Unknown-Unknown Analysis: Strategies for Identifying Compounds Not in Libraries Using Single Quadrupole GC/MS

Don Kuehl[1], Stacey Simonoff[1], Yongdong Wang[1]

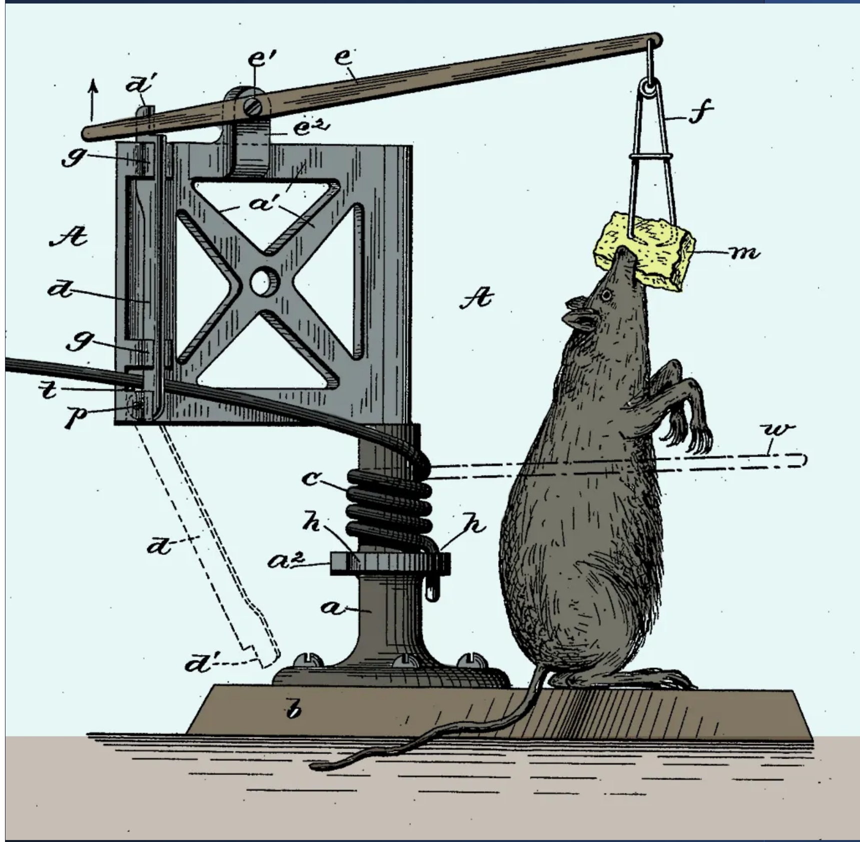
[1]Cerno Bioscience
Las Vegas, NV USA

Unknown-Unknowns

- Unknown-Unknowns are compounds in a GC/MS run that are not in the GC/MS library
  - Compound ID is important, but classic approaches to identification are tedious
    - Run high resolution MS to obtain possible molecular formula
      - May need to run ‘soft” ionization
    - Run MS/MS to help deduce the structure
    - Isolate the compound in quantity for analysis by other techniques (NMR, IR…)
    - Purchase/synthesize probable compound and run GC/MS and match retention time and MS

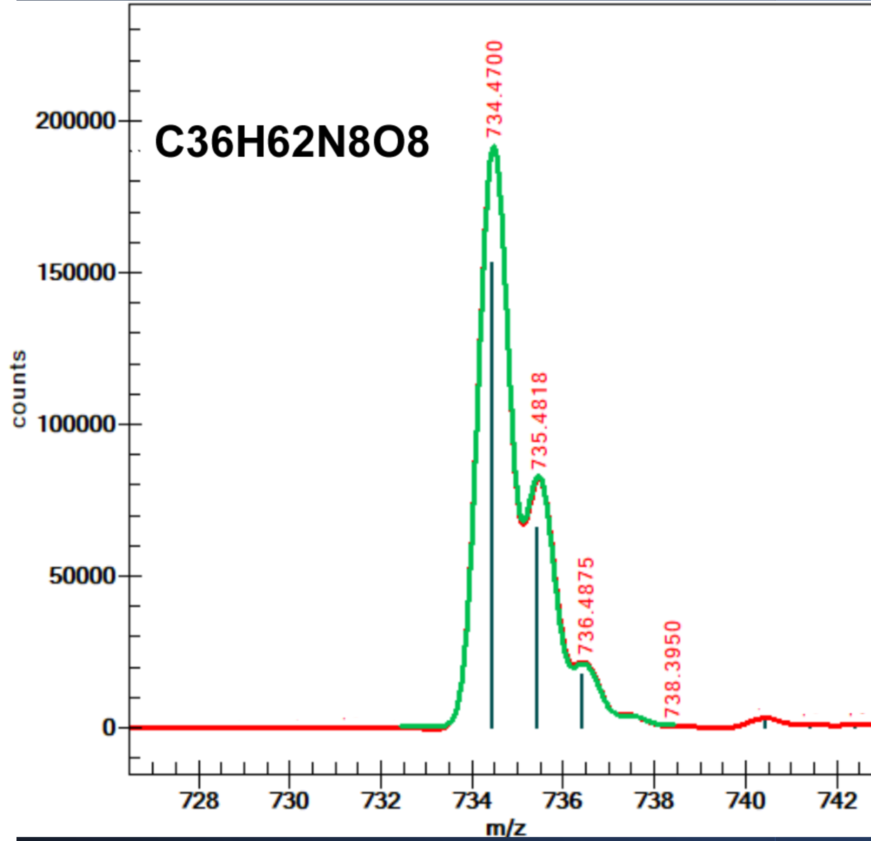
Is There a Faster, More Economical Approach?

- What methods are available from a single quad GC/MS run for performing U-U ID?
  - Obtain the molecular formula by leveraging calibrated MS data to obtain accurate mass/spectral accuracy
  - Perform NIST "Hybrid Search" which can search compounds not in the library, but are related
  - Confirm or filter by retention index of the U-U with AI predicted RI (NIST AIRI)



Formula ID on a Quadrupole

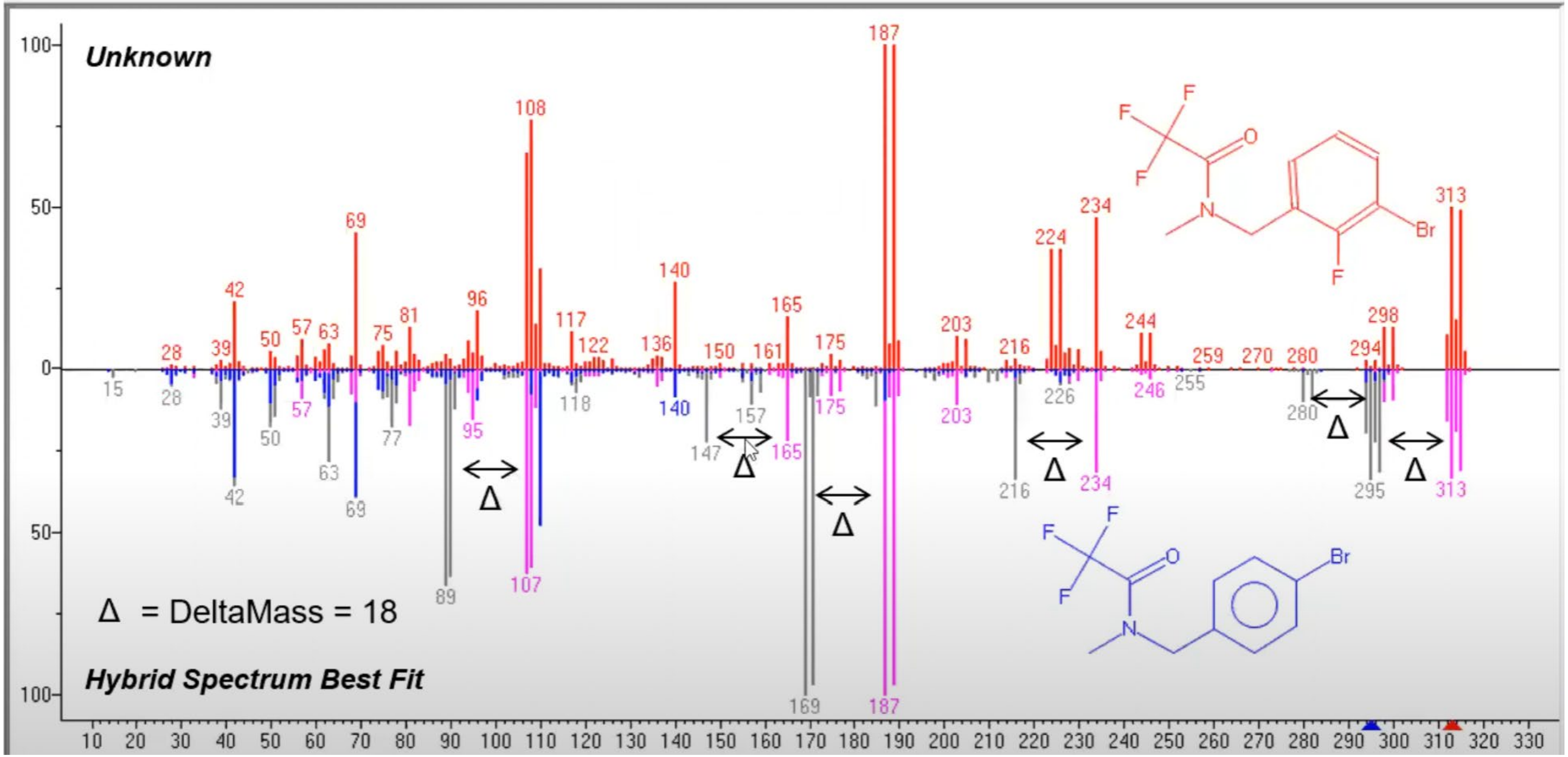
- Calibrated profile mode data provides good mass accuracy (~+/- 10mDa) which when combined with spectral accuracy provides effective mass accuracy of a few ppm.
- Simple to obtain
  - Run data in profile mode
  - Turn on cal gas (PFTBA) briefly at end of run
  - Software automatically calibrates data



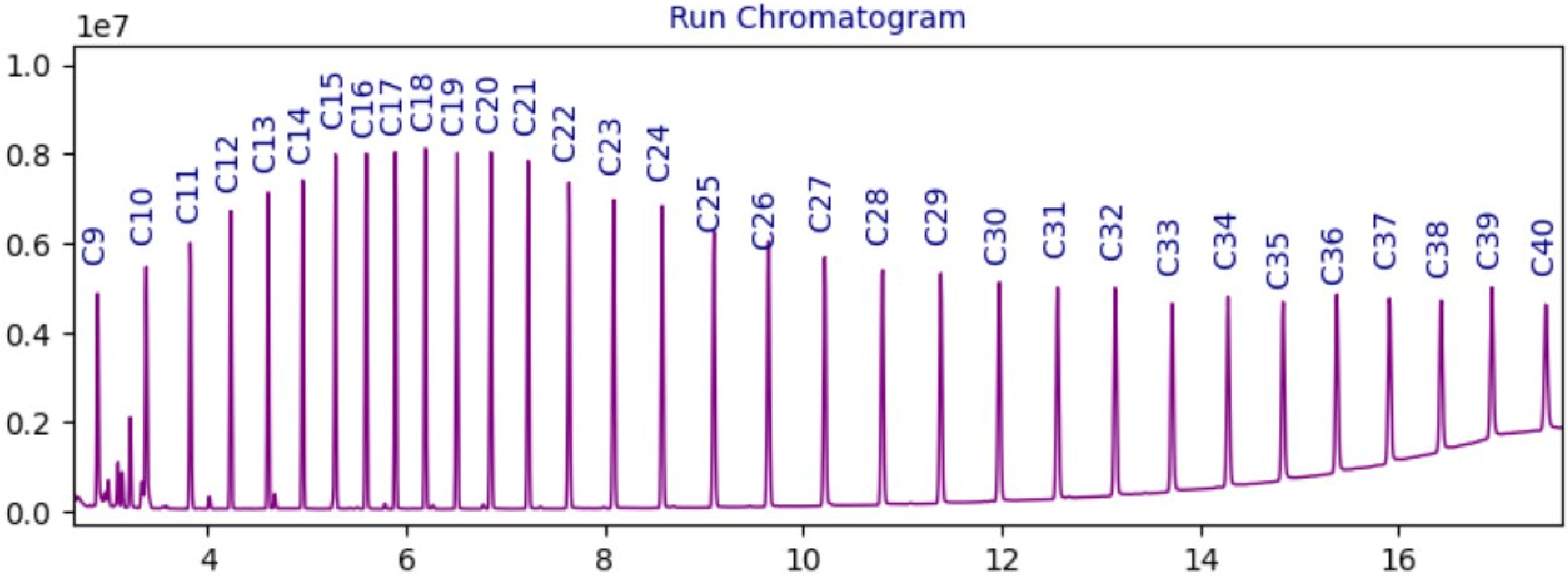
Hybrid Search

- Hybrid search matches fragments and neutral losses
- Extends the scope of the library by including “nearest neighbor” compounds
- Requires the presence of similar compounds in library
- You “must” know the nominal MW, and preferably the molecular formula
- Mass Difference MUST be confined to a single region of molecule with no significant alteration of fragmentation behavior
- DeltaMass represents to difference (modification) between library and the unknown molecule

Hybrid Search Display

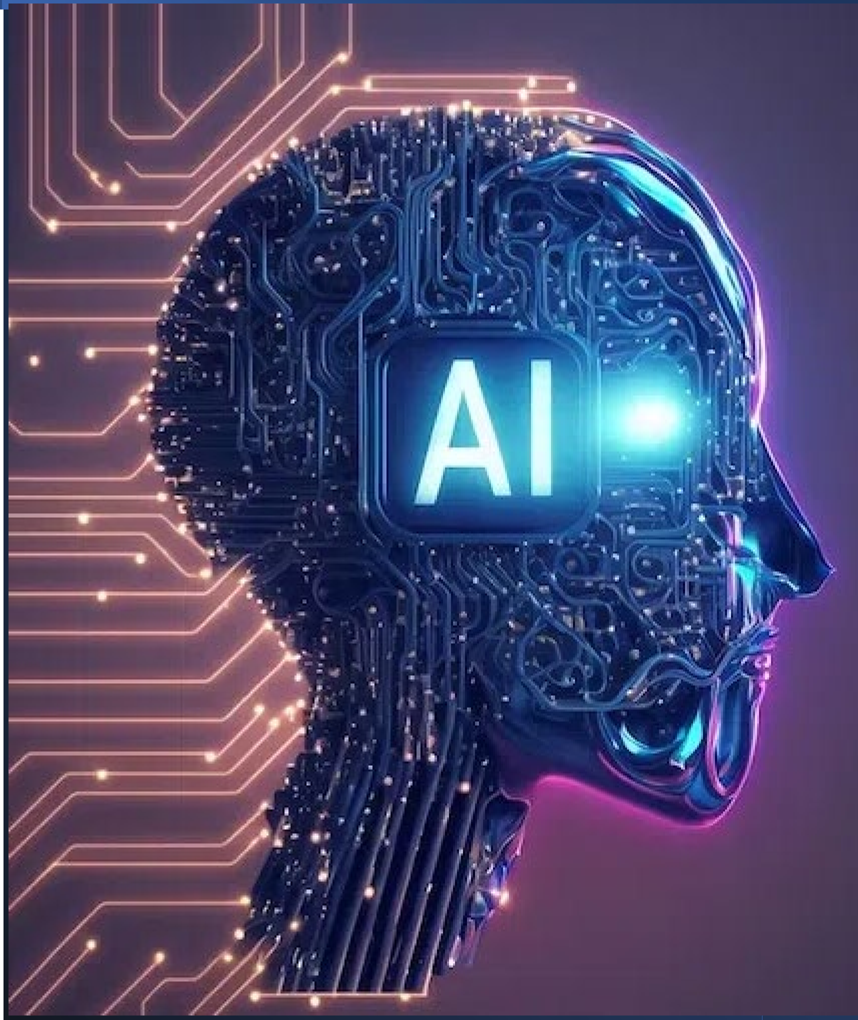


Calibrate for Retention Index



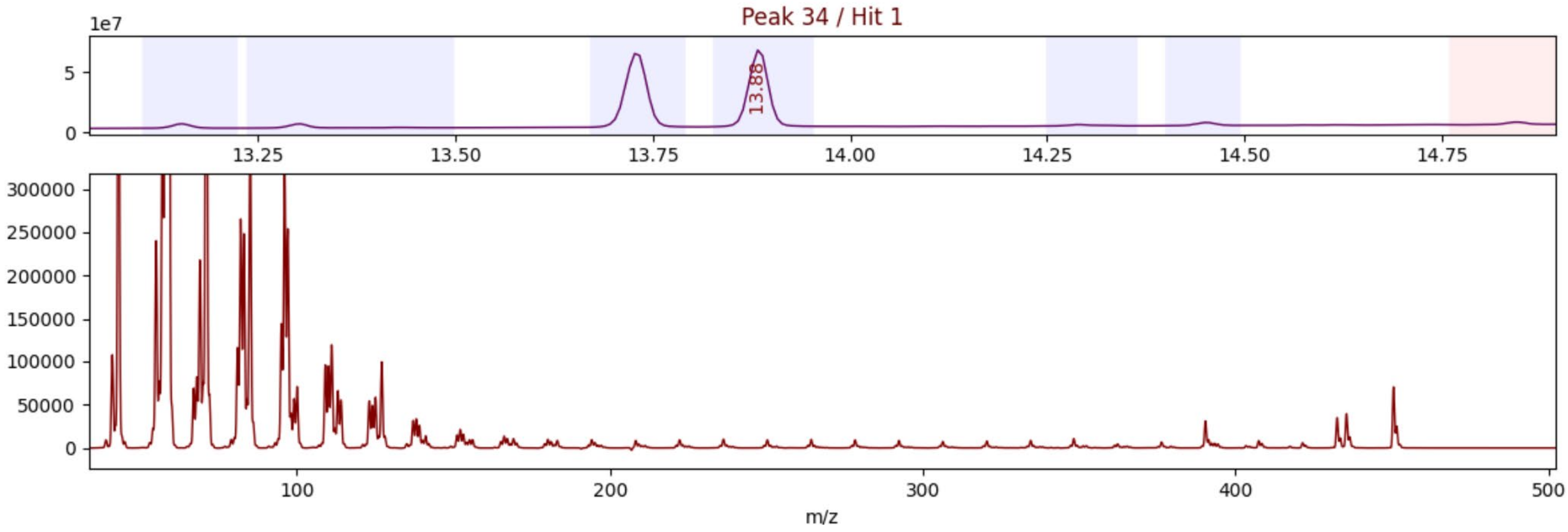
What is AIRI?

- NIST used there extensive database measured RI data as a training set for predicting RI of any compound by structure
- “AIRI” source code and models are published by NIST. But, it is tedious to get set up.
- Cerno is creating web-based app to predict RI from structures to make AIRI easily available to all



(1) AIRI: Predicting Retention Indices and Their Uncertainties Using Artificial Intelligence, Geer, Stein, Mallard, and Slotta, Journal of Chemical Information and Modeling 2024 64 (3), 690-696.

Unknown Peak at 13.88 is Strong and Pure



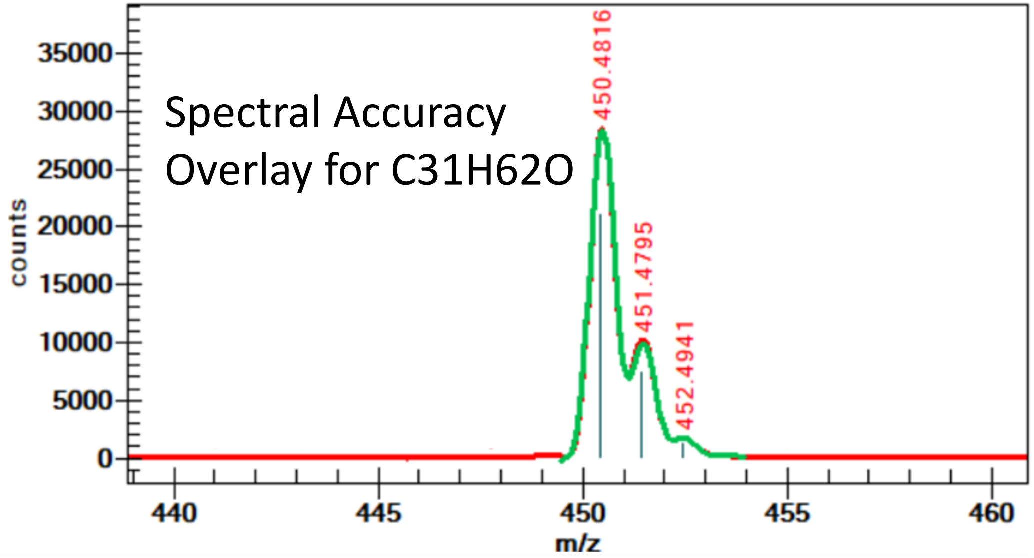
Despite that, the Search Results are Very Poor

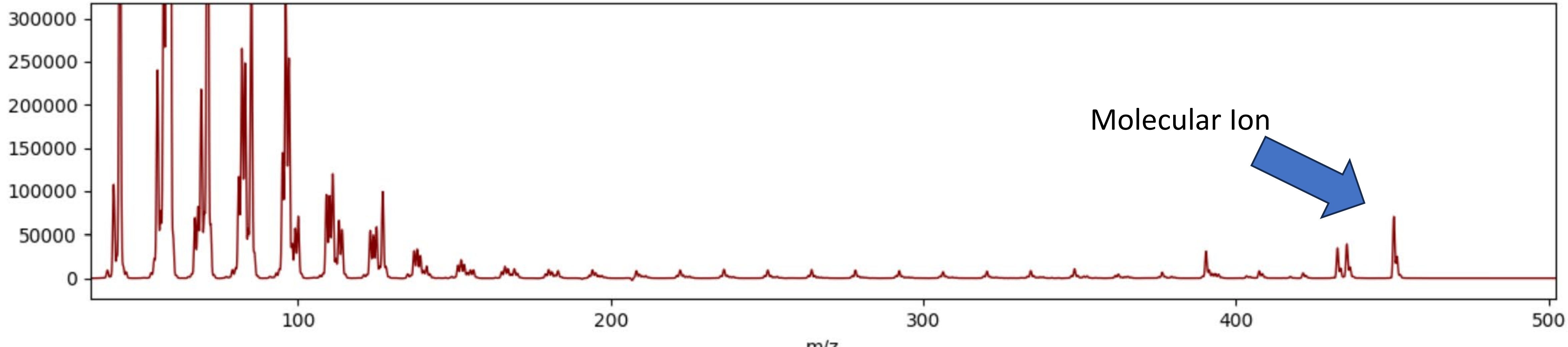
Hit	FwdS	RevS	SA-M	RI-Fit	RI	RI-Lib	MW	CAS	Formula	Name
1:	763	839	74.23	0	3329.8	2106	282.292	629-66-3	C19H38O	2-Nonadecanone
2:	739	756	66.4	0	3329.8	3531	478.511	75207-55-5	C33H66O	2-Tritriacontanone
3:	727	792	96.47	0	3329.8	1804	240.245	18787-63-8	C16H32O	2-Hexadecanone
4:	721	786	87.29	0	3329.8	1904	254.261	2922-51-2	C17H34O	2-Heptadecanone
5:	719	803	74.23	0	3329.8	2106	282.292	629-66-3	C19H38O	2-Nonadecanone
6:	717	786	97.59	0	3329.8	1698	226.23	2345-28-0	C15H30O	2-Pentadecanone
7:	712	733	15.84	0	3329.8	3725	506.543	77327-17-4	C35H70O	2-Pentatriacontanone
8:	707	793	96.79	0	3329.8	1597	212.214	2345-27-9	C14H28O	2-Tetradecanone
9:	706	767	96.47	0	3329.8	1804	240.245	18787-63-8	C16H32O	2-Hexadecanone
10:	701	817	96.47	0	3329.8	1804	240.245	18787-63-8	C16H32O	2-Hexadecanone
11:	698	769	92.28	0	3329.8	2715	366.386	75207-54-4	C25H50O	2-Pentacosanone
12:	695	751	74.23	0	3329.8	2106	282.292	629-66-3	C19H38O	2-Nonadecanone
13:	693	752	87.29	0	3329.8	1904	254.261	2922-51-2	C17H34O	2-Heptadecanone
14:	692	741	97.44	0	3329.8	3126	422.449	17600-99-6	C29H58O	2-Nonacosanone
15:	679	737	87.29	0	3329.8	1904	254.261	2922-51-2	C17H34O	2-Heptadecanone
16:	675	735	97.59	0	3329.8	1698	226.23	2345-28-0	C15H30O	2-Pentadecanone
17:	674	723	92.28	0	3329.8	2715	366.386	75207-54-4	C25H50O	2-Pentacosanone
18:	673	802	97.59	0	3329.8	1698	226.23	2345-28-0	C15H30O	2-Pentadecanone
19:	673	729	63.1	0	3329.8	2004	268.277	7373-13-9	C18H36O	Methyl n-hexadecyl ketone
20:	672	742	94.56	0	3329.8	1496	198.198	593-08-8	C13H26O	2-Tridecanone

We can Determine the MW formula Using our Calibrated Single Quad Data

Formula Search Results

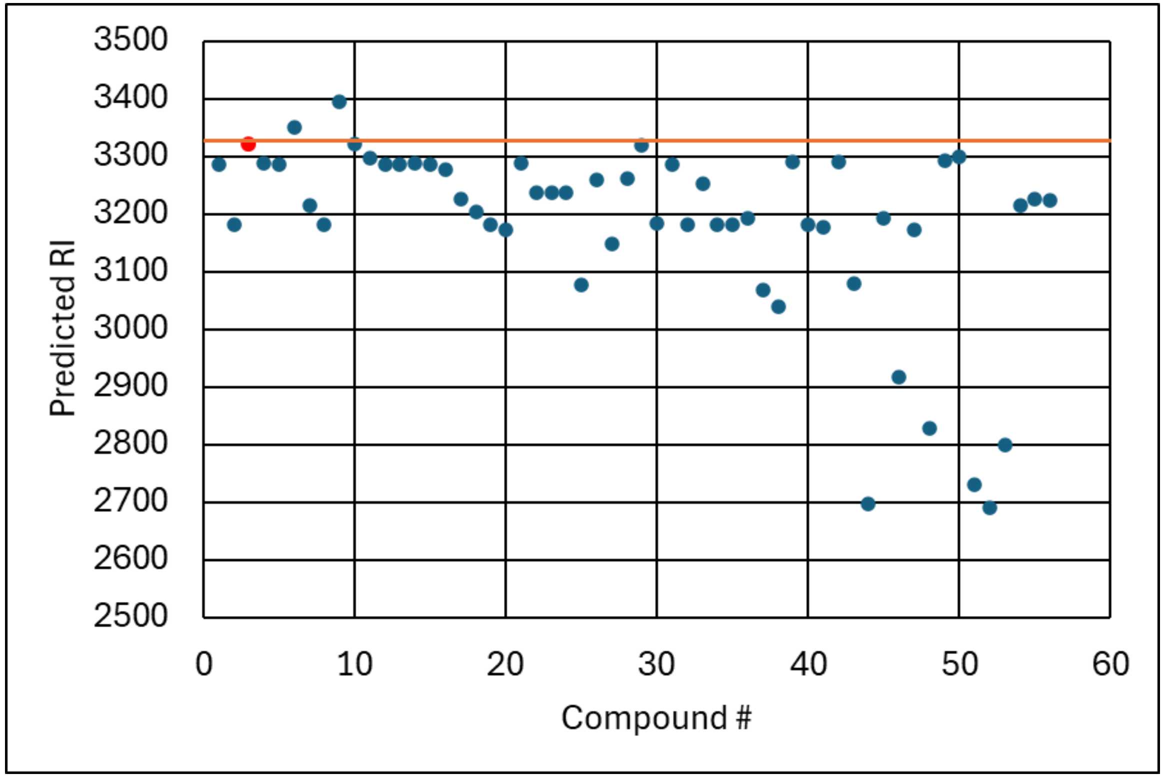
Row	Formula	Mono Isoto	Mass Error (mDa)	Mass Error (PPM)	Spectral Accuracy
1	C31H62O	450.4795	2.082	4.6216	98.0607
2	C30H62N2	450.4908	-9.1514	-20.3147	97.8866
3	C30H60ON	450.4669	14.658	32.5386	97.6968
4	C29H60N3	450.4782	3.4246	7.6022	97.4928
5	C28H58N4	450.4656	16.0007	35.5191	97.0491



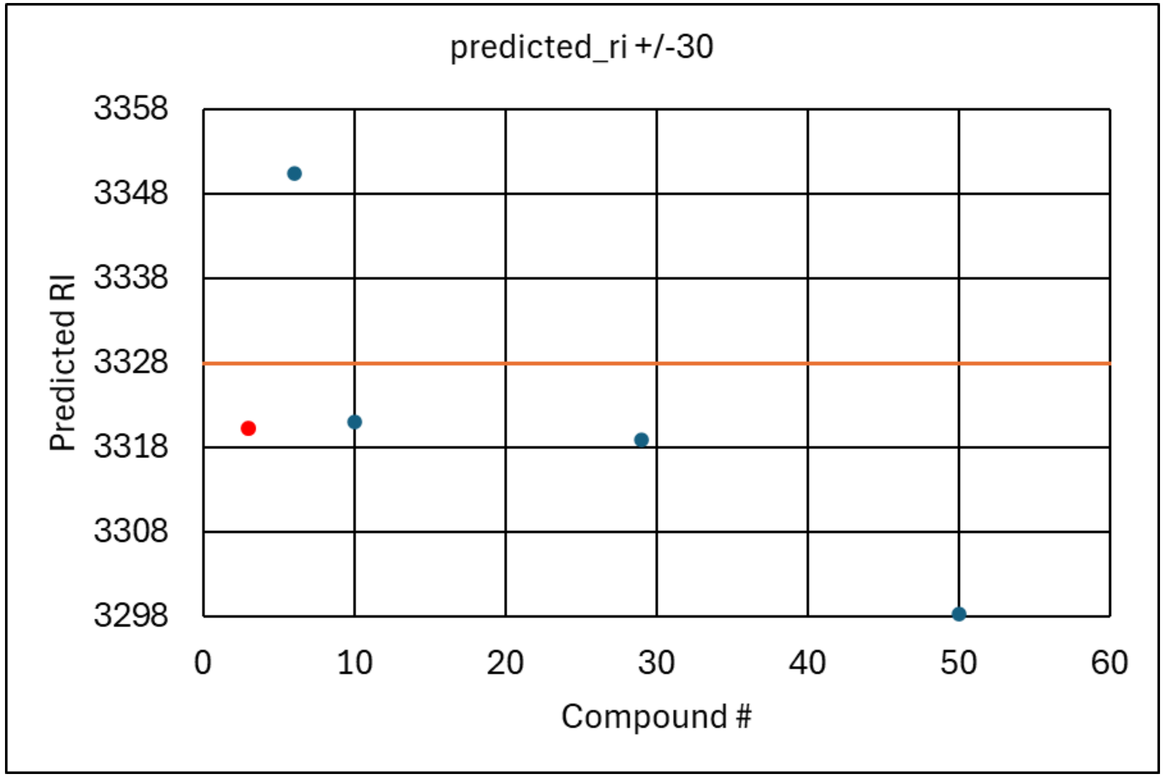


AI Predicted Retention Index

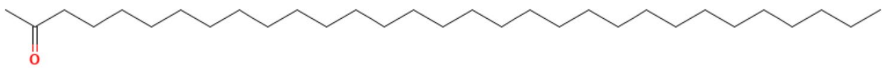
57 compounds C31H62O from PubChem Search



Filter all compounds to +/- 30 iu returns 4



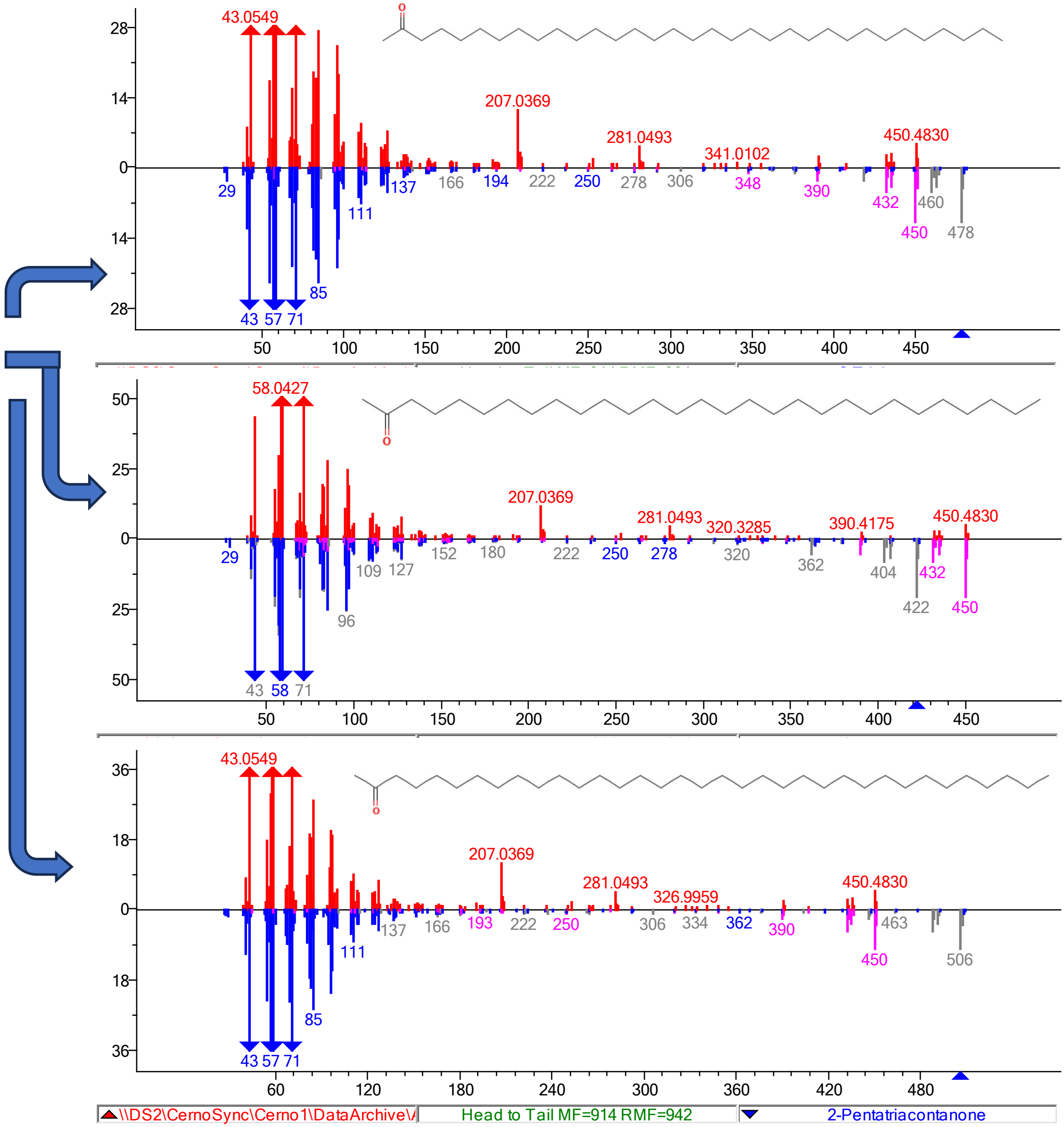
Hybrid Search



#	Lib.	Match	R.Match	RI	PSS.Match	DeltaMass	Name	MF	MW
1	M	914	964	3531	930	-28	2-Tritriacontanone	$C_{33}H_{66}O$	478
2	M	914	956	3126	935	28	2-Nonacosanone	$C_{29}H_{58}O$	422
3	M	914	942	3725	941	-56	2-Pentatriacontanone	$C_{35}H_{70}O$	506
4	M	903	925	2715	937	84	2-Pentacosanone		
5	M	899	930	2513	917	112	2-Tricosanone		
6	M	869	900	1698	896	224	2-Pentadecanone		
7	R	868	879	2106	892	168	2-Nonadecanone		
8	R	866	875	1904	889	196	2-Heptadecanone		
9	R	856	900	1804	880	210	2-Hexadecanone		
10	M	838	865	2004	864	182	Methyl n-hexadecyl ketone		
11	M	827	878	2306	897	140	2-Heneicosanone		

Compound	Formula	Exact Mass	Measured M/Z	PPM	SA%
2-Hentriacontanone	$C_{31}H_{62}O$	450.4795	450.4816	4.7	98.06

Compound	Measured	AIRI	Delta RI
2-Hentriacontanone	3338	3320	18



Summary

- Combining Formula ID, RI, Hybrid search is a simple, rapid approach to analyzing U-U in a single GC/MS run on a benchtop single quad
- For cases where hybrid search is not suitable, we can still filter the potential U-U compound candidates using RI+Formula ID
- Fragment analysis can also be used to validate ID (e.g. NIST MS Interpreter)
- Compound synthesis is still the gold standard