# Examples of Extracted Features

Supplement to:

Hegeman AD, Schulte CF, Cui Q, Lewis IA, Huttlin EL, Eghbalnia H, Harms AC, Ulrich EL, Markley JL, and Sussman MR (2007) Stable Isotope Assisted Assignment of Elemental Compositions for Metabolomics, *Anal. Chem.*

The purpose of this supplement is to provide samples of raw ESI-TOF data used for formulae assignment in the parent manuscript. As discussed therein, one of the most challenging aspects of LC-MS based metabolomic data processing automation concerns the assignment of features to metabolites that accurately capture the exact mass measurement, retention information and integrated signal intensity. The application that we used for this extraction is MassHunter (Agilent Technologies, Santa Clara, CA), which is currently under development and in its current version does not allow user input for optimizing chromatographic peak fitting or deselecting deisotoping routines. We recognize that this is sub optimal for analysis of isotopically labeled samples and provide the following set of examples to show how the MassHunter application assigned features in each of our labeling regimes [(I) ^{12}C & ^{14}N (natural abundance); (II) ^{12}C & ^{15}N; (III) ^{13}C & ^{14}N; or (IV) ^{13}C & ^{15}N].

Example #	Formulae	Which Strategy IDed?		14N12C			14N13C			15N12C			15N13C			SUMMARY data			
		database	calculated	feature ID#	RT (min.)	Query Mass	feature ID#	RT (min.)	Query Mass	feature ID#	RT (min.)	Query Mass	feature ID#	RT (min.)	Query Mass	ave RT	STDEV RT	Ave Mass error (ppm)	Ave. Absolute Value of Mass error
(i)	C9H15N	no	yes	86	38.41	137.1207	53	38.38	146.15	25	38.325	138.1179	42	38.43	147.1471	38.385	0.046	0.43	15.98
(ii)	C27H30O15	yes	yes	10	31.8	594.1586	41	31.75	621.2487	50	31.744	594.1588	58	31.85	621.249	31.786	0.051	0.86	5.13
(iii)	C14H20N4O2	yes	yes	67	14.81	276.159	43	15.43	290.2048	157	15.994	280.1466	127	16.3	294.1937	15.635	0.657	0.18	7.86
(iv)	C14H18N2O2	yes	yes	221	23.32	246.1371	378	23.32	260.1827	217	23.279	248.1312	338	23.41	262.1776	23.330	0.053	-0.15	11.45
(v)	C7H19N3	yes	no	120	2.971	145.1575	269	3.384	152.1804	49	3.188	148.1493	27	3.26	155.1717	3.201	0.173	3.06	4.07
(vi)	C54H60O28	no	no	–	33.8	1156.3281	–	33.8	1210.51	–	33.8	1156.328	–	33.8	1210.51	33.8	–	–	–

Six example species have been selected and are summarized in the table above. Three were identified by both database and calculated strategies, one was identified just the database strategy, one was identified by just the calculated strategy, and one was missed by both strategies. On the following pages will show two frames for each labeling regime with: 1) annotated mass spectra of the isotopic envelopes for each species, and 2) extracted ion chromatograms (EICs) for significant peaks in each envelope. EICs were limited to a minimum extraction width of 0.5 amu, which was used throughout. Monoisotopic peaks are enclosed in light blue boxes in the envelope spectra for each regime. If any feature was extracted within the relevant m/z and chromatographic time scale it is marked in the spectrum window with an asterisk (*) and annotated with the feature ID#, the m/z, and the retention time in minutes. The positions of the features are also mapped in the EICs. As mentioned in the text, often multiple features are extracted for the same m/z, within mass error, and within the examined chromatographic window. There will be brief comments associated with each page of annotated spectra and EICs and no further summary.

Examples of Extracted Features
Supplement to: Hegeman et al. 2007 Anal. Chem.

Page 1.

C$_9$H$_{15}$N
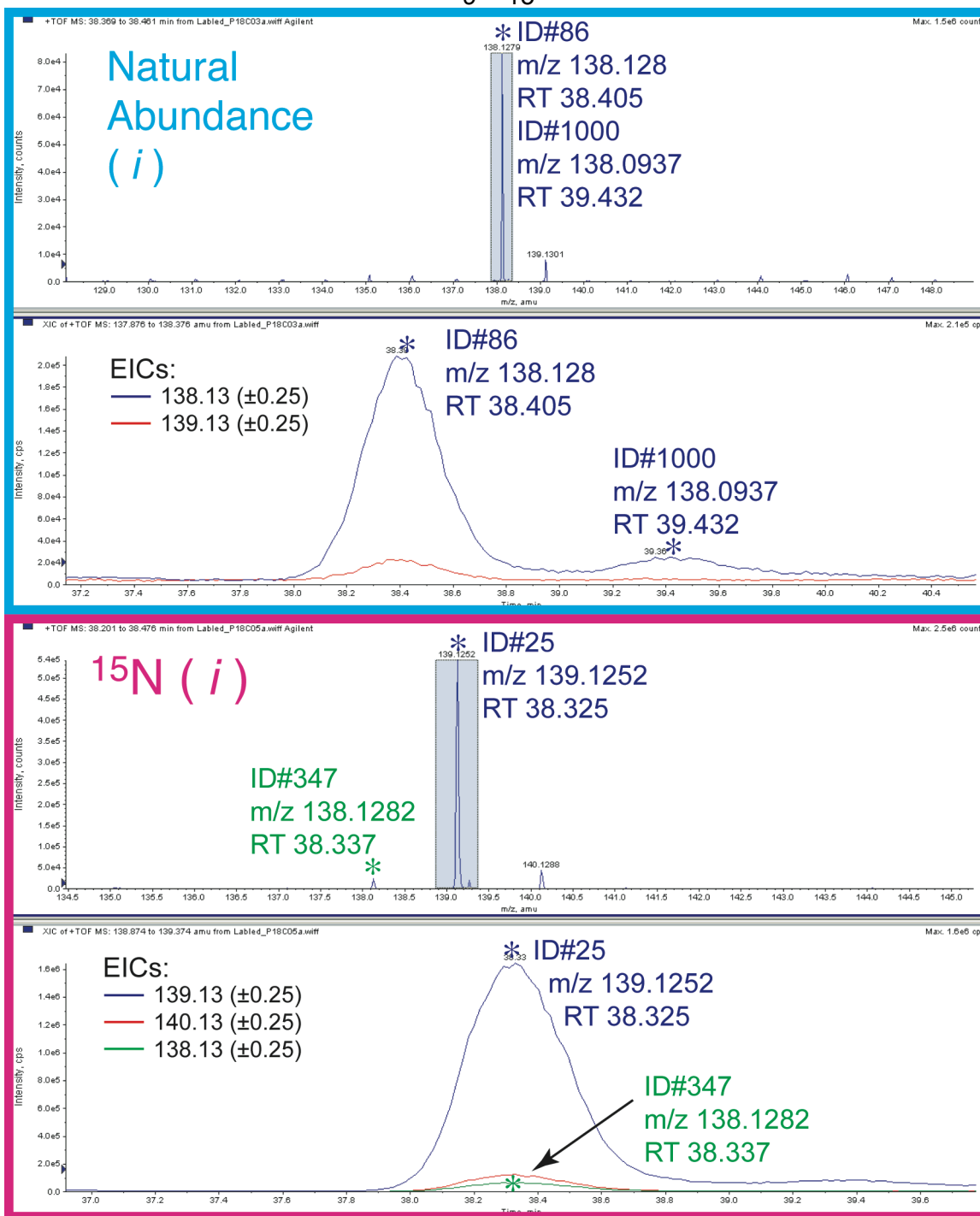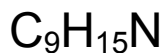
- Deisotoping works well with natural abundance as expected.
- 2 features assigned to natural abundance data for two species well resolved in both mass and chromatographic properties with well behaved elution properties.
- ^{15}N has an extra feature assigned to the coeluting –1 peak owing to 1-2% abundance of ^{14}N.

C$_9$H$_{15}$N continued…
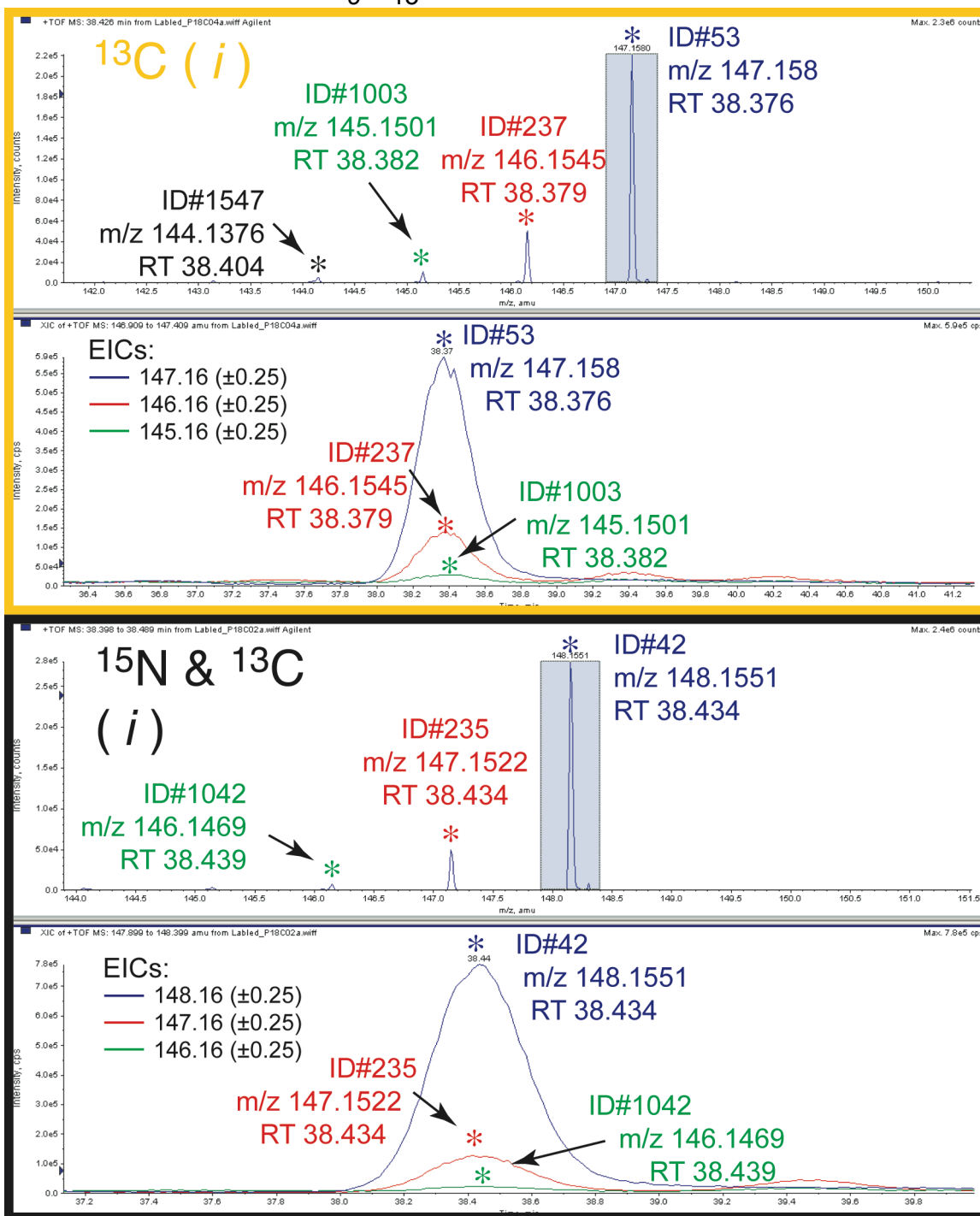
- Both the ^{13}C and double labeled spectra show lots of extra features assigned to isotopic envelope peaks that appear below the monoisotopic peak. In this example every major envelope peak is assigned. The ^{13}C appears to be significantly less well enriched in ^{13}C than the double labeled sample.
- The database strategy failed because it did not contain a compound with this formula.
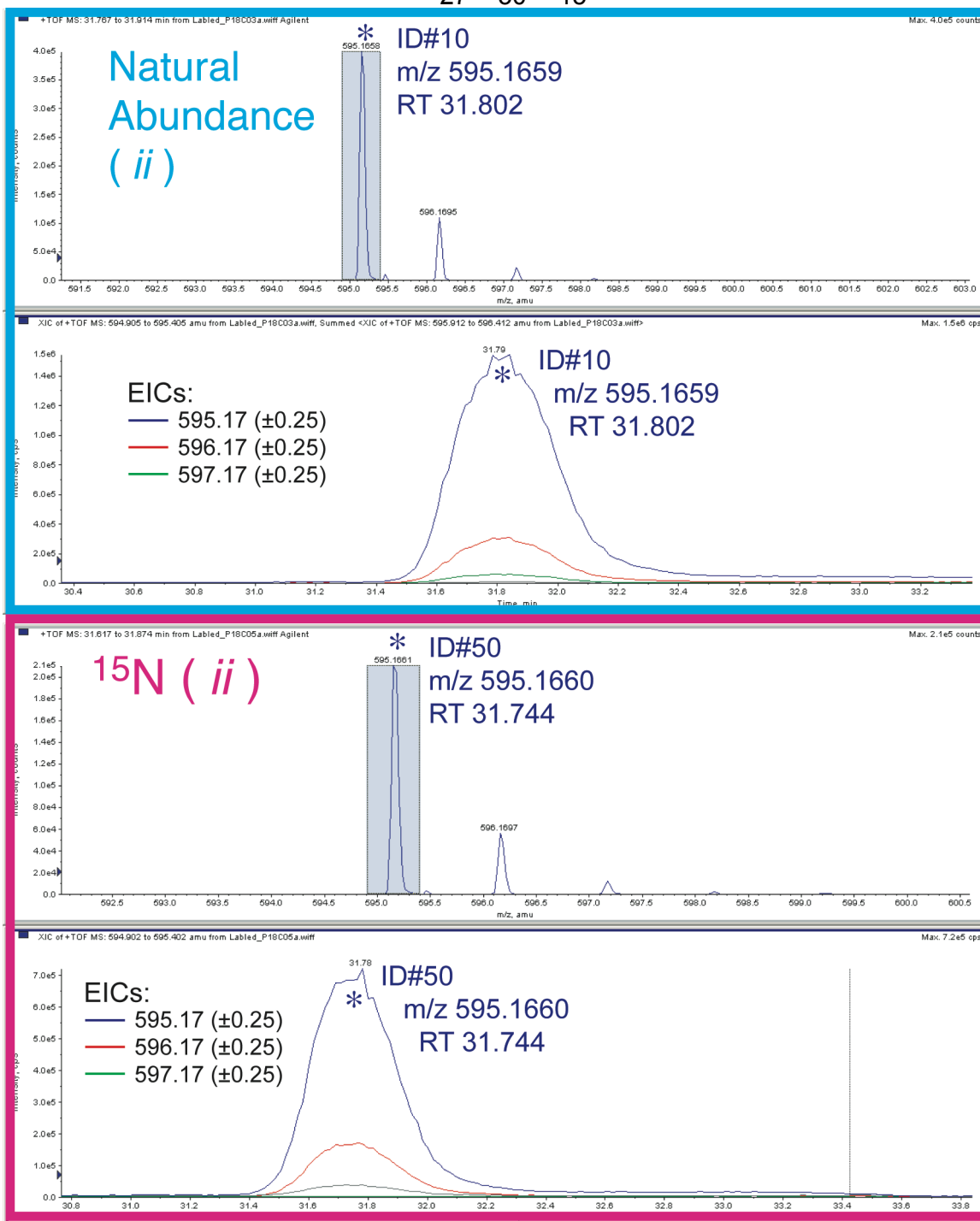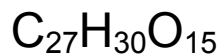
$C_{27}H_{30}O_{15}$
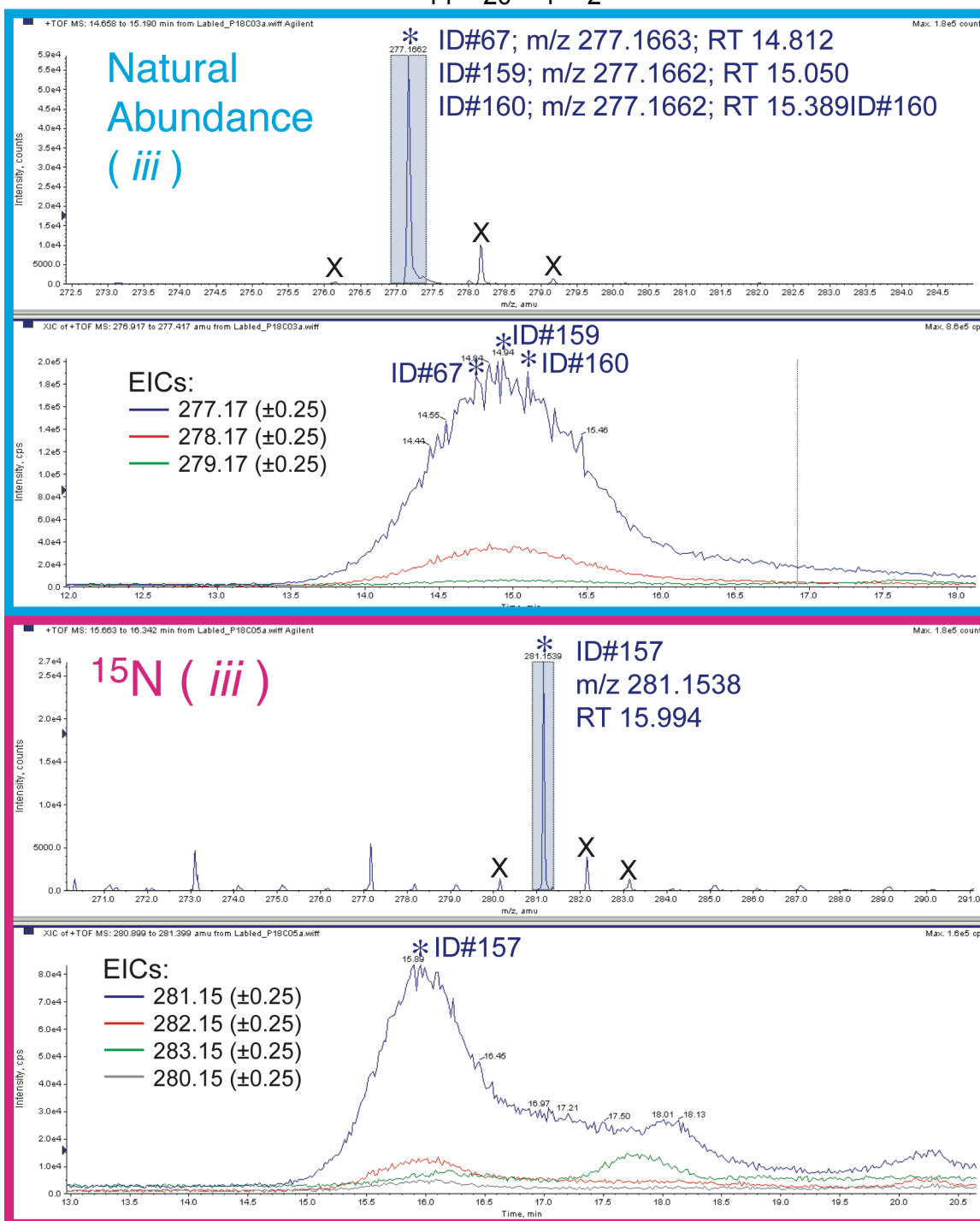

- This compound contains no nitrogen, so both spectra are pretty much the same. Monoisotopic peaks are the only assigned features.

Examples of Extracted Features   Page 4.

Supplement to: Hegeman et al. 2007 Anal. Chem.

C$_{27}$H$_{30}$O$_{15}$ continued…
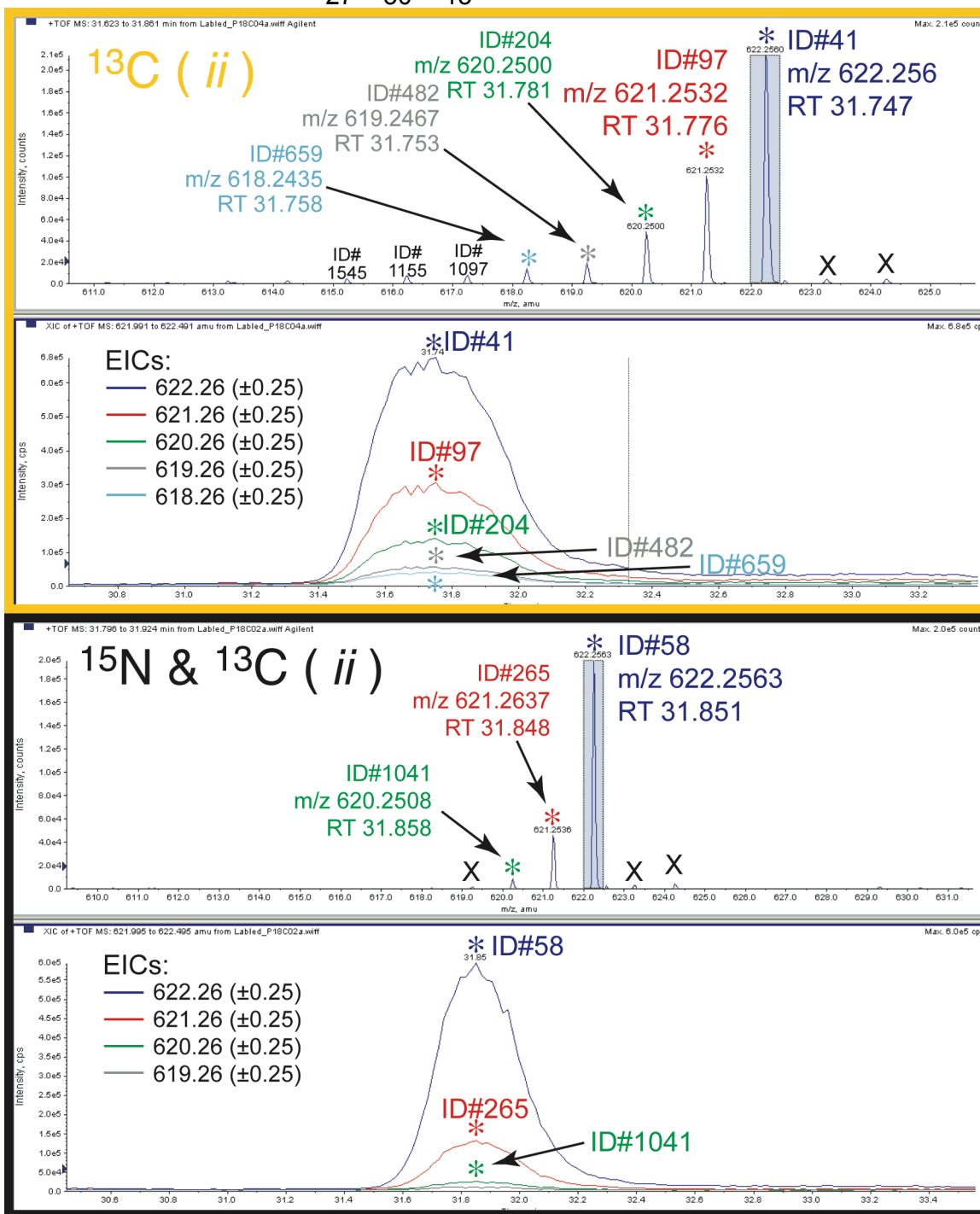
- Again, a huge string of peaks owing to ^{12}C contamination are assigned feature status in both ^{13}C-containing samples.
- The 'X's mark peaks that were double-checked for feature assignment and none was found.
- Both calculated and database strategies succeeded in assigning a formula.

$C_{14}H_{20}N_4O_2$

Natural Abundance (*iii*)

ID#67; m/z 277.1663; RT 14.812
ID#159; m/z 277.1662; RT 15.050
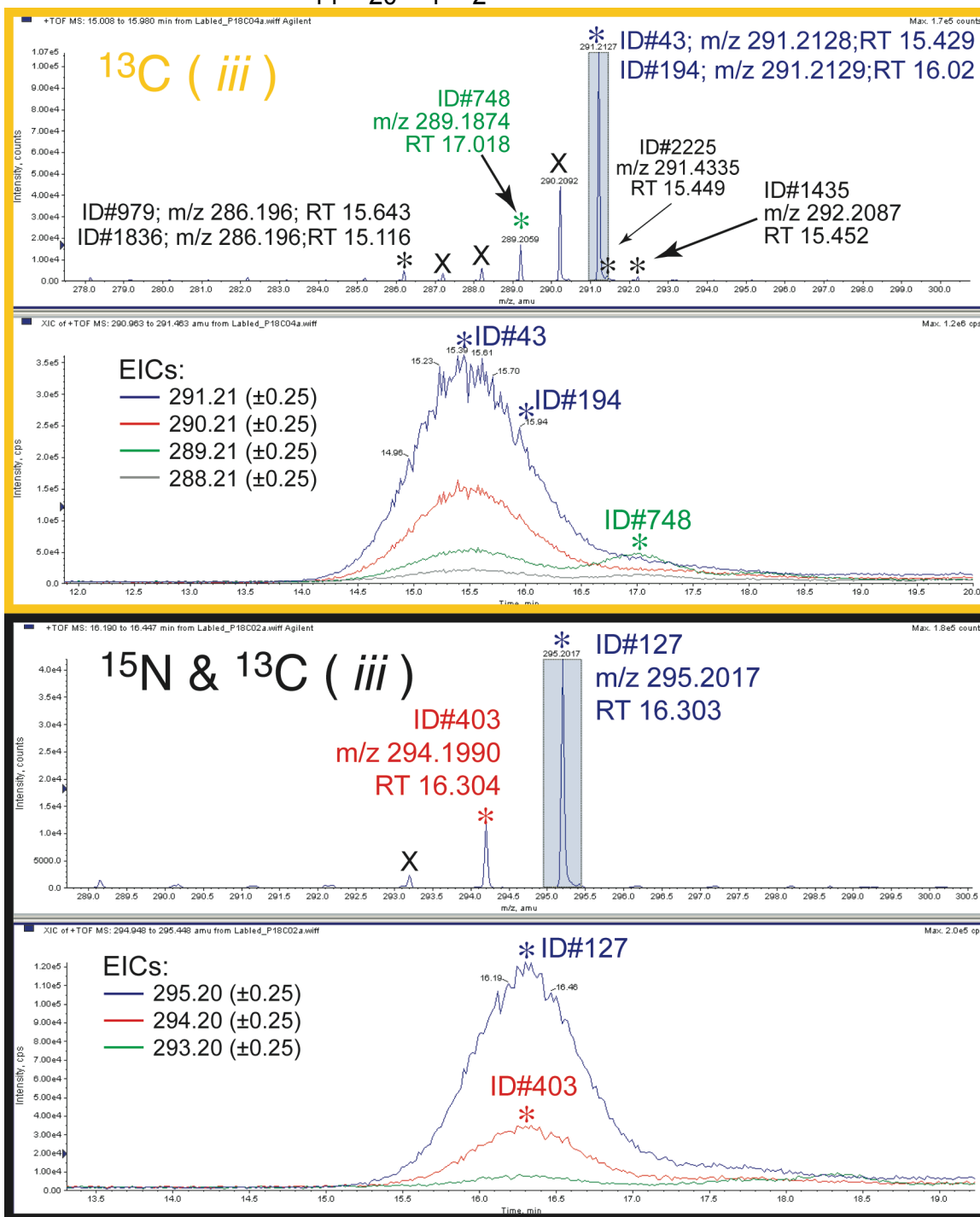ID#160; m/z 277.1662; RT 15.389ID#160

EICs:
— 277.17 (±0.25)
— 278.17 (±0.25)
— 279.17 (±0.25)

^{15}N (*iii*)

ID#157
m/z 281.1538
RT 15.994

EICs:
— 281.15 (±0.25)
— 282.15 (±0.25)
— 283.15 (±0.25)
— 280.15 (±0.25)
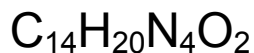
- The natural abundance sample shows how a peak that is clearly one compound can be split up into three separate assigned features. In this example, the jaggedness of the EIC is the result of lower signal intensity, and may indicate that smoothing is needed to prevent split assignments.
- The 'X's mark peaks that were double-checked for feature assignment and none was found.

Examples of Extracted Features
Supplement to: Hegeman et al. 2007 Anal. Chem.

Page 6.

$C_{14}H_{20}N_4O_2$ continued…

^{13}C (*iii*)

ID#43; m/z 291.2128; RT 15.429
ID#194; m/z 291.2129; RT 16.02

ID#748
m/z 289.1874
RT 17.018

ID#2225
m/z 291.4335
RT 15.449

ID#1435
m/z 292.2087
RT 15.452

ID#979; m/z 286.196; RT 15.643
ID#1836; m/z 286.196; RT 15.116

EICs:
— 291.21 (±0.25)
— 290.21 (±0.25)
— 289.21 (±0.25)
— 288.21 (±0.25)

ID#43
ID#194
ID#748

^{15}N & ^{13}C (*iii*)

ID#127
m/z 295.2017
RT 16.303

ID#403
m/z 294.1990
RT 16.304

EICs:
— 295.20 (±0.25)
— 294.20 (±0.25)
— 293.20 (±0.25)

ID#127
ID#403

- Low signal intensity results in some jaggedness and split feature assignment.
- Interestingly, the coeluting –1, –2, –3 and –4 peaks were not assigned feature status while other closely eluting species were (#748, #1435, #2225 etc.).
- Double labeled spectrum is typical of what we have seen already.
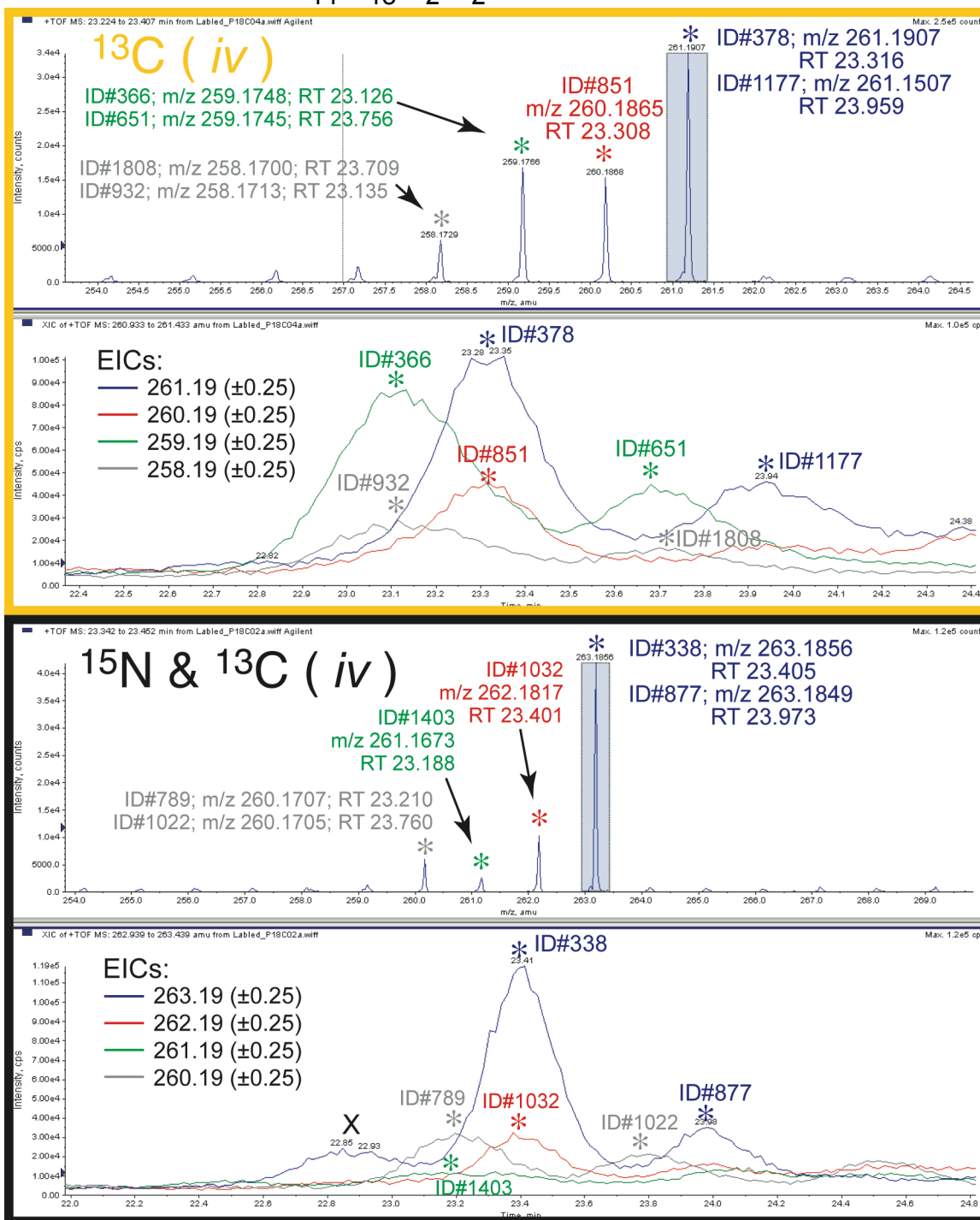- Both calculated and database strategies succeeded in assigning a formula.

$C_{14}H_{18}N_2O_2$

Natural Abundance (*iv*)

ID#221; m/z 247.1444; RT 23.318
ID#739; m/z 247.1441; RT 23.885

ID#248
m/z 250.1439
RT 23.17

EICs:
— 247.14 (±0.25)
— 248.14 (±0.25)
— 249.14 (±0.25)

ID#221
ID#739

^{15}N (*iv*)

ID#217; m/z 249.1385; RT 23.279
ID#525; m/z 249.1371; RT 23.877
ID#588; m/z 249.0916; RT 24.619

ID#920
m/z 248.1404
RT 21.263

ID#460; m/z 251.0819; RT 23.877
ID#560; m/z 251.1397; RT 23.676
ID#407; m/z 251.1402; RT 23.068

EICs:
— 249.13 (±0.25)
— 248.13 (±0.25)
— 247.13 (±0.25)
— 250.13 (±0.25)
— 251.13 (±0.25)

ID#217
ID#407
ID#560
ID#525
ID#588
ID#460
ID#920

- Here is a really busy part of the spectrum/chromatographic range.
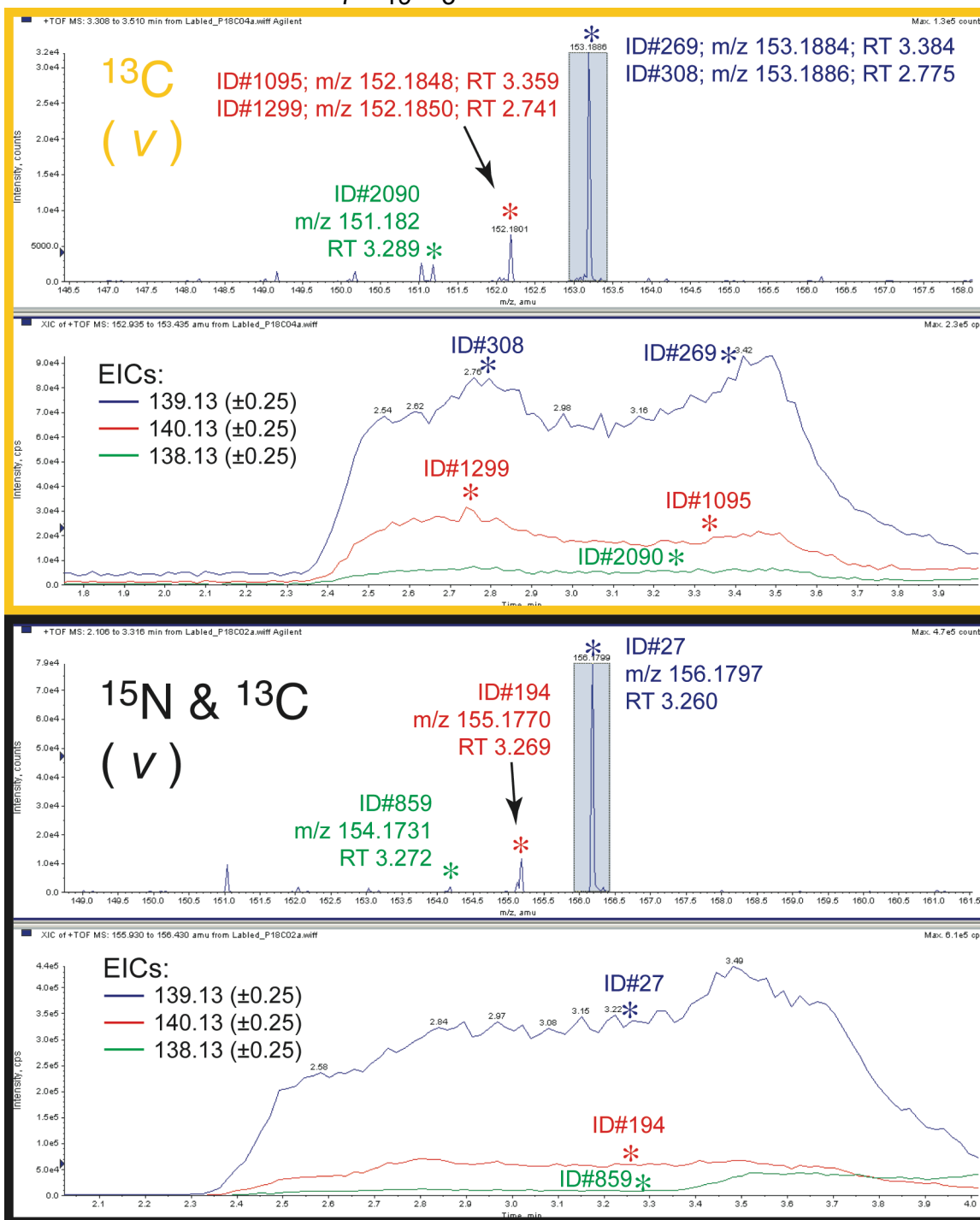
C₁₄H₁₈N₂O₂ continued…

- These are also both messy, but show similar patterns of –1, –2 etc peak assignment as before.
- Both calculated and database strategies succeeded in assigning a formula to the dominant species in this cluster.

Examples of Extracted Features

Page 9.

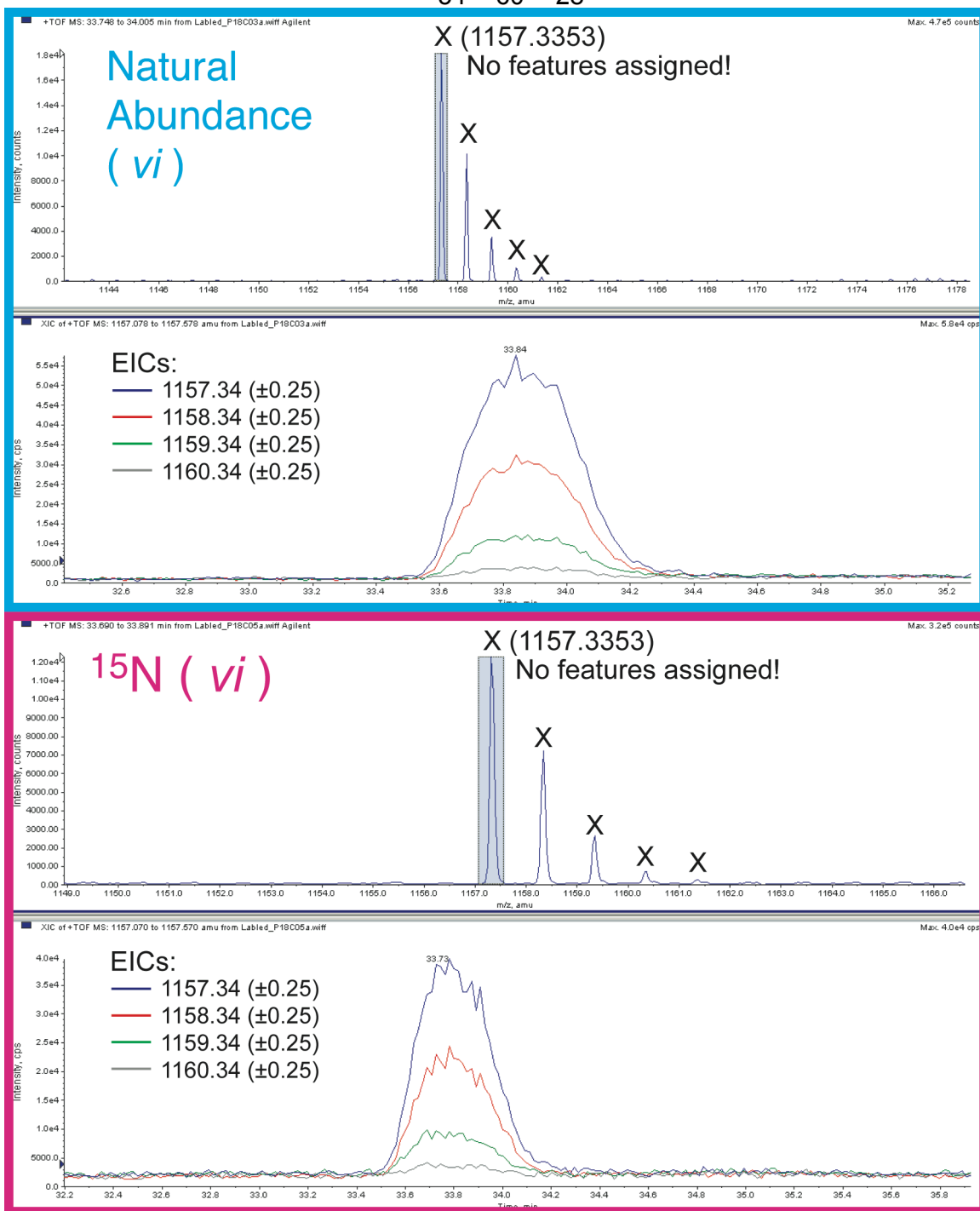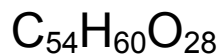Supplement to: Hegeman et al. 2007 Anal. Chem.
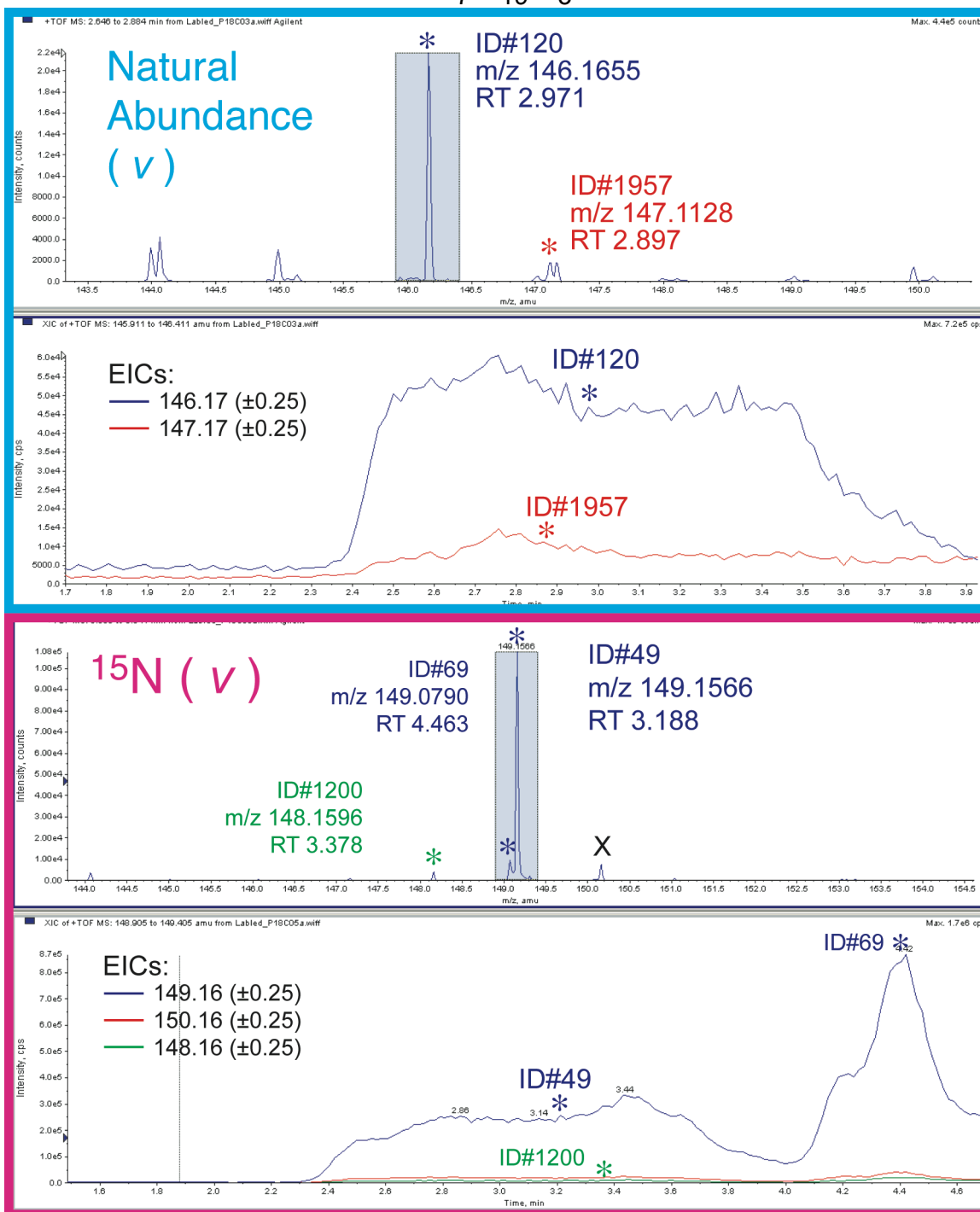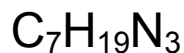
$C_7H_{19}N_3$

- This is a compound eluting in the flow through. The peak shape is terrible and there are lots of coeluting species suppressing ionization and overlapping in the spectra. Features are assigned within non-ideal elution profiles.

Examples of Extracted Features

Supplement to: Hegeman et al. 2007 Anal. Chem.

Page 10.

C$_7$H$_{19}$N$_3$ continued…

^{13}C
(v)

ID#1095; m/z 152.1848; RT 3.359
ID#1299; m/z 152.1850; RT 2.741

ID#269; m/z 153.1884; RT 3.384
ID#308; m/z 153.1886; RT 2.775

ID#2090
m/z 151.182
RT 3.289

EICs:
— 139.13 (±0.25)
— 140.13 (±0.25)
— 138.13 (±0.25)

ID#308   ID#269
ID#1299
ID#1095
ID#2090

^{15}N & ^{13}C
(v)

ID#27
m/z 156.1797
RT 3.260

ID#194
m/z 155.1770
RT 3.269

ID#859
m/z 154.1731
RT 3.272

EICs:
— 139.13 (±0.25)
— 140.13 (±0.25)
— 138.13 (±0.25)

ID#27
ID#194
ID#859

- The database strategy succeeds in this example while the calculated strategy fails. Possibly the large number of essentially coeluting species introduces additional stochastic elements limiting assignment via the calculated strategy. Accidentally matching features from this dense region will compete to with appropriate matches to increase false positive formula assignments.

$C_{54}H_{60}O_{28}$

Natural Abundance (*vi*)

X (1157.3353)
No features assigned!

EICs:
— 1157.34 (±0.25)
— 1158.34 (±0.25)
— 1159.34 (±0.25)
— 1160.34 (±0.25)

^{15}N (*vi*)

X (1157.3353)
No features assigned!

EICs:
— 1157.34 (±0.25)
— 1158.34 (±0.25)
— 1159.34 (±0.25)
— 1160.34 (±0.25)

- This example is a clear failure to assign <u>any</u> features to this species. It is well separated chromatographically and so manual assignment was trivial.

$C_{54}H_{60}O_{28}$ continued…
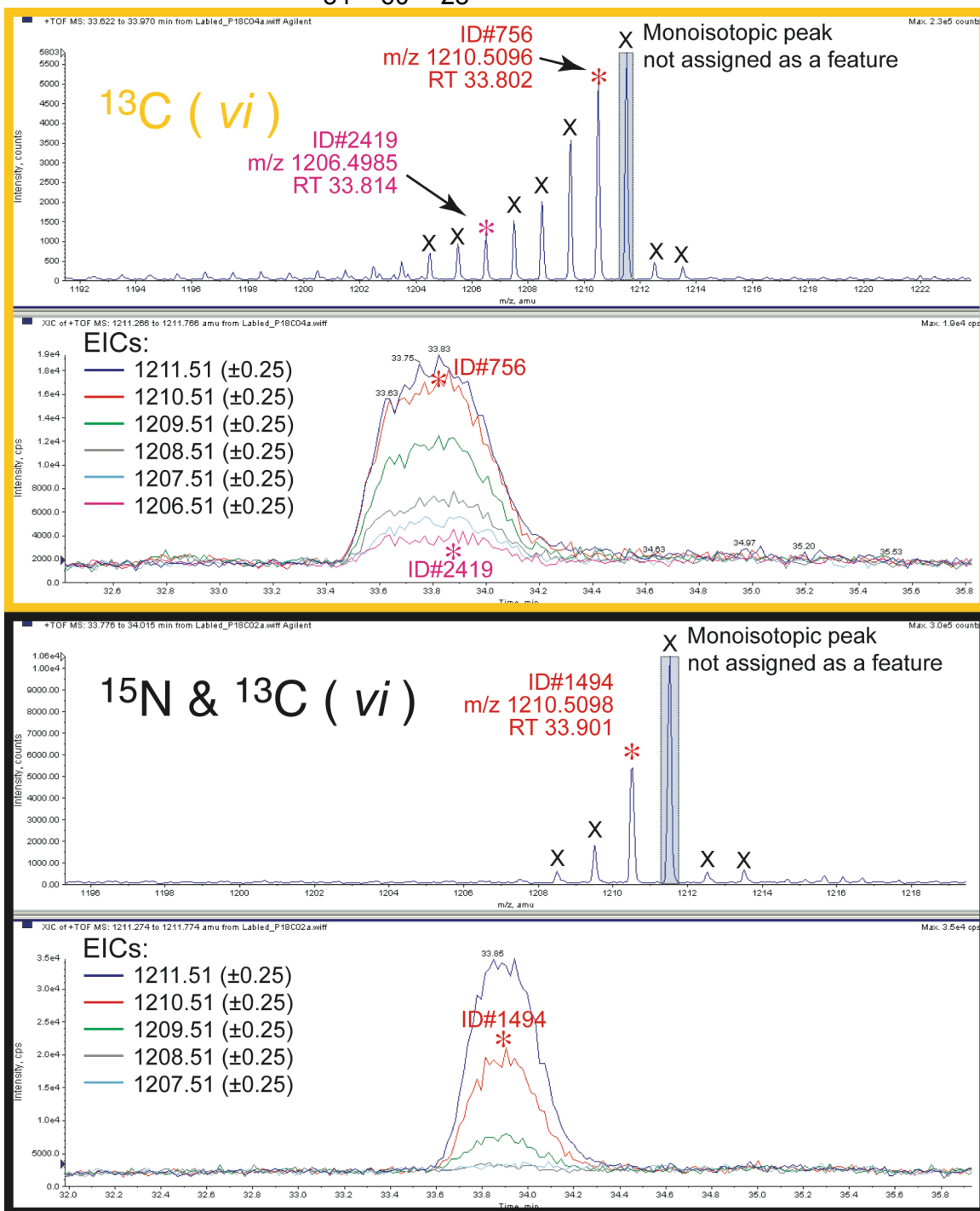
- The monoisotopic masses were missed in all four regimes, and a handful of less intense isotopic peaks were assigned in the ^{13}C and double-labeled samples. These serve as good internal controls for intensity limits, which in this case were not passed.
- We were disappointed to have missed this particular compound as it is a great example of the usefulness of our approach with a single formula being assigned to a mass as high as ~1157.