Assigning Adduct and Charge States to High-resolution Accurate-mass Mass Spectral Data Using Frequency of Assignment in Multiple Difference Networks

Thomas D. McClure, Matthew D. Kump, Michael Athanas
Thermo Fisher Scientific, San Jose, California

Thermo SCIENTIFIC

Overview

Purpose: Using a graph-theory based approach to determine adduct and neutral loss species within a mass spectrum as a part of a small molecule component detection workflow.

Methods: A graph theory based difference networks are used for determining adduct and charge states for component detection analysis of mass spectrometry data.

Results: This work shows that difference networks provide accurate results when applied to complicated collections of adduct and charge states in a mass spectrum**.**

Introduction

Liquid or gas chromatography coupled with mass spectrometry, has been demonstrated to be a powerful tool for characterizing small molecules in biological samples. Often the goal is to understand differences in the types and amounts of these small molecules in metabolomic studies, metabolism studies, and virtually any approach involving a complex matrix where untargeted profile information is desired. The data sets from these experiments are usually large and complex. Such complexity results, in no small part, from more than one signal for each compound detected. These multiple signals per compound come from the formation of multiple charge states, in-source neutral-loss fragmentations, chemical adducts, and the formation of gas-phase polymers. In addition, each of these species also produces a number of isotope signals.

In this presentation, we will demonstrate the utility of applying a graph-theory based algorithm for reducing complexity in an LC-MS experiment. Graph-theory mathematics is not a tool in the analysis of MS data. It has been used for analyzing mass spectrometric data in a number of applications including: *de novo* peptide sequencing[1,2], isotope assignment, protein identification and quantification. We will use this algorithm to properly assign and group signals arising from the presence of multiple signals from chemical adducts, in-source neutral-loss fragmentations and adduct related multiple charge states for individual compounds in a mixture of compounds.

Methods

Synthetic Data Creation for Algorithm Verification

An LC-MS data set was created containing thirty compounds each with three adducts ([M+H], [M+Na], and [M+NH4]) as well as the isotopes for each of these species. A mass sorted peak list was generated from the LC-MS data file described above using software that produces extracted ion current chromatograms which are then evaluated using parameter-less peak detection.

Amino Acid Mixture Sample Preparation

Commercially available standard mixture was obtained using a Thermo Scientific™ Pierce™ Amino Acid Standard H, P/N 20088. The concentration of the amino acids in this mixture was 2.5 µmol/ml, except cysteine at 1.25 µmol/ ml. A diluted stock solution was prepared using 100 µl of Amino Acid Standard H + 900 µl of water. The final sample was then prepared by diluting the stock in a dilution series to a final dilution of 1:1000 with HPLC-grade water. 2 µl was directly injected on to the HPLC column.

HPLC Conditions

A Thermo Scientific™ Dionex™ UltiMate™ 3000 RSLC system was used with a Thermo Scientific™ Hypersil GOLD™ HPLC column (150 × 2.1, 1.9 um, P/N 25002-152130). The HPLC solvents were: A – 0.1% formic acid in water; B – 0.1% formic acid in methanol. The elution gradient was: 0.5% B to 55% B in 5.5 min, 50% B to 98% B in 0.5 min, hold 98% B for 6 min. The flow rate was 450 µl/min and the column was heated to 55° C.

MS Conditions

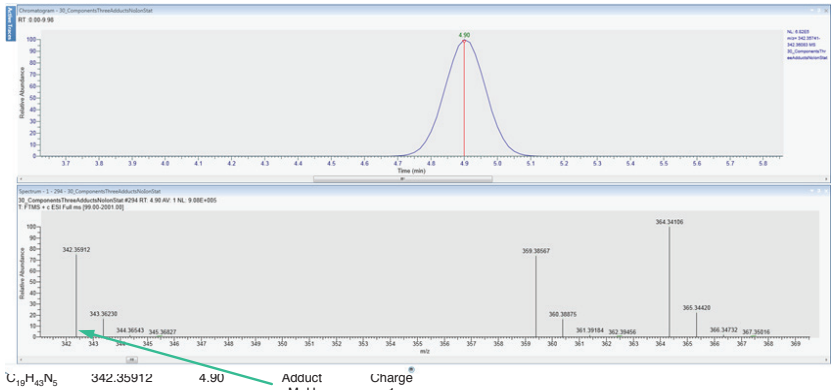
A Thermo Scientific™ Q Exactive™ mass spectrometer with a Thermo Scientific™ HESI-II source was used with the following gas settings: sheath was 45 and the aux gas was set to 8. The spray voltage was 3.8 kV, and the capillary temperature was 320° C. The HESI Heater temperature was 350° C. The mass analyzer had the following settings: positive polarity, full MS: 67–1000 AMU, AGC was set to 3+E6, the resolution was 70,000, and the maximum ion injection time was 100 ms.

Results

Algorithm Verification Using Synthetic Data

We verified the operation and accuracy of the assignments by the difference network using a synthetically generated raw file with 30 components, each having three adducts (MH+, M+Na+, and M+NH4+). Each adduct had a minimum of three, and usually four, isotopes. The two component examples are shown below.

Synthetic Example 1



$C_{19}H_{43}N_5$	342.35912	4.90	Adduct	Charge
			M+H	1

Isotope	MZ	Intensity	Peak Area
A0	342.35912	669135	5867777
A1	343.36230	147786	1295968
A2	344.36543	15569	135995
A3	345.36827	951	7155

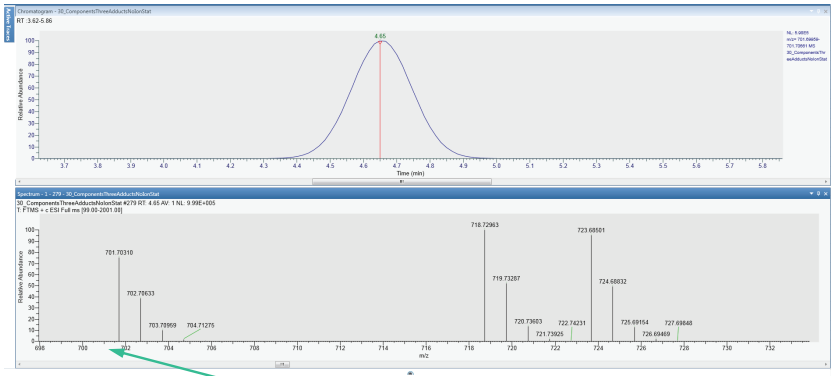
M+Na 1

Isotope	MZ	Intensity	Peak Area
A0	364.34106	891390	7816780
A1	365.34420	197788	1734444
A2	366.34732	20891	181364
A3	367.35016	1353	10581

M+NH4 1

Isotope	MZ	Intensity	Peak Area
A0	359.38567	656801	5759615
A1	360.38875	147162	1290490
A2	361.39184	15687	137002
A3	362.39456	1034	7830

Synthetic Example 2



$C_{45}H_{88}N_4O$	701.70310	4.65	Adduct	Charge
			M+H	1

Isotope	MZ	Intensity	Peak Area
A0	701.70310	586712	7854121
A1	702.70633	302768	4052986
A2	703.70959	77602	1038850
A3	704.71275	13145	176007
A4	705.71652	1271	17021

M+Na 1

Isotope	MZ	Intensity	Peak Area
A0	723.68501	743036	9947609
A1	724.68832	383766	5134197
A2	725.69154	98287	1316105
A3	726.69469	16661	223059
A4	727.69848	1611	21570

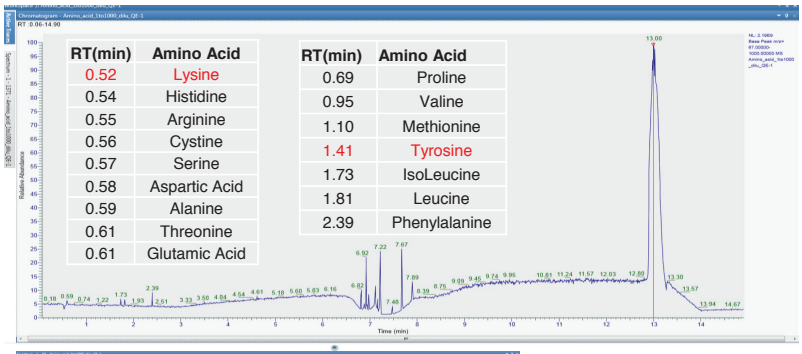
M+NH4 1

Isotope	MZ	Intensity	Peak Area
A0	718.72963	778223	10421120
A1	719.73287	404727	5414875
A2	720.73603	104345	1396982
A3	721.73925	17783	238128
A4	722.74231	1961	24241

In both cases, the adduct and charge states are correctly assigned.

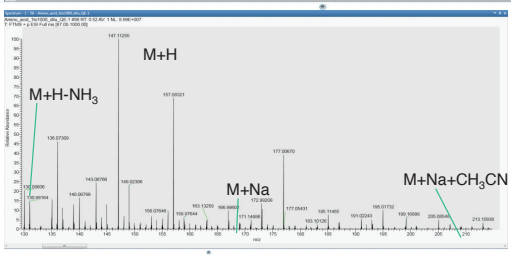
Analysis of a Dilute Amino Acid Mixture

The amino acid mixture previously described was analyzed using the difference network algorithm. The following chromatogram is labeled with the amino acids in the mixture and their elution times.



RT(min)	Amino Acid
0.52	Lysine
0.54	Histidine
0.55	Arginine
0.56	Cystine
0.57	Serine
0.58	Aspartic Acid
0.59	Alanine
0.61	Threonine
0.61	Glutamic Acid

RT(min)	Amino Acid
0.69	Proline
0.95	Valine
1.10	Methionine
1.41	Tyrosine
1.73	IsoLeucine
1.81	Leucine
2.39	Phenylalanine



Lysine

m/z	Apex RT
147.11255	0.46

Adduct	Charge
M+H	1

Isotope	MZ	Intensity	Peak Area
A0	147.11255	54321440	66288519
A1	148.11589	3604857	4250354
A1	148.10962	192584	316075
A2	149.11687	457894	2173745

M+H-NH3	1

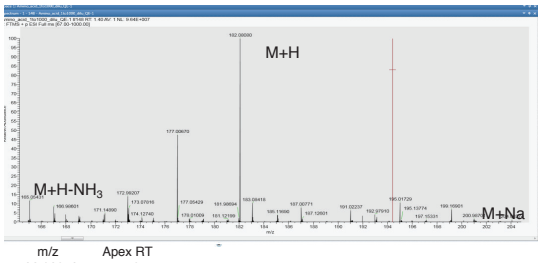
Isotope	MZ	Intensity	Peak Area
A0	130.08606	10109494	11013528
A1	131.08937	812226	559401

M+Na+CH3CN	1

Isotope	MZ	Intensity	Peak Area
A0	210.12015	195011	143738

M+Na	1

Isotope	MZ	Intensity	Peak Area
A0	169.09444	2110288	2464782
A1	170.09763	114523	72849



Tyrosine

m/z	Apex RT
182.08078	1.40

Adduct	Charge
M+H	1

Isotope	MZ	Intensity	Peak Area
A0	182.08078	67884942	130861879
A1	183.08418	7242606	11166390
A2	184.08751	306655	349548
A2	184.08508	370365	627524

M+H-NH3	1

Isotope	MZ	Intensity	Peak Area
A0	165.05429	7440865	12655904
A1	166.05765	753306	1231796

M+Na	1

Isotope	MZ	Intensity	Peak Area
A0	204.06277	766400	1321692

Analysis of a Complicated Mixture of Adducts

In our final example, an unpublished study, we analyzed a sample that contained 13 adducts in one mass spectrum. Of these 13, we detected the following 10 species: M+CaCOOH (z=1), M+MgCOOH (z=1), M+Fe-H (z=1), M+Na (z=1), M+H (z=1), M+H-2(H2O) (z=1), M+Ca+2(CH3CN) (z=2), M+Ca+CH3CN (z=2), M+Mg+CH3CN (z=2), M+Ca+H2O (z=2).

Notice there are both single and double charge states in this series of adducts.

Discussion

Adduct and Neutral-Loss Assignment Algorithm

We use the following formula to identify the important mass carrying components that are present in each species being detected by the mass spectrometer.

$$\Delta\ m/z\ =\ (n_1M_1 + M_{a1} + M_{cc1})\ /\ Z_1\ -\ (n_2M_2 + M_{a2} + M_{cc2})\ /\ Z_2 \qquad (1)$$

Where:

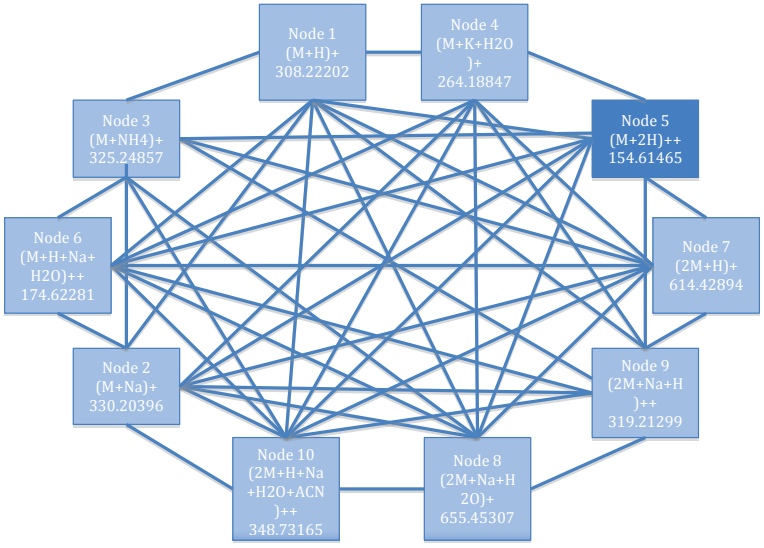
Δ ***m/z*** is the difference in mass-to-charge between two different species of the same molecule, **n** is the gas-phase cluster or polymeric number for the base molecule, **M** is the parent neutral molecule mass, **Ma** is the total mass that is contributed to the species by the adduct or adducts or neutral-loss or losses, **Mcc** is the total mass that is contributed to the species by the charge carrier or carriers, and **Z** is the total charge of the detected species.

1. Establishing a List of Known Adducted Species

Using the terms in equation 1, we construct a table of candidate "modifying" species with combinations of charge carriers, neutral adducts, and neutral losses that can combine with M to form additional signals. These candidate "modifying" species are used to generate a combinatorial table of mass differences.

2. Generating the Nodes for the Graph

Nodes are generated by matching differences in the list of candidate species with mass differences determined from monoisotopic *m/z* values from the mass spectrum being analyzed. The matching *m/z* values from the spectrum are added as nodes to the graph and an edge is drawn between the two nodes. The result is a large number of nodes and edges for the mass spectrum. This is illustrated below using hypothetical data.



3. Edge Trimming and Assignment of Species

Because of the large number of nodes and edges, ambiguities in assignment can arise. Nodes with multiple assignments are then ordered by frequency of assignment for M+H (positive ion mode) or M–H (negative ion mode). The algorithm also allows for weighting factors to be applied to predicted species should *a priori* information provide appropriate insight. The weighting factors can be node specific, which causes edges to move up or down on the ordered list.

4. Assignment of Species

The assignments of the charge carrier(s) and neutral adduct(s) or neutral loss(es) are then made according to final position in the ordered list. Should ambiguities still arise, all possibilities are reported.

5. Overview

The current approach is able to identify difference charge states as well as numerous uncommon adducts in a complicated mixture all within the same mass spectrum. This approach is also able to identify species resulting from in-source neutral losses, such as loss of water or ammonia, as shown by the amino acid samples.

6. Example Results

Using the synthetic data we were able to verify the operation of the algorithm and accurately assign both the neutral adducts and charged species for a number of combinations.

The amino acid data sets provided additional complexity showing that neutral losses, adducts, and charge carrying species are accurately assigned in the presences of other potentially interfering signals. The assignment of loss of ammonia from both Lysine and Tyrosine is consistent with the accurate mass determined elemental composition.

The assignment of the acetonitrile and sodium to lysine is questionable, given that the mobile phase used methanol and not acetonitrile and also the strength of the signal is comparatively quite low. In this case, improved accuracy could be achieved by using additional information such as the mobile phase composition to restrict the list of predicted adducts.

Uncommon adducts such as those containing calcium, magnesium, and iron were detected and labeled as shown in the final example. Some of the calcium and magnesium adducts also included neutral solvent species and were doubly-charged.

Conclusion

- Using the described graph-theory based algorithm to assign neutral adducts, in-source neutral losses, and charge species to mass spectral data produces useful results.
- Using the method described here complex mixtures of adducts with varied charge states and in-source neutral losses are detected and properly assigned.
- The complexity of the mass spectral information is reduced as a consequence of identification of the aforementioned species, which provides the analyst the capability to group compound related signals.
- The accuracy of this algorithm is further enhanced by incorporation of information such as mobile phase composition and knowledge of in-source fragmentation. Preference can be given to species known to occur through the use of weighting factors for the predicted adducts, charge carriers, and neutral loss species.

References

1. Taylor, J. A.; Johnson, R. S. *Rapid Commun. Mass Spectrom.* **1997,** *11* (9), 1067–1075.
2. Clauser, K. R.; Baker, P.; Burlingame, A. L. *Anal. Chem.* **1999,** *71* (14), 2871–2882.
3. Cox, J.; Mann, M. *Nat. Biotechnol.* **2008,** *26* (12), 1367–1372.

www.thermofisher.com

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Canada +1 800 530 8447
China 800 810 5118 (free call domestic)
400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 9 3291 0200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 45 453 9100
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Spain +34 914 845 965
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand

PN-64093-EN-0716S