Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS

Pablo Gago-Ferrero,[†] Emma L. Schymanski,[‡] Anna A. Bletsou,[†] Reza Aalizadeh,[†] Juliane Hollender,[‡,§] and Nikolaos S. Thomaidis*,[†]

[†]Laboratory of Analytical Chemistry, Department of Chemistry, University of Athens, Panepistimiopolis Zografou, 15771 Athens, Greece

[‡]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland

[§]Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, 8092, Zürich, Switzerland

Ⓢ Supporting Information

ABSTRACT: An integrated workflow based on liquid chromatography coupled to a quadrupole-time-of-flight mass spectrometer (LC-QTOF-MS) was developed and applied to detect and identify suspect and unknown contaminants in Greek wastewater. Tentative identifications were initially based on mass accuracy, isotopic pattern, plausibility of the chromatographic retention time and MS/MS spectral interpretation (comparison with spectral libraries, in silico fragmentation). Moreover, new specific strategies for the identification of metabolites were applied to obtain extra confidence including the comparison of diurnal and/or weekly concentration trends of the metabolite and parent compounds and the complementary use of HILIC. Thirteen of 284 predicted and literature metabolites of selected pharmaceuticals and nicotine were tentatively identified in influent samples from Athens and seven were finally confirmed with reference standards. Thirty four nontarget compounds were tentatively identified, four were also confirmed. The sulfonated surfactant diglycol ether sulfate was identified along with others in the homologous series ($SO_4C_2H_4(OC_2H_4)_xOH$), which have not been previously reported in wastewater. As many surfactants were originally found as nontargets, these compounds were studied in detail through retrospective analysis.
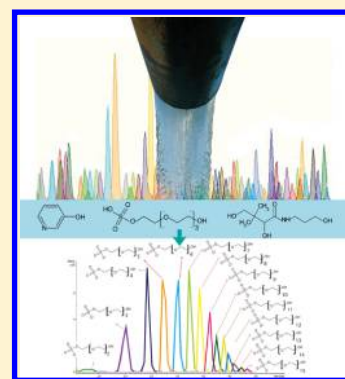
INTRODUCTION

The chemical pollutants that are internationally regulated represent a very small fraction of the universe of known and unknown anthropogenic chemicals occurring in the environment.[1] Despite this, most regulatory bodies responsible for water and wastewater treatment still assume that a few well-known compounds such as the priority pollutants of the Water Framework Directive are responsible for the most significant environmental, human health and economic risks, even though they only represent a tiny fraction of both known and yet-to-be identified chemicals.[1,2]

Waste water treatment plants (WWTPs) process numerous organic contaminants, including emerging pollutants, related transformation products (TPs)[3,4] and human metabolites. If removal is incomplete, these can be released into the aquatic environment, forming a major source of contamination. Thus, the fate of emerging pollutants and their TPs in WWTPs is of paramount environmental importance.

Existing target screening methods, where the chemicals are selected in advance, can only cover a relatively small proportion of these organic contaminants. This can result in bias (due to the preselection) and potential chemical stressors may be omitted. Therefore, one of the hottest trends in environmental analysis is high resolution mass spectrometry (HR-MS) coupled with liquid chromatography (LC) to screen samples for suspect pollutants, where some information is available but no reference standard, or even nontarget analysis, where no preselection is performed (generally called "non-target analysis").[5] For a holistic risk assessment, target-based environmental monitoring should be accompanied by nontargeted analysis.

Different comprehensive and semiautomated strategies for LC-HRMS (generally) combining target and suspect screening have been developed in the past few years to evaluate the presence of a larger number of substances without necessarily purchasing the standards for all of them. Suspect screening methods have been performed on pesticides,[6] pharmaceuticals,[7] iodinated contrast media photodegradation products[8] and transformation products[9] in surface or natural waters, as well as pesticides and pharmaceuticals in effluent wastewater.[10] Other studies considered a wide scope of suspects.[3,4,11−13] These strategies led to the detection of some TPs in the environment for the first time. Although LC-HRMS is often used to identify nontarget TPs in degradation experiments, few papers cover

"generic non-target screening" applied to the LC-HRMS/MS analysis of environmental samples. Schymanski et al.[4] and Hug et al.[3] successfully identified suspect and nontarget compounds in wastewater using an LTQ Orbitrap, including confirmation with reference standards in some cases. Van Leerman et al.[14] also identified some truly unknown compounds in river water through the use of LC-HR-MS combined with high resolution nuclear magnetic resonance spectroscopy (HR-NMR). In general, however, the identification of unknowns remains a very difficult and time-consuming task with no guarantee of success,[15] as shown by a recent collaborative trial between 17 institutions.[16] With the increasing interest in suspect and nontarget workflows, there is also a need to communicate the confidence in the identification in a way that reflects the evidence available.[17]

None of the aforementioned studies focused on the analysis of influent wastewaters (IWW). The analysis of micropollutants present in influents provides valuable information about patterns of human consumption of several substances (e.g pharmaceuticals, drugs of abuse).[18] Here, the analysis of human metabolites is of high importance. Thus, the main objectives of the present work were (i) the development of additional strategies to support identification of suspect and unknown organic compounds, (ii) the actual identification of these substances in wastewater using an integrated workflow based on liquid chromatography–quadrupole-time-of-flight mass spectrometry (LC–QToF-MS), and (iii) the use of retrospective screening in these and previous samples[4] to find additional evidence for the relevance of identified nontarget substances.

An in-house database with eight high consumption pharmaceuticals[18] (amisulpride, atenolol, clarithromycin, metformin, niflumic acid, ranitidine, tramadol, and venlafaxine), plus nicotine and their related human metabolites (284 suspects in total) was developed, based on a metabolite prediction software and pharmacokinetic literature. Apart from the usual criteria used in suspect analysis (e.g., mass accuracy or isotopic fit), new specific strategies were applied in the tentative identification in order to increase the identification confidence of these compounds. One investigated hypothesis was that both the parent compounds and their related metabolites follow similar diurnal or/and weekly concentration trends in influents. The complementary use of HILIC was also investigated as well as the comparison of the spectra obtained for a given substance in ESI(+) and ESI(−), when possible, and retention time plausibility, using an in-house developed QSRR prediction model. Nontarget peaks were selected and elucidated using in silico fragmentation prediction and their commercial importance, in addition to the strategies described above. Since a high proportion of surfactants were observed among the tentatively identified nontarget substances, a retrospective suspect screening was performed for these compounds.

MATERIALS AND METHODS

Chemicals, Sampling, and Analysis. 173 compounds substances with a wide range of physicochemical properties were selected to prepare an *artificial suspect mixture* to evaluate and validate the suspect screening approach. Details about all chemicals are given in the Supporting Information (SI, SI-1).

Influent and effluent wastewater samples (24 h composite flow-proportional samples) were collected from the Athens WWTP (Greece) on March 15, 2014 (Saturday). On the same day, 2 h flow proportional influent samples were collected every 2 h.

The WWTP of Athens is designed with primary sedimentation, activated sludge process with biological nitrogen and phosphorus removal and secondary sedimentation. The estimated sewage flow for the collected samples was 750 000 m^3 day^{-1}. The residential population connected to the WWTP is 3 700 000.

Wastewater was collected in precleaned high-density polyethylene (HDPE) bottles, filtered with glass fiber filters (pore size 0.7 μm) immediately upon arrival, then stored in the dark at −18 °C until analysis.

The samples were extracted using a slightly modified protocol from Kern et al.[9] Analyses were carried out using a UHPLC/QTOF-MS system (Dionex UltiMate 3000 RSLC, Thermo Fisher Scientific, Dreieich, Germany). Sample preparation and instrumental analysis are described extensively in the SI (SI, SI-2).

Suspect Substances Processing. Peak lists for suspect screening were obtained using Find Compounds−Chromatogram (TargetAnalysis, Bruker Daltonics, Bremen, Germany), which creates the base peak chromatograms for masses above the given intensity threshold, excluding the isotopic peaks. The criteria used for tentative identification are summarized a flowchart (Figure 1). The criteria included: (1) subtraction of
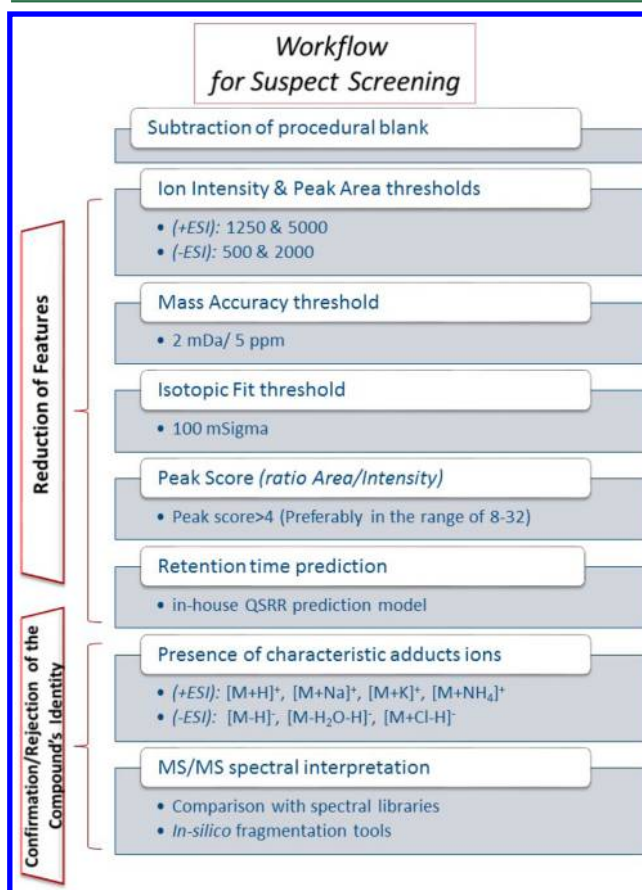

Figure 1. Suspect screening flowchart.

compounds present in the processed method-procedural blanks and those below a threshold of peak area and intensity, (2) a mass accuracy threshold of 2 mDa and 5 ppm on the monoisotopic peaks, (3) a threshold (≤100 mSigma) for the isotopic pattern fit, where mSigma represents the goodness of fit between the measured and theoretical isotopic pattern (mass and ion ratios): the smaller the better,[19] (4) the *peak score*, considering only peaks

with the Peak Area/Peak Intensity ratio greater than 4 (SI, SI-3), (5) the chromatographic retention time plausibility, using an in-house QSRR retention time prediction model (CRTPM).[20] The predicted t_R was considered to match if it was within $\pm 3\delta$ (standardized residual) of the measured value, as this covers 99.7% of normally distributed data. For most retention times, this is approximately equivalent to ± 2 min, (6) the presence of characteristics adducts: $[M + H]^+$, $[M + Na]^+$, $[M+K]^+$, and $[M+NH_4]^+$ in positive mode, and $[M-H]^-$, $[M-H_2O-H]^-$, $[M+Cl-H]^-$; and $[M+HCOOH-H]^-$ in negative mode, (7) MS/MS spectral interpretation, including comparisons with spectra from the MassBank database[21,22] and NIST 2014 library.[23] Spectral similarity values were calculated with the OrgMassSpecR package in R;[24,25] a threshold of 0.7 was used. Other strategies described below were applied to increase the identification confidence. For the tentatively identified compounds that were commercially available, the corresponding standard was purchased in order to confirm the identity of the substance (Level 1).

First, a validation of the suspect screening method was performed using known target compounds (see list in Table S1, SI, SI-1) to determine the successful identification rate. Second, a suspect database (hereafter "metabolite suspects") was created for nine substances and their related human metabolites. Eight high consumption pharmaceuticals (amisulpride, atenolol, clarithromycin, metformin, niflumic acid, ranitidine, tramadol and venlafaxine), and nicotine were selected. The related metabolites were obtained by reviewing the literature[26−28] and by using the MetabolitePredict software from Bruker (Version 2.0, Bruker Daltonics, Bremen, Germany). The full list, including the molecular formulas and the SMILES code[29] (where possible) is presented in the SI (Table S2, SI-3). For metabolites with several possible substitution isomers, the number of structural isomers is indicated and the SMILES code represents only the most likely substance according to our knowledge. When a molecular formula was detected during suspect screening, all possible isomers were considered for the evaluation of the identity of the compound.

A retrospective suspect screening was performed as it became evident that surfactant peaks dominated the nontarget results (see below). Most entries from the suspect surfactant list in Schymanski et al.[4] were used, including the linear alkylbenzyl sulfonates (LAS), sulfophenyl alkyl carboxylic acids (SPACs), sulfophenyl alkyl di-carboxylic acids (SPADCs), di-alkyl tetralin sulfonates (DATS), sulfo-tetralin alkyl carboxylic acids (STACs), sulfo-tetralin alkyl di-carboxylic acids (STADCs), alkyl sulfates (AS), alkyl ethoxy sulfates (AES), secondary alkyl sulfonates (SAS), and 15 nonylphenol ethoxylate (NPEO) sulfates (NPEO-S). PEGs and AGNs were not included as they were studied in detail using the nontarget approach.

Nontarget Substances Processing. Peak peaking for nontarget analysis was carried out using the molecular features algorithm, included in the DataAnalysis software (Bruker). This function clusters all related MS signals and reports them as compounds compiled in a large list for further evaluation. Relevant peaks that were not present in either the procedural blanks or existing target[16] and suspect lists were selected based on the intensity and the presence of distinctive isotopic patterns (as these were considered the most relevant substances with reasonable identification possibilities). For the selected peaks, the most plausible molecular formula(s) were determined using the same parameters for steps (2), (3), and (6) above. Where the formula was not clear, the Seven Golden Rules (SGR)[30] v46 and

GenForm[31] were used to select the molecular formula. GenForm was used with the parameter settings elements $=$ C, H, N, O, P, S (unless there was evidence of halogens), existence filter "exist", odd electron ions (oei), ppm = 5 and acc = 15 (MS1 and MS/MS accuracy settings in ppm). MS/MS spectral interpretation included using both the MassBank library and in silico fragmentation platforms (MS Fragmenter,[32] Mass Frontier[33] and MetFrag[34] via MetFusion[35]) to find candidates. To assess the plausibility of the candidates, a CRTPM was also applied (as Step 7 above). For selected cases, commercial importance criteria were also used, including the number of references and data sources from ChemSpider[36,37] and the number of patents from PubChem.[38] Finally, some compounds were confirmed by purchasing the corresponding standard and comparison of the t_R and MS/MS spectrum. The developed workflow for nontarget screening is summarized in Figure -2.
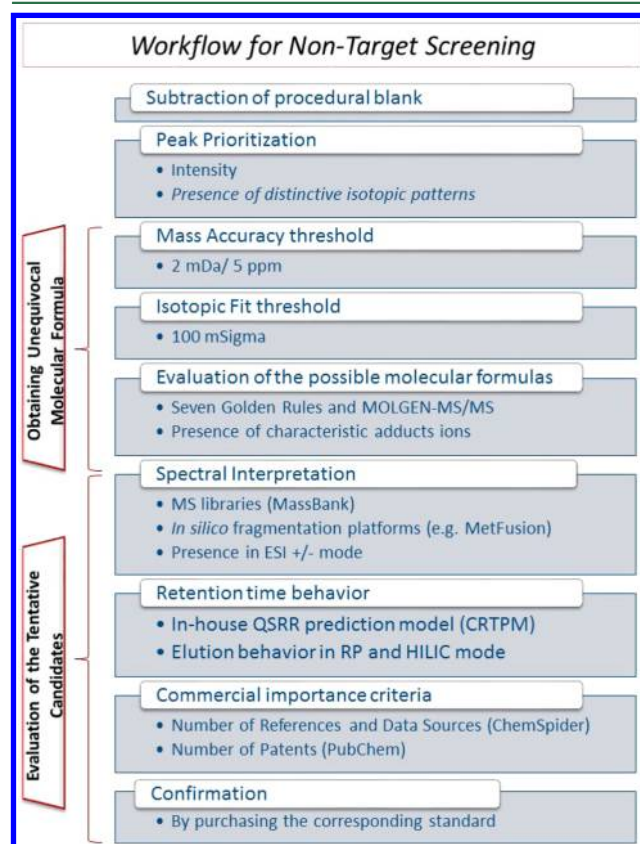

Figure 2. Nontarget screening flowchart.

The level of confidence for the identification of the detected compounds was used according to Schymanski et al.,[17] where Level 1 corresponds to *confirmed structures* (reference standard is available), level 2 to *probable structures*, level 3 for *tentative candidate(s)*, Level 4 to *unequivocal molecular formulas*, and level 5 to *exact mass(es) of interest*.

■ RESULTS AND DISCUSION

Evaluation of the Suspect Screening Approach. The suspect screening workflow was evaluated with 173 target compounds (SI, Table S1), applied as artificial suspects, using TargetAnalysis (Bruker). The only a priori information was the exact mass of the protonated and deprotonated ion ($[M + H]^+$, PI) and $[M-H]^-$, NI) calculated from the chemical formula.

Different parameters were optimized to obtain an acceptable percentage of false negatives while minimizing the number of false positives. First, the peak area and intensity thresholds were optimized using real wastewater spiked at 1 μg L^{-1} with the aim of succeeding false negative rate <5%. These values (area, 5000 (PI) and 2000 (NI); intensity, 1250 (PI) and 500 (NI)) were further applied to lower concentrations. Results are shown graphically in Figure S1 (SI, SI-4). As it is indicated in Figure S1, 94% of the compounds were retained at 0.05 μg L^{-1}. Then, mass accuracy (2 mDa and 5 ppm) and isotopic fit (100 mSigma) thresholds were also applied and the false negative rate was below 5% for both. The combination of the aforementioned filters led to an acceptable false negative rate of 10% at 0.05 μg L^{-1} as shown in Figure S1 (red line). The reduction of features (false positives) due to the applied filters described in this section is summarized in Table S3 (SI, SI-4). The cumulative reduction of hits ranged from 44% to 74% for the different evaluated concentrations. Further steps for the reduction of false positives will be discussed in the next section.
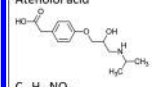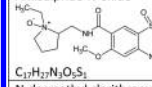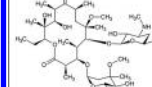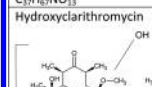
Suspect Screening for Pharmaceutical Metabolites. Screening the influent wastewater (IWW) for the "metabolite suspects" yielded 1660 hits in PI and 864 in NI applying only an accuracy threshold (2 mDa). After the application of steps 1−5 above (also Figure 1), this number decreased to 79 hits (PI, corresponding to 37 substances) and 71 hits (NI, corresponding to 21 substances). The remaining substances were investigated according to steps 6 and 7 for the presence of characteristic adducts and spectral features. After all steps, 13 suspect compounds in PI were tentatively identified fulfilling all criteria. None of the evaluated substances remained for NI, most probably due to the lower sensitivity of this ionization mode. The identified suspects corresponded to all of the selected parent compounds except niflumic acid, see Table 1.
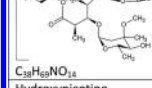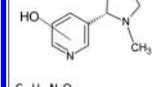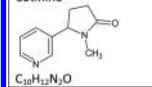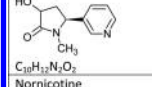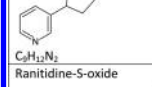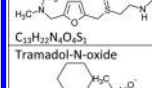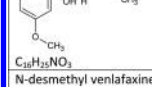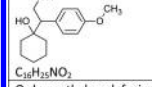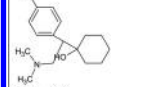
Where library mass spectra matched the measured spectrum, level 2a was assigned initially. Figure S2 (SI, SI-5) shows the suspect cotinine, a metabolite of nicotine, detected at high intensity in IWW and at low concentrations in EWW. The MS/MS spectrum in ESI(+) matched well with MassBank spectrum WA000998 (score = 0.99). After purchase of the commercial standard of cotinine, the identity was confirmed via MS, MS/MS and t_R to reach Level 1. Mass spectral matching also led to other confirmed identifications in suspect (e.g., atenolol acid) and nontarget screening (e.g., panthenol, see below), while two substances remained at level 2a.

For substances without library mass spectra, other evidence was pursued to increase the confidence of the suspect identification. The spectrum of the tentatively identified metabolite hydroxycotinine (Figure S3 (SI, SI-5)) had some peaks in common with cotinine (m/z 80.0493, 118.0647), whereas m/z 98.0602 and 146.0601 were missing and instead m/z 134.0591 and 149.0699 were present, indicating that the hydroxylation occurred on the pyridinyl group. The measured t_R in RP was within predicted range for this substance and the measured HILIC t_R was plausible according to its physicochemical properties. Two possible candidates remained (3- or 4-hydroxycotinine); 3-hydroxycotinine was confirmed through the purchase and analysis of the standard for this compound, reaching level 1.

The complementary nature of HILIC and RP elution as well as presence in influent and effluent samples was exploited to identify guanylurea, a metabolite of metformin. While guanylurea was only present at low levels in the influent (data not shown), it was present at much higher levels in the effluents at intensities

Table 1. Details on the 13 Suspect Metabolites

Name, Structure and Formula	Parent	Exp. t_R (Pred. t_R)[a]	Additional Evidence[b]	Level
Atenolol acid, $C_{18}H_{21}NO_4$	Atenolol	5.1 (5.5)	-Similarity 0.92 with MassBank record EA069710[c]. -Confirmation with standard	1
Amisulpride-N-oxide, $C_{17}H_{27}N_3O_5S_1$	Amisulpride	5.6 (6.5)	- Confirmation with standard	1
N-desmethyl clarithromycin, $C_{37}H_{67}NO_{13}$	Clarithromycin	10.1 (9.6)	-Intra-day trend consistent with the parent compound[d] -Intra-week trend consistent with the parent compound[e] -Confirmation with standard	1
Hydroxyclarithromycin, $C_{38}H_{69}NO_{14}$	Clarithromycin	8.5 (9.2)	-Intra-day trend consistent with the parent compound[d] -Intra-week trend consistent with the parent compound[e]	3
Hydroxynicotine, $C_{10}H_{14}N_2O_2$	Nicotine	3.1 (2.3)	-	3
Cotinine, $C_{10}H_{12}N_2O$	Nicotine	4.9 (5.1)	-Similarity 0.99 with MassBank record WA000998[c] -Confirmation with standard	1
Hydroxycotinine, $C_{10}H_{12}N_2O_2$	Nicotine	3.9 (4.0)	- HILIC/RP elution supports properties[f] -Confirmation with standard	1
Nornicotine, $C_9H_{12}N_2$	Nicotine	2.5 (2.6)	-Similarity 0.77 with NIST record 1185301[c]	2b
Ranitidine-S-oxide, $C_{13}H_{22}N_4O_4S_1$	Ranitidine	2.0 (2.9)	-Intra-week trend consistent with the parent compound[e] -Confirmed with reference standard	1
Tramadol-N-oxide, $C_{16}H_{25}NO_3$	Tramadol	6.0 (8.0)	-Intra-week trend consistent with the parent compound[e]	2b
N-desmethyl venlafaxine, $C_{16}H_{25}NO_2$	Venlafaxine	7.4 (6.3)	-Similarity 0.96 with MassBank record EA103410[c] -Intra-day trend consistent with the parent compound[d] -Intra-week trend consistent with the parent compound[e]	2a
O-desmethyl venlafaxine, $C_{16}H_{25}NO_2$	Venlafaxine	6.0 (6.4)	-Similarity 0.98 with MassBank record EA105304[c] -Intra-day trend consistent with the parent compound[d] -Intra-week trend consistent with the parent compound[e]	2a
Guanylurea, $C_2H_6N_4O$	Metformin	1.3 (1.2)	- HILIC/RP elution supports properties[f] Confirmation with standard	1

[a]All the compounds presented in this table showed feasible chromatographic retention times according to the model.[20] [b]Additional evidence apart from the visual evaluation of the MS/MS spectra and other previously discussed thresholds. [c]Calculated similarity between the experimental spectra and the one obtained in the MS/MS database. [d]By evaluating 2 h composite samples collected every 2 h during 24 h. [e]By evaluating 7 24-h composite samples corresponding to 7 consecutive days. [f]Plausible elution times according to the physicochemical properties and identical MS/MS spectra.
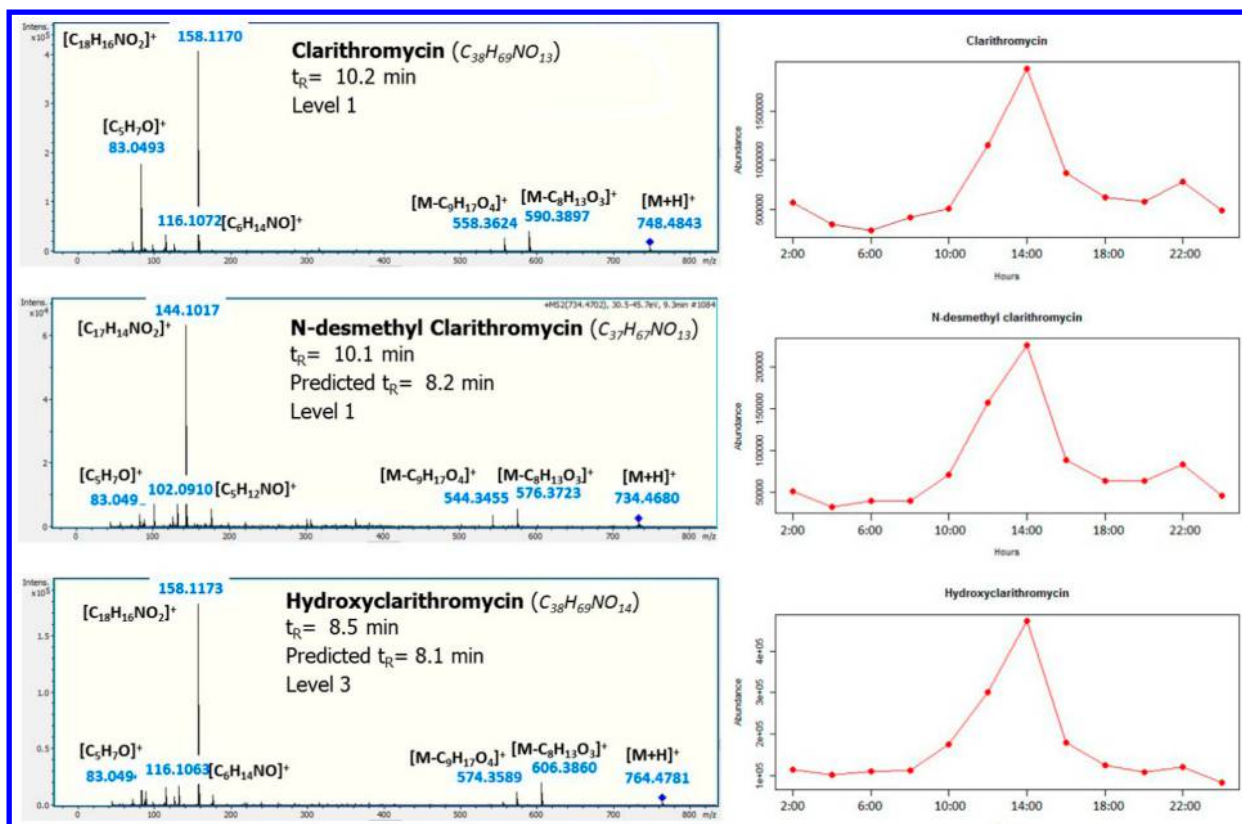
Figure 3. MS/MS spectra and intraday concentration profiles of clarithromycin and its metabolites *N*-desmethyl clarithromycin and hydroxyclarithromycin.

higher than the parent metformin (see Figure S4, SI, SI-5), demonstrating its formation during wastewater treatment. Guanylurea was detected at a low t_R (1.3 min) in the RP column, but at 6.3 min in HILIC, in line with its physicochemical properties. As the presence of the peak at m/z 60.0552 is also present in the metformin spectrum, the standard of guanylurea was purchased and the identification was confirmed to achieve Level 1.

The intraday concentration profile of parent and metabolites can also provide additional evidence for identification, as shown in Figure 3 for clarithromycin and two related metabolites, *N*-desmethyl clarithromycin and hydroxyclarithromycin. The two metabolites had plausible t_R and comprehensive MS/MS spectra that fit with the proposed structures. The concentrations in the IWW measured every 2 h follow identical profiles for clarithromycin and the metabolites. *N*-desmethyl clarithromycin was confirmed with a commercial standard reaching level 1, while the hydroxylated metabolite remains at Level 3. This procedure was also used in the case of venlafaxine and its two metabolites *N*-desmethyl venlafaxine and *O*-desmethyl venlafaxine (final level 2a), as shown in Figure S5A (SI, SI-5) where an excellent interrelation in the intraday profiles among parent/metabolites was observed. A good intraday interrelation between the two confirmed metabolites cotinine and hydroxycotinine (Figure S5B, SI) was also observed, but these profiles did not match the profile of the parent compound, nicotine, indicating that this strategy can provide valuable additional evidence for identification, but the results should be interpreted with caution. The absence of interrelation does not imply a false positive.

Similar conclusions were reached from the comparison of the intraweek concentration profiles among parent compounds and related metabolites. Very similar profiles were observed for the aforementioned metabolites of clarithromycin and venlafaxine and their corresponding parent compounds during 7 consecutive days (Figure S6A and S6B, SI, SI-5). Good interrelations were also found for the metabolites ranitidine-*S*-oxide and tramadol-*N*-oxide and their parent compounds (data not shown). However, poor correlations were found for the rest of investigated compounds (some of them confirmed), showing that this strategy may not always provide supporting evidence.

In the end, 7 of the 13 tentative candidates (amisulpride-*N*-oxide, atenolol acid, *N*-desmethyl clarithromycin, cotinine, norcotinine, ranitidine-*S*-oxide and guanylurea) were confirmed with a commercial standard to reach Level 1, while the rest remained tentative as shown in Table 1.

Nontarget Screening. The nontarget screening approach was applied to masses selected from among the most intense masses detected in the IWW from the WWTP of Athens as described in the flowchart (Figure 2). The workflow was tested by treating the target compounds metformin and acesulfame (both among the most intense peaks detected in the IWW in PI and NI, respectively) as unknown compounds to check the performance of the procedure. Both compounds were unambiguously correctly identified, as shown (only for metformin) in the SI (SI-6, Figure S7).

Nontarget Results: ESI(−) Mode. Table S4 (SI, SI-7) summarizes the results obtained for 25 of the masses detected among the most intense substances in NI. Three substances, panthenol, 4-dodecylbenzenesufonic acid and 3-hydroxypyridine (the latter also identified in ESI (+) mode, see below) were confirmed with the corresponding standard, reaching the confidence level 1. A further 13 substances were tentatively identified as probable structures (level 2) and five more as tentative candidates (level 3). For 3 substances it was not possible to go

beyond the determination of the unequivocal molecular formula (level 4) and only one mass remained as exact mass of interest (level 5). The nontarget results are exemplified using the surfactant compound 2-[2-[2-(2-hydroxyethoxy) ethoxy]ethoxy]ethyl hydrogen sulfate (diglycol ether sulfate, AGN-PC-03I93Q, CAS Number 767579-36-2) (Figure 4A). A peak corresponding to m/z 273.0651 (t_R 2.4 min) was detected among the most intense peaks in NI. After applying the mass accuracy and isotopic fit criteria, 11 possible molecular formulas were found; after applying the SGR, 3 plausible molecular formulas remained. Further processing with GenForm revealed that only one formula, $C_8H_{18}O_8S$, could explain all MS/MS fragments. Searching PubChem via MetFusion, three candidates were retrieved but only one explained all fragments in the MS/MS spectrum (see Figure 4A) and had the highest MetFrag and MetFusion score. The predicted retention time for this compound (3.8 min) was in agreement with the experimental t_R. Moreover, diglycol ether sulfate had the highest commercial importance, with several patents in the field of hair care and herbicide products, indicating potential wide use of this substance and therefore high chances to be present in IWW. A substance with this molecular formula was also detected in PI. Although the intensity was lower, a clear MS/MS spectrum was obtained, showing more detailed fragmentation of the ethoxylate chain. Again diglycol ether sulfate was the only compound that explained all the fragments obtained in the PI MS/MS spectrum. The experimental retention time in PI (2.2 min, as different chromatography was used), was also plausible according with the CRTPM. With all this evidence, an identification level of 2b was assigned.

The presence of other members in the corresponding homologous series ($SO_4C_2H_4(OC_2H_4)_xOH$) was then screened from X = 1 (m/z 186.0193) to X = 19 (m/z 978.4911) and 14 compounds were detected (Figure 4B), from $SO_4C_2H_4(OC_2H_4)_2OH$ to $SO_4C_2H_4(OC_2H_4)_{15}OH$, most of them at high intensity. Although the long glycol ether sulfate (GES) surfactants were out of the domain of the CRTPM in negative mode and thus the predicted t_Rs did not fit the experimental ones, the t_Rs increased constantly with the number of carbons (like the predicted t_Rs) and the chromatographic shape was always similar. This is consistent with observations made for other homologous series detected in other studies.[4,39] For most compounds the intensity was high enough to obtain clear MS/MS spectra, which are listed in Figure S8 (SI, SI-7). The spectra were very similar, showing in all the cases characteristic fragments at m/z 79.9562, m/z 96.9590 and m/z 122.974656, as seen in Figure 4A.

GES surfactants were the compounds detected at the highest intensity in NI, showing its high use in the evaluated area. Apart from the use of these substances in the field of hair care and herbicide products, GES could be also degradation intermediates of AES and other surfactants formed as a result of oxidative degradation processes where low molecular weight products containing oxygen (hydroxyl, peroxide, aldehyde, etc.) are formed either along polymer chain or in its end.[40] To the best of the authors' knowledge, it is the first time that these compounds are identified in wastewater. The presence of GES surfactants was assessed retrospectively in effluent samples from Switzerland and were detected in all samples but at greatly varying intensities, shown in Figure S9A and S9B (SI, SI-7). This variation indicates that either the input or removal of these compounds may vary between different WWTPs (no corresponding influent samples were available). This example illustrates the power of retrospective analysis using HRMS.

Nontarget rResults: ESI(+) Mode. Table S5 (SI, SI-8) summarizes the results obtained for 25 of the highest intensity masses detected in PI. Two substances, 4-PEG and 3-hydroxy-pyridine, were confirmed (level 1) with the corresponding standard. Another 13 PEGs were tentatively identified and structures were proposed (level 3, since different isomers are possible). Another 8 remained with an unequivocal molecular formula (level 4) and 3 masses remained as exact mass of interest (level 5). The identity of the adduct state was confirmed by the presence of other adducts where at least two of $[M + H]^+$, $[M+NH_4]^+$ and $[M + Na]^+$ were present.

The identification of 3-hydroxypyridine is summarized in Figure S10 (SI, SI-8). A very intense peak corresponding to m/z 96.0446 was detected at t_R 2.27 min, with only one plausible formula, C_5H_5NO, found after applying mass accuracy and isotopic fit filters. There were 38 compounds with this formula in ChemSpider, where only two of them (3-pyrinidol and 4-pyrinidol) had a good MetFusion score when evaluating the MS/MS spectra. Both compounds explained the two MS/MS fragments, had plausible t_R according to the CPTRM and were both of high commercial importance according to the number of data sources and references. ESI(+) MS/MS spectra were available in MassBank for 3-hydroxypyridine (similarity 0.997 with MassBank spectrum PR100333). Standards of 3-hydroxypyridine and 4-hydroxypyridine were purchased and the identity of 3-hydroxypyridine was confirmed by t_R.

The identification of tetraethylene glycol (4-PEG, m/z 195.1233, t_R 4.2 min) is summarized in Figure S11 (SI, SI-8). As above, only one plausible molecular formula remained after the accuracy and isotopic fit criteria: $C_8H_{18}O_5$. ChemSpider yielded 13 possible structures for this molecular formula, where only three had a MetFusion score above 0.5. The three substances were able to explain the six fragments found in the MS/MS spectrum (see Figure S11), with similar MetFusion scores and plausible t_R. However, the number of data sources and references indicated a greater commercial importance and thus likelihood of presence in IWW of 4PEG (379 references) compared with the other two candidate compounds (2 and 1 references, respectively). The identity of 4PEG was confirmed with a standard, showing the importance of considering the number of sources and references.

PEGs were the family of compounds present at the highest levels in PI mode in the evaluated wastewater samples, consistent with previous studies.[4,16] 5-PEG was identified following the same procedure described above, using the protonated adduct. From 6-PEG and onward, the $[M+NH_4]^+$ adduct was predominant and used to provide higher sensitivity for these compounds when increasing the length of the chain (Figure S12A, SI, SI-8). Figure S12B shows the extracted chromatograms for all the detected PEGs (using the ammonia adducts). Consistent peak shapes and constant increase of t_R are observed when increasing the chain length. In all cases, MS/MS spectra were similar and consistent with the proposed structures (data not shown), as well as with previous observations.[4,16]

Retrospective Suspect Screening of Surfactants. As a high number of tentatively identified but differing surfactant substances were among the most intense peaks, a retrospective suspect screening was performed using the surfactants list of 398 suspects described above. After applying the aforementioned thresholds of intensity, mass accuracy, isotopic pattern and peak score, 110 suspects (88 in NI and 22 in PI) remained and were further evaluated. Plausible t_R times among the homologue series, in RP and in HILIC mode, along with the evaluation of the
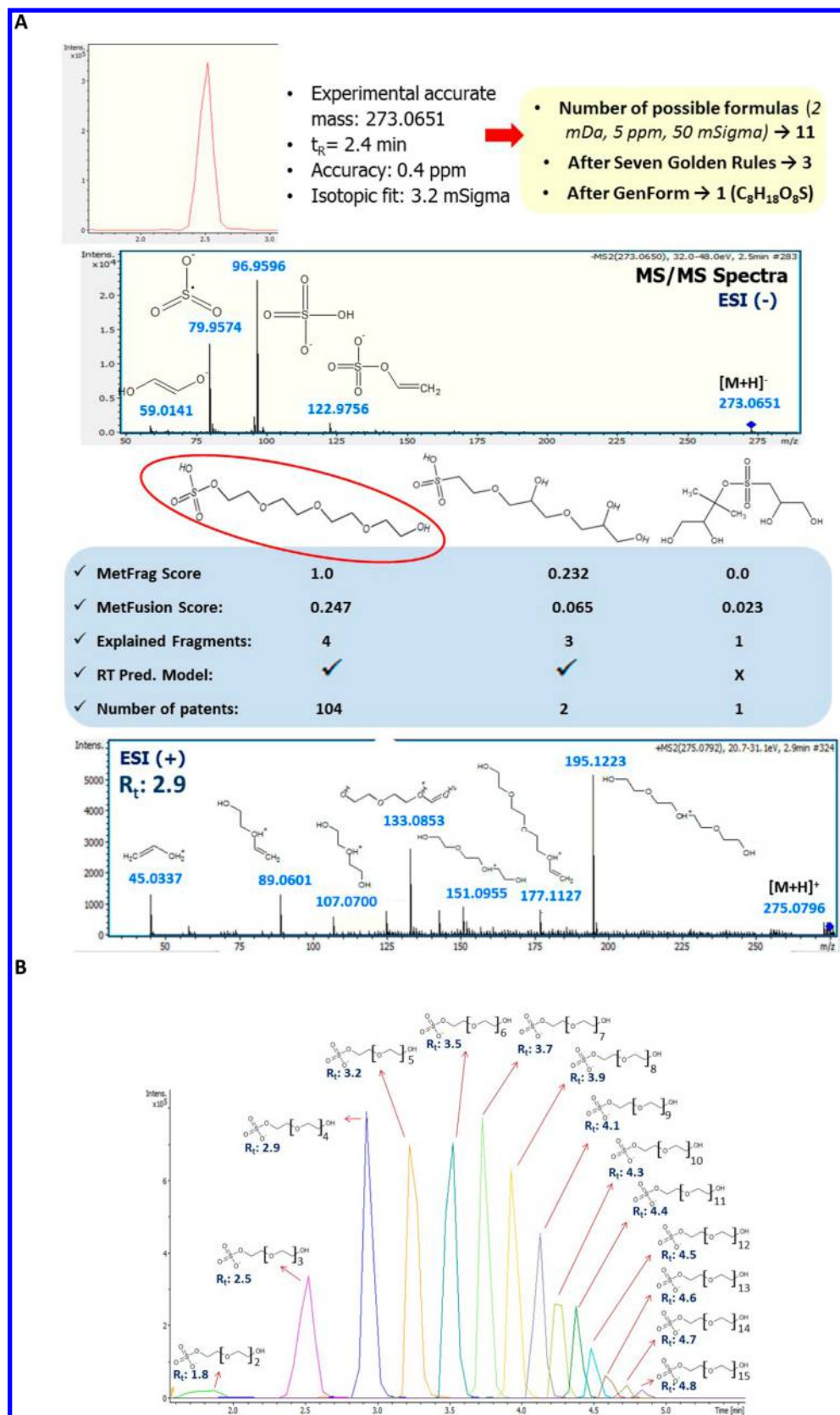
Figure 4. (A) Nontarget identification of the unknown compound diglycol ether sulfate. (B) Extracted ion chromatograms of the corresponding homologous series ($SO_4C_2H_4(OC_2H_4)_xOH$) (glycol ether sulfate (GES) surfactants).

MS/MS spectra were the main criteria used for the tentative identification of the suspects. This information along with the names, molecular formulas and exact masses is summarized in Table S6, in the SI (SI-9).

In total, 82 substances were tentatively identified, 38 with MS/MS evidence and 44 without MS/MS evidence, but with additional information (plausible t_R and chromatographic peak shape among the homologue series), supporting their presence. Nineteen substances were rejected on the basis that either the t_R or MS/MS did not match or simply due to the absence of evidence supporting their presence. The SPCs and LAS were present at high intensity. Ten SPCs, from C4-SPAC to C13-SPAC, were tentatively identified. Figure S13A (SI, SI-9) shows increasing 3t_R with increasing length of the alkyl chain and consistent peak shapes. Plausible MS/MS spectra were observed, including the characteristics fragments $m/z = 79.9574$ (SO_3^-) and $m/z = 183.0121$ ($C8H7SO_3^-$), as shown in Figure S13B for C8-SPAC. Some spectra are available in MassBank stemming from the previous study[4] and showed a good fit compared with those obtained experimentally (e.g., C8-SPAC, similarity 0.978 with spectrum ETS00018). This clearly shows the benefits of sharing also suspect spectra in public libraries. Other groups of surfactants were tentatively identified similarly, including DATS (7 substances), LAS (4 substances), AS (4 substances), AEOs (13 substances) or DEAs (3 substances). However, MS/MS spectra could not be obtained for some surfactant classes due to the low intensity of the peaks. The less transformed surfactants (e.g., DATS), were detected in higher number and higher intensity than their related TPs (e.g., STACs and STADCs), which is consistent with the evaluated matrix of IWW. The opposite trend was observed in a previous study performed with EWW,[4] showing that the comparison between these two matrices may provide additional evidence in the tentative identification of surfactants. Where MS/MS was absent, the tentative identification was based on chromatographic criteria as demonstrated for C13-AES in Figure S14A (SI, SI-9). Seven of these substances were identified based on the increasing t_R (from 13.2 to 13.9 min) when increasing the length of the alkyl chain, along with consistent peak shapes. A plausible chromatographic behavior was also observed for these compounds in HILIC mode, showing t_R from 1.2 to 1.4 min, shown in Figure S14B. The observed t_R in both RP and HILIC were also plausible when compared with those obtained for C12-AES, C14-AES, and C16-AES. Although the identification confidence is not as high as in the cases where MS/MS data is available, these results are supported by the fact that previous studies also detected these substances and a similar chromatographic behavior was reported.[4,39]

In view of these results, the new strategies applied here provide valuable additional evidence for the identification of suspect and unknown compounds in environmental samples. The comparison of the daily and/or weekly concentration trends helped increase the identification confidence in some cases. Thirty four nontarget compounds were tentatively identified and four were confirmed, showing the good performance of the developed approach. Consistent with previous studies, a high proportion of high intensity peaks corresponded to surfactants and the identification of the surfactant series GES is reported for the first time. The usefulness of retrospective analysis is clear as these GES were found subsequently in samples from another European geographical area.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.5b03454.

Additional information as noted in the text (chemicals; sample preparation and instrumental parameters; suspect workflow, validation data and more results; nontarget workflow, validation and more results) (PDF)

AUTHOR INFORMATION

Corresponding Author

*Phone: +30 210 7274317; fax: +30 210 7274750; e-mail: ntho@chem.uoa.gr.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Daughton, C. G. Document Non-regulated water contaminants: Emerging research. *Environ. Impact Assess. Rev.* **2004**, *24*, 711−732.

(2) European Commission. Directive 2013/39/EU of the European Parliament and the Council of 12. August 2013 amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy. *Off. J. Eur. Union* **2013**, L 226/1.

(3) Hug, C.; Ulrich, N.; Schulze, T.; Brack, W.; Krauss, M. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. *Environ. Pollut.* **2014**, *184*, 25−32.

(4) Schymanski, E. L.; Singer, H. P.; Longree, P.; Loos, M.; Ruff, M.; Stravs, M. A.; Ripolles-Vidal, C.; Hollender, J. Strategies to Characterize Polar Organic Contamination in Wastewater: Exploring the Capability of High Resolution Mass Spectrometry. *Environ. Sci. Technol.* **2014**, *48* (3), 1811−1819.

(5) Krauss, M.; Singer, H.; Hollender, J. LC−high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal. Bioanal. Chem.* **2010**, *397*, 943−951.

(6) Moschet, C.; Piazzoli, A.; Singer, H.; Hollender, J. Alleviating the reference standard dilemma using a systematic exact mass suspect screening approach with liquid chromatography-high resolution mass spectrometry. *Anal. Chem.* **2013**, *85* (21), 10312−10320.

(7) Vergeynst, L.; Van Langenhove, H.; Joos, P.; Demeestere, K. Suspect screening and target quantification of multi-class pharmaceuticals in surface water based on large-volume injection liquid chromatography and time-of-flight mass spectrometry. *Anal. Bioanal. Chem.* **2014**, *406* (11), 2533−2547.

(8) Zonja, B.; Delgado, A.; Pérez, S.; Barceló, D. LC-HRMS suspect screening for detection-based prioritization of iodinated contrast media photodegradates in surface waters. *Environ. Sci. Technol.* **2015**, *49* (6), 3464−3472.

(9) Kern, S.; Fenner, K.; Singer, H. P.; Schwarzenbach, R. P.; Hollender. Identification of transformation products of organic contaminants in natural waters by computer-aided prediction and high-resolution mass spectrometry. *Environ. Sci. Technol.* **2009**, *43*, 7039−7046.

(10) Nurmi, J.; Pellinen, J.; Rantalainen, A. L. J. Critical evaluation of screening techniques for emerging environmental contaminants based on accurate mass measurements with time-of-flight mass spectrometry. *J. Mass Spectrom.* **2012**, *47*, 303−312.

(11) Segura, P. A.; MacLeod, S. L.; Lemoine, P.; Sauvé, S.; Gagnon, C. Quantification of carbamazepine and atrazine and screening of suspect organic contaminants in surface and drinking waters. *Chemosphere* **2011**, *84*, 1085−1094.

(12) Chiaia-Hernandez, A. C.; Krauss, M.; Hollender. Screening of lake sediments for emerging contaminants by liquid chromatography atmospheric pressure photoionization and electrospray ionization coupled to high resolution mass spectrometry. *Environ. Sci. Technol.* **2012**, *47*, 976−986.

(13) Hernandez, F.; Ibanez, M.; Portoles, T.; Cervera, M. I.; Sancho, J. V.; Lopez, F. J. Advancing towards universal screening for organic pollutants in waters. *J. Hazard. Mater.* **2015**, *282*, 86−95.

(14) van Leerdam, J. A.; Vervoort, J.; Stroomberg, G.; de Voogt, P. Identification of unknown microcontaminants in Dutch river water by liquid chromatography-high resolution mass spectrometry and nuclear magnetic resonance spectroscopy. *Environ. Sci. Technol.* **2014**, *48* (21), 12791−12799.

(15) Zedda, M.; Zwiener, C. Is nontarget screening of emerging contaminants by LC-HRMS successful? A plea for compound libraries and computer tools. *Anal. Bioanal. Chem.* **2012**, *403* (9), 2493−2502.

(16) Schymanski, E.L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I. M.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; Thomaidis, N. S.; Bletsou, A.; Zwiener, C.; Ibáñez, M.; Portolés, T.; de Boer, R.; Reid, M. J.; Onghena, M.; Kunkel, U.; Schulz, W.; Guillon, A.; Noyon, N.; Leroy, G.; Bados, P.; Bogialli, S.; Stipaničev, D.; Rostkowski, P.; Hollender, J. Non-target screening with high resolution mass spectrometry: Critical review using a collaborative trial on water analysis. *Anal. Bioanal. Chem.* **2015**, *407* (21), 6237−6255.

(17) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014b**, *48* (4), 2097−2098.

(18) Thomaidis, N. S.; Gago-Ferrero, P.; Ort, C.; Borova, V.; Alygizakis, N.; Maragou, N.; Dasenaki, M.; Pistos, C. Effect of strong socio-economic changes on licit and illicit drug use patterns through sewage epidemiology *Environ. Sci. Technol.* **2015**, in preparation.

(19) https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/literature/ApplicationNotes/ET-21%20Metabolomics_ebook.pdf (accessed 8 July 2015).

(20) Aalizadeh, R.; Thomaidis, N. S.; Bletsou, A. A.; Gago-Ferrero, P. **2015**, in preparation.

(21) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010**, *45* (7), 703−714.

(22) MassBank (2014) NORMAN MassBank (www.massbank.eu) (accessed 8 July 2015).

(23) The NIST 14 mass spectral library. www.sisweb.com/software/ms/nist.htm (accessed 28 June 2015).

(24) Dodder, N. G. and with code contributions from Mullen, K. M. (2014). OrgMassSpecR: Organic Mass Spectrometry. R package version 0.4−4. http://CRAN.R-project.org/package=OrgMassSpecR.

(25) Stein, S. E.; Scott, D. R. Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 859−866.

(26) Trautwein, C.; Kümmerer, K. Incomplete aerobic degradation of the antidiabetic drug Metformin and identification of the bacterial dead-end transformation product Guanylurea. *Chemosphere* **2011**, *85*, 765−773.

(27) Helbling, D.; Hollender, J.; Kohler, H. P.; Singer, H.; Fenner, K. High-Throughput Identification of Microbial Transformation Products of Organic Micropollutants. *Environ. Sci. Technol.* **2010**, *44*, 6621−6627.

(28) Kern, S.; Baumgartner, R.; Helbling, D. E.; Hollender, J.; Singer, H.; Loos, M. J.; Schwarzenbach, R. P.; Fenner, K. A tiered procedure for assessing the formation of biotransformation products of pharmaceuticals and biocides during activated sludge treatment. *J. Environ. Monit.* **2010**, *12*, 2100−2111.

(29) Daylight (2012) SMILES - A Simplified Chemical Language . Daylight Chemical Information Systems Inc. http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html (accessed 8 July 2015).

(30) Kind, T.; Fiehn, O. Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinf.* **2007**, *8*, 105.

(31) Meringer, M.; Reinker, S.; Zhang, J. A.; Muller, A. MS/MS data improves automated determination of molecular formulas by mass spectrometry. *Match-Commum. Math. Co.* **2011**, *65* (2), 259−290.

(32) Advanced Chemistry Development, Inc (2010) ACD/MS Fragmenter. http://www.acdlabs.com/products/adh/ms/ms_frag/ (accessed 8 July 2015).

(33) HighChem (2015) Mass Frontier. 7.0 edn. HighChem Ltd./Thermo Scientific, Bratislava, Slovakia.

(34) Wolf, S.; Schmidt, S.; Muller-Hannemann, M.; Neumann, S. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinf.* **2010**, *11*, 148.

(35) Gerlich, M.; Neumann, S. MetFusion: integration of compound identication strategies. *J. Mass Spectrom.* **2013**, *48*, 291−298.

(36) Little, J. L.; Williams, A. J.; Pshenichnov, A.; Tkachenko, V. Identification of "Known Unknowns" Utilizing Accurate Mass Data and ChemSpider. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (1), 179−185.

(37) RSC (2015) ChemSpider. Royal Society of Chemistry. www.chemspider.com (accessed 28 April 2015).

(38) NCBI (2015) PubChem. National Center for Biotechnology Information https://pubchem.ncbi.nlm.nih.gov/ (accessed 8 July 2015).

(39) Lara-Martin, P. A.; Gonzalez-Mazo, E.; Brownawell, B. J. Multiresidue method for the analysis of synthetic surfactants and their degradation metabolites in aquatic systems by liquid chromatography-time-of-flight-mass spectrometry. *J. Chromatogr. A* **2011**, *1218* (30), 4799−4807.

(40) Celina, M. C. Review of polymer oxidation and its relationship with materials performance and lifetime prediction. *Polym. Degrad. Stab.* **2013**, *98*, 2419−2429.