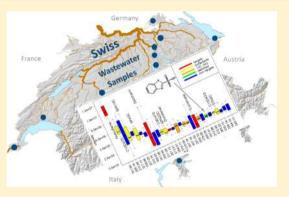# Strategies to Characterize Polar Organic Contamination in Wastewater: Exploring the Capability of High Resolution Mass Spectrometry

Emma L. Schymanski,[†] Heinz P. Singer,[†] Philipp Longrée,[†] Martin Loos,[†,§] Matthias Ruff,[†] Michael A. Stravs,[†,§] Cristina Ripollés Vidal,[‡] and Juliane Hollender[†,§,*]

[†]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland
[‡]University Jaume I, Department of Physical and Analytical Chemistry, Avda. Sos Baynat S/N 12071 Castellón, Spain
[§]Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, CH-8092, Zürich, Switzerland

S *Supporting Information*

ABSTRACT: Wastewater effluents contain a multitude of organic contaminants and transformation products, which cannot be captured by target analysis alone. High accuracy, high resolution mass spectrometric data were explored with novel untargeted data processing approaches (enviMass, nontarget, and RMassBank) to complement an extensive target analysis in initial "all in one" measurements. On average 1.2% of the detected peaks from 10 Swiss wastewater treatment plant samples were assigned to target compounds, with 376 reference standards available. Corrosion inhibitors, artificial sweeteners, and pharmaceuticals exhibited the highest concentrations. After blank and noise subtraction, 70% of the peaks remained and were grouped into components; 20% of these components had adduct and/or isotope information available. An intensity-based prioritization revealed that only 4 targets were among the

top 30 most intense peaks (negative mode), while 15 of these peaks contained sulfur. Of the 26 nontarget peaks, 7 were tentatively identified via suspect screening for sulfur-containing surfactants and one peak was identified and confirmed as 1,3-benzothiazole-2-sulfonate, an oxidation product of a vulcanization accelerator. High accuracy, high resolution data combined with tailor-made nontarget processing methods (all available online) provided vital information for the identification of a wider range of heteroatom-containing compounds in the environment.

■ INTRODUCTION

Wastewater effluents contain potentially tens of thousands of substances that are in daily use in households and industry at varying concentrations, forming a major point source for contamination of surface waters.[1] Thus, both parent compounds and their transformation products (TPs) can potentially accumulate in the environment and exert adverse effects.[2] Improvements in extraction, enrichment, and analytical procedures mean that increasing numbers of chemicals can be detected in samples. The evolution of high resolution (HR), high accuracy mass spectrometry (MS), coupled with liquid chromatography (LC, together LC-HRMS) has opened up new windows of opportunity for the detection of polar organic contaminants in complex samples. With this technology, many additional compounds in water matrixes that are not well amenable to gas chromatography (GC) without derivatization can be monitored, including those with functional groups such as acids, phenols, and amines. Furthermore, it is now possible to monitor expected and unexpected compounds together in a sample using LC-HRMS with full scan acquisition methods. Three major approaches for postmeasurement processing were detailed by Krauss et al.;[3] *target analysis* (with reference

standards), *suspect screening* (with suspected substances based on prior information but no reference standards), and finally *nontarget screening* (no prior information, no reference standards), modifying the terms introduced by Hernandez et al.[4]

For a comprehensive *target analysis*, a reference standard is necessary to determine the concentration in the sample and to match the measured retention time (RT) and, if available, tandem mass spectrum (MS/MS). An isotope-labeled internal standard (IS) should ideally be available for each target to assess the sample-specific response. A full calibration curve yields quantitative concentrations, otherwise the results are semiquantitative (one-point calibration). A complete target analysis cannot be performed for all compounds of potential environmental relevance, as this would involve the purchase and measurement of hundreds, if not thousands, of chemicals—for which reference standards are not always

available.[5] Thus, when analyzing complex samples, a balance is needed between extensive target analysis (ideally yielding concentrations of the most relevant compounds) and screening methods, which can assist in tentatively identifying other potentially relevant compounds.

Suspect screening with LC-HRMS(/MS) data relies on accurate mass and isotope information available for the precursor ion and additional evidence (see, e.g., Krauss et al.[3] and below for more) for tentative identification. Compounds that are expected to be in the samples (the "suspects") can be screened using the exact mass of their expected ions, calculated from the molecular formula.[5] LC-HRMS screening currently remains inherently different to gas chromatographic (GC) screening studies.[6] While extensive databases of comparable electron impact (EI) mass spectra are available to support tentative identification of compounds in GC-EI/MS investigations,[7] databases for LC-MS/MS are still relatively small[8] and the spectra are not as reproducible between instruments.[7] Thus, although exact mass screening methods are computationally rapid and many masses can be screened in a given sample, the gathering of evidence and confirmation of the screened masses remains very time-consuming.

Nontarget screening involves masses that are detected in the samples, but where no *a priori* information on the underlying compound is available beforehand.[3,9] Full identification of the nontarget mass is often difficult, with no guarantee of a successful outcome.[3,10] High accuracy, high resolution data improve the chances of a unique molecular formula assignment to detected masses.[11] Gonsior et al.[12] unambiguously assigned $C_xH_yO_zS_1$ formulas using mass and isotope patterns (including the sulfur peak) to 872 peaks in wastewater effluent samples measured in negative mode Fourier Transform Ion Cyclotron Resonance (FT-ICR) MS, but did not confirm structures for these masses. Incorporating high accuracy MS/MS information further improves formula assignment,[13] which in turn improves the results of compound database queries (e.g., ChemSpider[14]) by retrieval of fewer candidates to identify "known unknowns".[15] The isotopic signal available from high resolution instruments, which are capable of resolving especially ^{34}S, but also ^{15}N and ^{18}O (depending on the m/z) in addition to ^{37}Cl and ^{81}Br isotopes, provides additional vital information about composition, which can be used during the unknown identification. Identifying "unknown unknowns" (where the compound is not in a compound database) is even more challenging and beyond the scope here.

The goal of this study was to develop and apply the 3-fold approach of extensive target analysis, suspect screening and finally nontarget screening to perform a comprehensive characterization of polar compounds in wastewater effluents, which can enter the aquatic environment. To determine the composition of typical municipal wastewater effluent after state-of-the-art (tertiary) treatment, effluent samples were collected from 10 different municipal wastewater treatment plants (WWTPs) in Switzerland. The analytical methods included (a) multiresidue extraction, (b) reversed phase HPLC separation combined with HR-MS/MS, and (c) a comprehensive target list. The further characterization was prioritized using a peak inventory as well as intensity, isotope/adduct and data-dependent MS/MS information. Suspect screening (demonstrated for sulfur-containing surfactants with of ^{34}S isotope fine structure) and nontarget screening used software developed in-house, optimized on HR-MS data and available

online (enviMass[16] and the R packages nontarget[17] and RMassBank[8,18]).

■ EXPERIMENTAL SECTION

Sampling and LC-HRMS/MS Analysis. Flow-proportional effluent samples (24 h) were collected from 10 WWTPs in Switzerland with conventional tertiary treatment systems, summarized in Figure S1 and Table S1 of the Supporting Information (SI). Samples were collected by plant operators in February 2010 during dry weather, filled into 1 L glass bottles and stored in the dark at −20 °C prior to preparation. A sample volume of 0.25 L was adjusted to pH 6.5 with formic acid or ammonia and pressure-filtered through a 0.7 μm glass fiber filter (Whatman). Isotope-labeled IS (103 total, 100 ng, details in the SI) were spiked to each sample prior to enrichment with mixed-bed multilayer solid-phase extraction cartridges comprising Oasis HLB, Isolute ENV+, Strata-X-AW, and Strata-X-CW (exact details in Kern et al.[19]) via vacuum extraction at 10 mL/min. The analytes were then extracted from the dried cartridges with a 6 mL basic (2% of 25% ammonia) followed by a 3 mL acidic mixture (1.7% of 100% formic acid) of ethyl acetate/methanol (50:50 V/V). The neutral combined extract was concentrated to 100 μL under a gentle nitrogen stream, adjusted to 1 mL with HPLC-grade water, filtered through a 0.45 μm regenerated cellulose filter into a 2 mL vial and stored at 4 °C prior to analysis. While the filtering steps are necessary to avoid clogging the HPLC system, previous investigations indicated that this does not cause losses in the compounds of interest.[20]

High performance liquid chromatography (HPLC) analysis was performed on 20 μL of the extracts. The HPLC system consisted of a PAL Autosampler (CTC Analytics, Zwingen, Switzerland), a Rheos 2200 quaternary low pressure mixing pump (Flux Instruments, Basel, Switzerland), and an XBridge C18 column (3.5 μm, 2.1 × 50 mm) from Waters (Milford, U.S.) with a 2.1 × 10 mm precolumn of the same material. The gradient (water/methanol, both with 0.1% formic acid) was 90:10 at 0 min, to 50:50 at 4 min, to 5:95 at 17 min, held until 25 min then 90:10 at 25.1 to 30 min at a flow of 200 μL/min and a column temperature of 30 °C. Full scan MS detection was performed with an LTQ Orbitrap XL (resolution R = 60 000 at m/z 400, for m/z = 115 to 1000) from Thermo Fisher Scientific (San Jose, U.S.) with electrospray ionization (ESI) in positive and negative mode, with a spray voltage of +4 and −4 kV, respectively, and a capillary temperature of 300 °C. Data-dependent acquisition was used to record 6 MS/MS scans using a 1.5 Da isolation window between each full scan in the original measurements. Samples were remeasured on a Q-Exactive (Thermo Fisher Scientific, San Jose, U.S.) with an essentially identical experimental setup, using inclusion lists and targeted MS/MS experiments to obtain additional data for identification/confirmation. Further details are given in the SI. All HRMS data were peak picked using Formulator (Thermo Scientific, San Jose, U.S.).

Quantitative and Semiquantitative Target Analysis. A total of 364 target compounds were available, including pharmaceuticals, biocides, illicit drugs, industrial chemicals, perfluorinated compounds, food additives, corrosion inhibitors, personal care products, and plant protection products (hereafter called "pesticides" for simplicity). Of these, 91 were transformation products, predominantly from pesticides, pharmaceuticals, biocides, and corrosion inhibitors. The targets were selected specifically for Swiss conditions and sample

Table 1. Breakdown of Average Peak Numbers into Components in the 10 WWTP Effluent Samples, Plus/Minus Standard Deviation

	positive		negative	
	picked peaks (%)	components (%)	picked peaks (%)	components (%)
processing				
total	13400 ± 2900 (100)		14000 ± 1500 (100)	
sparks/noise	1890 ± 460 (14)		2970 ± 350 (21)	
blank peaks	2090 ± 340 (15)		1560 ± 160 (11)	
target and standards				
internal standards	174 ± 19 (1.3)	72 ± 2	62 ± 3 (0.45)	17 ± 1
targets	160 ± 14 (1.2)	84 ± 5	55 ± 5 (0.40)	22 ± 2
nontargets				
nontargets	9550 ± 2750 (71)	6700 ± 1670 (100)	9610 ± 1390 (69)	7120 ± 790 (100)
monoisotopic peaks		5300 ± 1200 (79)		5720 ± 580 (80)
isotope and/or adducts		1400 ± 470 (21)		1400 ± 200 (20)
- ^{13}C isotopes[a]		755 ± 235		861 ± 132
- ^{15}N isotopes[a]		7 ± 6		4 ± 3
- ^{34}S isotopes[a]		13 ± 7		123 ± 23
- ^{37}Cl isotopes[a]		24 ± 8		153 ± 22
- ^{81}Br isotopes[a]		13 ± 6		45 ± 11
- adduct signals		630 ± 170		269 ± 27
suspect/homologues				
suspects		133 ± 31 (1.9)		129 ± 14 (1.8)
homologues		800 ± 720 (12)		180 ± 70 (2.6)

[a]Includes isotopes within small error margins (±0.2 ppm) only; see SI for more details. Although one component can have several isotope signals present (e.g., ^{13}C and ^{34}S), multiple signals from the same element (e.g., ^{37}Cl) are only counted once.

type.[21] Suppliers for the target compounds and IS are given in the SI.

First, 125 target compounds (with 103 corresponding IS) were analyzed quantitatively, given in SI Table S2 with common names, CAS number, detection limits, and identification points (where applicable). Quantification was performed using Xcalibur (Thermo Scientific, v. 2.1.0.1139, 2009). Concentrations are given in SI Table S3, indicated with a Q.

The remaining 239 target compounds were screened with enviMass[16] (Eawag, Switzerland, version 1.0). First, a noise and blank subtraction step was performed, followed by detection of targets and internal standards with their associated isotope and adduct peaks. The resulting one-point calibration on peak-picked data yielded semiquantitative concentrations. These results (indicated with an S) are given in SI Table S3, along with a common name and CAS number. The enviMass parameters are given in SI Table S4.

Peak Inventory and Prioritization. Following target and internal standard detection, a nontarget mass list for each sample was compiled with enviMass. Isotope and adduct grouping ("componentization") of these nontarget masses was subsequently performed using the R package "nontarget"[17] (Eawag, Switzerland, version 1.0). This retrieved the most intense peak per unknown compound and grouped all related peaks together into one component to associate all adduct and isotope peaks with the compound for tracking during the suspect and nontarget screening and to avoid overestimation of the total number of unknown compounds. Additionally, the homologue detection feature of the nontarget package was used to link potential homologous series in the samples by searching for consistent mass differences. The parameters are given in SI Table S5.

An Excel macro was used to extract all nontarget masses above an intensity of 10^6 that occurred in at least one sample in either ionization mode. This intensity corresponds to approximately 20 ng/L of well-ionizable compounds (e.g., atrazine) in positive mode and allows sufficient intensity for MS/MS experiments and thus identification efforts. These masses were then screened across all samples; those that were present in all samples were prioritized by cumulative intensity and used to form inclusion lists for MS/MS acquisition.

Suspect Screening. Suspect lists, including several surfactant homologous series among others, were compiled from the literature[22−32] and summarized in SI Table S6. The corresponding molecular formulas were added to enviMass, and the isotopologue masses were calculated for the $[M + H]^+$ or $[M − H]^−$ species in positive and negative ionization modes, respectively (additional adducts were detected via the "nontarget" package as described above). The suspect screening was performed using enviMass with a retention time window spanning the entire chromatographic run. All other parameters were consistent with those in SI Table S4.

Evidence used to support (or reject) the tentative suspect identification included (1) occurrence of several masses in the homologous series, (2) expected retention time behavior (e.g., increasing with increasing length of the alkyl chain), (3) similar chromatographic peak shape determined visually for members of a homologous series, (4) a rational pattern of intensity distributions according to literature data, (5) occurrence of isotope peaks as expected, and (6) interpretation of the MS/MS. Selected literature spectra[12,22,24,26,27,29,32] were digitized and saved into a personal MassBank[33] database, including fragment structures, where available. The chromatographs of homologous mass traces identified via enviMass and the nontarget package were plotted using a script based on the R package RMassBank[8,18] to overlay the peaks and extract the corresponding MS/MS, where available. All results were cross-checked using Xcalibur (Thermo Scientific, version 2.1.0.1139, 2009).
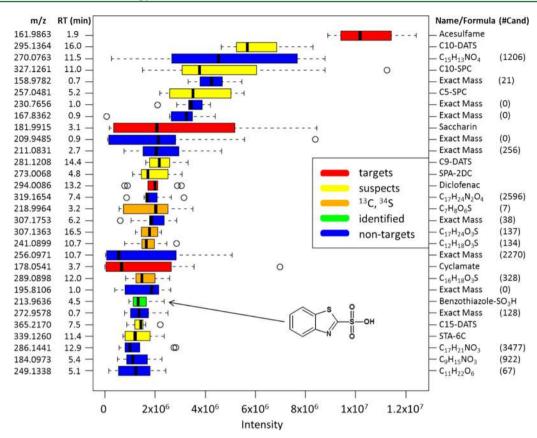
Figure 1. Top 30 components (ranked by intensity) in negative mode, colored according to target (red), suspect (yellow), or nontarget (blue). Orange indicates ^{34}S presence, green is a confirmed nontarget (insert). Left: m/z and RT(min). Right: Name or formula/exact mass and number of candidates.

Nontarget Screening. MOLGEN-MS/MS[34] was used to calculate molecular formulas from the exact mass and isotope patterns from the MS as well as MS/MS fragmentation information, if available. MetFusion[35] was used to perform parallel searches of compound databases and spectral libraries and perform in silico fragmentation of the candidate structures. The number of references, retrieved from ChemSpider[14] per compound was also used to rank candidates.[15] Further details are in the SI.

■ RESULTS AND DISCUSSION

Peak Inventory. The number of picked peaks in each of the ten wastewater samples ranged between 10 591 to 20 033 (positive mode) and 12 565 to 17 657 (negative mode). A peak inventory is given in Table 1, including the average number of peaks and components over all 10 samples with the standard deviation and a breakdown of the isotopes. The "picked peaks" columns refer to the (ungrouped) output from Formulator and can include multiple peaks (isotopes, adducts) associated with a compound. These peaks are grouped together to the respective most intense monoisotopic mass in the "components" column, such that each detected compound is only represented once in the list. For example, an average of 174 IS peaks are detected in each sample in positive mode and this corresponds with 72 IS components in positive mode (i.e., 72 of the 103 IS are detected in positive mode in the sample and ~2.4 isotope or adduct peaks are associated with each IS).

As shown in Table 1, only 1.2% of the peaks corresponded with any of the 364 targets, while approximately 70% (9550 in positive mode, 9610 in negative mode) were nontarget peaks.

These were grouped into 6700 and 7100 components, respectively. The majority of the components per sample were single monoisotopic peaks (i.e., neither isotope nor adduct peaks were associated with this m/z), generally of relatively low intensity and thus low priority for further identification efforts. The remaining 1400 components (in each mode) had isotope or adduct information (or both) available. The isotope breakdown gives the number of components with the given isotope signal present.

The discrepancy between the ^{34}S isotope signals detected in positive (13) and negative (123) modes in Table 1 indicated the potential presence of sulfonic acids and similar compounds, which ionize very well in negative ESI mode.

Target Results and Prioritization. The detailed target results, with quantitative and semiquantitative concentrations for compounds detected in at least one sample are given in SI Table S3 and summarized in SI Figure S2 using all concentrations detected in all samples, while the highest concentration targets detected in at least 4 samples are shown in SI Figure S3. Pharmaceuticals, corrosion inhibitors, and food additives (sweeteners) were present in high concentrations, but still within the range of reported concentrations from a recent EU-wide survey of effluents covering 18 countries.[36]

Although hundreds of targets were analyzed and many were present in very high concentrations, few targets were among the highest intensity compounds detected in all samples after prioritization, showing the need to complement target approaches with suspect and nontarget screening. Figure 1 shows the top 30 components by intensity for negative mode and includes only 4 target compounds (labeled by name), while

Table 2. Summary of Evidence for Homologous Series Suspects, Used in Tentative Identification (Details in SI)^a

series	no.	alkyl homologues	average highest I	RT range (min)	chro.	RT/I patt.	MS/MS avail.	MS/MS match	tent. ID
negative									
LAS	10	$C_{16-20}H_{26-34}O_3S$	4.4×10^6	17.7–25.0	Y	Y	Y	Y	Y
SPAC	10	$C_{9-21}H_{10-34}O_5S$	6.2×10^6	3.8–16.2	Y	Y	Y	Y	Y
SPADC	10	$C_{9-22}H_{8-34}O_7S$	1.9×10^6	4.2–12.9	?	N	Y	Y(1)	Y(1)
DATS	10	$C_{12-20}H_{16-32}O_3S$	6.1×10^6	9.4–21.6	Y	Y	Y	Y	Y
STAC	10	$C_{12-22}H_{14-34}O_5S$	2.8×10^6	5.6–14.9	Y	Y	Y	Y	Y
STADC	10	$C_{12-16}H_{12-20}O_7S$	9.2×10^4	3.9–5.4	N	N	Y	Y(1)	Y(1)
AS	10	$C_{12-16}H_{26-34}O_4S$	3.8×10^6	16.4–22.4	N	N	Y	?	N
SAS	10	$C_{10-16}H_{22-34}O_3S$	1.4×10^6	12.5–24.0	Y	Y	Y	Y	Y
C12-AES	10	$C_{14-32}H_{30-66}O_{5-14}S$	1.2×10^6	19.8–21.7	Y	Y	Y	Y	Y
C13-AES	10	$C_{15-25}H_{32-52}O_{5-10}S$	1.5×10^5	18.5–22.8	Y	Y	Y	Y	Y
O/NPEC	10	$C_{16-19}H_{24-30}O_{3-4}$	1.0×10^6	14.7–17.2	N/A	Y	Y	Y	Y
NPEC-S	10	$C_{17-31}H_{28-56}O_{5-12}S$	9.7×10^4	14.0–19.5	Y	Y	N	N/A	?
positive									
PEGs	10	$C_{6-28}H_{14-58}O_{4-15}$	3.7×10^6	1.6–6.7	Y	Y	N	N/A	Y
SPAC	8	$C_{14-19}H_{20-30}O_5S$	8.7×10^4	8.4–13.6	Y	Y	N	N/A	?
Cx-AES	10	$C_{15-32}H_{32-66}O_{5-12}S$	5.0×10^5	4.7–12.5	?	N	N	N/A	N
CxAEOx	10	$C_{14-42}H_{30-86}O_{2-15}$	3.9×10^6	7.4–17.7	Y	Y	N	N/A	N
CDEAs	10	$C_{12-22}H_{25-45}NO_3$	2.0×10^6	5.7–17.4	N	N	N	N/A	N
NPEOx	10	$C_{17-47}H_{28-88}O_{2-17}$	1.0×10^5	13.7–18.3	N	N	N	N/A	N

^{a}Y: yes, N: no, N/A: not applicable, ?: match unclear, (1): only applies to one suspect.

only two were present in the top 30 from positive mode (DEET and 4-acetylaminoantipyrine, see SI Figure S4). Half of the components in Figure 1 contained sulfur (acesulfame, saccharin, cyclamate, and the 12 components marked yellow and orange). The dominance of sulfur-containing compounds in negative measurements, which was highlighted by the isotopic fine-structure grouping performed using the "non-target" R package, directed subsequent efforts at suspect (surfactant) and nontarget identification, as outlined below.

Suspect Screening. MS/MS information acquired for the most intense sulfur-containing components indicated the presence of sulfonic compounds, with distinctive fragments at $m/z = 79.9574$ (SO_3^-) and $m/z = 183.0121$ $(C_8H_7SO_3^-).$[22,29] This directed a literature search and subsequent suspect screening. A full list of the 394 suspects screened is given in SI Table S6 and included 5 Linear Alkylbenzyl Sulfonates (LAS), 13 SulfoPhenyl Alkyl Carboxylic acids (SPACs), 15 SulfoPhenyl Alkyl Di-Carboxylic acids (SPADCs), 16 Di-Alkyl Tetralin Sulfonates (DATS), 12 Sulfo-Tetralin Alkyl Carboxylic acids (STACs), 16 Sulfo-Tetralin Alkyl Di-Carboxylic acids (STADCs), 5 Alkyl Sulfates (AS), 60 Alkyl Ethoxy Sulfates (AES), 12 Secondary Alkyl Sulfonates (SAS), and 15 Nonyl Phenol EthOxylate (NPEO) sulfates (NPEO-S). All 394 suspects were screened for the presence of $[M + H]^+$ (positive) and $[M - H]^-$ (negative mode) species; other species associated with the suspects were detected via the componentization in the "nontarget" package (see Table 1). In negative mode, 173 suspects were detected in at least one sample, while 240 were detected in positive mode. However, *a priori* it is quite likely that many of these are false positives (i.e., arising from peaks with coincidently the same mass). Moschet et al.[5] estimated the false positive rate for pesticide suspect screening at 30−50% after extensive filtering steps including an intensity cutoff, but a much higher false positive rate of over 89% before applying the filtering procedure. Thus, evidence to support (or reject) the tentative suspect identification was gathered and summarized for the homologous series screened in Table 2. Full details for each of these series, tables of masses and

retention times, as well as chromatograms and MS/MS spectra (where available) are given in the SI, Tables S7−S22 and Figures S4−S24. Tentatively identified and annotated spectra are included in NORMAN MassBank[37] in the Eawag Tentative Spectra (ETS00001 to ETS00021) database, along with the literature spectra used in the Literature Spectra database (LIT00001 to LIT00039).

Using the SPACs as an example, 13 suspect SPAC masses (SPA-3C to SPA-15C; note the number preceding the C indicates the alkyl carbons, not total C) were screened and detected in all 10 samples. An annotated MS/MS spectrum of SPA-9C and the extracted ion chromatograms for all SPAC masses are given in Figure 2. The retention times were generally very consistent over the samples (< ± 0.3 min for SPA-4C to SPA-14C) and increased with the number of carbons. The increasing number of carbons was supported by the accurate mass and the increasing relative intensity of the ^{13}C peak in the HRMS scan. The most intense homologues were SPA-6C to SPA-9C, in accordance with literature.[22] The chromatographic shape was also consistent across the homologous series. The LAS, parent compounds for the SPACs, were detected at higher retention times, as expected and were confirmed in the samples. The MS/MS spectrum of SPA-9C matched the literature spectrum of 6ϕ-SPA9C[38] (LIT00035 in MassBank) with a match value of 87%. Additional peaks in the MS/MS measured in this study could be reconciled to the structure with MOLGEN-MS/MS as indicated in Figure 2, or are considered noise peaks. Finally, the presence of SPACs in the samples was confirmed using reference standards for single homologue isomers of SPA-8C and SPA-10C, provided by J. Field (see SI, Figures S25−S33 for LAS and SPAC confirmation).

Similarly, multiple lines of evidence existed to tentatively identify LAS, DATS, STACS, SAS, and AES (primarily the C12-AES) series. Although the SPADC and STADC masses demonstrated very inconsistent mass traces, one member of each series could be tentatively identified using the MS/MS. The AS and SAS series were possibly present, but had very
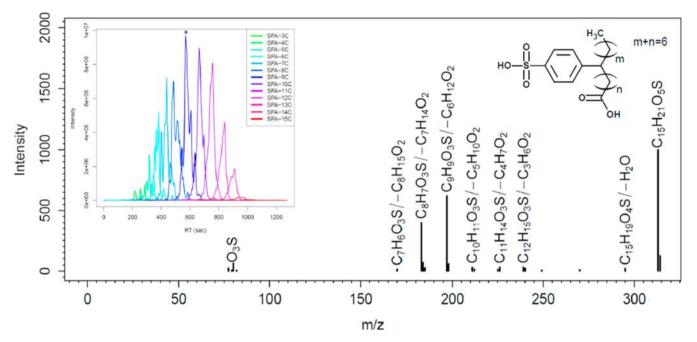
1815

dx.doi.org/10.1021/es4044374 | *Environ. Sci. Technol.* 2014, 48, 1811−1818

Figure 2. Annotated HCD 60%NCE (high collision energy dissociation, % normalized collision energy) MS/MS spectrum of SPA-9C homologues from one sample. Inset: Extracted ion chromatograms of the tentatively identified SPACs, ± 10 ppm, negative mode, for one sample; * indicates the SPA-9C peak. MS/MS peak intensities for $m/z < 300$ are multiplied by 10.

Table 3. Candidates for Nontarget m/z = 213.9637 and Associated Data

Score	0.924	0.924	0.75	0.924	0.75	0.858	0.858
References	38	29	1	2	2	0	0
Standard	No	Yes	No	No	No	No	No
Match		MSMS, RT					

uninformative mass spectra (e.g., only a SO_3 loss) and while the SAS showed a good chromatographic pattern, the AS did not. The NPEO-S compounds were of insufficient intensity for tentative identification, while two nonyl and octyl-phenol ethoxy-carboxylates were tentatively identified using MS/MS. For the positive homologous series, PEG masses were present in all samples, but with interferences in the MS/MS, while the SPACs showed weak traces in all samples at approximately the same retention times as the negative signals. The other series could not be tentatively identified and were not present with sufficient intensity to record MS/MS in such complex samples. Thus, 61 of the 173 suspects detected in negative mode (35%) but only 8 of the 240 suspects in positive mode (3%) could be considered tentatively identified using additional evidence. This was mainly due to the very high intensities and good MS/MS for the negative suspects; very few clean MS/MS spectra could be obtained for the positive suspects. Despite the amount of evidence present in many cases, the identities of these compounds cannot be confirmed unequivocally without reference standards, which are not easy to obtain for these compounds. A pure C12-LAS and a mixed C10−C14-LAS standard matched the LAS suspects retention pattern and MS/MS reasonably well (see SI Figures S25−S29). Although the peaks in the samples were broader and with higher intensities for the lower LAS isomers, this is likely due to the presence of additional isomers and transformation processes during treatment, consistent with the literature.[22]

Nontarget Identification. Nontarget identification was performed on selected masses from the top 30 most intense peaks. For example, the HR-MS information for m/z = 213.9637 revealed the presence of N, S and C in the formula, yielding an unambiguous formula assignment $C_7H_5NO_3S_2$ with MOLGEN-MS/MS. Two fragments were present in the MS/MS, corresponding with a loss of SO_2 (~10% intensity) and SO_3 (50% intensity) with $[M − H]^-$ as the base peak. Only 7 candidates were retrieved from ChemSpider via MetFusion using the molecular formula and are shown in Table 3, along with the MetFusion score and number of references. A standard was purchased for the second structure, 1,3-benzothiazole-2-sulfonate (given in Figure 1), which confirmed the tentative identification via matching RT and MS/MS spectra. This compound is an oxidation product of the high production volume vulcanization accelerator 2-mercaptobenzothiazole, which has been reported previously in industrial wastewaters.[39]

The results were varying for the remaining 18 nontarget masses in Figure 1. Formulas could be assigned unambiguously

to 9 of these 18 masses with MOLGEN-MS/MS and used for candidate searching with MetFusion, while the rest were searched using an exact mass window of 5 ppm. The number of possible candidates ranged from 7 to 3477, while 4 actually had zero candidates. These 4 nontargets eluted very early in the chromatographic regime (i.e., they are essentially unretained), and it is possible that these are organometallic compounds or salt clusters. Although we had considered using a retention time cutoff in the prioritization to avoid this (e.g., eliminating all compounds with a retention time below 1.5 min), some target compounds also elute early, such as iomeprol and iopamidol (both ~1.3 min). With regard to the identification, MS/MS was available for 15 of the 18 nontarget masses, but for two masses no candidates appeared to match the data. Subsequent measurements of the samples concentrated on obtaining MS/MS information for species with isotope information and suspect compounds, to increase the chances of identification success. There were no clear top candidates with standards available for further confirmation of the compounds in Figure 1 at this stage. These results reinforce the continuing need to develop additional identification methods for nontarget compounds.[10,40] Nontarget identification remains biased toward heteroatom-containing compounds. Although HR-MS data extend this to S- and N-containing compounds rather than only halogens, much work remains to be done to improve the identification success of nontarget methods.

Implications. A wealth of information is available in HR-MS(/MS) data, and this study demonstrates that such data can be used to perform rapid suspect screening to complement target analysis. In the end, the 4 targets identified originally in the top 30 negative mode peaks were complimented with 7 suspects and one nontarget, such that 12 compounds were tentatively or fully identified, a 3-fold improvement above target analysis alone. Thus, suspect screening and nontarget identification efforts are essential to capture and identify high concentration and potentially environmentally relevant compounds. The dominance of surfactants in wastewater effluents is clear, and it is shown that homologous series with many members can be screened and tentatively identified relatively quickly using software tailored for HR-MS data and available online.[16−18] Despite the extensive screening performed, intense peaks remained unidentified. The sharing of mass spectrometric data of target and suspect (tentatively identified) compounds via open spectral libraries such as MassBank would ensure rapid progress in the identification of compounds detected frequently in different laboratories, although attention should be given to ensuring that high quality data are uploaded.[8]

ASSOCIATED CONTENT

Ⓢ Supporting Information

Additional tables (Tables S1−S22), figures (Figures S1−S34) and text to support the experimental and results detailing sample locations, target compounds, parameters for software, prioritization, suspect confirmation, and additional references. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +41 58 765 5493; fax: +41 58 765 5893; e-mail: juliane.hollender@eawag.ch.

Notes

ACKNOWLEDGMENTS

REFERENCES

(1) Schwarzenbach, R. P.; Escher, B. I.; Fenner, K.; Hofstetter, T. B.; Johnson, C. A.; von Gunten, U.; Wehrli, B. The challenge of micropollutants in aquatic systems. *Science* **2006**, *313* (5790), 1072−1077.

(2) Escher, B. I.; Fenner, K. Recent advances in environmental risk assessment of transformation products. *Environ. Sci. Technol.* **2011**, *45* (9), 3835−3847.

(3) 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* (3), 943−951.

(4) Hernandez, F.; Pozo, O. J.; Sancho, J. V.; Lopez, F. J.; Marin, J. M.; Ibanez, M. Strategies for quantification and confirmation of multi-class polar pesticides and transformation products in water by LC-MS2 using triple quadrupole and hybrid quadrupole time-of-flight analyzers. *Trends Anal. Chem.* **2005**, *24* (7), 596−612.

(5) 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.

(6) Hoh, E.; Dodder, N. G.; Lehotay, S. J.; Pangallo, K. C.; Reddy, C. M.; Maruya, K. A. Nontargeted comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry method and software for inventorying persistent and bioaccumulative contaminants in marine environments. *Environ. Sci. Technol.* **2012**, *46* (15), 8001−8008.

(7) Stein, S. Mass spectral reference libraries: An ever-expanding resource for chemical identification. *Anal. Chem.* **2012**, *84* (17), 7274−7282.

(8) Stravs, M. A.; Schymanski, E. L.; Singer, H. P.; Hollender, J. Automatic recalibration and processing of tandem mass spectra using formula annotation. *J. Mass Spectrom.* **2013**, *48* (1), 89−99.

(9) Hernandez, F.; Sancho, J. V.; Ibanez, M.; Abad, E.; Portoles, T.; Mattioli, L. Current use of high-resolution mass spectrometry in the environmental sciences. *Anal. Bioanal. Chem.* **2012**, *403* (5), 1251−1264.

(10) 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.

(11) Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinform.* **2006**, *7*, 234−244.

(12) Gonsior, M.; Zwartjes, M.; Cooper, W. J.; Song, W. H.; Ishida, K. P.; Tseng, L. Y.; Jeung, M. K.; Rosso, D.; Hertkorn, N.; Schmitt-Kopplin, P. Molecular characterization of effluent organic matter identified by ultrahigh resolution mass spectrometry. *Water Res.* **2011**, *45* (9), 2943−2953.

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

(14) Royal Society of Chemistry: ChemSpider www.chemspider.com (accessed 19/12/2013)

(15) 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.

(16) Loos, M.; Ruff, M.; Singer, H. enviMass version 1.2 target screening software, http://www.eawag.ch/forschung/uchem/software/enviMass1_2. (accessed 20/12/2013).

(17) Loos, M. R Package nontarget: Detecting, combining and filtering isotope, adduct and homologue series relations in high-resolution mass spectrometry (HRMS) data. http://www.eawag.ch/forschung/uchem/software/R_package_start. (accessed 20/12/2013).

(18) Stravs, M. A.; Schymanski, E. L. RMassBank package http://www.eawag.ch/forschung/uchem/software/RMassBank. (accessed 20/12/2013).

(19) Kern, S.; Fenner, K.; Singer, H. P.; Schwarzenbach, R. P.; Hollender, J. Identification of transformation products of organic contaminants in natural waters by computer-aided prediction and high-resolution mass spectrometry. *Environ. Sci. Technol.* **2009**, *43* (18), 7039−7046.

(20) Singer, H.; Jaus, S.; Hanke, I.; Luck, A.; Hollender, J.; Alder, A. C. Determination of biocides and pesticides by on-line solid phase extraction coupled with mass spectrometry and their behaviour in wastewater and surface water. *Environ. Pollut.* **2010**, *158* (10), 3054−3064.

(21) Götz, C. W.; Stamm, C.; Fenner, K.; Singer, H.; Scharer, M.; Hollender, J. Targeting aquatic microcontaminants for monitoring: exposure categorization and application to the Swiss situation. *Environ. Sci. Pollut. Res. Int.* **2010**, *17* (2), 341−354.

(22) Lara-Martin, P. A.; Gonzalez-Mazo, E.; Brownawell, B. J. Multi-residue 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.

(23) Corada-Fernandez, C.; Lara-Martin, P. A.; Candela, L.; Gonzalez-Mazo, E. Tracking sewage derived contamination in riverine settings by analysis of synthetic surfactants. *J. Environ. Monit.* **2011**, *13* (7), 2010−2017.

(24) Lara-Martin, P. A.; Corada-Fernandez, C.; Gomez-Parra, A.; Gonzalez-Mazo, E. Seasonal changes in the concentration of anionic surfactants in estuarine sediments from the River Guadalete (Cadiz, Spain). *Sci. Mar.* **2010**, *74*, 125−131.

(25) Gonzalez, S.; Petrovic, M.; Radetic, M.; Jovancic, P.; Ilic, V.; Barcelo, D. Characterization and quantitative analysis of surfactants in textile wastewater by liquid chromatography/quadrupole-time-of-flight mass spectrometry. *Rapid Commun. Mass Spec.* **2008**, *22* (10), 1445−1454.

(26) Di Corcia, A.; Costantino, A.; Crescenzi, C.; Marinoni, E.; Samperi, R. Characterization of recalcitrant intermediates from biotransformation of the branched alkyl side chain of nonylphenol ethoxylate surfactants. *Environ. Sci. Technol.* **1998**, *32* (16), 2401−2409.

(27) Di Corcia, A.; Crescenzi, C.; Marcomini, A.; Samperi, R. Liquid chromatography electrospray mass spectrometry as a valuable tool for characterizing biodegradation intermediates of branched alcohol ethoxylate surfactants. *Environ. Sci. Technol.* **1998**, *32* (5), 711−718.

(28) Di Corcia, A.; Capuani, L.; Casassa, F.; Marcomini, A.; Samperi, R. Fate of linear alkyl benzenesulfonates, coproducts, and their metabolites in sewage treatment plants and in receiving river waters. *Environ. Sci. Technol.* **1999**, *33* (22), 4119−4125.

(29) Di Corcia, A.; Casassa, F.; Crescenzi, C.; Marcomini, A.; Samperi, R. Investigation of the fate of linear alkyl benzenesulfonates and coproducts in a laboratory biodegradation test by using liquid chromatography/mass spectrometry. *Environ. Sci. Technol.* **1999**, *33* (22), 4112−4118.

(30) Field, J. A.; Field, T. M.; Poiger, T.; Giger, W. Determination of secondary alkane sulfonates in sewage wastewaters by solid-phase extraction and injection-port derivatization gas-chromatography mass-spectrometry. *Environ. Sci. Technol.* **1994**, *28* (3), 497−503.

(31) Field, J. A.; Field, T. M.; Poiger, T.; Siegrist, H.; Giger, W. Fate of secondary alkane sulfonate surfactants during municipal waste-water treatment. *Water Res.* **1995**, *29* (5), 1301−1307.

(32) Castillo, M.; Riu, J.; Ventura, F.; Boleda, R.; Scheding, R.; Schroder, H. F.; Nistor, C.; Emneus, J.; Eichhorn, P.; Knepper, T. P.; Jonkers, C. C. A.; de Voogt, P.; Gonzalez-Mazo, E.; Leon, V. M.; Barcelo, D. Inter-laboratory comparison of liquid chromatographic techniques and enzyme-linked immunosorbent assay for the determination of surfactants in wastewaters. *J. Chromatogr. A* **2000**, *889* (1−2), 195−209.

(33) 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.

(34) Meringer, M.; Schymanski, E. L. Small molecule identification with MOLGEN and mass spectrometry. *Metabolites* **2013**, *3* (2), 440−462.

(35) Gerlich, M.; Neumann, S. MetFusion: Integration of compound identication strategies. *J. Mass Spectrom.* **2013**, *48* (3), 291−298.

(36) Loos, R.; Carvalho, R.; Antonio, D. C.; Comero, S.; Locoro, G.; Tavazzi, S.; Paracchini, B.; Ghiani, M.; Lettieri, T.; Blaha, L.; Jarosova, B.; Voorspoels, S.; Servaes, K.; Haglund, P.; Fick, J.; Lindberg, R. H.; Schwesig, D.; Gawlik, B. M. EU-wide monitoring survey on emerging polar organic contaminants in wastewater treatment plant effluents. *Water Res.* **2013**, *47* (17), 6475−6487.

(37) NORMAN Association, NORMAN MassBank, www.massbank.eu. (accessed 20/12/2013).

(38) Lara-Martin, P. A.; Gomez-Parra, A.; Sanz, J. L.; Gonzalez-Mazo, E. Anaerobic degradation pathway of linear alkylbenzene sulfonates (LAS) in sulfate-reducing marine sediments. *Environ. Sci. Technol.* **2010**, *44* (5), 1670−1676.

(39) De Wever, H.; Besse, P.; Verachtert, H. Microbial trans-formations of 2-substituted benzothiazoles. *Appl. Microbiol. Biot.* **2001**, *57* (5−6), 620−625.

(40) Schymanski, E. L.; Neumann, S. CASMI: And the winner is. *Metabolites* **2013**, *3* (2), 412−439.

■ **NOTE ADDED AFTER ASAP PUBLICATION**

This article published January 14, 2014 without the last author affiliation. The correct version published January 15, 2014.