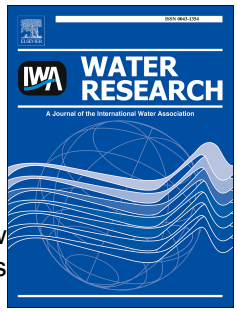
Accepted Manuscript



Effect-based assessment of toxicity removal during wastewater treatment

Pia Välitalo, Riccardo Massei, Ilse Heiskanen, Peter Behnisch, Werner Brack, Andrew J. Tindall, David Du Pasquier, Eberhard Küster, Anna Mikola, Tobias Schulze, Markus Sillanpää

PII: S0043-1354(17)30747-9

DOI: 10.1016/j.watres.2017.09.014

Reference: WR 13208

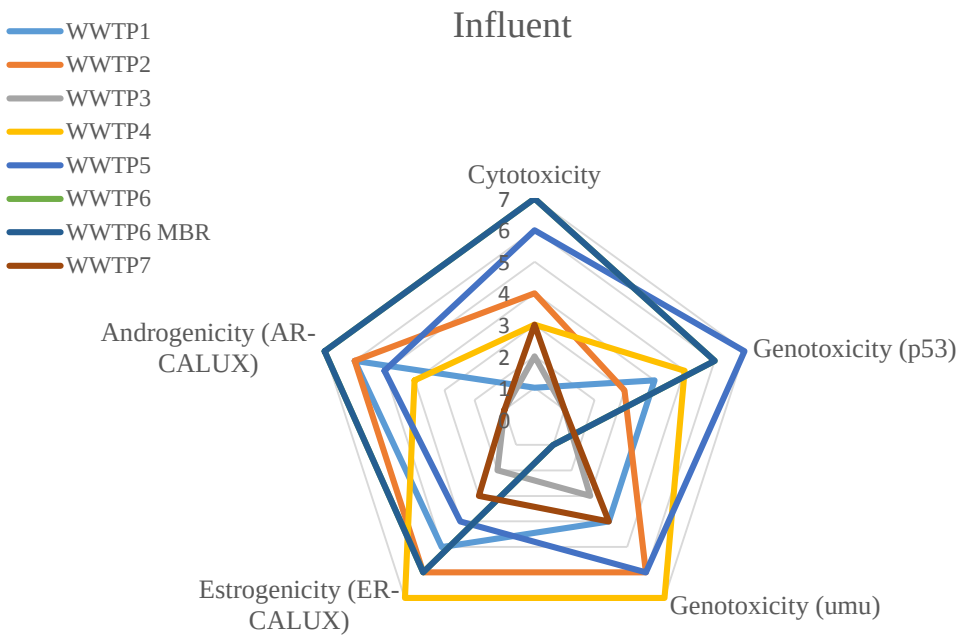
To appear in: *Water Research*

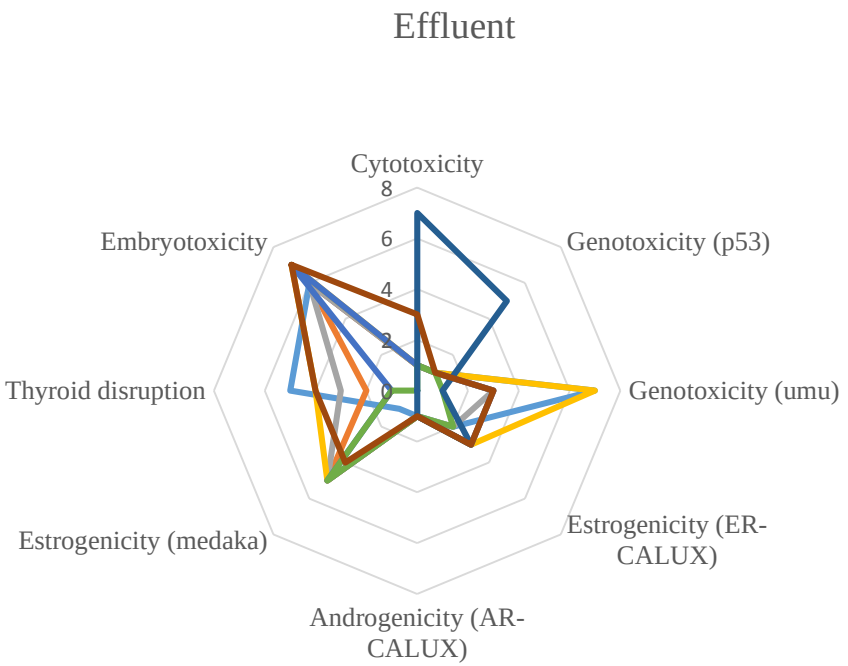
Received Date: 26 May 2017

Revised Date: 24 August 2017

Accepted Date: 4 September 2017

Please cite this article as: Välitalo, P., Massei, R., Heiskanen, I., Behnisch, P., Brack, W., Tindall, A.J., Du Pasquier, D., Küster, E., Mikola, A., Schulze, T., Sillanpää, M., Effect-based assessment of toxicity removal during wastewater treatment, *Water Research* (2017), doi: 10.1016/j.watres.2017.09.014.


Influent
WWTP1
WWTP2
WWTP3
WWTP4
WWTP5
WWTP6
WWTP6 MBR
WWTP7
Cytotoxicity
Genotoxicity (p53)
Genotoxicity (umu)
Estrogenicity (ER-CALUX)
Androgenicity (AR-CALUX)
0
1
2
3
4
5
6
7


Effluent
Cytotoxicity
Genotoxicity (p53)
Genotoxicity (umu)
Estrogenicity (ER-CALUX)
Androgenicity (AR-CALUX)
Estrogenicity (medaka)
Thyroid disruption
Embryotoxicity
0
2
4
6
8

Effect-based assessment of toxicity removal during wastewater treatment

Pia Välitalo[a,b,*], Riccardo Massei[c,d] , Ilse Heiskanen[a], Peter Behnisch[e], Werner Brack[c,d], Andrew J. Tindall[f], David Du Pasquier[f], Eberhard Küster[c], Anna Mikola[b], Tobias Schulze[c] and Markus Sillanpää[a]

a. Finnish Environment Institute, Laboratory Centre, Hakuninmaantie 6, 00430 Helsinki, Finland
b. Aalto University, Department of Civil and Environmental Engineering, Tietotie 1E, 02150 Espoo, Finland
c. UFZ - Helmholtz Centre for Environmental Research GmbH, Leipzig, German
d. Institute for Environmental Research (Biology V), RWTH Aachen University, Aachen, Germany.
e. BioDetection Systems b.v., Amsterdam, The Netherlands
f. WatchFrog S.A., 1 rue Pierre Fontaine, 91000 Evry, France

* Corresponding author: pia.valitalo@aalto.fi

Abstract

Wastewaters contain complex mixtures of chemicals, which can cause adverse toxic effects in the receiving environment. In the present study, the toxicity removal during wastewater treatment at seven municipal wastewater treatment plants (WWTPs) was investigated using an effect-based approach. A battery of eight bioassays was applied comprising of cytotoxicity, genotoxicity, endocrine disruption and fish embryo toxicity assays. Human cell-based CALUX assays, transgenic larval models and the fish embryo toxicity test were particularly sensitive to WWTP effluents. The results indicate that most effects were significantly reduced or completely removed during wastewater treatment (76 – 100%), while embryo toxicity, estrogenic activity and thyroid disruption were still detectable in the effluents suggesting that some harmful substances remain after treatment. The responsiveness of the bioassays was compared and the human cell-based CALUX assays showed highest responsiveness in the samples. Additionally, the fish embryo toxicity test and the transgenic larval models for endocrine disrupting effects showed high responsiveness at low sample concentrations in nearly all of the effluent samples. The results showed a similar effect pattern among all WWTPs investigated, indicating that the wastewater composition could be rather similar at different locations. There were no considerable differences in the toxicity removal efficiencies of the treatment plants and no correlation was observed with WWTP characteristics, such as process configuration or sludge age. This study demonstrated that a biotest battery comprising of multiple endpoints can serve as a powerful tool when assessing water quality or water treatment efficiency in a holistic manner. Rather than analyzing the concentrations of a few selected chemicals, bioassays can be used to complement traditional methods of monitoring in the future by assessing sum-parameter based effects, such as mixture effects, and tackling chemicals that are present at concentrations below chemical analytical detection limits.

Keywords: Wastewater; bioassays; toxicity removal; removal efficiency; in vitro; in vivo

1. Introduction

An increasing number of harmful chemicals are detected in wastewater treatment plant (WWTP) effluents and there is strong evidence that their discharge can lead to adverse environmental effects in the receiving waters (Goudreau et al., 1993; Kolpin et al., 2002; Vajda et al., 2008; Malaj et al., 2014; Prasse et al., 2015). Numerous studies analyzing micro-pollutants in WWTP effluents have highlighted insufficient removal of harmful substances (Halling-Sørensen et al., 1998; Ternes et al., 1999; Joss et al., 2005; Stasinakis et al., 2013; Luo et al., 2014; Välitalo et al., 2016). Despite rapid developments in analytical chemistry, it is not possible to analyze and identify all of the pollutants in wastewater due to limitations (e.g. cost and time). In addition, chemical analytical data does not provide information on the cumulative effects of complex compound mixtures in wastewater or on possible environmental effects. Thus, in order to get a more holistic view of the hazards posed by WWTP effluents, effect-based monitoring approaches are required to provide important complementary information to chemical analysis.

There are numerous effect-based tools available for water quality monitoring, including *in vitro* and *in vivo* bioassays (Escher et al., 2013; Leusch et al., 2014; Jia et al., 2015; König et al., 2017). However, despite this most studies investigating the removal efficiency of wastewater treatment plants (WWTPs) have focused on a few specific substances or toxicological endpoints, which is clearly insufficient for estimating the efficiency of hazard reduction by treatment processes. Previous studies have employed effect-based approaches to assess wastewater treatment efficiency on a laboratory scale or full-scale (Ma

et al.,2005; Margot et al., 2013; Wigh et al., 2016). Macova et al. (2011) applied an effect-based approach comprising of six endpoints to monitor organic pollutants across an indirect potable reuse scheme, including samples from one WWTP. However, a comprehensive bioassay battery has not been used to assess and compare multiple WWTPs.

In the present study, a battery of eight bioanalytical tools was applied to assess the toxicity of influent and effluents samples collected from seven municipal WWTPs in Finland. The selected methods cover multiple toxicological endpoints, such as cytotoxicity, genotoxicity, endocrine disruption and fish embryo toxicity. The test battery contained standardized assays and bioassays that were modified for wastewater analysis. The extensive ecotoxicological analysis was possible due to the use of a recently developed automated large-volume solid-phase extraction device (LVSPE50), enabling the extraction of large volumes of influent and effluent efficiently and relatively cost-effectively. In addition, at one WWTP the biological test battery was used to assess the performance of a newly installed membrane bioreactor (MBR) pilot facility compared to the conventional treatment process. The main goals of this research were to estimate how efficiently multiple toxic effects are reduced during wastewater treatment at typical Finnish WWTPs and to assess the water quality of influent and effluent based on their ecotoxicological profile. In addition, the most relevant toxicological endpoints were identified and the responsiveness of the selected bioassays was assessed.

2. Materials and Methods

2.1 Sample collection

Influent and effluent samples were collected from seven municipal WWTPs in Finland. The selected WWTPs represent typical treatment plants in Finland, where the most common secondary treatment process is activated sludge with enhanced biological nitrogen removal and simultaneous phosphorus precipitation. Tertiary treatment in order to improve phosphorus removal is also widely applied and chemical precipitation followed by sand filtration is the most common tertiary treatment step. Three of the selected treatment plants also have significant industrial loading. One of the studied treatment plants employs a pretreatment with an attached growth bioreactor and one operates a mixed-bed bioreactor (MBBR) in combination with dissolved air flotation (DAF). A detailed description of the WWTPs is given in Table 1. The following parameters describing the performance and the characteristics of the WWTPs were determined: average flow, sludge age, volumetric loading, suspended solids in influent and effluent, nitrification rate and the share of industrial loading. The samples (sample volumes presented in Table 2) were collected as 24-hour composite samples with the treatment plants' automated samplers between February and March 2015. The influent and effluent samples were collected according to the WWTP's hydraulic retention time in order to sample the "same" water in theory. The samples were transferred immediately to the laboratory for further sample treatment. At WWTP 6 there was an ongoing membrane bioreactor (MBR) pilot, thus two effluent samples were collected (after activated sludge process and after MBR treatment).

2.2 Large volume solid phase extraction

The influent and effluent samples were extracted in the laboratory by an automated large volume solid phase extraction device (LVSPE50), which was recently developed for the extraction of large volumes

93 (50 – 1000 L) of surface waters (Brack et al., 2016; Schultze et al., 2017). The principles of the device
94 and the approach are introduced in Schulze et al. (2017), however some modifications were made to
95 optimize the extraction process for wastewater samples. In short, a large SPE cartridge packed with 10 g
96 of HR-X sorbent material (Macherey Nagel Chromabond® HR-X) between two glass filter plates
97 (THOMAPOR® 50 mm) was applied. The samples were pre-filtered prior to extraction with Sartopure
98 GF+ Midicap filters, therefore particle bound contaminants are not considered in the present work. The
99 sorbent material was conditioned with 200 mL of ethyl acetate, 200 mL of methanol and 100 mL of
100 deionized water. The maximal volume of each sample was extracted with the device, which depended
101 on the rate at which the filters became clogged. The samples were extracted sequentially in portions of
102 500 mL. The extracted sample volumes are presented in Table 2.
103
104 After each sample extraction with the LVSPE50, the cartridge was dried overnight under a nitrogen
105 stream. After drying, the compounds of interest were eluted from the sorbent material with a sequential
106 elution scheme into four different fractions (100 mL ethyl acetate, 100 ml methanol, 100 mL methanol
107 with 1% formic acid, 100 mL methanol with 2% 7N-ammonia in methanol).
108
109 After elution, the acidic and basic fractions were neutralized to pH 7 ± 0.5 and all of the fractions were
110 filtered through filter paper (Whatman GF/F) to remove any residual interfering particles or salts. Each
111 fraction was evaporated to dryness with rotary evaporation and an EZ-Envi centrifugal evaporator
112 (Genevac Ltd, Ipswich, UK) and then re-dissolved in MeOH resulting in a final concentration factor of
113 5000x. These eluates were stored in the freezer (-20ºC) prior to analysis.
114
115

116 **Table 1.** Information on the seven WWTPs in Finland selected for sampling of effluent.

Location	Population equivalent	Average flow (m3/d)	Industrial influent % of the total and type	Secondary treatment	Tertiary treatment	Receiving water	Sludge age (d)	Volumetric loading (kgBOD/m³/d)	Influent suspended solids (kg/d)	Effluent suspended solids (SS mg/L)	Nitrification rate (%)
WWTP 1	40 000	17 000	4%, miscellaneous	Activated sludge	No	Baltic Sea	14	0.32	4098	3	93.5
WWTP 2	330 000	83 000	7%, miscellaneous	Activated sludge	Sand filtration	Baltic Sea	17.25	0.375	27579	2	97.6
WWTP 3	13 000	4 500	0%	Activated sludge	Sand filtration	Baltic Sea	17	0.081	670	7.5	98
WWTP 4	1 100 000	264 000	17%, miscellaneous	Activated sludge	Denitrifying filters	Baltic Sea	9	0.55	57138	5	98
WWTP 5	94 000	12 500	18%, dairy	MBBR + activated sludge	Sand filtration	River	30	0.37	2850	2.3	100
WWTP 6	16 000	2700	0%	Activated sludge and MBR pilot	No	River	16	0.15	900	1.9	100
WWTP 7	50 000	8000	85%, paper mill and meat processing	MBBR + flotation	No	River	14	0.17	4800	5.3	98

117

118

For the bioassays, the four fractions from each water sample were combined. The samples were divided into aliquots depending on the concentration factor required for each test. The combined eluates were evaporated to dryness with an EZ-Envi centrifugal evaporator and re-eluted in MeOH or DMSO depending on the test. Ten liters of LC-MS grade water (Chromasolv, Sigma-Aldrich) was extracted in the same way as the wastewater samples and used as an operational blank. The operational blank was analyzed in all of the biotests to check for possible background contamination from the sample treatment process.

Table 2. The amount of influent and effluent extracted from each WWTP by the LVSPE50 device.

Location	Influent (extracted volume, L)	Effluent (extracted volume, L)
WWTP 1	13	41
WWTP 2	15	39.5
WWTP 3	19.5	43
WWTP 4	22.5	40
WWTP 5	10	35
WWTP 6	6	13 (2.5 MBR pilot)
WWTP 7	16	36

For the bioassays the concentrations of the wastewater extracts were calculated as relative enrichment factors (REFs) according to (Macova et al., 2010). In short, the REF values were calculated by multiplying the dilution factor of each bioassay by the enrichment factor of the extracted sample (Eq. 1). The value represents the enrichment or dilution of the original water sample in each bioassay. The equations for calculating the dilution factor and the enrichment factor are presented in the Supplementary Information (SI, 1).

$$REF = dilution\ factor_{bioassay} \times enrichment\ factor_{SPE} \quad \text{(Eq. 1)}$$

142

143 **2.3 Biological analysis**

144

145 The samples extracted with the LVSPE50 device were analyzed with a battery of biological
146 toxicity tests comprising of bioassays for different toxicological endpoints (Table 3). Influents
147 and effluents were analyzed with five and eight bioassays, respectively. The selected assays
148 included organism-level assays and *in vitro* tests. A detailed description of the methods is
149 provided in the Supplementary Information (SI, 2). Briefly, the acute cytotoxic effects of the
150 influent and effluent samples were investigated by using the neutral red retention (NRR) assay
151 with a rainbow trout liver cell line RTL-W1 (Lee et al., 1993). Endocrine disrupting effects were
152 analyzed with multiple assays covering androgenic effects, estrogenic effects and thyroid
153 disruption. Both *in vitro* (AR-CALUX®, ER-CALUX®) and organism-level approaches were
154 applied (transgenic eleuthero-embryonic models for estrogen and thyroid axis activity).
155 Genotoxicity of the samples was evaluated with the standardized umuC assay and a newly
156 developed p53-CALUX® assay. Embryo toxic effects (lethal and sub-lethal effects) were
157 investigated with the standardized fish embryo toxicity test (FET). The sub-lethal effects that
158 were analyzed are presented in SI, Table 1.

159

160

161

162

163

164

165

Table 3. Toxicity assays selected for the biological analysis of influent and effluent samples.

Bioassay	Type	Toxicological endpoint	Influent samples analyzed	Effluent samples analyzed	Reference
NRR-retention test (RTL-W1 cells)	*In vitro*	Acute cytotoxicity	x	x	Lee et al., 1993
AR-CALUX®	*In vitro*	Androgenic activity	x	x	van der Linden et al., 2008
ER-CALUX®	*In vitro*	Estrogenic activity	x	x	van der Linder et al., 2008
Rapid estrogen activity in vivo (REACTIV) medaka assay	*In vivo*	Estrogenic activity		x	Spirhanzlova et al., 2016
Xenopus embryonic thyroid assay (XETA)	*In vivo*	Thyroid disruption		x	Fini et al., 2007
umuC assays	*In vivo*	Genotoxicity	x	x	ISO 13829 (2000)
p53-CALUX®	*In vitro*	Genotoxicity	x	x	van der Linden et al., 2014
Zebrafish embryo toxicity	*In vivo*	Embryotoxicity (survival, sublethal effects)		x	OECD TG 236 (2013)

3. Results and discussion

3.1 Biological effects of influent and effluent

All five toxicological endpoints were triggered in a majority of the seven influent samples and seven out of eight endpoints were found active in at least one out of eight WWTP effluent samples. The operational blank did not show positive responses in any of the bioassays at the tested concentrations, indicating that there was no background toxicity due to sample processing. The figures for the toxicities, which were detected at relatively low sample REFs after wastewater treatment process, are presented here. For the other endpoints, the figures and tables can be found in the supplementary information.

Cytotoxicity. The results of the NRR assay revealed cytotoxic potential of all the influent samples, especially when sample concentrations were REF10 or higher (SI, Table 2). Influent samples from WWTP 2, 5 and 6 were most cytotoxic with no viable cells at the three highest sample concentrations, whereas the cytotoxicity of the influent was clearly lower in samples from WWTP 1, 3 and 7 with more than 50 % viable cells at the highest test concentration. The cytotoxicity of the effluent samples was considerably lower compared to the raw wastewater samples, but minor effects were detected in six out of eight effluent samples. The most cytotoxic samples were the influent sample from WWTP 6 and the effluent sample from the MBR-pilot (WWTP 6).

Androgenic activity. Androgenic activity was detected in five (WWTP 1, 2, 4, 5, 6) out seven influent samples with dihydrotestosterone (DHT) equivalents ranging between 14 – 67 ng/L (SI, Fig. 2). None of the effluent samples showed androgenic activity above the detection limit. The highest activity was observed in WWTP 6 influent (67 ng DHT eq./L). Previous studies have reported DHT equivalents ranging from below the detection limit to 370 ng/L (Svenson and Allard 2004; Bain et al., 2014; Leusch et al., 2014). In this study, large differences between androgenic and estrogenic activities in influent samples were not observed, which support the previous findings of van der Linden et al. (2008) and Leusch et al. (2014).

Estrogenic activity. Estrogenic activity was detected in all of the influent and effluent samples with the ERα-CALUX® assay (Figure 1). The highest estrogenicity was observed in the influent sample from WWTP 4 (42 ng E2 eq./L), which is the largest WWTP. For the influent samples from WWTP 1, 2, 5 and 6 the estrogenic activity was in the range of 14 – 32 ng E2 eq./L.

Influent samples from WWTP 3 and 7 had the lowest estrogenic potential (0.45 – 1.6 ng E2 eq./L). Overall, estrogenic activity was substantially reduced in the effluent samples, where the results ranged between 0.61 – 3.1 ng E2 eq./L. The samples were tested only in agonistic mode in the ER-CALUX® assay. Therefore, the estrogenic activity of the samples may be underestimated as the presence of antagonists in the samples may decrease the net response (Ihara et al. 2014, Neale and Leusch 2015).

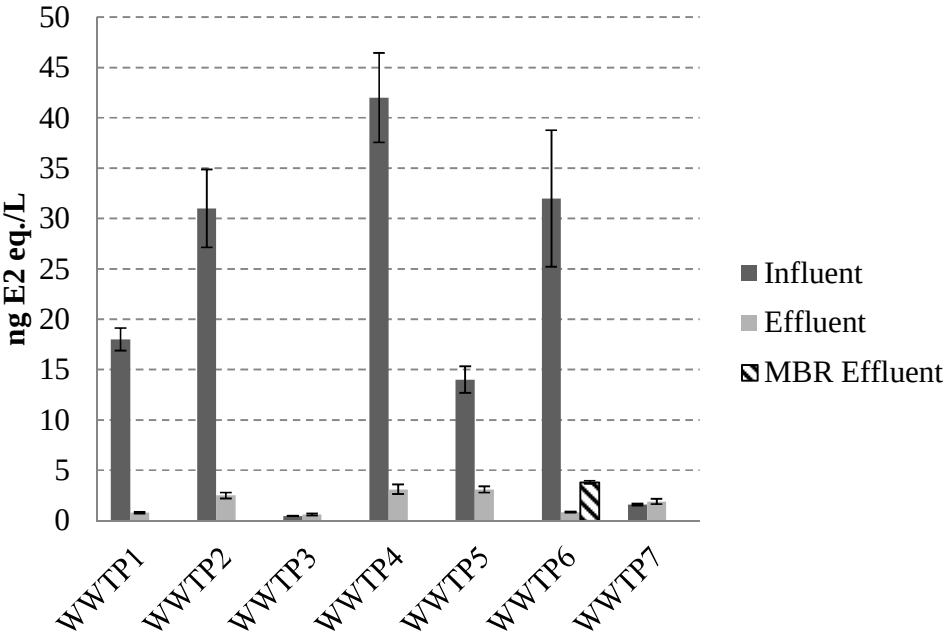


Figure 1. The estrogenic activity of influent and effluent samples from seven WWTPs analyzed with ERα-CALUX®. The error bars represent standard deviation of the bioassay replicates.

The effluent samples (excluding WWTP 6 MBR pilot) were also tested for estrogen disruption in the presence (spiked mode) and absence (unspiked mode) of testosterone with transgenic medaka embryos (see SI, section 2.2.2 for a detailed description of the assay). All of the effluent samples caused 100 % mortality of the embryos at REF10. A ten-fold lower concentration (REF1) was therefore selected for the estrogen disruption test. Significant estrogenic activity was found in samples from all of the WWTPs in the absence of testosterone, except WWTP 1 (SI,

Fig. 3). The results are similar to those obtained with the ER-CALUX®, however slight differences can be observed. In both of the assays, WWTP 4 effluent samples show highest estrogenic potency. However, in the ER-CALUX® assay WWTP 3 has the lowest estrogenic activity, whereas in the in vivo medaka assay, the estrogenicity of WWTP 3 effluent is not notably lower than the other samples. This might indicate that the assays respond in a different way to some specific compounds. This is likely as the *in vivo* transgenic medaka assay can detect the effects of compounds not acting directly on the estrogen receptor as well as compounds requiring metabolic activation. The samples were also tested for antagonistic effects and modulation of aromatase enzyme by spiking the samples with testosterone in the medaka assay. No significant effects were detected in any of the samples in spiked mode (data not shown).

Thyroid disruption. Most studies on endocrine disrupting potency of wastewaters have focused on estrogenic and androgenic activities. However, thyroid hormones (THs) are important modulators of development and physiology and identification of adverse effects on TH signaling is important when considering wildlife health. Effluent samples (excluding WWTP 6 MBR) were tested for thyroid disruption with a transgenic line of *Xenopus* embryos (see SI, section 2.2.2 for a detailed description of the assay). Thyroid disruption was detected in the unspiked effluent samples from WWTP 1, 2, 3, 4 and 7 (Figure 2) at REF1 and triiodothyronine T3 equivalents ranged between undetected to 1.34 µg/L. The samples were also tested in the presence of T3 (spiked mode), to reveal antagonistic effects and more complex effects such as disruption of thyroid hormone metabolism. Significant pro-thyroid effects were observed for the same WWTPs showing effects in unspiked mode. However, more marked effects were observed in spiked mode ranging from undetected to 3.71 µg/L T3 equivalents.

246

247 The results of the present study are somewhat unexpected compared to previous *in vitro* studies,
248 where effluents have not shown high potential for thyroid disruption activity. Jugan et al. (2009)
249 detected thyroidal activity in influent samples but not effluent samples with cell-based luciferase
250 reporter gene assays. Additionally, Escher et al. (2013) did not detect any responses in thyroid
251 receptor based *in vitr*o assays in effluent samples. However, it has been well established that
252 thyroid hormone disruptors in particular act via non-receptor based mechanisms of action. It is
253 therefore particularly important for thyroid disruption to evaluate non-receptor based thyroid
254 disruption which is unlikely to be detected by *in vitro* (Wegner et al. 2016).

255

256 Castillo et al. (2013) studied thyroid disruption of wastewaters with the same transgenic *Xenopus*
257 *laevis* embryos as used in the present study and observed thyroid disruption in most influent
258 samples and some effluent samples. However, the thyroid disruption potential of the effluent
259 samples was clearly lower than the untreated wastewaters. It is well recognized that some
260 brominated flame-retardants and the antimicrobial compound triclosan can cause thyroid-
261 disrupting effects (Veldhoen et al., 2006; Crofton et al., 2007; Boas et al., 2012), and these
262 chemicals have been frequently detected in Finnish wastewater samples (Huhtala et al., 2011;
263 Vieno, 2014). In the next phase of this project, the same WWTP samples will be analyzed to
264 determine the concentrations of various organic contaminants, and perhaps the results from the
265 chemical analysis will provide more insight on the effluent thyroid disruption activity (data not
266 yet published).

267

268

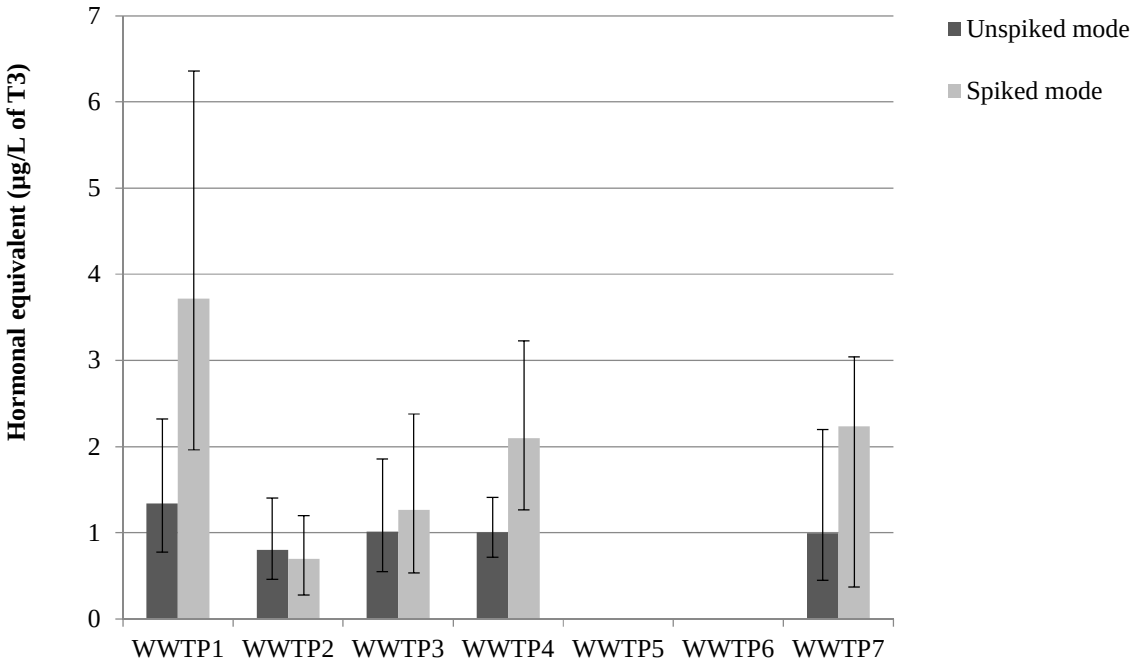


Figure 2. Thyroid disrupting effects of the effluents samples analyzed with transgenic *Xenopus laevis* embryos (XETA assay) in spiked and unspiked modes. Results are expressed as T3 hormonal equivalents (µg/L). The results for WWTP5 and 6 were below the limit of detection. The error bars represent 95% confidence interval.

Genotoxicity. The genotoxicity of the influent and effluent samples was analyzed with a newly developed p53-CALUX® assay (van der Linden et al., 2014) and a commonly employed *Salmonella typhimurium* TA 1535 pSK1002 (umuC-assay). Six of the influent samples showed activation of p53 protein in the test with metabolic activation with S9. No effects were detected in the tests without S9. The genotoxic activity varied greatly between the WWTPs (61 – 6200 µg cyclophosphamide eq./L) (SI, Fig. 4). The genotoxic potency was considerably higher in the influent sample from WWTP 5 compared to the other samples. The genotoxic effects were reduced to below the limit of detection (<53 µg cyclophosphamide eq./L) in all of the effluent samples except the MBR pilot effluent, where the cyclophosphamide equivalent value was 540 µg/L. The samples did not show high genotoxic potential in the umuC-assay. Genotoxic effects

in the influent and effluent samples were detected only in the most concentrated samples (>REF20) (SI, Table 3). Many of the influent samples were cytotoxic to the bacteria in the highest sample concentration (REF40), thus the induction ratio could not be calculated in those cases. Influent samples from WWTP 4 and WWTP 5 were the most cytotoxic with bacterial growth factors under 0.5 also at REF20. Cytotoxic effects were observed only in tests without metabolic activation with S9. None of the influent or effluent samples showed genotoxic activity at any of the lower concentrations (REF10 and REF5). Only two effluent samples (WWTP 1 and WWTP 4) had induction ratios exceeding the threshold value at REF20 in addition to REF40. In all of the cases, samples had higher induction ratios in the tests without metabolic activation.

Zebrafish embryo toxicity. The use of fish acute toxicity test in environmental risk assessment is becoming a routine in several European countries (Scholz et al., 2008). The assessment of environmental quality can include acute effects (i.e. lethality) and interference with development, growth and reproduction (Embry et al., 2010). In the present study, *Danio rerio* embryos showed high sensitivity towards the tested wastewater effluent extracts. Toxic effects were observed in all of the samples with considerable mortality (20 – 43 %) even at the lowest exposure concentration (REF2.5). Mortality was evidently the primary effect induced by the samples, since lethal endpoints accounted for 95.5 % on average of all the observed endpoints throughout the tested samples and dilutions. Mortality in the negative controls was below 15% in all three replicates.

Toxic effects were detected in the first 24 h with 85.5 % embryo mortality at REF10. Embryos showed also several malformations such as scoliosis, lordosis and pericardial edema at different

evaluated time points and effluent samples (Fig. 3 and SI, Fig. 5). However, none of the malformation proved to be site-specific but rather a general stress response of fish embryos to environmental mixtures. These effects can be induced by many compounds and it was not possible to detect any specific responses. Due to the extracted sample volumes, only the effluent samples (excluding WWTP 6) were analyzed. No considerable differences were detected between the different WWTPs (Fig. 4). Oxygen levels and pH were stable for the duration of the exposure.

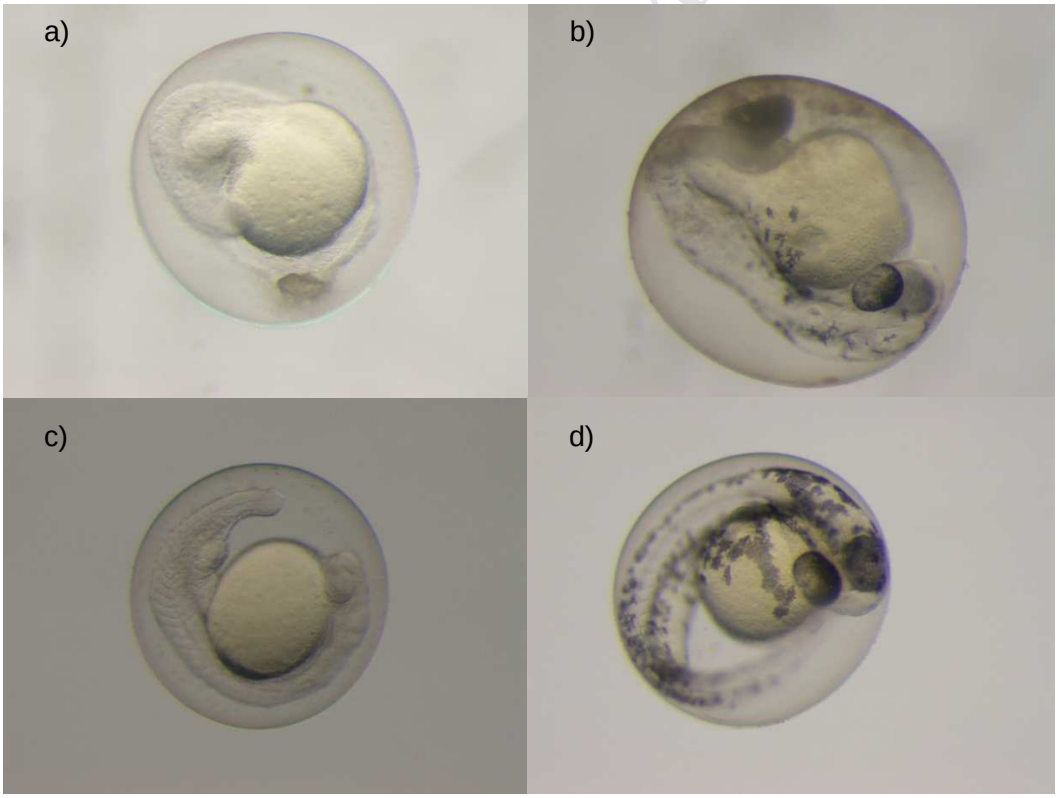


Figure 3. Examples of malformations in *D. rerio* embryos exposed to the wastewater samples after a) 24 h b) 48 h. C and d are examples of normal embryos at 24 and 48 h, respectively.

The successful application of early life stages of fish for wastewater toxicity testing has been demonstrated in many studies. As an example, Zha and Wang (2006) already showed the use of larval stages of Japanese medaka (*Oryzias latipes*) to assess the toxicity from a banknote printing plant after a 9-day exposure. They observed several lesions in the embryos, such as pericardial edema and hemostasis, at concentrations as low as 12.5% of the native samples. Also Galus et al. (2013) exposed zebrafish embryos and adult fish to municipal wastewater from Ontario, Canada. The exposure to a higher concentration (25% diluted) of wastewater significantly increased the incidence of developmental abnormities in adults. These studies were conducted by exposing the embryos directly to the water sample without previous extraction. As demonstrated in the present study, the FET test can also be applied to analyze extracts dissolved in a carrier solvent.

One of the major advantages of using organic carrier solvents is the possibility to concentrate the samples by several orders of magnitude without changing the final volume of the exposure media (Tanneberger et al., 2010). By concentrating the samples, it is possible to obtain acute toxicity data also from samples where toxic substances are present at low concentrations. The information obtained can be applied for prolonged toxicity studies (Arome and Chinedu 2013). Moreover, sample enrichment allows the percentage of solvent in the test media to be reduced and minimizes its potential toxic effect during exposure (Hallare et al. 2005).

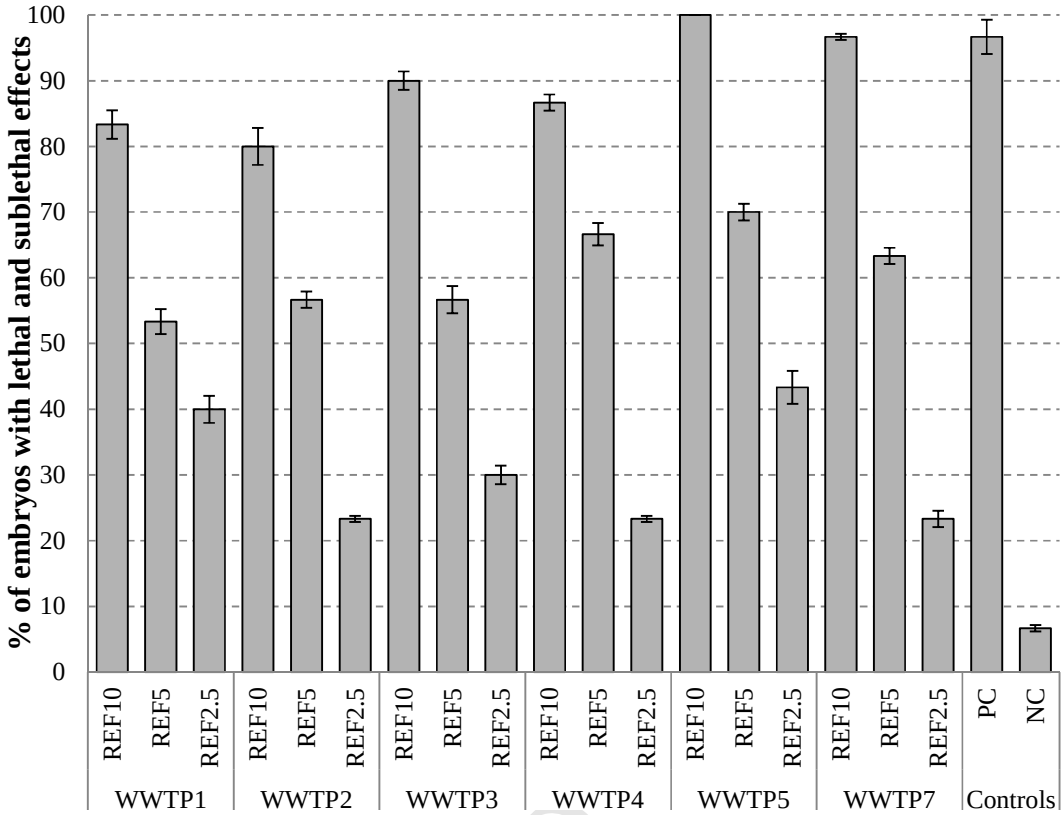


Figure 4. Total embryotoxic effects (lethal and sublethal) observed in effluent samples from 6 WWTPs at three different sample concentrations (REF10, 5 and 2.5) as well as positive and negative controls (PC and NC respectively).

3.2 Responsiveness of the bioassays and their use as screening tools

The responsiveness of the bioassays varied depending on the assay method, sample and endpoint (Fig. 5). Figure 5 displays an overview of the responsiveness of the selected assays in all of the tested influent and effluent samples as a heat map. Color coding indicates the ratio between the lowest sample enrichment (REF) and the lowest negative control enrichment eliciting a toxic response. Red indicates sample effects at a low enrichment (high potency) and dark green for

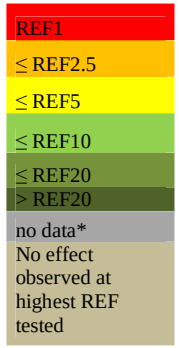
sample effects at a high enrichment (low potency). Naturally, responsiveness is related to assay sensitivity. However, even the most sensitive assay cannot detect toxicity in the absence of the chemicals that can activate the bioassay endpoint. The results of assay responsiveness can provide useful information regarding the suitability of bioassays for monitoring purposes or for assessing the efficiency of wastewater treatment processes (Escher et al. 2013).

Comparisons between the samples show that the more polluted influent samples induced more positive results at lower sample concentrations than the effluent samples. Positive responses were detected in all of the five endpoints that were tested with influents. The p53-CALUX® test was more sensitive than the umuC assay, because in the CALUX assay genotoxic effects were observed in the influent samples even at REF1, whereas none of the samples showed any genotoxic activity at concentrations lower than REF20 in the umuC assay. Interestingly, the results related to metabolic activation with S9 differed between the assays. In the p53-CALUX® assay genotoxic effects were only seen in the +S9 test, whereas in the umuC assay more genotoxicity was observed in the tests without metabolic activation, which might indicate that the tests are responding to different compounds. In addition, the differences in the results could be partly explained by physiological differences between the test organisms, e.g. bacteria do not have a nuclear envelope protecting the DNA as opposed to the eukaryotic cells used in the CALUX assays.

Based on the overall results, the key endpoints related to wastewater toxicity were estrogenic activity, thyroid disruption and fish embryo toxicity. These endpoints were activated in the majority of the samples and responses were detected at low sample concentrations indicating high toxic potency. The human cell-based CALUX® assays showed highest responsiveness to the

379 influent samples with positive results detected at REF1. The ER-CALUX® assay for estrogenic
380 activity was the most responsive cell-based assay, as in all of the samples an estrogenic response
381 was detected at low exposure concentrations. The effluent samples induced embryotoxicity in the
382 FET assay in all of the samples at REF2.5, suggesting that the toxic effects may be caused by
383 chemicals that are typically present in all municipal wastewaters. Additionally, positive
384 responses were detected in the transgenic larval models for endocrine disruption at low sample
385 concentrations in several effluent samples.

386 The heat map forms a bioanalytical fingerprint for each sample, which can be used to assess the
387 most relevant endpoints related to the risks posed by effluent discharges to the aquatic
388 environments concerning assay responsiveness. A battery of bioassays selected to cover relevant
389 biological endpoints can be used as a comprehensive tool for indicating water quality. The
390 battery should include endpoints for detecting general toxicity such as cytotoxicity, as well as
391 bioassays with more specific endpoints, such as the receptor-mediated tests for endocrine
392 disruption. As shown in this study, the inclusion of a sensitive *in vivo* assay such as the FET test
393 can also be beneficial. Escher et al. (2013) also suggested that specific receptor-mediated modes
394 of action for endocrine disruption and assays for reactive modes of action, such as umuC for
395 genotoxicity, are promising assays for screening water quality. However, the most sensitive
396 methods should be applied as the genotoxic potency of effluents may be low as indicated in the
397 present study.



	WWTP1		WWTP2		WWTP3		WWTP4		WWTP5		WWTP6			WWTP7	
Endpoint/Bioassay	Influent	Effluent	Influent	Effluent	Influent	Effluent	Influent	Effluent	Influent	Effluent	Influent	Effluent	MBR Effluent	Influent	Effluent
Cytotoxicity (NRR)															
Estrogenic activity (ER-CALUX)															
Estrogenic activity (medaka) unspiked mode		>REF1													
Estrogenic activity (medaka) spiked mode		>REF1		>REF1		>REF1		>REF1		>REF1		>REF1			>REF1
Androgenic activity (AR-CALUX)															
Genotoxicity (p-53CALUX)															
Genotoxicity (umuC-assay)															
Embryotoxicity (FET)															
Thyroid disruption (*Xenopus*) unspiked mode										>REF5		>REF5			
Thyroid disruption (*Xenopus*) spiked mode										>REF5		>REF5			

Figure 5. The heat map of all the bioassays for influent and effluent samples from 7 WWTPs. The effect concentrations are plotted as the lowest sample concentrations (REF) where a toxic effect was compared to the negative control. The colors indicate the level of sample enrichment: red indicates an assay response with low enrichment, whereas dark green corresponds to low potency with toxic effects detected only at high sample enrichment.

411 **3.3 Toxicity removal during wastewater treatment**

412

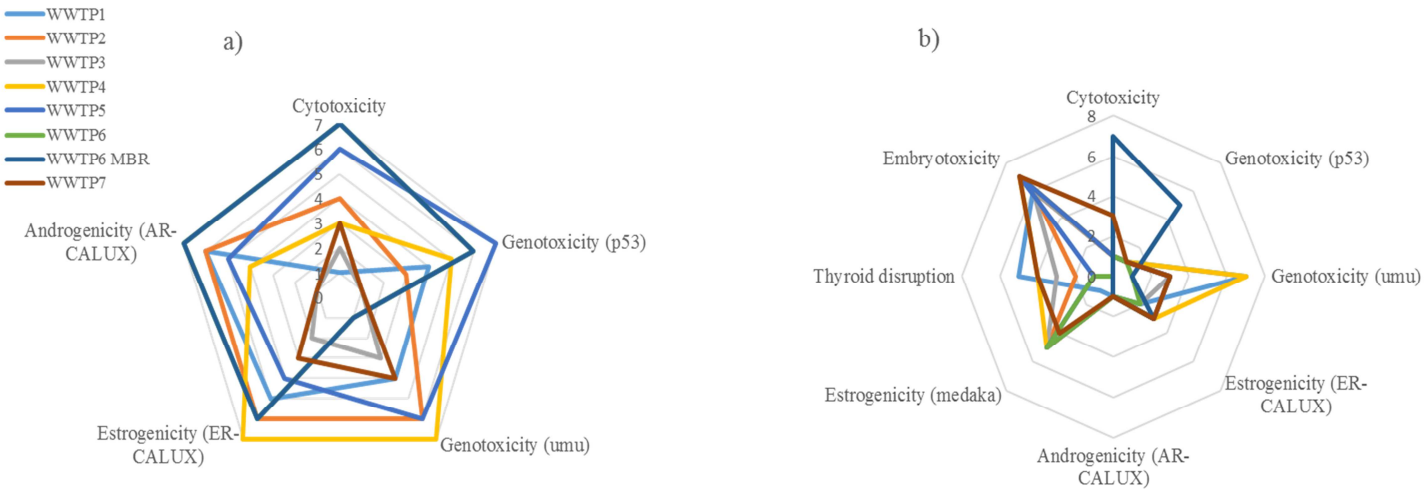
413 The efficiency of the WWTPs to reduce toxicity was calculated by comparing the toxicity of the
414 influent and effluent samples. There was no correlation between toxicity removal efficiencies
415 and the presence or absence of a tertiary treatment step at the WWTP. However, it should be
416 noted that these treatment steps were originally designed for the removal of phosphorus and
417 suspended solids rather than micro-pollutants. Additionally, other WWTP parameters, such as
418 sludge age or nitrification rate, did not correlate with toxicity removal efficiency either. To our
419 knowledge, there are no previous studies on the correlation between sludge age or sludge
420 retention time (SRT) and general toxicity removal. Previous studies have been focused on
421 individual chemical compounds, such as pharmaceuticals or hormones. Higher biodegradation
422 rates for micro-pollutants have been observed in processes with higher sludge retention times or
423 sludge age, such as membrane bioreactors (Vieno and Sillanpää, 2014; Kruglova et al., 2016).
424 The benefits related to increased degradation rates of micro-pollutants with higher sludge age
425 seems to reach an optimal level at approximately 20 – 25 days (Zeng et al., 2013; Fålas et al.,
426 2016). However, in the present study clear differences in toxicity removal between WWTPs with
427 the lowest sludge age (9 days at WWTP 4) or the highest sludge age (30 days at WWTP 5) were
428 not seen. It is possible that other operational parameters and factors (e.g. temperature, organic
429 loading rates) could greatly affect the removal efficiency (Kruglova et al., 2014), and thus further
430 research is needed in order to draw distinct conclusions. In agreement with the present study,
431 some previous investigations have failed to find a correlation between operational parameters
432 and pollutant removal. For example, Joss et al. (2005) observed no clear dependency between
433 sludge age, temperature or reactor configuration and compound removal. They concluded that

sludge age unexpectedly showed no significant impact on the transformation efficiency of the seven pharmaceuticals analyzed.

Three out of four WWTPs with a tertiary treatment step had sand filtration as the final treatment step and one WWTP had denitrifying filters. The results suggest that sand filtration does not provide conclusive advantages related to toxicity removal. To our knowledge, there are no previous studies that have investigated the removal efficiency of sand filtration as a tertiary treatment step related to multiple toxic effects. Previous research had focused on determining the removal of specific substances during sand filtration, and those studies have shown that sand filtration does not significantly improve pollutant removal, which support the findings of the present study. Okuda et al. (2008) concluded that during sand filtration process, the total concentrations of 66 pharmaceuticals did not decrease significantly. Nakada et al. (2007) showed that the removal of pharmaceutically active compounds was generally inefficient during sand filtration, perhaps due to the hydrophilic nature of the selected target compounds supporting the findings of the present study. Koh et al. (2008) showed that biological processes play the most important role in removing estrogenic activity through biotransformation and biodegradation, indicating that sand filtration does not significantly improve the removal of estrogens. Also according to Leusch et al. (2005) estrogenic activity remained unchanged following sand filtration. Other tertiary treatment steps, such as ozonation and activated carbon, have shown more promising results related to toxicity removal (Reungoat et al., 2011; Altmann et al., 2014; Luo et al., 2014; Mathon et al., 2017).

456 The findings of the present study suggest that the removal efficiency was more related to each
457 toxicological endpoint than characteristics of the WWTPs. The toxicity removal related to all of
458 the selected endpoints is summarized as radar charts in Figure 6. In these charts, the toxicity of
459 each sample is evaluated by a simple scoring system, where a value between 1 and 7 is given, 7
460 indicating higher toxicity and 1 no toxicity. The score bands for each bioassay are presented in
461 SI, Table 4. The toxicity pattern of the influent samples was similar between the WWTPs, as
462 almost all of the samples induced toxic effects in the majority of the bioassays. Furthermore,
463 there was some variation depending on the endpoint since the influent of certain WWTPs was
464 clearly more androgenic, cytotoxic, genotoxic or estrogenic than the others. Based on the overall
465 results, influent samples from WWTP3 and WWTP7 were the least toxic and influent samples
466 from WWTP4, WWTP5 and WWTP6 were the most toxic. The radar diagrams for effluent
467 toxicity clearly show that the remaining toxicities after treatment are embryo toxicity, estrogenic
468 activity, thyroid disruption and genotoxicity (umuC).
469
470 When looking at the different toxicological endpoints in more detail, some variation in removal
471 efficiency between the WWTPs can be observed. The cytotoxicity was substantially or
472 completely reduced during the wastewater treatment process in all of the WWTPs, except
473 WWTP 7 and the MBR pilot plant at WWTP 6, where there was no significant reduction in
474 cytotoxic effects. In general cytotoxicity reduction was high (76 – 89 %) and in the cases of
475 WWTP 5 and WWTP 6 the toxic effects were completely removed in the highest sample
476 concentration tested.
477
478

479



480 **Figure 6.** The toxicity of the influent (a) and effluent (b) samples related to all of the selected
481 endpoints. The toxicity of each sample is scored by giving a value between 1 and 7, 7 indicating
482 higher toxicity and 1 representing no toxicity. The results from WWTP 6 influent sample follow
483 the same line as for WWT6 MBR.

484

485 The results of this study suggest that the activated sludge process is the most effective treatment
486 step at removing cytotoxicity from the studied WWTPs. Toxicity removal of cytotoxic effects
487 was at the lowest level at WWTP 7, which employs the MBBR + DAF process. WWTP 7 had
488 the highest industrial loading, which could partly explain the outcome as industrial influent may
489 contain more compounds that are less biodegradable compared to typical municipal wastewaters.
490 At WWTP 5 where the MBBR process is combined with activated sludge, the toxic effects were
491 completely removed. Having sand filtration as a tertiary treatment step did not improve the
492 removal of cytotoxic effects, since there was no clear correlation between better removal
493 efficiency and sand filtration. Previous studies have also shown that influents are typically highly
494 cytotoxic and that toxicity is significantly reduced during conventional wastewater treatment
495 processes (Smital et al., 2011; Stalter et al., 2011). Similar results have also been shown with

bacterial assays. For example, Castillo et al. (2001) observed a substantial decrease of the inhibition of bacteria, from 70–80 % down to 15–20 % when analyzing the WWTP influent versus effluent. Cytotoxicity of wastewater influents has also been linked to linear alkylbenzene-sulfonates, which are surfactants mainly used in laundry products (Castillo et al., 2001). Surfactants are typically present at high concentrations in wastewaters (Smital et al., 2011), however cytotoxicity of the samples may also be linked to other substances.

Androgenic effects were most efficiently removed during the wastewater treatment processes and no androgenic activity was detected in any of the effluent samples. This suggests that androgenic endocrine disruption is of less concern than estrogenic endocrine disruption in regard to organisms in WWTP effluent receiving waters. The results from the present study support previous findings showing that androgenic effects were efficiently removed during conventional wastewater treatment (Bain et al., 2014; Leusch et al., 2014). However as no *in vivo* androgen assays were included in the test battery, it cannot be excluded that androgen disruptors remained that required metabolic activation or did not interact directly with the androgen receptor.

A similar trend was observed in the case of genotoxicity. In the adaptive stress response assay (p53-CALUX®) the genotoxic effects were reduced to below the limit of detection (<53 µg cyclophosphamide eq./L) in all of the effluent samples except the MBR pilot effluent, where the genotoxic potency was reduced only by 16 %. This finding suggests that the pilot plant was not operating at the targeted level and more sampling would have been necessary to draw further conclusions. Overall, the results based on the p53-CALUX assay indicate that the compounds causing genotoxic effects are removed efficiently during the conventional treatment processes.

519 However, the results based on the umuC-assay suggest that the genotoxic effects are not reduced
520 during wastewater treatment, although effects are only detected at high sample concentrations.
521 The results from previous studies have also presented varying results. Al-Saleh et al. (2017)
522 showed that effluents still had high genotoxic potential after wastewater treatment process.
523 Additionally, genotoxic potential of wastewater effluents was demonstrated in a study by Escher
524 et al. (2014) and Jolibois and Guerbet (2005) with several assays based on reactive modes of
525 action (e.g. umuC assay and SOS chromotest). Magdeburg et al. (2014) demonstrated significant
526 genotoxic effects in samples taken after secondary sedimentation, which were effectively
527 reduced by an ozonation process but were not further reduced by sand filtration following the
528 ozonation step. In a study by Žegura et al. (2009) genotoxic effects were not observed in influent
529 samples but were detected in some of the corresponding effluent samples, which may be due to
530 the formation of genotoxic compounds during the biological treatment of wastewaters.
531
532 The reduction in estrogenic activity was between 78 – 97 % due to the water treatment in
533 majority of the WWTPs. In WWTP 3 and 7 estrogenic activity was not removed at all, however
534 in those samples the estrogenic potential of the influent was low to begin with. The removal
535 efficiency of the MBR pilot in WWTP 6 was lower than the efficiency of the normal treatment
536 process (normal 97 %, MBR 88 %). A number of studies have investigated the removal of
537 steroid hormones using membrane bioreactors. Some of the studies have shown that MBR
538 removes estrogens more efficiently than conventional activated sludge process (Zuehlke et al.,
539 2006; Maletz et al., 2013). On the other hand, Ivashechkin et al. (2004) and Weber et al. (2005)
540 did not find an appreciable difference in removal of estrogens between membrane activated
541 sludge or conventional activated sludge systems. In the present study, any conclusions on the

542 removal efficiency of the MBR process compared to conventional activated sludge are difficult
543 to draw, because the MBR system was a newly installed pilot and the operational parameters
544 might not have been fully optimized as indicated by the other results from this study. In addition,
545 it should be acknowledged that the present study is based on one sampling event and the results
546 can vary depending on the time, temperature and other varying parameters. The majority of the
547 previous studies focusing on estrogens in wastewaters have used chemical analytical tools to
548 analyze the removal efficiencies of estrogenic compounds (Ternes et al., 1999; Johnson,
549 Belfroid, & Di Corcia, 2000; Ying, Kookana, & Kumar, 2008; Xu et al., 2012; Luo et al., 2014;).
550 Some previous studies have employed the ERα-CALUX® to study removal efficiencies of
551 estrogenic activity during wastewater treatment processes (Murk et al., 2002; Maletz et al., 2013;
552 Bain et al., 2014) and their findings support the results of the present study. Murk et al. (2002)
553 showed substantial reductions (90 – 95 %) of estrogenic potency in effluents compared to
554 influents in municipal WWTPs. Similar results were shown in a more recent study with
555 reductions between 89 – 100 % for estrogenic activity in three Australian WWTPs (Bain et al.,
556 2014). Their results suggest that tertiary treatment processes (flocculation, tertiary clarification,
557 dual-media filtration, chlorine disinfections and dechlorination) contribute markedly to the
558 enhanced reduction of estrogenic potency following conventional treatment.
559
560 The estimation of the risks posed by the treated effluents to the receiving waters is challenging as
561 many factors, such as dilution and flow rate of the receiving water, affect the actual risk. The
562 chemicals causing toxic effects in wastewater effluents are typically present at low
563 concentrations and concentration of the samples is often necessary to observe ecotoxicological
564 effects in acute tests. In the receiving waters, the effluents are diluted, but the exposure is

565 typically constant and long-term. One approach to assess the bioassay results in terms of risk
566 context, is to apply effect-based trigger (EBT) values (Escher et al., 2013; Jarošová et al., 2014;
567 Escher et al., 2015; van der Oost et al., 2017; Leusch et al., 2017). These trigger values have
568 been developed to assess whether the detected effect in a particular bioassay is at an acceptable
569 or a safe level (Leusch et al., 2017).
570
571 The available effect based trigger values in literature for the ER-CALUX® assay vary between
572 0.2-2.0 ng/L EEQ depending on the sample type (effluent/surface/potable water) and exposure
573 duration (Jarošová et al., 2014; Escher et al., 2015; Leusch et al., 2017). The EEQ values in the
574 present study were higher than the lowest calculated trigger value (0.2 ng/L EEQ) in all of the
575 effluent samples suggesting that the effluents may pose a risk to the receiving waters. The EEQ
576 values for samples from WWTP 2, 4 and 5 exceeded also the highest trigger value calculated for
577 the ER-CALUX® assay.
578
579 EBT values for nonspecific toxicity are determined by using a different approach. Van der Oos
580 et al. (2017) derived EBT values for nonspecific toxicity based on the assumption that acute
581 toxicity in a concentrated sample is an indication of chronic effects in the original sample. They
582 determined that for nonspecific toxicity effects measured below a REF 20 are considered
583 indicative of chronic effects, whereas REFs above 20 translate to a lower risk. In the present
584 study, significant lethal and sublethal effects (> 20 % of embryos with lethal and sublethal
585 effects) were detected at REF2.5 suggesting that chronic effects would likely be seen in the
586 original sample.
587

588 Genotoxicity bioassays are typically not easily quantifiable, therefore calculating biological
589 equivalent values is difficult (van der Oos et al., 2017). In addition, current guidelines for
590 genotoxic substances assume that there is no safe level, even though the likelihood of adverse
591 effects decreases at lower exposure levels. Considering the theoretical risk which is always
592 present with genotoxic substances, van der Oos et al. (2017) proposed an EBT value of 0.005
593 genotoxic units, which means genotoxic effects observed at REF200. In the present study,
594 genotoxic effects were not observed in the effluents samples with the p53-CALUX assay,
595 although it was impossible to test the samples at REF200 due to cytotoxicity. However,
596 genotoxicity was observed at REF40 and REF20 with the umuC-assay, suggesting that based on
597 the EBT value some risks persist after the treatment process.
598
599 Comparing the results from the present study to the EBT values available in the literature
600 suggests that the remaining toxicities after wastewater treatment are at a level, which is not
601 considered acceptable in terms of risks.

602 **4. Conclusions**

603 This study demonstrates the successful application of an effect-based approach to assess water
604 quality and toxicity removal at seven WWTPs. The analysis of the biological effects of influent
605 and effluent samples revealed that within the investigated endpoints the key effects were
606 estrogenic activity, thyroid disruption and fish embryo toxicity. These toxicities remained in the
607 effluent after wastewater treatment process in nearly all of the sampled WWTPs. Comparison of
608 results to published EBT values suggests that receiving waters may be at risk. Assays for
609 genotoxicity, androgenic activity and cytotoxicity revealed the high toxic potency of influent

610 samples, but were not responsive in the less polluted effluent samples indicating that these
611 toxicities were efficiently removed during the conventional treatment process. Interestingly, the
612 toxicity removal efficiency of the WWTPs did not show dependency between the operational
613 parameters or WWTP characteristics, but rather showed similar patterns for each toxicological
614 endpoints. These findings suggest that the toxicological effect pattern or composition of
615 municipal wastewaters is very similar within the sampled WWTPs and that the chemicals
616 causing the observed effects are not completely removed by activated sludge processes
617 regardless of the WWTP characteristics. The results of the present study are based on one
618 sampling event, thus further research is needed to draw further conclusions. For future
619 perspectives, it can be concluded that in order to reduce the toxic potency of effluents and the
620 risks to the receiving environments more advanced treatment methods should be applied. In
621 addition, concerning future challenges in monitoring water quality, effect-based tools are clearly
622 required to analyze the net effects of environmental samples.

623 **5. Acknowledgements**

624 The authors would like to thank Maj ja Tor Nessling foundation, Maa- ja vesitekniikan tuki ry.,
625 Ekokem and the Finnish Ministry of the Environment for funding this research. Additionally, the
626 authors thank FP7-SOLUTIONS for funding the tests conducted by WatchFrog. We thank all of
627 the WWTPs for participating in the sampling. The following are acknowledged for their advice
628 and guidance during the project: Eija Schultz from SYKE, Nicole Schweiger and Martin Krauss
629 from UFZ.

630

6. References

Al-Saleh, I., Elkhatib, R., Al-Rajoudi, T., & Al-Qudaihi, G. (2017). Assessing the concentration of phthalate esters (PAEs) and bisphenol A (BPA) and the genotoxic potential of treated wastewater (final effluent) in Saudi Arabia. *Science of The Total Environment*, *578*, 440-451.

Altmann, J., Ruhl, A. S., Zietzschmann, F., & Jekel, M. (2014). Direct comparison of ozonation and adsorption onto powdered activated carbon for micropollutant removal in advanced wastewater treatment. *Water Research*, *55*, 185-193.

Arome, D., & Chinedu, E. (2013). The importance of toxicity testing. *J. Pharm. Bio Sci*, *4*, 146-148.

Bain, P. A., Williams, M., & Kumar, A. (2014). Assessment of multiple hormonal activities in wastewater at different stages of treatment. *Environmental Toxicology and Chemistry*, *33*(10), 2297-2307.

Boas, M., Feldt-Rasmussen, U., & Main, K. M. (2012). Thyroid effects of endocrine disrupting chemicals. *Molecular and cellular endocrinology*, *355*(2), 240-248.

Brack, W., Ait-Aissa, S., Burgess, R. M., Busch, W., Creusot, N., Di Paolo, C., . . . Krauss, M. (2016). Effect-directed analysis supporting monitoring of aquatic environments — an in-depth overview. *Science of the Total Environment*, *544*, 1073-1118.

Castillo, M., Alonso, M. C., Riu, J., Reinke, M., Klöter, G., Dizer, H., ... & Barceló, D. (2001). Identification of cytotoxic compounds in European wastewaters during a field experiment. *Analytica Chimica Acta*, *426*(2), 265-277.

Castillo, L., Seriki, K., Mateos, S., Loire, N., Guédon, N., Lemkine, G. F., ... & Tindall, A. J. (2013). In vivo endocrine disruption assessment of wastewater treatment plant effluents with small organisms. *Water Science and Technology*, *68*(1), 261-268.

Crofton, K. M., Paul, K. B., DeVito, M. J., & Hedge, J. M. (2007). Short-term in vivo exposure to the water contaminant triclosan: evidence for disruption of thyroxine. *Environmental Toxicology and Pharmacology*, *24*(2), 194-197.

Embry, M. R., Belanger, S. E., Braunbeck, T. A., Galay-Burgos, M., Halder, M., Hinton, D. E., ... & Whale, G. (2010). The fish embryo toxicity test as an animal alternative method in hazard and risk assessment and scientific research. *Aquatic Toxicology*, *97*(2), 79-87.

Escher, B.I., Allinson, M., Altenburger, R., Bain, P.A., Balaguer, P., Busch, W., Crago, J., Denslow, N.D., Dopp, E. and Hilscherova, K. (2013) Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays. Environmental Science & Technology 48(3), 1940-1956.

665 Escher, B. I., van Daele, C., Dutt, M., Tang, J. Y., & Altenburger, R. (2013). Most oxidative
666 stress response in water samples comes from unknown chemicals: the need for effect-based
667 water quality trigger values. *Environmental science & technology*, *47*(13), 7002-7011.

668 Escher, B. I., Neale, P. A., & Leusch, F. D. (2015). Effect-based trigger values for in vitro
669 bioassays: Reading across from existing water quality guideline values. *Water research*, *81*,
670 137-148.

671 Falås, P., Wick, A., Castronovo, S., Habermacher, J., Ternes, T. A., & Joss, A. (2016). Tracing
672 the limits of organic micropollutant removal in biological wastewater treatment. *Water*
673 *research*, *95*, 240-249.

674 Fini, J. B., Le Mével, S., Turque, N., Palmier, K., Zalko, D., Cravedi, J. P., & Demeneix, B. A.
675 (2007). An in vivo multiwell-based fluorescent screen for monitoring vertebrate thyroid
676 hormone disruption. *Environmental science & technology*, *41*(16), 5908-5914.

677 Galus, M., Jeyaranjaan, J., Smith, E., Li, H., Metcalfe, C., & Wilson, J. Y. (2013). Chronic
678 effects of exposure to a pharmaceutical mixture and municipal wastewater in zebrafish.
679 *Aquatic Toxicology*, *132–133*, 212-222.

680 Goudreau, S. E., Neves, R. J., & Sheehan, R. J. (1993). Effects of wastewater treatment plant
681 effluents on freshwater mollusks in the upper Clinch River, Virginia, USA. *Hydrobiologia*,
682 *252*(3), 211-230.

683 Hallare, A., Nagel, K., Köhler, H. R., & Triebskorn, R. (2006). Comparative embryotoxicity and
684 proteotoxicity of three carrier solvents to zebrafish (Danio rerio) embryos. *Ecotoxicology*
685 *and environmental safety*, *63*(3), 378-388.

686 Halling-Sørensen, B., Nielsen, S. N., Lanzky, P. F., Ingerslev, F., Lützhøft, H. H., & Jørgensen,
687 S. E. (1998). Occurrence, fate and effects of pharmaceutical substances in the environment-
688 A review. *Chemosphere*, *36*(2), 357-393.

689 Huhtala, S., Munne, P., Nakari, T., Nuutinen, J., Perkola, N., Sainio, P., Schultz, E., & Schultz,
690 L. (2011). WP3 Innovative Approaches to Chemical Controls of Hazardous Substances.
691 *National Report Finland. COHIBA–project (Control of Hazardous Substances in the Baltic*
692 *Sea Region)*.

693 Ihara, M., Ihara, M. O., Kumar, V., Narumiya, M., Hanamoto, S., Nakada, N., ... & Tanaka, H.
694 (2014). Co-occurrence of estrogenic and antiestrogenic activities in wastewater: quantitative
695 evaluation of balance by in vitro ERα reporter gene assay and chemical analysis.
696 *Environmental science & technology*, *48*(11), 6366-6373.

697 ISO. Water quality—Determination of the genotoxicity of water and waste water using the umu-
698 test. ISO 13829:2000.

699 Ivashechkin, P., Corvini, P. X., & Dohmann, M. (2004). Behaviour of endocrine disrupting
700 chemicals during the treatment of municipal sewage sludge. *Water Science and Technology*,
701 *50*(5), 133-140.

702 Jarošová, B., Bláha, L., Giesy, J. P., & Hilscherová, K. (2014). What level of estrogenic activity
703 determined by in vitro assays in municipal waste waters can be considered as safe?.
704 *Environment international*, *64*, 98-109.

705 Jia, A., Escher, B. I., Leusch, F. D. L., Tang, J. Y. M., Prochazka, E., Dong, B., . . . Snyder, S. A.
706 (2015). In vitro bioassays to evaluate complex chemical mixtures in recycled water. *Water*
707 *Research*, *80*, 1-11.

708 Johnson, A. C., Belfroid, A., & Di Corcia, A. (2000). Estimating steroid oestrogen inputs into
709 activated sludge treatment works and observations on their removal from the effluent.
710 *Science of the Total Environment*, *256*(2–3), 163-173.

711 Joss, A., Keller, E., Alder, A. C., Göbel, A., McArdell, C. S., Ternes, T., & Siegrist, H. (2005).
712 Removal of pharmaceuticals and fragrances in biological wastewater treatment. *Water*
713 *research*, *39*(14), 3139-3152.

714 Jugan, M. L., Oziol, L., Bimbot, M., Huteau, V., Tamisier-Karolak, S., Blondeau, J. P., & Levi,
715 Y. (2009). In vitro assessment of thyroid and estrogenic endocrine disruptors in wastewater
716 treatment plants, rivers and drinking water supplies in the greater Paris area (France).
717 *Science of the Total Environment*, *407*(11), 3579-3587.

718 Kirk, L. A., Tyler, C. R., Lye, C. M., & Sumpter, J. P. (2002). Changes in estrogenic and
719 androgenic activities at different stages of treatment in wastewater treatment works.
720 *Environmental Toxicology and Chemistry*, *21*(5), 972-979.

721 Koh, Y. K. K., Chiu, T. Y., Boobis, A., Cartmell, E., Scrimshaw, M. D., & Lester, J. N. (2008).
722 Treatment and removal strategies for estrogens from wastewater. *Environmental*
723 *Technology*, *29*(3), 245-267.

724 Kolpin, D. W., Furlong, E. T., Meyer, M. T., Thurman, E. M., Zaugg, S. D., Barber, L. B., &
725 Buxton, H. T. (2002). Pharmaceuticals, hormones, and other organic wastewater
726 contaminants in US streams, 1999-2000: A national reconnaissance. *Environmental science*
727 *& technology*, *36*(6), 1202-1211.

728 Kruglova, A., Ahlgren, P., Korhonen, N., Rantanen, P., Mikola, A., & Vahala, R. (2014).
729 Biodegradation of ibuprofen, diclofenac and carbamazepine in nitrifying activated sludge
730 under 12 C temperature conditions. *Science of the Total Environment*, *499*, 394-401.

731 Kruglova, A., Kråkström, M., Riska, M., Mikola, A., Rantanen, P., Vahala, R., & Kronberg, L.
732 (2016). Comparative study of emerging micropollutants removal by aerobic activated sludge
733 of large laboratory-scale membrane bioreactors and sequencing batch reactors under low-
734 temperature conditions. *Bioresource technology*, *214*, 81-88.

König, M., Escher, B. I., Neale, P. A., Krauss, M., Hilscherová, K., Novák, J., ... & Ahlheim, J. (2017). Impact of untreated wastewater on a major European river evaluated with a combination of in vitro bioassays and chemical analysis. *Environmental Pollution*, *220*, 1220-1230.

Lee, L. E., Clemons, J. H., Bechtel, D. G., Caldwell, S. J., Han, K., Pasitschniak-Arts, M., . . . Bols, N. C. (1993). Development and characterization of a rainbow trout liver cell line expressing cytochrome P450-dependent monooxygenase activity. *Cell Biology and Toxicology*, *9*(3), 279-294.

Leusch, F. D., Chapman, H. F., Körner, W., Gooneratne, S. R., & Tremblay, L. A. (2005). Efficacy of an advanced sewage treatment plant in southeast Queensland, Australia, to remove estrogenic chemicals. *Environmental science & technology*, *39*(15), 5781-5786.

Leusch, F. D., Chapman, H. F., van den Heuvel, M. R., Tan, B. L., Gooneratne, S. R., & Tremblay, L. A. (2006). Bioassay-derived androgenic and estrogenic activity in municipal sewage in Australia and New Zealand. *Ecotoxicology and Environmental Safety*, *65*(3), 403-411.

Leusch, F. D., Khan, S. J., Gagnon, M. M., Quayle, P., Trinh, T., Coleman, H., ... & Reitsema, T. (2014). Assessment of wastewater and recycled water quality: a comparison of lines of evidence from in vitro, in vivo and chemical analyses. *Water research*, *50*, 420-431.

Leusch, F. D., Neale, P. A., Hebert, A., Scheurer, M., & Schriks, M. C. (2017). Analysis of the sensitivity of in vitro bioassays for androgenic, progestagenic, glucocorticoid, thyroid and estrogenic activity: Suitability for drinking and environmental waters. *Environment international*, *99*, 120-130.

Luo, Y., Guo, W., Ngo, H. H., Nghiem, L. D., Hai, F. I., Zhang, J., . . . Wang, X. C. (2014). A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Science of the Total Environment, 473–474*, 619-641.

Ma, M., Li, J., & Wang, Z. (2005). Assessing the detoxication efficiencies of wastewater treatment processes using a battery of bioassays/biomarkers. *Archives of environmental contamination and toxicology*, *49*(4), 480-487.

Macova, M., Escher, B. I., Reungoat, J., Carswell, S., Chue, K. L., Keller, J., & Mueller, J. F. (2010). Monitoring the biological activity of micropollutants during advanced wastewater treatment with ozonation and activated carbon filtration. *Water Research, 44*(2), 477-492.

Macova, M., Toze, S., Hodgers, L., Mueller, J. F., Bartkow, M., & Escher, B. I. (2011). Bioanalytical tools for the evaluation of organic micropollutants during sewage treatment, water recycling and drinking water generation. *Water research*, *45*(14), 4238-4247.

Magdeburg, A., Stalter, D., Schlüsener, M., Ternes, T., & Oehlmann, J. (2014). Evaluating the efficiency of advanced wastewater treatment: target analysis of organic contaminants and (geno-) toxicity assessment tell a different story. *Water research*, *50*, 35-47.

Malaj, E., von der Ohe, P. C., Grote, M., Kühne, R., Mondy, C. P., Usseglio-Polatera, P., . . . Schäfer, R. B. (2014). Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. *Proceedings of the National Academy of Sciences*, *111*(26), 9549-9554.

Maletz, S., Floehr, T., Beier, S., Klümper, C., Brouwer, A., Behnisch, P., . . . Hollert, H. (2013). In vitro characterization of the effectiveness of enhanced sewage treatment processes to eliminate endocrine activity of hospital effluents. *Water Research*, *47*(4), 1545-1557.

Margot, J., Kienle, C., Magnet, A., Weil, M., Rossi, L., De Alencastro, L. F., ... & Barry, D. A. (2013). Treatment of micropollutants in municipal wastewater: ozone or powdered activated carbon?. *Science of the total environment*, *461*, 480-498.

Mathon, B., Coquery, M., Miege, C., Penru, Y., & Choubert, J. M. (2017). Removal efficiencies and kinetic rate constants of xenobiotics by ozonation in tertiary treatment. *Water Science and Technology*, wst2017114.

Murk, A. J., Legler, J., van Lipzig, M. M. H., Meerman, J. H. N., Belfroid, A. C., Spenkelink, A., . . . Vethaak, D. (2002). Detection of estrogenic potency in wastewater and surface water with three in vitro bioassays. *Environmental Toxicology and Chemistry*, *21*(1), 16-23.

Nakada, N., Shinohara, H., Murata, A., Kiri, K., Managaki, S., Sato, N. and Takada, H. (2007) Removal of selected pharmaceuticals and personal care products (PPCPs) and endocrine-disrupting chemicals (EDCs) during sand filtration and ozonation at a municipal sewage treatment plant. Water Research 41(19), 4373-4382.

Neale, P. A., & Leusch, F. D. (2015). Considerations when assessing antagonism in vitro: Why standardizing the agonist concentration matters. *Chemosphere*, *135*, 20-23.

OECD, 2013. Guideline for Testing of Chemicals, 236. Fish Embryo Acute Toxicity (FET) Test. OECD, Paris, France. Available at: <http://www.oecd.org>.

Okuda, T., Kobayashi, Y., Nagao, R., Yamashita, N., Tanaka, H., Tanaka, S., ... & Houwa, I. (2008). Removal efficiency of 66 pharmaceuticals during wastewater treatment process in Japan. *Water Science and Technology*, *57*(1), 65-71.

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

Reungoat, J., Escher, B. I., Macova, M., & Keller, J. (2011). Biofiltration of wastewater treatment plant effluent: effective removal of pharmaceuticals and personal care products and reduction of toxicity. *Water research, 45*(9), 2751-2762.

Scholz, S., Fischer, S., Gündel, U., Küster, E., Luckenbach, T., & Voelker, D. (2008). The zebrafish embryo model in environmental risk assessment—applications beyond acute toxicity testing. *Environmental Science and Pollution Research, 15*(5), 394-404.

Schulze, T., Ahel, M., Ahlheim, J., Aït-Aïssa, S., Brion, F., Di Paolo, C., ... & Hu, M. (2017). Assessment of a novel device for onsite integrative large-volume solid phase extraction of water samples to enable a comprehensive chemical and effect-based analysis. *Science of The Total Environment*.

Smital, T., Terzic, S., Zaja, R., Senta, I., Pivcevic, B., Popovic, M., . . . Ahel, M. (2011). Assessment of toxicological profiles of the municipal wastewater effluents using chemical analyses and bioassays. *Ecotoxicology and Environmental Safety, 74*(4), 844-851.

Spirhanzlova, P., Leleu, M., Sébillot, A., Lemkine, G.F., Iguchi, T., Demeneix, B.A., and Tindall, A.J. (2016). Oestrogen reporter transgenic medaka for non-invasive evaluation of aromatase activity. Comparative Biochemistry and Physiology, Part C, 179, 64–71.

Stalter, D., Magdeburg, A., Wagner, M., & Oehlmann, J. (2011). Ozonation and activated carbon treatment of sewage effluents: Removal of endocrine activity and cytotoxicity. *Water Research, 45*(3), 1015-1024.

Stasinakis, A. S., Thomaidis, N. S., Arvaniti, O. S., Asimakopoulos, A. G., Samaras, V. G., Ajibola, A., . . . Lekkas, T. D. (2013). Contribution of primary and secondary treatment on the removal of benzothiazoles, benzotriazoles, endocrine disruptors, pharmaceuticals and perfluorinated compounds in a sewage treatment plant. *Science of the Total Environment, 463–464*, 1067-1075.

Svenson, A., & Allard, A. S. (2004). Occurrence and some properties of the androgenic activity in municipal sewage effluents. *Journal of Environmental Science and Health, Part A, 39*(3), 693-701.

Tanneberger, K., Rico-Rico, A., Kramer, N. I., Busser, F. J., Hermens, J. L., & Schirmer, K. (2010). Effects of solvents and dosing procedure on chemical toxicity in cell-based in vitro assays. *Environmental science & technology, 44*(12), 4775-4781.

Ternes, T. A., Stumpf, M., Mueller, J., Haberer, K., Wilken, R. -., & Servos, M. (1999). Behavior and occurrence of estrogens in municipal sewage treatment plants — I. investigations in germany, canada and brazil. *Science of the Total Environment, 225*(1–2), 81-90.

Vajda, A. M., Barber, L. B., Gray, J. L., Lopez, E. M., Woodling, J. D., & Norris, D. O. (2008). Reproductive disruption in fish downstream from an estrogenic wastewater effluent. *Environmental science & technology*, *42*(9), 3407-3414.

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. *Environmental science & technology*, *42*(15), 5814-5820.

van der Linden, S. C., von Bergh, A. R. M., van Vught-Lussenburg, B. M. A., Jonker, L. R. A., Teunis, M., Krul, C. A. M., & van der Burg, B. (2014). Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, *760*, 23-32.

van der Oost, R., Sileno, G., Suárez-Muñoz, M., Nguyen, M. T., Besselink, H. and Brouwer, A. (2017), 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. doi:10.1002/etc.3836

Veldhoen, N., Skirrow, R. C., Osachoff, H., Wigmore, H., Clapson, D. J., Gunderson, M. P., ... & Helbing, C. C. (2006). The bactericidal agent triclosan modulates thyroid hormone-associated gene expression and disrupts postembryonic anuran development. *Aquatic Toxicology*, *80*(3), 217-227.

Vieno, N., 2014. Haitalliset Aineet Jätevedenpuhdistamoilla - Hankkeen Loppuraportti. Vesilaitosyhdistys, Helsinki, p. 273.

Vieno, N., & Sillanpää, M. (2014). Fate of diclofenac in municipal wastewater treatment plant—a review. *Environment international*, *69*, 28-39.

Välitalo, P., Perkola, N., Seiler, T., Sillanpää, M., Kuckelkorn, J., Mikola, A., . . . Schultz, E. (2016). Estrogenic activity in finnish municipal wastewater effluents. *Water Research*, *88*, 740-749.

Weber, S., Leuschner, P., Kämpfer, P., Dott, W., & Hollender, J. (2005). Degradation of estradiol and ethinyl estradiol by activated sludge and by a defined mixed culture. *Applied Microbiology and Biotechnology*, *67*(1), 106-112.

Wegner, S., Browne, P., & Dix, D. (2016). Identifying reference chemicals for thyroid bioactivity screening. *Reproductive Toxicology*, *65*, 402-413.

Wigh, A., Devaux, A., Brosselin, V., Gonzalez-Ospina, A., Domenjoud, B., Aït-Aïssa, S., ... & Bony, S. (2016). Proposal to optimize ecotoxicological evaluation of wastewater treated by conventional biological and ozonation processes. *Environmental Science and Pollution Research*, *23*(4), 3008-3017.

Xu, N., Xu, Y. -., Xu, S., Li, J., & Tao, H. -. (2012). Removal of estrogens in municipal wastewater treatment plants: A chinese perspective. *Environmental Pollution, 165*, 215-224.

Ying, G., Kookana, R. S., & Kumar, A. (2008). Fate of estrogens and xenoestrogens in four sewage treatment plants with different technologies. *Environmental Toxicology and Chemistry*, *27*(1), 87-94.

Žegura, B., Heath, E., Černoša, A., & Filipič, M. (2009). Combination of in vitro bioassays for the determination of cytotoxic and genotoxic potential of wastewater, surface water and drinking water samples. *Chemosphere*, *75*(11), 1453-1460.

Zeng, Q., Li, Y., & Yang, S. (2013). Sludge retention time as a suitable operational parameter to remove both estrogen and nutrients in an anaerobic–anoxic–aerobic activated sludge system. *Environmental engineering science*, *30*(4), 161-169.

Zha, J., & Wang, Z. (2006). Acute and early life stage toxicity of industrial effluent on japanese medaka (oryzias latipes). *Science of the Total Environment, 357*(1–3), 112-119.

Zuehlke, S., Duennbier, U., Lesjean, B., Gnirss, R., & Buisson, H. (2006). Long-term comparison of trace organics removal performances between conventional and membrane activated sludge processes. *Water Environment Research, 78*(13), 2480-2486.

- Toxicity removal during wastewater treatment at 7 municipal WWTPs was investigated
- Removal efficiency was assessed by an effect-based approach comprising of multiple endpoints
- Large volumes of influent and effluent samples were extracted with a novel device
- Embryo toxicity, estrogenic activity and thyroid disruption were detected in effluent samples
- The results showed a similar effect pattern among all the WWTPs