Facilitating high resolution mass spectrometry data processing for screening of environmental water samples: An evaluation of two deconvolution tools

Richard Bade[a,1], Ana Causanilles[b], Erik Emke[b], Lubertus Bijlsma[a], Juan V. Sancho[a], Felix Hernandez[a], Pim de Voogt[b,c*]

[a] *Research Institute for Pesticides and Water, University Jaume I, Avda. Sos Baynat s/n, E-12071 Castellón, Spain.*

[b] *KWR Watercycle Research Institute, Chemical Water Quality and Health, P.O. Box 1072, 3430 BB, Nieuwegein, The Netherlands*

[c] *Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, P.O. Box 94248, 1090 GE, Amsterdam, The Netherlands*

[*] Corresponding author: w.p.devoogt@uva.nl Tel.: +31 20 5256565; fax: +31 20 5257431

[1] *Visiting researcher at KWR Watercycle Research Institute*

Abstract

A screening approach was applied to influent and effluent wastewater samples. After injection in a LC-LTQ-Orbitrap, data analysis was performed using two deconvolution tools, MsXelerator (modules MPeaks and MS Compare) and Sieve 2.1. The outputs were searched incorporating an in-house database of more than 200 pharmaceuticals and illicit drugs or ChemSpider. This hidden target screening approach led to the detection of numerous compounds including the illicit drug cocaine and its metabolite benzoylecgonine and the pharmaceuticals carbamazepine, gemfibrozil and losartan. The compounds found using both approaches were combined, and isotopic pattern and retention time prediction were used to filter out false positives. The remaining potential positives were reanalysed in MS/MS mode and their product ions were compared with literature and/or mass spectral libraries. The inclusion of the chemical database ChemSpider led to the tentative identification of several metabolites, including paraxanthine, theobromine, theophylline and carboxylosartan, as well as the pharmaceutical phenazone. The first three of these compounds are isomers and they were subsequently distinguished based on their product ions and predicted retention times. This work has shown that the use deconvolution tools facilitates non-target screening and enables the identification of a higher number of compounds.

Keywords: Non-target screening, peak-picking, hidden target screening, software, high resolution mass spectrometry, aquatic environment

Introduction

The investigation of emerging contaminants has become prevalent in analytical environmental chemistry circles. The use of pharmaceuticals, personal care products and illicit drugs is increasing worldwide, due to the growing population and the rise in available products and the amount of these contaminants entering the aquatic environment is of concern (Fatta-Kassinos et al., 2011). There is no blanket removal process able to be undertaken by wastewater treatment plants (WWTPs) for all compounds, leading to poor removal rates and detection of many of these compounds in effluent wastewaters (EWW) and consequently in surface waters (Bijlsma et al., 2012; Luo et al., 2014; van der Aa et al., 2013).

High Resolution Mass Spectrometry (HRMS) instruments, such as Time of Flight (TOF) and Orbitrap have revolutionized the investigation of emerging contaminants in the aquatic environment due to their high sensitivity in full scan mode, their increased mass accuracy and the possibility to distinguish the isotopic pattern. HRMS instruments have the ability to screen for unknowns due to exact mass measurements and these are unique characteristics compared to other mass spectrometry instruments. Hybrid systems i.e. HRMS hyphenated to a quadrupole or linear ion trap (LTQ), such as the LTQ-Orbitrap, combines the tandem mass spectrometric capability associated with the LTQ with the high mass resolving power (up to 100,000 FWHM) and mass accuracy capability of the Orbitrap (de Voogt et al., 2011; Makarov and Scigelova, 2010). These hybrid configurations based on HRMS allow reliable interpretation of MS/MS spectra and are very valuable when dealing with complex environmental matrices, such as wastewater, where co-elution of analytes with matrix interferences can result in ambiguous peaks (Hogenboom et al., 2009). By utilising the ultra-high resolution capabilities, isobaric compounds can easily be differentiated (Hernández et al., 2012).

In the literature, three different approaches are described for the detection and/or identification of compounds: target, suspect/post-target and non-target (Aceña et al., 2015; Bletsou et al., 2015; Gago-Ferrero et al., 2015; Hernández et al., 2015b, 2014, 2005; Krauss et al., 2010; Leendert et al., 2015). Target methods are limited to a restricted number of compounds, for all of which reference standards must be obtained and, therefore, information on the occurrence of other unknown, relevant micropollutants may be missed. Suspect screening takes advantage of a database of "known" compounds, including molecular formulae, fragmentation and retention time, which can then be

computationally correlated to spectral HRMS data to give potential positive compounds. As the concept of suspect screening implies that reference standards are not necessarily available, the tentative identification of potential positives needs to be confirmed by the use of reference standards (and MS/MS injections, if required) in a final step.

The third, non-target, approach is of increasing interest but notoriously difficult to undertake, as, strictly speaking, no *a priori* information is available (Krauss et al., 2010; Schymanski et al., 2014b; Zedda and Zwiener, 2012). Even with the help of automated peak-picking software, thousands of peaks can be detected in an individual sample (Hug et al., 2014; Kern et al., 2009). Consequently, subsequent steps must then be made to reduce the number of peaks to a more manageable number, including molecular formula derivation, isotopic pattern, mass defect analysis and retention time prediction (Gago-Ferrero et al., 2015; Helbling et al., 2010; Kind and Fiehn, 2007). Further confidence in the "potential positives" remaining can be gained through the use of fragmentation in a subsequent MS/MS injection and comparison with *in silico* fragmentation and/or mass spectral libraries (Bletsou et al., 2015; Gerlich and Neumann, 2013; Herrera-Lopez et al., 2014; Hug et al., 2014; Little et al., 2012), with the latter referred to as "hidden targets" (Letzel et al., 2015). In these situations, it is of prime importance and for ease of the analyst to have software capable of fulfilling most (if not all) of these steps automatically. Most manufacturers have software specific for their instrument and data, which can automatically extract analytes of interest from the raw data, to facilitate suspect screening approaches. However, despite the tremendous advances in software for metabolite/transformation product detection and further non-target work, sometimes not all required information is available in one platform, leading users to manufacturer-independent software, such as the Eawag open-source R-code packages enviMass, enviPick, nontarget and RMassBank (Schollée et al., 2015; Schymanski et al., 2014a) which can enable the incorporation of additional parameters, such as the steps outlined above. In spite of these problems, non-target screening is necessary to identify new or unknown relevant pollutants, which is why efforts need to be made in developing proper software and efficient identification tools.

This work portrays the combination of non-target data processing and hidden target searching of environmental water samples after injection in an LC-LTQ-Orbitrap. Two computational programs were utilized: MsXelerator (MsMetrix) and Sieve 2.1 (Thermo Scientific). An in-house database of more than 200 pharmaceuticals, personal care

products and illicit drugs was incorporated in both programs. Additionally, Sieve 2.1 was used in combination with the ChemSpider search feature. The main objective was to demonstrate the utility and additional value of these software packages for screening. This led to the detection of numerous compounds across both programs. The compounds detected by both methods were then reinjected to obtain MS/MS fragmentation, leading to the tentative identification of 24 compounds. Ultimately, this work shows that the combined use of two deconvolution tools combined with two hidden target screening approaches provides more information than either one used individually.

2. Materials and Methods

2.1 Reagents

HPLC-grade methanol (MeOH), and formic acid (>98 % w/w) were purchased from Mallinckrodt Baker (Deventer, The Netherlands). The ultrapure water was obtained by purifying demineralized water in a Milli-Q system from Millipore (Bedford, MA, USA). SPE cartridges used were Oasis HLB 3 mL (60 mg) from Waters (Milford, MA, USA). Polytyrosine-1,3,6 standard used for mass axis calibration was purchased from Cs Bio Co. (Menlo Park, CA, USA). Mixed cellulose ester membrane filters (0.45 μm) were purchased from Whatman (Dassel, Germany).

2.2 Water samples and extraction procedure

Seven influent wastewater (IWW) and seven effluent wastewater (EWW) 24-hour composite samples were collected over seven consecutive days in March 2014. They were stored in high density polystyrene bottles, immediately centrifuged and stored in the dark at -20°C. Analyses were performed as soon as possible after collection in order to keep biotic or abiotic degradation to a minimum (Llorca et al. 2014).

A solid phase extraction (SPE) step was applied prior to analysis to pre-concentrate the samples. All samples were filtered through a mixed cellulose ester membrane filter (0.45 μm). SPE was performed using Oasis HLB cartridges (60 mg). The water samples (EWW 100 mL, with IWW four times diluted (i.e. 25 mL sample diluted to 100mL by adding Milli-Q water)) were loaded onto the cartridges and reconstructed in 1 mL of 10:90 MeOH:H_2O after elution with MeOH (5 mL). A procedural blank was also made, following the steps above but using Milli-Q water. Analyses were performed by injecting 20 μL of the final extract (in triplicate) into the LC-LTQ FT Orbitrap. For further information on the SPE procedure, see (Hernández et al., 2015a).

2.3 Liquid Chromatography

The HPLC system, consisted of a Surveyor auto sampler model Plus and a Surveyor quaternary gradient HPLC-pump (Thermo Fisher Scientific, Breda, The Netherlands). Chromatographic separation of the compounds was made using an XBridge C18 column (150 mm × 2.1 mm I.D., particle size 3.5 μm) (Waters). The pre-column used was a 4.0 mm × 2.0 mm I.D. Phenomenex Security Guard column (Bester, Amsterdam, the

Netherlands). The analytical column and the guard column were maintained at a temperature of 21ºC in a column thermostat. A gradient was used at a constant flow rate of 0.3 mL min^{-1} using Milli-Q water (Solvent A) and MeOH (Solvent B) both with 0.05% formic acid. The percentage of organic modifier (B) was changed linearly as follows: 0 min, 5%; 40 min, 100%; 45 min, 100%; 47 min, 5%. Between consecutive runs, the analytical column was re-equilibrated for 5 min.

2.4 LTQ-FT Orbitrap mass spectrometry

An LTQ FT Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany) was used. The LTQ part of this system was equipped with a Heated Ion Max Electrospray Ionization (HESI) probe and operated in the positive ion mode. The conditions were: source voltage 3.0 kV, heated capillary temperature 300ºC, vaporizer temperature 350ºC, capillary voltage 13 V and tube lens 70 V. Products ions were generated in the LTQ trap at a normalized collision energy setting of 35% and using an isolation width of 2 Da.

Full-scan accurate mass spectra (mass range from 50 to 1300 Da) were obtained at a mass resolution of 60,000. The total cycle time depends upon the resolution; at the selected resolution the total cycle time is 0.5 s. The instrument was initially set to operate in full-scan ('survey') mode with accurate mass measurements. When an ion exceeded a preset threshold, the instrument switched to product-ion scan mode in the ion trap part. Further details on instrument operating conditions can be found elsewhere (Bijlsma et al., 2013)

All data were acquired and processed using Xcalibur version 2.1 software. A second MS/MS injection was made by incorporating an inclusion list of masses (see Supporting Information (S.I.) **Table S1** for list) with a retention window of ±2 minutes and collision energy of 35%. Since MS/MS fragmentation was carried out in the ion trap, only nominal mass was measured.

Mass axis calibration was performed with every batch run just prior to starting the batch by using flow injection of a polytyrosine-1,3,6 solution ([M+H]+ 182.01170/508.20783 and 997.39781) at a flow rate of 10 µL min^{-1} .

161 *2.5 Settings of the Deconvolution Tools*

162 MsXelerator (MsMetrix)

163 MS Compare and MPeaks are modules within MsXelerator. MS Compare is specifically
164 designed for comparing MS spectra, whereas the MPeaks module picks peaks, with the
165 "keep largest C13 peaks only" and the "peak cluster" algorithms used to help discard
166 some, the latter performing componentization which groups together all peaks (i.e.
167 isotopes and adducts) arising from a single retention time (Table S2). All samples were
168 uploaded individually and later investigated as triplicates, corresponding to the three
169 triplicate injections of each sample. Procedural blank samples were initially processed
170 using the optimized software settings (see below) to subtract identical peaks from each
171 wastewater sample.

172 The "peak picking" was carried out by MPeaks on each individual sample using the
173 following parameters and values: Base peak width =11 (arbitrary units); spike width = 5
174 scans; peak separation = 5 scans; peak threshold = 0.5% (vs. largest peak); smoothness
175 threshold = 0.65%, signal/noise ratio = 20. The sensitivity value, which helps the user
176 find more or less sensitive parameters for the previous three parameters, was set relatively
177 low, at 2 (out of a maximum setting of 6). The peaks picked using these parameters were
178 further reduced by only keeping peaks relating to an $[M+H]^+$ charge state.

179 Using the second module, MS Compare, all samples were subjected to the following
180 LC/MS settings (in accurate mass mode) of the module for peak picking across multiple
181 samples: No baseline correction; FWHM (scans) = 3 (min) - 40 (max); min peak height
182 = 10,000 counts; delete spikes; *m/z* range: 100-650; Max. shift between peaks for
183 grouping: 20 scans; Time window for XIC: 0.25min; Mass accuracy: 10 ppm.

184 Sieve 2.1 (Thermo Scientific)

185 Sieve 2.1 combines the power of the two modules from MsXelerator. After an initial
186 peak-picking process using the settings described below, it compares MS spectra of the
187 procedural blank samples and the studied wastewater samples. Only compounds with an
188 *m/z* between 150 and 500, and only protonated molecules ($[M+H]^+$) were considered. The
189 list of potential positives were then search by either the in-house database or ChemSpider,
190 which were incorporated in the software. The results of the ChemSpider search were
191 exported into Microsoft Excel, and positive "hits" were considered based on their mass

error (< 2 ppm), and if the compound commercially existed i.e. hits which only represented chemical formula were excluded.

The "control compare trend" feature of Sieve 2.1 (Thermo Scientific) was used with the following parameters: peak intensity threshold = 250,000 (62,500 for IWW); m/z range = 100-650; m/z width = 10ppm; retention time range = 3-40 minutes; maximum number of frames = 5,000; frame time width = 1.00 minute; align bypass = true. For the hidden target screening, either the database used in the analyses of MPeaks and MS Compare or ChemSpider was incorporated.

2.6 General workflow

The general workflow followed in this work (pictorialized in Figure 1) falls into the "hidden target" area of non-target screening, hypothesized by Letzel *et al.* (Letzel et al., 2015), wherein non-target techniques (i.e. peak picking) are originally applied, but a database (i.e. in-house database or ChemSpider) is used for identification.

All samples were injected in triplicate and the data were processed with the different software packages of MsXelerator or Sieve 2.1. Only peaks in all three injections were further investigated. This resulted in a list of chromatographic peaks, based on the accurate masses of their protonated molecules. To gain a list of potential positives, two hidden target identification methods were used: 1) an in-house database, containing more than 200 parent compounds and metabolites and the online database ChemSpider. False positives were manually removed after investigating the isotopic pattern (for the characteristic patterns of sulfur- and chlorine-containing species) and retention time prediction. A final "target list" was investigated by reinjecting the samples in MS/MS mode, to get product ions. Fragmentation was then compared with online databases and literature, which allowed the tentative identification of several compounds.

3. Results and Discussion

In this study, in order to show the progression through confidence levels of identification, the terminology proposed in the literature by Hernández *et al.* (Hernández et al., 2015) and Schymanski *et al.* (Schymanski et al., 2014) were followed. It must be noted that potential positives and detected compounds, differentiated in this work, would both be level 3 tentative candidate in the terminology of Schymanski *et al.* The final, tentatively identified compounds are of a higher confidence level (level 2a). However, in order to have total confirmation (level 1), reference standards are necessary. As no reference standards were utilized in this work, this level could not be attained.

3.1 Optimization of the workflow

All samples were injected and processed in triplicate, which were compared together, with only peaks in all three injections being further investigated. Procedural blank samples were processed first, to subtract identical peaks from each subsequent IWW and EWW sample.

The *m/z* range of MS Compare was made quite narrow as the compounds of interest in this study and in the in-house database (small pharmaceutical/drug molecules) would be within that range. The retention time range was reduced just to 3-40 mins to reduce the likelihood of erroneously detecting species that elute very early and late due to the high/low ratio of organic modifier, with the vast majority of all peaks in the total ion chromatogram falling within this range. In spite of the known mass accuracy capability of the Orbitrap, the mass accuracy was set at 10ppm to ensure that no compound would be missed. After this processing, a list of masses common within each triplicate set was made, with compounds being detected using the same peak peaking parameters and database used in the final step of the MPeaks analysis.

Sieve 2.1 used the same *m/z* range, *m/z* width and retention time range as MS Compare for better ease of results comparison. The peak intensity threshold was originally set quite high for both IWW and EWW samples, but it was later found that IWW gave fewer peaks, possibly due to the complexity of these samples and stronger matrix effects, mostly leading to ionization suppression. The threshold was thus reduced to one quarter to account for this. The maximum number of components was raised to 10,000 to ensure that no compounds would be missed, leading to more than 5,000 components being detected in the IWW and EWW samples (**Table 1**). These were reduced by including

compounds with only a *m/z* 150-500 and $[M+H]^+$ =1. The in-house database used by the previous two modules within the "Accurate Mass Identification Parameters" of Sieve was then used to gain a list of potential positives.

The ChemSpider database (with 10ppm mass accuracy threshold) was also used within Sieve and was begun after the initial component optimization was completed (**Figure 2**). The threshold was made quite high, for optimal "hidden target" analysis, where the detected peaks should correspond to compounds which are commonly and/or highly used. The peak lists of both IWW and EWW results with all data pertaining to mass error, *m/z* and intensity were exported into Microsoft Excel. From these lists, several thresholds were set and outlying peaks removed: only compounds between m/z 150-500; only $[M+H]^+$; mass error under 2.0ppm; all "hits" just representing a chemical structure, rather than a generic/known name. This final step is rather pragmatic but makes for a more optimal non-target screening, where the remaining compounds should be the more common and/or highly used, as emphasized by having a high intensity threshold. However, this could lead to some less intense peaks being missed and not noted as a possible emerging contaminant in the environment.

3.2 Identification with in-house database

Both programs incorporated an in-house database of more than 200 pharmaceuticals, illicit drugs and metabolites (**Table S3**) to get a list of potential positives. All samples were first processed with MsXelerator (modules MPeaks and MSCompare) and Sieve 2.1 using the parameters outlined in Section 2.5. **Table 2** shows the compounds detected by each.

There was very little difference between the compounds found with Sieve and MS Compare, while MPeaks detected somewhat fewer compounds. This could be due to their apparent uses: MPeaks is for pick-peaking, MS Compare for comparing samples, while Sieve does both, resulting in the latter two have more similar results. The fact that all compounds detected by MPeaks were also found with MS Compare leads to the preferential use of the latter module for screening. However, by optimizing the peak-peaking parameters of MPeaks, specifically the sensitivity value, this module could also be of future use in suspect and/or non-target screening.

Two methods were used to remove potential false positive peaks: isotopic pattern (for chlorine- and sulfur-containing species) and retention time prediction. Only three of the above compounds (losartan, sulfamethoxazole and temazepam) had a chlorine or sulfur atom, giving rise to a characteristic isotopic pattern. Extracted ion chromatograms were extracted from the initial full-scan data of the Orbitrap and investigated manually. Both sulfamethoxazole and losartan showed the characteristic isotopic pattern, while temazepam did not. Temazepam was thus considered as a false positive and removed from further investigation.

A retention time predictor was made, based on artificial neural networks, as in (Miller et al., 2013; Munro et al., 2015) and in our previous work (Bade et al., 2015a). A retention time window of ±11% of total run time was used to find compounds to focus on, based on the window used in our previous work. Of the 25 potential positives investigated, four were removed using this method (benzocaine, ibuprofen, lincomycin and salbutamol) with predicted retention times between 11.5-16.8 minutes (24-36% of the total run time) away from the experimental times. While only four compounds were removed using this technique, it does simplify the identification process, and provides greater confidence in the compounds remaining.

3.3 Identification with ChemSpider

To make a more comprehensive analysis of the samples, an investigation was made using the ChemSpider database search feature of Sieve (**Figure 2**).. The introduction of ChemSpider, while removing many components, had the added complication of isobaric and isomeric compounds, with most distinct *m/z* values having more than one compound associated, as seen in step 5 of **Table 1**. To further refine this list, the mass error was limited to 2ppm (step 6) and all compounds having a formula-only entry were deleted, leaving just compounds with generic names (step 7), leaving up to 100 components in the samples. The literature was then searched to determine whether or not their detection in wastewater could be expected, leading to approximately 30 components and up to 34 isomeric/isobaric compounds in the samples. The literature search was made using the Scopus database, and search terms were the generic name of interest, “HRMS”, “LC”, “environment” and “water”. If there were no suitable papers concerning the generic name of interest, the compound was removed from further investigation. To determine which of the isomeric/isobaric compounds the compound within the sample was, the molecular

formula was manually searched on ChemSpider, with the compound having the highest number of references deemed to be the compound of interest. Step 7 and the literature search, while pragmatic, were employed to ensure that the compounds detected were those of high consumption/prescription and could therefore be more easily identified in the later *in silico* fragmentation comparison. Finally, eighteen (including three isomers) and eight compounds were finally deemed as potential positives using this non-target approach for IWW and EWW samples respectively (**Table S4**).

It is worth noting that by using this approach, most of the same compounds were found as with the in-house database (**Table 2** and **S4**). With such great similarities between the set of potential positive compounds, only 2-hydroxy carbamazepine, desvenlafaxine, adenosine, albendazole, phenazone and the three isomers theophylline, paraxanthine and theobromine required further investigation. Albendazole was the only compound that required an investigation of isotopic pattern as it contains one sulfur atom, which was inconsistent with the mass spectrum, leading to its removal as a false positive. The remaining seven compounds were subjected to retention time prediction based on the time given by Sieve 2.1, and all were found within the set ±11% of total run time retention time window.

3.4 Tentative identification

The potential positives found using both hidden target screening approaches were combined, less those removed in previous steps, leaving 28 compounds to investigate (**Table S1**). These compounds were added to a target list and several IWW and EWW samples were reinjected to see if fragment ions from these compounds could give further confidence to their identification. Metfusion and MassBank were used to help provide further confidence to the fragment ions. As has been mentioned in previous suspect and non-target studies (Agüera et al., 2013; Herrera-Lopez et al., 2014; Zedda and Zwiener, 2012), the use and improvement of mass spectral databases, such as MassBank, is extremely important in the tentative identification of compounds for which standards are unavailable. In the end, 22 compounds were able to be tentatively identified (**Table 3**) with at least one fragment ion, while the other six were removed as false positives due to having incorrect fragment ions.

One interesting finding was the detection of three isomers (paraxanthine, theobromine and theophylline). These three isomers are all metabolites of caffeine, accounting for

80%, 11% and 4% of total metabolism, respectively (Miners and Birkett, 1996). Conventionally, isomeric compounds, separated chromatographically, would be distinguished by retention time. However, as no standards were available, the best way to order the peaks was with retention time prediction. The approach outlined in Sieve 2.1 combines the power of the two modules from MsXelerator as described above. After an initial peak-picking process using the settings described in Section 2.5, it compares MS spectra of the procedural blank samples and the studied wastewater samples. Only compounds with an m/z between 150 and 500, and only protonated molecules ($[M+H]^+$) were considered. The list of potential positives were then search by either the in-house database or ChemSpider, which were incorporated in the software. The results of the ChemSpider search were exported into Microsoft Excel, and positive "hits" were considered based on their mass error (< 2 ppm), and if the compound commercially existed i.e. hits which only represented chemical formula were excluded. predicted retention times of 8.48 min, 9.34 min and 9.41 min for theobromine, paraxanthine and theophylline, respectively. While these times are 1-2 minutes from the experimental retention time, they do provide an idea for the order of the isomers. To give more confidence to this information, the fragment ions were checked. As seen in **Figure 3**, the peak at 6.27 min had fragment ions of *m/z* 163 and 138 while the peaks at 7.98 and 8.37 both had one major peak of *m/z* 124. These fragment ions were checked and compared with MassBank and the literature (Bianco et al., 2009; Gómez et al., 2010; Horai et al., 2010).. Theobromine was found to have fragment ions of *m/z* 163 and 138, while both paraxanthine and theophylline were found to have a main fragment ion of *m/z* 124. The losses leading to each fragment ion is defined in **Figure 3**. To differentiate the latter two, the initial retention time predictions led to paraxanthine being the larger peak at 7.98 min and theophylline the small peak at 8.37 min.

While 22 compounds were tentatively identified using the workflow outlined throughout this paper, it must be noted that even incorporating the false positive removal strategies of retention time prediction and isotopic pattern as well as fragment ions, the final confirmation of the identity of compounds requires the use of reference standards. Nevertheless, the addition of advanced deconvolution tools (MsXelerator and Sieve) to the HRMS data of the Orbitrap has been shown to be of great value, and the results show how far one can go without the need to purchase reference standards. The information obtained with this strategy circumvents the cost and problems associated with the storage

and expiry dates of standards in the laboratories, as the purchase can be directed only towards those compounds that have been previously tentatively identified in the samples.

Conclusion

This work has shown that, following initial unbiased, non-target oriented deconvolution using two tools(MsXelerator and Sieve), allowed relevant peaks of interest to be attained. The complementary use of an in-house database or ChemSpider facilitated detection and enabled the identification of more compounds than using just one of these databases.

The combination of deconvolution tools and high resolution mass spectrometry, without the use of any reference standards, has enabled 22 compounds to be tentatively identified in environmental water samples. The majority of compounds that were identified in wastewater samples were pharmaceuticals, including the metabolites 4-formylamino antipyrine, 4-acetylamino antipyrine, theobromine, theophylline, paraxanthine and carboxylosartan.

It is worth noting that the two hidden target approaches primarily found the same compounds, with some exceptions. Furthermore, when applying small databases it is often easier to analyse the raw data directly. Whereas, when a much larger database is incorporated, these software tools will facilitate searching as well as reducing processing time. With further improvements to these computational programs non-target analysis will become more enticing and easier for laboratories to use in everyday screening methods.

Acknowledgements

Richard Bade and Ana Causanilles acknowledge the European Union for their Early Stage Researcher (ESR) contracts as part of the EU-International Training Network SEWPROF (Marie Curie- PEOPLE Grant #317205)

Part of this work was supported by the COST Action ES1307 “SCORE - Sewage biomarker analysis for community health assessment”.

The financial support of Generalitat Valenciana (Prometeo II 2014/023) and of the Spanish Ministry of Economy and Competitiveness (Project ref CTQ2015-65603) is also acknowledged by the authors of University Jaume I.

References

Aceña, J., Stampachiacchiere, S., Pérez, S., Barceló, D., 2015. Advances in liquid chromatography–high-resolution mass spectrometry for quantitative and qualitative environmental analysis. Anal. Bioanal. Chem. 6289–6299. doi:10.1007/s00216-015-8852-6

Agüera, A., Martínez Bueno, M.J., Fernández-Alba, A.R., 2013. New trends in the analytical determination of emerging contaminants and their transformation products in environmental waters. Environ. Sci. Pollut. Res. Int. 20, 3496–515. doi:10.1007/s11356-013-1586-0

Bade, R., Bijlsma, L., Miller, T.H., Barron, L.P., Sancho, J.V., Hernández, F., 2015a. Suspect screening of large numbers of emerging contaminants in environmental waters using artificial neural networks for chromatographic retention time prediction and high resolution mass spectrometry data analysis. Sci. Total Environ. 538, 934–941. doi:10.1016/j.scitotenv.2015.08.078

Bade, R., Rousis, N.I., Bijlsma, L., Gracia-Lor, E., Castiglioni, S., Sancho, J. V., Hernandez, F., 2015b. Screening of pharmaceuticals and illicit drugs in wastewater and surface waters of Spain and Italy by high resolution mass spectrometry using UHPLC-QTOF MS and LC-LTQ-Orbitrap MS. Anal. Bioanal. Chem. 407, 8979–8988. doi:10.1007/s00216-015-9063-x

Bianco, G., Abate, S., Labella, C., Cataldi, T.R.I., 2009. Identification and fragmentation pathways of caffeine metabolites in urine samples via liquid chromatography with positive electrospray ionization coupled to a hybrid quadrupole linear ion trap (LTQ) and Fourier transform ion cyclotron resonance mass spec. Rapid Commun. Mass Spectrom. 23, 1065–1074. doi:10.1002/rcm.3969

Bijlsma, L., Emke, E., Hernandez, F., de Voogt, P., 2012. Investigation of drugs of abuse and relevant metabolites in Dutch sewage water by liquid chromatography coupled to high resolution mass spectrometry. Chemosphere 89, 1399–1406. doi:10.1016/j.chemosphere.2012.05.110

Bijlsma, L., Emke, E., Hernández, F., de Voogt, P., 2013. Performance of the linear ion trap Orbitrap mass analyzer for qualitative and quantitative analysis of drugs of abuse and relevant metabolites in sewage water. Anal. Chim. Acta 768, 102–110. doi:10.1016/j.aca.2013.01.010

Bletsou, A.A., Jeon, J., Hollender, J., Archontaki, E., Thomaidis, N.S., 2015. Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. Trends Anal. Chem. 66, 32–44.

de Voogt, P., Emke, E., Helmus, R., Panteliadis, P., van Leerdam, J.A., 2011. Determination of illicit drugs in the water cycle by LC-Orbtitrap MS, in: Castiglioni, S., Zuccato, E., Fanelli, R. (Eds.), Illicit Drugs in the Environment: Occurence, Analysis, and Fate Using Mass Spectrometry. John Wiley & Sons, Ltd., pp. 87–114.

Fatta-Kassinos, D., Meric, S., Nikolaou, A., 2011. Pharmaceutical residues in environmental waters and wastewater: Current state of knowledge and future research. Anal. Bioanal. Chem. 399, 251–275. doi:10.1007/s00216-010-4300-9

Gago-Ferrero, P., Schymanski, E.L., Bletsou, A.A., Aalizadeh, R., Hollender, J., Thomaidis, N.S., 2015. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. Environ. Sci. Technol. 49, 12333–12341. doi:10.1021/acs.est.5b03454

Gerlich, M., Neumann, S., 2013. MetFusion: integration of compound identification strategies. J. Mass Spectrom. 48, 291–8. doi:10.1002/jms.3123

Gómez, M.J., Gómez-Ramos, M.M., Malato, O., Mezcua, M., Férnandez-Alba, A.R., 2010. Rapid automated screening, identification and quantification of organic micro-contaminants and their main transformation products in wastewater and river waters using liquid chromatography-quadrupole-time-of-flight mass spectrometry with an accurate-mass . J. Chromatogr. A 1217, 7038–54. doi:10.1016/j.chroma.2010.08.070

Helbling, D.E., Hollender, J., Kohler, H.-P.E., Singer, H., Fenner, K., 2010. High-throughput identification of microbial transformation products of organic micropollutants. Environ. Sci. Technol. 44, 6621–7. doi:10.1021/es100970m

Hernández, F., Ibáñez, M., Bade, R., Bijlsma, L., Sancho, J.V., 2014. Investigation of pharmaceuticals and illicit drugs in waters by liquid chromatography-high-resolution mass spectrometry. TrAC Trends Anal. Chem. 63, 140–157. doi:10.1016/j.trac.2014.08.003

Hernández, F., Ibáñez, M., Botero-Coy, A.-M., Bade, R., Bustos-López, M.C., Rincón, J., Moncayo, A., Bijlsma, L., 2015a. LC-QTOF MS screening of more than 1,000 licit and illicit drugs and their metabolites in wastewater and surface waters from the area of Bogotá, Colombia. Anal. Bioanal. Chem. 407, 6405–6416. doi:10.1007/s00216-015-8796-x

Hernández, F., Ibáñez, M., Portolés, T., Cervera, M.I., Sancho, J. V, López, F.J., 2015b. Advancing towards universal screening for organic pollutants in waters. J. Hazard. Mater. 282, 86–95. doi:10.1016/j.jhazmat.2014.08.006

Hernández, F., Pozo, Ó.J., Sancho, J. V., López, F.J., Marín, J.M., Ibáñez, M., 2005. Strategies for quantification and confirmation of multi-class polar pesticides and transformation products in water by LC–MS2 using triple quadrupole and hybrid quadrupole time-of-flight analyzers. TrAC Trends Anal. Chem. 24, 596–612. doi:10.1016/j.trac.2005.04.007

Hernández, F., Sancho, J. V, Ibáñez, M., Abad, E., Portolés, T., Mattioli, L., 2012. Current use of high-resolution mass spectrometry in the environmental sciences. Anal. Bioanal. Chem. 403, 1251–64. doi:10.1007/s00216-012-5844-7

Herrera-Lopez, S., Hernando, M.D., García-Calvo, E., Fernández-Alba, a. R., Ulaszewska, M.M., 2014. Simultaneous screening of targeted and non-targeted contaminants using an LC-QTOF-MS system and automated MS/MS library searching. J. Mass Spectrom. 49, 878–893. doi:10.1002/jms.3428

Hogenboom, A.C., van Leerdam, J.A., de Voogt, P., 2009. Accurate mass screening and identification of emerging contaminants in environmental samples by liquid chromatography–hybrid linear ion trap Orbitrap mass spectrometry. J. Chromatogr. A 1216, 510–519. doi:10.1016/j.chroma.2008.08.053

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda,

F., Sawada, Y., Hirai, M.Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, K., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., Nishioka, T., 2010. MassBank: a public repository for sharing mass spectral data for life sciences. J. Mass Spectrom. 45, 703–14. doi:10.1002/jms.1777

Hug, C., Ulrich, N., Schulze, T., Brack, W., Krauss, M., 2014. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. Environ. Pollut. 184, 25–32. doi:10.1016/j.envpol.2013.07.048

Kern, S., Fenner, K., Singer, H.P., Schwarzenbach, R.P., Hollender, J., 2009. Identification of Transformation Products of Organic Contaminants in Natural Waters by Computer-Aided Prediction and High-Resolution Mass Spectrometry. Environ. Sci. Technol. 43, 7039–7046. doi:10.1021/es901979h

Kind, T., Fiehn, O., 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics 8, 105. doi:10.1186/1471-2105-8-105

Krauss, M., Singer, H., Hollender, J., 2010. LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns. Anal. Bioanal. Chem. 397, 943–951. doi:10.1007/s00216-010-3608-9

Leendert, V., Van Langenhove, H., Demeestere, K., 2015. Trends in liquid chromatography coupled to high-resolution mass spectrometry for multi-residue analysis of organic micropollutants in aquatic environments. TrAC Trends Anal. Chem. 67, 192–208. doi:10.1016/j.trac.2015.01.010

Letzel, T., Bayer, A., Schulz, W., Heermann, A., Lucke, T., Greco, G., Grosse, S., Schüssler, W., Sengl, M., Letzel, M., 2015. LC – MS screening techniques for wastewater analysis and analytical data handling strategies : Sartans and their transformation products as an example. Chemosphere 137, 198–206. doi:10.1016/j.chemosphere.2015.06.083

Little, J.L., Williams, A.J., Pshenichnov, A., Tkachenko, V., 2012. Identification of "Known Unknowns" Utilizing Accurate Mass Data and ChemSpider. J. Am. Soc. Mass Spectrom. 23, 179–185. doi:10.1007/s13361-011-0265-y

Luo, Y., Guo, W., Ngo, H.H., Nghiem, L.D., Hai, F.I., Zhang, J., Liang, S., Wang, X.C., 2014. A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. Sci. Total Environ. 473-474, 619–41. doi:10.1016/j.scitotenv.2013.12.065

Makarov, A., Scigelova, M., 2010. Coupling liquid chromatography to Orbitrap mass spectrometry. J. Chromatogr. A 1217, 3938–45. doi:10.1016/j.chroma.2010.02.022

Miller, T.H., Musenga, A., Cowan, D.A., Barron, L.P., 2013. Prediction of chromatographic retention time in high-resolution anti-doping screening data using artificial neural networks. Anal. Chem. 85, 10330–7. doi:10.1021/ac4024878

Miners, J.O., Birkett, D.J., 1996. The use of caffeine as a metabolic probe for human drug metabolizing enzymes. Gen. Pharmacol. 27, 245–249. doi:10.1016/0306-3623(95)02014-4

Munro, K., Miller, T.H., Martins, C.P.B., Edge, A.M., Cowan, D. a., Barron, L.P., 2015. Artificial neural network modelling of pharmaceutical residue retention times in

wastewater extracts using gradient liquid chromatography-high resolution mass spectrometry data. J. Chromatogr. A 1396, 33–44. doi:10.1016/j.chroma.2015.03.063

Schollée, J.E., Schymanski, E.L., Avak, S.E., Loos, M., Hollender, J., 2015. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. Anal. Chem. acs.analchem.5b02905. doi:10.1021/acs.analchem.5b02905

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014a. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. Environ. Sci. Technol. 48, 2097–2098. doi:10.1021/es5002105

Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés Vidal, C., Hollender, J., 2014b. Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry. Environ. Sci. Technol. 48, 1811–8. doi:10.1021/es4044374

van der Aa, M., Bijlsma, L., Emke, E., Dijkman, E., van Nuijs, A.L.N., van de Ven, B., Hernández, F., Versteegh, A., de Voogt, P., 2013. Risk assessment for drugs of abuse in the Dutch watercycle. Water Res. 47, 1848–57. doi:10.1016/j.watres.2013.01.013

Zedda, M., Zwiener, C., 2012. Is nontarget screening of emerging contaminants by LC-HRMS successful? A plea for compound libraries and computer tools. - Anal. Bioanal. Chem. 2493–2502. doi:10.1007/s00216-012-5893-y

566 **Table 1:** *Number of components after each step of the Sieve hidden target identification approaches..*

567

568

Step	**IWW components**	**EWW components**
1.	6690	5091
2.	5158	3528
3.	2014	2175
4.	18	16
NON-TARGET SCREENING		
Number of distinct *m/z* (total number of compounds)		
5.	239 (437)	441 (677)
6.	166 (362)	308 (543)
7.	100 (150)	64 (108)
Final	18	8

569

570

571 **Table 2:** *All compounds detected (in at least one sample) by each program following suspect screening.*

Compound	IWW			EWW		
	Sieve	MSCompare	Mpeaks	Sieve	MSCompare	Mpeaks
4-acetylamino antipyrine						
4-formylamino-antipyrine						
Acetaminophen						
Benzocaine						
Benzoylecgonine						
Caffeine						
Carbamazepine						
Cocaine						
Cotinine						
Gemfibrozil						
Ibuprofen						
Irbesartan						
Ketoprofen						
Lidocaine						
Lincomycin						
Losartan						
Metoprolol						
Naproxen						
Phenacetin						
Phenytoin						
Salbutamol						
Sulfamethoxazole						
Temazepam						
Trimethoprim						
Valsartan						

572

	Detected by the program
	Not detected by the program

573

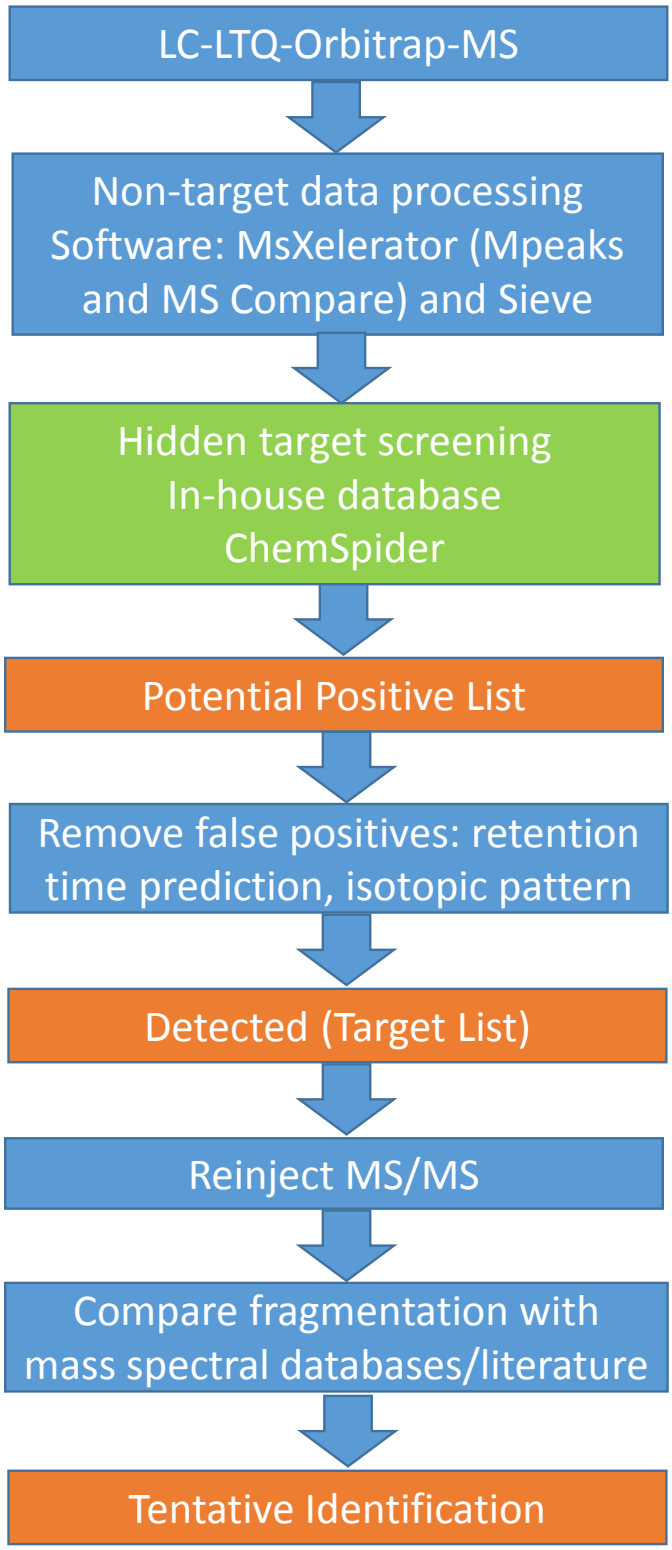
574

575 **Table 3:** *All compounds tentatively identified (level 2a), together with retention time and fragment ions*

Compound	m/z	RT	Fragment ions			IWW	EWW
4-acetylaminoantipyrine[1]	246.1234	9.43	228.1	204.1		X	X
4-formylaminoantipyrine[2]	232.1086	9.29	214.1	204.1		X	X
Acetaminophen[3]	152.0706	6.09	110.1	134.0		X	X
Adenosine[2]	268.1035	3.58	136.0				X
Benzoylecgonine[2]	290.1385	12.22	168.1			X	X
Caffeine[1]	195.0876	10.4	138.1			X	X
Carbamazepine[3]	237.1022	22.56	194.1	152.9		X	X
Carboxylosartan[1]	437.1480	26.94	207.1	235.1	365.3	X	X
Cocaine[2]	304.1543	13.84	182.2			X	
Ketoprofen[3]	255.1014	16.78	237.1	209.1		X	X
Lidocaine[1]	235.1807	10.12	86.1			X	
Losartan[1]	423.1695	25.58	405.0	207.2	377.2	X	X
Metoprolol[2]	268.1908	13.71	218.1	191.1	159.1	X	
Naproxen[3]	231.1016	27.24	185.1			X	X
Paraxanthine[2]	181.0721	7.98	124.1			X	X
Phenacetin[2]	180.1030	17.22	138.1	110.0		X	X
Phenazone[2]	189.1022	12.15	161.2	146.1	131.1	X	X
Sulfamethoxazole[2]	254.0594	12.41	235.8	188.1	156.1	X	X
Theobromine[2,4]	181.0721	6.28	163.1	137.1	138.1	X	X
Theophylline[2,4]	181.0721	8.37	124.1			X	X
Trimethoprim[2]	291.1454	9.91	230.2	123.2	261.1	X	X
Valsartan[2]	436.2341	28.96	335.1	265.2	155.1	X	X

576 RT = retention time (minutes)
577 The parent ions were recorded at accurate mass (full-scan mode) while the fragment
578 ions were recorded as part of a product ion scan in the ion trap part with nominal mass
579 measurement.
580 [1] Information on fragment ions from Hernández *et al* (Hernández et al., 2015a)
581 [2] Information on fragment ions from MassBank (Horai et al., 2010)
582 [3] Information on fragment ions from Bade *et al* (Bade et al., 2015b)
583 [4] Information on fragment ions from Gómez *et al* (Gómez et al., 2010)
584

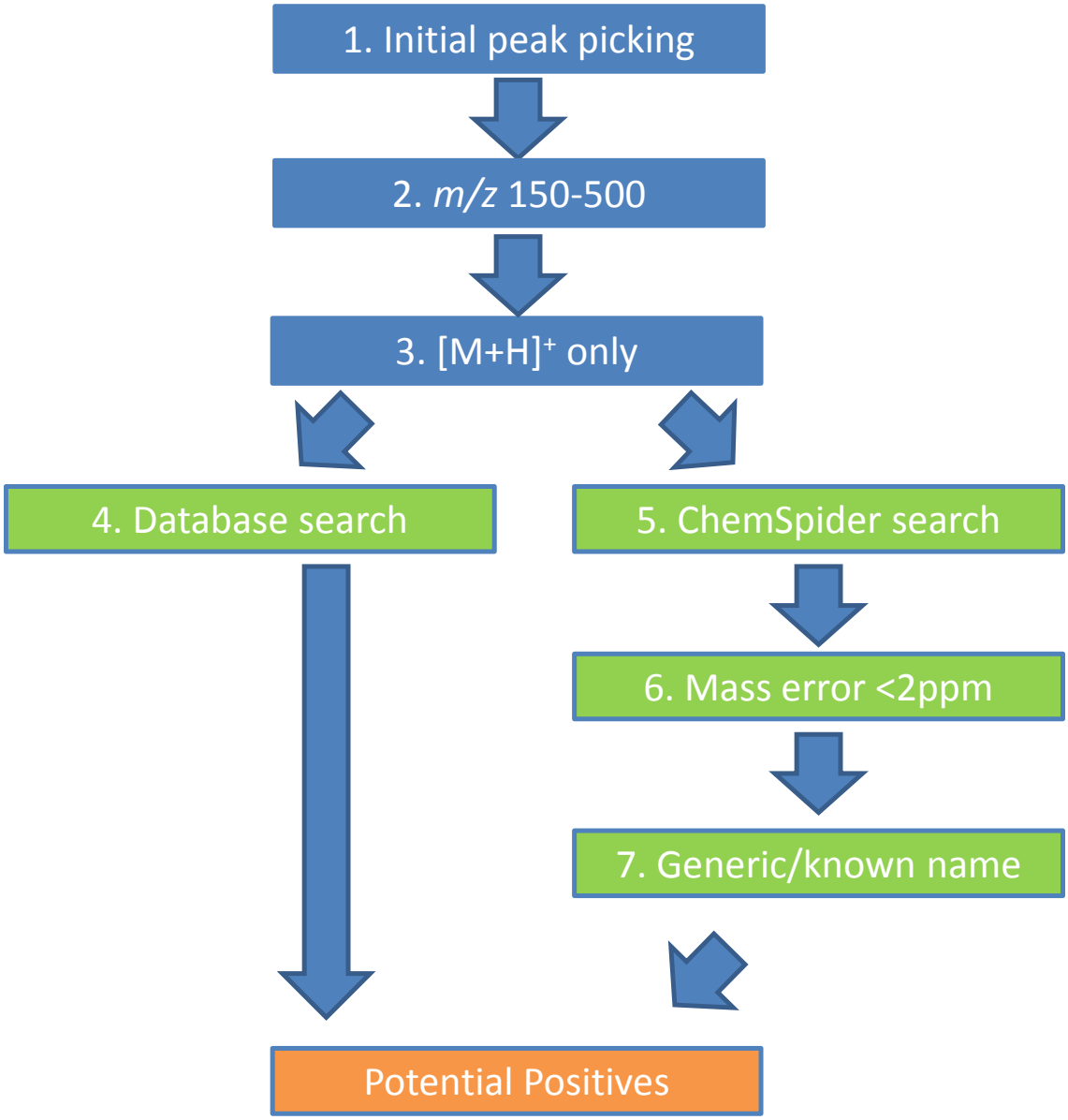
585



586

587 **Figure 1:** *Workflow for screenings using the deconvolution tools MsXelerator and Sieve. All orange levels*
588 *represent specific identification confidence levels*

589



590

591 **Figure 2:** *Sieve workflow for the two hidden target identification approaches.*

592

593

594

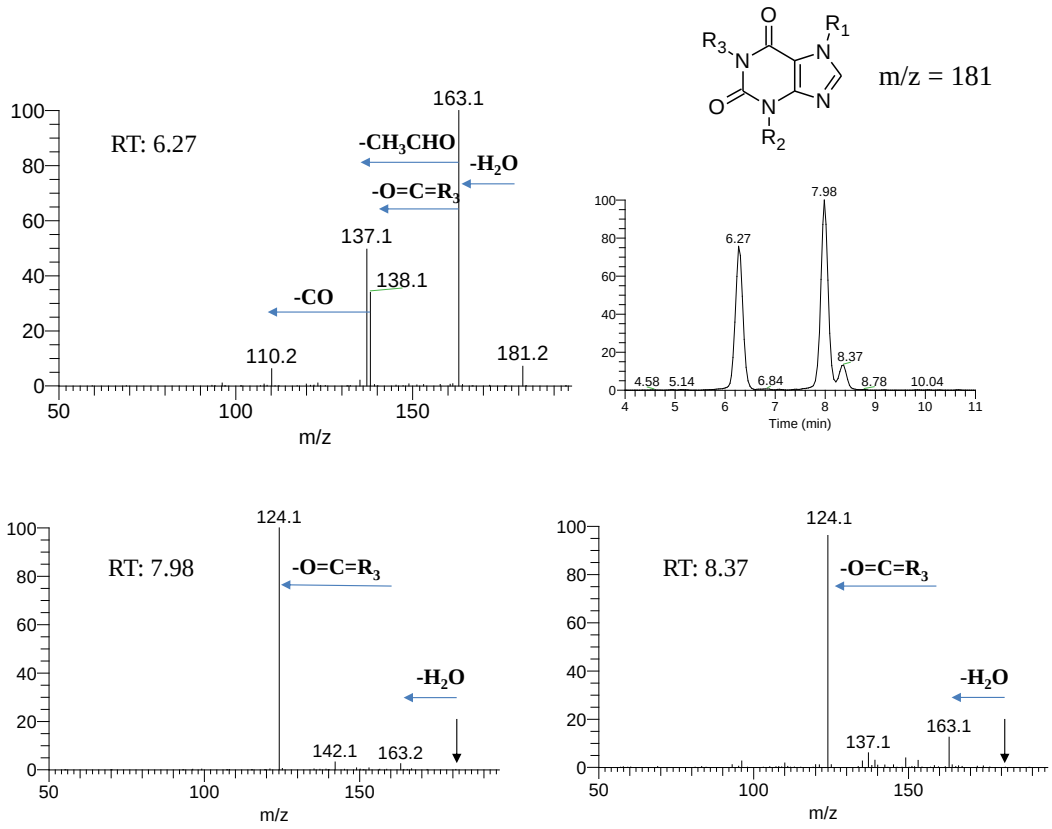


Figure 3: *Tentative identification of theobromine (top left), paraxanthine (bottom left) and theophylline (bottom right), with chromatographic peaks (top right). The generic structure has been shown in the top right corner, where R_1, R_2 and R_3 differ for the metabolites as follows: theobromine: R_1 =CH_3, R_2 = CH_3 and R_3 = H; theophylline: R_1 =H, R_2 = CH_3 and R_3 = CH_3; paraxanthine: R_1 =CH_3, R_2 = H and R_3 = CH_3*