



NORMAN digital sample freezing platform: A European virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in “digitally frozen” environmental samples

Nikiforos A. Alygizakis^{a, b, *}, Peter Oswald^a, Nikolaos S. Thomaidis^{b, **},
Emma L. Schymanski^{c, d}, Reza Aalizadeh^b, Tobias Schulze^e, Martina Oswaldova^a,
Jaroslav Slobodnik^a

^a Environmental Institute, Okružná 784/42, 97241 Koš, Slovak Republic

^b Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Panepistimiopolis Zografou, 15771 Athens, Greece

^c Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland

^d Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6, avenue du Swing, L-4367 Belvaux, Luxembourg

^e UFZ Helmholtz Centre for Environmental Research GmbH, Permoserstraße 15, 04318 Leipzig, Germany

ARTICLE INFO

Article history:

Available online 12 April 2019

Keywords:

Digital sample freezing platform
High resolution mass spectrometry
Digital archiving
Suspect screening
Antibiotics
REACH
Black sea

ABSTRACT

A platform for archiving liquid chromatography high-resolution mass spectrometry (LC-HRMS) data was developed for the retrospective suspect screening of thousands of environmental pollutants with the ambition of becoming a European and possibly global standard. It was termed Digital Sample Freezing Platform (DSFP) and incorporates all the recent developments in the HRMS screening methods within the NORMAN Network. In the workflow, raw mass spectral data are converted into mzML, then mass spectral and chromatographic information on thousands of peaks of each sample is extracted into Data Collection Templates. The 'digitally frozen' samples can be retrospectively screened for the presence of virtually any compound amenable to LC–MS using a combination of information on its (i) exact mass, (ii) predicted retention time window in the chromatogram, (iii) isotopic fit and (iv) qualifier fragment ions. Its potential was demonstrated on monitoring of 670 antibiotics and 777 REACH chemicals from the Joint Black Sea Surveys (JBSS).

© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Tens of thousands of chemical substances are produced in Europe and worldwide in large amounts with potential to enter the environment [1,2]. Many of these substances and their transformation products are potentially toxic to flora, fauna and humans, but the scientific tools to establish a critical mass of evidence for this ever increasing chemical impact to support the respective legal

regulatory tools are not yet sufficiently in place [2,3]. The NORMAN network, as an interface between science and policy-making, has been working to establish a retrospective system able to detect any contaminants of environmental concern (CECs) that may be harmful to environmental or human health for over a decade [4]. In 2017, over 70 NORMAN members in Europe and North America decided to expand from considering only hundreds to tens of thousands substances in their activities. A domain of NORMAN suspected pollutants was established [5], evidence of their occurrence in the environment was collected [6], and a toxicity threshold was assigned to each of the substances [7]. This information was intended for use in identifying/prioritising compounds (exceeding the toxic threshold values at many sites) that should be considered for regulation in Europe [8].

* Corresponding author. Environmental Institute, Okružná 784/42, 97241 Koš, Slovak Republic.

** Corresponding author.

E-mail addresses: alygizakis@ei.sk (N.A. Alygizakis), ntho@chem.uoa.gr (N.S. Thomaidis).

HRMS instruments such as quadrupole-time-of-flight (Q-TOF) and hybrid quadrupole or ion trap Orbitraps acquire accurate mass and high-resolution MS and MS/MS (or in some cases MSⁿ) full-scan spectra and can be used to perform comparative sample evaluation [9,10]. In the margins of the analytical conditions (e.g. sampling, enrichment method and solvents used) and instrumental limitations (e.g. ionizability, selectivity, sensitivity and resolution), the full mass spectral information of detectable compounds is stored in raw data files. However often the measurements are only evaluated partially (e.g. for the targeted analysis of a given list of CECs) and most of the detected peaks remain unannotated and thus unknown [11–13]. One main reason why stored HRMS data are still underexploited is the lack of software tools for data archiving, quality control and exchange [2,14,15]. This reduces the potential for use of such data in regulation and thus limits the general ability to perform in-depth investigations into environmental contamination. The second main reason is due to a general lack of large LC–HRMS/MS mass spectral libraries and the inability of current libraries to cover all suspected CECs and their transformation products [16]. The NORMAN MassBank (<https://massbank.eu/MassBank>), established by the NORMAN network in 2011 [17], is currently populated with 48,822 experimental mass spectra on more than 10,000 substances [18], while other public HRMS libraries include up to a few tens of thousands substances, with some overlap among them [16,19]. The potential of a public mass spectral platform for raw data to search for the non-regulated emerging substances through the use of retrospective suspect screening with high-resolution mass spectrometry was first tested in NormaNEWS pilot study [20,21], in which laboratories around Europe were asked to check a pre-defined list of newly-identified CECs (https://comptox.epa.gov/dashboard/chemical_lists/normanews) in mass chromatograms of environmental samples stored locally in participating laboratories.

In this critical discussion article, the NORMAN Digital Sample Freezing Platform (DSFP) is presented for archiving, processing, analysing, data mining and retrieving information from the large amount of environmental mass spectral information derived by the community of environmental scientists and deposited at NORMAN. DSFP incorporates all the latest developments in HRMS screening [2] and offers an integrated tool for wide-scope screening of CECs in the environment. The primary intended uses are retrospective analysis of newly emerging substances and comparison of mass spectral data across compartments (e.g. water, biota, sediment, indoor environment, air), however additional uses and exploration of other potential applications are strongly encouraged. For example, the outcomes can be used to indicate the occurrence and spatial distribution of a certain substance within a geographical region (country, river basin, pan-European scale, etc.) or to prioritise unidentified features or compounds for future identification efforts. DSFP promotes automation of retrospective screening, enhances the transparency of LC–HRMS data and serves as a tool for drafting future policy recommendations related to chemicals management in the environment. It was tested using samples from the Joint Black Sea Survey (JBSS) covering seawater, sediment and biota matrices [22]. All JBSS samples were investigated retrospectively for a list of 670 antibiotics (list S6, ITNAN-TIBIOTIC, at <https://www.norman-network.com/?q=node/236>). Antibiotics were selected due to their specific adverse effects on flora and fauna, which are exhibited even at low concentrations [23,24]. Their presence in the environment may also induce antibiotic resistance [25,26], which has been identified as one of major global threats to the society [27,28]. Additionally, samples from the JBSS were screened for 777 compounds extracted from the REACH registry in cooperation with the European Chemicals Agency (ECHA) in 2017.

2. Databases

2.1. Candidate substances

The NORMAN Network (www.norman-network.com) is a unique network of reference laboratories, research centres and related organisations for monitoring of CECs in the environment. A working group (WG) on non-target screening (NTS) was established in 2015 with the aim to promote and integrate all the research efforts in NTS of environmental samples within Europe and beyond. One of the main efforts of the NTS WG was the development of DSFP. DSFP was designed for archiving LC–HRMS chromatograms and for the retrospective screening of polar and semi-polar CECs in various environmental matrices. Such methods should include a generic sample preparation methods, such as those presented for wastewater [29], sludge and sediments [30] and biota [31]. A recent collaborative trial revealed that suspect screening was a very common and efficient way to expedite non-target screening [9]. As a result, the NORMAN Suspect Exchange initiative was founded (<http://www.norman-network.com/?q=node/236>) and members were encouraged to submit their suspect lists. To date, 40,053 highly varying substances have been collected and curated. All preparation and curation was done within the network using open access cheminformatics toolkits and formed one merged list (SusDat) [5]. Antibiotics and their major transformation products (TPs) is part of SusDat and the compounds were mined from the literature, compound databases and chemical catalogues. A list of the reviewed sources can be found in Table S2 (SI). The list of 670 antibiotics (termed “ITNAN-TIBIOTIC”) is accessible online at NORMAN Suspect List Exchange [32], which contains environmentally relevant substances collected and contributed by NORMAN members. The literature review showed that only a small fraction of the collected antibiotics (66 out of 670) was mentioned in the literature (Table S2), whereas for the rest of antibiotics occurrence data in environmental samples is rather scarce (Fig. S2A and B). The spectral libraries mzCloud [33] and NORMAN MassBank [17,18] were queried for these substances. Experimental mass spectra, including information on preferable ionization of the compound, screened adduct and qualifier fragment ions were found for 159 antibiotics. The rest of substances were screened assuming the presence of $[M + H]^+$ and $[M - H]^-$ ions using *in silico* predicted fragments calculated using CFM-ID [34].

2.2. Data collection templates

Each sample was stored digitally in DSFP as a separate Excel spreadsheet-based “Data Collection Template” (DCT), which is linked to one or more raw mass spectral files (mzML file format) according to the number of different HRMS methods applied for analysis of the sample. Each DCT consists of six sections; organization details, retention time index (RTI) information, sample description, analysis, instrumental metadata and fragment peak list (https://norman-data.eu/DCT_NTS.xlsx). This DCT was used to report NTS results from reference laboratories participating in collaborative trials organised by NORMAN in 2015 and 2016 [9,35]. DCT includes important sample information organised for each environmental matrix and crucial instrumental parameters. Analysis of sample extracts by LC–HRMS result in binary data, which have different structure and format, so that files are accessible for processing by the respective vendor's software. To allow for interoperability among results obtained by instruments from different vendors, files are converted to a commonly agreed format (mzML) [36] by converters, among which the most widespread is ProteoWizard [37]. For exporting JBSS chromatograms in mzML format in

this study, CompassXport 3.0.9.2. (Bruker Daltonics, Bremen, Germany) was used.

2.3. Repository

To acquire suitable LC–HRMS data and use DSFP to its full potential, the procedure described in Section S3 (SI) should be followed. The procedure involves actions to be taken before, during and after the instrumental analysis. For future import of new samples, contributors should follow the same scheme and, while the choice of mobile phases, gradients and reversed-phase columns can be according to their protocol, they should (i) assure that their equipment is clean and well-calibrated, (ii) inject RTI mixture during the sequence (Fig. S3–5, SI), and (iii) analyse sample extracts in data-independent and data-dependent modes, as explained in Fig. S3–6 (SI) for optimal results.

The R-based workflow for importing new samples into DSFP is represented in Fig. 1. LC–HRMS chromatograms converted to mzML can be processed with community driven workflows [36]. There are many workflows available, but the final output is always a component list [38–40]. DSFP uses the centWave algorithm (through XCMS R-package v. 3.4.1) for peak picking with previously optimized *ppm* and *peakwidth* parameters [41,42]. The peak picking algorithm searches for consecutive masses within a mass error threshold forming peak shape in chromatographic dimension [39]. The next step is componentization, which is a procedure for grouping peaks coming from the same compound (adducts, isotopic peaks, in-source fragments) [43–45]. For this purpose, functions from nontarget R-package (v. 1.9) were used [43].

The next step involves calculation of an experimental RTI for every feature detected based on the retention time observed and the calibration curve equation ($RT = f(RTI)$), derived from the retention time of the standard calibration mixture [46]. This is followed by the extraction of HRMS/MS fragments that were isolated and fragmented, in case the sample was analysed using a data-dependent method. Finally, the output is shaped to the DCT format after adding organization details, information on sampling site/date/matrix and instrumental metadata. The chromatograms

are contributed via a web interface (Fig. S3–4 (SI)3 (SI)), which automatically generates a DCT for each sample and facilitates the upload of the respective sample-specific mzML files to the server in a harmonised format ready for processing.

3. Overview of the screening process

Once DCTs and mzML files are imported into DSFP, the user may search these and/or other samples for the presence of a single substance (Fig. S4–1, SI) or for many compounds included in SusDat (40,053 suspected CECs as of November 2018) [5]. Individual searches for unknown compounds based on their exact mass are also available. SusDat contains important information for the screening of CECs in DSFP (exact mass of a molecular adduct ions $[M + H]^+$ and $[M - H]^-$, predicted RTI) supplemented by masses of experimentally observed or predicted fragment ions. If no experimental fragments were available in MassBank for a given compound, then *in silico* predicted fragments were used instead, calculated with CMF-ID [34]. The fragmentation pattern of all NORMAN MassBank compounds is integrated to SusDat (list S1 “MASSBANK” at <https://www.norman-nework.com/?q=node/236>). Therefore, the platform automatically suggests to search for the exact mass of a compound using a preferred specific ionization mode (positive or negative or both) and an expected adduct (like $[M + H]^+$, $[M + NH_4]^+$, $[M + Na]^+$ in positive ionization mode and $[M - H]^-$, $[2M - H]^-$ in negative ionization mode) according to the available experimental data. For tentative identification (indicated by summary of identification evidences), the platform considers mass accuracy, plausible (window of) retention time in the chromatogram, isotopic fit and a presence of matching fragments. If both high-energy collision dissociation (HCD) and collision-induced dissociation (CID) fragments at different collision energies are available for a compound, DSFP will adapt to the closest acquisition conditions used for the sample chromatogram. Unknown substances not included in SusDat can also be searched for occurrence over multiple samples by exact mass. In all cases, the user must specify the mass accuracy error, the tolerance in

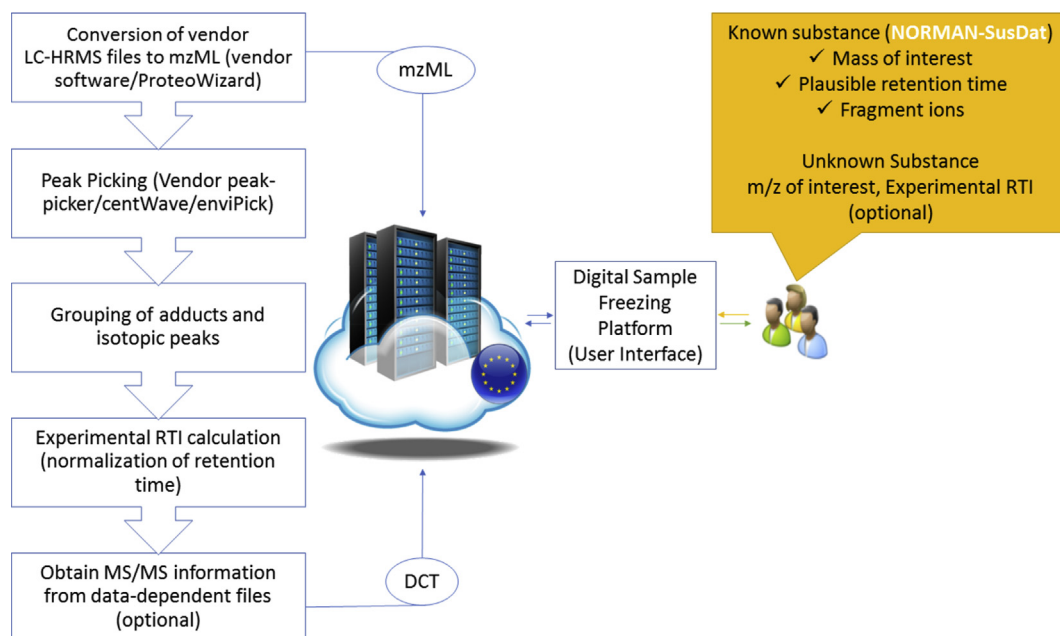


Fig. 1. Adopted workflow for obtaining harmonised mzML raw data formats (provided by users) and automatically generated Data Collection Templates (DCTs) accessible to users through the Digital Sample Freezing Platform (DSFP) interface.

plausibility of retention time (by default values are proposed) and fragments.

Compound querying is a rather straightforward procedure, in which user selects (i) a compound of interest from a drop-down list of SusDat substances and (ii) samples in which the compound should be searched. Compound selection is possible using any of the following identifiers: common name, CAS number or InChIKey (Fig. S4-2, SI). Mass chromatograms can be filtered down to those desired to be submitted for investigation based on the country, matrix type and project (Fig. S4-3, SI). Then a search can start, in which the exact mass of the ionized form of the substance is searched for a match in the DCTs of all selected mass chromatograms. Features that pass mass accuracy and fit into the expected RTI window are presented in an interactive and downloadable table (Fig. S4-4, SI). A column with detected fragment ions is presented in a format “exact mass/retention time/absolute maximum intensity”, which allows for a quick check whether the proposed identification is plausible (Fig. S4-4, SI). DSFP offers the possibility to change parameters such as mass accuracy tolerance or RTI tolerance (Fig. S4-1b, SI). As an example, RTI tolerance could be set to 100% in situations where one does not wish to consider the RTI values and instead rely only on exact mass ions-based identification.

From the results table, it is possible to go back to the raw mass chromatograms stored in the repository and perform extracted ion chromatogram (EIC) search and/or have a look at the full scan spectra of the selected chromatographic peak (Fig. 2). DSFP will automatically retrieve fragments for compounds with experimental spectral information available in SusDat (2219 compounds as of December 2018) and add them in an interactive table to help users verify the identity of the compound they searched. For example, in Fig. 2 the identity of the antibiotic sulfamethoxazole was supported by the presence of molecular ion adduct and two matching fragments. The EIC table is interactive, i.e. rows with different exact masses of interest can be added manually and thus visualized on the screen, whereas mass accuracy and mass chromatograms selected for screening can be changed according to the choice of the user.

4. Investigation of spatial distribution of detected compounds

The results of the search can also be visualized in an interactive map (Fig. 3). Observed intensities of the compounds are normalized based on the maximum observed intensity over all samples. Moreover, the enrichment factor is also considered in case the sample was enriched, e.g. by solid phase extraction (this information is mandatory during the contribution procedure). Finally, the intensities of a compound are shown on the map as scaled circles; the bigger the size and the more intense the colour - the higher the signal and presumably concentration of the substance. The scaled mapping of the normalized intensities allows for a user-friendly visual assessment of the spatial distribution of substances of interest and possible sources of pollution.

Fig. 3 shows the example of DEET, the spatial distribution of which suggests inputs from various diffuse sources. This visualization method is suitable for results obtained for samples using the same sample preparation and analytical conditions and coming from the same instrument but may not be accurate for comparing samples coming from different instruments. Nevertheless, even in this case, it will give a quick rough overview on the presence of investigated chemicals in different locations. As a further example, in Fig. S4-5(SI), the highest normalized intensities of sulfamethoxazole were detected in three Black Sea sampling stations close to the Danube delta during JBSS 2016 [22], suggesting the Danube as a pollution source. This is in line with observation obtained in the

Joint Danube Survey 3, conducted earlier in 2013, where sulfamethoxazole was reported at relatively high concentration levels 15 and 16 ng L⁻¹ in sampling stations Sf. Gheorghe arm and Vilkovo, respectively, in the areas close to the above Black Sea sampling stations [47].

5. Batch mode module and interactive heatmap visualization

The batch mode module provides the possibility to search for up to thousands of compounds included in SusDat in a single batch in all or a selected number of samples stored in DSFP. Again, the user specifies compounds and samples to be investigated and obtains detailed and summarized results by a single click of ‘Submit’ button (Fig. S4-7, SI). DSFP returns a summary spreadsheet file containing absolute maximum intensity of the observed signals, experimental retention time, mass error/accuracy (in mDa and ppm) and information on detected fragments including exact mass/retention time/absolute maximum intensity (Figs. S5a–e). For each analyte there is a common name, molecular formula, CAS No., SMILES, InChI, InChIKey (all retrieved from SusDat), a column with the identification evidence (i.e. mass accuracy, isotopic fit, plausibility of retention time and number of fragments) and a column whether fragments are predicted *in silico* or obtained experimentally (extracted from available HRMS libraries). A detailed report is provided in a format of multi-sheet spreadsheet file, in which each sheet represents one compound with the same content as obtained from the single compound search (Fig. S5-f). It is encouraged that batch-mode requests include procedural blank and quality control samples, so that the user can subtract and evaluate the instrumental signals that happen unintentionally because of contamination during sample preparation or instrumental analysis. The total processing time depends on the number of selected compounds, samples and computational power. Typical processing time for screening of 1000 suspects in 86 JBSS samples was 13 ± 1 min for an Intel® Core™ i7-4702MQ CPU processor at 2.20 GHz.

The batch mode tool is equipped with its separate interactive graphical presentation tool – a heatmap, such as the one presented in Fig. 4 and Fig. S5-2 (SI). Here, the selected compounds are presented in rows and samples in columns. White colour in the heatmap means that compound was not detected in the sample, whereas blue means positive detection. In a simplified scoring system, compounds that satisfy the mass accuracy criterium and have a plausible retention time (observed within the RTI window) receive one point and one additional point for each matching fragment ion. As shown in Fig. S5-2 (SI), the user can customize which compounds appear in the heatmap (compounds with predicted or experimental fragments, compounds that exceed a specified score), as well as the appearance of heatmap.

In this study, an identification was considered as having substantial supporting evidence if three or more fragments were detected for compounds with library spectra available (unless available library spectra contained less fragments) and if five or more fragments were detected for compounds with *in silico* predicted mass spectra, in addition a match of exact mass of the molecular ion and a plausible RTI. The number of fragments is critical to distinguish between false positive and false negative identifications. It should be made clear here that DSFP is providing supporting evidence for tentative identification and does not aim at assigning identification levels. While all proposed identifications remain technically as a Level 3 [48], those with substantial supporting evidence are clearly higher confidence as those with only an exact mass match. All plausible identifications should be verified with an exact library match (to obtain a Level 2a status) or confirmation of retention time and fragment information with a

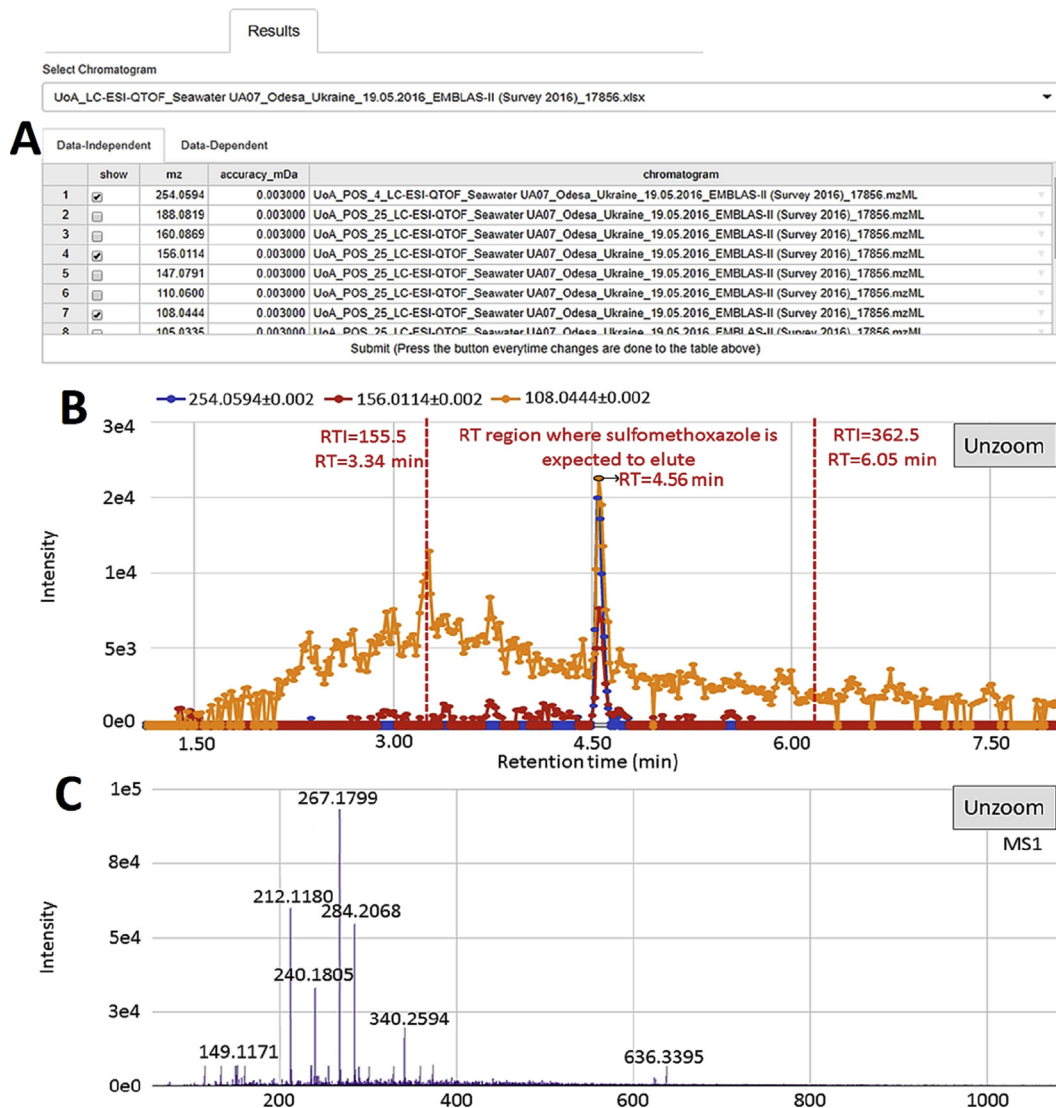


Fig. 2. (A) Editable fragment list retrieved automatically for sulfamethoxazole (B) Extracted ion chromatograms (m/z [M + H]⁺ 254.0504, fragments from MassBank m/z 156.0144, 108.0444) indicating the presence of sulfamethoxazole in the Black Sea water sample from the area close to the delta of the Danube river. The molecular ion and two structure characteristic fragments at high collision energy (25 eV) support the identification of the structure, the RTI window from 155.5 to 362.5 serves as complementary supporting evidence (Level 3; [42]). The compound was eventually confirmed with a reference standard (Level 1). (C) Full-scan spectrum generated by clicking at the maximum data point of the peak with retention time 4.56 min.

reference standard for Level 1. While DSFP is not able to perform this directly at this stage, instead DSFP offers the users an opportunity to further explore the data by going back to raw mass chromatograms by performing extracted ion chromatograms and by visualizing the MSⁿ spectra (e.g. as shown in Fig. 2). DSFP is not designed to solve problems of identifying structural isomers, as it is an inherent disadvantage of the current state-of-the-art LC–HRMS instrumentation. Structural isomers may have a similar retention time and common fragments. It is the responsibility of the user to further evaluate the output using raw mass spectra based on other evidence such as ion ratio, or searching for diagnostic ions using expert knowledge.

6. Screening of antibiotics in the Black Sea samples

Before the application of DSFP screening in real case studies, the results were benchmarked against the results of a target screening method. The JBSS samples were subjected to wide-scope target screening of 2248 compounds (list “UATHTARGETS” [32,49]). The

validation results are summarized in section S6 (SI). DSFP was used to screen antibiotics in JBSS samples using the “ITNANTIBIOTIC” list. Twelve out of 670 suspect antibiotics were detected. Following further evaluation of DSFP results, nine compounds were confirmed by analysis of reference standard at level 1, three were identified at level 2a (probable structure by library spectrum match) (Table S6) [48]. Antibiotics with the highest frequency of detection were sulfamethoxazole (44 out of 86 samples, mainly in seawater), aminosalicylic acid (34 out of 86 JBSS samples), chlorhexidine (15 out of 86 samples, mainly in sediment) and 8-hydroxyquinolin (11 out of 86 samples, mainly in biota), while the other nine antibiotics detected sporadically were mainly found in seawater samples close to the Danube delta. Macrolide monensin, griseofulvin, lincomycin and the sulfonamides sulfadiazine and sulfapyridine, all confirmed at level 1, were detected in the sample from the Ukrainian shelf (Sampling station UA07; [22]) and some of them even at stations more distant from the Danube delta (Sampling stations UA06, UA05). Sulfadiazine was also detected on the

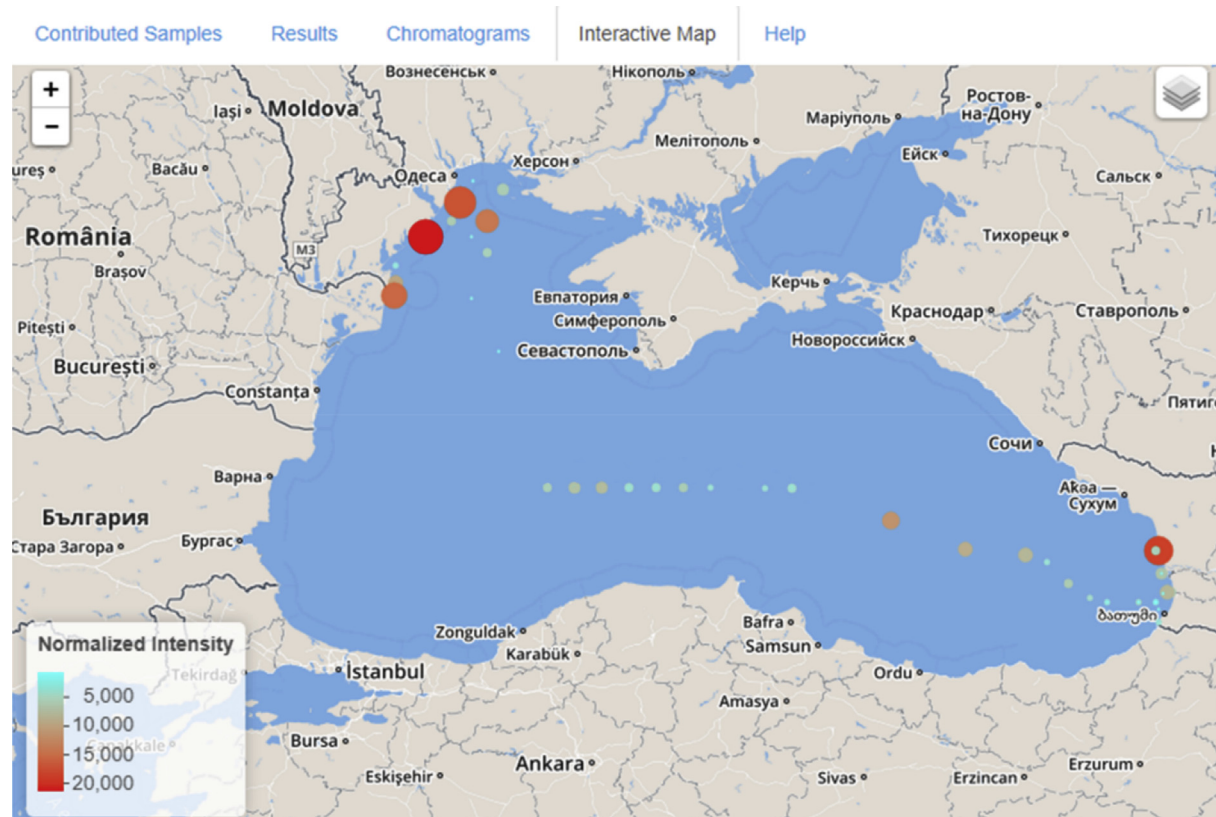


Fig. 3. Spatial distribution of DEET in seawater samples from the Black Sea [17]. The profile reveals that there is land input of DEET to the Black Sea. DEET was detected in 39 (out of 54) water samples analysed by LC–HRMS and using the NTS workflow (for details, see Methodology). The circles indicate sampling stations; small, light blue circles relate to low normalized intensities; large red circles indicate high normalized intensities of DEET.

Georgian coastline together with fluconazole, the presence of which was confirmed from mass chromatograms obtained in both positive and negative ionization (Fig. 4).

Chlorhexidine was detected in 15 out of 19 sediment samples, including samples taken from the seabed of the Black Sea at a depth of more than 2000 m, confirming its widespread occurrence and persistence. Another compound with widespread occurrence was aminosalicic acid (level 1), which was detected in seven out of 19 samples in all investigated matrices. 8-hydroxychinolin, aminosalicic acid and sulfamethoxazole were detected in biota samples. 8-hydroxychinolin has a wide range of applications and a tendency for bioaccumulation and persistence in the environment [50]. The presence of this compound in all tested biota sample should be further explored. Aminosalicic acid, an antibiotic primarily used to treat tuberculosis, was detected in five out of twelve biota samples. The presence of this substance in all three investigated matrices (seawater, sediment and biota) is potentially alarming. Sulfamethoxazole was detected also in one biota sample and its presence in the marine ecosystem deserves attention. Only 12 out of 670 antibiotics were detected in the investigated samples. However, occurrence of these antibiotics in such a diluted matrix as sea water and their accumulation in sediments and biota far from the sources of pollution should draw the attention of both ecologists and regulators.

7. Screening of REACH compounds in the Black Sea samples

Out of ca. 68,000 substances that can be found on the official website of ECHA containing registered REACH compounds (<https://echa.europa.eu/information-on-chemicals/registered-substances>),

only 777 compounds had experimentally observed mass spectra stored in the NORMAN MassBank. The lack of coverage of REACH chemicals in MassBank highlights the need for further support for the development of HRMS libraries. The JBSS samples were screened for the presence of these compounds using DSFP in batch mode. 80 out of the 777 substances were detected. The relatively high detection rate (10.3%) indicates that the screened compounds were of widespread use.

Industrial chemicals such as phthalate esters and phosphates, pharmaceuticals (phenazone and its TP, carbamazepine, fenbendazole) and pesticides (atrazine, terbutylazine, chloridazon, ametryn, metolachlor, simazine, diuron) appeared to be the most widely detected CECs in the samples. 60 out of 80 detected CECs were monitored by wide-scope target screening and their occurrence, spatial distribution and risk assessment are discussed in detail in the final scientific report of the JBSS 2016 [22]. The remaining 20 CECs included industrial chemicals such as phosphates and phthalate esters (triisobutyl phosphate, tris(2-butoxyethyl) phosphate (TBEP), dicyclohexyl phthalate), surfactants (N,N-bis(2-hydroxyethyl)dodecanamide, lauric isopropanolamide), and industrial intermediates (benzenesulfonamide, N-butylbenzenesulphonamide) and pharmaceuticals-antimicrobial substances (8-hydroxychinolin).

The occurrence of these 20 compounds is represented in the SI, section S7. Triisobutyl phosphate, with an annual production tonnage of 1000–10,000 t/a, was detected in all sediment samples and in 45 out of 55 seawater samples, while it was not detected in any biota sample. The spatial distribution of this CEC indicates input from Ukrainian and Georgian shores to the Black Sea (Fig. S4–6, SI). The plasticizer N-butylbenzenesulphonamide, produced in the same

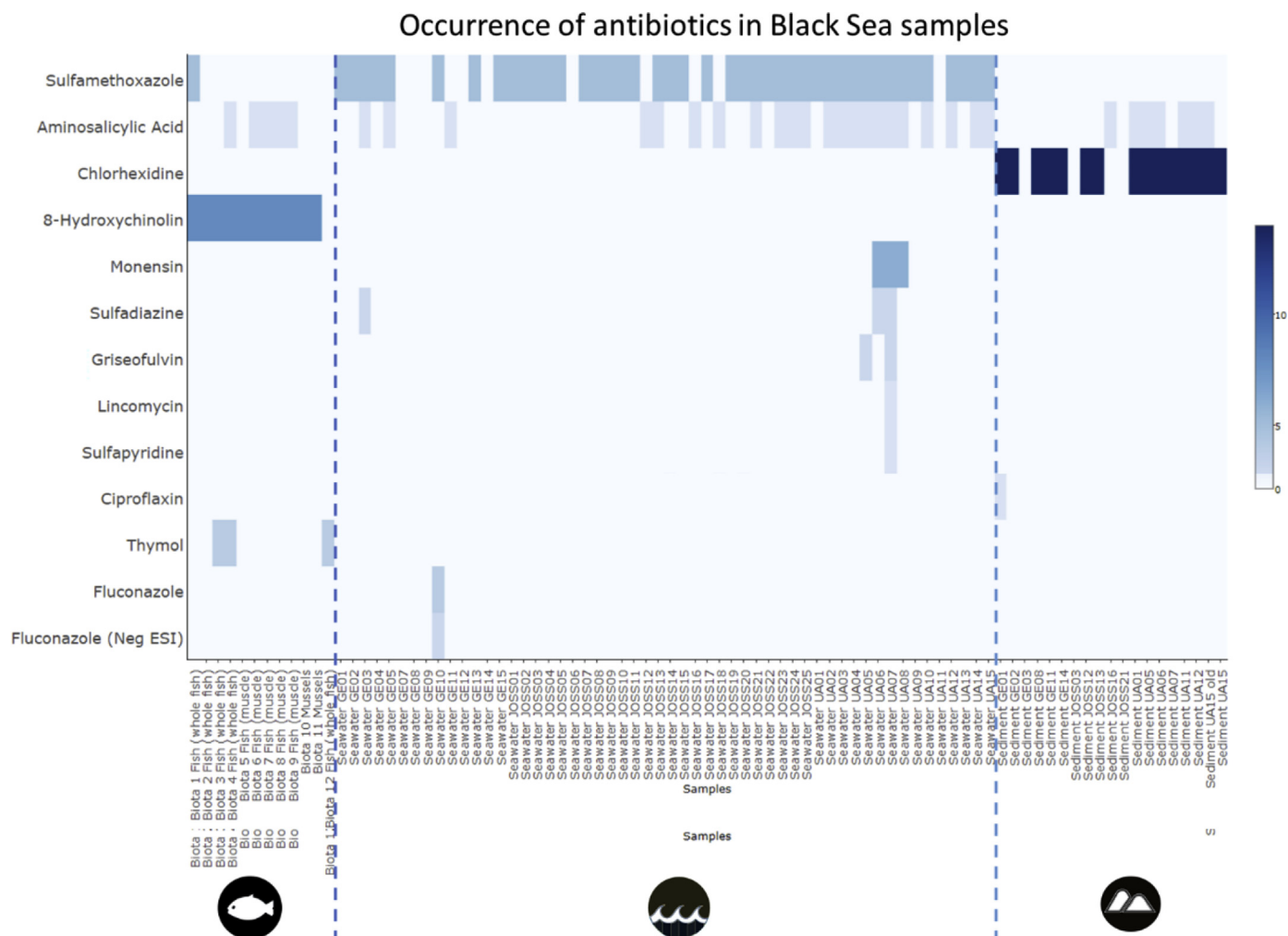


Fig. 4. Occurrence of antibiotics in the Black Sea samples collected within the JBSS 2016 [22]. Blue coloured rectangles indicate that a compound was detected in the sample; the darker the colour, the more structure-related fragments were detected in the corresponding mass spectra. The first 12 samples (x-axis) are fish and molluscs, the following 54 samples are seawater followed by 19 sediments samples. The bar on the right-hand side indicates colour code related to the score of identification. Compounds are ordered by increasing frequency of appearance. For more details, see text.

tonnage range, was detected in 13 out of 55 seawater and in 17 out of 19 sediment samples. TBEP, produced in the same tonnage range, was detected in 7 out of 55 seawater and 10 out of 19 sediment samples. Finally, the surfactant lauric isopropanolamide was detected in 32 seawater samples, whereas its presence was not observed in biota or sediment samples.

8. Conclusions and future developments

An open, integrated, interactive and intuitive platform for archiving, processing and data mining from a large amount of LC–HRMS data produced by the environmental community was developed and thoroughly tested. The platform allows for the retrospective suspect screening of the presence of tens of thousands of CECs and their transformation products in environmental samples in a systematic and consistent way, with a goal of becoming a European and possibly global standard. The platform integrates tools for storing raw HRMS chromatograms of samples, each containing typically several thousands of compounds, in a uniform mzML format independent from vendor software. Both single substance and batch queries are possible across selected or all of the samples stored in the platform.

The results of the NTS workflow used in DSFP were validated against the outcomes of the target screening of 2248 substances in the same samples. The compounds identified by both approaches overlapped in 97% of cases for seawater, 94% for biota and 106% (more compounds detected by DSFP) for biota samples. The applicability and the potential use of DSFP was demonstrated in the screening of 670 antibiotics and 777 REACH chemicals in Black Sea seawater, sediment and biota samples. Thus, DSFP incorporates all the state-of-the-art developments of HRMS screening methodologies.

Continuous improvements of the data processing capabilities of the platform with the addition of modules for trend analysis, elemental analysis, mass defect analysis etc. is a priority. Next steps agreed within the NORMAN network will aim to collect a critical mass of raw mass chromatograms so as to improve the spatial coverage in all environmental compartments for comprehensive screening of the presence of major polluting compounds across Europe and beyond. Moreover, the application of the updated NORMAN NTS prioritisation algorithm will help to identify the most relevant suspect and unknown substances to be further investigated for unambiguous elucidation of their identity or evaluations as substances in need of regulatory measures. Finally,

the integration of GC-APCI-HRMS and GC-EI-HRMS data is in progress as a significant upgrade towards a unified global platform for storing, viewing and screening of environmental pollutants in a much wider analytical window.

At present only those organisations and researchers contributing their data can use the platform. It is planned to open DSFP to the public after a thorough testing of all of its functionalities with a reference set of 'big data' and when a critical mass of samples is achieved. Any organization can become a member of the NORMAN network according to its Statutes. Researchers willing to join the activities of the network are invited to contribute to further development of DSFP and participate in network activities, such as collaborative trials, workshops, expert group meetings etc. (see e.g. <https://www.norman-network.net/> 'Upcoming events'). DSFP contains unique datasets which will allow environmentalists to change the current paradigm, which is mainly based on tracing of individual environmental pollutants when they become regulated, to simultaneous screening and retrospective assessment of knowns and unknowns in complex matrices. In the medium-term, the NORMAN network will attempt to maintain and manage DSFP as a harmonised platform at the European level. NORMAN operates in close contact with the European Commission via DG ENV and its services, including ECHA and JRC. NORMAN experts involved in the European project on human biomonitoring (HBM4EU; <https://www.hbm4eu.eu/>) will ensure full harmonisation and interoperability with the standards that are going to be developed for chemical screening of human samples. Screening for the presence of all relevant pollutants across all environmental compartments and their potential occurrence in humans are recognised as a priority.

Disclaimer

The content of this article reflects only the authors' views and the Research Executive Agency is not responsible for any use that may be made of the information it contains.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 675530 and from the NORMAN Association. Authors would like to acknowledge EU/UNDP EMBLAS-II project ENPI/2013/313-169 (<http://emblasproject.org/>) for providing the samples from the Joint Black Sea Survey 2016. Authors would like to acknowledge Maria-Christina Nika, Aikaterini Psoma and Dimitrios Damalas for sample preparation and analysis of the samples from JBSS. Authors would like to acknowledge the contribution of Valeria Dulio. ELS is supported by the Luxembourg National Research Fund (FNR) for ATTRACT project 12341006.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.trac.2019.04.008>.

References

[1] European Commission, Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission

Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC OJ L 396 of 30 December 2006, Off. J. Eur. Union 136 (2007) 3–280.

[2] J. Hollender, E.L. Schymanski, H.P. Singer, P.L. Ferguson, Nontarget screening with high resolution mass spectrometry in the environment: ready to go? *Environ. Sci. Technol.* 51 (2017) 11505–11512.

[3] W. Brack, V. Dulio, M. Agerstrand, I. Allan, R. Altenburger, M. Brinkmann, D. Bunke, R.M. Burgess, I. Cousins, B.I. Escher, F.J. Hernandez, L.M. Hewitt, K. Hilscherova, J. Hollender, H. Hollert, R. Kase, B. Klauer, C. Lindim, D.L. Herraes, C. Miede, J. Munthe, S. O'Toole, L. Posthuma, H. Rudel, R.B. Schafer, M. Sengl, F. Smedes, D. van de Meent, P.J. van den Brink, J. van Gils, A.P. van Wezel, A.D. Vethaak, E. Vermeirssen, P.C. von der Ohe, B. Vrana, Towards the review of the European Union Water Framework Directive: recommendations for more efficient assessment and management of chemical contamination in European surface water resources, *Sci. Total Environ.* 576 (2017) 720–737.

[4] V. Dulio, B. van Bavel, E. Brorström-Lundén, J. Harmsen, J. Hollender, M. Schlabach, J. Slobodnik, K. Thomas, J. Koschorreck, Emerging pollutants in the EU: 10 years of NORMAN in support of environmental policies and regulations, *Environ. Sci.* 30 (2018).

[5] NORMAN SusDat. <https://www.norman-network.com/nds/susdat/>, 2018 (accessed 10.12.18).

[6] NORMAN EMPDAT, Database of Geo-Referenced Monitoring and Bio-Monitoring Data on Emerging Substances in Water, Sediments, Biota, SPM, Soil, Sewage Sludge and Air, 2018. <https://www.norman-network.com/nds/empdat/> (accessed 10.12.18).

[7] NORMAN, Ecotoxicology Database, 2018. <https://www.norman-network.com/nds/ecotox/> (accessed 10.12.18).

[8] V. Dulio, J. Slobodnik, In response: the NORMAN perspectives on prioritization of emerging pollutants, *Environ. Toxicol. Chem.* 34 (2015) 2183–2185.

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

[10] Y. Verkh, M. Rozman, M. Petrovic, A non-targeted high-resolution mass spectrometry data analysis of dissolved organic matter in wastewater treatment, *Chemosphere* 200 (2018) 397–404.

[11] B.Y.L. Peisl, E.L. Schymanski, P. Wilmes, Dark matter in host-microbiome metabolomics: tackling the unknowns – a review, *Anal. Chim. Acta* 1037 (2018) 13–27.

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

[13] A.C. Chiaia-Hernandez, E.L. Schymanski, P. Kumar, H.P. Singer, J. Hollender, Suspect and nontarget screening approaches to identify organic contaminant records in lake sediments, *Anal. Bioanal. Chem.* 406 (2014) 7323–7335.

[14] K. Haug, R.M. Salek, P. Conesa, J. Hastings, P. de Matos, M. Rijnbeek, T. Mahendrakar, M. Williams, S. Neumann, P. Rocca-Serra, E. Maguire, A. Gonzalez-Beltran, S.A. Sansone, J.L. Griffin, C. Steinbeck, MetaboLights – an open-access general-purpose repository for metabolomics studies and associated meta-data, *Nucleic Acids Res.* 41 (2013) D781–D786.

[15] S. Samanipour, K. Langford, M.J. Reid, K.V. Thomas, A two stage algorithm for target and suspect analysis of produced water via gas chromatography coupled with high resolution time of flight mass spectrometry, *J. Chromatogr. A* 1463 (2016) 153–161.

[16] M. Vinaixa, E.L. Schymanski, S. Neumann, M. Navarro, R.M. Salek, O. Yanes, Mass spectral databases for LC/MS- and GC/MS-based metabolomics: state of the field and future prospects, *Trends Anal. Chem.* 78 (2016) 23–35.

[17] T. Schulze, E.L. Schymanski, M. Stravs, S. Neumann, M. Krauss, H. Singer, C. Hug, G. Gallampos, J. Hollender, J. Slobodnik, B. Werner, NORMAN MassBank, Towards a community-driven, open-access accurate mass spectral database for the identification of emerging pollutants, *NORMAN Netw Bull* 3 (2012) 9–10.

[18] NORMAN MassBank. <http://massbank.eu/MassBank/>, 2018 (accessed 10.12.18).

[19] C. Frainay, E.L. Schymanski, S. Neumann, B. Merlet, R.M. Salek, F. Jourdan, O. Yanes, Mind the gap: mapping mass spectral databases in genome-scale metabolic networks reveals poorly covered areas, *Metabolites* 8 (2018).

[20] N.A. Alygizakis, S. Samanipour, J. Hollender, M. Ibañez, S. Kaserzon, V. Kokkali, J. Van Leerdam, J.F. Mueller, M. Pijnappels, R.M.J.E.L. Schymanski, J. Slobodnik, N.S. Thomaidis, K.V. Thomas, Exploring the potential of a global emerging contaminant early warning network through the use of retrospective suspect screening with high-resolution mass spectrometry, *Environ. Sci. Technol.* 52 (2018) 5135–5144.

[21] NormaNEWS, The Norman Early Warning System – a New Retrospective Analysis Exchange Platform for Laboratories Active in Non-target Screening, 2018. <http://www.normadata.eu/?q=node/244> and https://comptox.epa.gov/dashboard/chemical_lists/normanews (accessed 10.12.18).

[22] J. Slobodnik, B. Alexandrov, V. Komorin, A. Mikaelyan, A. Guchmanidze, M. Arabidze, A. Korshenko, S. Moncheva (Editors), National Pilot Monitoring Studies and Joint Open Sea Surveys in Georgia, Russian Federation and Ukraine, 2016. Final Scientific Report, EU/UNDP Project: Improving Environmental Monitoring in the Black Sea – Phase II (EMBLAS-II) ENPI/2013/313-

- 169, http://emblasproject.org/wp-content/uploads/2017/04/Joint-Black-Sea-Surveys-2016_16.pdf (accessed 10.12.18).
- [23] G. McEneff, L. Barron, B. Kelleher, B. Paull, B. Quinn, A year-long study of the spatial occurrence and relative distribution of pharmaceutical residues in sewage effluent, receiving marine waters and marine bivalves, *Sci. Total Environ.* 476–477 (2014) 317–326.
- [24] N.A. Alygizakis, P. Gago-Ferrero, V.L. Borova, A. Pavlidou, I. Hatzianestis, N.S. Thomaidis, Occurrence and spatial distribution of 158 pharmaceuticals, drugs of abuse and related metabolites in offshore seawater, *Sci. Total Environ.* 541 (2016) 1097–1105.
- [25] B. Berglund, Environmental dissemination of antibiotic resistance genes and correlation to anthropogenic contamination with antibiotics, *Infect. Ecol. Epidemiol.* 5 (2015) 28564.
- [26] C.M. Manaia, J. Rocha, N. Scaccia, R. Marano, E. Radu, F. Biancullo, F. Cerqueira, G. Fortunato, I.C. Iakovides, I. Zammit, I. Kampouris, I. Vaz-Moreira, O.C. Nunes, Antibiotic resistance in wastewater treatment plants: tackling the black box, *Environ. Int.* 115 (2018) 312–324.
- [27] World Health Organization, Antibiotic Resistance – a Threat to Global Health Security, 2018. http://www.who.int/drugresistance/activities/wha66_side_event/en/ (accessed 10.12.18).
- [28] B. Piña, J.M. Bayona, A. Christou, D. Fatta-Kassinos, E. Guillon, D. Lambropoulou, C. Michael, F. Polesel, S. Sayen, On the contribution of reclaimed wastewater irrigation to the potential exposure of humans to antibiotics, antibiotic resistant bacteria and antibiotic resistance genes – NEREUS COST Action ES1403 position paper, *J. Environ. Chem. Eng.* (2018). <https://doi.org/10.1016/j.jece.2018.01.011>.
- [29] S. Kern, K. Fenner, H. Singer, R. Schwarzenbach, J. Hollender, Identification of transformation products of organic contaminants in natural waters by computer-aided prediction and high-resolution mass spectrometry, *Environ. Sci. Technol.* 43 (2009) 7039–7046.
- [30] P. Gago-Ferrero, V. Borova, M.E. Dasenaki, N.S. Thomaidis, Simultaneous determination of 148 pharmaceuticals and illicit drugs in sewage sludge based on ultrasound-assisted extraction and liquid chromatography–tandem mass spectrometry, *Anal. Bioanal. Chem.* 407 (2015) 4287–4297.
- [31] M.E. Dasenaki, N.S. Thomaidis, Multi-residue determination of 115 veterinary drugs and pharmaceutical residues in milk powder, butter, fish tissue and eggs using liquid chromatography–tandem mass spectrometry, *Anal. Chim. Acta* 880 (2015) 103–121.
- [32] NORMAN Suspect List Exchange, 2018. <https://www.norman-network.com/?q=node/236> (accessed 10.12.18).
- [33] mzCloud Mass Spectral Database, HighChem, 2018. <https://mzcloud.org/> (accessed 10.12.18).
- [34] F. Allen, R. Greiner, D. Wishart, Competitive fragmentation modeling of ESI–MS/MS spectra for putative metabolite identification, *Metabolomics* 11 (2014) 98–110.
- [35] P. Rostkowski, P. Haglund, R. Aalizadeh, A.N. Alygizakis, N.S. Thomaidis, J. Beltran Arandes, P. Booi, H. Budzinski, A. Covaci, S. Grosse, R. Hindle, I.M. Ipolyi, K. Jobst, S. Kaserzon, The strength in numbers: comprehensive characterization of house dust using complementary mass spectrometric techniques, *Anal. Bioanal. Chem.* 411 (2019) 1957–1977.
- [36] L. Martens, M. Chambers, M. Sturm, D. Kessner, F. Levander, J. Shofstahl, W.H. Tang, A. Römpp, S. Neumann, A.D. Pizarro, L. Montecchi-Palazzi, N. Tasman, M. Coleman, F. Reisinger, P. Souda, H. Hermjakob, P.-A. Binzg, E.W. Deutsch, mzML—a community standard for mass spectrometry data, *Mol. Cell. Proteomics* 10 (2011).
- [37] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egerton, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T.A. Baker, M.Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S.L. Seymour, L.M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E.W. Deutsch, R.L. Moritz, J.E. Katz, D.B. Agus, M. MacCoss, D.L. Tabb, P. Mallick, A cross-platform toolkit for mass spectrometry and proteomics, *Nat. Biotechnol.* 30 (2012) 918–920.
- [38] M. Katajamaa, J. Miettinen, M. Oresic, MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data, *Bioinformatics* 22 (2006) 634–636.
- [39] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Sluzdak, XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, *Anal. Chem.* 78 (2006) 779–787.
- [40] M. Sturm, A. Bertsch, C. Gropl, A. Hildebrandt, R. Hussong, E. Lange, N. Pfeifer, O. Schulz-Triegelaff, A. Zerck, K. Reinert, O. Kohlbacher, OpenMS – an open-source software framework for mass spectrometry, *BMC Bioinformatics* 9 (2008) 163.
- [41] G. Libiseller, M. Dvorzak, U. Kleb, E. Gander, T. Eisenberg, F. Madeo, S. Neumann, G. Trausinger, F. Sinner, T. Pieber, C. Magnes, IPO: a tool for automated optimization of XCMS parameters, *BMC Bioinformatics* 16 (2015) 118.
- [42] R. Tautenhahn, C. Bottcher, S. Neumann, Highly sensitive feature detection for high resolution LC/MS, *BMC Bioinformatics* 9 (2008) 504.
- [43] M. Loos, nontarget: Detecting Isotope, Adduct and Homologue Relations in LC–MS Data, R package version 1.9, 2016, <https://CRAN.R-project.org/package=nontarget>.
- [44] C.D. Broeckling, F.A. Afsar, S. Neumann, A. Ben-Hur, J.E. Prenni, RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data, *Anal. Chem.* 86 (2014) 6812–6817.
- [45] C. Kuhl, R. Tautenhahn, C. Bottcher, T.R. Larson, S. Neumann, CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets, *Anal. Chem.* 84 (2012) 283–289.
- [46] R. Aalizadeh, et al., Development and prediction of liquid chromatographic retention time indices, in preparation.
- [47] I. Liška, F. Wagner, M. Sengl, K. Deutsch, J. Slobodnik (Editors), A Comprehensive Analysis of Danube Water Quality. Joint Danube Survey 3, ICPDR – International Commission for the Protection of the Danube River, Vienna, Austria, 2015. http://www.danubesurvey.org/jds-files/nodes/documents/jds3_final_scientific_report_1.pdf.
- [48] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying small molecules via high resolution mass spectrometry: communicating confidence, *Environ. Sci. Technol.* 48 (2014) 2097–2098.
- [49] P. Gago-Ferrero, A. Bletsou, D.E. Damalas, R. Aalizadeh, N. Alygizakis, H. Singer, J. Hollender, N.S. Thomaidis, Wide-scope quantitative target screening of 2248 emerging contaminants in wastewater samples with UPLC–Q-ToF–HRMS/MS, *Anal. Bioanal. Chem.* (2019). submitted.
- [50] S. Yan, L. Chen, X. Dou, M. Qi, Q. Du, Q. He, M. Nan, Z. Chang, P. Nan, Toxicity of 8-hydroxyquinoline in *Cyprinus carpio* using the acute toxicity test, hepatase activity analysis and the comet assay, *Bull. Environ. Contam. Toxicol.* 95 (2015) 171–176.