Automating GC Retention Index Calibration to Enable more Confident GC/MS Search

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

[1]Cerno Bioscience
Las Vegas, NV USA

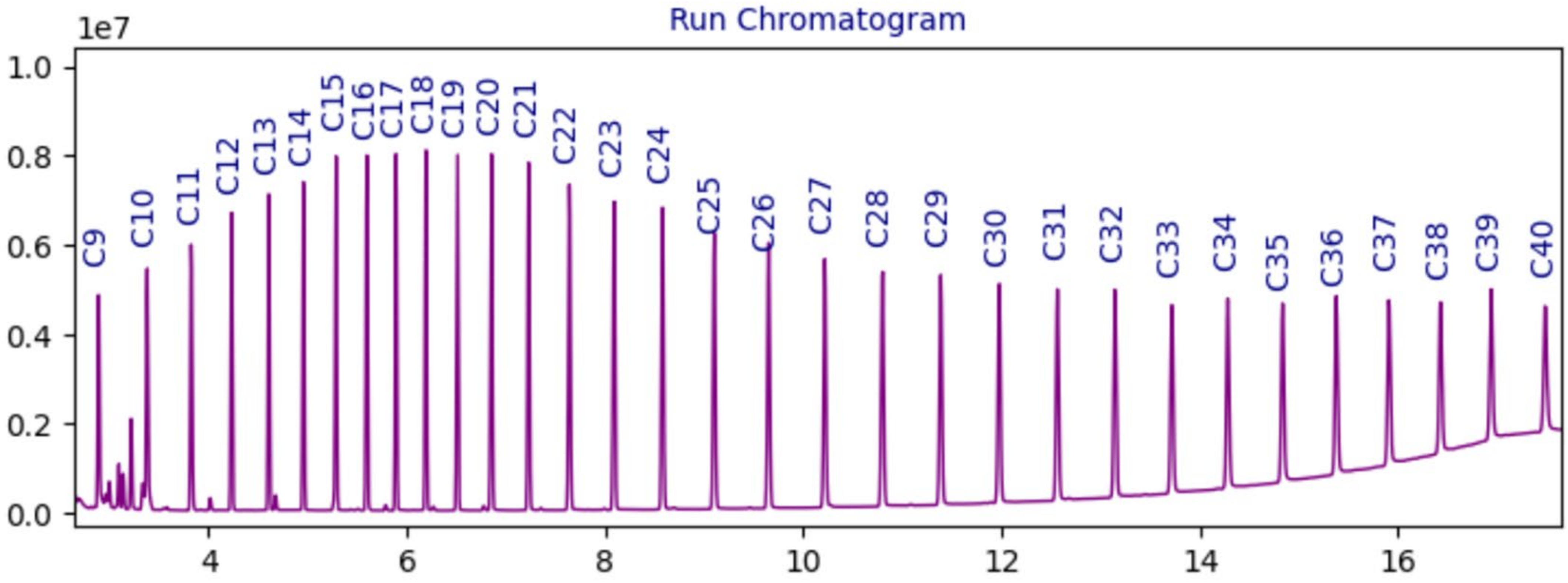
The Importance or Retention Index

- Retention Index (RI) matching is likely the most important complementary metric for identifying unknown compounds with GC/MS library search
- **Up to 30% of compounds in a run can be miss-identified without RI matching!**
- Combining search and RI can dramatically improve the confidence of unknown compound identification
- New NIST databases (NIST23) now contain measured and AI generated RI values for virtually every compound
  - **But few software products take full advantage of this information!**

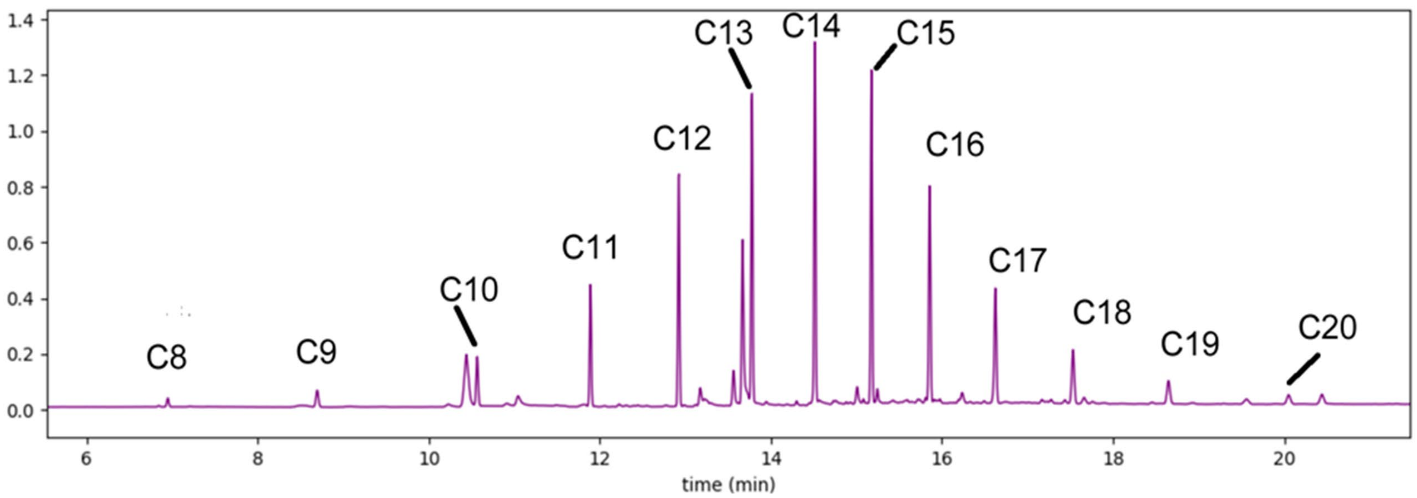
Calibrating Your GC for RI

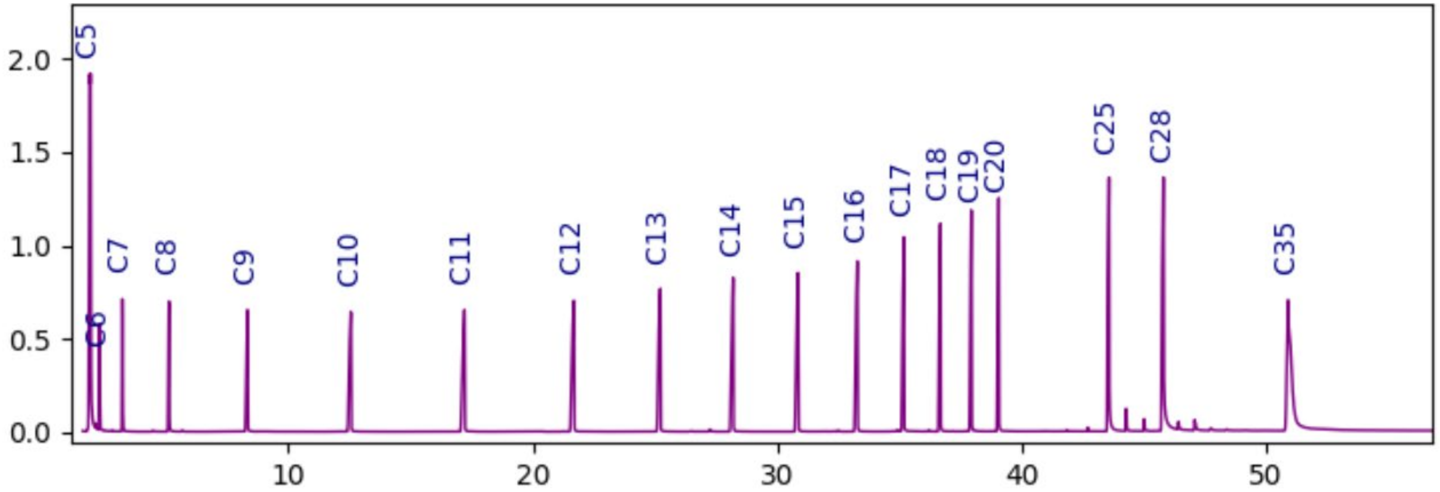
- Of course, one must calibrate the GC to take advantage of published RI values
- This is commonly done using a series of n-alkanes (n-alkane ladder)
- However, assigning the n-alkane peaks in a calibration run can be a complex, manual procedure due to the following:
    - Contaminate peak interferences, in particular, when using SPME and LVI
    - Carry over and old "dirty" columns
    - Impurities in n-alkane standards, typically branched alkanes
    - Solvent peaks and solvent impurities
    - User errors inputting incorrect n-alkane standards (e.g. standards without consecutive n-alkanes)
    - "Ghost" peaks
- The biggest barrier to using RI is the Calibration!

We Like to See RI Calibrations Like this:



In Reality, many Things Can Go Wrong, Some Examples



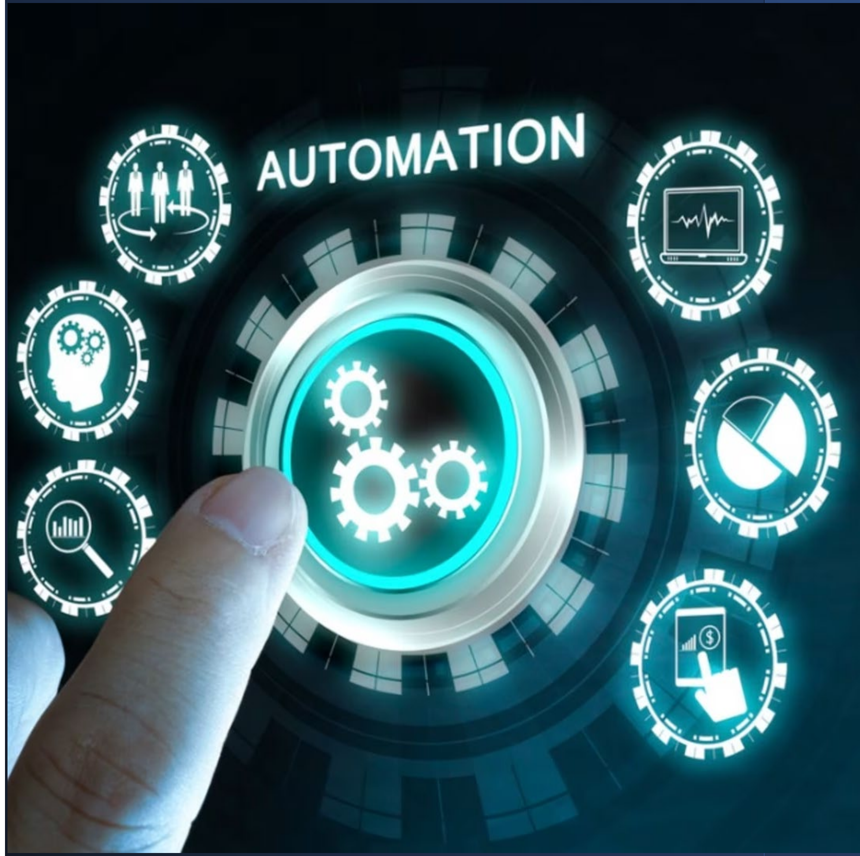


Some typical “issues” with RI calibrations

- Top: Contaminant peaks present in SPME run, many which are much larger than calibrants, very small calibrant peaks
- Bottom: Standard is not contiguous, potential for user input error

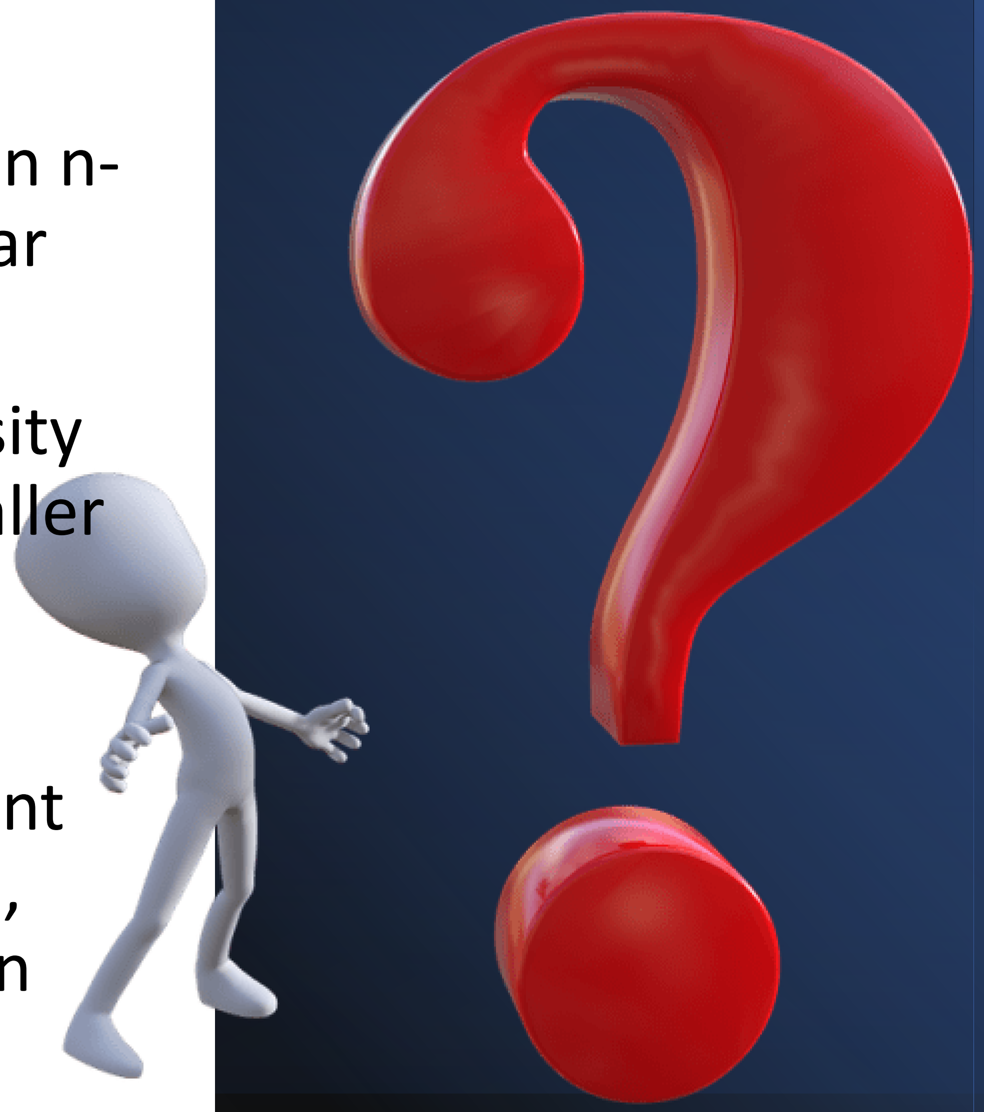
How to Easily and Automatically Locate Standards

- Find the N largest peaks where N=#calibrants and assign in order of retention time?
  - We just showed that won't work!
- Let's use library search to identify the n-alkanes and assign them!
  - Nope, that does not work. Library search algorithms are incapable of properly identifying heavier n-alkanes (~n=20-30 and greater depending on S/N)
  - What! **Search cannot properly identify compounds at large concentrations**????
    - For n-alkanes we cannot…

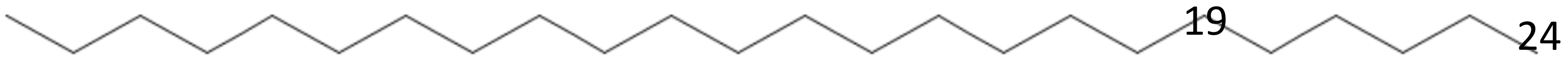


Why?

- n-alkanes are a linear, saturated hydrocarbon chain so the fragments produced for any given n-alkane have identical M/Z, up to the molecular ion.
- As the molecular weight increases, the intensity of the heavier fragments get smaller and smaller
- As the heavy fragments get smaller, the search algorithm weights them less and less
- At some point, other variations in the fragment intensities “masks” the influence of the small, heavy fragments and the search results return an incorrect identification



For Example, MS of C19 and C24 with Search Results

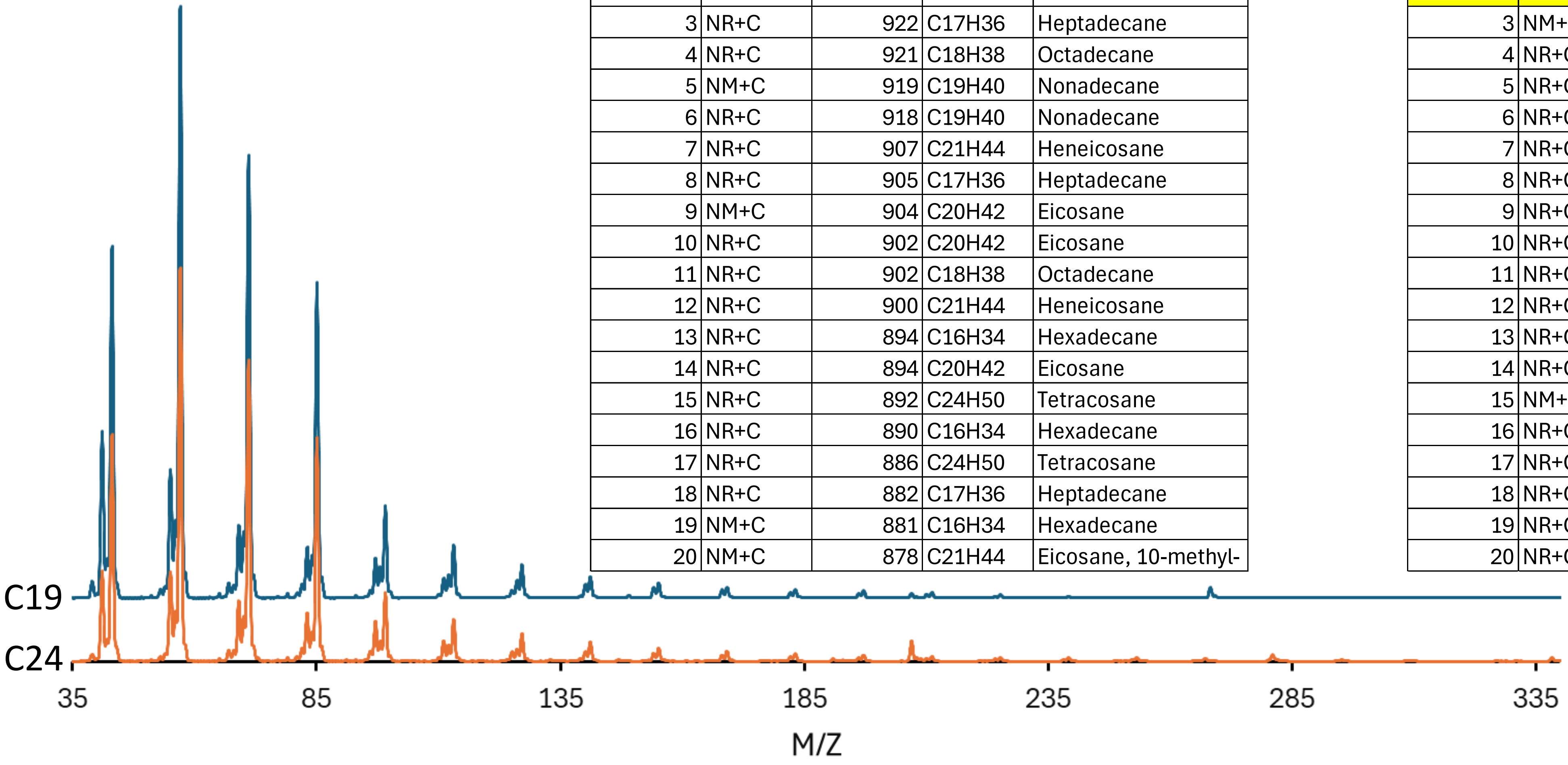


C19 MW=268

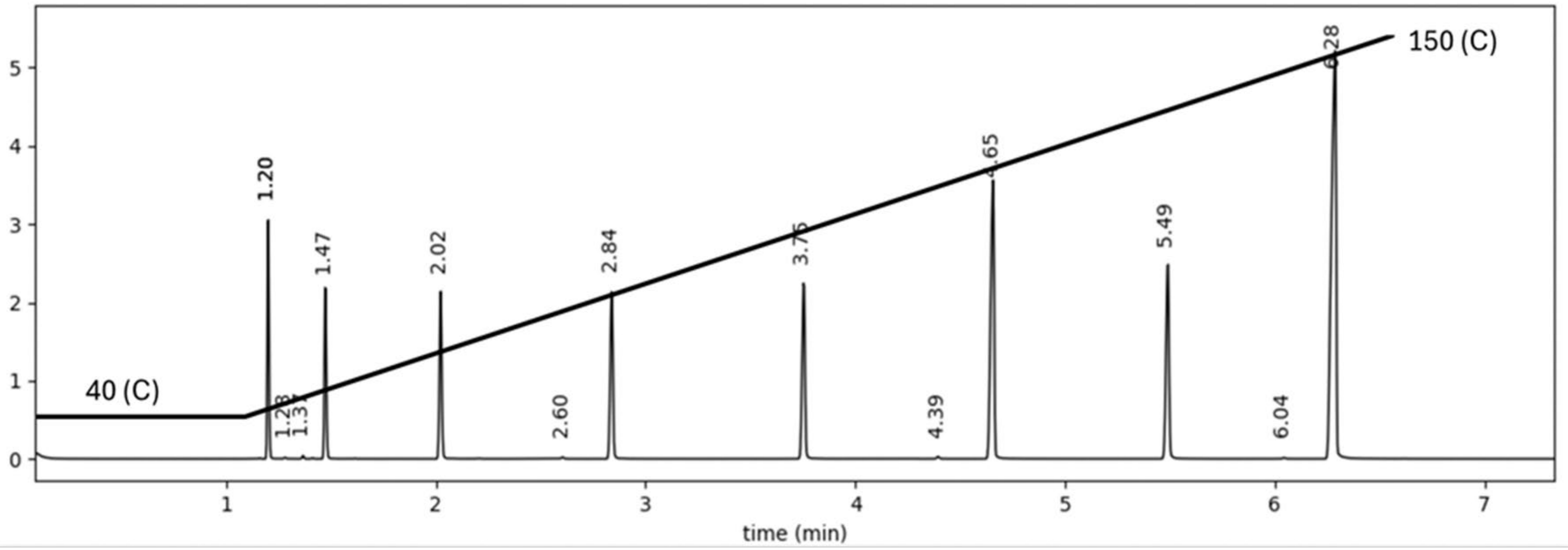
Hit	Lib	FwdS	Formula	Name
1	NR+C	933	C19H40	Nonadecane
2	NR+C	923	C19H40	Nonadecane
3	NR+C	922	C17H36	Heptadecane
4	NR+C	921	C18H38	Octadecane
5	NM+C	919	C19H40	Nonadecane
6	NR+C	918	C19H40	Nonadecane
7	NR+C	907	C21H44	Heneicosane
8	NR+C	905	C17H36	Heptadecane
9	NM+C	904	C20H42	Eicosane
10	NR+C	902	C20H42	Eicosane
11	NR+C	902	C18H38	Octadecane
12	NR+C	900	C21H44	Heneicosane
13	NR+C	894	C16H34	Hexadecane
14	NR+C	894	C20H42	Eicosane
15	NR+C	892	C24H50	Tetracosane
16	NR+C	890	C16H34	Hexadecane
17	NR+C	886	C24H50	Tetracosane
18	NR+C	882	C17H36	Heptadecane
19	NM+C	881	C16H34	Hexadecane
20	NM+C	878	C21H44	Eicosane, 10-methyl-

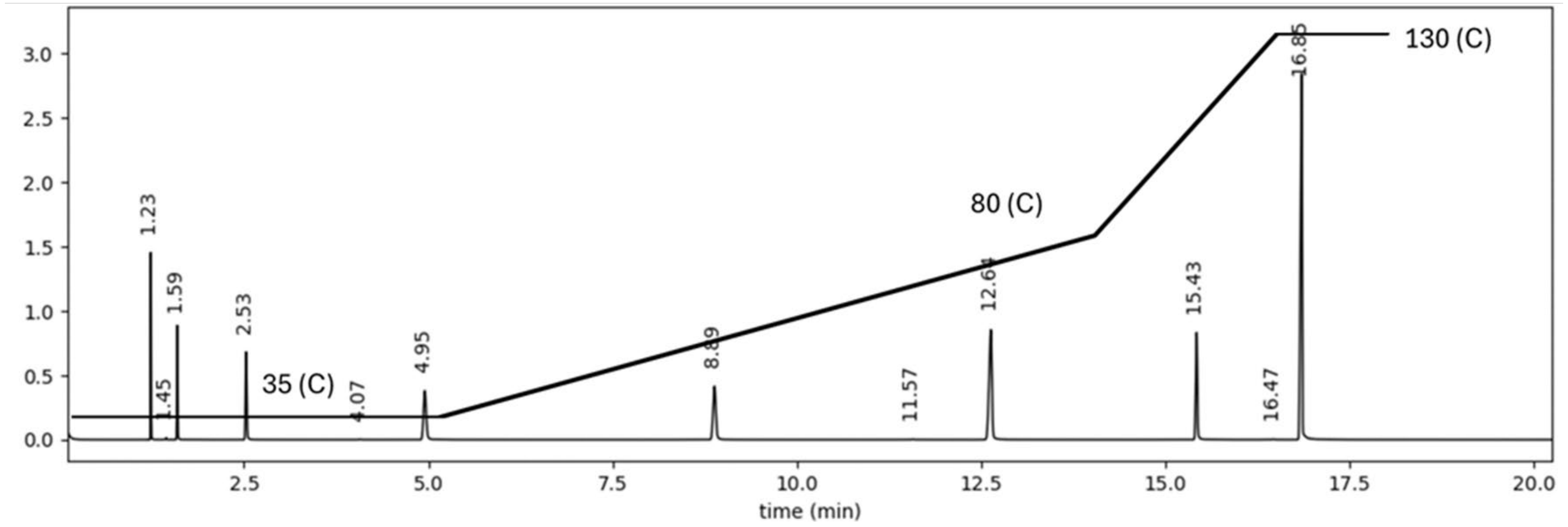
C24 MW=338

Hit	Lib	FwdS	Formula	Name
1	NR+C	940	C21H44	Heneicosane
2	NR+C	927	C24H50	Tetracosane
3	NM+C	924	C20H42	Eicosane
4	NR+C	922	C24H50	Tetracosane
5	NR+C	921	C21H44	Heneicosane
6	NR+C	920	C26H54	Hexacosane
7	NR+C	913	C28H58	Octacosane
8	NR+C	911	C17H36	Heptadecane
9	NR+C	911	C20H42	Eicosane
10	NR+C	910	C27H56	Heptacosane
11	NR+C	909	C20H42	Eicosane
12	NR+C	907	C18H38	Octadecane
13	NR+C	905	C25H52	Pentacosane
14	NR+C	905	C19H40	Nonadecane
15	NM+C	903	C19H40	Nonadecane
16	NR+C	903	C21H44	Heneicosane
17	NR+C	899	C27H56	Heptacosane
18	NR+C	896	C19H40	Nonadecane
19	NR+C	895	C22H46	Docosane
20	NR+C	895	C23H48	Tricosane



Can we take Advantage of Elution Times?





- Elution times are highly dependent on temperature programming profile
- But, if we can identify the first 3 n-alkanes, we can use velocity and acceleration to predict the location of the next n-alkane in the series and include it for predicting the next n-alkane, etc.

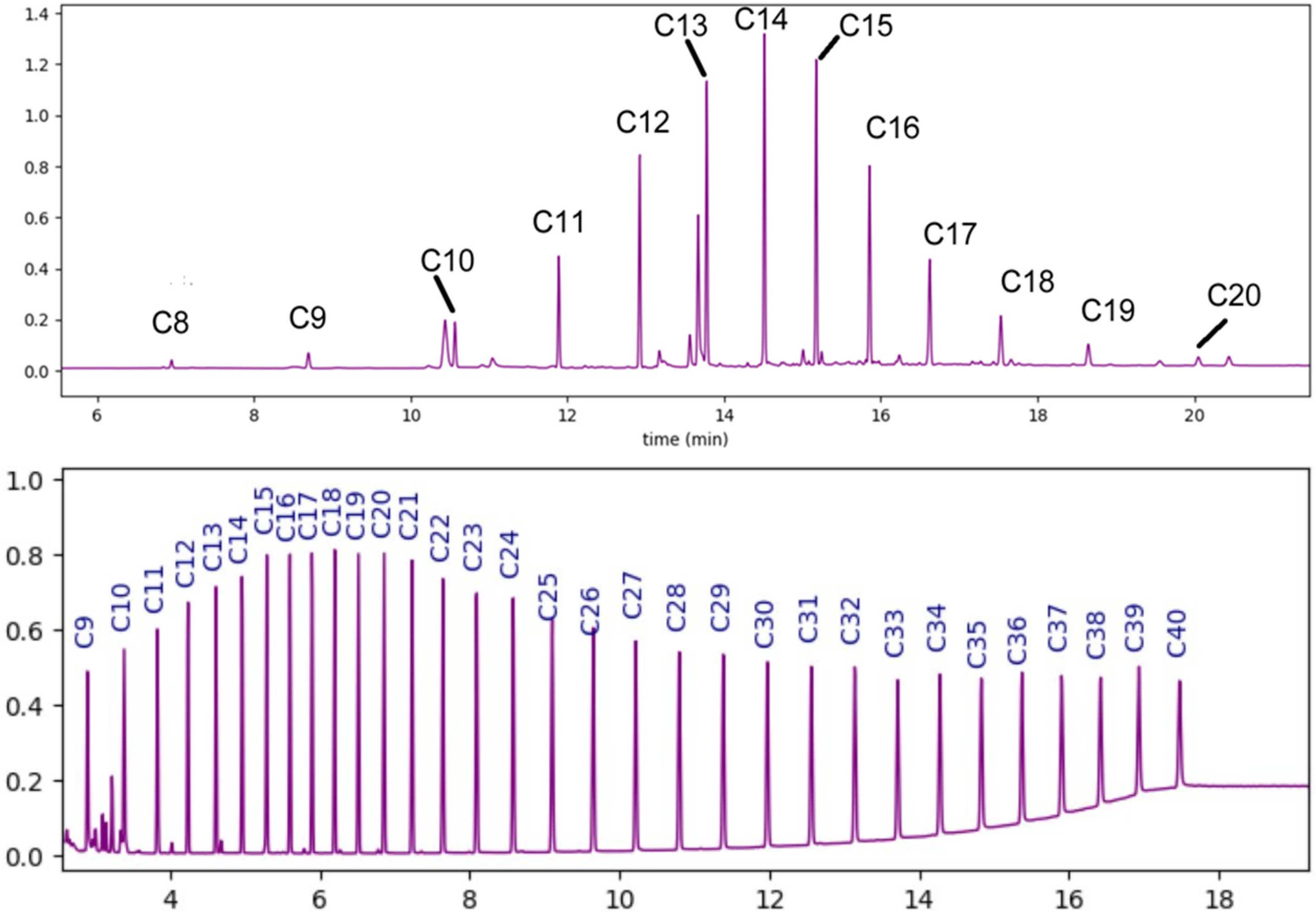
Putting it All Together

- Reject all peaks that are not in the n-alkane family
- The light n-alkanes can be reliably identified by search (up to ~C15)
- We can use these peaks to <u>project the elution time of the next n-alkane</u> above ~C15
- If the next peak identifies as an n-alkane (even if search says it is the "wrong" n-alkane) and falls within the projected retention time, its good
- If no peak is present in the n+1 window, look for an identifiable n-alkane at N+2…N+3 etc. (N=number of carbons)
- Recalculate the projection when next n-alkane is found
- Continue to march along until we identify all n-alkanes in the run

The End Result

- We can now reliably identify all n-alkanes for RI calibration, automatically
  - Handles cases with consecutive n-alkanes
  - Handles cases with non-consecutive n-alkanes (e.g. C6-C20, C25,C28,C35)
  - Rejects all interference peaks not n-alkane
  - “Expert” logic handles solvent peaks, “ghost” peaks
  - We also remove “constant” background interference such as column bleed to improve n-alkane identification
  - It can use user “expected n-alkanes” to enhance accuracy, but also can correct for errors and interferences

Example "Difficult" Cases Solve



Some typical "issues" with RI calibrations

- Top: Contaminant peaks present in SPME run, may which are much larger than calibrants, very small calibrant peaks
- Bottom: Column bleed interferes with ID of n-alkanes, must remove

Summary

- Retention index coupled with library search dramatically improves compound ID
- The difficulty in GC RI calibration can be a huge barrier to RI matching
- This robust, automated system for RI calibration (AutoRIX) eliminates the barriers for RI matching usage
- Every run can now automatically take advantage of the rich RI information provided in modern search libraries

Cerno Bioscience Booth #1911