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


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
A comparison of toxicity tests for influent and effluent from different WWTPs. The tests include:

- Cytotoxicity
- Genotoxicity (p53)
- Genotoxicity (umu)
- Estrogenicity (ER-CALUX)
- Androgenicity (AR-CALUX)
- Thyroid disruption
- Embryotoxicity
- Estrogenicity (medaka)
- Androgenicity (AR-CALUX)

The figure shows the relative levels of toxicity for each test across different WWTPs, with WWTP1, WWTP2, WWTP3, WWTP4, WWTP5, WWTP6, WWTP6 MBR, and WWTP7 represented by different lines and colors.
Effect-based assessment of toxicity removal during wastewater treatment

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

\textsuperscript{a}. Finnish Environment Institute, Laboratory Centre, Hakuninmaantie 6, 00430 Helsinki, Finland
\textsuperscript{b}. Aalto University, Department of Civil and Environmental Engineering, Tietotie 1E, 02150 Espoo, Finland
\textsuperscript{c}. UFZ - Helmholtz Centre for Environmental Research GmbH, Leipzig, Germany
\textsuperscript{d}. Institute for Environmental Research (Biology V), RWTH Aachen University, Aachen, Germany.
\textsuperscript{e}. BioDetection Systems b.v., Amsterdam, The Netherlands
\textsuperscript{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; Väitalo 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 (LV SPE50), 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
(50 – 1000 L) of surface waters (Brack et al., 2016; Schultze et al., 2017). The principles of the device and the approach are introduced in Schulze et al. (2017), however some modifications were made to optimize the extraction process for wastewater samples. In short, a large SPE cartridge packed with 10 g of HR-X sorbent material (Macherey Nagel Chromabond® HR-X) between two glass filter plates (THOMAPOR® 50 mm) was applied. The samples were pre-filtered prior to extraction with Sartopure GF+ Midicap filters, therefore particle bound contaminants are not considered in the present work. The sorbent material was conditioned with 200 mL of ethyl acetate, 200 mL of methanol and 100 mL of deionized water. The maximal volume of each sample was extracted with the device, which depended on the rate at which the filters became clogged. The samples were extracted sequentially in portions of 500 mL. The extracted sample volumes are presented in Table 2.

After each sample extraction with the LVSPE50, the cartridge was dried overnight under a nitrogen stream. After drying, the compounds of interest were eluted from the sorbent material with a sequential elution scheme into four different fractions (100 mL ethyl acetate, 100 ml methanol, 100 mL methanol with 1% formic acid, 100 mL methanol with 2% 7N-ammonia in methanol).

After elution, the acidic and basic fractions were neutralized to pH 7 ± 0.5 and all of the fractions were filtered through filter paper (Whatman GF/F) to remove any residual interfering particles or salts. Each fraction was evaporated to dryness with rotary evaporation and an EZ-Envi centrifugal evaporator (Genevac Ltd, Ipswich, UK) and then re-dissolved in MeOH resulting in a final concentration factor of 5000x. These eluates were stored in the freezer (-20°C) prior to analysis.
### Table 1. Information on the seven WWTPs in Finland selected for sampling of effluent.

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

<table>
<thead>
<tr>
<th>Location</th>
<th>Influent (extracted volume, L)</th>
<th>Effluent (extracted volume, L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WWTP 1</td>
<td>13</td>
<td>41</td>
</tr>
<tr>
<td>WWTP 2</td>
<td>15</td>
<td>39.5</td>
</tr>
<tr>
<td>WWTP 3</td>
<td>19.5</td>
<td>43</td>
</tr>
<tr>
<td>WWTP 4</td>
<td>22.5</td>
<td>40</td>
</tr>
<tr>
<td>WWTP 5</td>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>WWTP 6</td>
<td>6</td>
<td>13 (2.5 MBR pilot)</td>
</tr>
<tr>
<td>WWTP 7</td>
<td>16</td>
<td>36</td>
</tr>
</tbody>
</table>

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).

\[
REF = \text{dilution factor}_{\text{bioassay}} \times \text{enrichment factor}_{\text{SPE}}
\]  

(Eq. 1)
2.3 Biological analysis

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

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

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

Castillo et al. (2013) studied thyroid disruption of wastewaters with the same transgenic *Xenopus laevis* embryos as used in the present study and observed thyroid disruption in most influent samples and some effluent samples. However, the thyroid disruption potential of the effluent samples was clearly lower than the untreated wastewaters. It is well recognized that some brominated flame-retardants and the antimicrobial compound triclosan can cause thyroid-disrupting effects (Veldhoen et al., 2006; Crofton et al., 2007; Boas et al., 2012), and these chemicals have been frequently detected in Finnish wastewater samples (Huhtala et al., 2011; Vieno, 2014). In the next phase of this project, the same WWTP samples will be analyzed to determine the concentrations of various organic contaminants, and perhaps the results from the chemical analysis will provide more insight on the effluent thyroid disruption activity (data not yet published).
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\(^\circledR\) 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, Zhan 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 abnormalities 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 (REF 10, 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
influent samples with positive results detected at REF1. The ER-CALUX® assay for estrogenic activity was the most responsive cell-based assay, as in all of the samples an estrogenic response was detected at low exposure concentrations. The effluent samples induced embryotoxicity in the FET assay in all of the samples at REF2.5, suggesting that the toxic effects may be caused by chemicals that are typically present in all municipal wastewaters. Additionally, positive responses were detected in the transgenic larval models for endocrine disruption at low sample concentrations in several effluent samples.

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

The efficiency of the WWTPs to reduce toxicity was calculated by comparing the toxicity of the influent and effluent samples. There was no correlation between toxicity removal efficiencies and the presence or absence of a tertiary treatment step at the WWTP. However, it should be noted that these treatment steps were originally designed for the removal of phosphorus and suspended solids rather than micro-pollutants. Additionally, other WWTP parameters, such as sludge age or nitrification rate, did not correlate with toxicity removal efficiency either. To our knowledge, there are no previous studies on the correlation between sludge age or sludge retention time (SRT) and general toxicity removal. Previous studies have been focused on individual chemical compounds, such as pharmaceuticals or hormones. Higher biodegradation rates for micro-pollutants have been observed in processes with higher sludge retention times or sludge age, such as membrane bioreactors (Vieno and Sillanpää, 2014; Kruglova et al., 2016).

The benefits related to increased degradation rates of micro-pollutants with higher sludge age seems to reach an optimal level at approximately 20 – 25 days (Zeng et al., 2013; Fälas et al., 2016). However, in the present study clear differences in toxicity removal between WWTPs with the lowest sludge age (9 days at WWTP 4) or the highest sludge age (30 days at WWTP 5) were not seen. It is possible that other operational parameters and factors (e.g. temperature, organic loading rates) could greatly affect the removal efficiency (Kruglova et al., 2014), and thus further research is needed in order to draw distinct conclusions. In agreement with the present study, some previous investigations have failed to find a correlation between operational parameters and pollutant removal. For example, Joss et al. (2005) observed no clear dependency between 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).
The findings of the present study suggest that the removal efficiency was more related to each toxicological endpoint than characteristics of the WWTPs. The toxicity removal related to all of the selected endpoints is summarized as radar charts in Figure 6. In these charts, the toxicity of each sample is evaluated by a simple scoring system, where a value between 1 and 7 is given, 7 indicating higher toxicity and 1 no toxicity. The score bands for each bioassay are presented in SI, Table 4. The toxicity pattern of the influent samples was similar between the WWTPs, as almost all of the samples induced toxic effects in the majority of the bioassays. Furthermore, there was some variation depending on the endpoint since the influent of certain WWTPs was clearly more androgenic, cytotoxic, genotoxic or estrogenic than the others. Based on the overall results, influent samples from WWTP3 and WWTP7 were the least toxic and influent samples from WWTP4, WWTP5 and WWTP6 were the most toxic. The radar diagrams for effluent toxicity clearly show that the remaining toxicities after treatment are embryo toxicity, estrogenic activity, thyroid disruption and genotoxicity (umuC).

When looking at the different toxicological endpoints in more detail, some variation in removal efficiency between the WWTPs can be observed. The cytotoxicity was substantially or completely reduced during the wastewater treatment process in all of the WWTPs, except WWTP 7 and the MBR pilot plant at WWTP 6, where there was no significant reduction in cytotoxic effects. In general cytotoxicity reduction was high (76 – 89 %) and in the cases of WWTP 5 and WWTP 6 the toxic effects were completely removed in the highest sample concentration tested.
Figure 6. The toxicity of the influent (a) and effluent (b) samples related to all of the selected endpoints. The toxicity of each sample is scored by giving a value between 1 and 7, 7 indicating higher toxicity and 1 representing no toxicity. The results from WWTP 6 influent sample follow the same line as for WWT6 MBR.

The results of this study suggest that the activated sludge process is the most effective treatment step at removing cytotoxicity from the studied WWTPs. Toxicity removal of cytotoxic effects was at the lowest level at WWTP 7, which employs the MBBR + DAF process. WWTP 7 had the highest industrial loading, which could partly explain the outcome as industrial influent may contain more compounds that are less biodegradable compared to typical municipal wastewaters. At WWTP 5 where the MBBR process is combined with activated sludge, the toxic effects were completely removed. Having sand filtration as a tertiary treatment step did not improve the removal of cytotoxic effects, since there was no clear correlation between better removal efficiency and sand filtration. Previous studies have also shown that influents are typically highly cytotoxic and that toxicity is significantly reduced during conventional wastewater treatment 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.
However, the results based on the umuC-assay suggest that the genotoxic effects are not reduced during wastewater treatment, although effects are only detected at high sample concentrations. The results from previous studies have also presented varying results. Al-Saleh et al. (2017) showed that effluents still had high genotoxic potential after wastewater treatment process. Additionally, genotoxic potential of wastewater effluents was demonstrated in a study by Escher et al. (2014) and Jolibois and Guerbet (2005) with several assays based on reactive modes of action (e.g. umuC assay and SOS chromotest). Magdeburg et al. (2014) demonstrated significant genotoxic effects in samples taken after secondary sedimentation, which were effectively reduced by an ozonation process but were not further reduced by sand filtration following the ozonation step. In a study by Žegura et al. (2009) genotoxic effects were not observed in influent samples but were detected in some of the corresponding effluent samples, which may be due to the formation of genotoxic compounds during the biological treatment of wastewaters.

The reduction in estrogenic activity was between 78 – 97 % due to the water treatment in majority of the WWTPs. In WWTP 3 and 7 estrogenic activity was not removed at all, however in those samples the estrogenic potential of the influent was low to begin with. The removal efficiency of the MBR pilot in WWTP 6 was lower than the efficiency of the normal treatment process (normal 97 %, MBR 88 %). A number of studies have investigated the removal of steroid hormones using membrane bioreactors. Some of the studies have shown that MBR removes estrogens more efficiently than conventional activated sludge process (Zuehlke et al., 2006; Maletz et al., 2013). On the other hand, Ivashechkin et al. (2004) and Weber et al. (2005) did not find an appreciable difference in removal of estrogens between membrane activated sludge or conventional activated sludge systems. In the present study, any conclusions on the
removal efficiency of the MBR process compared to conventional activated sludge are difficult
to draw, because the MBR system was a newly installed pilot and the operational parameters
might not have been fully optimized as indicated by the other results from this study. In addition,
it should be acknowledged that the present study is based on one sampling event and the results
can vary depending on the time, temperature and other varying parameters. The majority of the
previous studies focusing on estrogens in wastewaters have used chemical analytical tools to
analyze the removal efficiencies of estrogenic compounds (Ternes et al., 1999; Johnson,
Belfroid, & Di Corcia, 2000; Ying, Kookana, & Kumar, 2008; Xu et al., 2012; Luo et al., 2014;).
Some previous studies have employed the ERα-CALUX® to study removal efficiencies of
estrogenic activity during wastewater treatment processes (Murk et al., 2002; Maletz et al., 2013;
Bain et al., 2014) and their findings support the results of the present study. Murk et al. (2002)
showed substantial reductions (90 – 95 %) of estrogenic potency in effluents compared to
influents in municipal WWTPs. Similar results were shown in a more recent study with
reductions between 89 – 100 % for estrogenic activity in three Australian WWTPs (Bain et al.,
2014). Their results suggest that tertiary treatment processes (flocculation, tertiary clarification,
dual-media filtration, chlorine disinfections and dechlorination) contribute markedly to the
enhanced reduction of estrogenic potency following conventional treatment.

The estimation of the risks posed by the treated effluents to the receiving waters is challenging as
many factors, such as dilution and flow rate of the receiving water, affect the actual risk. The
chemicals causing toxic effects in wastewater effluents are typically present at low
concentrations and concentration of the samples is often necessary to observe ecotoxicological
effects in acute tests. In the receiving waters, the effluents are diluted, but the exposure is
typically constant and long-term. One approach to assess the bioassay results in terms of risk
context, is to apply effect-based trigger (EBT) values (Escher et al., 2013; Jarošová et al., 2014;
Escher et al., 2015; van der Oost et al., 2017; Leusch et al., 2017). These trigger values have
been developed to assess whether the detected effect in a particular bioassay is at an acceptable
or a safe level (Leusch et al., 2017).

The available effect based trigger values in literature for the ER-CALUX® assay vary between
0.2-2.0 ng/L EEQ depending on the sample type (effluent/surface/potable water) and exposure
duration (Jarošová et al., 2014; Escher et al., 2015; Leusch et al., 2017). The EEQ values in the
present study were higher than the lowest calculated trigger value (0.2 ng/L EEQ) in all of the
effluent samples suggesting that the effluents may pose a risk to the receiving waters. The EEQ
values for samples from WWTP 2, 4 and 5 exceeded also the highest trigger value calculated for
the ER-CALUX® assay.

EBT values for nonspecific toxicity are determined by using a different approach. Van der Oos
et al. (2017) derived EBT values for nonspecific toxicity based on the assumption that acute
toxicity in a concentrated sample is an indication of chronic effects in the original sample. They
determined that for nonspecific toxicity effects measured below a REF 20 are considered
indicative of chronic effects, whereas REFs above 20 translate to a lower risk. In the present
study, significant lethal and sublethal effects (> 20 % of embryos with lethal and sublethal
effects) were detected at REF2.5 suggesting that chronic effects would likely be seen in the
original sample.
Genotoxicity bioassays are typically not easily quantifiable, therefore calculating biological equivalent values is difficult (van der Oos et al., 2017). In addition, current guidelines for genotoxic substances assume that there is no safe level, even though the likelihood of adverse effects decreases at lower exposure levels. Considering the theoretical risk which is always present with genotoxic substances, van der Oos et al. (2017) proposed an EBT value of 0.005 genotoxic units, which means genotoxic effects observed at REF200. In the present study, genotoxic effects were not observed in the effluents samples with the p53-CALUX assay, although it was impossible to test the samples at REF200 due to cytotoxicity. However, genotoxicity was observed at REF40 and REF20 with the umuC-assay, suggesting that based on the EBT value some risks persist after the treatment process.

Comparing the results from the present study to the EBT values available in the literature suggests that the remaining toxicities after wastewater treatment are at a level, which is not considered acceptable in terms of risks.

4. Conclusions

This study demonstrates the successful application of an effect-based approach to assess water quality and toxicity removal at seven WWTPs. The analysis of the biological effects of influent and effluent samples revealed that within the investigated endpoints the key effects were estrogenic activity, thyroid disruption and fish embryo toxicity. These toxicities remained in the effluent after wastewater treatment process in nearly all of the sampled WWTPs. Comparison of results to published EBT values suggests that receiving waters may be at risk. Assays for genotoxicity, androgenic activity and cytotoxicity revealed the high toxic potency of influent
samples, but were not responsive in the less polluted effluent samples indicating that these toxicities were efficiently removed during the conventional treatment process. Interestingly, the toxicity removal efficiency of the WWTPs did not show dependency between the operational parameters or WWTP characteristics, but rather showed similar patterns for each toxicological endpoints. These findings suggest that the toxicological effect pattern or composition of municipal wastewaters is very similar within the sampled WWTPs and that the chemicals causing the observed effects are not completely removed by activated sludge processes regardless of the WWTP characteristics. The results of the present study are based on one sampling event, thus further research is needed to draw further conclusions. For future perspectives, it can be concluded that in order to reduce the toxic potency of effluents and the risks to the receiving environments more advanced treatment methods should be applied. In addition, concerning future challenges in monitoring water quality, effect-based tools are clearly required to analyze the net effects of environmental samples.

5. Acknowledgements

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


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.


• 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