Leveraging MS1 Data in NTA Workflows to Improve PFAS Discovery

David Schiessel
Babcock Labs, Inc.
dschiessel@babcocklabs.com

Outline

The Lop-Sided Nature of NTA Data Features

Case of No MS/MS data

Recent Developments in PFAS Prioritization

Kendrick, Kaufmann, and Zweigle

RT Models and Molecular Formula Decomposition

Conclusion

Non-Targeted Analyses generate results with varying confidence levels.

What is NTA?
Non-Targeted Analysis

BP4NTA - The characterization of the chemical composition of any given sample without the use of a priori knowledge regarding the sample's chemical content.

The **framework** by which a defined **chemical space** is investigated within a sample without a priori knowledge for the primary purpose of **chemical discovery**.

NTA Feature Distribution

Body of Knowledge Limited

Features in Electrospray may be chemicals, in-source fragments, or clusters/adducts

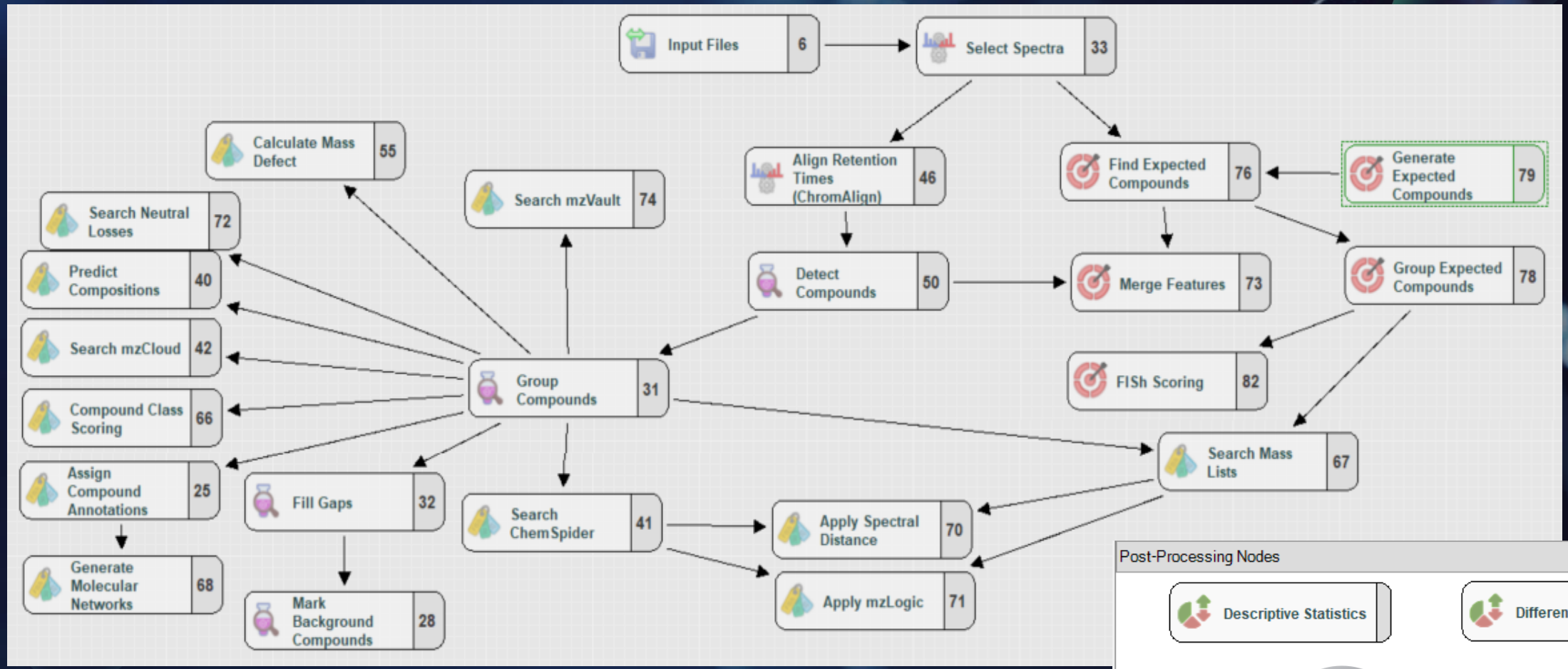
Chemical space of full scale NTA study is vast
(+9000 for drinking water)

More difficult to deconvolute In-source fragments than clusters
(once a bond breaks, the neutral is lost)

Confidence Annotation of features naturally heavily weighted on lower confidence scores
(Few level 1 and many Level 4/5)

Complete NTA Workflow

Example – Thermo Compound Discoverer 3.3



No MS/MS data
Conundrum of ddMS2 Workflows

Typical Full Scan NTA acquisition has a data-dependent MS2 component to it.

Triggered only when chemical feature is above threshold

Trade-off of ddMS2 over All ions fragmentation (AIF) is ease of deconvolution

Ultimately, many features do not have MS/MS data

Can we increase identification confidence for the many features that have no MS2 data?

Kendrick Mass / Mass Defect

Plot of Kendrick Mass Defect (KMD) vs. MW

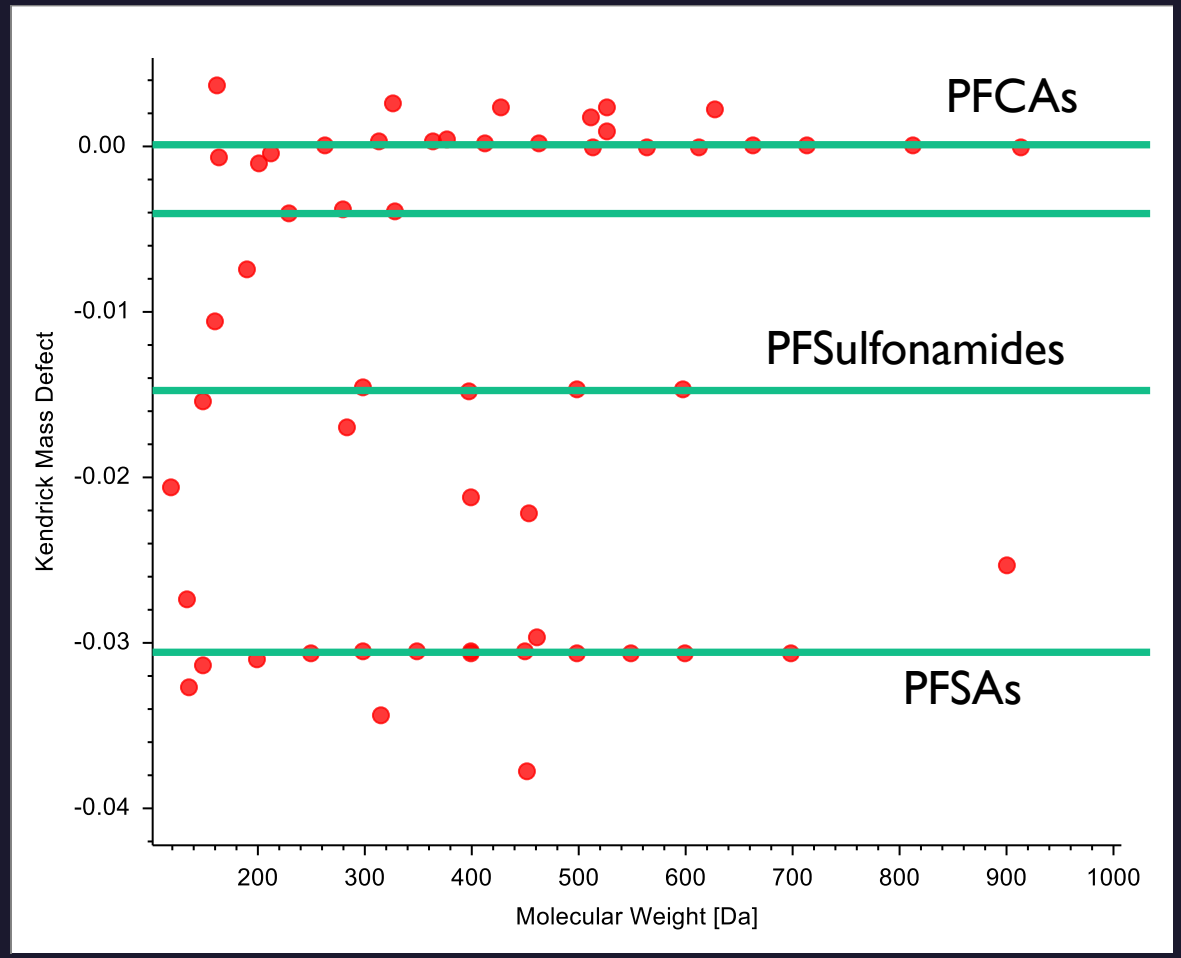
KMD(CF2) = M * round(KFM) / KFM – round(M)

Where KFM = 49.9968 and M = mass(feature)

Produces horizontal lines where CF2 homologs occur

Mass Defect = MW – **floor**(MW)

Disadvantage: Scales poorly to non CF2 homologs



What's in a M/Z

Example: [M-H]- ion for PFOA

<u>For PFAS</u>:

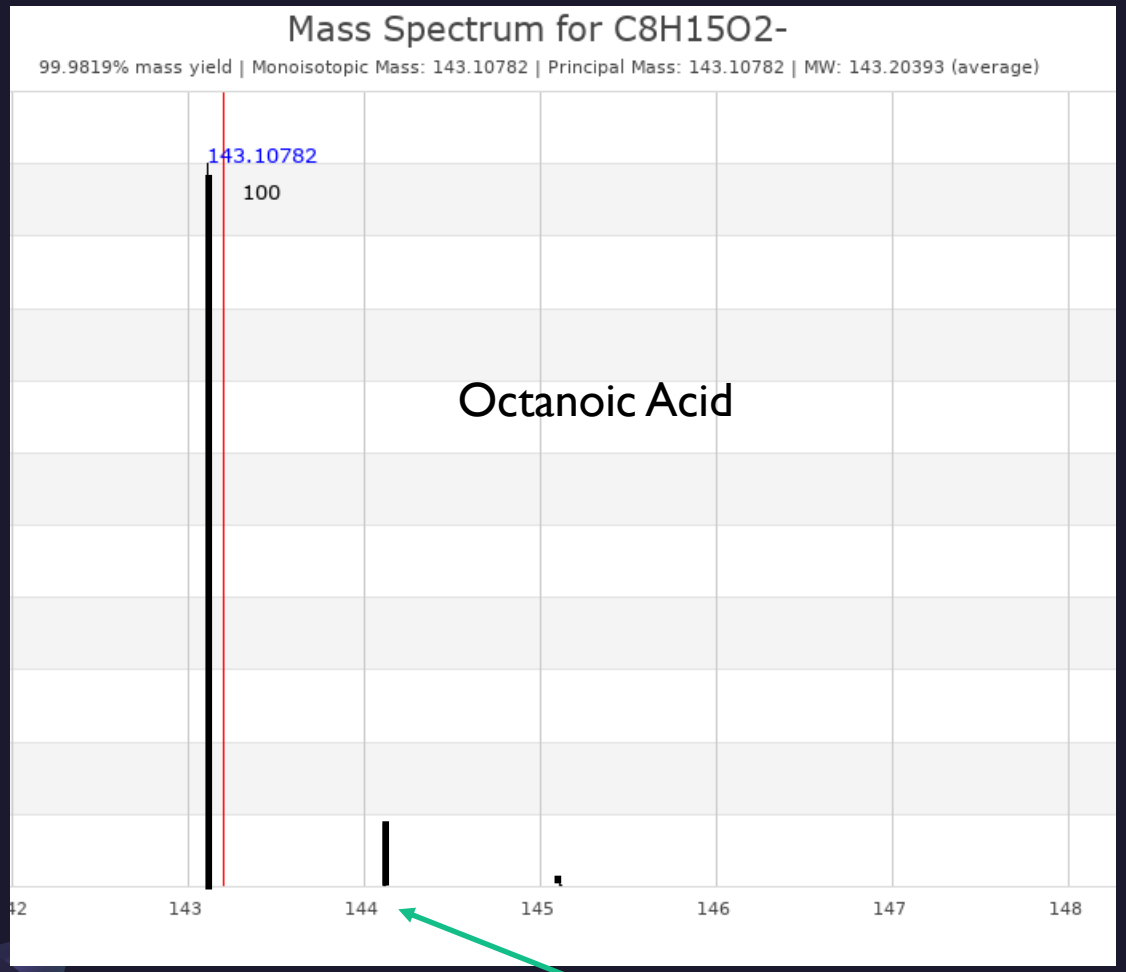
- m/z is larger because F atoms have replaced H atoms

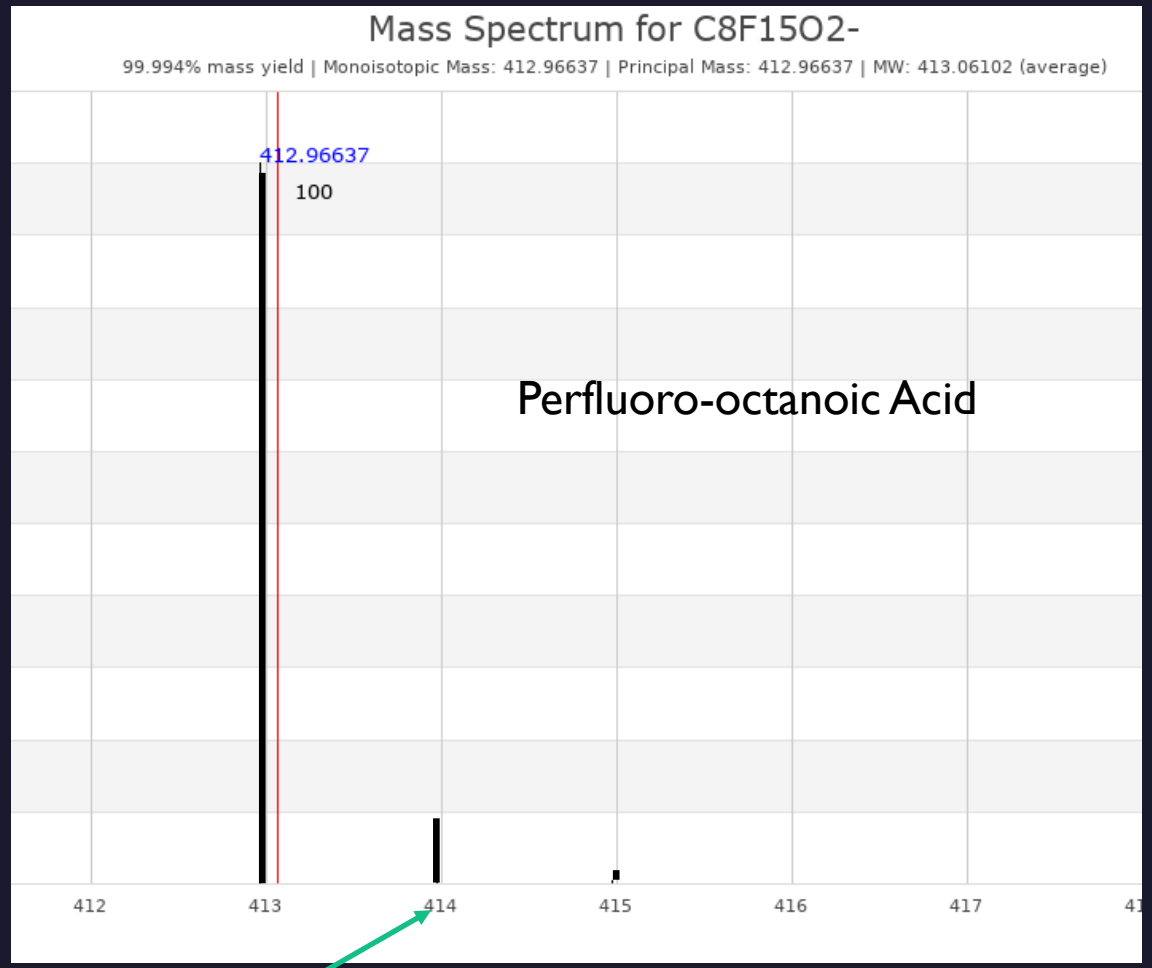
- Mass defect becomes larger as more F atoms added

412.96637

Mass Defect – Δ from integer

What's in a Mass Spectrum (MS1 Full Scan)





If Carbon 13 abundance is 8.73%, this feature has 8 carbons (independent of F atoms)

What’s in a Mass Spectrum (MS1 Full Scan)

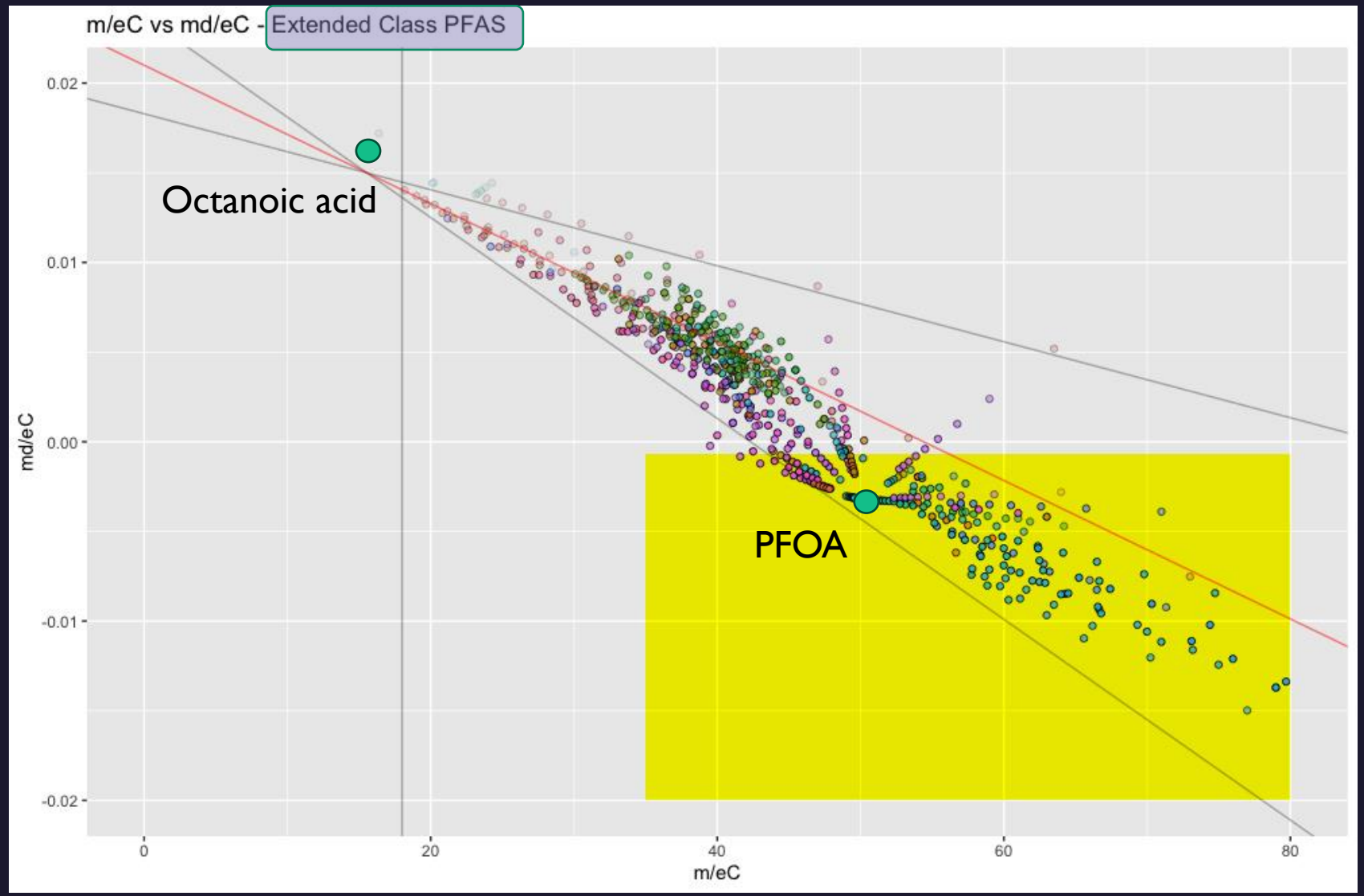
OCTANOIC ACID

- MF = C8 H16 O2
- MW ~ 143
- MW/Carbon ~ 18
- Mass Defect = 0.10782
- MD/Carbon = 0.0135

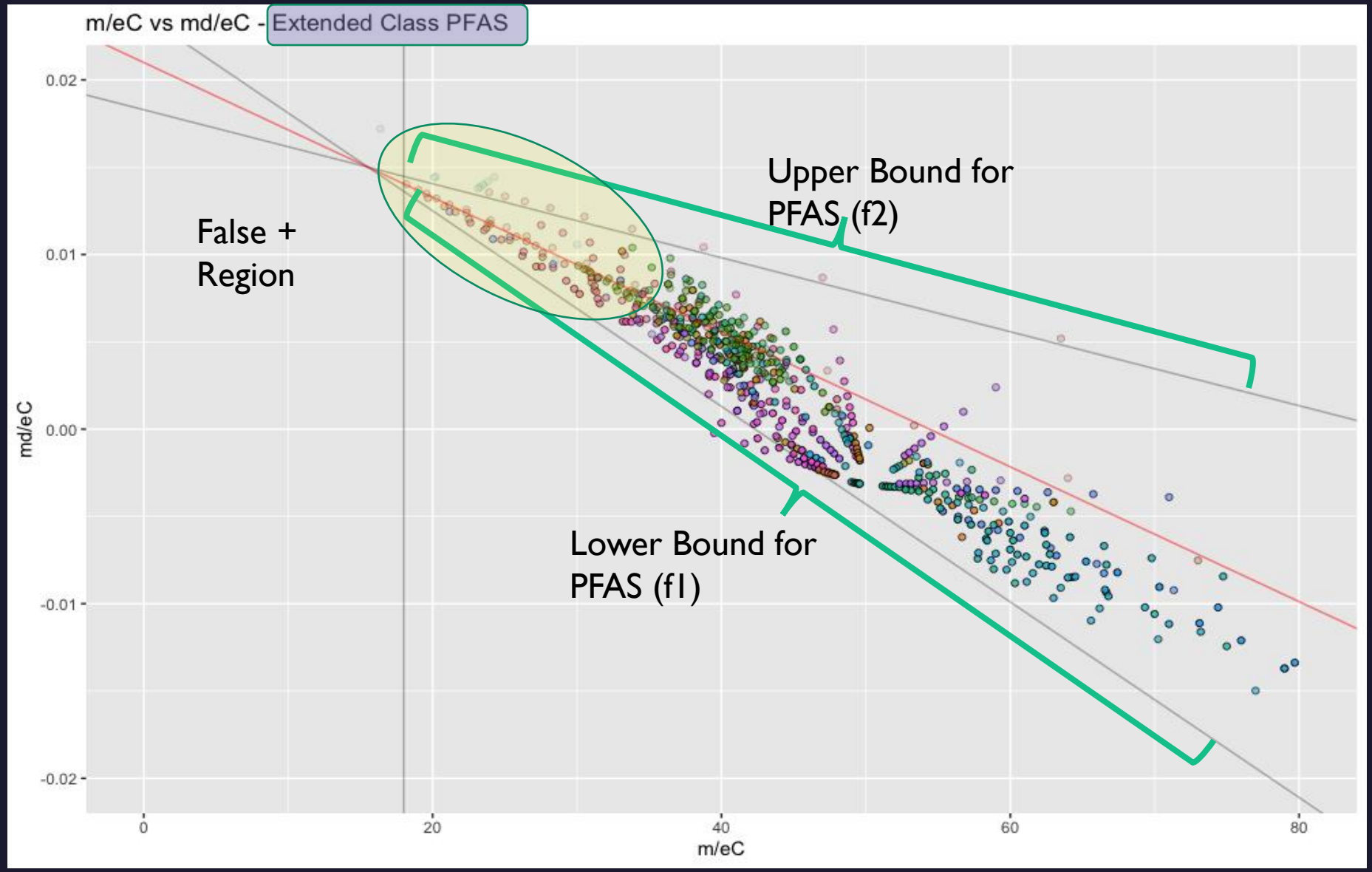
• PFOA

- MF = C8 H F15 O2
- MW ~ 413
- MW/Carbon ~ 52
- Mass Defect = -0.03363
- MD/Carbon = -0.0042

Kaufmann Plot of PFAS (md/C vs. m/C)



Kaufmann Plot of PFAS (md/C vs. m/C)



PFAS-ness – Upper and Lower Bounds

Tags	Checked	Name	Formula	RT [min]	m/z	Area (Max.)	Annot. ΔMass [ppm]	MS2	Annot. Source	mzVault Library Matches	Mass List Matches	eC	m/C	md/C	f1PFAS	f2PFAS	Calc. MW
	☑	TFMSi	C H F3 O2 S	1.246	132.95675	18576503	-6.79					1.07535	124.5772	-0.0334(	12.60436	-25.34853	133.96402
	☑	TFMS	C H F3 O3 S	1.327	148.95173	110842740	-5.65					1.09651	136.7597	-0.0373(	15.49424	-26.69815	149.95900
	☑	TFMSA	C H2 F3 N O2 S	1.767	147.96772	42968824	-5.64					1.16133	128.2793	-0.0215	26.60083	-12.64038	148.97499
	☑	PFPrA	C3 H F5 O2	2.176	162.98162	157987627	-5.32					2.80207	58.52420	-0.0040(	5.07650	-9.88992	163.98880
	☑	PFEtS	C2 H F5 O3 S	3.477	198.94889	124526386	-2.43					0.55187	362.3254	-0.0794	99.78122	-20.90803	199.95617
	☑	PFBA	C4 H F7 O2	4.393	168.98864	523352731	-2.83					3.03458	70.51583	-0.0046(	11.13253	-8.00697	213.98587
	☑	PFPrS	C3 H F7 O3 S	4.769	248.94611	172347615	-0.31					2.78246	89.83194	-0.0167	9.85090	-16.01061	249.95338
	☑	PFMPA, PFOPeA	C4 H F7 O3	4.823	228.97388	41856299	-1.02					3.93709	58.4140	-0.0047	4.22659	-10.70149	229.98116
	☑	Bistriflimide	C2 H F6 N O4 S2	4.925	279.91786	172916023	0.08					0.69323	405.2421	-0.1079	95.24803	-40.37625	280.92514
	☑	PFPeA	C5 H F9 O2	5.353	262.97598	167080174	-0.59					5.16132	51.14645	-0.0032	1.67347	-10.72550	263.98313
	☑	PFBS	C4 H F9 O3 S	5.501	298.94303	328331590	0.14					3.74778	80.03408	-0.0132(	7.86058	-14.59128	299.95031
	☑	PFMBA, PFOHxA	C5 H F9 O3									4.93940	56.68268	-0.0044	3.62880	-10.69677	279.97820
	☑	3:3-FTCA	C6 H5 F7 O2									6.48017	37.34740	0.00271	-0.07565	-7.67255	242.01756
	☑	H-PFHpA	C7 H2 F12 O2									7.61739	45.42055	-0.0018(	-0.06563	-10.47198	345.98628

eC	m/C	md/C	f1PFAS	f2PFAS
1.07535	124.5772	-0.0334(	12.60436	-25.34853
1.09651	136.7597	-0.0373(	15.49424	-26.69815
1.16133	128.2793	-0.0215	26.60083	-12.64038
2.80207	58.52420	-0.0040(	5.07650	-9.88992
0.55187	362.3254	-0.0794	99.78122	-20.90803
3.03458	70.51583	-0.0046(	11.13253	-8.00697
2.78246	89.83194	-0.0167	9.85090	-16.01061
3.93709	58.41401	-0.0047	4.22659	-10.70149
0.69323	405.2421	-0.1079	95.24803	-40.37625
5.16132	51.14645	-0.0032	1.67347	-10.72550
3.74778	80.03408	-0.0132(	7.86058	-14.59128
4.93940	56.68268	-0.0044	3.62880	-10.69677

Zweigle et al did similar work calculating a single value using a more complicated calculation

Machine Learning Approach to RT Modeling

RT Modeling – How and Why

Using R packages: rcdk (QSAR), neuralnet to generate a 7:4 MLP

Pretty good prediction +/- 1.0 min, some outliers

Not universal and not directly transferable

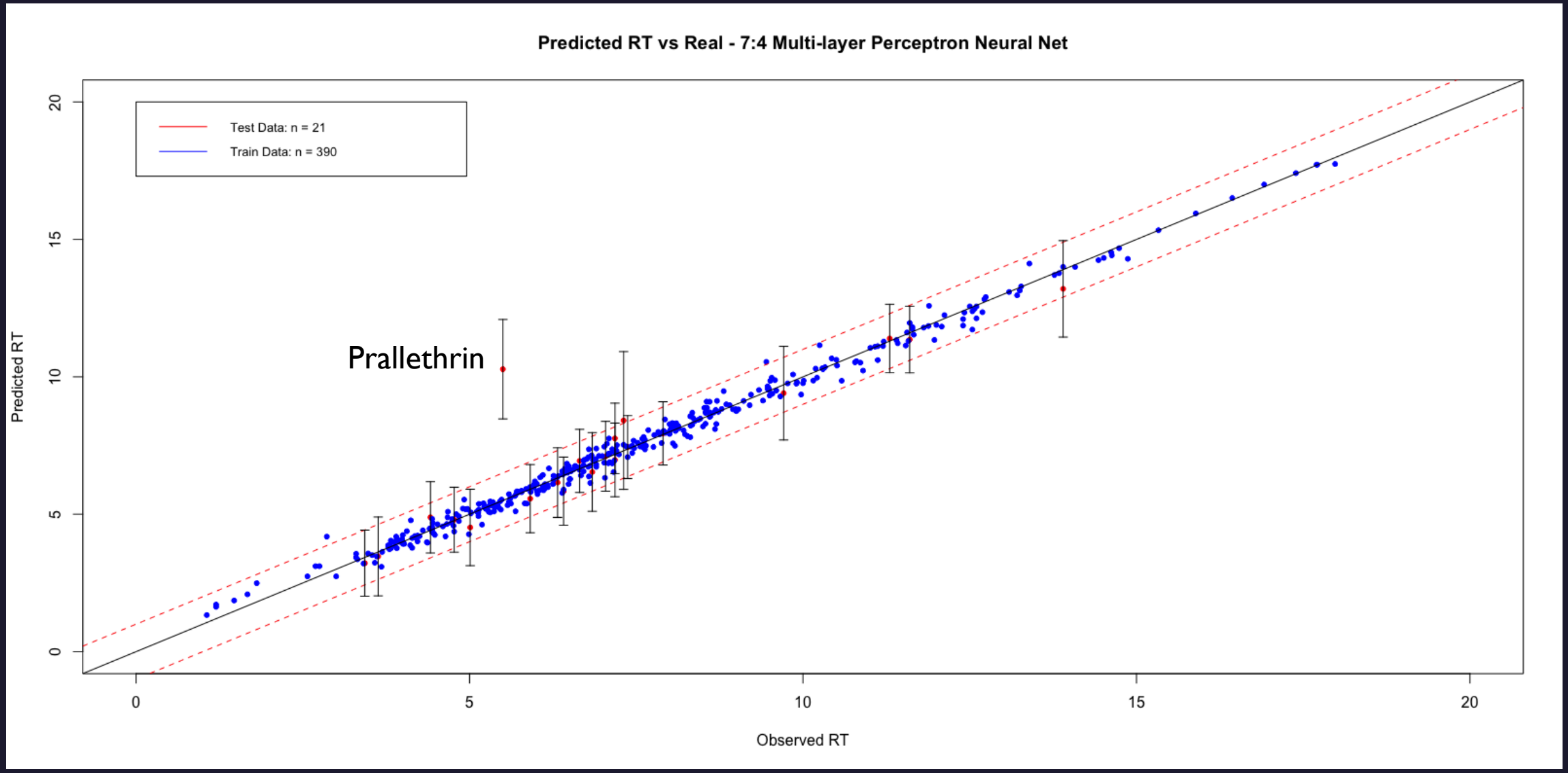
Overly trained for PFAS (80+ PFAS) but includes pesticides, CECs

Did initial Cross-Validation from 10% to 90% training, run with 5 replicates of randomized data. RMSE minimization confirmed.

Dominant QSARs that predict RT are usual suspects (eg: AlogP)

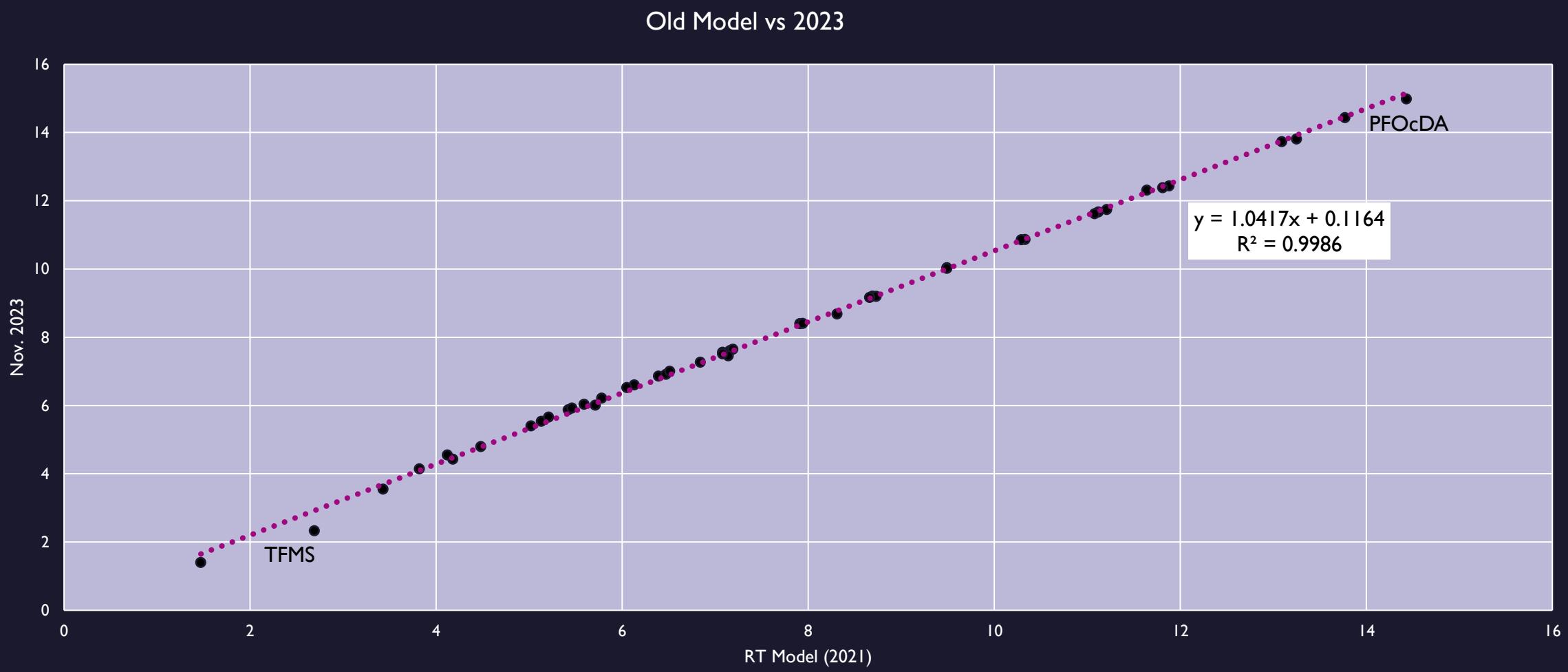
Another tool – to help increase ID confidence

RT Model – Observed/Predicted



RT Model (2021) vs. 2023

47 compounds – mostly PFAS

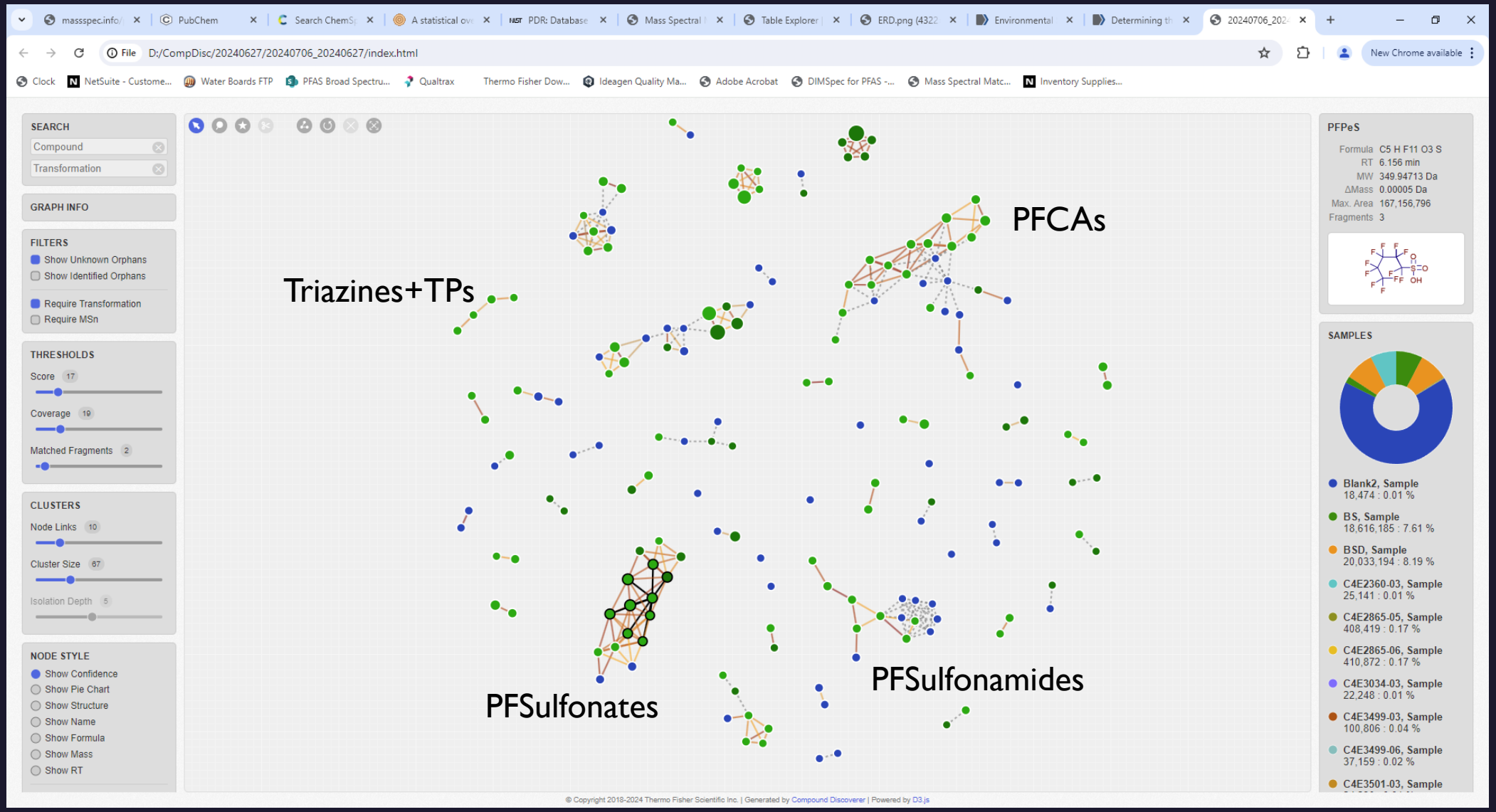


What's Required for an RT Model

- Structures of all molecules (SMILES)
- Specific Quantum Structure Activity Relationship (QSAR) calculations. Examples: ALogP, nHB Donors, nHB Acceptors, Elemental counts, polarizability)
- Train the model
- Validate the Model using various test/train ratios. Determine if RMSE goes down as train ratio increases
- Store the model
- Use model on Feature candidate lists (one feature, many possible chemicals)

Molecular Networks To Explore Feature Relationships

Molecular Networks (Thermo CD)



Informed DeNovo Molecular Formula Generation

Molecular Formula Generation

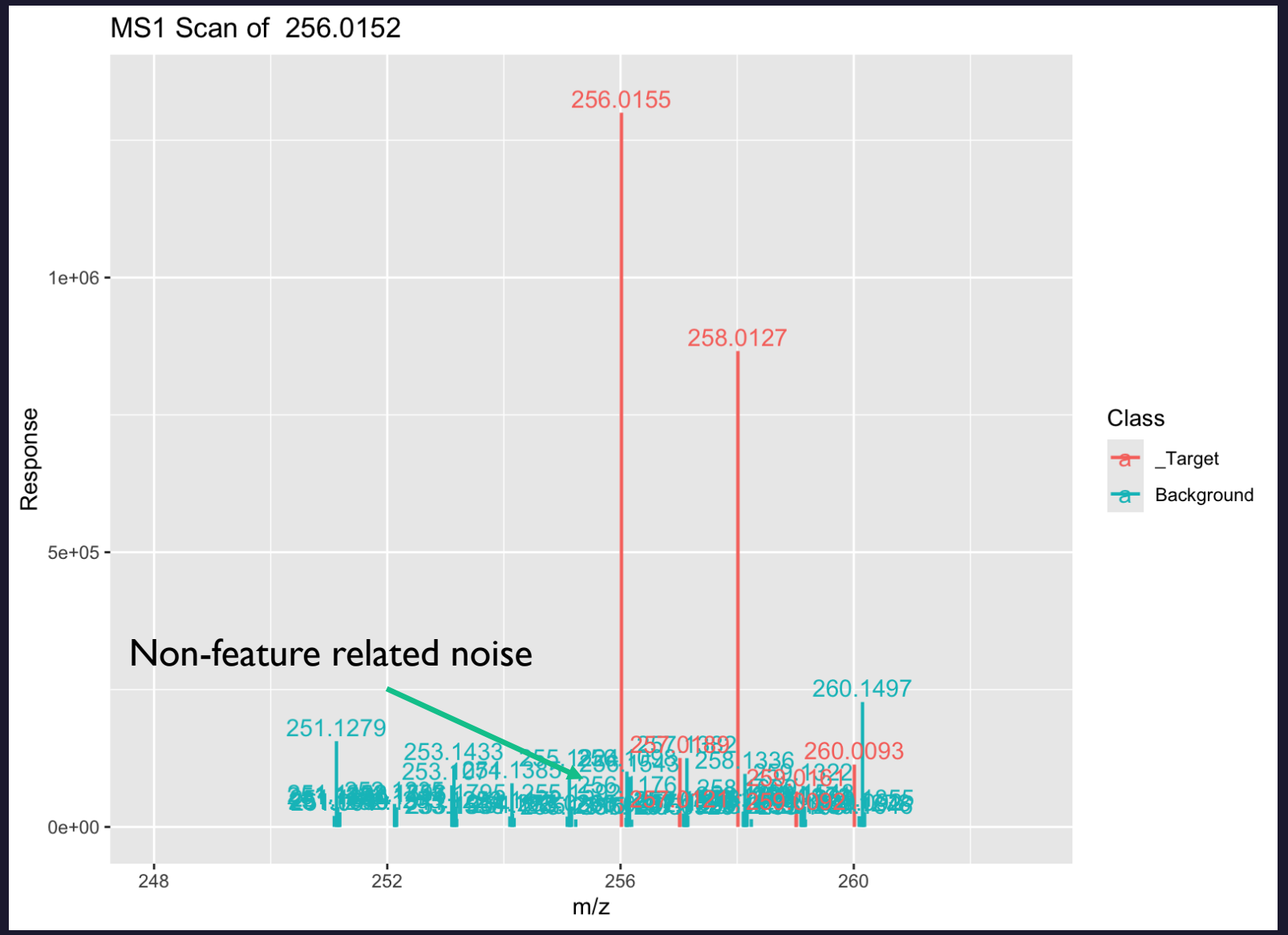
- DeNovo – "from the beginning"
- Decomposition computationally expensive / time consuming (last resort)
- Often produces more junk than useful information

	mz	MF	charge	RdisopScore	unsat	parity	error	nrule	ppm	SENIOR3	HtoC	NtoC
1	371.1009	C6H31O7N2S4	1	9.183108e-02	-7.5	e	-2.288724e-05	Valid	-0.0616739	3	5.1666667	0.3333333
5	371.1006	C13H23O10S	1	5.182191e-02	2.5	e	-2.393812e-04	Valid	-0.6450570	9	1.7692308	0.0000000
7	371.1006	C11H11N14S	1	5.010691e-02	13.5	e	-2.500732e-04	Valid	-0.6738686	45	1.0000000	1.2727272
8	371.1006	H150O10N2S	1	4.081205e-02	-73.0	e	-3.119322e-04	Valid	-0.8405591	-140	Inf	In
10	371.1012	C8H146O8	1	4.587981e-02	-64.0	e	3.399638e-04	Valid	0.9160953	-128	18.2500000	0.0000000
12	371.1005	C5H19O13N6	1	2.704001e-02	-0.5	e	-4.223192e-04	Valid	-1.1380173	5	3.8000000	1.2000000
14	371.1005	C3H7O3N20	1	2.590080e-02	10.5	e	-4.330112e-04	Valid	-1.1668289	41	2.3333333	6.6666667
17	371.1013	C6H23O8N6S2	1	3.147397e-02	-1.5	e	4.460708e-04	Valid	1.2020202	11	3.8333333	1.0000000
21	371.1014	CH154O5N2S3	1	2.005891e-02	-74.0	e	5.564578e-04	Valid	1.4994784	-134	154.0000000	2.0000000
23	371.1015	C14H27O5S3	1	1.442732e-02	1.5	e	6.290088e-04	Valid	1.6949805	15	1.9285714	0.0000000
26	371.1017	C7H35O2N2S6	1	4.583770e-03	-8.5	e	8.455028e-04	Valid	2.2783636	9	5.0000000	0.2857142
27	371.1000	C5H27O12N2S2	1	2.348783e-03	-6.5	e	-8.912772e-04	Valid	-2.4017114	-3	5.4000000	0.4000000
29	371.1000	C3H15O2N16S2	1	2.191293e-03	4.5	e	-9.019692e-04	Valid	-2.4305230	33	5.0000000	5.3333333
31	371.1018	C6H15O9N10	1	3.007650e-03	4.5	e	9.150288e-04	Valid	2.4657143	19	2.5000000	1.6666666
34	371.0999	C18H11O2N8	1	1.755212e-03	17.5	e	-9.355252e-04	Valid	-2.5209459	43	0.6111111	0.4444444
37	371.0999	C5H138O2N10	1	1.063501e-03	-58.0	e	-1.008076e-03	Valid	-2.7164480	-106	27.6000000	2.0000000
39	371.1019	CH146O6N6S	1	1.459270e-03	-68.0	e	1.025416e-03	Valid	2.7631725	-126	146.0000000	6.0000000

Showing 1 to 17 of 64 entries, 22 total columns

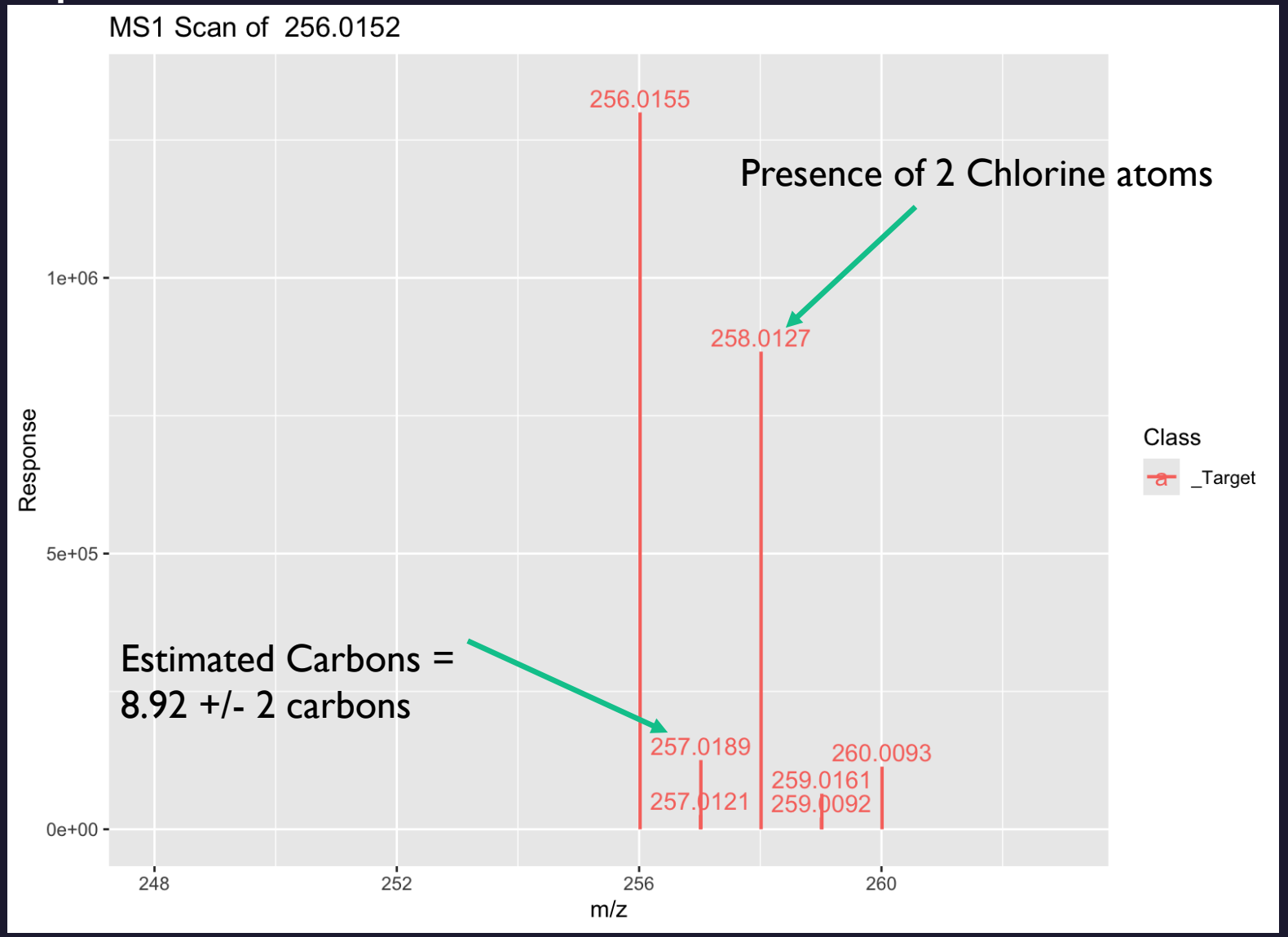
Molecular Formula Elemental Bounds

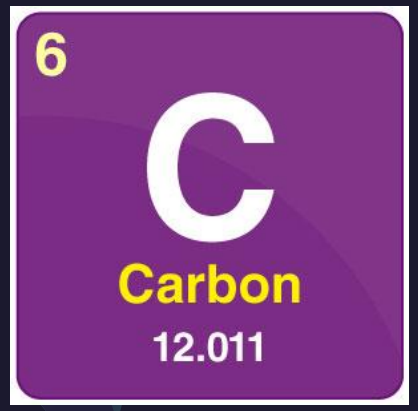
Raw MS1 Spectrum

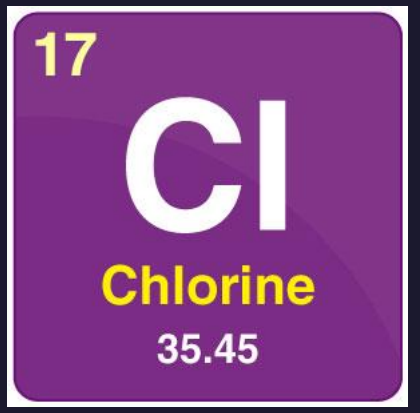


Molecular Formula Elemental Bounds

Filtered MS1 Spectrum

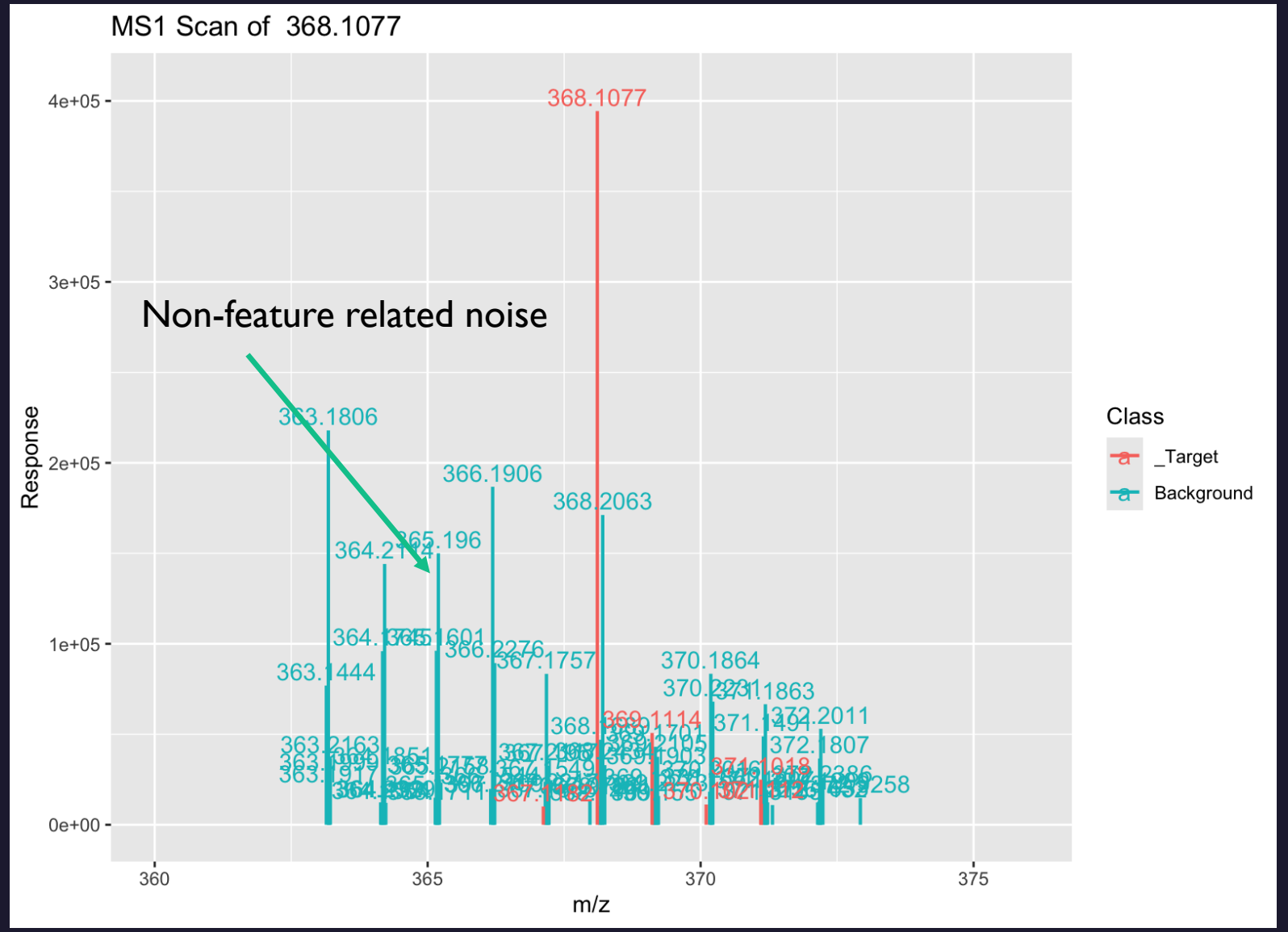






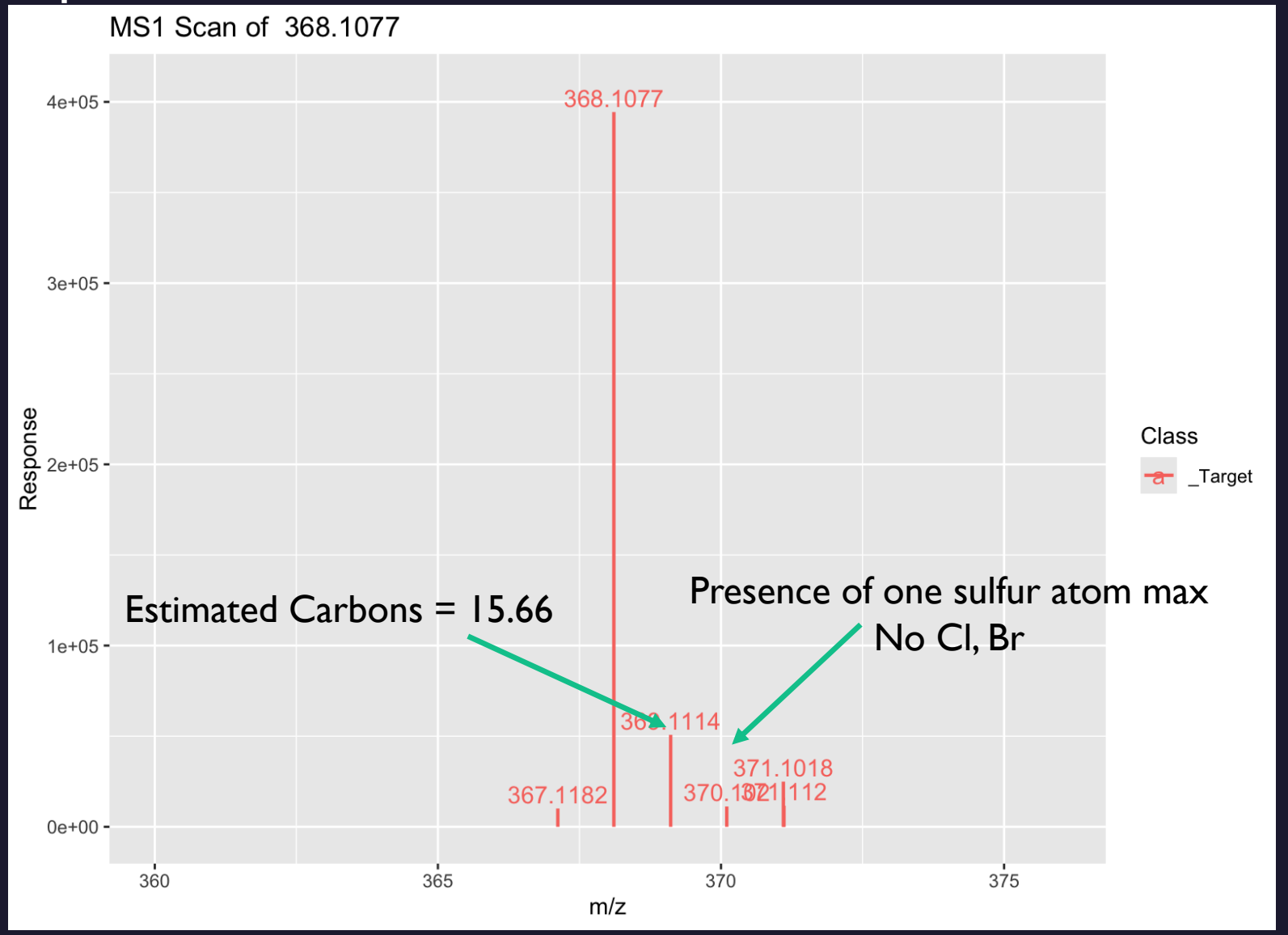
Molecular Formula Elemental Bounds

Raw MS1 Spectrum



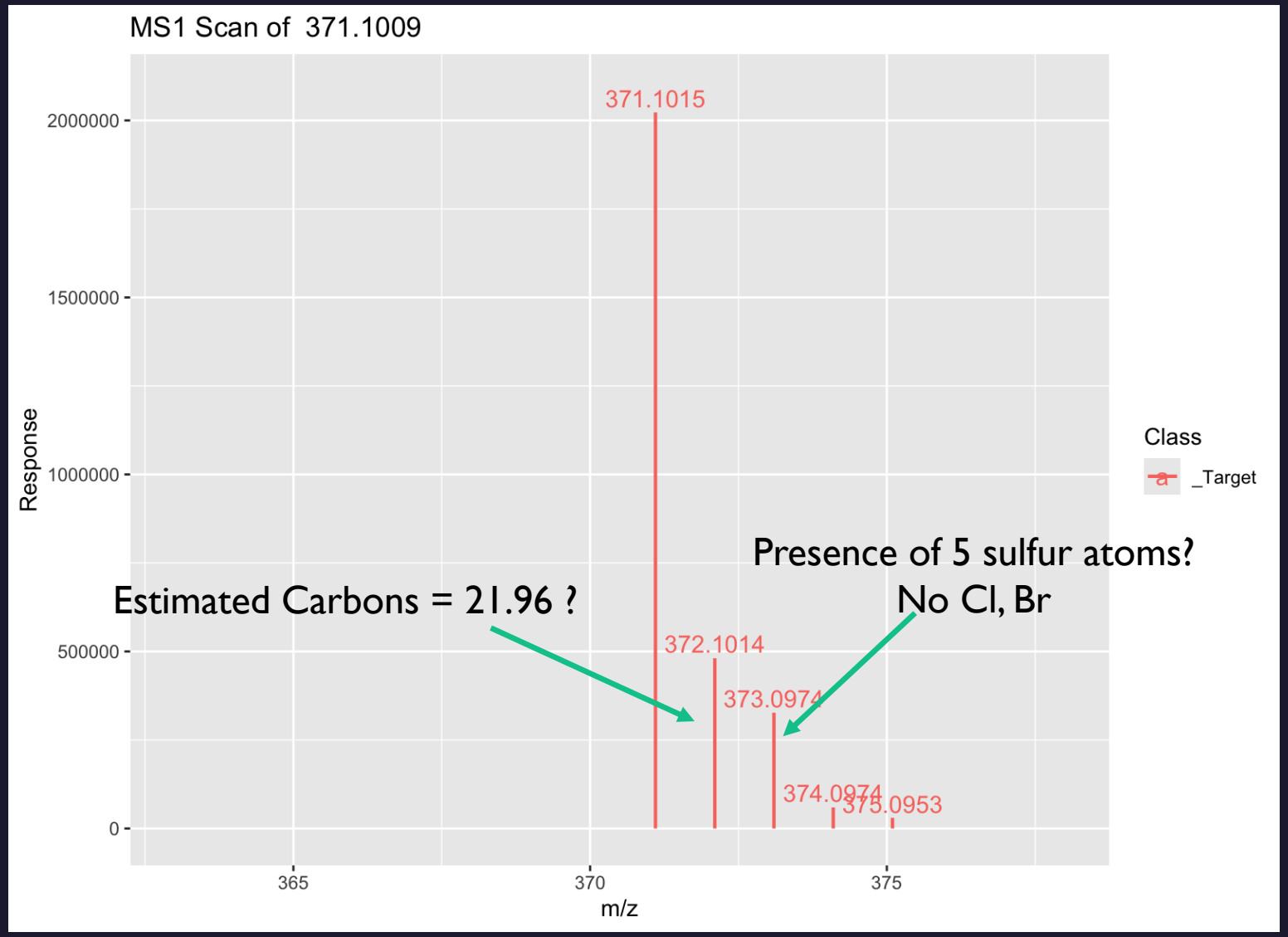
Molecular Formula Elemental Bounds

Filtered MS1 Spectrum



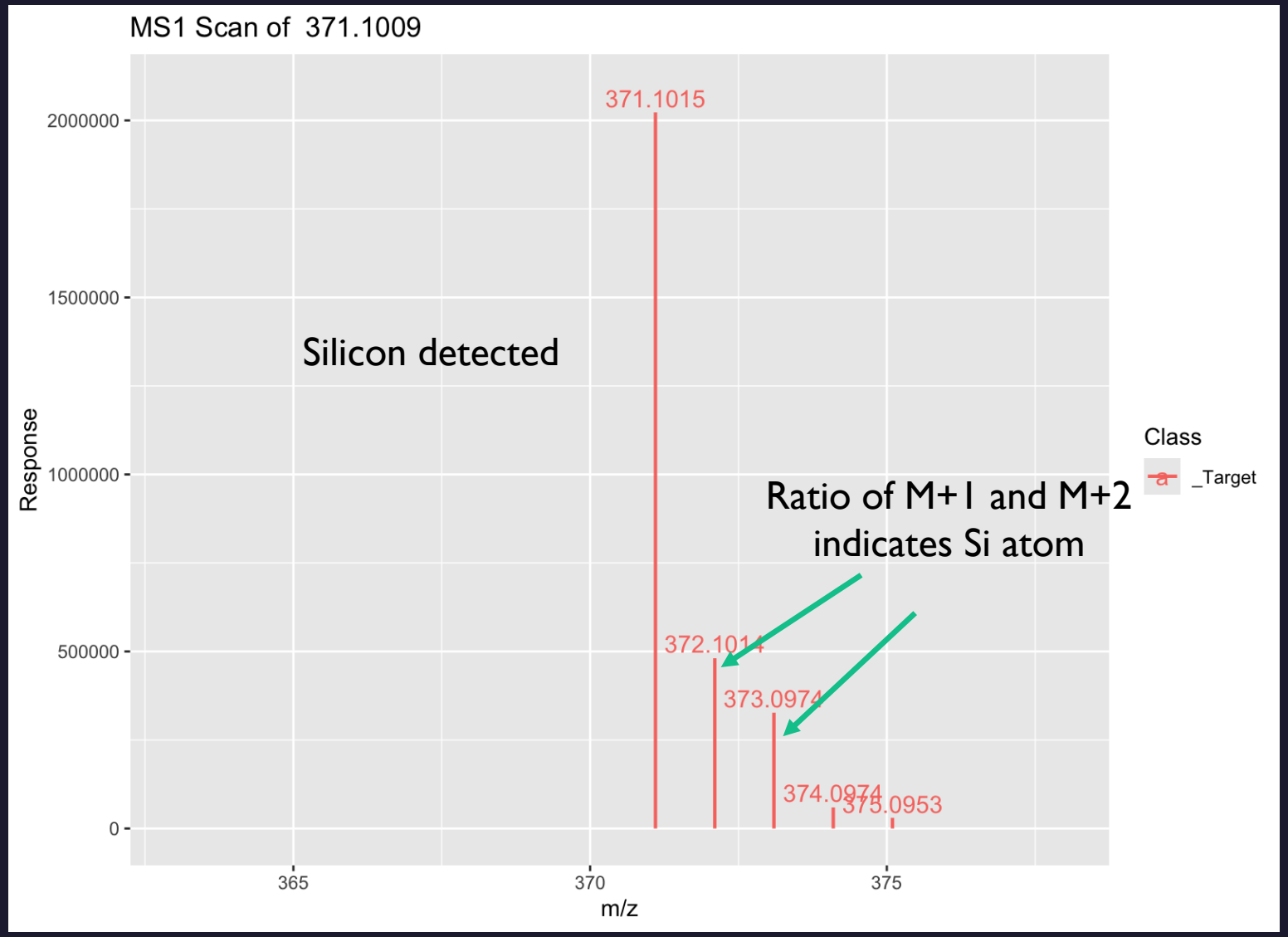
Molecular Formula Elemental Bounds

Filtered MS1 Spectrum



Molecular Formula Elemental Bounds

Filtered MS1 Spectrum



Presence of Silicon throws off M/Carbon ratio and easily identified

Conclusions

- In the absence of MS/MS data, MS1 spectra can be interrogated for more information
- Using High Res Mass Spec intrinsic values like m/z, mass defect, and [13]C ratios can be used to calculate "PFAS-ness" and can be prioritized.
- Retention time prediction models provide an orthogonal technique to confirm or reject potential features
- Plotting data with intrinsically determined values is very useful for PFAS prioritization (Kendrick or Kaufmann)
- Informed Molecular Formula predictions that properly decompose MS1 spectra with elemental bounds more effective than agnostic ones

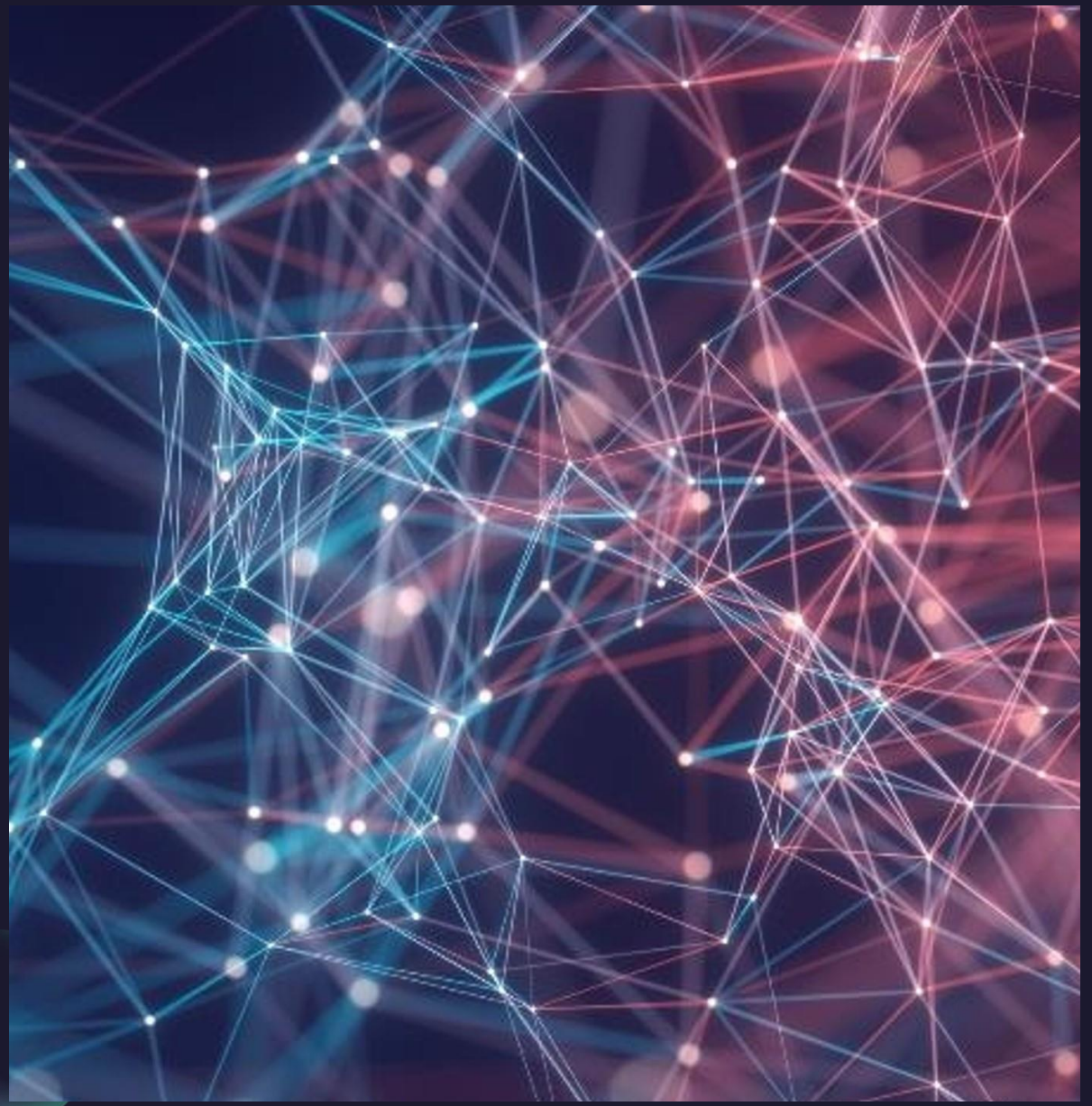
Thank you

David Schiessel

951 289 5278

dschiessel@babcocklabs.com

www.babcocklabs.com



References

- Bugsel, B.; Zwiener, C. LC-MS Screening of Poly- and Perfluoroalkyl Substances in Contaminated Soil by Kendrick Mass Analysis. *Anal Bioanal Chem* **2020**, *412* (20), 4797–4805. https://doi.org/10.1007/s00216-019-02358-0.

- Kaufmann, A.; Butcher, P.; Maden, K.; Walker, S.; Widmer, M. Simplifying Nontargeted Analysis of PFAS in Complex Food Matrixes. *Journal of AOAC INTERNATIONAL* **2022**, *105* (5), 1280–1287. https://doi.org/10.1093/jaoacint/qsac071.

- Zweigle, J.; Bugsel, B.; Zwiener, C. Efficient PFAS Prioritization in Non-Target HRMS Data: Systematic Evaluation of the Novel MD/C-m/C Approach. *Anal Bioanal Chem* **2023**, *415* (10), 1791–1801. https://doi.org/10.1007/s00216-023-04601-1.