Identification of Unknowns by GC-MS and LC-MS Using NIST Search with Commercial and User Libraries

James Little
Research Fellow, Eastman Chemical Company (41 years)
Consultant, Mass Spectral Interpretation Services (4 years)
tvasailor@gmail.com
https://littlemsandsailing.wordpress.com/
Kingsport, TN

ACS Conference, April 8, 2021
4:00-7:00 PM EST
Compound Identification with LC/MS and GC/MS: Experimental Method, Data Analysis, and Applications
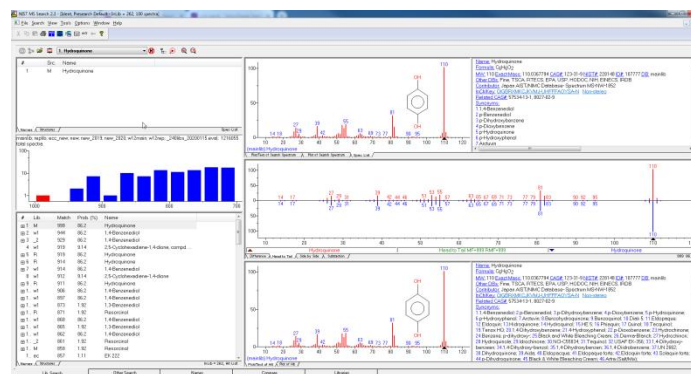Organizer: Xiaoyu (Sara) Yang [*NIST*]

NIST Search and Libraries at Eastman Chemical[1]

Eastman Chemical Co., Main Site, Kingsport, TN
50 Manufacturing Sites Worldwide, 13,500 Employees

NIST Search Software

~50 GC/MS's, LC/MS's Networked _Worldwide_

Eastman Use of NIST Search for Unknown Identification[1]

▪*Majority* of samples analyzed in remote laboratories by chemists in chromatography, environmental, and synthetic laboratories using Agilent GC-MSD's

▪User experience varies, but typically *not* "highly skilled" in mass spectrometry interpretation

▪*If* results not found by search, files are sent to "experts" in central MS lab for identification[2]

▪"Experts" report results *promptly* via e-mail and add to corporate library

▪Eastman corporate library *automatically* built and distributed *daily* to users

▪*If necessary*, samples sent to central MS lab for accurate mass, chemical ionization, advanced sample preparation, micro-synthesis, *etc.*

Free Web-Based Training for NIST Search Software[3]

-Two separate courses[3]
-GC/MS and LC/MS

https://littlemsandsailing.wordpress.com/

Handouts PDF format
Screenshots of settings
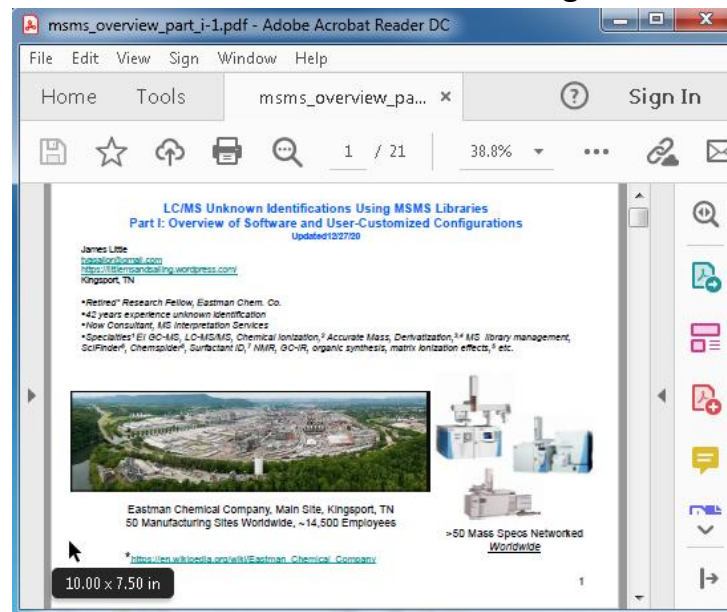
Table of Contents

Libraries Searched

EI Libraries	Entries	Comment
NIST20	350K	commercial
Wiley12	815K	commercial
Wiley Specialty (11 separate)	91K	commercial
MoNA[14,15]	19K	Free, "crowd-sourced"
Eastman Corporate	60K	proprietary
Total	1,335K	

| MSMS Libraries	Entries	Comment
NIST20 Tandem	1,300K	commercial
Wiley Tandem	25K	commercial
MoNA[14,15]	145K	Free, "crowd-sourced"
Eastman Corporate	3k	proprietary
Total	1,473K	

Types of Components Added to Eastman Library

▪*Anything* a user would find useful, thus ***much more*** "diverse" than purchased commercial EI and MS/MS libraries

▪***Thus,*** users must realize Eastman library is an ***aid*** *to identification* and should be used accordingly

▪Most entries high quality with high confidence and exact structure

▪Some entries will have "?" or "??" in front of name to show some uncertainty

▪Some entries added with no exact identity with reference to manufacturing process, R&D project name, and detailed sample history

Importance of Library Search and Comments in User Library

Google Aerial View of Kingsport Site
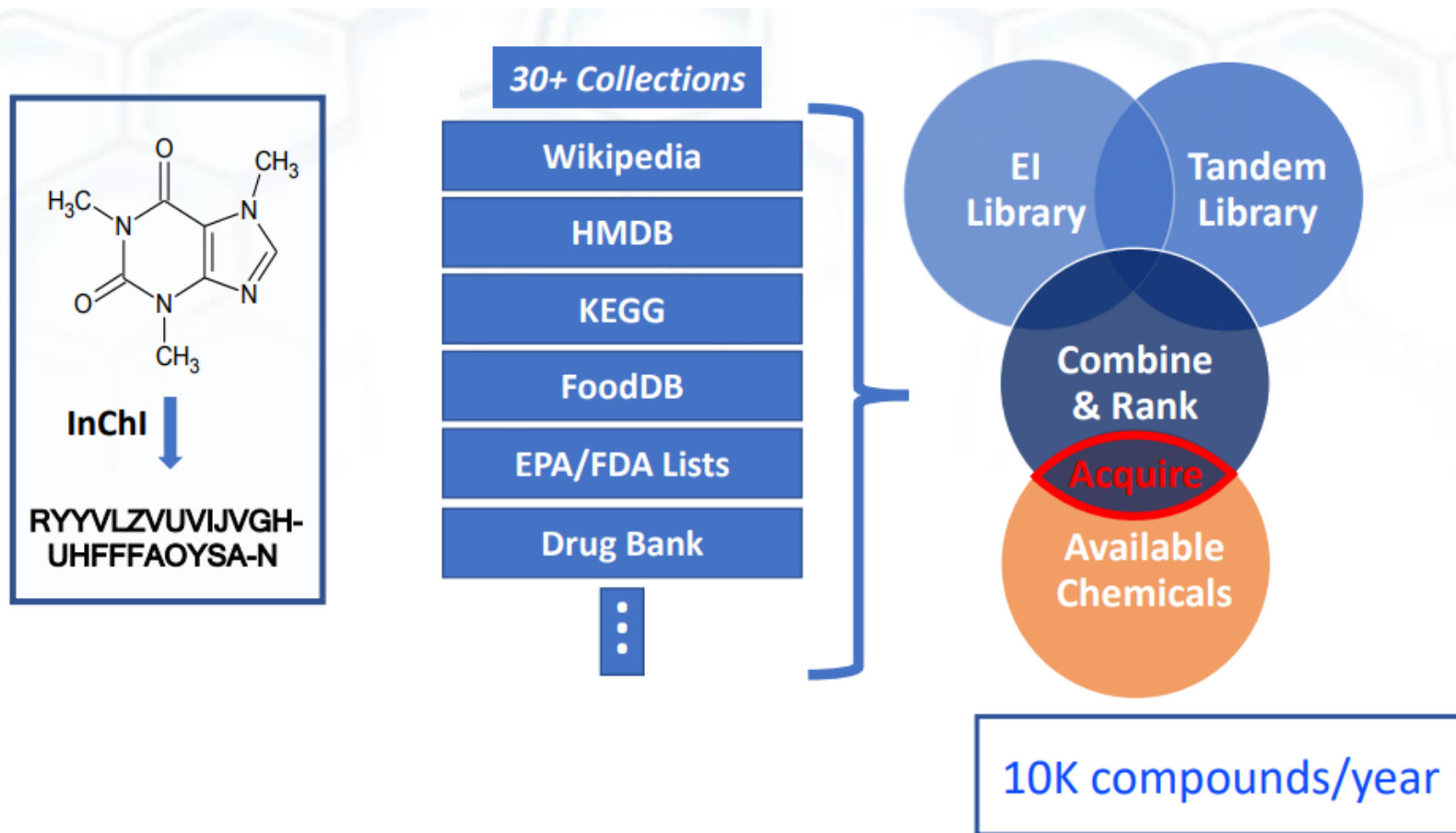4000 acres, 7,000 employees, 550 Bldgs.

■GC-MS's used in conjunction with TOC (total organic carbon) water monitors

■Comments Quickly target source of spills in manufacturing or water treatment plant upsets
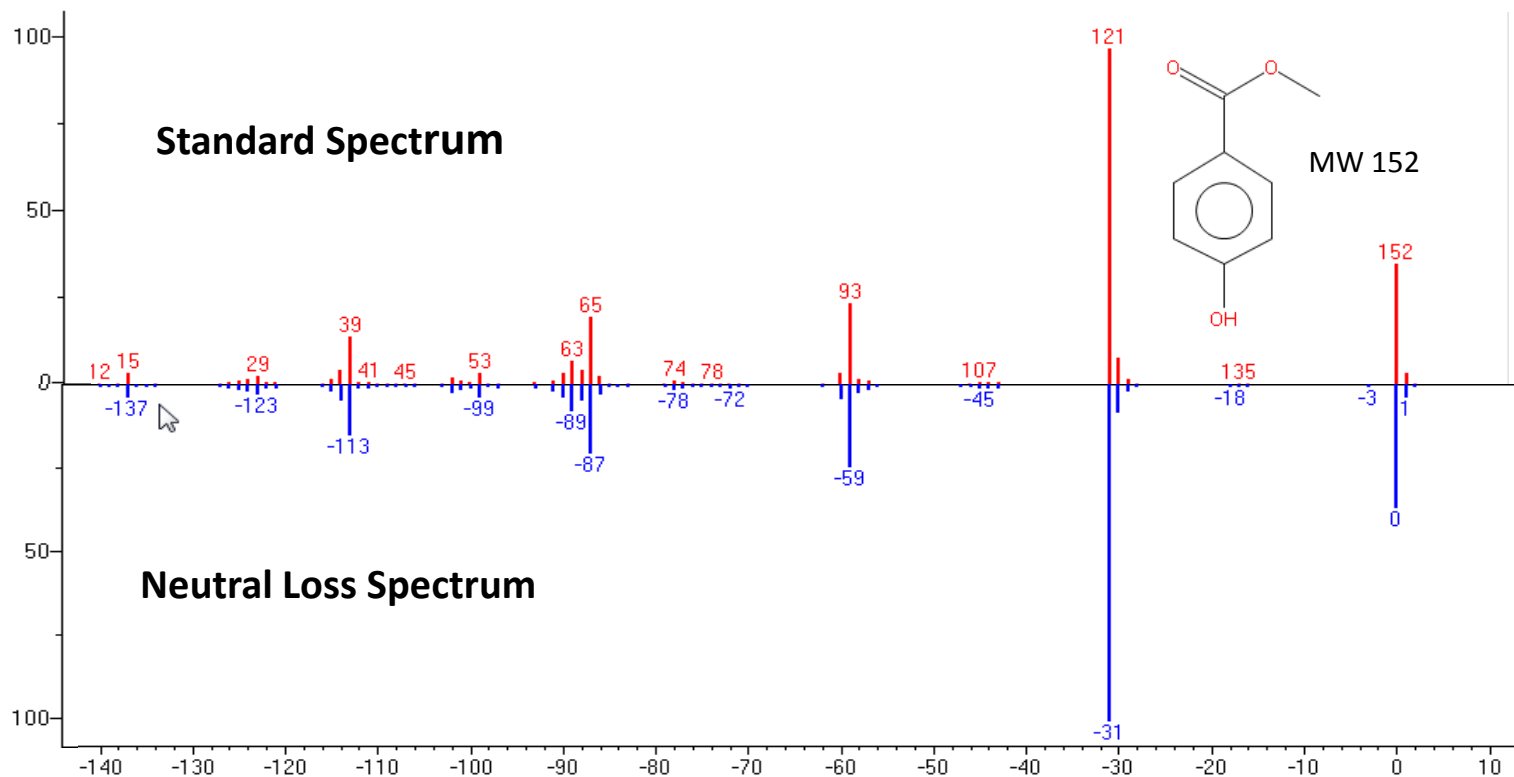
■Exact structure useful, but not necessary
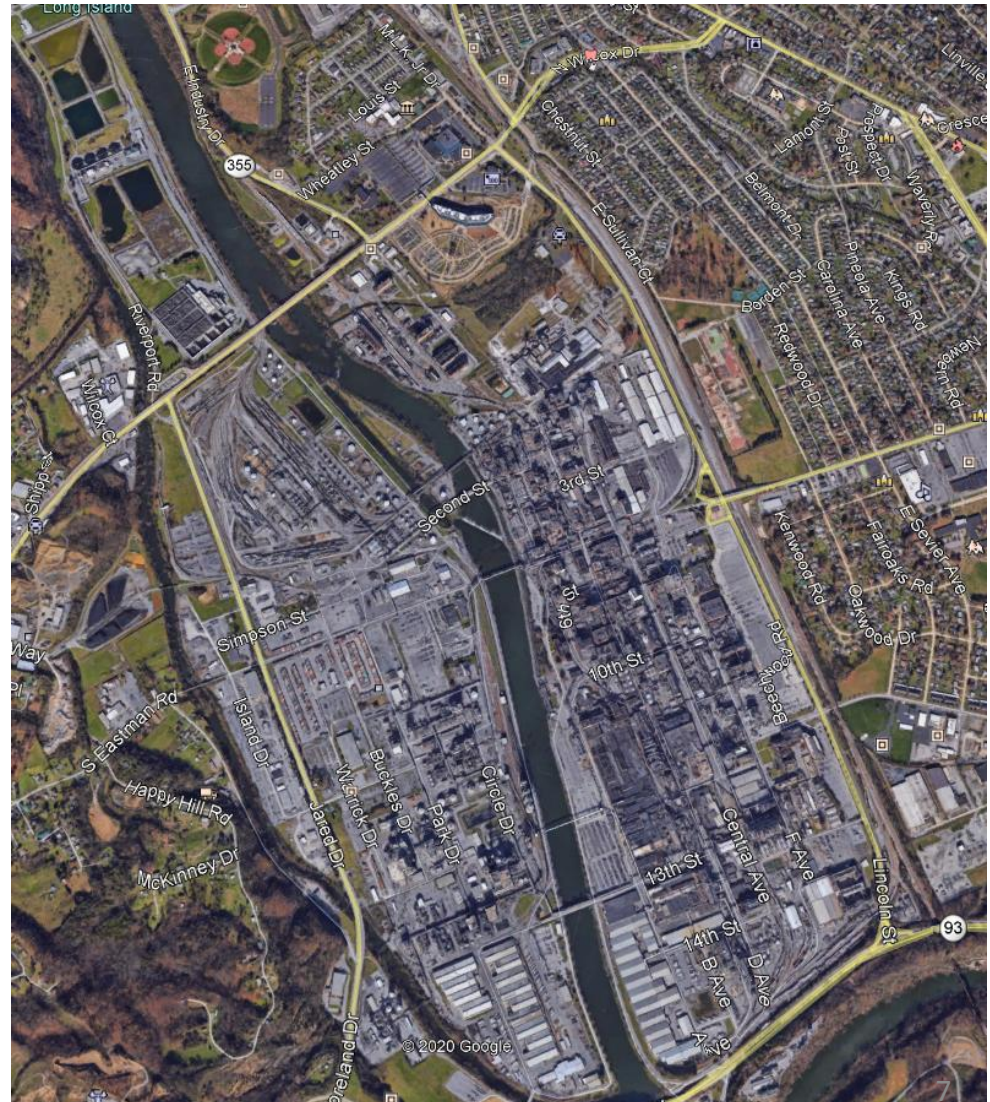
Some of the New Things for NIST in Last 6 Years

- Hybrid search for EI[4,5] and MS/MS[6,7]

- Purchased >50,000 high priority compounds[8]

- Compounds analyzed by EI (underivatized) and as Me, TFA, TMS, and TBDMS derivatives

- Infusion MS/MS at *multiple* energies (critical for good MSMS searches)[8]

- Quality of spectra checked by 2 or more evaluators

- MS interpreter using accurate mass for better correlation of observed ions to substructures[9-12]

Ambitious "Pipeline" for Targeting and Acquiring New Spectra[8]

InChI

RYYVLZVUVIJVGH-UHFFFAOYSA-N

30+ Collections

- Wikipedia
- HMDB
- KEGG
- FoodDB
- EPA/FDA Lists
- Drug Bank

EI Library

Tandem Library

Combine & Rank

Acquire

Available Chemicals

10K compounds/year

Hybrid Search of EI[4,5] and MSMS[6,7] Libraries

Standard Spectrum

Neutral Loss Spectrum

MW 152

"Simplified" Program Description:

- Standard search to generate match factor and hit list
- Searches the neutral loss spectrum of unknown against libraries neutral loss spectra
- Calculates a "hybrid score" based on standard match and neutral loss match
- Determines a *DeltaMass* which is the difference in the MW of unknown and library hit

10

# Hybrid Search of EI[4,5] and MSMS[6,7] Libraries

Program Advantages:

- Greatly extends the scope of all current libraries
- Success requires the presence of similar compounds in the library
- Hits sorted by either standard search match factor or "Hybrid" match factor
- *DeltaMass* reflects the modification of the molecule

My Personal Experience:

- Personally used for over >30,000 EI and 500 MS/MS searches
- *Routinely amazed* by the results
- Useful results obtained which are *not* noted in standard search
- Utility in identifying unknowns, finding similar model compounds, and supporting fragmentation mechanisms
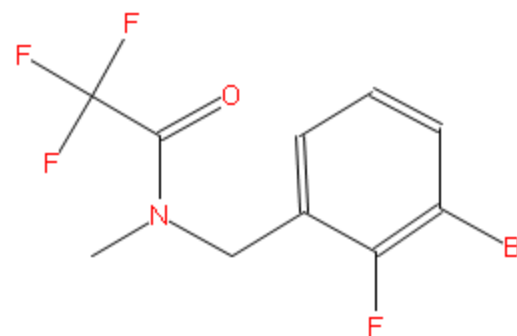
Example of "Hybrid" EI Search

"Unknown" Spectrum

"Hybrid" Best Result
Match 908 of 1000
"Standard search" 232 of 1000
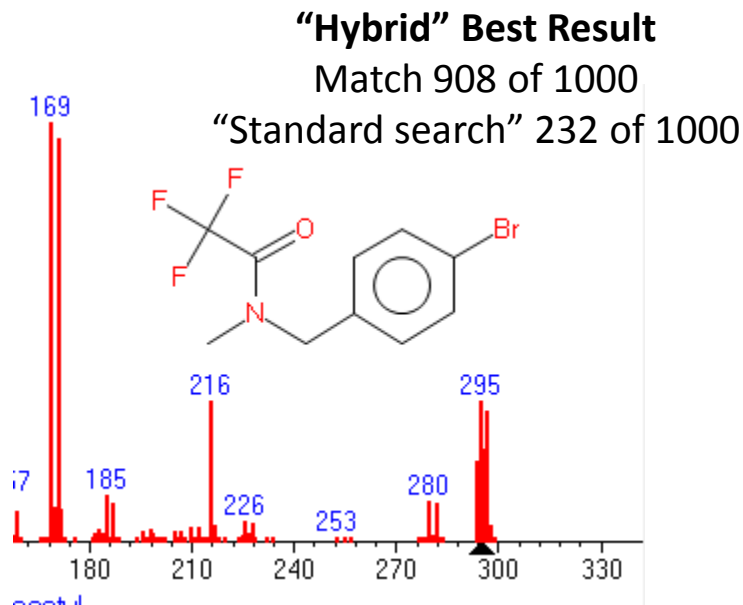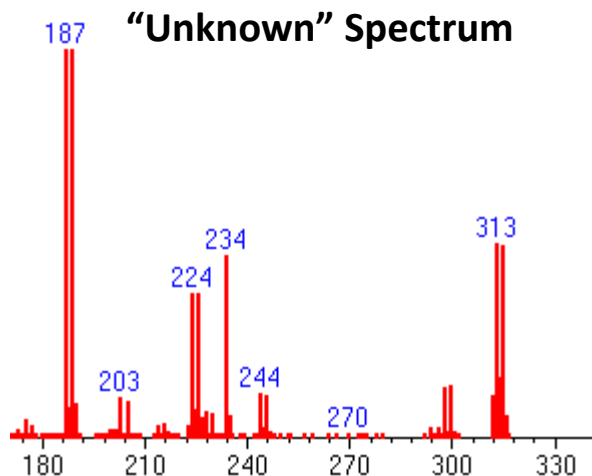
"Standard search" top 50 hits show this substructure
Match 559 to 413 of 1000

m/z 187
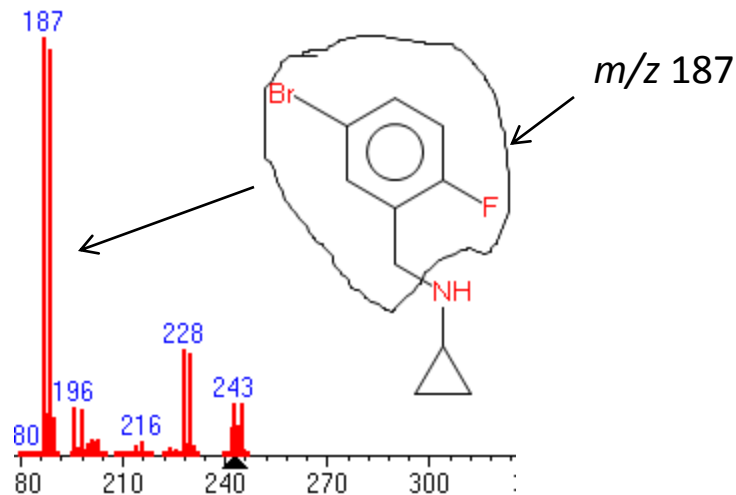
"Identity" of Unknown (or isomer)

DeltaMass = **313-295**

= F − H = 19 − 1

= 18

# NIST Novel Display of Hybrid Search Results
Top is Unknown, Bottom is "Hybrid" Spectrum
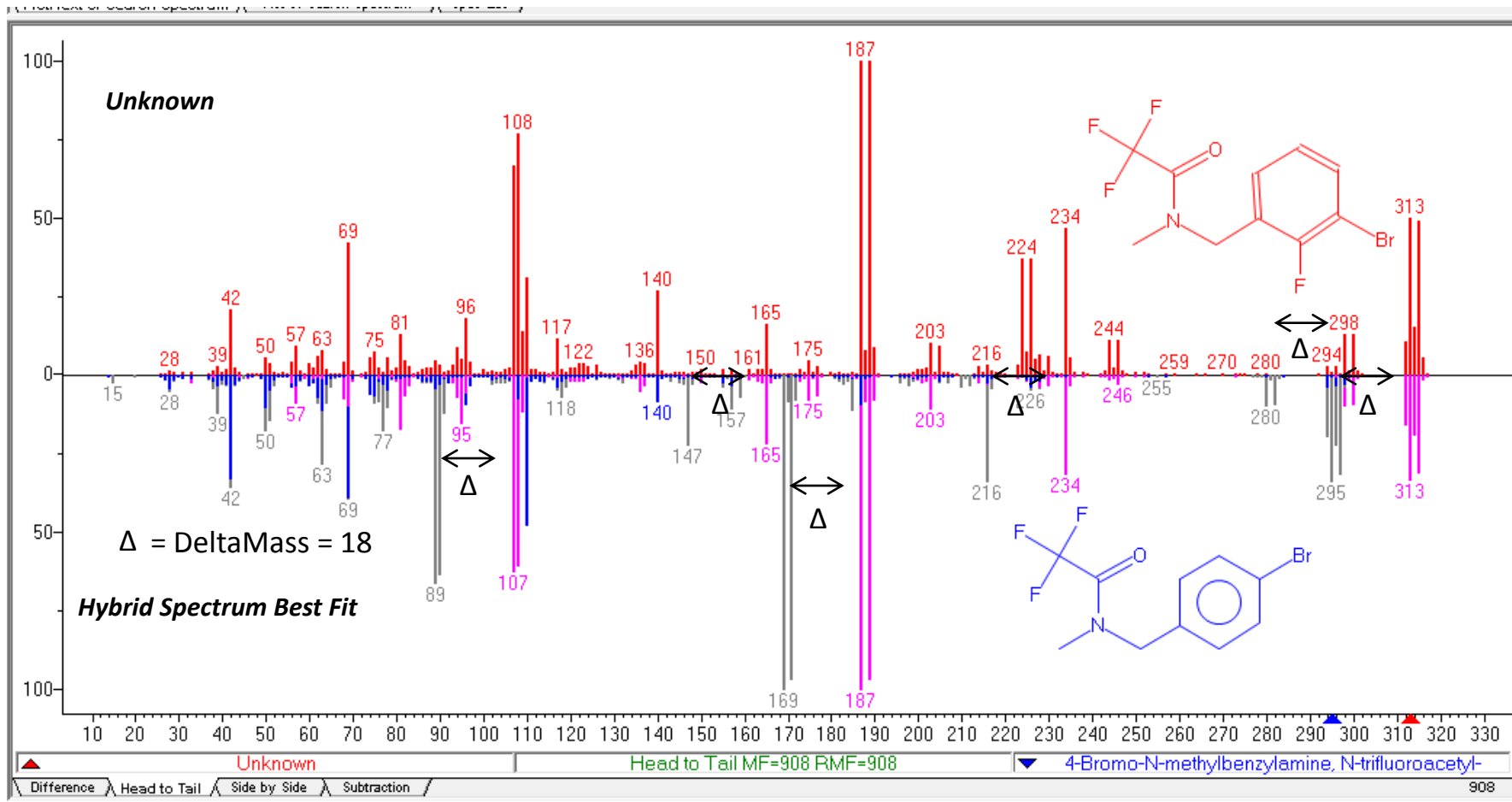
-Bottom spectrum, original ions in grey shifted to magenta by DeltaMass (Δ) 18 for visual comparisons, blue fragment ions not shifted

# Associating Some *Simple* Structures with DeltaMass Values[13]

- Some simple *small* MW compounds to illustrate types of substructural information
- These substructures can be a parts of *much larger* molecules
- *Note:* Odd values of DeltaMass contain one nitrogen change in structure, thus *"Nitrogen Rule"*
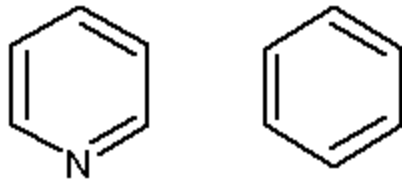- Isotope ratios and/or accurate mass helpful with redundancies
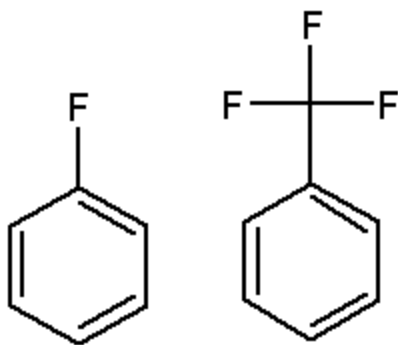
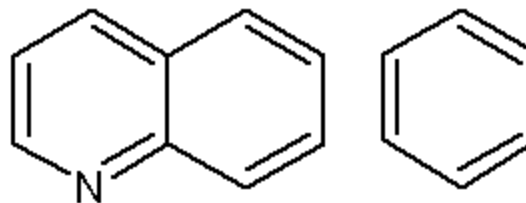MW 96 MW 146

DeltaMass 50
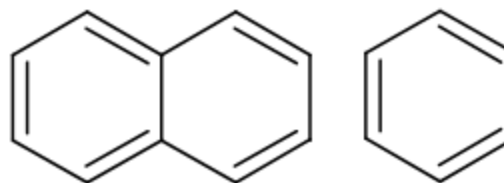
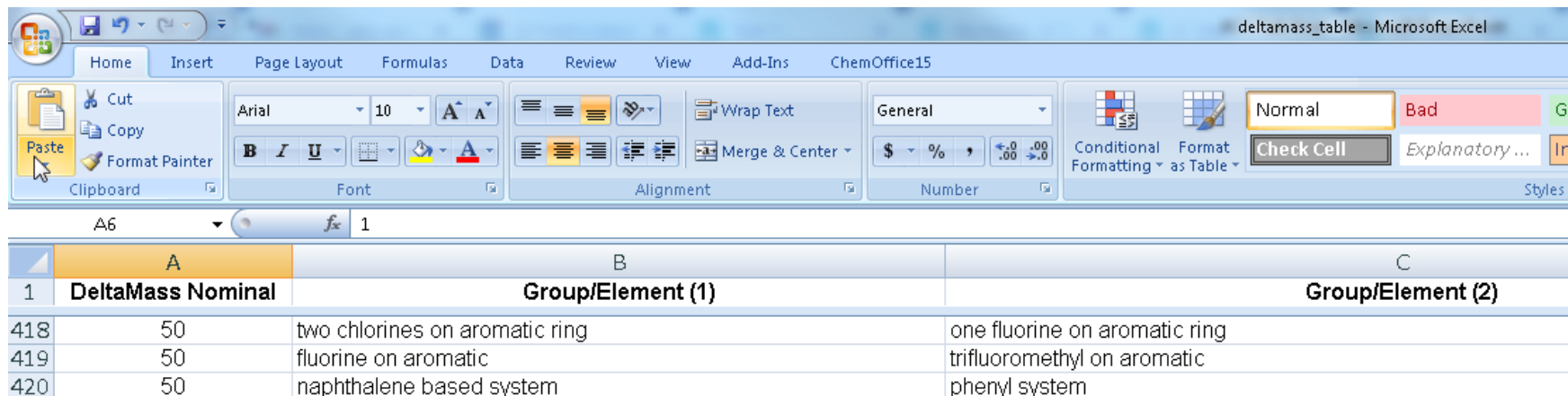MW 128 MW 78

DeltaMass 50

MW 79 MW 78

DeltaMass 1

MW 129 MW 78

DeltaMass 51

14

DeltaMass Table[13]

- >550 Hybrid DeltaMass values noted during >30,000 searches
- *Updated **Regularly** on my web site*

DeltaMass Nominal	Group/Element (1)	Group/Element (2)
50	two chlorines on aromatic ring	one fluorine on aromatic ring
50	fluorine on aromatic	trifluoromethyl on aromatic
50	naphthalene based system	phenyl system

Summary

- NIST Search powerful tool for identifying unknowns by GC-MS and LC-MS

- Essential to Eastman R&D, manufacturing, and environmental functions

- Utilized by both MS experts and less experienced users

- Training courses essential for effective utilization

- NIST continues to add large numbers of important spectra each year

- Hybrid search greatly increases the effectiveness of *all* available EI and MS/MS libraries

Presentation References (*Internet Links*)

1. *Importance of NIST Search at Eastman*
2. *Eastman General Approach for Unknown Identification*
3. *Free Training Courses GC-MS and LC-MS Unknown Identification with NIST Search*
4. *EI Hybrid Video*
5. *EI Hybrid Handout*
6. *Tandem Hybrid Video*
7. *Tandem Hybrid Handout*
8. *NIST Pipeline for Improving EI and Tandem Libraries*
9. *MS Interpreter Correlating Structure to EI Spectrum Video*
10. *MS Interpreter Correlating Structure to EI Spectrum Handout*
11. *MS Interpreter Correlating Structure to MSMS Spectrum Video*
12. *MS Interpreter Correlating Structure to MSMS Spectrum Handout*
13. *Delta Mass Table of Values*
14. *MoNA (MassBank of North America)*
15. *Handout Describing MoNA*

Acknowledgements

- Adam Howard (Eastman)

- Curt Cleven (Eastman)

- Stephen Stein (NIST)

- Dmitri Tchekhovskoi (NIST)

- Yuri Mirokhin (NIST)

- Arun Moorthy (NIST)

- David Sparkman (NIST Contractor)

- Gary Mallard (NIST Contractor)

- Sandy Markey (NIST Contractor)