COMPONENTIZATION FOLLOWING 3D-PEAK DETECTION IN THE UNIFI SCIENTIFIC INFORMATION SYSTEM

Jeff Goshawk, David Eatough, and Michelle Wood

The UNIFI® Scientific Information System from Waters contains a proprietary and powerful peak detection algorithm that transforms complex full scan data into components that can subsequently be used for targeted and non-targeted analysis. This white paper focuses on the analysis of the plant alkaloid scopolamine, using the Waters Forensic Toxicology Screening Application Solution with UNIFI,[1] and aims to provide a high-level explanation of the componentization process and its benefits for screening and discovery.

Waters

THE SCIENCE OF WHAT'S POSSIBLE.®

INTRODUCTION

Broad, comprehensive screening relies on the acquisition of a complete and unrestricted dataset, such that all data is available for both targeted and non-targeted processing. The utility of MS[E], a novel acquisition mode comprising the collection of full scan data under low and high-energy conditions, has been demonstrated in food and environmental,[2] and toxicological[3] testing. This type of data provides a wealth of information, for example the low-energy data provides the accurate mass of the precursor ion and the isotope cluster, whereas the high-energy data contains related fragment ions. Though this data is complex, UNIFI utilizes a proprietary algorithm to organize the data into a set of components.

PEAK DETECTION AND COMPONENTIZATION

Figure 2 shows an example of the full-scan continuous data acquired during a MS[E]-based toxicology screening analysis. A mixture of drug substances was analyzed which included the plant alkaloid scopolamine. The data can be considered as two mountain ranges; one for the low-energy data, and one for the high-energy data, respectively.

The 3D peak detection algorithm within UNIFI locates the summits of the peaks in a given mountain range*. The peak summits, or apexes, detected for the data shown in Figure 2 correspond to the dots overlaid on the data in Figure 3. Each dot represents a single ion characterized by a *m/z* value and a retention time (RT), together with a measure of intensity which is derived from the volume of the peak. Thus a unique *m/z*-RT pair is derived for each ion.

Figure 1. Datura stramonium, also known as Jimson weed or Devil's snare. A plant in the nightshade family from which the plant alkaloid scopolamine can be extracted.
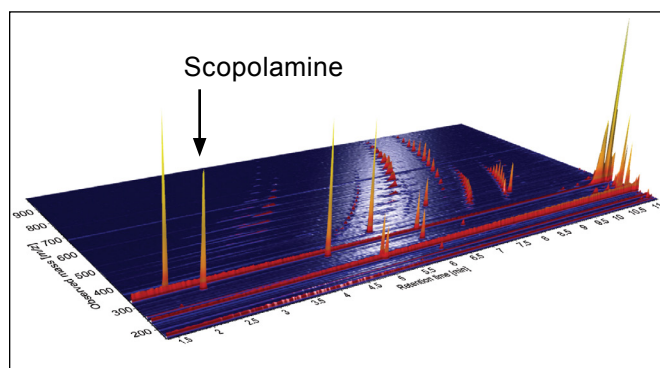

Figure 2. Analysis of a sample containing the plant alkaloid scopolamine: the full-scan continuous data is displayed in three dimensions.
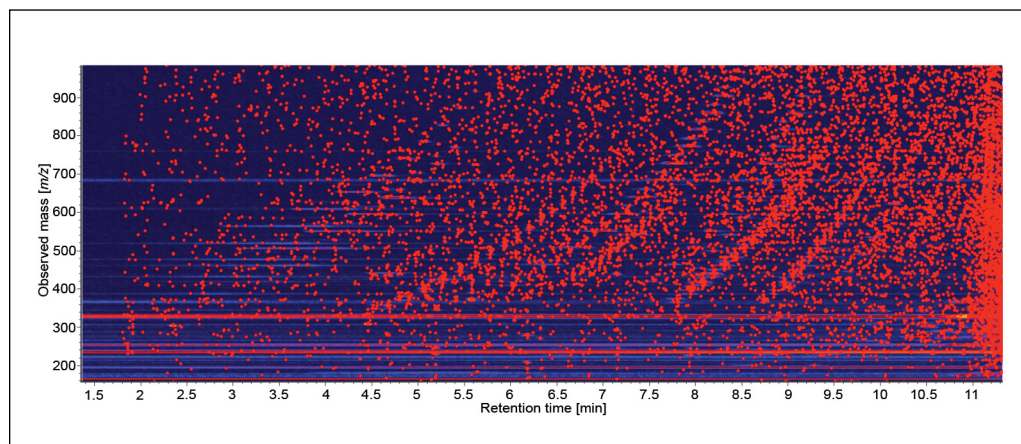

Figure 3. The peaks detected in the toxicology data shown in Figure 2.

*The intensity down to which the peaks are detected is controlled by parameters in the 3D Peak Detection section of the UNIFI analysis method.

Figure 4 shows the profile of a single peak in the data corresponding to the protonated ion of scopolamine and Figure 5 shows that a single ion is detected for that peak when using the 3D peak detection algorithm in UNIFI. This contrasts to considering the data scan-by-scan, which results in an ion being recorded many times across the elution profile of a chromatographic peak, as shown in Figure 6.

Following the detection of the individual ions in UNIFI they are organized into "candidate components". Every ion detected in the low-energy data can become a candidate with the exception of those ions that are deemed to be isotopes of other ions.[**] Associated with each candidate component is a low-energy spectrum, which contains ions with a peak apex at the same retention time as the candidate. For MS^E data, a high-energy spectrum is also associated with the candidate and contains high-energy ions with an apex at the same retention time.[†]
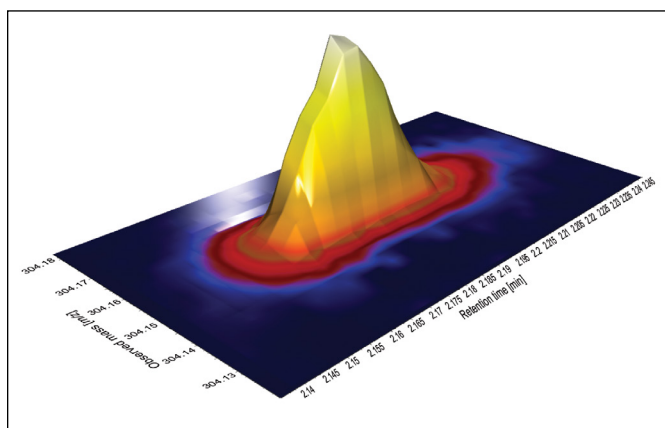

Figure 4. A single peak for the monoisotopic protonated ion of scopolamine.
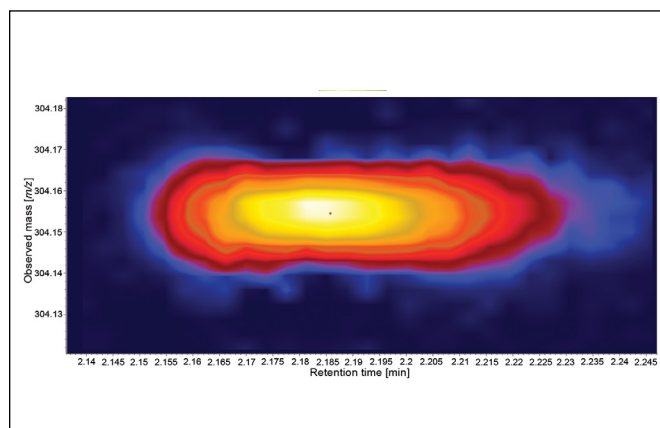

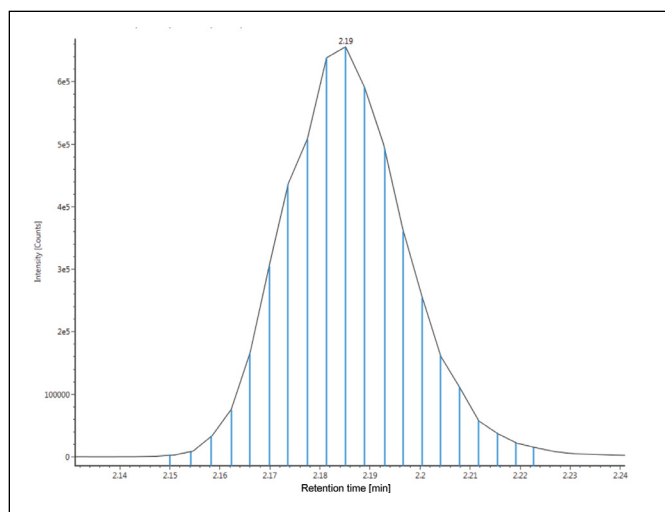Figure 5. A single ion detected for the peak for scopolamine shown in Figure 4.

Figure 6. An ion detected multiple times across the chromatographic peak for scopolamine (m/z 304.154). Each vertical line represents the measured ion for m/z 304.154 at the corresponding retention time.

**The number and charge of the isotopes is controlled by parameters in the 3D Isotope Clustering section of the UNIFI analysis method.

†The retention-time range over which low- and high-energy ions are aligned in spectra corresponding to a component is controlled by parameters in the 3D Isotope Clustering section of the UNIFI analysis method.

Since the spectra associated with each candidate component only contain ions that correspond to the apex of a peak, and have the same retention time as the candidate, they contain far fewer ions than the measured spectrum at that retention time. This is demonstrated in Figure 7 by comparing the measured continuum spectrum for scolopamine, and a centroid version of that spectrum, with the corresponding spectrum associated with the component for scolopamine. The simplification of the candidate spectrum is more apparent in Figure 8 which shows the lower *m/z* range for the three spectra.
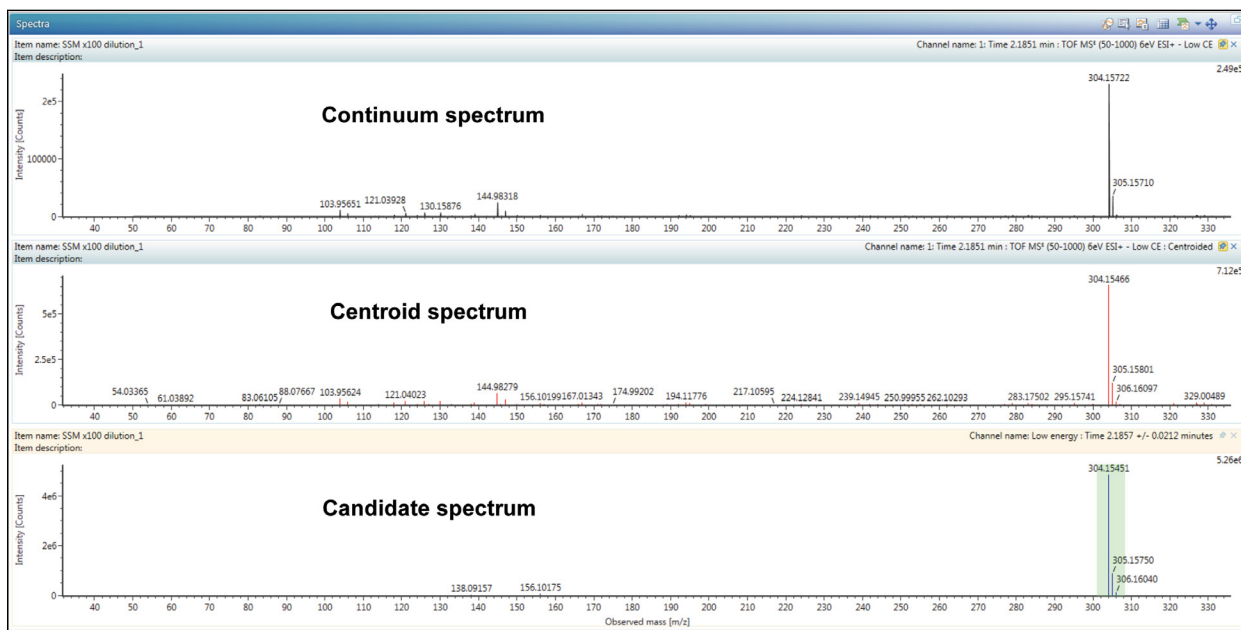

Figure 7. Comparison between the raw continuum spectrum, the centroid spectrum, and the candidate spectrum for scopolamine.
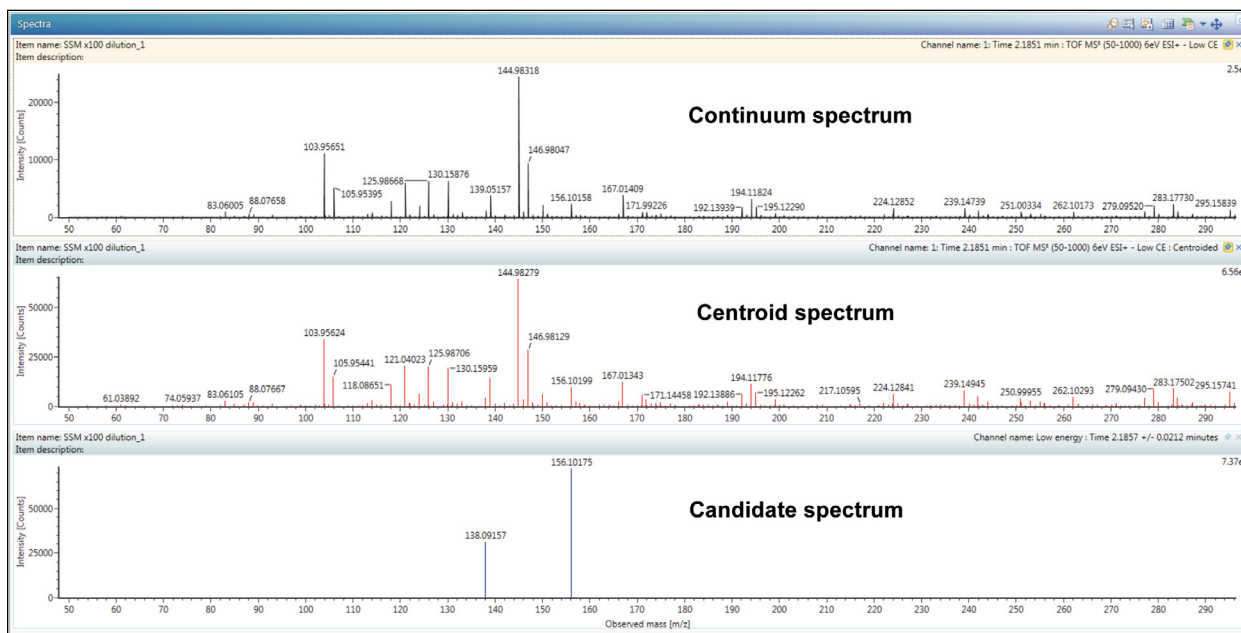

Figure 8. Comparison between the raw continuum spectrum, the centroid spectrum, and the candidate spectrum for scopolamine highlighting the spectral clean-up offered by componentization.

ASSOCIATING TARGETS TO COMPONENTS

A target is a substance sought in the measured data and may be defined by a number of properties. At the very basic level, a target can be the expected neutral mass of the substance or a chemical formula from which the neutral mass can be determined. Expected *m/z* values of adducts of the substance can subsequently be determined from the neutral mass. Additionally, if a standard of the substance has been acquired under conditions consistent with those of subsequent experiments, then supporting retention time and fragmentation data may also be known.

Since the componentization within UNIFI produces a list of candidate components, each of which has a unique *m/z*-RT pair, UNIFI can determine whether any targets match any of the candidate components to within *m/z*, and optionally, retention time tolerance. Further, if a target has associated fragments then, for MS^E data, the *m/z* values of the expected fragments can be compared to the *m/z* values of the ions in the high-energy data to determine whether they are within an acceptable tolerance and thereby support the observation of the target.

It would not be feasible to search for targets within the data in this way without creating components that have associated spectra only containing ions representative of peak apexes, since the number of false-positive assignments would be overwhelming.

EXTRACTED ION CHROMATOGRAMS VERSUS COMPONENTIZATION

A common method of looking for target substances in measured data is to simply extract a mass chromatogram using the exact mass of the appropriate adduct of the target and determine whether there is a peak at the expected retention time. However in some cases, this method of determining the presence or absence of a target can result in false positive observations since a peak on a chromatogram may not represent the *m/z* value of the corresponding peak in three dimensions.

This can be illustrated from the same set of data. Consider bufalin, another plant alkaloid; this substance has the chemical formula $C_{24}H_{34}O_4$ and therefore the *m/z* of the protonated species is 387.2530. When the accurate mass chromatogram corresponding to the protonated species of bufalin is extracted from the same sample (Figure 9) it is clear that there is a distinct peak on the chromatogram even though bufalin is not present in the sample. Therefore, if the data had been processed using an approach based on extracting ion chromatograms, the contaminant in the sample would have resulted in bufalin being reported as a false-positive.
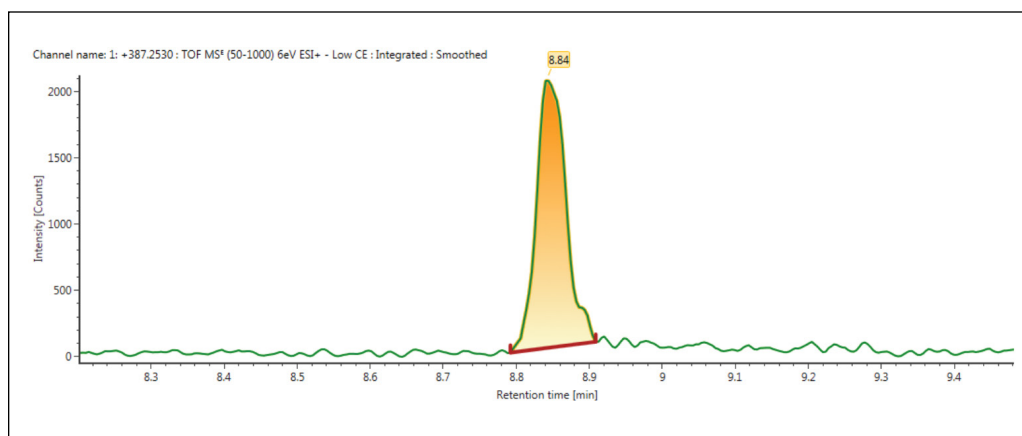

Figure 9. Extracted mass chromatogram corresponding to the protonated adduct of the plant alkaloid bufalin.

In contrast, the three-dimensional representation of the same peak is shown in Figure 10, together with a schematic representation of the accurate mass chromatogram shown in Figure 9. It is evident that the *m/z* value of the apex of the peak is 387.239, 36 ppm lower than the expected *m/z* of the protonated adduct of bufalin.

It should be noted, that for this example, the *m/z* tolerance would have to be greater than 36 ppm for bufalin to be reported as a false-positive observation when using componentization in UNIFI. In the event that the tolerance was set so generously, the mass error on the measurement would also be reported in UNIFI.
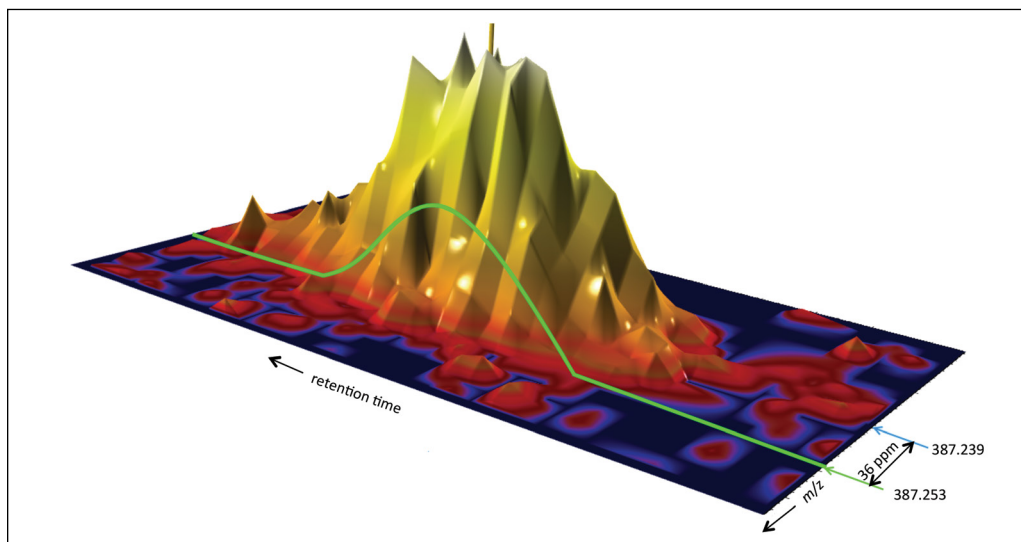

Figure 10. A 3D peak together with a representation of the extracted mass chromatogram for bufalin that deviates from the accurate m/z value of the 3D peak by 36 ppm.

COMPONENTIZATION AND *IN SILICO* FRAGMENTATION

One of the challenges faced by the modern toxicology laboratory is ensuring that analytical methods are kept up-to-date and relevant. It is particularly problematic nowadays owing to the constant proliferation of novel psychoactive substances (NPS).

The ability to screen for new drug substances by nominal mass (LC-MS/MS) instrumentation requires access to reference material, however certified material is not always available, and custom synthesis, while possible, can be extremely costly. In contrast, one of the key benefits of accurate mass instrumentation is an ability to tentatively screen without reference material, i.e. on the basis of expected elemental formula only.

Furthermore, in the absence of reference material, *in silico* fragmentation techniques may prove highly-valuable since the prediction of the *m/z* values of fragment ions from the structure of a substance can also be used as additional support to any putative identification; it might also help to justify purchase of material.

The fragment match tool within UNIFI takes the structure of a target substance, in the form of a mol file, and performs systematic bond disconnections to produce substructures. The *m/z* values of the theoretical substructures can then be compared to the *m/z* values of ions in the high-energy data to determine whether the ions are fragments of the substance.

However, for this to be effective it is crucial that the apexes of any fragment ions have the same retention time as the apex of the precursor ion, as is the case with spectra associated to components.

If interferences are not removed from the high-energy data, the only way to determine whether the fragment ions maximize at the same retention time are the precursor, is to extract a mass chromatogram for each high-energy ion with a *m/z* value corresponding to that of a proposed substructure, and determine whether there is a peak at the correct retention time. Not only would this be a computationally expensive approach but, as previously highlighted, false-positive results can be generated by extracting mass chromatograms and performing 2D peak detection.

COMPONENTIZATION AND DISCOVERY

The componentization in UNIFI facilitates the selection of candidate components which may represent unexpected substances within a sample. An accurate *m/z* value is determined for every candidate component together with a retention time, an intensity measurement, and associated spectra from which interfering ions have been removed. There are a number of ways in UNIFI in which this information can be used to select potentially important, unidentified components for additional investigation.

Components can be interrogated to determine whether they contain specific fragment ions or show specific neutral losses. This functionality relies on the fact that the spectra associated with each component contain only those ions that maximize at the retention time of the component. If interfering ions are not removed from the spectra, the likelihood of incorrectly designating that a peak contains a specific fragment ion or exhibits a specific neutral loss is increased.

It is also possible to compare the low-energy spectrum, associated with each component, to specific theoretical isotope distributions, particularly those containing the distinctive patterns for chlorine and bromine. More confidence can be applied to the results from such a comparison if interfering ions, which may lead to false-positive assignment, have been removed from the measured spectrum.

If the most intense components within a given retention time range are deemed to be the most appropriate for further investigation, then a list of such components can be generated by setting up an appropriate filter within UNIFI.

Upon selecting candidate components that require further investigation, elemental compositions corresponding to the neutral mass of the precursor ion can be determined. The theoretical isotope pattern each elemental composition can be compared to the peaks in the low-energy spectrum associated with the component. Since interferences have been removed from the spectrum, the comparison between the theoretical and measured data will be more accurate. Promising elemental compositions can be submitted to web searches and structures of prospective substances for the candidate components can be retrieved. The structures can undergo *in silico* fragmentation to determine whether the *m/z* of substructures match the *m/z* of ions in the high-energy spectrum associated with each component. This process would also lead to false-positive assignments if the interfering ions had not been removed from the high-energy spectra.

CONCLUSION

The analysis of a drug mixture, using the Forensic Toxicology Screening Application Solution with UNIFI presented an opportunity to illustrate the benefits of UNIFI's 3D peak detection; complex data is reduced down to components that can subsequently be used for targeted and non-targeted analysis. For MS^E data, the components have associated low- and high-energy spectra from which interfering ions have been eliminated. Such clean spectra permit both the confirmation and assignment of fragment ions and neutral losses, and allow for accurate comparison of measured data to theoretical isotopes patterns.

These aspects reduce the burden of data interpretation for the analyst, reduce potential false-positive assignments, and allow results to be presented clearly and concisely.

References

1. Forensic Toxicology Screening Application Solution with UNIFI. Waters Brochure (p/n 720004830EN).

2. López MG, Fussell RJ, Stead SL, Roberts D, McCullagh M, and Rao R. Evaluation and validation of an accurate mass screening method for the analysis of pesticides in fruits and vegetables using liquid chromatography–quadrupole-time of flight–mass spectrometry with automated detection. *J Chromatogr A.* 2014 Dec;1373:40–50.

3. Rosano TG, Wood M, Ihenetu K, and Swift TA. Drug screening in medical examiner casework by high resolution mass spectrometry (UPLC-MS[E]-TOF). *J Anal Toxicol.* 2013 Oct;37:580–93.

Waters

THE SCIENCE OF WHAT'S POSSIBLE.®