

Two Nontargeted Screening Approaches for Examination of Drinking Water Before and After Treatment

KC Hyland¹, Amy Heffernan²

¹SCIEX, USA ²SCIEX, Australia

Two workflows based on liquid chromatography coupled to a quadrupole-time-of-flight mass spectrometer (LC-MS) were applied to detect and identify suspect and unknown species in raw and treated water samples collected from drinking water treatment trains. Candidate structure assignments were made based on experimentally derived high-resolution accurate mass data and MS/MS spectral interpretation (including comparison to spectral databases and *in silico* fragmentation predictions). Data were collected using a novel combination of Data Independent Acquisition (SWATH[®] Acquisition) and Data Dependent



Acquisition (IDA) scan types in a single acquisition. Corrosion inhibitors, artificial sweeteners, and pharmaceuticals were among the components to be detected and identified in the samples. Differences in contaminant profiles were observed between raw, treated, and advanced treated waters. High resolution-accurate mass (HRAM) data combined with processing and statistical software tools for nontargeted screening are critical to achieve candidate structure identification. The first of the two workflows described is a suspect screening approach which attempts to identify all features at once with a MS/MS library search. The second approach involves initial utilization of a statistical approach to pinpoint important differentiating features before attempting to assign candidate structural identification with library screening.

Key Advantages of the Nontargeted Screening Approaches

- SWATH Acquisition, employed in both approaches, ensures that MS/MS information will be available for all features detectable in the sample.
- Screening against the SCIEX All-in-One with NIST library allows for broad compound coverage in the first-pass suspect screening. Having greater coverage during the suspect screen reduces the manual structural elucidation needed by initially achieving more suggested candidate structure matches.
- MarkerView Software can be used to identify which features are unique to the different types of samples. Statistical tools in this platform allow for nuanced investigation of differences between sample sets, and characterization of the constituent profile of complex unknowns.

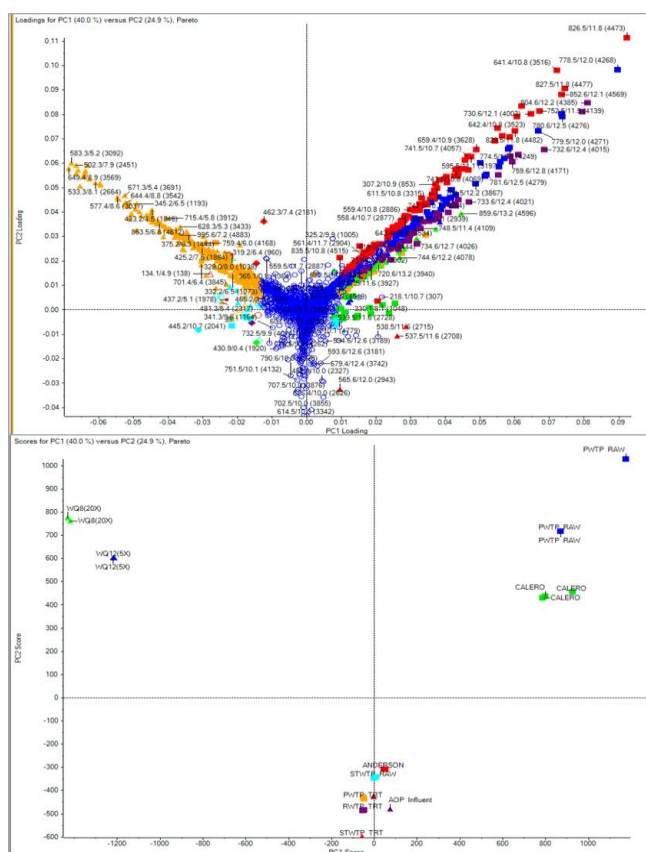


Figure 1. TOF MS Data (Positive Ionization Mode) in MarkerView Software. Loadings plot and PCA plots can be highly informative in identifying distinguishing features between sample sets. Loadings plot (top): data were normalized using the MLR method. Each point represents a peak feature, while the colors represent related feature groups. Principle Component Analysis plot (bottom) shows different samples clustered on a plot of PC1 vs. PC2. PC1 vs PC2 explained the greatest amount of variation, resulting in the most distinction between the sample type groups.

Experimental

Samples and Sample Preparation: Several different waters were screened for trace organic constituents.

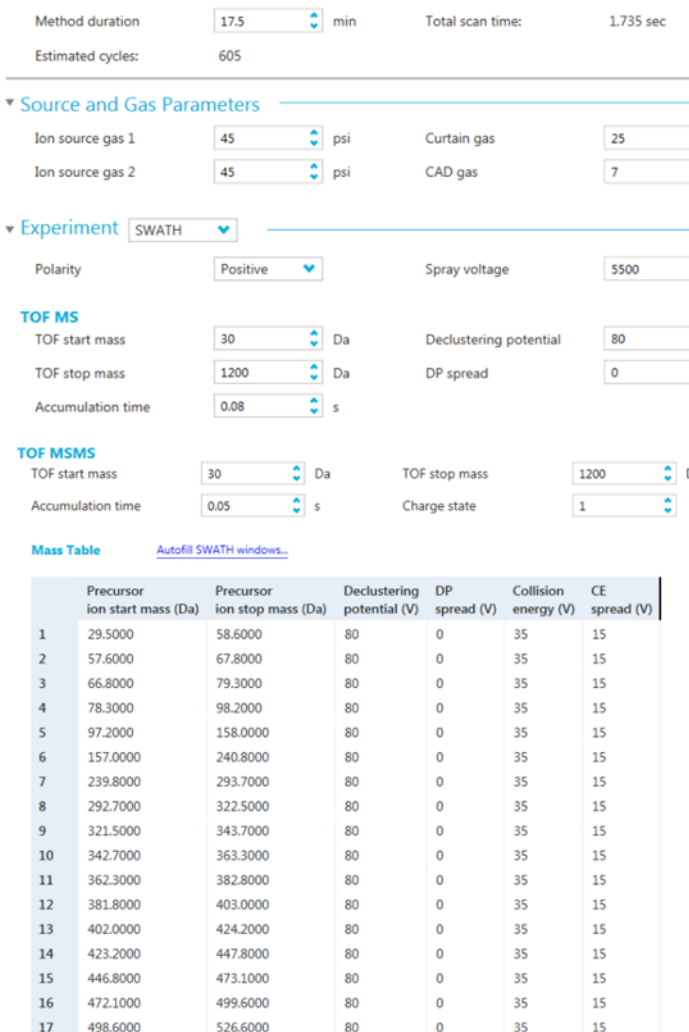
1. AOP Influent and Effluent: Influent or effluent of Advanced Oxidation Process (AOP) treatment.
2. Anderson, Calero: Reservoir water from two different California reservoirs
3. WQ12: Post microfiltration / pre-reverse osmosis (RO)
4. WQ8: Backflush of RO filter
5. PWTP or STWTP RAW: Drinking water treatment influent sampled from one of two different water treatment plants ("P" or "ST"). The influent is sourced from reservoirs.
6. PWTP or STWTP TRT: Drinking water treatment effluent sampled from one of two different water treatment plants ("P" or "ST").

Samples were concentrated using solid phase extraction (SPE) with Waters Oasis HLB SPE cartridges. The collected water sample (500 mL) was loaded onto the conditioned cartridge, rinsed, and eluted. This extract was dried under nitrogen and reconstituted to a volume of 100 μ L. The extract was then diluted with LC mobile phase for injection and analysis.

HPLC Conditions: LC separation was achieved using an ExionLC™ AD system with Phenomenex Kinetex 2.6 μ m C18 100Å 100 \times 2.1 mm column held at 30°C. A 16 minute gradient of eluent A (water with 0.1% formic acid) and eluent B (methanol with 0.1% formic acid) was used.

MS Parameters: SCIEX X500R QTOF LC-MS/MS system with Turbo V™ source and Electrospray Ionization (ESI) probe were used. Positive and negative modes were both acquired as separate injections. SWATH Acquisition was used to ensure that MS/MS spectra would be collected for all detectable ions, and the variable Q1 windows were defined in the SWATH acquisition method for optimizing the specificity of the collected MS/MS spectra in the mass regions where most of the sample constituents would be expected, best practices for application of SWATH acquisition in highly complex samples. Figure 2 illustrates the SWATH Acquisition method setup.

Data Processing: Library searching was performed using SCIEX OS Software. Statistical analysis of the SWATH acquisition data was performed using MarkerView Software (Principal Component Analysis (PCA) with Principal Component Variable Grouping (PCVG) and t-test), and differentiating features were then identified using Formula Finder in SCIEX OS Software.



Method duration: 17.5 min Total scan time: 1.735 sec
 Estimated cycles: 605

Source and Gas Parameters

Ion source gas 1: 45 psi Curtain gas: 25
 Ion source gas 2: 45 psi CAD gas: 7

Experiment SWATH

Polarity: Positive Spray voltage: 5500

TOF MS

TOF start mass: 30 Da Declustering potential: 80
 TOF stop mass: 1200 Da DP spread: 0
 Accumulation time: 0.08 s

TOF MSMS

TOF start mass: 30 Da TOF stop mass: 1200 Da
 Accumulation time: 0.05 s Charge state: 1

Mass Table [Autofill SWATH windows...](#)

	Precursor ion start mass (Da)	Precursor ion stop mass (Da)	Declustering potential (V)	DP spread (V)	Collision energy (V)	CE spread (V)
1	29.5000	58.6000	80	0	35	15
2	57.6000	67.8000	80	0	35	15
3	66.8000	79.3000	80	0	35	15
4	78.3000	98.2000	80	0	35	15
5	97.2000	158.0000	80	0	35	15
6	157.0000	240.8000	80	0	35	15
7	239.8000	293.7000	80	0	35	15
8	292.7000	322.5000	80	0	35	15
9	321.5000	343.7000	80	0	35	15
10	342.7000	363.3000	80	0	35	15
11	362.3000	382.8000	80	0	35	15
12	381.8000	403.0000	80	0	35	15
13	402.0000	424.2000	80	0	35	15
14	423.2000	447.8000	80	0	35	15
15	446.8000	473.1000	80	0	35	15
16	472.1000	499.6000	80	0	35	15
17	498.6000	526.6000	80	0	35	15

Figure 2. Nontargeted MS Acquisition Method Utilizing SWATH Acquisition with Variable Window Widths. From top to bottom, MS parameters listed include the ion source parameters (such as temperature, spray voltage, and gas flow), the TOF MS parameters (such as mass range and declustering potential, DP), and finally MS/MS collection parameters. These include the varying Q1 isolation windows listed as "Precursor start/stop mass" and the fragment ion mass range. The windows can also have individually designated declustering potential and collision energy settings, but for these experiments the DP / DP spread and Collision energy (CE) / CE spread values were kept to generic settings.

Approach #1: Suspect Screening

Suspect screening, or suspected-target screening, refers to the nontargeted-type screening workflow in which the data acquisition does not define target analytes, but the resulting data are assessed for qualitative identification of previously characterized constituents. These constituents are sometimes referred to as “Known Unknowns” – the species known in the chemical literature or MS reference databases but not defined in the acquisition method¹. After acquisition, the constituents or suspects are tentatively identified using high resolution-accurate mass information and MS/MS spectral data for comparison to analytical and chemical databases (Figure 3). This approach is advantageous in that candidate structures can be tentatively identified via mass spectrometric databases and/or reconciliation with *in silico* fragmentation predictions, even in the absence of a reference standard.

The data were processed using in SCIEX OS Software, using the Analytics module. An extraction blank was used as a control comparison, and the nontargeted feature-finding was set to process those features in the unknown samples which exceed the blank signal by at least 3x. Levels of confidence in compound identification achieved during the Suspect Screen can be defined in the data processing method using parameters such as mass error, fit scores to MS/MS spectra, retention time (if it is known),

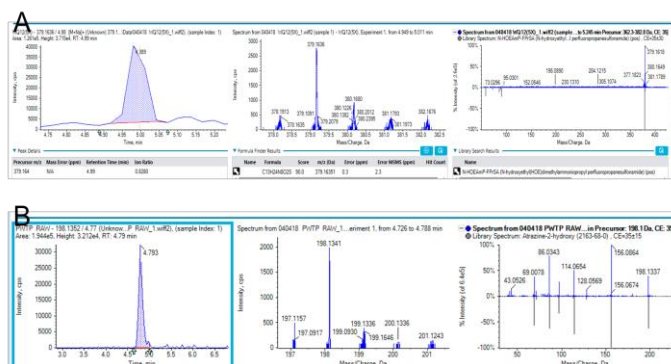


Figure 3. Suspect Screening Workflow. Utilizing a suspect screening approach resulted in identifying several candidate components in the various water samples which were screened. The resulting data from the nontargeted peak finding and subsequent MS/MS library searching includes, from left to right, the chromatographic peak which was found, the TOF MS spectrum and corresponding Formula Finder results, and lastly the acquired MS/MS spectrum and its corresponding candidate matches from the MS/MS database. In the above examples, these compounds were identified with excellent library match scores. A) A novel perfluorinated compound, N-HOEAmp-FPrSA, was tentatively identified based on MS/MS in the reverse osmosis (RO) backflush water. The match was made using the verified SCIEX Fluoros library v2.0. B) A transformation product of the common triazine herbicide, atrazine, was identified in the raw water at the start of the treatment train. Parent compound atrazine was not detected, so it's possible this transformation may have occurred in the environmental source water before arriving to the treatment plant where it was sampled.

and isotopic pattern. SCIEX OS software allows the user to set tolerance values for what is defined as a “match,” displayed as a green check mark for rapid visualization, filtering, and flagging. In this study, for library database searching, a Purity score of 70% or greater was defined as a match, and a mass error of 2ppm or less was defined as a match for Formula Finder results.

Approach #2: Statistical Analysis

While the workflow for Suspect Screening is straightforward and approachable, requiring little input for setup or comprehensive searching, it often results in a daunting amount of information that may be cumbersome to compare across multiple samples or sample groups. In recognition of this, a second workflow was applied which aims to first narrow down the list of detected peak features to be identified in the unknown samples. In this approach, the MarkerView software was utilized to facilitate a statistics-based approach for the characterization of the different water samples. The high resolution-accurate mass QTOF data were loaded into the MarkerView software to identify those mass features which differentiate the water samples from each other; these features were then populated into a Peaks of Interest list, to be used in conjunction with SCIEX OS software to suggest candidate identifications for these characteristic features. The workflow is outlined in Figure 4.

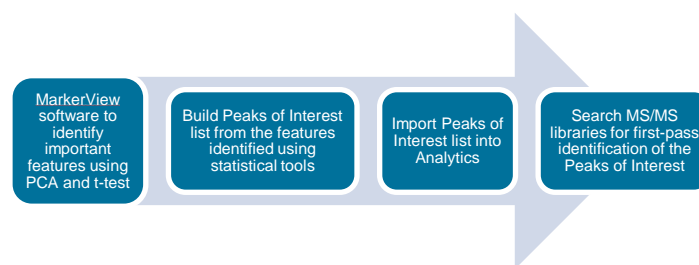


Figure 4. Statistical-Based Workflow. Workflow for using MarkerView software to mine the data for the most significant distinguishing features before performing a library search. In this workflow, the aim is to only ID those species which differentiate the different types of samples from one another.

The MarkerView software is a critical component of this workflow². The TOF MS data acquired for all samples can be loaded into this software, which then can extract all relevant detected features (unique m/z and retention time combinations) to produce a feature list which can then be subjected to a variety of statistical tests within the software interface. In Figure 1, the Scores and Loadings plots from the PCA analysis for the different water samples is shown. From these it becomes clear that the sample groups differ from each other based on the feature profiles of each. The next step is to investigate further which of these features are unique to one or more sample groups of interest. Figure 5 demonstrates one example of this, in

which features on the extreme ends of the loading plots are extracted to produce a profile plot. The features which are plotted at the farthest ends of the loading plot represent those which are most unique to the that PCA feature group (in this case, present in one sample group vs. relatively absent in another). This valuable information allows for the development of a list of peaks which, when identified (either tentatively or with more rigorous confirmation), will provide information on what differences or changes exist between samples.

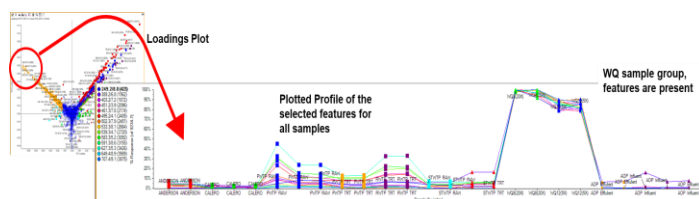


Figure 5. Loadings Plot to Quickly Find Features That Differentiate Samples. In this figure, features circled in the top right of the loadings plot have been highlighted. Extraction of these features shows the feature identifier (m/z plus retention time) for each of them in the legend, and they are all plotted with the intensity of the feature in each samples of the dataset. The plotted features are distinctly present in the WQ sample group relative to any of the others. These features are therefore identified as unique, unshared constituents of the WQ water samples, and may be of interest for further investigation/identification.

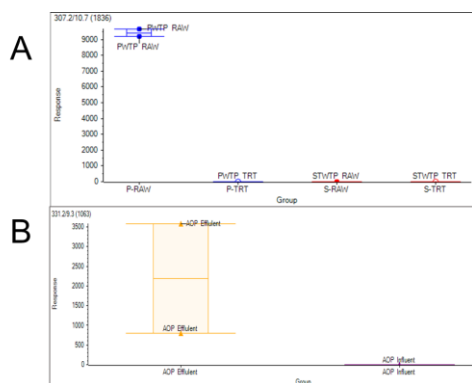


Figure 6. Box and Whisker Plots Showing Individual Peak Features Between Compared Sample Groups. A) Feature at m/z 307.2 and RT 10.7min is shown to be higher in the PWTP Raw vs. the PWTP treated. This species seems to be transformed during the PWTP treatment process. B) Feature at m/z 331.2 and RT 9.3min is shown to be higher in the AOP Effluent than the AOP Influent. This species may be one which is being generated during the oxidation process.

In this study, there was a specific interest to compare sample groups which represent a “before treatment” and an “after treatment” scenario. For example, the PWTP raw water versus the PWTP treated water, or the AOP influent versus the AOP effluent. These direct comparisons aim to discern what trace species are being transformed during the treatment process; which species are depleted during treatment, and which transformation products may be generated during treatment.

Figure 6 shows an example of two chemical features revealed by a t-test to be significantly different in the PWTP and AOP “before treatment” and “after treatment” scenarios. These are examples of features which would be added to the Peaks of Interest list for qualitative investigation. The t-test is also performed within the MarkerView software and can compare any sample group to another sample group, or to the rest of the groups.

Once the Peaks of Interest List is populated with the features extricated using the statistical tools in MarkerView Software, this list of “targets” can be processed in SCIEX OS software. Processing these features follows the same suspect screening logic; high resolution TOF MS data are used to generate candidate empirical formula for the features, and the MS/MS spectra are searched against databases for potential matches. Figure 7 shows how some example features from the t-test comparisons were matched to potential candidate structures.

Summary

Two high resolution mass spectrometric workflows were implemented with the goal of characterizing different trace- level organic constituents in water as it passes through various treatment processes. The first is a Suspect Screening workflow which finds feature peaks in a sample as compared to a control. Those features are investigated primarily through screening the acquired MS/MS information against a database and reporting library matches as candidate identifications. The second workflow has an additional step of initially employing statistical software to first narrow down the peaks list to only those most relevant to the sample set. Distinguishing, unique features are found with the aid of a PCA and t-tests and added to a Peaks of Interest list. These Peaks of Interest can then be investigated to achieve candidate structure information.

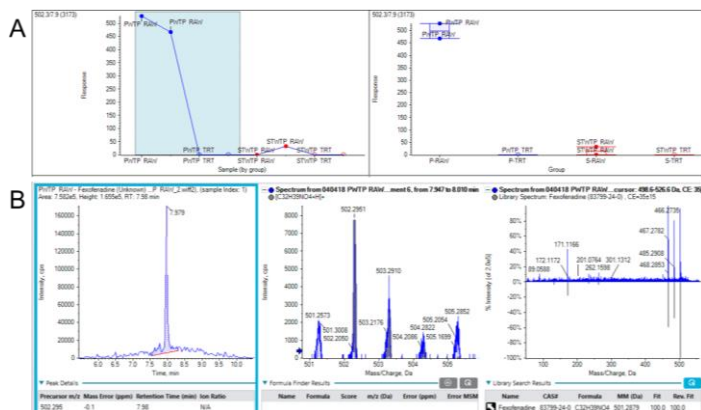


Figure 7. Compound Identification of Peaks from Interest List.

Features significantly different between PWTP raw and treated samples were processed in SCIEX OS software for candidate structure matches. A) Feature at m/z 502.3 and RT 7.9 min is shown to be higher in the PWTP Raw vs. the PWTP treated. This species seems to be transformed during the PWTP treatment process. B) From left to right is shown the chromatographic peak for the feature at RT 7.9 min; the TOF MS spectrum with its candidate result for empirical formula $[C_{32}H_{39}NO_4+H]^+$; the MS/MS spectrum shown with its match to library spectrum for the pharmaceutical compound, Fexofenadine. The empirical formula matches to Fexofenadine, and the product ion spectrum shows an excellent fit (Fit score of 100).

Both workflows represent valid approaches to a nontargeted analysis of contaminants in water samples. Depending on the search parameters, desired level of final detail in characterization, and sample complexity, the Suspect Screening may produce thousands of features for screening and identification from a single sample alone. This presents a potential challenge due to an overwhelming excess of data, making it more difficult to find the key differences between the sample sets (in this study, for example, differences between influent and effluent groups). By employing the MarkerView software to first find significantly distinctive features, the workflow becomes more tailored and the data processing more focused on characterization of those unknowns which are differentiating between samples in the study.

Some results from this study include the candidate identification of some structures in the collected wastewater samples, and some interesting findings regarding species which may be transformed. For example, during the Suspect Screening approach, a fluorinated species N-HOEAmp-FPrSA, was tentatively identified based on MS/MS in the reverse osmosis

AB Sciex is doing business as SCIEX.

© 2019 AB Sciex. For Research Use Only. Not for use in diagnostic procedures. The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.

Document number: RUO-MKT-02-8998-A

(RO) backflush water, a compound not routinely monitored. An atrazine metabolite, 2-hydroxy-atrazine, was detected in raw water entering the treatment plant PWTP and was also tentatively identified using Suspect Screening and MS/MS spectral matching. Interestingly, some tentative feature identifications in the PWTP raw water which appeared to be removed or transformed during treatment included multiple anthropogenic pharmaceutical compounds such as the Fexofenadine shown in Figure 7, as well as Epioxandrolone (a metabolite of pharmaceutical steroid hormone Oxandrolone) and Alproazolam, a sedative.

In summary, the best approach to take for any nontargeted screening or analysis will depend strongly on the questions being investigated. If there is a need to characterize every component in a complex sample, it may be most reasonable to do a suspect screen on all features by searching the acquired data against MS/MS libraries directly and comparing to a reasonable field control. However, if the more important question is really “what is different between these sample sets,” first utilizing statistical software tools to narrow down the feature list for investigation may be a more advantageous and informative workflow.

Acknowledgment

SCIEX thanks Renee Huang and the Santa Clara Valley Water District for collecting and preparing water samples for this study.

References

1. Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ru, M., Singer, H. P., Hollender, J. (2014) Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. 2097–2098.
2. A. Schreiber, N. Pace. *Identifying Unexpected Environmental Contaminants with High-Resolution, Accurate Mass LC–MS-MS*. LCGC Chromatography Online (2010).