Smart Elemental Composition by Automatically Limiting the Element Set

David A. Wright and Ralf Tautenhahn,
Thermo Fisher Scientific, San Jose, CA, USA

Overview

Purpose: This approach to elemental composition exploits the wealth of isotope information present in very high resolution data to limit the elements, and their frequency of occurrence, in order to improve the elemental composition result.

Methods: As part of the component detection process, each scan is processed to show only ions related to a particular component, and then examined for isotope patterns correspond to exotic (any element other than Carbon) isotope ions. The processing removes 99% of the ions in the spectrum, greatly reducing the probability that any particular ion is present by chance. The presence, and intensity relative to the 13C isotopologue, are used to limit the element set.

Results: This method can automatically detect most of the common elements relevant to small-molecule metabolism and EFS, including N, O, Si, S, Ca, Cl, and Br, at 120K resolution.

Introduction

It's well known that very high resolution scans show isotope fine structure and that structure can help in elemental composition by providing smarter limits for the element set. A new component detection algorithm developed at Thermo Fisher Scientific, in addition to being faster than the previous algorithm, detects and identified hetero atoms in spectra without user-intervention. No complicated calculations are needed for this result – a full, computationally complex, least-squares fit to all possible formulae is reserved for the final step in processing, after the uninteresting signals have been eliminated. Furthermore, this approach is untargeted – the algorithm does not look for specific elements, but rather looks for ions that are related and then tries to match them to a specific element. Ions that are related, but not identified by this approach are included in the isotope cluster for elemental composition calculation. It has been our experience that the combination of sub 5 ppm mass accuracy and the ability to limit the element set by this procedure results in the correct formula often being the highest scoring possible result.

Methods

Mass Spectrometry

Data collected on various models of Thermo Scientific™ Orbitrap™ based Instruments.

Data Analysis

Raw files were processed using alpha versions of an in-house Component Detection algorithm. As part of the feature detection phase, this algorithm analyzes the isotope patterns for the presence of "exotic" isotopes. The algorithm is able to accomplish this since it looks at every ion in every scan, and by pattern analysis can determine if a particular ion belongs to an isotope cluster, and which element it is from. Although it is only able to assign a probability to this identification, testing has shown that ion assignments above a 90% confidence level are accurate.

For a given resolution, mass and intensity, the algorithm knows whether or not a particular exotic should be detected, and at what abundance relative to Carbon. These limits are then used to define the element set and abundances for the final calculation of the probable formulae and their likelihood scores.

Results

A sample spiked with the 500 pesticide standard was run on a Thermo Scientific™ Orbitrap Velos Pro™ ion trap mass spectrometer at 120K resolution. After component detection the chromatographic peak from the pesticide Carboxin was selected.

The table shows that the exotic isotopes O18, S33 and S34 were detected for this component. Restricting the element set to CHNOSPF (F and P were included since they have no isotopes and cannot be excluded by our method.), and restricting the number of Carbons to at least 5, and Sulfurs to at least 1, gives 12 matches within 10 ppm.

Furthermore, the Nitrogen Rule requires an odd number of Nitrogens for this mass, and the failure to detect the A1 N15 ion limits when the slightly more intense S33 ion is seen means that is practically impossible that the correct formula has greater than 1 Nitrogen. This makes the correct match #2, since 5 Nitrogens is not possible from the A1 data.

FIGURE 1. Mixture of 500 pesticides run on an Orbitrap Velos Pro MS. Carboxin (detected m/z 236.0734) component detected. Top cell is TIC chromatogram, middle is a portion of the component list, and the bottom cell is the extracted Ion Chromatogram (XIC) for a single component
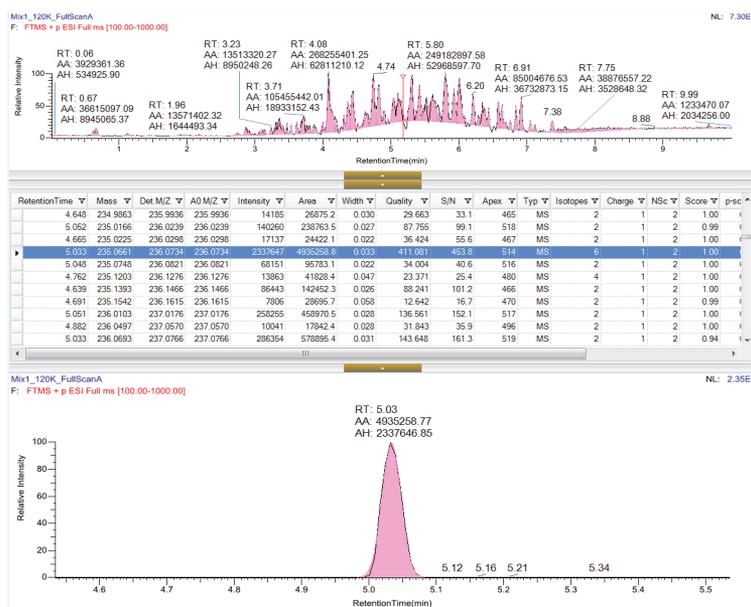
TABLE 1. The isotopes found by the algorithm for Carboxin

Number	RT	Monoisotopic M/Z	Detected M/Z	Intensity	Formula	Type	Label	Charge	Apex
6	5.033	236.0734	236.0734	2337647	C12 H13 N O2 S	A0	A0	1	514
6	5.033	236.0734	237.0766	298483		isotope	A1	1	
6	5.033	236.0734	238.0798	17902		isotope	A2	1	
6	5.033	236.0734	239.0829	2340		isotope	A3	1	
6	5.028	236.0734	238.077	4415		isotope	A2 O18	1	
6	5.033	236.0734	238.0689	76310		isotope	A2 S34	1	
6	5.033	236.0734	239.0723	8208		isotope	A3 S34	1	
6	5.047	236.0734	237.0725	6106		isotope	A1 S33	1	

FIGURE 1a. Simulation of A1 ions. Top cell – simulated pattern. Bottom cell – (top to bottom) incorrect formula, correct formula (1 Nitrogen), experimental data
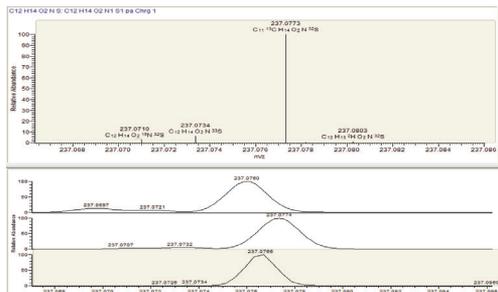
TABLE 1a. Formula matches to observed mass, with element set CHNOSPF

Idx	Formula	RDB	Delta ppm
1	C6 H15 O N5 PS	2.5	2.059
2	C12 H14 O2 NS	6.5	-2.313
3	C5 H11 O N7 FS	3.5	4.221
4	C5 H14 N7 S2	2.5	-5.216
5	C9 H15 O3 N FS	2.5	-7.153
6	C10 H13 N F2S	3.5	8.043
7	C8 H10 N7S	7.5	9.062

Elements in use

| Isotope	Min	Max	DB eq.	Mass
14 N	0	10	0.5	14.003
16 O	0	15	0.0	15.995
12 C	5	30	1.0	12.000
1 H	5	60	-0.5	1.008
32 S	1	2	0.0	31.972
19 F	0	3	-0.5	18.998
31 P	0	2	0.5	30.974

Detection of O18 and N15

This data is from a sample that has been spiked with 17 Amino Acids, and run on a Thermo Scientific™ Q Exactive™ Plus MS at 120K resolution in negative mode. The algorithm detects the 15N ion in the A1 spectrum of this mass 173 ion, which is Arginine.

TABLE 2. The isotopologues near the mass for Arginine

| Number	RT	Monoisotopic M/Z	DetectedM/Z	Intensity	Formula	Type	Label	Charge	Apex
5	0.797	172.958	173.9623	10038		isotope	A1	1	
6	0.765	173.0068	173.0068	211336 ???		A0	A0	1	172
6	0.765	173.0068	174.0103	15943		isotope	A1	1	
7	0.772	173.1044	173.1044	10485511	C6H13N4O2	A0	A0	1	172
7	0.772	173.1044	174.1078	689127		isotope	A1	1	
7	0.772	173.1044	175.1111	21582		isotope	A2	1	
7	0.772	173.1044	175.1087	43160		isotope	A2 O18	1	
7	0.774	173.1044	174.1015	137602		isotope	A1 N15	1	
12	0.791	174.0271	174.0271	54941 ???		A0	A0	1	176
14	0.822	173.8427	174.8397	322012 ???		A0	A0	2	183
15	0.835	174.8272	174.9644	132819 ???		A0	A0	7	186
15	0.835	174.8272	174.8272	507050		isotope	A0	7	
15	0.835	174.8272	174.8272	507050		isotope	A0	7	
15	0.835	174.8272	174.8272	507050		isotope	A0	7	
16	0.671	174.983	174.983	23844 ???		A0	A0	4	150

FIGURE 2a. The A1 ions, with stick spectrum on top, simulated spectrum in the middle, and experimental data at 120K resolution at the bottom, with 15N resolved.

FIGURE 2b. The A2 ions as in Fig. 2a, with 18O resolved.
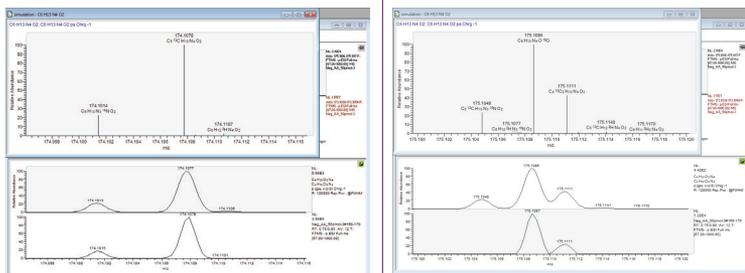
In this case, the correct formula is at the top of the list of possibilities returned by Elemental Composition (not shown). The mass accuracy is 0.4 ppm. If there were doubt about the formula assignment, the A1 15N intensity is only compatible with a C/N ratio of 3/2., as shown in Fig 2c. The other two possibilities (albeit far removed in mass value) have ratios of 8/1 and 4/7, and clearly don't match the data. This is shown in Fig. 2c below.

FIGURE 2c. Simulations of the 3 possible formulae for this mass 174 A1 ion. Experimental data at top, below are the 3 formulae shown in the combo box at left and the spectral view to the right. The correct formula is second trace from the top.
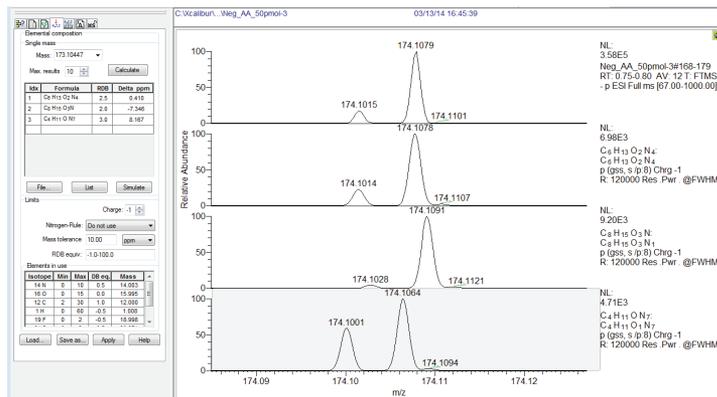
A Tale of 3 Similar Masses

This example dataset has 500 pesticides spiked into an onion matrix, and run on a Thermo Scientific™ Q Exactive™ Plus at 120K resolution. Sorting by mass, and picking a region, one can find components of similar mass but different found exotic isotopes. The first cluster (353.0778) in this region shows Chlorine isotopes but no Sulfur, the second (353.2214) contains neither, and the third (354.1314) shows Sulfur.

TABLE 3. The isotopologues near these masses of interest.

Number	RT	Monoisotopic M/Z	DetectedM/Z	Intensity	Formula	Type	Label	Charge	Apex
4538	5.015	353.0778	353.0778	1994074	C17H18ClO6	A0	A0	1	512
4538	5.015	353.0778	354.081	366965		isotope	A1	1	
4538	5.015	353.0778	355.0829	23560		isotope	A2	1	
4538	5.015	353.0778	355.0746	640723		isotope	A2 Cl37	1	
4538	5.015	353.0778	356.0777	93238		isotope	A3 Cl37	1	
4538	5.015	353.0778	357.0815	7110		isotope	A4 Cl37	1	
4538	5.015	353.0778	358.0832	1722		isotope	A5 Cl37	1	
4547	6.071	353.2214	353.2214	763985 ???		A0	A0	1	649
4547	6.071	353.2214	354.2248	187443		isotope	A1	1	
4547	6.071	353.2214	355.2278	16666		isotope	A2	1	
4547	6.071	353.2214	356.231	1560		isotope	A3	1	
4558	6.579	354.1314	354.1314	10893113	C14H29NO3PS2	A0	A0	1	712
4558	6.579	354.1314	355.1339	1921815		isotope	A1	1	
4558	6.579	354.1314	356.1362	97585		isotope	A2	1	
4558	6.579	354.1314	357.1396	3966		isotope	A3	1	
4558	6.579	354.1314	356.1268	946809		isotope	A2 S34	1	
4558	6.579	354.1314	357.1294	121981		isotope	A3 S34	1	
4558	6.579	354.1314	358.1324	4361		isotope	A4 S34	1	

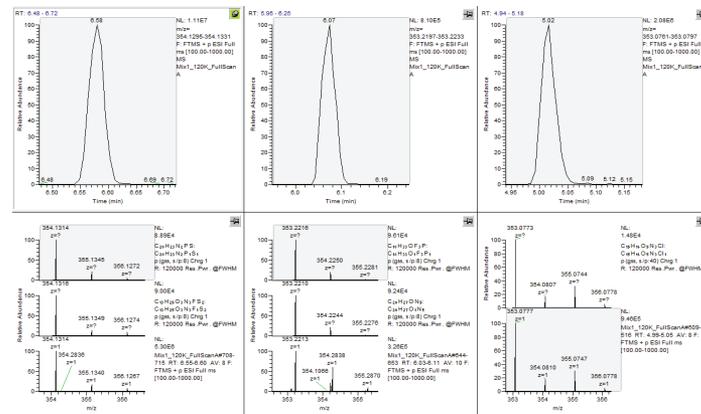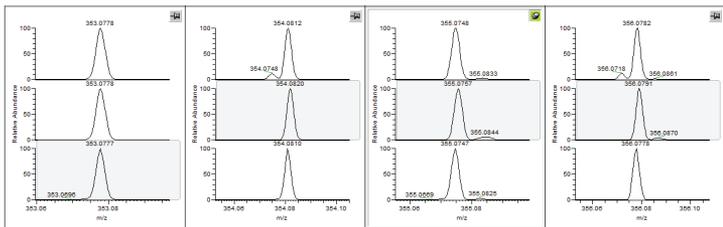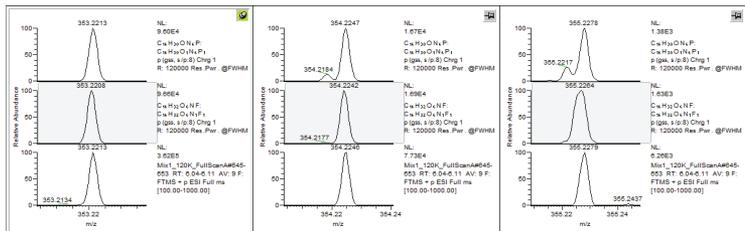**FIGURE 3. Three Chromatographic peaks near 350 M/Z.**

FIGURE 3a. Isotopes for the 353.0778 peak

In Fig. 3a above, the data is the bottom row of A0, A1, A2, and A3 isotopes, the correct formula (C17 H18 O6 Cl) in the middle row, and a formula (one of many), that are ~~closer in mass but contain Nitrogen~~ in the top row. Since the ~~correct~~ is ~~visible~~

FIGURE 3b. Isotopes for the 353.2214 peak

In Fig. 3b above, the observed data is the bottom row of A0, A1, and A2 isotopes. While the correct formula is unknown, in the middle row is the best match that has a low enough count of 15N to be consistent with the observed lack of a 15N A1 isotope. That formula has a mass error of 1.3 ppm. All of the better matches in mass value have several Nitrogen in the formula, and the top row is a the spectrum of one of those. While the A2 is weak enough that the 15N peak may not be visible, it should be for the A1 if the top formula is correct.

FIGURE 3c. Isotopes for the 354.1314 peak
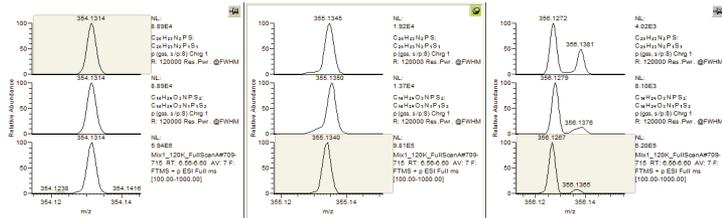
In Fig. 3c above, the observed data is the bottom row of A0, A1, and A2 isotopes, the correct formula (C14 H28 N O3 P S2) in the middle row, and a formula (one of many), that are closer in mass but contain more Nitrogen, is in the top row. Our algorithm did not detect an A1 15N isotopologue and computed the C/N ratio at 10/1 or greater. With this constraint on the element set, the expected formula of Piperophos (at -1.9 ppm error) is fourth on the list of possible formulae, computed by elemental composition.

Conclusion

- Use of very high resolution spectra is an important tool to constrain the element set and the relative abundances in the determination of the correct chemical formula.

- The automatic determination of elements which must (or must not) be used in elemental composition is always preferable to manual inspection.

- This method automatically provides an element set, and limits to abundances for formula calculation, on a per-feature basis.

- This method has been shown to work best within the limits of resolution and M/Z shown in the table below. As the maximum useable M/Z is approached, the accuracy of the limits value decreases until it becomes a 'present/not present' detector.

resolution	Isotope detected	maximum M/Z
120,000		
	15N	400
	18O	400
	41K	700
	34S	800

Thermo
SCIENTIFIC
A Thermo Fisher Scientific Brand