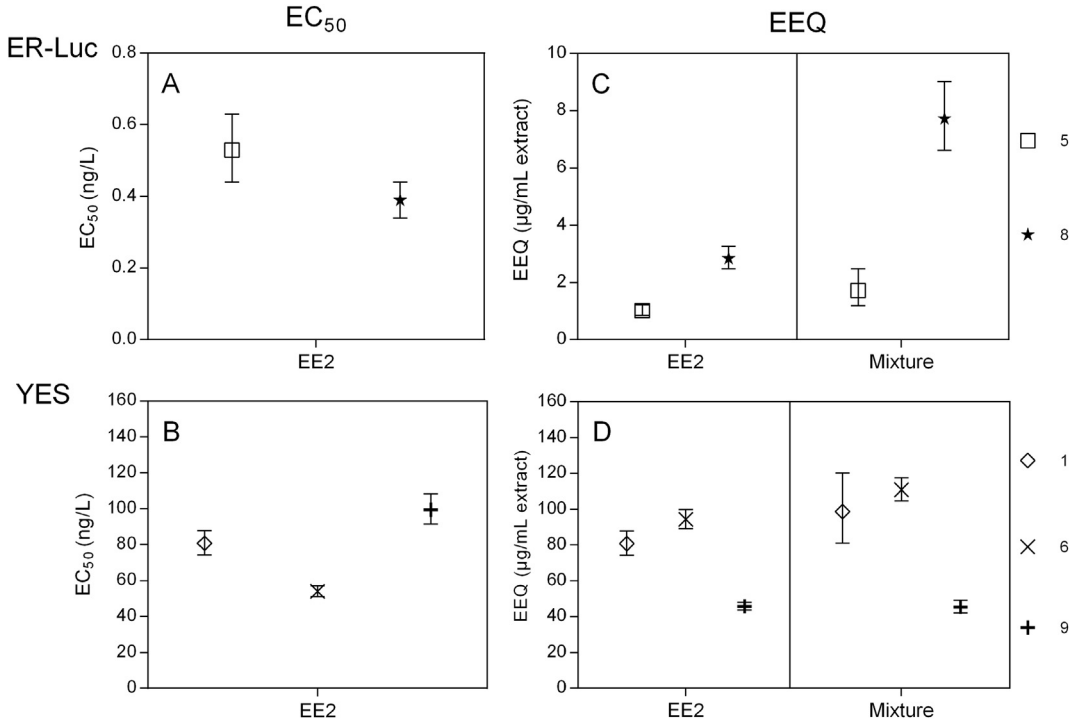


Fig. 3. EC_{50} (ng/L) values for EE2 in the ER-Luc (A) and YES (B) assays, and EEQ values obtained for the EE2 and the mixture spikes in the ER-Luc (C) and the YES (D) assays. EC_{50} values (symbols) and 95% C.I. (error bars) for respective sample. Results are presented according to laboratory code numbers (Table S1). Biological models are: T47D-kbLuc (5) BG1Luc4E2 (8), β-galactosidase recombinant yeast by McDonnell et al. (1991) (1), β-galactosidase recombinant yeast by Routledge and Sumpter 1996 (6), and luciferase recombinant yeast by Leskinen et al., 2003 (9).

exposure vessels. However, triclosan concentrations decreased to circa half even in 1 L of water after 24 h adult medaka exposure (Ishibashi et al., 2004). Therefore other factors could play a role such as phototransformation, which can be minimized by incubation in dark. Among sublethal effects, reduced growth and delayed development were prevalent, similarly to effects in *Xenopus laevis* embryos (Harada et al., 2008). Triclosan was also related to delayed swim-up behaviour initiation and reduced survival in rainbow trout early-life stages (Orvos et al., 2002) and to disrupted swimming and predator avoidance in fathead minnow larvae (Cherednichenko et al., 2012; Fritsch et al., 2013). We observed increased heartbeat rates at 96 h in zebrafish exposed to 1.0 (47.0 beats/20 s, $p < 0.01$) and 1.3 mg/L (48.7 beats/20 s, $p < 0001$) compared to water and solvent controls, concentrations which caused none and circa 10% ($p < 001$) cumulative effects, respectively. Since triclosan can impair the excitation-contraction coupling of cardiac and skeletal muscle (Cherednichenko et al., 2012; Fritsch et al., 2013), increased compensatory heartbeat rate could have occurred. Therefore the assessment of sublethal endpoints can support the identification of toxic effects other than lethality (Di Paolo et al., 2015a; Jonas et al., 2015).

For acridine, FET 96 h $LC_{50\text{-nom}}$ (0.71–1.28 mg/L, n.s.) were circa three times lower than those from *Daphnia* and algae tests. Present values are slightly higher than previous measured 48 h LC_{50} performed in 24-well plates (Peddinghaus et al., 2012). That can be related to possible acridine losses before and during experiments, since concentrations were shown to decrease to less than half of nominal values (Peddinghaus et al., 2012). Performance of semi-static exposure with solution renewal could be a possible solution to maintain exposure concentrations (OECD, 2013b).

Considering the $EC_{50\text{-single}}$:$EC_{50\text{-mixture}}$ (Fig. 2), triclosan tended to present lower values when compared to acridine, indicating it was prevalent in the mixture toxicity. EE2 effects are considered to be negligible, since its highest exposure concentration (0.1 mg/L) was 50 times lower than previous NOEC (5 mg/L) (Goto and Hiromi, 2003). For 3-NBA, although no information was found in the literature, acute effects are considered to be negligible.

3.2. Estrogenicity assessment

Although differences occurred between different estrogenicity assays and models, relative induction $EC_{50\text{-nom}}$ values were comparable to the literature, and obtained EEQ for the EE2 spike are in good agreement with previous values for ER-Luc and YES (Fig. 3).

3.2.1. ER-Luc assay

Among all assays performed by the coordinator (Table S3), the non-spiked water extract was active only in the ER-Luc (ER-CALUX), with an EEQ of 0.17 ± 0.01 ng/L_{water} for the enrichment factor of 1. EE2 spike induction EC_{50} (0.53 and 0.39 ng/L_{medium}, n.s.) were within the range of previously reported values for EE2 (Legler et al., 2002; Murk et al., 2002; Wilson et al., 2004; Bermudez et al., 2012; OECD, 2012). Although EEQ values showed some variation (Fig. 3C), which could be related to differences in assay protocol or model sensitivity (Jarošová et al., 2014), EEQ determination showed to be a reliable measurement for sample content.

Considering the mixture spike, concentrations ≥0.5 $\mu L_{extract}/mL_{medium}$ caused cytotoxicity and were excluded from regression analysis. This effect is considered to be caused by triclosan concentrations (≥0.5 mg/L_{medium}) in the cytotoxic range for human cells (Henry and Fair, 2013); while no acridine cytotoxicity is indicated (Brinkmann et al., 2014). Tendency for higher EEQ values was observed for the mixture spike (Fig. 3C). It could be discussed that such response is related to estrogen receptor binding by other chemicals in mixture, since acridine induction in T47Dluc assay produced an estradiol equivalency factor (EEF) of $2.5.10^{-7}$

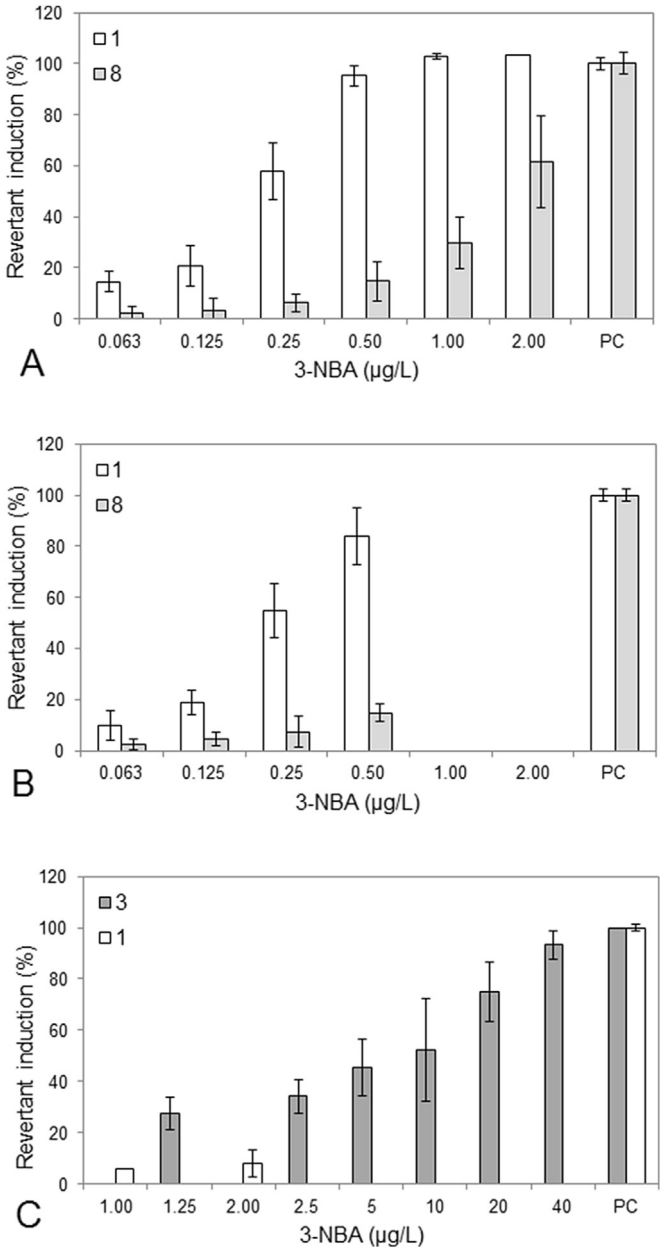


Fig. 4. Revertant induction versus 3-NBA concentrations (µg/L) contained in (A) 3-NBA spike in TA98-S9, (B) mixture spike in TA98 -S9, and (C) 3-NBA spike in TA100 -S9; plus respective positive control (PC) conditions. Average values (bars) and standard deviations (error bars) for two to three experiments. Results are presented using laboratory code numbers (Table S1).

(Brinkmann et al., 2014). However there is no evidence of triclosan agonism in estrogen-receptor reporter gene cell-based assays (own results) (Ahn et al., 2008). More likely, non-specific effects on cellular membranes or metabolism (Ajao et al., 2015) could have interfered with induction.

3.2.2. YES assay

Our induction EC_{50} for the EE2 spike varied up to 2.5-fold (54.1–132.7 ng/L, $p < 0.01$ to 0.0001), in similar range to literature data (Table S4). The lowest EE2 spike EC_{50} was produced by the Routledge/Sumpter strain (1996), in agreement with previous studies (Van den Belt et al., 2004; Balsiger et al., 2010), while the bioluminescent strain (Leskinen et al., 2005) produced the highest value. For the McDowell/ISO assay (ISO, 2013), the EC_{50} of 99.5 ng/L was slightly higher than the EC_{50} obtained for the standard curve (80.4 ng/L), which also uses EE2 in this assay. EEQ values varied circa 2-fold (45.8–94.3 $\mu g/mL_{extract}$), which can be related to the fact that different yeast strains and protocols can produce different EEF values (Svobodová et al., 2009; Jarošová et al., 2014). Therefore for the application of estrogenicity assays in water quality, effect-concentrations for the standard chemical, main estrogens and investigated samples should be determined using the same model and protocol (Jarošová et al., 2014; Kunz et al., 2015).

The highest mixture spike test concentrations ($\geq$0.1 $\mu L_{extract}/mL_{medium}$) caused cytoxicity to the yeast cells and were excluded from regression analysis. This is attributed mostly to triclosan ($\geq$0.1 mg/L_{medium}), since acridine concentrations are not expected to be toxic to the yeast cells (Brinkmann et al., 2014). No differences occurred between respective EEQ values for single and mixture spikes (Fig. 3D). Previously, acridine was not identified as estrogenic by the lyticase YES assay (Brinkmann et al., 2014). Although triclosan was active in the Routledge/Sumpter strain, the compound was not identified as estrogenic in the bioluminescent YES (Svobodová et al., 2009).

3.3. Mutagenicity assessment by the Ames fluctuation assay

Strong revertant induction occurred following 3-NBA spike incubation with the TA98 strain in the absence of S9 fraction (-S9) (Fig. 4A), which was of lower magnitude after metabolic transformation and for TA100 -S9 (Fig. 4B–C). 3-NBA spike revertant induction EC_{50} values were 0.21 and 1.56 µg/L ($p < 0.01$) for TA98 -S9; and 5.73 µg/L for TA100 -S9. Therefore the compound was clearly identified as mutagenic, although further improvement might be needed if precise effect-concentration values are required. Such results are in agreement with previous studies describing 3-NBA as a strong direct-acting mutagen in the TA98 strain, and the fact that it is less active in TA100 suggests that it causes frameshift-type mutations (Enya et al., 1997; IARC, 2014). Further, there are indications that 3-NBA is also genotoxic *in vitro* and in *in vivo* (Watanabe et al., 2005b). 3-NBA is a major mutagen in diesel particles, sediments, and surface soils (Enya et al., 1997; Watanabe et al., 2005a; Lübcke-von Varel et al., 2012) and concentrations up to 2.6 ng/L were identified in rainwater (Murahashi et al., 2003).

For the mixture spike, test concentrations $\geq$0.5 $\mu L_{extract}/mL_{medium}$ caused toxic effects in –S9 exposures (attributed to triclosan 50 ng/mL medium), which were excluded from regression analysis (Fig. 4B, Fig. S7). Cytotoxic effects were reduced by the S9 fraction incubation (Fig. S7), suggesting that resulting triclosan metabolites present less toxic effects than the parent compound. Our results showed that neither triclosan nor acridine caused increase in the number of revertants (Table S4), in agreement with previous studies investigating their mutagenicity through the Ames plate incorporation method (Eisentraeger et al., 2008; SCCP, 2009).

3.4. Bioassay battery strategy

Bioassay battery assessment of water quality is based on the consideration that one single bioassay does not provide an overview on potential effects on different organisms and toxicity mechanisms. Since sensitivity to different toxicants varies between organisms, multi-taxa assessment supports the comprehension of toxicant effects on aquatic communities (Guillen et al., 2012). The organism-level assays proposed in the present study investigate population-level effects in freshwater algae as primary producers, acute toxicity to the filter-feeder invertebrate *Daphnia*, and acute toxicity to fish individuals. Multi-taxa toxicity assessment is applied for EQS derivation within the WFD, which requires

evaluation of acute and chronic data for (i) alga/macrophyte, (ii) *Daphnia*/another invertebrate, and (iii) fish (EC, 2011). Similar strategy is applied in REACH to evaluate aquatic pelagic toxicity (ECHA 2014). The suitability of the algae, *Daphnia* and FET assays to compose a basic (eco)toxicity test battery was evaluated for hazard waste, wastewater effluent, freshwater and drinking water assessment (Keddy et al., 1995; Diaz-Baez et al., 2002; Manusadžianas et al., 2003; Pandard et al., 2006; Gartiser et al., 2009; Römbke and Moser, 2009); and for effect-directed analysis (Brack et al., 2013, 2016; Di Paolo et al., 2015b). Therefore the assays are expected to be already established in diverse laboratories worldwide. Finally, the followed miniaturized assay performance has already been investigated in comparison with higher-volume methods and with adult fish for the FET (Eisentraeger et al., 2003; Knobel et al., 2012; Baumann et al., 2014).

Complementary, mechanism-specific bioassays can provide information on modes-of-action that are intrinsically of concern for ecosystems and health. For example, the photosynthesis inhibition endpoint of the performed combined algae test covers many current WFD priority compounds and emerging compounds. Furthermore, endocrine disruption and mutagenicity are of particular relevance for population-level effects and humans (EC, 2000, 2011; ECHA 2014). For estrogens, regulatory strategies involving bioassays are reinforced after the recent inclusion of estrogenic pharmaceuticals in the WFD watch list (Hecker and Hollert, 2011; EC, 2013). In fact, both ER-Luc and YES assays have been recommended for estrogen monitoring in water bodies (Loos, 2012). Regarding mutagenicity, the Ames fluctuation assay round-robin study was the first step towards its regulatory implementation in water legislation (Wolz et al., 2010; Reifferscheid et al., 2012). Moreover, the Ames and umu tests are recommended as mutagenicity and genotoxicity methods for the waste ecotoxicological characterization (Römbke and Moser, 2009). Due to their environmental and health relevance, estrogenicity and mutagenicity assays are also established in many laboratories.

The present results complement previous validation studies of the organism-level and mechanism-specific tests by demonstrating the good performance of methods not only with single chemicals but also to evaluate water extracts spiked with emerging contaminants. Our approach can provide useful information to link chemical testing and field studies with those assays. A relevant aspect to consider is that the assays can be applied to evaluate not only water extracts but raw water samples and effluents. In this sense the proposed bioassay battery presents a flexible setup for diverse applications in the context of water quality monitoring.

3.5. Stepping-stones towards the establishment of bioassays in water quality monitoring

Currently there are diverse European initiatives towards bioassay application in water quality assessment, such as the Technical Report on effect-based tools in the context of the WFD (Wernersson et al., 2015) and activities for the validation of low volume, high-throughput bioassay batteries (Brack et al., 2013, 2015; Altenburger et al., 2015; Neale et al., 2015; Schulze et al., 2015). Such applied studies will be of high relevance for the decision on a basic battery for water monitoring. Similarly to our approach, these initiatives tend to focus on assays that allow relatively fast performance. Consequently, only acute toxicity is evaluated in fish and daphnids, while mechanism-specific methods are investigated in the *in vitro* level. However, after the setup of such basic battery, its composition can certainly be expanded according to regional requirements or specific investigation. For instance, when chronic fish toxicity is suspected, the decision on whether to perform chronic tests can be supported by toxicity assays with fish early-life stages (OECD, 2013a; Villeneuve et al., 2014; Di Paolo et al., 2015a). In cases when freshwater sediments present a concern, whole-sediment toxicity assays with different organisms are available. Ring tests have demonstrated the good performance of tests evaluating macrophyte growth impairment (Feiler et al., 2014); and growth and reproduction effects on interstitial water nematodes (Hoss et al., 2012). Recent studies include also a methodological investigation of a freshwater ostracod sub-chronic test (Casado-Martinez et al., 2016); and a tiered strategy for sediment risk assessment integrating different toxicity tests (Diepens et al., 2016).

Importantly, the investigation of additional mechanism-specific toxicities can rely on diverse reporter-gene assays, for which effect-based trigger values to support decisions on water quality assessment are being established (Loos, 2012; Brand et al., 2013; Escher et al., 2015). In parallel to these tests, it is necessary to investigate the occurrence of non-specific toxicity caused by sample components, which can interfere with the performance of assays and even mask mechanism-specific effects (Brack et al., 2016). That was demonstrated in our study for the ubiquitous contaminant triclosan, which was cytotoxic to human cells, yeast and bacteria at concentrations representative of water samples or extracts (von der Ohe et al., 2012). Finally, further studies can investigate remaining aspects of relevance for bioassays screening of water sample and extracts. For instance, different conditions of sample storage can partially affect chemical composition, including of endocrine disruptors (Aboulfadl et al., 2010). In the future, the influence of sample shipping and storage conditions should be evaluated not only through chemical analysis but also regarding effects on bioassay performance and results.

4. Conclusions and outcomes

The battery of miniaturized bioassays presented complementary sensitivity and specificity to the water extract spikes containing four emerging pollutants as single-chemicals or mixtures. Aquatic organism sensitivity varied following exposure to different chemicals, confirming the complementary role of the tests with the three taxa for water quality assessment. Estrogenicity and mutagenicity assays identified with high precision the respective mechanism-specific effects of spikes, even though non-specific toxicity of mixture compounds affected the evaluation of higher test concentrations. Since differences in experimental protocols, model organisms, and data analysis can affect the determination of effect-concentrations, respective standard methods and harmonized procedures should be followed when implementing bioassays in water monitoring. Together with other ongoing activities for the validation of a basic battery of bioassays, the present study is an important step towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring.

Acknowledgements

We would like to thank the colleagues at RWTH Aachen Simone Hotz, Monika Lam, Christine Schönlau and Josef Koch for the contribution to the preliminary tests of the water extract and chemicals; and Yvonne Müller and Sarah Johann for their help in organizing the project meeting in Aachen. Thanks to Dr. Brigitte von Danwitz from the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (LANUV) in Germany for providing useful literature to support the discussion of results. Thanks to Dr. Robert Kase from the Swiss Centre for Applied Ecotoxicology Eawag-EPFL in Switzerland for his suggestions on the manuscript draft. This project was developed by the Working Group 2 Bioassays and biomarkers in water quality monitoring of

the NORMAN Association (http://www.norman-network.net). It was funded and supported by the European Marie Curie Initial Training Network EDA-EMERGE (grant agreement no. 290100; www.ufz.de/eda-emerge), the European FP7 Collaborative Project SOLUTIONS (grant agreement no. 603437; www.solutions-project.eu), and the NORMAN Association (www.norman-network.net). CDP was supported by the EDA-EMERGE Marie Curie ITN (ESR1). SA and EM (INERIS) were supported by the French Ministry of Ecology (grant P181) and by the ONEMA.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.watres.2016.08.018.

References

Aboulfadl, K., De Potter, C., Prévost, M., Sauvé, S., 2010. Time-dependent integrity during storage of natural surface water samples for the trace analysis of pharmaceutical products, feminizing hormones and pesticides. Chem. Central J. 4 (1), 1–8.

Ahn, K.C., Zhao, B., Chen, J., Cherednichenko, G., Sanmarti, E., Denison, M.S., Lasley, B., Pessah, I.N., Kültz, D., Chang, D.P.Y., Gee, S.J., Hammock, B.D., 2008. Vitro biologic activities of the antimicrobials triclocarban, its analogs, and triclosan in bioassay screens: receptor-based bioassay screens. Environ. Health Perspect. 116 (9), 1203–1210.

Ajao, C., Andersson, M.A., Teplova, V.V., Nagy, S., Gahmberg, C.G., Andersson, L.C., Hautaniemi, M., Kakasi, B., Roivainen, M., Salkinoja-Salonen, M., 2015. Mitochondrial toxicity of triclosan on mammalian cells. Toxicol. Rep. 2, 624–637.

Altenburger, R., Ait-Aissa, S., Antczak, P., Backhaus, T., Barceló, D., Seiler, T.-B., Brion, F., Busch, W., Chipman, K., de Alda, M.L., de Aragão Umbuzeiro, G., Escher, B.I., Falciani, F., Faust, M., Focks, A., Hilscherova, K., Hollender, J., Hollert, H., Jäger, F., Jahnke, A., Kortenkamp, A., Krauss, M., Lemkine, G.F., Munthe, J., Neumann, S., Schymanski, E.L., Scrimshaw, M., Segner, H., Slobodnik, J., Smedes, F., Kughathas, S., Teodorovic, I., Tindall, A.J., Tollefsen, K.E., Walz, K.-H., Williams, T.D., Van den Brink, P.J., 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 (0), 540–551.

Balsiger, H.A., de la Torre, R., Lee, W.-Y., Cox, M.B., 2010. A four-hour yeast bioassay for the direct measure of estrogenic activity in wastewater without sample extraction. Conc. Steriliz. Sci. Total Environ. 408 (6), 1422–1429.

Baumann, J., Sakka, Y., Bertrand, C., Koser, J., Filser, J., 2014. Adaptation of the Daphnia sp. acute toxicity test: miniaturization and prolongation for the testing of nanomaterials. Environ. Sci. Pollut. Res. Int. 21 (3), 2201–2213.

Bermudez, D.S., Gray, L.E., Wilson, V.S., 2012. Modelling defined mixtures of environmental oestrogens found in domestic animal and sewage treatment effluents using an in vitro oestrogen-mediated transcriptional activation assay (T47D-KBluc). Int. J. Androl. 35 (3), 397–406.

Besselink, H.T., 2015. Validation of the (Anti-)ERα CALUX Bioassay U2-OS Cells Transcriptional ERalpha CALUX Assay for the Detection of Estrogenic and Anti-estrogenic Chemicals for Inclusion in the OECD TG 455.

Blaylock, B.G., Frank, M.L., McCarthy, J.F., 1985. Comparative toxicity of copper and acridine to fish, Daphnia and algae. Environ. Toxicol. Chem. 4 (1), 63–71.

Brack, W., Ait-Aissa, S., Burgess, R.M., Busch, W., Creusot, N., Di Paolo, C., Escher, B.I., Mark Hewitt, L., Hilscherova, K., Hollender, J., Hollert, H., Jonker, W., Kool, J., Lamoree, M., Muschket, M., Neumann, S., Rostkowski, P., Ruttkies, C., Schollee, J., Schymanski, E.L., Schulze, T., Seiler, T.B., Tindall, A.J., De Aragao Umbuzeiro, G., Vrana, B., Krauss, M., 2016. Effect-directed analysis supporting monitoring of aquatic environments - an in-depth overview. Sci. Total Environ. 544, 1073–1118.

Brack, W., Altenburger, R., Schuurmann, G., Krauss, M., Lopez Herraez, D., van Gils, J., Slobodnik, J., Munthe, J., Gawlik, B.M., van Wezel, A., Schriks, M., Hollender, J., Tollefsen, K.E., Mekenyan, O., Dimitrov, S., Bunke, D., Cousins, I., Posthuma, L., van den Brink, P.J., Lopez de Alda, M., Barcelo, D., Faust, M., Kortenkamp, A., Scrimshaw, M., Ignatova, S., Engelen, G., Massmann, G., Lemkine, G., Teodorovic, I., Walz, K.H., Dulio, V., Jonker, M.T., 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.

Brack, W., Govender, S., Schulze, T., Krauss, M., Hu, M., Muz, M., Hollender, J., Schirmer, K., Schollee, J., Hidasi, A., Slobodnik, J., Rabova, Z., Ait-Aissa, S., Sonavane, M., Carere, M., Lamoree, M., Leonards, P., Tufi, S., Ouyang, X., Schriks, M., Thomas, K., de Almeida, A., Froment, J., Hammers-Wirtz, M., Ahel, M., Koprivica, S., Hollert, H., Seiler, T.-B., Paolo, C., Tindall, A., Spirhanzlova, P., 2013. EDA-EMERGE: an FP7 initial training network to equip the next generation of young scientists with the skills to address the complexity of environmental contamination with emerging pollutants. Environ. Sci. Eur. 25 (1), 18.

Brand, W., de Jongh, C.M., van der Linden, S.C., Mennes, W., Puijker, L.M., van Leeuwen, C.J., van Wezel, A.P., Schriks, M., Heringa, M.B., 2013. Trigger values for investigation of hormonal activity in drinking water and its sources using CALUX bioassays. Environ. Int. 55, 109–118.

Brinkmann, M., Maletz, S., Krauss, M., Bluhm, K., Schiwy, S., Kuckelkorn, J., Tiehm, A., Brack, W., Hollert, H., 2014. Heterocyclic aromatic hydrocarbons show estrogenic activity upon metabolization in a recombinant transactivation assay. Environ. Sci. Technol. 48 (10), 5892–5901.

Carvalho, R.N., Arukwe, A., Ait-Aissa, S., Bado-Nilles, A., Balzamo, S., Baun, A., Belkin, S., Blaha, L., Brion, F., Conti, D., Creusot, N., Essig, Y., Ferrero, V.E.V., Flander-Putrle, V., Fürhacker, M., Grillari-Voglauer, R., Hogstrand, C., Jonáš, A., Kharlyngdoh, Joubert B., Loos, R., Lundebye, A.-K., Modig, C., Olsson, P.-E., Pillai, S., Polak, N., Potalivo, M., Sanchez, W., Schifferli, A., Schirmer, K., Sforzini, S., Stürzenbaum, S.R., Søfteland, L., Turk, V., Viarengo, A., Werner, I., Yagur-Kroll, S., Zounková, R., Lettieri, T., 2014. Mixtures of chemical pollutants at european legislation safety concentrations: how safe are they? Toxicol. Sci. 141 (1), 218–233.

Casado-Martinez, M.C., Burga-Pérez, K.F., Bebon, R., Férard, J.-F., Vermeirssen, E.L.M., Werner, I., 2016. The sediment-contact test using the ostracod Heterocypris incongruens: effect of fine sediments and determination of toxicity thresholds. Chemosphere 151, 220–224.

Cherednichenko, G., Zhang, R., Bannister, R.A., Timofeyev, V., Li, N., Fritsch, E.B., Feng, W., Barrientos, G.C., Schebb, N.H., Hammock, B.D., Beam, K.G., Chiamvimonvat, N., Pessah, I.N., 2012. Triclosan impairs excitation–contraction coupling and Ca2+ dynamics in striated muscle. Proc. Natl. Acad. Sci. 109 (35), 14158–14163.

de Voogt, P., Laane, R.W.P.M., 2009. Assessment of azaarenes and azaarones (oxidized azaarene derivatives) in the Dutch coastal zone of the North Sea. Chemosphere 76 (8), 1067–1074.

Di Paolo, C., Groh, K.J., Zennegg, M., Vermeirssen, E.L.M., Murk, A.J., Eggen, R.I.L., Hollert, H., Werner, I., Schirmer, K., 2015a. Early life exposure to PCB126 results in delayed mortality and growth impairment in the zebrafish larvae. Aquat. Toxicol. 169, 168–178.

Di Paolo, C., Seiler, T.-B., Keiter, S., Hu, M., Muz, M., Brack, W., Hollert, H., 2015b. The value of zebrafish as an integrative model in effect-directed analysis - a review. Environ. Sci. Eur. 27 (1), 1–11.

Diaz-Baez, M.C., Sanchez, W.A., Dutka, B.J., Ronco, A., Castillo, G., Pica-Granados, Y., Castillo, L.E., Ridal, J., Arkhipchuk, V., Srivastava, R.C., 2002. Overview of results from the WaterTox intercalibration and environmental testing phase II program: part 2, ecotoxicological evaluation of drinking water supplies. Environ. Toxicol. 17 (3), 241–249.

Diepens, N.J., Koelmans, A.A., Baveco, H., Brink, P.J., Heuvel-Greve, M.J. and Brock, T.C.M. (2016), pp. 1–77, Springer New York, New York, NY.

Dijkman, N.A., van Vlaardingen, P.L.A., Admiraal, W.A., 1997. Biological variation in sensitivity to N-heterocyclic PAHs; effects of acridine on seven species of microalgae. Environ. Pollut. 95 (1), 121–126.

EC, 2000. Directive 2000/60/EC of the European Parliament and of the council of 23 October 2000 establishing a framework for Community action in the field of water policy. OJ L 327, 1–73, 22.12.2000.

EC, 2011. European Commission WFD-CIS Guidance Document No. 27 Technical Guidance on Deriving Environmental Quality Standards. Technical report 2011–055.

EC, 2013. Official J. Eur. Union L 266, 1–17.

ECHA, E.C.A, 2014. Guidance on Information Requirements and Chemical Safety Assessment Chapter R.7b: Endpoint Specific Guidance.

Eisentraeger, A., Brinkmann, C., Hollert, H., Sagner, A., Tiehm, A., Neuwoehner, J., 2008. Heterocyclic compounds: toxic effects using algae, daphnids, and the Salmonella/microsome test taking methodical quantitative aspects into account. Environ. Toxicol. Chem. 27 (7), 1590–1596.

Eisentraeger, A., Dott, W., Klein, J., Hahn, S., 2003. Comparative studies on algal toxicity testing using fluorometric microplate and Erlenmeyer flask growth-inhibition assays. Ecotoxicol. Environ. Saf. 54 (3), 346–354.

Enya, T., Suzuki, H., Watanabe, T., Hirayama, T., Hisamatsu, Y., 1997. 3-Nitrobenzanthrone, a powerful bacterial mutagen and suspected human carcinogen found in diesel exhaust and airborne particulates. Environ. Sci. Technol. 31 (10), 2772–2776.

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

Escher, B.I., Bramaz, N., Mueller, J.F., Quayle, P., Rutishauser, S., Vermeirssen, E.L.M., 2008. Toxic equivalent concentrations (TEQs) for baseline toxicity and specific modes of action as a tool to improve interpretation of ecotoxicity testing of environmental samples. J. Environ. Monit. 10 (5), 612–621.

Escher, B.I., Neale, P.A., Leusch, F.D.L., 2015. Effect-based trigger values for in vitro bioassays: reading across from existing water quality guideline values. Water Res. 81, 137–148.

Feiler, U., Ratte, M., Arts, G., Bazin, C., Brauer, F., Casado, C., Doren, L., Eklund, B., Gilberg, D., Grote, M., Gonsior, G., Hafner, C., Kopf, W., Lemnitzer, B., Liedtke, A.,

Matthias, U., Okos, E., Pandard, P., Scheerbaum, D., Schmitt-Jansen, M., Stewart, K., Teodorovic, I., Wenzel, A., Pluta, H.J., 2014. Inter-laboratory trial of a standardized sediment contact test with the aquatic plant Myriophyllum aquaticum (ISO 16191). Environ. Toxicol. Chem. 33 (3), 662–670.
Feldmannová, M., Hilscherová, K., Maršálek, B., Bláha, L., 2006. Effects of N-heterocyclic polyaromatic hydrocarbons on survival, reproduction, and biochemical parameters in Daphnia magna. Environ. Toxicol. 21 (4), 425–431.
Fritsch, E.B., Connon, R.E., Werner, I., Davies, R., Beggel, S., Feng, W., Pessah, I.N., 2013. Triclosan impairs swimming behavior and alters expression of excitation contraction coupling proteins in fathead minnow (Pimephales promelas). Environ. Sci. Technol. 47 (4), 2008–2017.
Gartiser, S., Hafner, C., Oeking, S., Paschke, A., 2009. Results of a "Whole Effluent Assessment" study from different industrial sectors in Germany according to OSPAR's WEA strategy. JEM 11 (2), 359–369.
Goto, T., Hiromi, J., 2003. Toxicity of 17alpha-ethynylestradiol and norethindrone, constituents of an oral contraceptive pill to the swimming and reproduction of cladoceran Daphnia magna, with special reference to their synergetic effect. Mar. Pollut. Bull. 47 (1–6), 139–142.
Guillen, D., Ginebreda, A., Farre, M., Darbra, R.M., Petrovic, M., Gros, M., Barcelo, D., 2012. Prioritization of chemicals in the aquatic environment based on risk assessment: analytical, modeling and regulatory perspective. Sci. Total Environ. 440, 236–252.
Hamers, T., Legler, J., Blaha, L., Hylland, K., Marigomez, I., Schipper, C.A., Segner, H., Vethaak, A.D., Witters, H., de Zwart, D., Leonards, P.E.G., 2013. Expert opinion on toxicity profiling—report from a NORMAN expert group meeting. Integr. Environ. Assess. Manag. 9 (2), 185–191.
Harada, A., Komori, K., Nakada, N., Kitamura, K., Suzuki, Y., 2008. Biological effects of PPCPs on aquatic lives and evaluation of river waters affected by different wastewater treatment levels. Water science and technology. J. Int. Assoc. Water Pollut. Res. 58 (8), 1541–1546.
Hartnik, T., Norli, H.R., Eggen, T., Breedveld, G.D., 2007. Bioassay-directed identification of toxic organic compounds in creosote-contaminated groundwater. Chemosphere 66 (3), 435–443.
Hecker, M., Hollert, H., 2011. Endocrine disruptor screening: regulatory perspectives and needs. Environ. Sci. Eur. 23 (1), 15.
Henry, N.D., Fair, P.A., 2013. Comparison of in vitro cytotoxicity, estrogenicity and anti-estrogenicity of triclosan, perfluorooctane sulfonate and perfluorooctanoic acid. J. Appl. Toxicol. 33 (4), 265–272.
Hoss, S., Ahlf, W., Bergtold, M., Bluebaum-Gronau, E., Brinke, M., Donnevert, G., Menzel, R., Mohlenkamp, C., Ratte, H.T., Traunspurger, W., von Danwitz, B., Pluta, H.J., 2012. Interlaboratory comparison of a standardized toxicity test using the nematode Caenorhabditis elegans (ISO 10872). Environ. Toxicol. Chem. 31 (7), 1525–1535.
IARC, I.A.f.R.o.C, 2014. IARC Working Group on the Evaluation of Carcinogenic Risk to Humans. Diesel and Gasoline Engine Exhausts and Some Nitroarenes.
Ishibashi, H., Matsumura, N., Hirano, M., Matsuoka, M., Shiratsuchi, H., Ishibashi, Y., Takao, Y., Arizono, K., 2004. Effects of triclosan on the early life stages and reproduction of medaka Oryzias latipes and induction of hepatic vitellogenin. Aquat. Toxicol. 67 (2), 167–179.
ISO, 2012a. ISO 6341:2012 Water Quality – Determination of the Inhibition of the Mobility of Daphnia Magna Straus (Cladocera, Crustacea) – Acute Toxicity Test.
ISO, 2012b. ISO 8692:2012 Water Quality – Fresh Water Algal Growth Inhibition Test with Unicellular Green Algae.
ISO, 2012c. ISO 11350:2012 Water Quality – Determination of the Genotoxicity of Water and Waste Water – Salmonella/microsome Fluctuation Test (Ames Fluctuation Test).
ISO, 2013. ISO/TC 147/SC 5 N 804 Water Quality — Determination of the Estrogenic Potential of Water and Waste Water — Part 1: Yeast Estrogen Screen (Saccharomyces cerevisiae).
Jarošová, B., Bláha, L., Giesy, J.P., Hilscherová, K., 2014. What level of estrogenic activity determined by in vitro assays in municipal waste waters can be considered as safe? Environ. Int. 64, 98–109.
Jonas, A., Scholz, S., Fetter, E., Sychrova, E., Novakova, K., Ortmann, J., Benisek, M., Adamovsky, O., Giesy, J.P., Hilscherova, K., 2015. Endocrine, teratogenic and neurotoxic effects of cyanobacteria detected by cellular in vitro and zebrafish embryos assays. Chemosphere 120, 321–327.
Keddy, C.J., Greene, J.C., Bonnell, M.A., 1995. Review of whole-organism bioassays: soil, freshwater sediment, and freshwater assessment in Canada. Ecotoxicol. Environ. Saf. 30 (3), 221–251.
Knobel, M., Busser, F.J., Rico-Rico, A., Kramer, N.I., Hermens, J.L., Hafner, C., Tanneberger, K., Schirmer, K., Scholz, S., 2012. Predicting adult fish acute lethality with the zebrafish embryo: relevance of test duration, endpoints, compound properties, and exposure concentration analysis. Environ. Sci. Technol. 46 (17), 9690–9700.
Kunz, P.Y., Kienle, C., Carere, M., Homazava, N., Kase, R., 2015. In vitro bioassays to screen for endocrine active pharmaceuticals in surface and waste waters. J. Pharm. Biomed. Analysis 106 (0), 107–115.
Legler, J., Zeinstra, L.M., Schuitemaker, F., Lanser, P.H., Bogerd, J., Brouwer, A., Vethaak, A.D., de Voogt, P., Murk, A.J., van der Burg, B., 2002. Comparison of in vivo and in vitro reporter gene assays for short-term screening of estrogenic activity. Environ. Sci. Technol. 36 (20), 4410–4415.
Leskinen, P., Michelini, E., Picard, D., Karp, M., Virta, M., 2005. Bioluminescent yeast assays for detecting estrogenic and androgenic activity in different matrices. Chemosphere 61 (2), 259–266.
Leskinen, P., Virta, M., Karp, M., 2003. One-step measurement of firefly luciferase activity in yeast. Yeast Chichester Engl. 20 (13), 1109–1113.
Loos, R., 2012. EU JRC Report. 2012: "Analytical Methods Relevant to the European Commission's 2012 Proposal on Priority Substances under the Water Framework Directive. ISBN 978-92-79-26642-3.
Loos, R., Gawlik, B.M., Locoro, G., Rimaviciute, E., Contini, S., Bidoglio, G., 2009. EU-wide survey of polar organic persistent pollutants in European river waters. Environ. Pollut. 157 (2), 561–568.
Lübcke-von Varel, U., Bataineh, M., Lohrmann, S., Löffler, I., Schulze, T., Flückiger-Isler, S., Neca, J., Machala, M., Brack, W., 2012. Identification and quantitative confirmation of dinitropyrenes and 3-nitrobenzanthrone as major mutagens in contaminated sediments. Environ. Int. 44 (0), 31–39.
Maes, H.M., Maletz, S.X., Ratte, H.T., Hollender, J., Schaeffer, A., 2014. Uptake, elimination, and biotransformation of 17α-Ethinylestradiol by the freshwater alga Desmodesmus subspicatus. Environ. Sci. Technol. 48 (20), 12354–12361.
Maletz, S., Floehr, T., Beier, S., Klumper, C., Brouwer, A., Behnisch, P., Higley, E., Giesy, J.P., Hecker, M., Gebhardt, W., Linnemann, V., Pinnekamp, J., Hollert, H., 2013. In vitro characterization of the effectiveness of enhanced sewage treatment processes to eliminate endocrine activity of hospital effluents. Water Res. 47 (4), 1545–1557.
Manusadžianas, L., Balkelyt≐, L., Sadauskas, K., Blinova, I., Põllumaa, L., Kahru, A., 2003. Ecotoxicological study of Lithuanian and Estonian wastewaters: selection of the biotests, and correspondence between toxicity and chemical-based indices. Aquat. Toxicol. 63 (1), 27–41.
McDonnell, D.P., Nawaz, Z., O'Malley, B.W., 1991. In situ distinction between steroid receptor binding and transactivation at a target gene. Mol. Cell. Biol. 11 (9), 4350–4355.
Murahashi, T., Iwanaga, E., Watanabe, T., Hirayama, T., 2003. Determination of the mutagen 3-Nitrobenzanthrone in rainwater collected in Kyoto, Japan. J. Health Sci. 49 (5), 386–390.
Murk, A.J., Legler, J., van Lipzig, M.M.H., Meerman, J.H.N., Belfroid, A.C., Spenkelink, A., van der Burg, B., Rijs, G.B.J., Vethaak, D., 2002. Detection of estrogenic potency in wastewater and surface water with three in vitro bioassays. Environ. Toxicol. Chem. 21 (1), 16–23.
Neale, P.A., Ait-Aissa, S., Brack, W., Creusot, N., Denison, M.S., Deutschmann, B., Hilscherova, K., Hollert, H., Krauss, M., Novák, J., Schulze, T., Seiler, T.B., Serra, H., Shao, Y., Escher, B.I., 2015. Linking in vitro effects and detected organic micropollutants in surface water using mixture toxicity modeling. Environ. Sci. Technol. 49 (24), 14614–14624.
OECD, 2004. Test No. 202: Daphnia sp. Acute Immobilisation Test. OECD Publishing.
OECD, 2011. Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test. OECD Publishing.
OECD, 2012. Test No. 457: BG1Luc Estrogen Receptor Transactivation Test Method for Identifying Estrogen Receptor Agonists and Antagonists. OECD Publishing.
OECD, 2013a. Test No. 210: Fish, Early-life Stage Toxicity Test. OECD Publishing.
OECD, 2013b. Test No. 236: Fish Embryo Acute Toxicity (FET) Test.
Oliveira, R., Domingues, I., Koppe Grisolia, C., Soares, A.M., 2009. Effects of triclosan on zebrafish early-life stages and adults. Environ. Sci. Pollut. Res. 16 (6), 679–688.
Orvos, D.R., Versteeg, D.J., Inauen, J., Capdevielle, M., Rothenstein, A., Cunningham, V., 2002. Aquatic toxicity of triclosan. Environ. Toxicol. Chem. 21 (7), 1338–1349.
OSPAR, 2007. OSPAR Hazardous Substances Committee. OSPAR Practical Study 2005 on Whole Effluent Assessment.
Pandard, P., Devillers, J., Charissou, A.-M., Poulsen, V., Jourdain, M.-J., Férard, J.-F., Grand, C., Bispo, A., 2006. Selecting a battery of bioassays for ecotoxicological characterization of wastes. Sci. Total Environ. 363 (1–3), 114–125.
Peddinghaus, S., Brinkmann, M., Bluhm, K., Sagner, A., Hinger, G., Braunbeck, T., Eisentrager, A., Tiehm, A., Hollert, H., Keiter, S.H., 2012. Quantitative assessment of the embryotoxic potential of NSO-heterocyclic compounds using zebrafish (Danio rerio). Reprod. Toxicol. Elmsford, N.Y. 33 (2), 224–232.
Peng, Y., Luo, Y., Nie, X.P., Liao, W., Yang, Y.F., Ying, G.G., 2013. Toxic effects of triclosan on the detoxification system and breeding of Daphnia magna. Ecotoxicol. Lond. Engl. 22 (9), 1384–1394.
Reifferscheid, G., Maes, H.M., Allner, B., Badurova, J., Belkin, S., Bluhm, K., Brauer, F., Bressling, J., Domeneghetti, S., Elad, T., Fluckiger-Isler, S., Grummt, H.J., Gurtler, R., Hecht, A., Heringa, M.B., Hollert, H., Huber, S., Kramer, M., Magdeburg, A., Ratte, H.T., Sauerborn-Klobucar, R., Sokolowski, A., Soldan, P., Smital, T., Stalter, D., Venier, P., Ziemann, C., Zipperle, J., Buchinger, S., 2012. International round-robin study on the Ames fluctuation test. Environ. Mol. Mutagen. 53 (3), 185–197.
Reiss, R., Mackay, N., Habig, C., Griffin, J., 2002. An ecological risk assessment for triclosan in lotic systems following discharge from wastewater treatment plants in the United States. Environ. Toxicol. Chem. 21 (11), 2483–2492.
Rogers, J.M., Denison, M.S., 2000. Recombinant cell bioassays for endocrine disruptors: development of a stably transfected human ovarian cell line for the detection of estrogenic and anti-estrogenic chemicals. In vitro Mol. Toxicol. 13 (1), 67–82.
Rojíčková, R., Dvořáková, D., Maršálek, B., 1998. The use of miniaturized algal bioassays in comparison to the standard flask assay. Environ. Toxicol. Water Qual. 13 (3), 235–241.
Römbke, J., Moser, H., 2009. Ecotoxicological Characterization of Waste - Results and Experiences of an International Ring Test. Springer Science + Business Media Verlag, New York.
Rosal, R., Rodea-Palomares, I., Boltes, K., Fernández-Piñas, F., Leganés, F., Petre, A., 2010. Ecotoxicological assessment of surfactants in the aquatic environment:

combined toxicity of docusate sodium with chlorinated pollutants. Chemosphere 81 (2), 288–293.
Routledge, E.J., Sumpter, J.P., 1996. Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. Environ. Toxicol. Chem. 15 (3), 241–248.
SCCP, S.C.o.C.P., 2009. Scientific Committee on Consumer Products SCCP Opinion on Triclosan COLIPA n° P32.
Schulze, T., Ahel, M., Ahlheim, J., Aït-Aïssa, S., Brion, F., Di Paolo, C., Hollender, J., Hollert, H., Kloß, A., Koprivica, S., Krauss, M., Oswald, P., Schollée, J., Seiler, T., Shao, Y., Slobodnik, J., Sonavane, M., Tousova, Z., Walz, K., Brack, W., 2016. A Novel Device for Onsite Integrative Large-volume Solid Phase Extraction for the Chemical and Effect-based Screening Analyses of Water Resources in preparation.
Schulze, T., Krauss, M., Novak, J., Hilscherova, K., Ait-Aissa, S., Creusot, N., Macova, M., Neale, P., Escher, B.I., Gomes, T., Tollefsen, K.E., Tarcai, Z., Shao, Y., Deutschmann, B., Seiler, T.B., Hollert, H., Tarabek, P., Tousova, Z., Slobodník, J., Walz, K.-H., Brack, W., 2015. Joint Danube Survey 3-A comprehensive analysis of Danube water quality. In: Liska, I., Wagner, F., Sengl, M., Deutsch, K., Slobodník, J. (Eds.), ICDPR - International Commission for the Protection of the Danube River, Vienna, pp. 284–295.
Svobodová, K., Plačková, M., Novotná, V., Cajthaml, T., 2009. Estrogenic and androgenic activity of PCBs, their chlorinated metabolites and other endocrine disruptors estimated with two in vitro yeast assays. Sci. Total Environ. 407 (22), 5921–5925.
Tamura, I., Kagota, K.-i., Yasuda, Y., Yoneda, S., Morita, J., Nakada, N., Kameda, Y., Kimura, K., Tatarazako, N., Yamamoto, H., 2013. Ecotoxicity and screening level ecotoxicological risk assessment of five antimicrobial agents: triclosan, triclocarban, resorcinol, phenoxyethanol and p-thymol. J. Appl. Toxicol. 33 (11), 1222–1229.
Tang, J.Y.M., Escher, B.I., 2014. Realistic environmental mixtures of micropollutants in surface, drinking, and recycled water: herbicides dominate the mixture toxicity toward algae. Environ. Toxicol. Chem. 33 (6), 1427–1436.
Tatarazako, N., Ishibashi, H., Teshima, K., Kishi, K., Arizono, K., 2004. Effects of triclosan on various aquatic organisms. Environmental sciences. Int. J. Environ. Physiol. Toxicol. 11 (2), 133–140.
Tixier, C., Singer, H.P., Canonica, S., Müller, S.R., 2002. Phototransformation of triclosan in surface waters: a relevant elimination process for this widely used BiocideLaboratory studies, field measurements, and modeling. Environ. Sci. Technol. 36 (16), 3482–3489.
Van den Belt, K., Berckmans, P., Vangenechten, C., Verheyen, R., Witters, H., 2004. Comparative study on the in vitro/in vivo estrogenic potencies of 17β-estradiol, estrone, 17α-ethynylestradiol and nonylphenol. Aquat. Toxicol. 66 (2), 183–195.
Villeneuve, D., Volz, D.C., Embry, M.R., Ankley, G.T., Belanger, S.E., Leonard, M., Schirmer, K., Tanguay, R., Truong, L., Wehmas, L., 2014. Investigating alternatives to the fish early-life stage test: a strategy for discovering and annotating adverse outcome pathways for early fish development. Environ. Toxicol. Chem. 33 (1), 158–169.
von der Ohe, P., Schmitt-Jansen, M., Slobodnik, J., Brack, W., 2012. Triclosan—the forgotten priority substance? Environ. Sci. Pollut. Res. 19 (2), 585–591.
von der Ohe, P.C., Dulio, V., Slobodnik, J., De Deckere, E., Kühne, R., Ebert, R.-U., Ginebreda, A., De Cooman, W., Schüürmann, G., Brack, W., 2011. A new risk assessment approach for the prioritization of 500 classical and emerging organic microcontaminants as potential river basin specific pollutants under the European Water Framework Directive. Sci. Total Environ. 409 (11), 2064–2077.
Watanabe, T., Ohe, T., Hirayama, T., 2005a. Occurrence and origin of mutagenicity in soil and water environment. Environmental sciences. Int. J. Environ. Physiol. Toxicol. 12 (6), 325–346.
Watanabe, T., Tomiyama, T., Nishijima, S., Kanda, Y., Murahashi, T., Hirayama, T., 2005b. Evaluation of genotoxicity of 3-Amino-, 3-Acetylamino- and 3-Nitrobenzanthrone using the ames Salmonella assay and the comet assay. J. Health Sci. 51 (5), 569–575.
Wernersson, A.-S., Carere, M., Maggi, C., Tusil, P., Soldan, P., James, A., Sanchez, W., Dulio, V., Broeg, K., Reifferscheid, G., Buchinger, S., Maas, H., Van Der Grinten, E., O'Toole, S., Ausili, A., Manfra, L., Marziali, L., Polesello, S., Lacchetti, I., Mancini, L., Lilja, K., Linderoth, M., Lundeberg, T., Fjallborg, B., Porsbring, T., Larsson, D., Bengtsson-Palme, J., Forlin, L., Kienle, C., Kunz, P., Vermeirssen, E., Werner, I., Robinson, C.D., Lyons, B., Katsiadaki, I., Whalley, C., den Haan, K., Messiaen, M., Clayton, H., Lettieri, T., Carvalho, R.N., Gawlik, B.M., Hollert, H., Di Paolo, C., Brack, W., Kammann, U., Kase, R., 2015. The European technical report on aquatic effect-based monitoring tools under the water framework directive. Environ. Sci. Eur. 27 (1), 7.
Wilson, V.S., Bobseine, K., Gray, L.E., 2004. Development and characterization of a cell line that stably expresses an estrogen-responsive luciferase reporter for the detection of estrogen receptor agonist and antagonists. Toxicol. Sci. 81 (1), 69–77.
Wolz, J., Fleig, M., Schulze, T., Maletz, S., Lubcke-von Varel, U., Reifferscheid, G., Kuhlers, D., Braunbeck, T., Brack, W., Hollert, H., 2010. Impact of contaminants bound to suspended particulate matter in the context of flood events. J. Soils Sediments 10, 1174–1185.
Yang, L.H., Ying, G.G., Su, H.C., Stauber, J.L., Adams, M.S., Binet, M.T., 2008. Growth-inhibiting effects of 12 antibacterial agents and their mixtures on the freshwater microalga Pseudokirchneriella subcapitata. Environ. Toxicol. Chem. 27 (5), 1201–1208.