Environment International 170 (2022) 107585

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Full length article

Enhancing spectral quality in complex environmental matrices: Supporting suspect and non-target screening in zebra mussels with ion mobility

Frank Menger [a,*], Alberto Celma [b], Emma L. Schymanski [c], Foon Yin Lai [a], Lubertus Bijlsma [b], Karin Wiberg [a], Félix Hernández [b], Juan V. Sancho [b], Lutz Ahrens [a,*]

[a] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences (SLU), SE-75007 Uppsala, Sweden
[b] Environmental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water, University Jaume I, Avda. Sos Baynat s/n, E-12071 Castellón, Spain
[c] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6, Avenue du Swing, L-4367 Belvaux, Luxembourg

ARTICLE INFO

Handling Editor: Olga Kalantzi

Keywords:
Polar organic chemical integrative sampler (POCIS)
Mobility alignment
Collision cross-section (CCS)
Mass spectrum quality

ABSTRACT

Identification of bioaccumulating contaminants of emerging concern (CECs) via suspect and non-target screening remains a challenging task. In this study, ion mobility separation with high-resolution mass spectrometry (IM-HRMS) was used to investigate the effects of drift time (DT) alignment on spectrum quality and peak annotation for screening of CECs in complex sample matrices using data independent acquisition (DIA). Data treatment approaches (Binary Sample Comparison) and prioritisation strategies (Halogen Match, co-occurrence of features in biota and the water phase) were explored in a case study on zebra mussel (*Dreissena polymorpha*) in Lake Mälaren, Sweden's largest drinking water reservoir. DT alignment evidently improved the fragment spectrum quality by increasing the similarity score to reference spectra from on average (±standard deviation) 0.33 ± 0.31 to 0.64 ± 0.30 points, thus positively influencing structure elucidation efforts. Thirty-two features were tentatively identified at confidence level 3 or higher using MetFrag coupled with the new PubChemLite database, which included predicted collision cross-section values from CCSbase. The implementation of predicted mobility data was found to support compound annotation. This study illustrates a quantitative assessment of the benefits of IM-HRMS on spectral quality, which will enhance the performance of future screening studies of CECs in complex environmental matrices.

1. Introduction

High-resolution mass spectrometry (HRMS) is now recognized as a key technology for the identification of unknown chemicals. In environmental science, HRMS has led to the successful identification of different contaminants of emerging concern (CECs) in the aquatic environment, *e.g.*, pesticides, pharmaceuticals, illicit drugs and their metabolites/transformation products (Kiefer et al., 2019; Moschet et al., 2014; Tian et al., 2020; Wang et al., 2020; Fabregat-Safont et al., 2021; Hernández et al., 2019), industrial chemicals (Schlüsener et al., 2015), novel per- and polyfluoroalkyl substances (PFAS) (Barrett et al., 2021; Jacob et al., 2021; Wang et al., 2018) and a highly toxic storm water contaminant (quinone transformation product) (Peter et al., 2018; Tian et al., 2021). Software and instrumental capabilities in the field have rapidly improved in recent years, enabling large-scale screening of CECs (Gago-Ferrero et al., 2020; Hollender et al., 2017). Sample preparation steps should be mild and kept at a minimum in non-target screening studies to avoid losses of potential compounds of interest (Baduel et al., 2015; Dürig et al., 2020). However, this becomes a challenge for samples that have complex matrix composition, such as biota, where sample clean-up is required to minimize matrix interferences. Hence, careful prioritisation and critical manual reviewing often become necessary (Dom et al., 2018), which makes data treatment and peak annotation arduous and time-consuming. These challenges are especially pronounced for data acquired at low concentrations and in data independent acquisition (DIA) modes, where multiple compounds often co-elute and contribute to one fragment spectrum.

Ion mobility (IM) separation hyphenated to HRMS (IM-HRMS) is an instrumental setup that has the potential to improve CEC identification performance compared to HRMS instruments, especially for samples with complex matrices (D'Atri et al., 2018; Belova et al., 2021; Stephan et al., 2016). In IMS, ions pass through an inert gas under an electrical

* Corresponding authors at: Box 7050, SE-750 07 Uppsala, Sweden.
E-mail addresses: frank.menger@slu.se (F. Menger), lutz.ahrens@slu.se (L. Ahrens).

https://doi.org/10.1016/j.envint.2022.107585
Received 24 June 2022; Received in revised form 11 October 2022; Accepted 13 October 2022
Available online 14 October 2022

field and are separated based on their size, shape and charge. The resulting separation parameter, the drift time (DT), can be used to align mass spectra and to determine the collision cross section (CCS) of each ion. DT alignment groups signals according to their DT and, as a result, provides cleaner spectra since interferences from co-eluting compounds with different DT are removed (Celma et al., 2020). Since DT adds additional complexity to the HRMS data, open-source solutions for DT alignment are limited. DT alignment performed in vendor software (*e.g.* UNIFI) is often carried out as a "black box" operation. CCS values are not affected by matrix interferences or chromatographic separation, and thus experimental CCS values are generally robust across different conditions and instruments (Celma et al., 2020; Hinnenkamp et al., 2018; Regueiro et al., 2016; Celma et al., 2021), although they can vary with drift gas and exact set-up. Several recently developed machine-learning CCS prediction approaches have proven capable of predicting CCS values within a relative error of < 6 % (95th percentile) and with a median error of < 2 % (Bijlsma et al., 2017; Mollerup et al., 2018; Ross et al., 2020; Zhou et al., 2016). Predicted CCS values can be used analogously to predicted retention time (RT) information in candidate selection (Mollerup et al., 2018; Bijlsma et al., 2019), helping to narrow the list of candidate structures that could potentially explain the empirical observations. IM-HRMS offers three key advantages for HRMS-based screening studies; 1) an additional separation dimension that allows for more confident peak annotation using *e.g.* library CCS values (Celma et al., 2021), 2) DT alignment for HRMS data (similar to retention time alignment), which reduces signals of co-eluting compounds and endogenous matrix interferences (Celma et al., 2020; Kaufmann et al., 2020), therefore reducing the complexity of the (fragment) spectra, and 3) separation of isobaric compounds (*e.g.* positional variants of peptides) and thus increased separation power of the analysis (D'Atri et al., 2018).

MetFrag is an *in silico* fragmentation approach developed for small molecule identification, which uses a chemical database to retrieve candidate structures based on exact mass or molecular formula (Wolf et al., 2010; Ruttkies et al., 2016). Candidates are then ranked according to fit of the predicted fragmentation compared to experimental data (FragmenterScore) and other scoring parameters if available, *e.g.* number of patents (Patent_count) and number of literature references (PubMed_count) (Ruttkies et al., 2016). Recently, PubChemLite was developed as a new database that uses specific information (*e.g.*, agrochemical, use and manufacturing information) to extract the most relevant compounds for non-target small molecule identification workflows from PubChem (LCSB-ECI, 2021; Schymanski et al., 2021). The PubChemLite version published in January 2021 contains ~ 380,000 compounds compared to 109 million compounds in PubChem (January 2021), and includes predicted CCS values from CCSbase (https://ccsbase.net/) (Ross et al., 2020; LCSB-ECI, 2021). Using databases like PubChemLite with (predicted) CCS values in combination with MetFrag allows the coupling of CCS values with high throughput compound annotation on a highly relevant selection of chemicals.

Detection of CECs in biota is alarming as bioaccumulation may lead to biomagnification and toxic effects in exposed organisms (Nilsen et al., 2019). Known examples of such chemicals are polychlorinated biphenyls (PCBs) (Madgett et al., 2022), organochlorinated pesticides (Mrema et al., 2013) and certain PFAS (Munoz et al., 2022). Traditionally, these studies are performed using gas chromatography (GC-) HRMS, as hydrophobic contaminants accumulating in fatty tissue are of primary focus (Goto et al., 2020; Rebryk and Haglund, 2021). However, recently liquid chromatography (LC-) HRMS based approaches have successfully been employed to detect more polar CECs like PCB metabolites (Liu et al., 2018) and diverse chemicals in highway runoff (Du et al., 2017). While in target analysis selective extraction and clean-up procedures can be applied, non-discriminatory sample treatment is required in wide-scope screening applications (Dürig et al., 2020; Grabicova et al., 2018). Consequently, very complex datasets are recorded, and highly selective prioritisation approaches are needed to minimise false positives. Examples of selective prioritisation approaches applied in biota screening are the use of reference samples for background subtraction (Chen et al., 2016; Xing et al., 2017) to detect co-occurring (Du et al., 2017) or unique features (Heffernan et al., 2017), time trend analysis (Plassmann et al., 2018; Dürig et al., 2022) and determination of characteristic halogen isotope patterns to detect organohalogen compounds (Goto et al., 2020; Baygi et al., 2021). Suspect and non-target screening for CECs in biota remains challenging and time-consuming and only few studies have been performed to date, which highlights the need for more research to address existing barriers.

Zebra mussels (*Dreissena polymorpha*) are a bivalve species that have been used to analyse trace contaminants (*e.g.*, metals and other bioaccumulating contaminants) in the aquatic environment (Bervoets et al., 2005; Kannan et al., 2005) and to investigate chronic exposure effects to aquatic biota (Binelli et al., 2009; Parolini et al., 2015). Mussels can be collected and transplanted into new environments, where they accumulate bioavailable pollutants from the surrounding water within weeks, making them useful for monitoring a wide range of pollutants (Parolini et al., 2015; Bervoets et al., 2004). In this study, zebra mussels were screened for bioaccumulating CECs in Lake Mälaren, Sweden's most important drinking water source, where these mussels can be found in great numbers along the shorelines (Hallstan et al., 2010). As a complement, we used polar organic chemical integrative sampler (POCIS), *i.e.* passive samplers that provide integrated chemical exposure of a water body over time (Menger et al., 2020).

The aim of our study was to evaluate the benefits of IMS in LC-HRMS for suspect and non-target screening of bioaccumulating CECs in a field experiment with caged zebra mussels and POCIS sampling in Lake Mälaren. Specific objectives included i) investigation of effects of DT alignment on full scan fragment spectrum quality and its consequences for peak annotation during screening in mussel and water (POCIS) and ii) development and assessment of prioritisation approaches and tools for suspect and non-target screening for bioaccumulated CECs in transplanted zebra mussels exposed to Lake Mälaren as an example.

2. Material and methods

2.1. Sampling

Each mussel sample consisted of a pool of 50 individual zebra mussels kept in a handmade steel mesh cage (1 dm^3) for distribution. The mussels were collected from a single site (site 2 in Fig. 1), which was selected based on its high abundance of zebra mussels and no (known) point sources nearby (Rehrl et al., 2020). Only adult mussels with shell lengths between 18 and 22 mm were used in order to achieve a pool of mussels with approximately same age (Bervoets et al., 2005; Bervoets et al., 2004). One mussel sample from site 2 was not deployed, but instead directly stored at −20 °C and later extracted as a reference sample together with the other samples (sample d0_2). The reference sample together with the common origin of the mussels from the same site and the approximate same age of the mussels were especially important for our study, as it allowed binary sample comparison during data treatment.

The POCIS disks were packed with 200 mg Oasis® hydrophilic-lipophilic balanced (HLB) sorbent (particle Ø 29.4 μm; surface area 800 m^2/g; Waters Corporation) between two polyethersulfone (PES) membranes (Ahrens et al., 2015). Two POCIS discs were prepared for each site. POCIS were wrapped in aluminium foil, placed in a zip-bag, and stored at −20 °C before deployment.

Zebra mussel samples (one cage with ~ 50 mussels per site) and POCIS ($n = 2$ per site) were distributed at 10 sampling sites in Lake Mälaren (Fig. 1). Sampling covered the full length of the lake and included sites close to urban areas, *viz.* Uppsala (site 1) and Västerås (site 9), outside urban areas, *viz.* Stockholm (site 3), Västerås (site 8), Enköping (site 6) and Märsta (site 2), and more remote sites, *viz.* sites 4, 5, 7 and 10 (Rehrl et al., 2020). The deployment period was 17 days

Fig. 1. Indicated sampling locations in Lake Mälaren, Sweden.

(18th September to 5th October 2018), and samples were deployed from shore at ~ 40 cm below the surface level and at ~ 10 m from the shoreline. Water temperature and pH were measured, and grab samples were collected for analysis of dissolved organic carbon (DOC) after placement (18th September 2018), during the deployment (26th September 2018) and at collection of the samples (5th Oct 2018, see Table S1 in Supporting Information A (SI_A)). After sample collection, the samples were stored cold during transport (max 1 h) and then at −20 °C until sample preparation.

2.2. Sample preparation

After collection, soft parts of visually healthy mussels were separated from shells and ~ 20 mussels per site were homogenized using a Precellys tissue homogenizer (Bertin Technologies) (15 mL lysing tubes with ceramic beats, 5000 rpm, 2 × 40 s, 20 s break). Soft part homogenates were then extracted according to a biota extraction method developed for multi-residue HRMS-based screening (Text S1 in SI_A) (Dürig et al., 2020; Grabicova et al., 2018). Preparation of the POCIS discs is explained in Text S2 in SI_A.

2.3. Instrumental analysis using UPLC-IM-QTOF

Instrumental analysis was performed according to Celma et al. (2020) (Celma et al., 2020). In short, samples were injected on an Acquity ultra performance liquid chromatography (UPLC) system coupled to a VION IM-QTOF mass analyser with electrospray ionisation (ESI) interface (Waters, USA). Separate injections were performed for positive and negative ESI (PI and NI, respectively). The UPLC was operated under standard reverse phase conditions using CORTECS C_{18} fused core column (2.1 × 100 mm, particle diameter 2.7 μm), using a gradient with H_2O (A) and methanol (B) as mobile phases (both with 0.01 % formic acid) during an 18 min run at a flow rate of 300 μL min^{-1}. B increased linearly from 10 % (start) to 90 % over 14 min and remained isocratic for 2 min before returning to starting conditions in 0.1 min and equilibrating for 2 min. Injection volume was 5 μL. Mass spectra were acquired from *m/z* 50 to 1000 using High Definition MSE ($HDMS^E$) acquisition mode, *i.e.* data independent acquisition including separation by ion mobility, with low collision energy (CE) scans at 6 eV and high collision energy scans at a ramp of 28–56 eV. IM drift gas was N_2, wave velocity 250 m s^{-1}, and wave height ramp 20–50 V. Leucine enkephalin lock-spray was used for mass correction.

2.4. Effects of drift time (DT) alignment on spectrum quality

The effects of DT alignment on the fragment spectrum quality and on the performance of peak annotation were assessed using final extracts spiked with target compounds (hereafter denoted 'spiked post-extraction matrix'). The final extracts of mussel and POCIS samples (mussels from the reference sample d0_2 and an additional procedural POCIS blank) were spiked with target compounds (n = 129) at 50 ng mL^{-1}. After analysis, the target compounds and their respective high CE spectra were exported from UNIFI (the vendor software of the instrumental system), to investigate the effects of DT alignment. This export was performed twice (with and without DT alignment - see also Text S3 in SI_A and (Menger, 2021). A relative intensity cut-off of 5 % was applied to remove most of the background interferences. The effects of DT alignment on the fragment spectrum quality were then investigated by running this data through MetFrag (Ruttkies et al., 2016; Ruttkies et al., 2019) (see section MetFrag settings) and the subsequent comparison of the results from the DT-aligned and non-DT-aligned spectra of the target compounds.

2.5. Pre-processing and prioritisation

Standard data pre-processing, including peak picking, spectra deconvolution and componentisation was adopted from Celma *et al.* (2020) and was performed in UNIFI (v1.8.2 and v1.9.4) (Celma et al., 2020). After pre-processing, the response threshold was adjusted according to the chromatographic noise (>5,000 in PI and > 2,000 in NI). General complexity of the HRMS biota data was reduced by means of binary sample comparison (BSC) with the reference sample (see also Text S4 in SI_A). BSC was used to remove matrix-features by considering only features that were 'unknown unique', *i.e.* present in a deployed sample but not in the reference, or present in both samples, but with an intensity in the deployed sample of at least twice the intensity of the reference sample. These were hypothesized to be features of exogenous origin rather than endogenous compounds. Features of interest were then extracted from the still complex HRMS biota data following a three-way prioritisation approach (details in Text S5 in SI_A):

A) matching a list of compounds of interest (combined target and suspect screening), B) detecting characteristic isotope patterns of

halogenated (brominated and chlorinated) compounds using the 'Halogen Match' option within UNIFI and C) co-occurrence in both biota and water (mussels and POCIS) from the same sampling site. Prioritization in UNIFI could only be performed on data without DT alignment. All prioritised features were annotated using DT aligned data and MetFrag (see next section), and further processed for (tentative) identification (Fig. 2).

All candidates (*i.e.* prioritised features) were exported from UNIFI for further investigation using MetFrag (Ruttkies et al., 2016). Chromatographic information was exported as component tables and included, *e.g.*, measured exact mass, measured RT, measured CCS, predicted halogens, response, and sample. Fragment information was exported as DT-aligned high CE spectra with a relative intensity cut-off at 5 % (for details, see Text S3 in SI and (Menger, 2021); when candidates were detected in several samples, only the fragment information of the feature with the highest response was exported.

2.6. MetFrag settings, structure elucidation and validation

All exported features were run through MetFrag via batch processing in R using the ReSOLUTION package (Schymanski, 2021). PubChemLite (version 01Jan2021; includes predicted CCS values from CCSbase) was used as local database (Ross et al., 2020; LCSB-ECI, 2021). Candidates (retrieved via exact mass within 10 ppm relative deviation) were ranked using FragmenterScore, spectrum similarity to reference spectra in the MoNA database (http://mona.fiehnlab.ucdavis.edu) calculated based on cosine similarity (MoNAScore, aka "Exact Spectral Similarity (MoNA)" as on the website), PubMed_Count, Patent_Count and number of annotation categories (AnnoTypeCount, see (Schymanski et al., 2021) for details) (weightings of 1, 1, 0.33, 0.33 and 0.33; max score = 3). Predicted CCS values from CCSbase were included in the results files but considered only during the final peak annotation, which was performed manually considering all evidence at hand.

A spectrum similarity score (SimScore) was calculated as a cosine similarity between the introduced high CE spectrum and the fragment peaks explained by MetFrag (Lai et al., 2021) (for details, see Text S6 in SI_A). In case all measured fragment peaks (regardless of the number) were explained by the *in silico* prediction, this score would have the maximum score of 1. The SimScore was used to estimate the prevalence of interference peaks during the assessment of the effects of DT alignment on the spectrum quality. A schematic overview of how the challenges of screening complex samples were addressed in this study is included in Fig. S1 in SI_A.

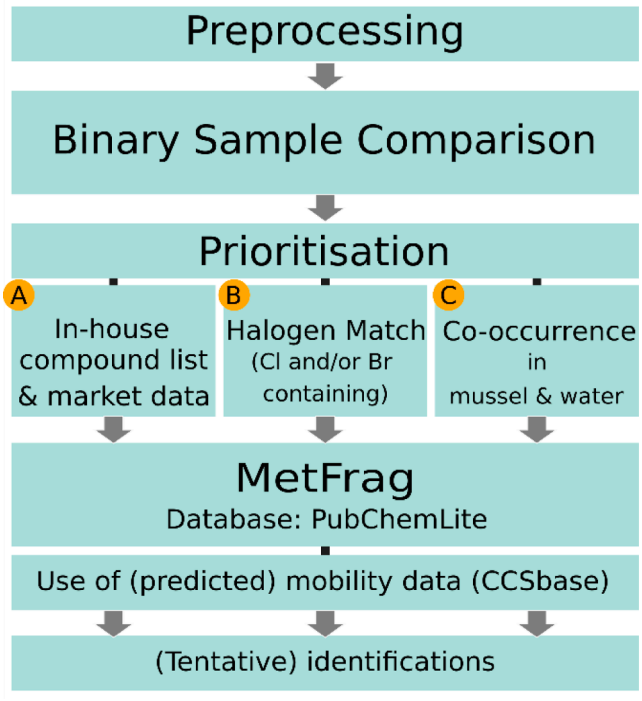

Fig. 2. Schematic overview of the data treatment incorporating three prioritisation approaches A-C.

2.7. Quality control and quality assurance

In brief, procedural blanks and instrument blanks were included, samples were spiked with mass labelled internal standards, the performance of the zebra mussel extraction was investigated (data on Table S_B1 in SI_B (Excel)) and the injection sequence was set up to avoid artificially elevated response ratios during BSC (details in Text S7 in SI_A).

3. Results and discussion

3.1. Effects of DT alignment on fragment spectrum quality

Only target compounds that were detected in both spiked post-extraction matrices (n = 117) were considered for evaluation of the effects of DT alignment on the fragment spectrum quality, meaning that those not detected in mussel (n = 12) or POCIS (n = 1) or neither (n = 3) were not considered (details in Table S_B2 in SI_B). For some compounds, sodium adducts were exported from UNIFI as the main adduct (n = 14) but were removed from the evaluation to ensure best possible comparability (as sodium adducts often yield fragment-poor spectra). Of the remaining 102 compounds (75 in PI and 28 in NI), the target compound was ranked 1st by MetFrag in all four tested instances (in mussel and POCIS matrix, with and without DT alignment) in 91 % of the cases (n = 94; 66 in PI and 28 in NI). This proved that, even in matrix rich samples and without DT alignment, MetFrag with PubChemLite was robustly able to correctly rank the most likely compound as the top candidate, exceeding the previous performance evaluations (Ruttkies et al., 2016; Schymanski et al., 2021; Schymanski et al., 2017). However, it should be highlighted that most target analytes are well-known compounds, meaning that reference spectra and/or metadata are readily available, which, consequently, can point at correct annotations even when their fragment spectra are of poor quality (this also applies to the evaluation sets). As an example, acetamiprid, a known neonicotinoid insecticide (Barbosa et al., 2016), clearly ranked first in POCIS matrix when DT-aligned data was processed (score = 2.51, 2nd rank score = 1.24), with the highest number of explained peaks (5 out of 6) and a high MoNAScore (0.95; max score = 1). However, when non-DT-aligned data of the same feature was processed, acetamiprid still ranked 1st (score = 1.57, 2nd rank score = 1.04) but with only one explained fragment and no MoNAScore. The reason for these low scores of the non-DT-aligned data can be explained by the fact that several compounds co-eluted with acetamiprid at higher intensities, which, consequently, led to a high CE spectrum with many (more intense) interference peaks (Fig. S2 in SI_A). This resulted in diminished relative signal intensities of the acetamiprid fragments, *e.g.*, of its main fragment (m/z 126.01050), which had an intensity of 100 % in the DT-aligned data, but only an intensity of 4 % in the not aligned data. In fact, the one fragment that was matched by MetFrag for the non-DT-aligned data (m/z 150.09123) originated from a co-eluting compound but not from acetamiprid itself, as can be seen by its absence in the DT-aligned spectra. This shows the clear beneficial effect that DT alignment can have on the high CE spectrum quality of some compounds and, therefore, the performance of fragment information evaluations, especially in matrix rich samples. For six of the nine target compounds that did not rank 1st in all four instances, they were in fact ranked 1st in at least one of the DT-aligned data sets (mussel and/or POCIS) and only were scored below rank 1 when non-DT-aligned data was processed, again highlighting the

usefulness of DT alignment.

The average effect of DT alignment on the high CE spectrum quality was assessed by investigating the SimScores and MoNAScores across all considered target compounds (Fig. 3 and S_B2 in SI_B (Excel)). The average SimScores (±standard deviation) of the 97 considered target compounds in DT-aligned data were 0.77 (±0.26) and 0.78 (±0.29) for mussel and POCIS matrix, respectively, whereas the scores for non-DT-aligned data were 0.40 (±0.26) and 0.46 (±0.33), respectively. This highlights that interfering fragments were efficiently removed by the DT alignment as the share of fragments explained by MetFrag clearly increased for DT-aligned data compared to the respective non-DT-aligned data. This confirms the previously observed reduction of matrix-endogenous interferences by ion mobility (Celma et al., 2020; Regueiro et al., 2016). DT alignment improved the MoNAScores on average by 0.31 (±0.30) and 0.20 (±0.27) points in the mussel and POCIS matrix, respectively. This improved similarity to reference spectra can likely be explained by the removal of otherwise interfering co-eluting compounds producing high intensity fragment signals (as has been shown in the acetamiprid example above). Further, 7 (mussel) and 12 (POCIS) target compounds that had MoNAScores > 0 in DT-aligned data had no MoNAScores in non-DT-aligned data. During elucidation of unknown features, the simple presence of a MoNAScore can already flag compounds of interest (MoNAScores are available for only a small portion of compounds) and, consequently, in these cases attention would only be drawn to the true compound when DT-aligned data was used. In 4 (mussel) and 10 (POCIS) cases, MoNAScores were higher in the non-DT-aligned data than in the DT-aligned data. These at first sight surprising cases are likely explained by falsely matched fragments originating from co-eluting interference peaks (as was the case for the one falsely matched fragment in the acetamiprid example). It is, therefore, conceivable that falsely elevated MoNAScores can occasionally occur, especially in matrix rich samples and in non-DT-aligned data. In three cases a MoNAScore > 1 was calculated, which is potentially caused by the presence of high intensity isotope peaks in the concerned spectra.

The improvement of high CE spectrum quality by DT alignment clearly benefits the performance of scoring terms that depend on this information and, consequently, reduces the number of false annotations in real-world applications. An alternative strategy to obtain high quality fragment information is re-injection of the samples after prioritisation using a data-dependent acquisition (DDA) mode with inclusion list (*e.g.*, Du et al., 2017). However, this approach is not feasible when fragment information is to be considered during prioritisation, or when degradation might have occurred or when screening digitally archived data.

3.2. Performance of the prioritisation approaches

The performance of the different data treatment steps up to, and including, prioritisation (Fig. 2) is exemplified here for the mussel sample from sampling site 1. After pre-processing, 9711 (PI) and 8235 (NI) features were detected above the adjusted response thresholds. BSC reduced these numbers by roughly 50 % to 4529 (PI) and 3553 (NI), which shows that this approach can be used to remove common background compounds, such as endogenous compounds. Omnipresent compounds can potentially be lost during BSC, *e.g.* irbesartan, a compound included in the target screening and detected in most mussel samples, was in no samples detected at a response greater than twice the response of the reference sample and would have, consequently, been missed in this approach if it was not already a target. Considering the remaining high number of features, three additional (highly selective) prioritisation approaches were applied:

A) The combined target and suspect screening detected 4 (PI) and 1 (NI) targets and produced 20 (PI) and 37 (NI) suspect screening hits.

B) Halogen Match detected 427 (PI) and 175 (NI) features containing Br and/or Cl atoms. Many of these features had masses with $m/z > 600$ and were detected towards the end of the chromatographic run, indicating low polarity and, therefore, high probability of being compounds that were not of interest for this study. Hence, a mass threshold of $m/z < 600$ was used to eliminate these cases. This reduced the numbers prioritised via Halogen Match to 203 (PI) and 82 (NI) features. Low signal intensities were expected to negatively affect the performance of Halogen Match, and it was decided to increase the response cut-off to $>$ 10,000 (PI) and $>$ 5,000 (NI) for this approach, which reduced the numbers further to 27 (PI) and 19 (NI).

C) Checking for co-occurrence in mussels and POCIS highlighted 119 (PI) and 67 (NI) co-occurring features. These co-occurring features had to be manually labelled for export in UNIFI, and, to reduce workload, the response cut-off for these features was increased to $>$ 20,000 (PI and NI), which reduced the numbers to 46 (PI) and 18 (NI).

Following the above-described prioritization approaches, 205 (PI) and 123 (NI) unique features were prioritised in total across all 10 mussel samples (from here discussed together) and were exported and processed in MetFrag. In 4 (PI) and 12 (NI) cases, MetFrag did not match any of the measured fragment masses to any of the possible candidate structures, likely because of poor fragment information (in 8 cases all high CE peaks had higher masses than the precursor ion). These 16 cases were excluded from further evaluation. The top candidates of the remaining MetFrag results were used as 'suspects' for reprocessing in UNIFI without response threshold to determine occurrence according to UNIFI's component tables and to further investigate the corresponding features by accessing and comparing extracted ion chromatograms (XICs) across samples. This highlighted 24 (PI) and 12 (NI) features occurring in blank samples at intensities comparable to the intensities in the mussel samples (intensities within one order of magnitude), and 49 (PI) and 28 (NI) features that were, in fact, noise according to visual

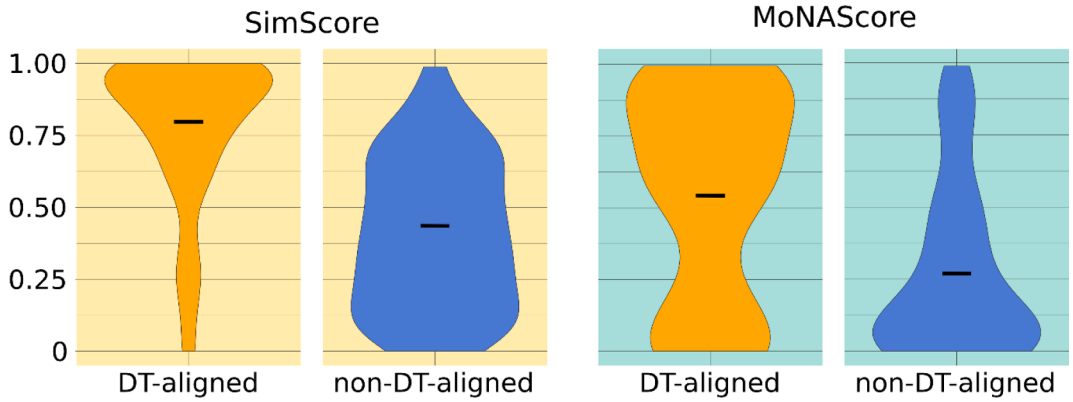

Fig. 3. Violin plots with indicated averages (black dashes) of spectrum similarity scores based on the fragment peaks explained by MetFrag's *in silico* fragmentation (SimScore) and based on the similarity to reference spectra from the MoNA database (MoNAScore) between drift time (DT) aligned and non-DT-aligned high collision energy spectra in zebra mussel matrix spiked with target compounds (n = 102).

investigation. These features were excluded, accordingly. Comparing the Halogen Match results across samples highlighted 78 (PI) and 25 (NI) features that were not consistently predicted, with < 50 % accordance across samples. This indicated that Halogen Match introduced a considerable number of false positives, and these features were excluded. The algorithm had shown reliable performance when tested on spiked matrix samples, with < 1 % false predictions (Table S3 in SI_A). Distinct isotopic signatures of Cl and Br have been used in several other studies to reliably prioritise compounds, *e.g.*, (Badea et al., 2020; Fernando et al., 2018). Our findings show that even a small rate of false positive predictions can lead to a considerable number of false positives, despite using a highly selective tool, particularly if the matrix is complex.

Finally, 35 features in PI (11 suspect hits, 4 halogenated, 20 co-occurring in POCIS and mussel) and 38 features in NI (6 suspect hits, 18 halogenated, 14 co-occurring in POCIS and mussel) passed all checks and quality criteria and were manually investigated in depth (described further below).

3.3. Use of predicted ion mobility data during peak annotation

Experimental CCS values can be used as an additional identification criterion (Regueiro et al., 2016). A CCS error of ± 2 % has been suggested as an appropriate margin for confident identifications (comparing experimental values with other experimental values) during screening studies (Celma et al., 2020). This threshold of ± 2 % was met in this study for all spiked target compounds with reference values available (Celma et al., 2019). When reference standards are not available, predicted CCS values can also be of great help because they provide additional confidence and aid the decision making during peak annotation; however, predicted values come with a specific error defined by the prediction method, *e.g.*, CCSbase reported a performance of > 80 % of predictions within a 3 % error (Ross et al., 2020). This, consequently, leaves ~ 20 % of predictions with an error exceeding 3 %. The deviations between experimental and predicted CCS values of the spiked target compounds in our study were in accordance with the reported performance (Ross et al., 2020), with 80 of 94 predicted values (85 %) below a 3 % error - excluding three compounds with no predicted values and six PFAS, which had exceptionally high errors between 6 and 35 % due to insufficient training data at the time (Table SI_B2 in SI_B (Excel)). This confirms that a < 3 % deviation between predicted and experimental CCS values can be expected in many cases and may be considered additional evidence. However, it also highlights that a deviation > 3 % should be expected in some cases too. For example, considering only candidates with < 3 % predicted CCS deviation to the experimental CCS would lead to discarding of the MetFrag candidate ranked 1st in 25 of the 84 cases of peak annotation investigated in this study. Further, predicted CCS values of different candidates can be close to each other, in which cases decision making is not helped by these values, *e.g.* the top three MetFrag candidates for 39 of the features prioritised in this study (n = 84) had all predicted CCS values within 3 % of the experimental values, and in 9 cases all values exceeded 3 %. In its current state, careful consideration of predicted CCS values is advised over automated implementation of cut-off values to avoid introducing false results, and instead CCS deviation trigger values could be used to, *e.g.*, flag cases that require special attention.

3.4. Peak annotation

Every feature that passed the manual investigations was initially considered tentatively identified with the top MetFrag candidate at a confidence level 3 (Celma et al., 2020; Schymanski et al., 2014), as long as there was no clear indication of a different compound being more likely (Table SI_B4 in SI_B (Excel)). For example, for features that were prioritised based on their predicted halogenated structure, only candidates with exact matching Cl and/or Br number were considered for annotation. Predicted CCS values from CCSbase were considered during manual evaluation and aided the decision making, *e.g.*, in cases where candidates were closely ranked. One such example was the tentatively identified structure octahydro-1-(5-fluoro-1H-indol-3-yl)-2H-quinolizine (PubChem CID: 355887) with a CCS deviation of 0.2 %. In this case, the only other candidate (ranked 1st) had a noticeably higher CCS deviation (−6.7 %), and the structure ranked 2nd was hence deemed as the most likely. Few candidates had matching reference spectra available in MoNA and three naturally occurring compounds could be assigned confidence level 2a, L-arginine and L-phenylalanine (false positive suspect hits), and guanosine (co-occurring naturally in mussel and POCIS). Ultimately, reference standards could be purchased for 18 of the 32 tentatively identified compounds and one compound (guanosine) was confirmed.

Eleven features that were prioritised using Halogen Match occurred across several samples and could be assigned tentative identities; however, only two corresponding reference standards were commercially available (both false positives). Although the tentatively assigned structure for the feature with *m/z* 265.9816 in PI was a false positive, the fact remains that a feature containing one bromine with the likely formula $C_{11}H_8BrNO_2$ was detected in every mussel sample. Similarly, a feature with *m/z* 277.0593 (NI) containing one chlorine was detected in all mussel samples, which was tentatively assigned the identity (Z)-4-[(2-amino-3-methylbutanoyl)amino]-3-chloropent-2-enedioic acid (PubChem CID: 6913469) after considering predicted CCS values and estimating ionisation behaviour. Co-occurrence in biota and POCIS samples led to the prioritisation of 22 compounds, several of which turned out to be of likely natural origin. Considering the remaining uncertainties of the tentative identifications of these compounds, the relevance of these findings remains unclear, and it is possible that the stringent prioritization criteria have caused an "over-prioritization".

In addition, target screening confirmed eight compounds (AMOZ, irbesartan, 5-hydroxyomeprazole, epitestosterone, β-zearalanol, α-nandrolone, β-nandrolone, gemfibrozil) in the mussel samples (Table S4 in SI_A).

3.5. Challenges and perspectives

In this study, ion mobility separation and advanced data processing strategies illustrated valuable to minimize or overcome current challenges of suspect and non-target screening in biota samples (Fig. S1 in SI_A). One key challenge is matrix interferences from endogenous compounds that lead to poor spectrum quality, comparatively low signal intensities of compounds of interest, and an overall high risk of false positives. This study quantitatively illustrated the benefits of ion mobility (drift time) alignment in enhancing spectral quality (similarity to predicted and reference spectra) and improving peak annotation using MetFrag. This appears important for data archiving for retrospective screening, *e.g.* to avoid loss of qualifier fragments (as was showcased for acetamiprid). DT alignment of low CE spectra could play an important role for the performance of approaches for the deconvolution of isotopic profiles, and further studies are advised. The performance of drift time alignment likely is influenced by matrix type as well as by analyte concentration, which are interesting research questions that could not be addressed in this study. Predicted CCS values were found to provide additional evidence during peak annotation in some cases; however, manual evaluations are recommended over hard cut-off values currently.

Another challenge is the complexity of the generated HRMS biota data, which is especially true for data acquired in a data-independent acquisition mode. In this study, Binary Sample Comparison to a reference sample was used to remove common background compounds and notably reduce data complexity. A limitation of this approach is that omnipresent compounds can be lost, and other means for reducing data complexity are needed for such compounds. BSC's performance heavily relies on the reference sample, which, together with the limitation that

no replicate information can be fed into the algorithm in UNIFI, highlights the importance of a carefully selected reference sample. Feature prioritisation in complex datasets relies on highly selective prioritisation approaches, and suspect screening using a carefully crafted suspect list is a well-established strategy. Moreover, to detect unknown CECs, non-targeted prioritisation approaches are needed. Detecting isotope patterns of certain halogenated structures is a highly selective approach for these compounds, but as seen in this study, even a small rate of false matches can lead to a considerable number of false detections when screening complex biota samples. This highlights the challenges of such screening studies, namely a high workload and often a low success rate. Considering additional criteria like consistency of the detected patterns across replicates or several samples can reduce this number, but implementation of such criteria can be challenging due to software limitations. Prioritisation of features co-occurring in different matrices (or samples, in general) can be another selective prioritisation approach. However, its feasibility depends on the specific study design and objectives, and there seems to be currently a lack of sufficiently flexible software solutions that allow implementation of this type of data treatment.

Working with complex HRMS biota data also comes with a high workload caused by numerous features that need to be investigated, many of which will turn out to be false flags. Setting up (semi-) automated data processing workflows can help reduce the need for drastic data reduction at bottlenecks of the workflow (often manual tasks), which will otherwise inevitably lead to loss of data of interest. For example, avoiding the shortcomings of one software (only manual comparison of features across samples possible in UNIFI) by performing a desired operation in another software (finding features co-occurring in POCIS and mussel matrix using R) can allow processing of substantially more data once the workaround is set up. However, this only truly becomes a solution when data can readily be exported and imported at different stages of the workflow – a limitation often encountered, especially when working with vendor software. Over the course of this study, the capacity to investigate prioritised features continuously grew thanks to continued advances made in software and database resources (MetFrag, PubChemLite and CCSbase), and today the number of features that could have been investigated in a timely manner clearly exceeds the number anticipated and prioritised earlier in the study. Unfortunately, some bottlenecks remain in the workflow (lack of exporting and importing functionalities) that limit throughput and can force an "overprioritisation". Once these shortcomings are addressed, thousands of unknown features (and samples) can be investigated in a reasonable time.

In conclusion, great data quantities can today be investigated and suspect and non-target screening in complex biota samples is possible. New instrument and software solutions (IM-HRMS, CCS prediction) together with new database resources (PubChemLite) increasingly help to reduce data complexity and can provide rapid results with a wealth of information. However, confident identification of unknown features to this day remains a major challenge, especially for compounds that do not have reference information (and/or standards) readily available, and more work is needed to better distinguish between endogenous compounds and compounds of concern.

CRediT authorship contribution statement

Frank Menger: Conceptualization, Data curation, Investigation, Software, Visualization, Writing – original draft. **Alberto Celma:** Data curation, Validation, Writing – review & editing. **Emma L. Schymanski:** Conceptualization, Data curation, Software, Writing – review & editing. **Foon Yin Lai:** Methodology, Writing – review & editing. **Lubertus Bijlsma:** Supervision, Writing – review & editing. **Karin Wiberg:** Funding acquisition, Writing – review & editing. **Félix Hernández:** Resources, Writing – review & editing. **Juan V. Sancho:** Supervision, Validation, Writing – review & editing. **Lutz Ahrens:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

F. Menger, K. Wiberg and L. Ahrens acknowledge the Swedish Research Council for Sustainable Development (FORMAS) for funding this study under the registration number 2016-01173 (LakePOPs). A. Celma acknowledges the Spanish Ministry of Economy and Competitiveness for his predoctoral grant (BES-2016-076914), and L. Bijlsma acknowledges the financial support by the Swedish Research Council (FORMAS) for his post-doctoral grant in "Effect-directed analysis as a tool toward a nontoxic environment-identification of mixture effects and toxicity drivers in water (DANTE)" project (2018-02256). J. V. Sancho acknowledges the financial support from Universitat Jaume I (UJI-B2020-19), and E. L. Schymanski acknowledges funding support by the Luxembourg National Research Fund (FNR) for project A18/BM/12341006. The Research Institute for Pesticides and Water acknowledges the financial support of Generalitat Valenciana as Research Group of Excellence (Prometeo 2019/040). We gratefully acknowledge Sigtuna Naturskola, Oksana Golovko, Wiebke Dürig, Mikael Isaksson, Vera Franke, and Svenja Hoffmeister for their help during sample collection, experimental work, and discussion, as well as Evan Bolton, Jeff Zhang, Paul Thiessen, Sunghwan Kim and the PubChem team (NCBI/NLM/NIH) plus Libin Xu, Ally Krinsky and Dylan Ross from CCSbase (University of Washington) for contributions to and associated discussions around the PubChemLite CCSbase predicted values.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2022.107585.

References

Ahrens, L., Daneshvar, A., Lau, A.E., Kreuger, J., 2015. Characterization of five passive sampling devices for monitoring of pesticides in water. J. Chromatogr. A. 1405, 1–11. https://doi.org/10.1016/j.chroma.2015.05.044.

Badea, S.-L., Geana, E.-I., Niculescu, V.-C., Ionete, R.-E., 2020. Recent progresses in analytical GC and LC mass spectrometric based-methods for the detection of emerging chlorinated and brominated contaminants and their transformation products in aquatic environment. Sci. Total Environ. 722, 137914 https://doi.org/10.1016/j.scitotenv.2020.137914.

Baduel, C., Mueller, J.F., Tsai, H., Gomez Ramos, M.J., 2015. Development of sample extraction and clean-up strategies for target and non-target analysis of environmental contaminants in biological matrices. J. Chromatogr. A. 1426, 33–47. https://doi.org/10.1016/j.chroma.2015.11.040.

Barbosa, M.O., Moreira, N.F.F., Ribeiro, A.R., Pereira, M.F.R., Silva, A.M.T., 2016. Occurrence and removal of organic micropollutants: An overview of the watch list of EU Decision 2015/495. Water Res. 94, 257–279. https://doi.org/10.1016/j.watres.2016.02.047.

Barrett, H., Du, X., Houde, M., Lair, S., Verreault, J., Peng, H., 2021. Suspect and nontarget screening revealed class-specific temporal trends (2000–2017) of poly- and perfluoroalkyl substances in St. Lawrence Beluga Whales. Environ. Sci. Technol. 55, 1659–1671. https://doi.org/10.1021/acs.est.0c05957.

Baygi, S.F., Fernando, S., Hopke, P.K., Holsen, T.M., Crimmins, B.S., 2021. Nontargeted discovery of novel contaminants in the great lakes region: a comparison of fish fillets and fish consumers. Environ. Sci. Technol. 55, 3765–3774. https://doi.org/10.1021/acs.est.0c08507.

Belova, L., Caballero-Casero, N., van Nuijs, A.L.N., Covaci, A., 2021. Ion mobility-high-resolution mass spectrometry (IM-HRMS) for the analysis of contaminants of emerging concern (CECs): database compilation and application to urine samples. Anal. Chem. 93, 6428–6436. https://doi.org/10.1021/acs.analchem.1c00142.

Bervoets, L., Voets, J., Chu, S., Covaci, A., Schepens, P., Blust, R., 2004. Comparison of accumulation of micropollutants between indigenous and transplanted zebra mussels (dreissena polymorpha). Environ. Toxicol. Chem. 23, 1973. https://doi.org/10.1897/03-365.

Bervoets, L., Voets, J., Covaci, A., Chu, S., Qadah, D., Smolders, R., Schepens, P., Blust, R., 2005. Use of transplanted zebra mussels (*Dreissena polymorpha*) to assess the bioavailability of microcontaminants in flemish surface waters. Environ. Sci. Technol. 39, 1492–1505. https://doi.org/10.1021/es049048t.

Bijlsma, L., Bade, R., Celma, A., Mullin, L., Cleland, G., Stead, S., Hernandez, F., Sancho, J.V., 2017. Prediction of collision cross-section values for small molecules: application to pesticide residue analysis. Anal. Chem. 89, 6583–6589. https://doi.org/10.1021/acs.analchem.7b00741.

Bijlsma, L., Berntssen, M.H.G., Merel, S., 2019. A refined nontarget workflow for the investigation of metabolites through the prioritization by in silico prediction tools. Anal. Chem. 91, 6321–6328. https://doi.org/10.1021/acs.analchem.9b01218.

Binelli, A., Cogni, D., Parolini, M., Riva, C., Provini, A., 2009. In vivo experiments for the evaluation of genotoxic and cytotoxic effects of Triclosan in Zebra mussel hemocytes. Aquat. Toxicol. 91, 238–244. https://doi.org/10.1016/j.aquatox.2008.11.008.

Celma, A., Fabregat-Safont, D., Ibàñez, M., Bijlsma, L., Hernandez, F., Sancho, J.V., 2019. S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI, http://doi.org/10.5281/zenodo.3966751.

Celma, A., Sancho, J.V., Schymanski, E.L., Fabregat-Safont, D., Ibáñez, M., Goshawk, J., Barknowitz, G., Hernández, F., Bijlsma, L., 2020. Improving target and suspect screening high-resolution mass spectrometry workflows in environmental analysis by ion mobility separation. Environ. Sci. Technol. 54, 15120–15131. https://doi.org/10.1021/acs.est.0c05713.

Celma, A., Ahrens, L., Gago-Ferrero, P., Hernández, F., López, F., Lundqvist, J., Pitarch, E., Sancho, J.V., Wiberg, K., Bijlsma, L., 2021. The relevant role of ion mobility separation in LC-HRMS based screening strategies for contaminants of emerging concern in the aquatic environment. Chemosphere 280, 130799. https://doi.org/10.1016/j.chemosphere.2021.130799.

Chen, C., Wohlfarth, A., Xu, H., Su, D., Wang, X., Jiang, H., Feng, Y., Zhu, M., 2016. Untargeted screening of unknown xenobiotics and potential toxins in plasma of poisoned patients using high-resolution mass spectrometry: Generation of xenobiotic fingerprint using background subtraction. Anal. Chim. Acta. 944, 37–43. https://doi.org/10.1016/j.aca.2016.09.034.

D'Atri, V., Causon, T., Hernandez-Alba, O., Mutabazi, A., Veuthey, J.-L., Cianferani, S., Guillarme, D., 2018. Adding a new separation dimension to MS and LC–MS: What is the utility of ion mobility spectrometry? J. Sep. Sci. 41, 20–67. https://doi.org/10.1002/jssc.201700919.

Dom, I., Biré, R., Hort, V., Lavison-Bompard, G., Nicolas, M., Guérin, T., 2018. Extended targeted and non-targeted strategies for the analysis of marine toxins in mussels and oysters by (LC-HRMS). Toxins. 10, 375. https://doi.org/10.3390/toxins10090375.

Du, B., Lofton, J.M., Peter, K.T., Gipe, A.D., James, C.A., McIntyre, J.K., Scholz, N.L., Baker, J.E., Kolodziej, E.P., 2017. Development of suspect and non-target screening methods for detection of organic contaminants in highway runoff and fish tissue with high-resolution time-of-flight mass spectrometry. Environ. Sci. Process. Impacts. 19, 1185–1196. https://doi.org/10.1039/c7em00243b.

Dürig, W., Kintzi, A., Golovko, O., Wiberg, K., Ahrens, L., 2020. New extraction method prior to screening of organic micropollutants in various biota matrices using liquid chromatography coupled to high-resolution time-of-flight mass spectrometry. Talanta 219, 121294. https://doi.org/10.1016/j.talanta.2020.121294.

Dürig, W., Alygizakis, N.A., Menger, F., Golovko, O., Wiberg, K., Ahrens, L., 2022. Novel prioritisation strategies for evaluation of temporal trends in archived white-tailed sea eagle muscle tissue in non-target screening. J. Hazard. Mater. 424, 127331 https://doi.org/10.1016/j.jhazmat.2021.127331.

Fabregat-Safont, D., Ibáñez, M., Bijlsma, L., Hernández, F., Waichman, A.V., de Oliveira, R., Rico, A., 2021. Wide-scope screening of pharmaceuticals, illicit drugs and their metabolites in the Amazon River. Water Res. 200, 117251 https://doi.org/10.1016/j.watres.2021.117251.

Fernando, S., Renaguli, A., Milligan, M.S., Pagano, J.J., Hopke, P.K., Holsen, T.M., Crimmins, B.S., 2018. Comprehensive analysis of the great lakes top predator fish for novel halogenated organic contaminants by GC×GC-HR-ToF mass spectrometry. Environ. Sci. Technol. 52, 2909–2917. https://doi.org/10.1021/acs.est.7b05999.

Gago-Ferrero, P., Bletsou, A.A., Damalas, D.E., Aalizadeh, R., Alygizakis, N.A., Singer, H. P., Hollender, J., Thomaidis, N.S., 2020. Wide-scope target screening of >2000 emerging contaminants in wastewater samples with UPLC-Q-ToF-HRMS/MS and smart evaluation of its performance through the validation of 195 selected representative analytes. J. Hazard. Mater. 387, 121712 https://doi.org/10.1016/j.jhazmat.2019.121712.

Goto, A., Tue, N.M., Isobe, T., Takahashi, S., Tanabe, S., Kunisue, T., 2020. Nontarget and target screening of organohalogen compounds in mussels and sediment from Hiroshima Bay, Japan: occurrence of novel bioaccumulative substances. Environ. Sci. Technol. 54, 5480–5488. https://doi.org/10.1021/acs.est.9b06998.

Grabicova, K., Vojs Staňová, A., Koba Ucun, O., Borik, A., Randak, T., Grabic, R., 2018. Development of a robust extraction procedure for the HPLC-ESI-HRPS determination of multi-residual pharmaceuticals in biota samples. Anal. Chim. Acta. 1022, 53–60. https://doi.org/10.1016/j.aca.2018.04.011.

Hallstan, S., Grandin, U., Goedkoop, W., 2010. Current and modeled potential distribution of the zebra mussel (Dreissena polymorpha) in Sweden. Biol. Invasions. 12, 285–296. https://doi.org/10.1007/s10530-009-9449-9.

Heffernan, A.L., Gómez-Ramos, M.M., Gaus, C., Vijayasarathy, S., Bell, I., Hof, C., Mueller, J.F., Gómez-Ramos, M.J., 2017. Non-targeted, high resolution mass spectrometry strategy for simultaneous monitoring of xenobiotics and endogenous compounds in green sea turtles on the Great Barrier Reef. Sci. Total Environ. 599–600, 1251–1262. https://doi.org/10.1016/j.scitotenv.2017.05.016.

Hernández, F., Bakker, J., Bijlsma, L., de Boer, J., Botero-Coy, A.M., Bruinen de Bruin, Y., Fischer, S., Hollender, J., Kasprzyk-Hordern, B., Lamoree, M., López, F.J., ter Laak, T.L., van Leerdam, J.A., Sancho, J.V., Schymanski, E.L., de Voogt, P., Hogendoorn, E.A., 2019. The role of analytical chemistry in exposure science: Focus on the aquatic environment. Chemosphere 222, 564–583. https://doi.org/10.1016/j.chemosphere.2019.01.118.

Hinnenkamp, V., Klein, J., Meckelmann, S.W., Balsaa, P., Schmidt, T.C., Schmitz, O.J., 2018. Comparison of CCS values determined by traveling wave ion mobility mass spectrometry and drift tube ion mobility mass spectrometry. Anal. Chem. 90, 12042–12050. https://doi.org/10.1021/acs.analchem.8b02711.

Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017. Nontarget screening with high resolution mass spectrometry in the environment: ready to go? Environ. Sci. Technol. 51, 11505–11512. https://doi.org/10.1021/acs.est.7b02184.

Jacob, P., Barzen-Hanson, K.A., Helbling, D.E., 2021. Target and nontarget analysis of per- and polyfluoralkyl substances in wastewater from electronics fabrication facilities. Environ. Sci. Technol. 55, 2346–2356. https://doi.org/10.1021/acs.est.0c06690.

Kannan, K., Tao, L., Sinclair, E., Pastva, S.D., Jude, D.J., Giesy, J.P., 2005. Perfluorinated compounds in aquatic organisms at various trophic levels in a great lakes food chain. Arch. Environ. Contam. Toxicol. 48, 559–566. https://doi.org/10.1007/s00244-004-0133-x.

Kaufmann, A., Butcher, P., Maden, K., Walker, S., Widmer, M., 2020. Does the ion mobility resolving power as provided by commercially available ion mobility quadrupole time-of-flight mass spectrometry instruments permit the unambiguous identification of small molecules in complex matrices? Anal. Chim. Acta. 1107, 113–126. https://doi.org/10.1016/j.aca.2020.02.032.

Kiefer, K., Müller, A., Singer, H., Hollender, J., 2019. New relevant pesticide transformation products in groundwater detected using target and suspect screening for agricultural and urban micropollutants with LC-HRMS. Water Res. 165, 114972 https://doi.org/10.1016/j.watres.2019.114972.

Lai, A., Singh, R.R., Kovalova, L., Jaeggi, O., Kondić, T., Schymanski, E.L., 2021. Retrospective non-target analysis to support regulatory water monitoring: from masses of interest to recommendations via in silico workflows. Environ. Sci. Eur. 33, 43. https://doi.org/10.1186/s12302-021-00475-1.

LCSB-ECI, Schymanski, Emma, Kondic, Todor, PubChem Team, Bolton, Evan, Thiessen, Paul, Zhang, Jeff, CCSbase Team, Krinsky, Ally, Ross, David H., Xu, Libin, PubChemLite for Exposomics (1 Jan 2021) + predicted CCS from CCSbase, (2021). http://doi.org/10.5281/zenodo.4456208.

Liu, Y., Richardson, E.S., Derocher, A.E., Lunn, N.J., Lehmler, H.-J., Li, X., Zhang, Y., Cui, J.Y., Cheng, L., Martin, J.W., 2018. Hundreds of unrecognized halogenated contaminants discovered in polar bear serum. Angew. Chem. Int. Ed. 57, 16401–16406. https://doi.org/10.1002/anie.201809906.

Madgett, A.S., Yates, K., Webster, L., McKenzie, C., Brownlow, A., Moffat, C.F., 2022. The concentration and biomagnification of PCBs and PBDEs across four trophic levels in a marine food web. Environ. Pollut. 309, 119752 https://doi.org/10.1016/j.envpol.2022.119752.

Menger, F., Gago-Ferrero, P., Wiberg, K., Ahrens, L., 2020. Wide-scope screening of polar contaminants of concern in water: A critical review of liquid chromatography-high resolution mass spectrometry-based strategies. Trends Environ. Anal. Chem. 28, e00102.

Menger, F., 2021. Re-structuring UNIFI data for use in MetFrag. https://github.com/MengerFrankC/UNIFI_SpectrumExport (accessed May 6, 2021).

Mollerup, C.B., Mardal, M., Dalsgaard, P.W., Linnet, K., Barron, L.P., 2018. Prediction of collision cross section and retention time for broad scope screening in gradient reversed-phase liquid chromatography-ion mobility-high resolution accurate mass spectrometry. J. Chromatogr. A. 1542, 82–88. https://doi.org/10.1016/j.chroma.2018.02.025.

Moschet, C., Wittmer, I., Simovic, J., Junghans, M., Piazzoli, A., Singer, H., Stamm, C., Leu, C., Hollender, J., 2014. How a complete pesticide screening changes the assessment of surface water quality. Environ. Sci. Technol. 48, 5423–5432. https://doi.org/10.1021/es500371t.

Mrema, E.J., Rubino, F.M., Brambilla, G., Moretto, A., Tsatsakis, A.M., Colosio, C., 2013. Persistent organochlorinated pesticides and mechanisms of their toxicity. Toxicology 307, 74–88. https://doi.org/10.1016/j.tox.2012.11.015.

Munoz, G., Mercier, L., Duy, S.V., Liu, J., Sauvé, S., Houde, M., 2022. Bioaccumulation and trophic magnification of emerging and legacy per- and polyfluoroalkyl substances (PFAS) in a St. Lawrence River food web. Environ. Pollut. 309, 119739 https://doi.org/10.1016/j.envpol.2022.119739.

Nilsen, E., Smalling, K.L., Ahrens, L., Gros, M., Miglioranza, K.S.B., Picó, Y., Schoenfuss, H.L., 2019. Critical review: Grand challenges in assessing the adverse effects of contaminants of emerging concern on aquatic food webs. Environ. Toxicol. Chem. 38, 46–60. https://doi.org/10.1002/etc.4290.

Parolini, M., Magni, S., Castiglioni, S., Zuccato, E., Binelli, A., 2015. Realistic mixture of illicit drugs impaired the oxidative status of the zebra mussel (Dreissena polymorpha). Chemosphere 128, 96–102. https://doi.org/10.1016/j.chemosphere.2014.12.092.

Peter, K.T., Tian, Z., Wu, C., Lin, P., White, S., Du, B., McIntyre, J.K., Scholz, N.L., Kolodziej, E.P., 2018. Using high-resolution mass spectrometry to identify organic contaminants linked to urban stormwater mortality syndrome in coho salmon. Environ. Sci. Technol. 52, 10317–10327. https://doi.org/10.1021/acs.est.8b03287.

Plassmann, M.M., Fischer, S., Benskin, J.P., 2018. Nontarget time trend screening in human blood. Environ. Sci. Technol. Lett. 5, 335–340. https://doi.org/10.1021/acs.estlett.8b00196.

Rebryk, A., Haglund, P., 2021. Non-targeted screening workflows for gas chromatography–high-resolution mass spectrometry analysis and identification of biomagnifying contaminants in biota samples. Anal. Bioanal. Chem. 413, 479–501. https://doi.org/10.1007/s00216-020-03018-4.

Regueiro, J., Negreira, N., Berntssen, M.H.G., 2016. Ion-mobility-derived collision cross section as an additional identification point for multiresidue screening of pesticides in fish feed. Anal. Chem. 88, 11169–11177. https://doi.org/10.1021/acs.analchem.6b03381.

Rehrl, A.-L., Golovko, O., Ahrens, L., Köhler, S., 2020. Spatial and seasonal trends of organic micropollutants in Sweden's most important drinking water reservoir. Chemosphere 249, 126168. https://doi.org/10.1016/j.chemosphere.2020.126168.

Ross, D.H., Cho, J.H., Xu, L., 2020. Breaking down structural diversity for comprehensive prediction of ion-neutral collision cross sections. Anal. Chem. 92, 4548–4557. https://doi.org/10.1021/acs.analchem.9b05772.

Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J. Cheminformatics. 8, 3. https://doi.org/10.1186/s13321-016-0115-9.

Ruttkies, C., Neumann, S., Posch, S., 2019. Improving MetFrag with statistical learning of fragment annotations. BMC Bioinf. 20, 376. https://doi.org/10.1186/s12859-019-2954-7.

Schlüsener, M.P., Kunkel, U., Ternes, T.A., 2015. Quaternary triphenylphosphonium compounds: A new class of environmental pollutants. Environ. Sci. Technol. 49, 14282–14291. https://doi.org/10.1021/acs.est.5b03926.

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Technol. 48, 2097–2098. https://doi.org/10.1021/es5002105.

Schymanski, E.L., Ruttkies, C., Krauss, M., Brouard, C., Kind, T., Dührkop, K., Allen, F., Vaniya, A., Verdegem, D., Böcker, S., Rousu, J., Shen, H., Tsugawa, H., Sajed, T., Fiehn, O., Ghesquière, B., Neumann, S., 2017. Critical assessment of small molecule identification 2016: automated methods. J. Cheminformatics. 9, 22. https://doi.org/10.1186/s13321-017-0207-1.

Schymanski, E.L., Kondić, T., Neumann, S., Thiessen, P.A., Zhang, J., Bolton, E.E., 2021. Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. J. Cheminformatics. 13, 19. https://doi.org/10.1186/s13321-021-00489-0.

Schymanski, E., 2021. schymane/ReSOLUTION, https://github.com/schymane/ReSOLUTION.

Stephan, S., Hippler, J., Köhler, T., Deeb, A.A., Schmidt, T.C., Schmitz, O.J., 2016. Contaminant screening of wastewater with HPLC-IM-qTOF-MS and LC+LC-IM-qTOF-MS using a CCS database. Anal. Bioanal. Chem. 408, 6545–6555. https://doi.org/10.1007/s00216-016-9820-5.

Tian, Z., Peter, K.T., Gipe, A.D., Zhao, H., Hou, F., Wark, D.A., Khangaonkar, T., Kolodziej, E.P., James, C.A., 2020. Suspect and nontarget screening for contaminants of emerging concern in an urban estuary. Environ. Sci. Technol. 54, 889–901. https://doi.org/10.1021/acs.est.9b06126.

Tian, Z., Zhao, H., Peter, K.T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., Cortina, A.E., Biswas, R.G., Kock, F.V.C., Soong, R., Jenne, A., Du, B., Hou, F., He, H., Lundeen, R., Gilbreath, A., Sutton, R., Scholz, N.L., Davis, J.W., Dodd, M.C., Simpson, A., McIntyre, J.K., Kolodziej, E.P., 2021. A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. Science 371 (6525), 185–189.

Wang, Y., Yu, N., Zhu, X., Guo, H., Jiang, J., Wang, X., Shi, W., Wu, J., Yu, H., Wei, S., 2018. Suspect and nontarget screening of per- and polyfluoroalkyl substances in wastewater from a fluorochemical manufacturing park. Environ. Sci. Technol. 52, 11007–11016. https://doi.org/10.1021/acs.est.8b03030.

Wang, X., Yu, N., Yang, J., Jin, L., Guo, H., Shi, W., Zhang, X., Yang, L., Yu, H., Wei, S., 2020. Suspect and non-target screening of pesticides and pharmaceuticals transformation products in wastewater using QTOF-MS. Environ. Int. 137, 105599 https://doi.org/10.1016/j.envint.2020.105599.

Wolf, S., Schmidt, S., Müller-Hannemann, M., Neumann, S., 2010. In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinf. 11, 148. https://doi.org/10.1186/1471-2105-11-148.

Xing, J., Zang, M., Liu, H., 2017. The application of a novel high-resolution mass spectrometry-based analytical strategy to rapid metabolite profiling of a dual drug combination in humans. Anal. Chim. Acta. 993, 38–46. https://doi.org/10.1016/j.aca.2017.08.047.

Zhou, Z., Shen, X., Tu, J., Zhu, Z.-J., 2016. Large-scale prediction of collision cross-section values for metabolites in ion mobility-mass spectrometry. Anal. Chem. 88, 11084–11091. https://doi.org/10.1021/acs.analchem.6b03091.