Creation and Annotation of a Recurrent Spectral Library of CHO Cell Metabolites and Media Components

Kelly H. Telu, Ramesh Marupaka, Nirina R. Andriamaharavo, Yamil Simón-Manso, Yuxue Liang, Yuri A. Mirokhin, Xinjian Yan, Tallat H. Bukhari and Stephen E. Stein

67th ASMS Conference on Mass Spectrometry and Allied Topics

June 5th, 2019

National Institute of Standards and Technology
U.S. Department of Commerce

Mass Spectrometry Data Center

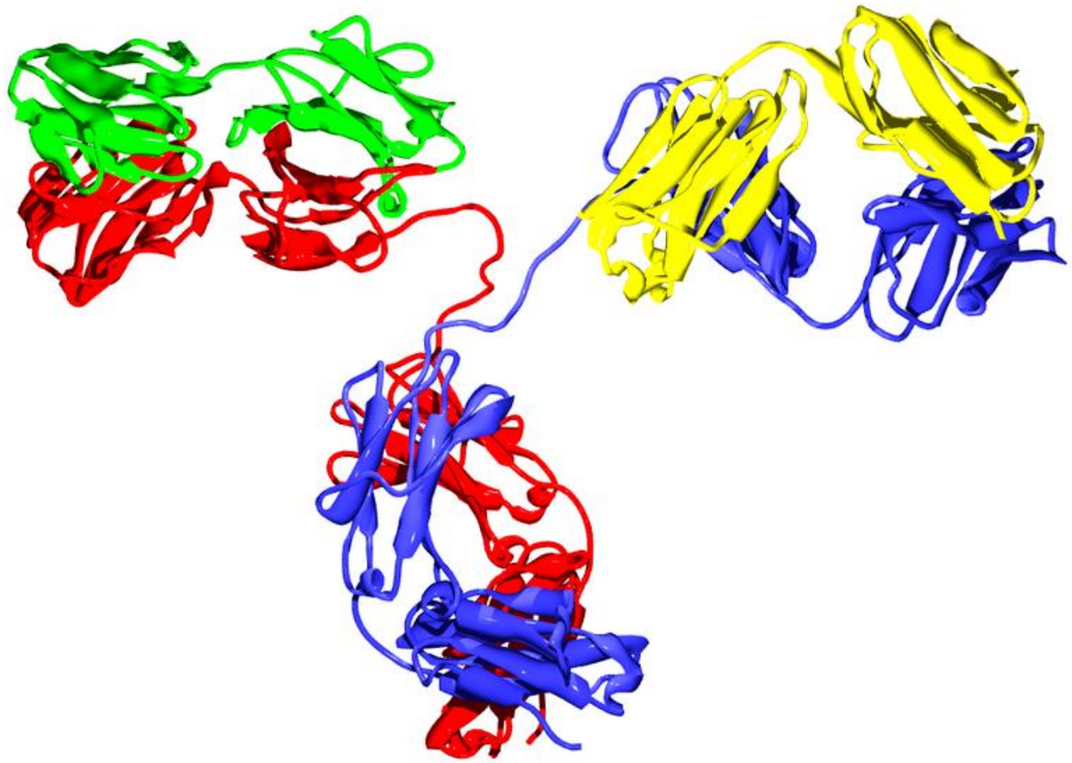
Project Goal and Talk Outline

- Goal: Identify as many metabolites as possible

- Outline:
  1. Introduction (Intro)
  2. Recurrent spectral library for CHO cell metabolites and media (Library)
  3. Improving identification accuracy (ID)
  4. Strategy to annotate unidentified spectra (Annotate)
  5. Strategy to assign confidence to MS/MS IDs (Confidence)

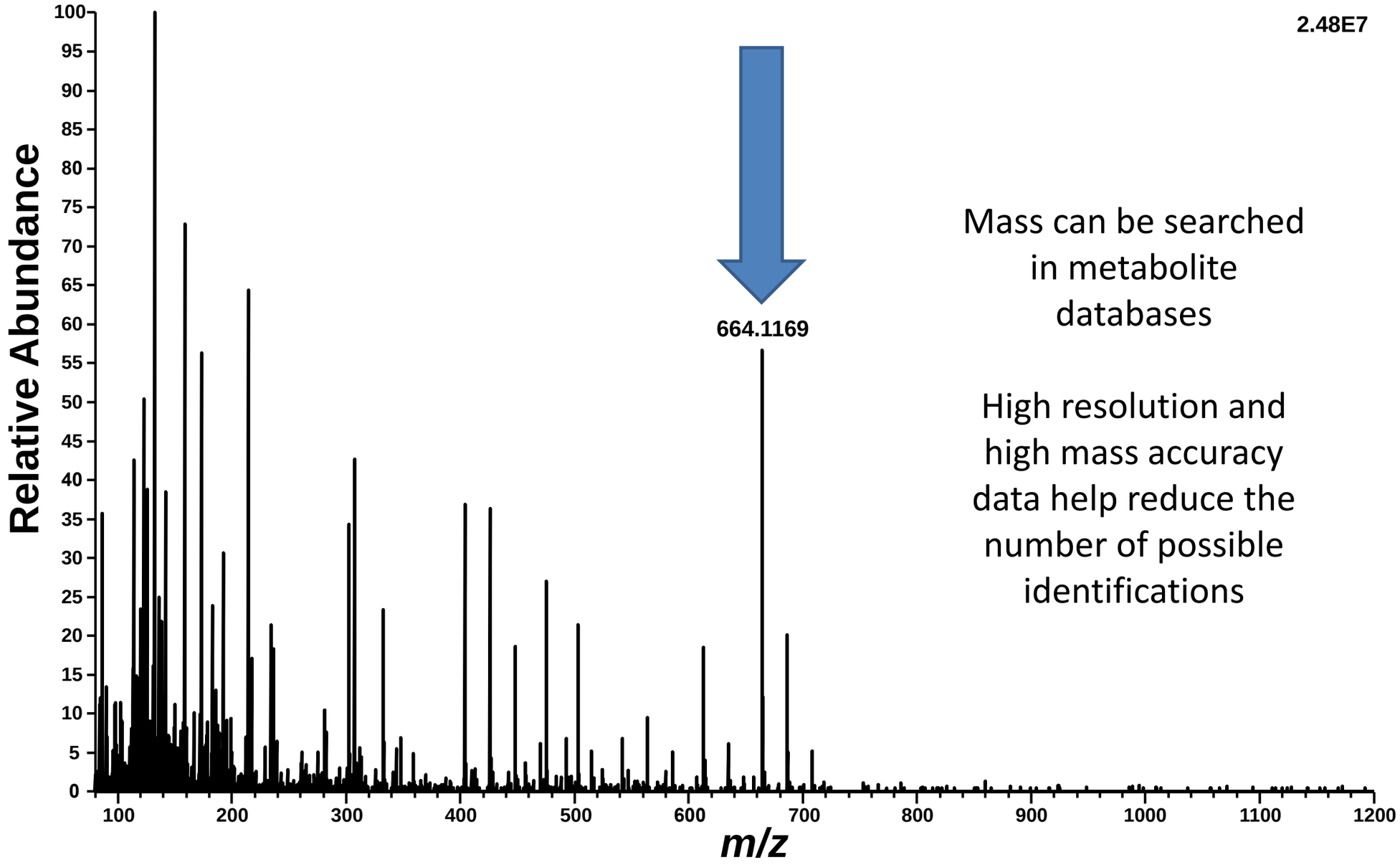
Biomanufacturing

- We analyzed Chinese hamster ovary (CHO) cell metabolites and growth media components using LC-MS/MS as part of NIST's Biomanufacturing Initiative.
- CHO cells are the predominant host cells for monoclonal antibody (mAb) production.[1]
- mAbs are a class of biopharmaceutical estimated to be worth nearly $125 billion in global sales by 2020.[2]
- Metabolic profiles of CHO cells and culture media can be used to monitor process variability, improve biologic output and identify contamination.



1. Walsh, G. (2018). Biopharmaceutical benchmarks 2018. Nature biotechnology, 36(12), 1137.
2. Ecker, D. M., Jones, S. D., & Levine, H. L. (2015, January). The therapeutic monoclonal antibody market. In MAbs (Vol. 7, No. 1, pp. 9-14).
3. Image: https://commons.wikimedia.org/wiki/File:Antibody_IgG2.png

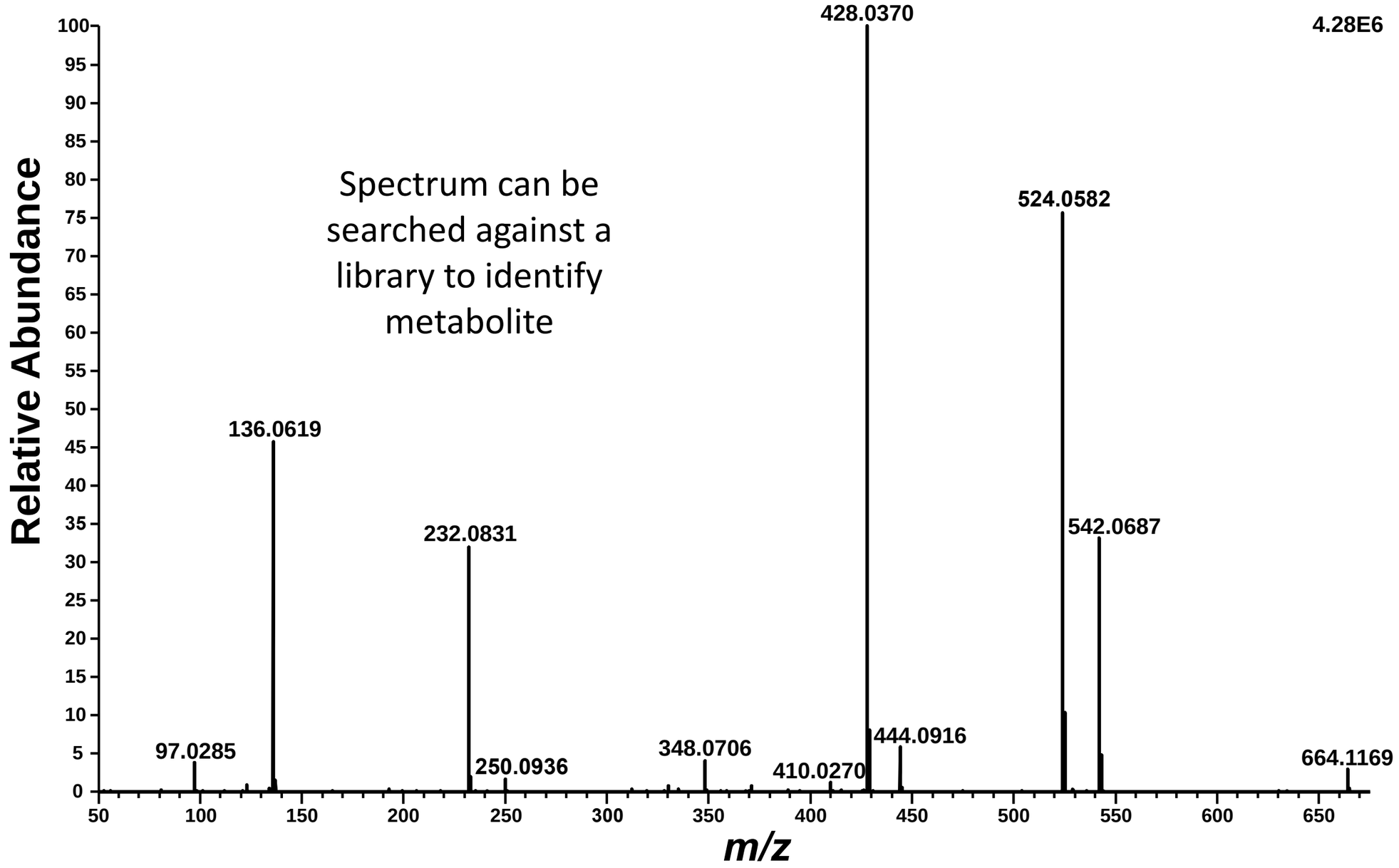
Metabolite Identification using MS



Mass can be searched in metabolite databases

High resolution and high mass accuracy data help reduce the number of possible identifications

Metabolite Identification using MS/MS

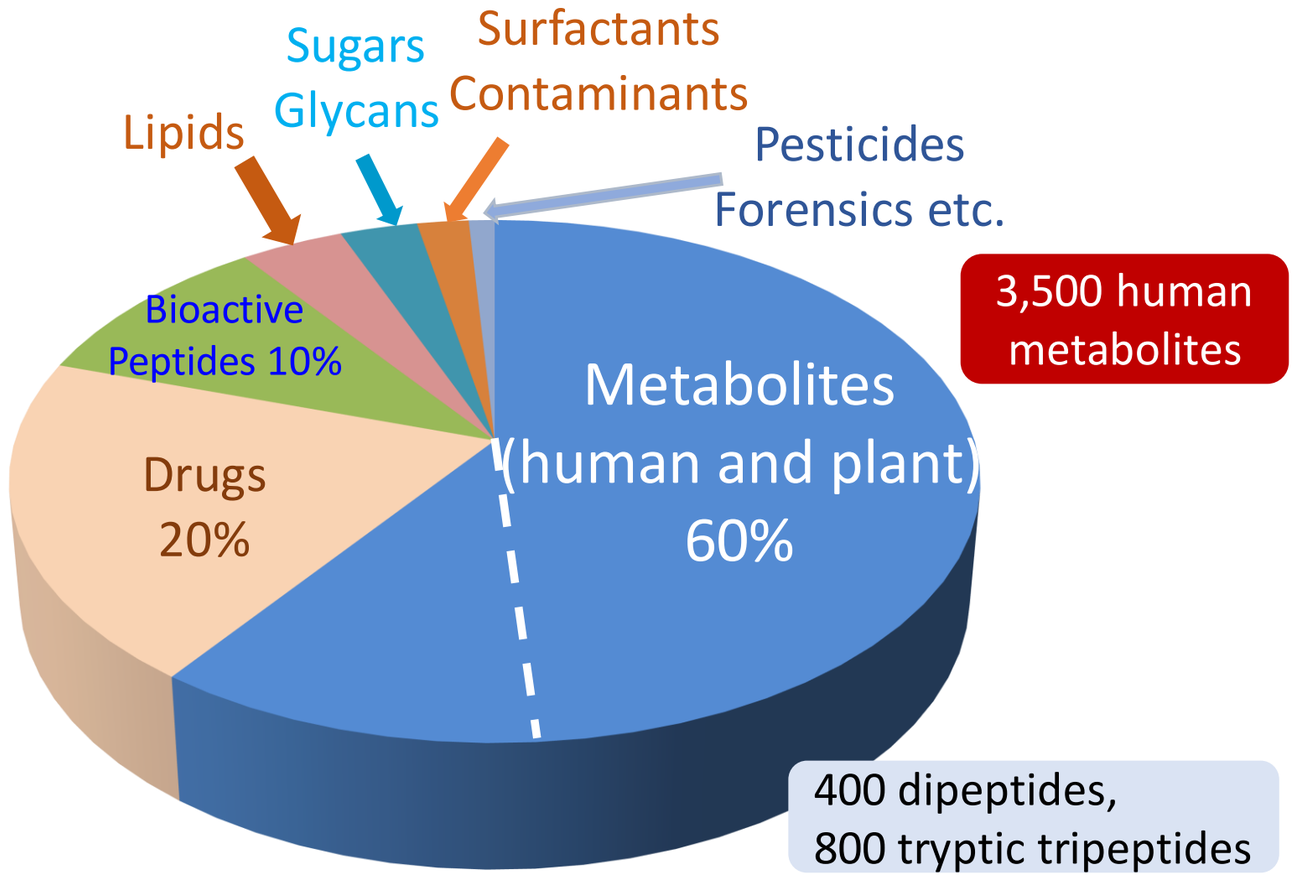


NIST 17 Tandem Mass Spectral Library

15,243 compounds

123,881 precursor ions

574,826 spectra



80% (+), 20% (-)

27% MS^2 of in-source

17% MS^n

https://chemdata.nist.gov/

Characterizing Product Ions in a Reference Tandem Mass Spectral Library – WP 422

Library Searching

NIST MS Search 2.3 - [MS/MS, Presearch Default - 3 spectra]

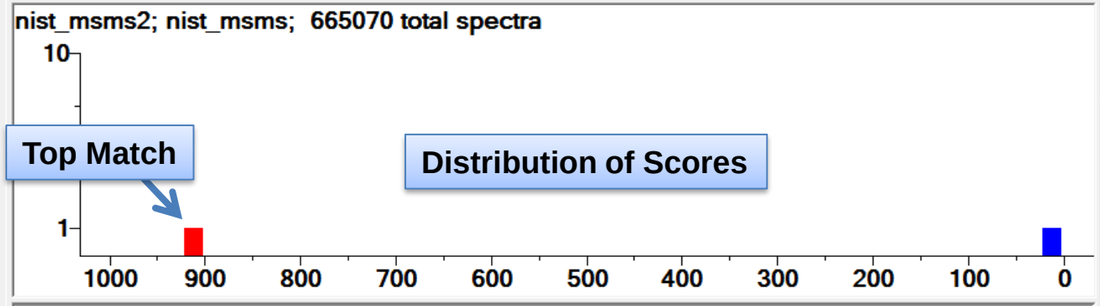
File Search View Tools Options Window Help

1. Scan:1093 RT:2.168 PrecursorScan:

...	S..	Name
1	L	Scan:1215 RT:2.414 PrecursorScan:1207? nMSN:1055 ...
2	L	Scan:294 RT:0.610 PrecursorScan:290? nMSN:254 Mo...
3	L	Scan:292 RT:0.607 PrecursorScan:290? nMSN:252 Mo...
4	L	Scan:417 RT:0.823 PrecursorScan:411? nMSN:361 Mo...
5	L	Scan:614 RT:1.182 PrecursorScan:612? nMSN:532 Mo...
6	L	Scan:1093 RT:2.168 PrecursorScan:1092? nMSN:948 M...
7	L	Cluster_000038 NS=1244 NR=60 Ch=-1 NI=223/Top=11

Names | Structures | Spec List

Spectrum List

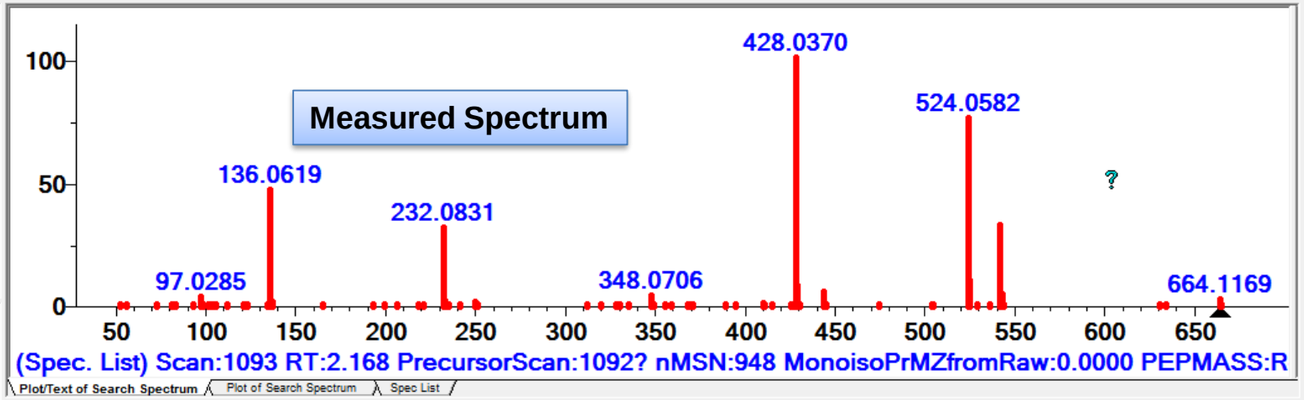


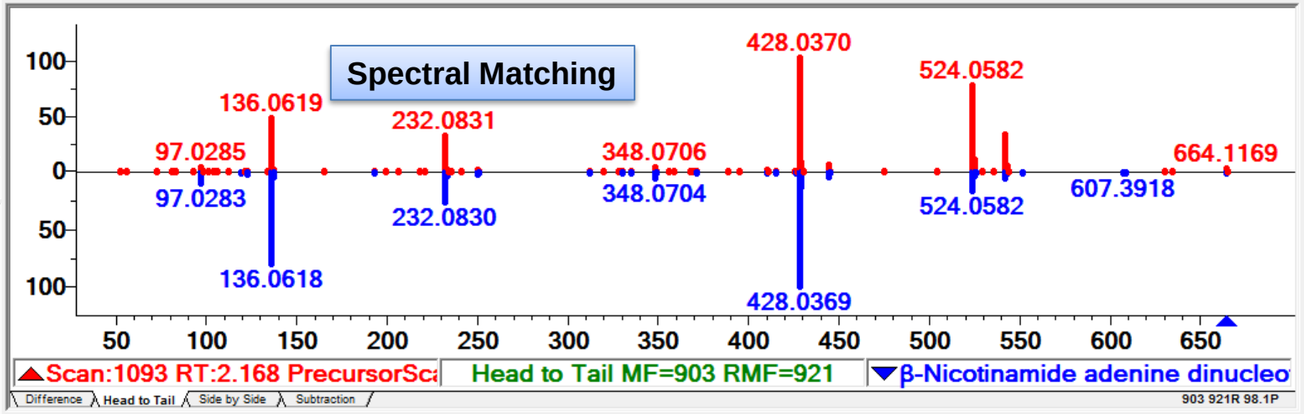
Top Match

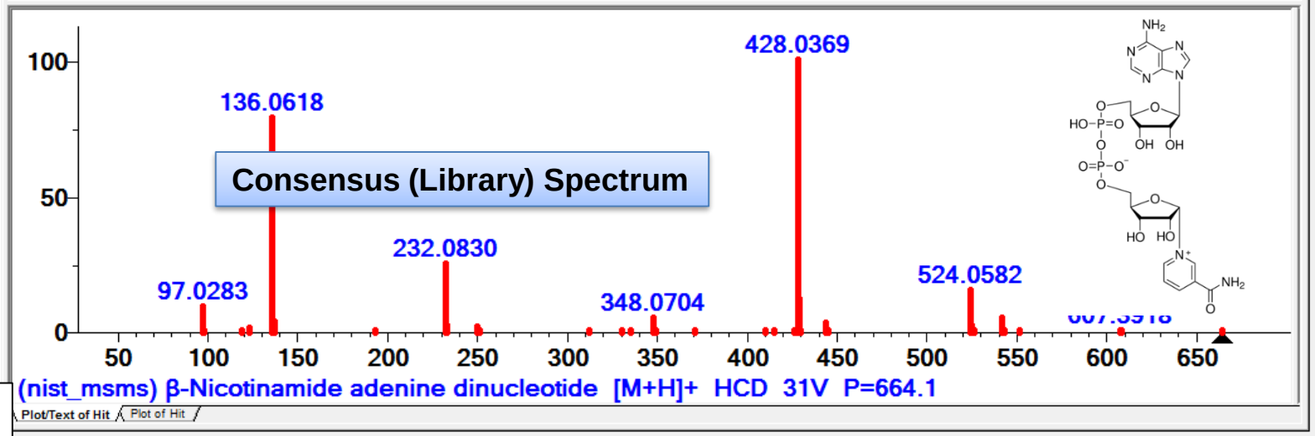
Distribution of Scores

#	Sc...	Do...	R...	Pr...	T/...	N	S...	Name
1	903	921	925	98.1	6.21	2.	2	β-Nicotinamide adenine dinucleotide [M+H]+
2	1	14	98.2	0.98	-1...	6	2	β-Nicotinamide adenine dinucleotide, reduced
3	0	1	23.9	0.94	-1...	1	0	Adrenomedullin (1-52), porcine [M+9H]9+ QQ(

Hit List







Evaluation of NIST Library Search Software – WP 311

Definitions

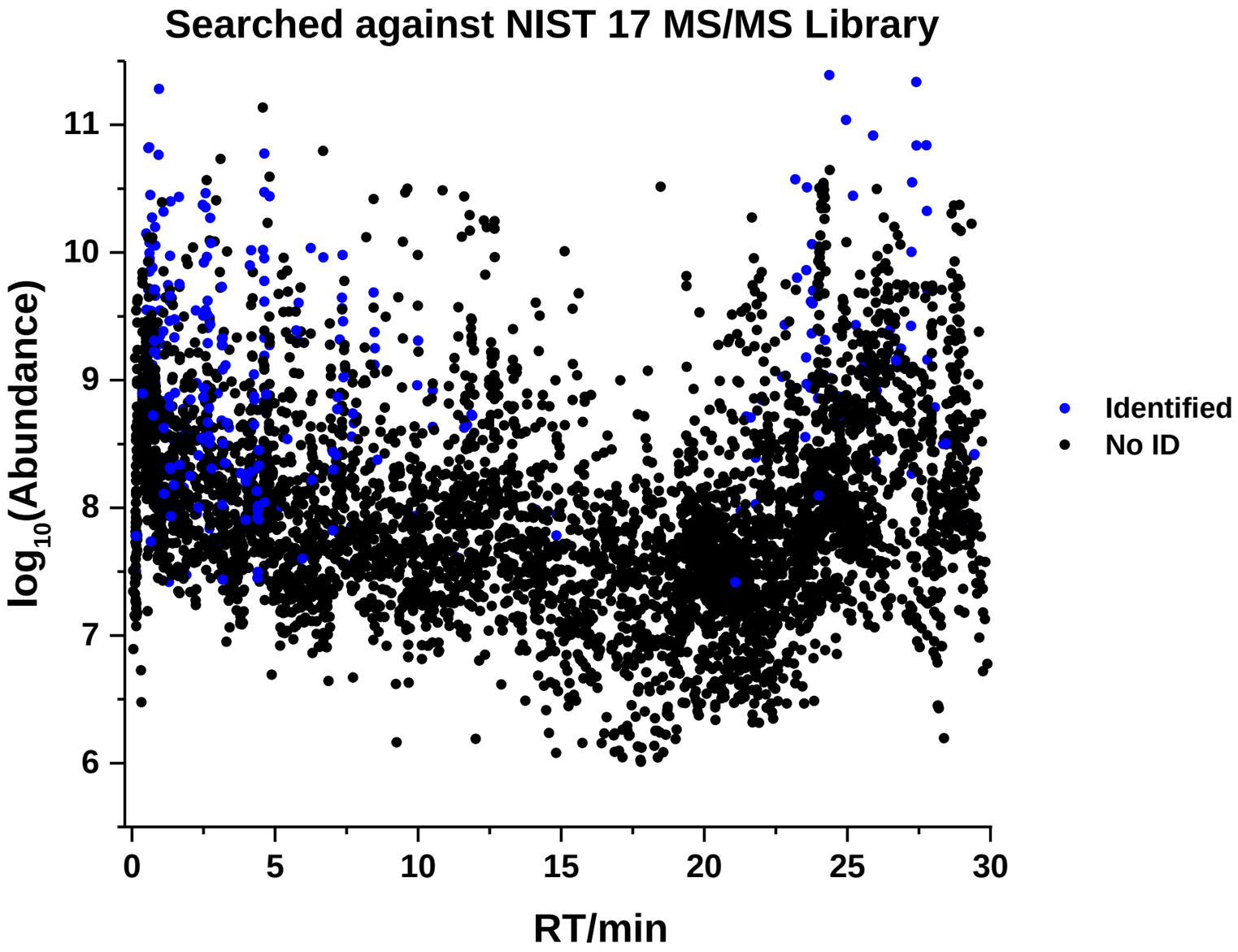
- **Consensus Spectra** – Spectra that are a result of processing similar experimental spectra to produce a representative spectrum to put in a library.

- **Recurrent Spectra** – Spectra that occur repeatedly in the sample.

- **Hybrid Search** – A type of MS/MS search that finds ions that differ by an inert chemical group, hence can often match unidentified spectra with members of the same chemical classes that are present in the library. The term Delta Mass is used to represent the difference in mass between the query spectrum and library entry.
  - Burke, M.C., *et al*., *J. Proteome Res*. 2017, *16*(5), 1924-1935
  - Blaženović, I., *et al*., *Analytical chemistry* 2019, *91*(3), 2155-2162

Ion Plot of CHO Metabolite Extraction Sample

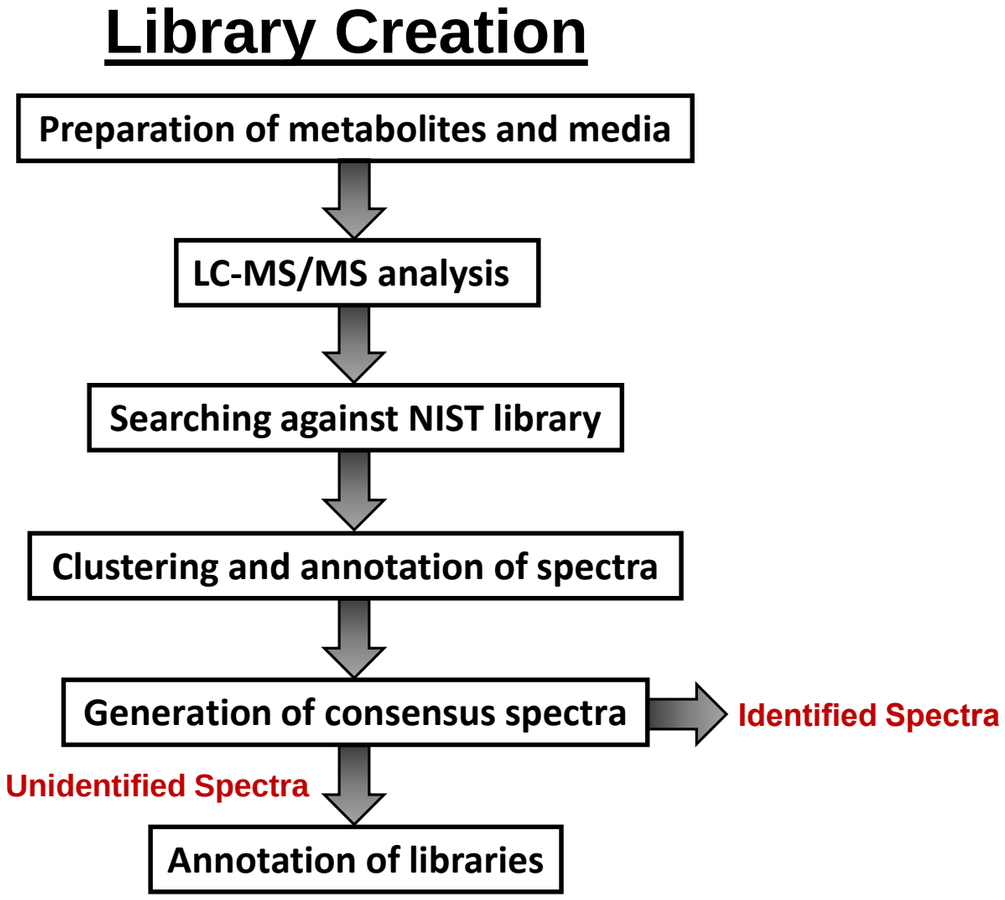
- Each circle is a cluster of similar MS/MS spectra
- **94% of these clusters are not identified**
  - Direct MS/MS Search
  - Score cutoff of 400



Recurrent Unidentified Spectral Libraries

- Uses:
  1. Identify any ion by fragmentation fingerprint
  2. Has ion been seen before?
     - How often, and under what conditions?
  3. Assign class ID for compounds not in library
     - Or not commercially available
  4. **Enables library 'evolution' via user feedback**
     - **Identify once = Identify always**

- Developed libraries derived from:
  - CHO cells and culture media
  - 7 Urine SRMs
  - 8 Serum/Plasma SRMs
  - Glycans



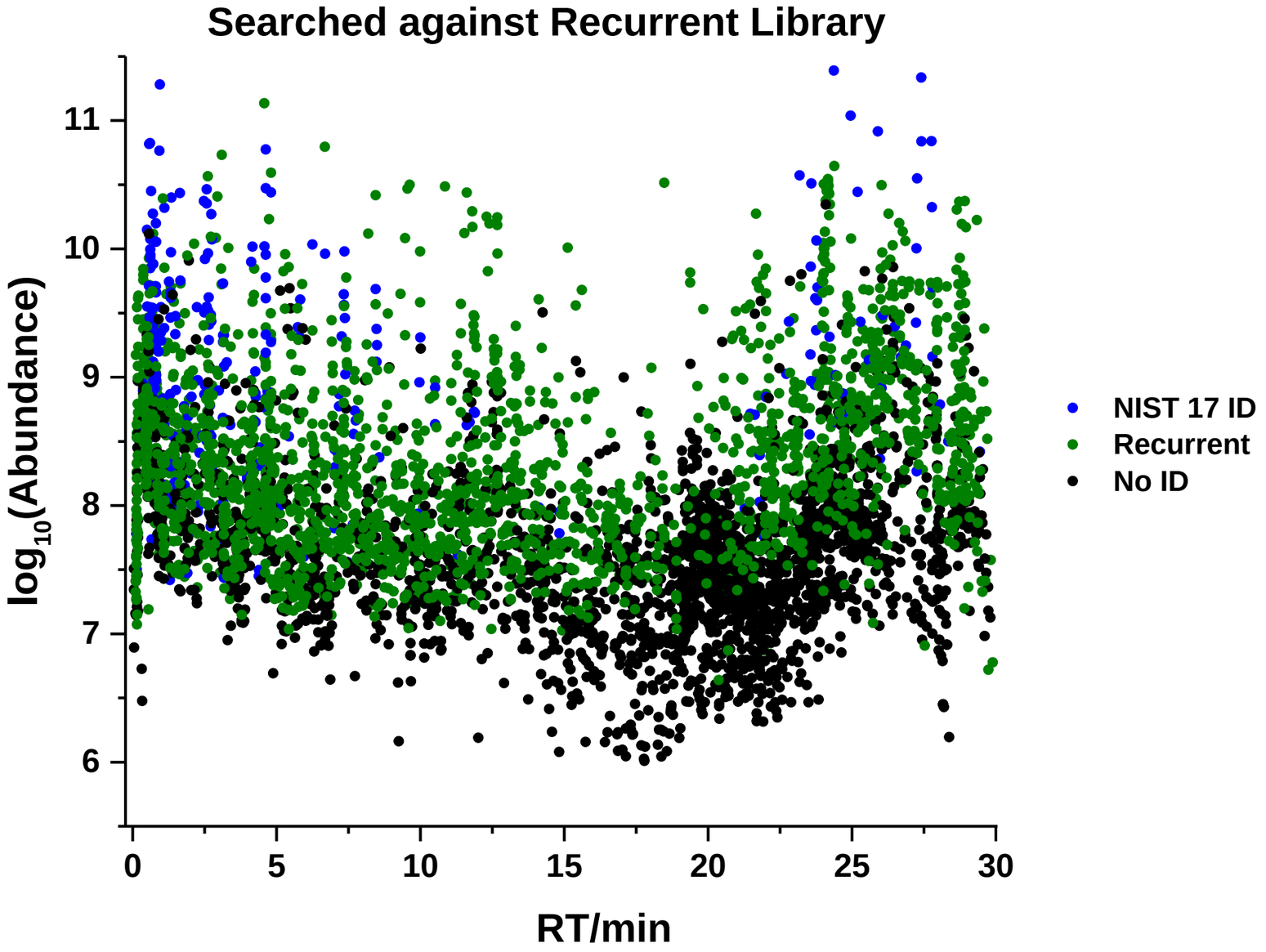
https://chemdata.nist.gov/

Materials and Methods

- Metabolite Extraction Methods
  - Resuspension in 50% Acetonitrile; Methanol/Chloroform/Water; Freeze/Thaw in Methanol
- Fresh/Spent Media Preparation
  - Precipitation in 80% methanol, Drying completely, Resuspension in 100% methanol or 50% acetonitrile
- LC Methods
  - Reversed phase and HILIC LC methods – Aqueous phase of metabolite extractions and media samples
  - Lipid LC method – Organic phase of metabolite extraction
- MS Methods
  - Positive and Negative mode
  - Beam-type Collison Cell (BTCC) with wide energy range and ion trap (IT) fragmentation

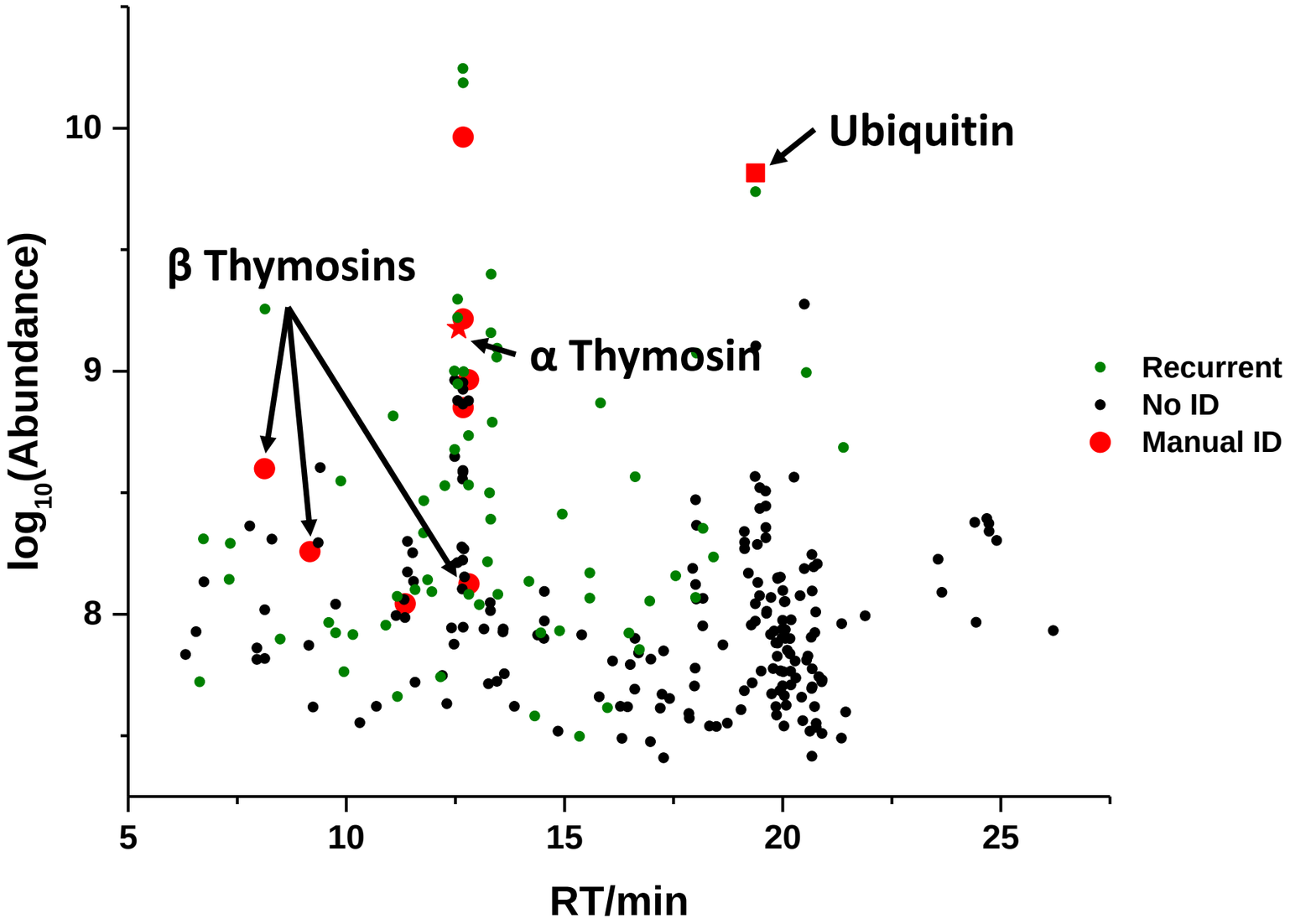
Searched Against Recurrent Unidentified Spectral (RUS) Library

- **Now 49% of clusters are not identified.**
- RUS Library
  - Represent all detectable metabolites, both known and unknown.
  - Cover all known fragmentation conditions and precursors.
  - Combines CHO metabolites and media components because it is difficult to completely separate. Spectra origin is labeled (CHO, media, both).
- Differences from Tandem Library:
  - Contains unidentified spectra
  - Metabolite extracts vs single compound analyzed



High Charge (4^+ or Higher) Clusters

- Present only in the 50% acetonitrile extractions
- Unexpected
  - 79% are 1^+
  - 9% are 2^+
  - 6% are 3^+
- Suspected large peptides/small proteins
- Red ones manually identified



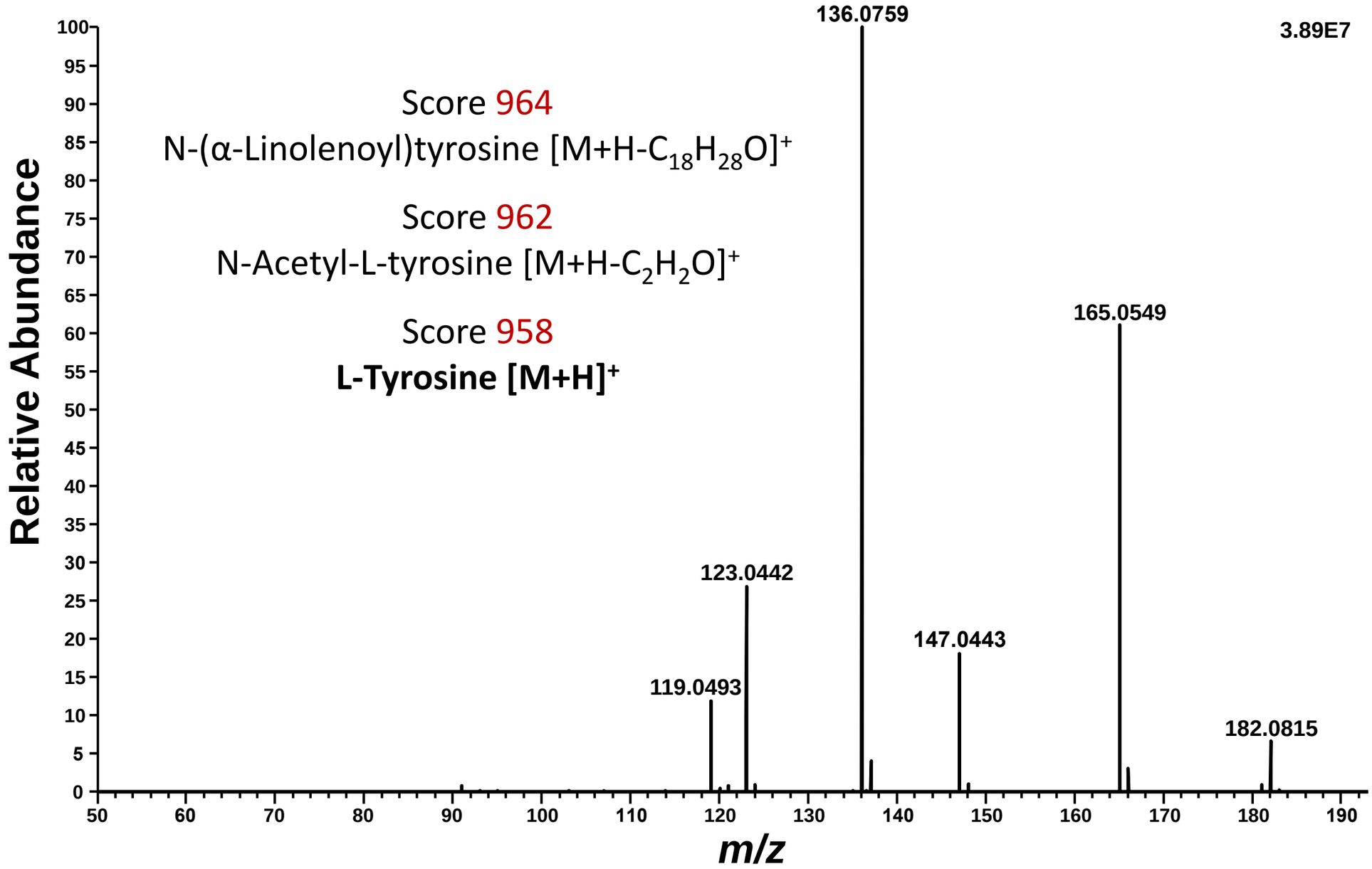
Improving MS/MS Identification Scoring

- Initially score was only based on quality of spectral match.

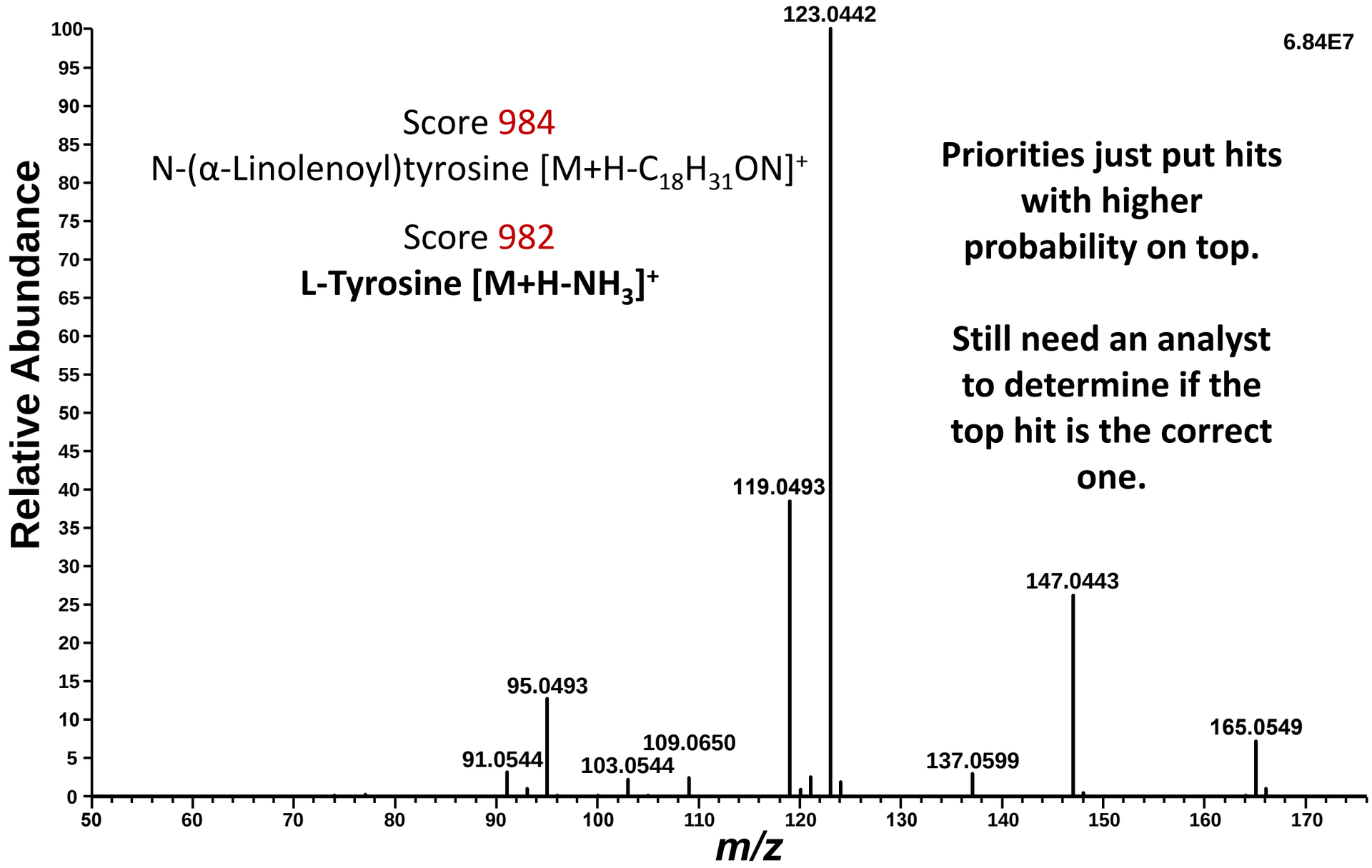
- Led to 3 types of match problems
  1. In-source fragmentation match with a higher score compared to the non in-source match
  2. Uncommon loss match with a higher score than a common neutral loss match
  3. Hybrid search match with a higher score than direct match

- Solution – Changed scoring to give greater priority to some identifications with higher probability of being correct:
  1. $[M+H]^+$ precursor matches
  2. $[M+Na]^+$ and $[M+H-NH_3]^+$ precursor matches
  3. $[M+H-H_2O]^+$ and $[M]^+$ precursor matches
  4. Hybrid search matches
  5. Other in-source precursor matches

In-source Fragmentation Match with Higher Score



Uncommon Loss Match with Higher Score



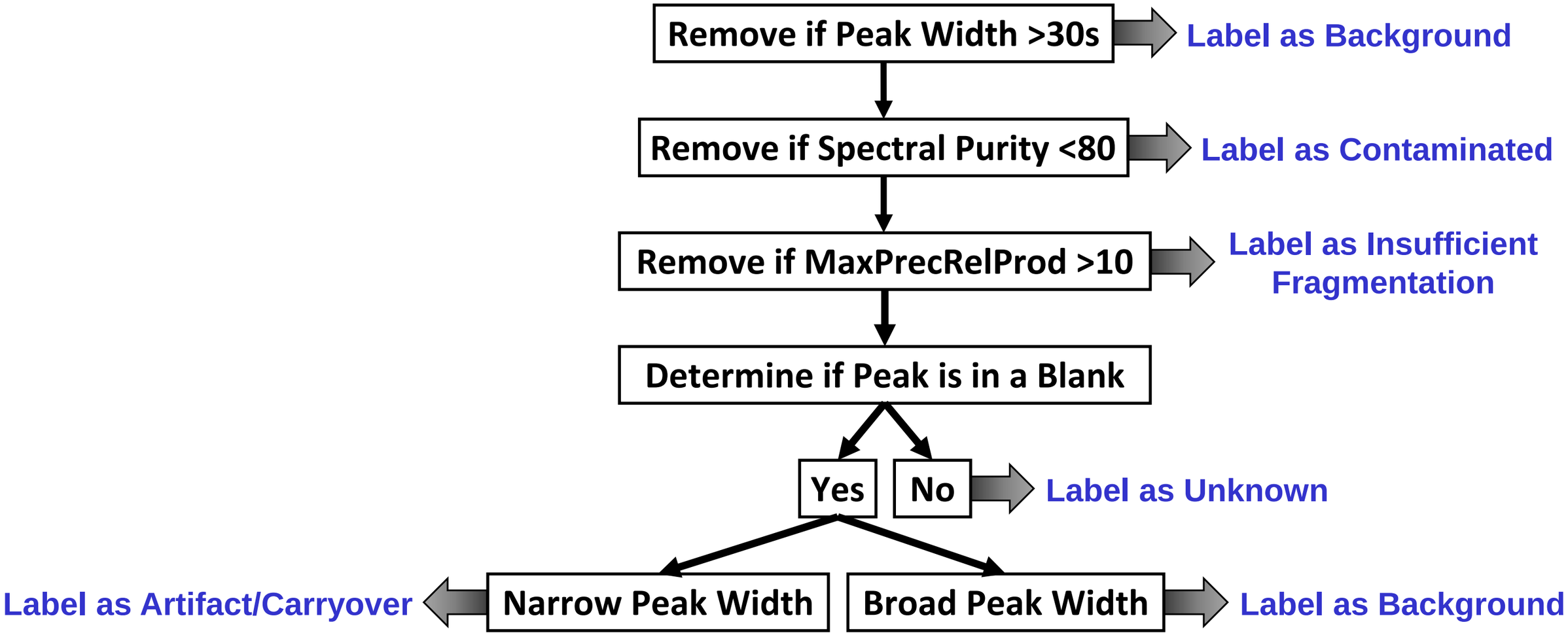
Identification Reassignment

- Incorrect match types that improving scoring did not resolve:
  1. Uncommon metabolite/not a metabolite match with a higher score compared to a common one
  2. Too few peaks in the spectra to result in a match above the score cutoff of 400

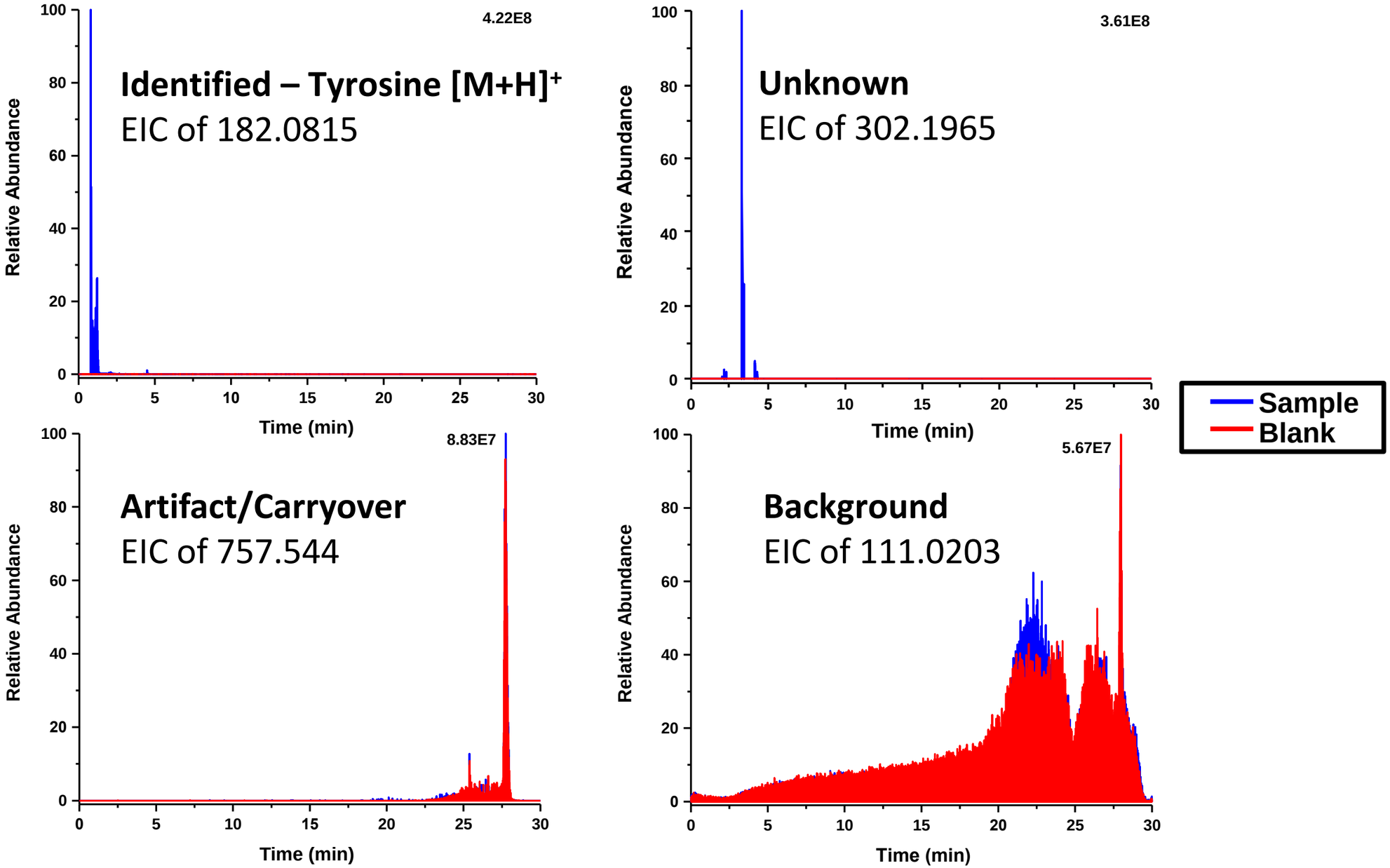
- Incorrect match type introduced by improved scoring:
  - protonated uncommon metabolite match with a higher score than a common metabolite with a common neutral loss
    - Ex: protonated pyroglutamic acid instead of protonated glutamic acid with a water loss

- Solution:
  - semi-automated approach that would re-assign an identification based upon a manually curated list of masses and their appropriate identification
    - IDs discovered by hybrid search also included

Annotation of Recurrent Unidentified Spectra

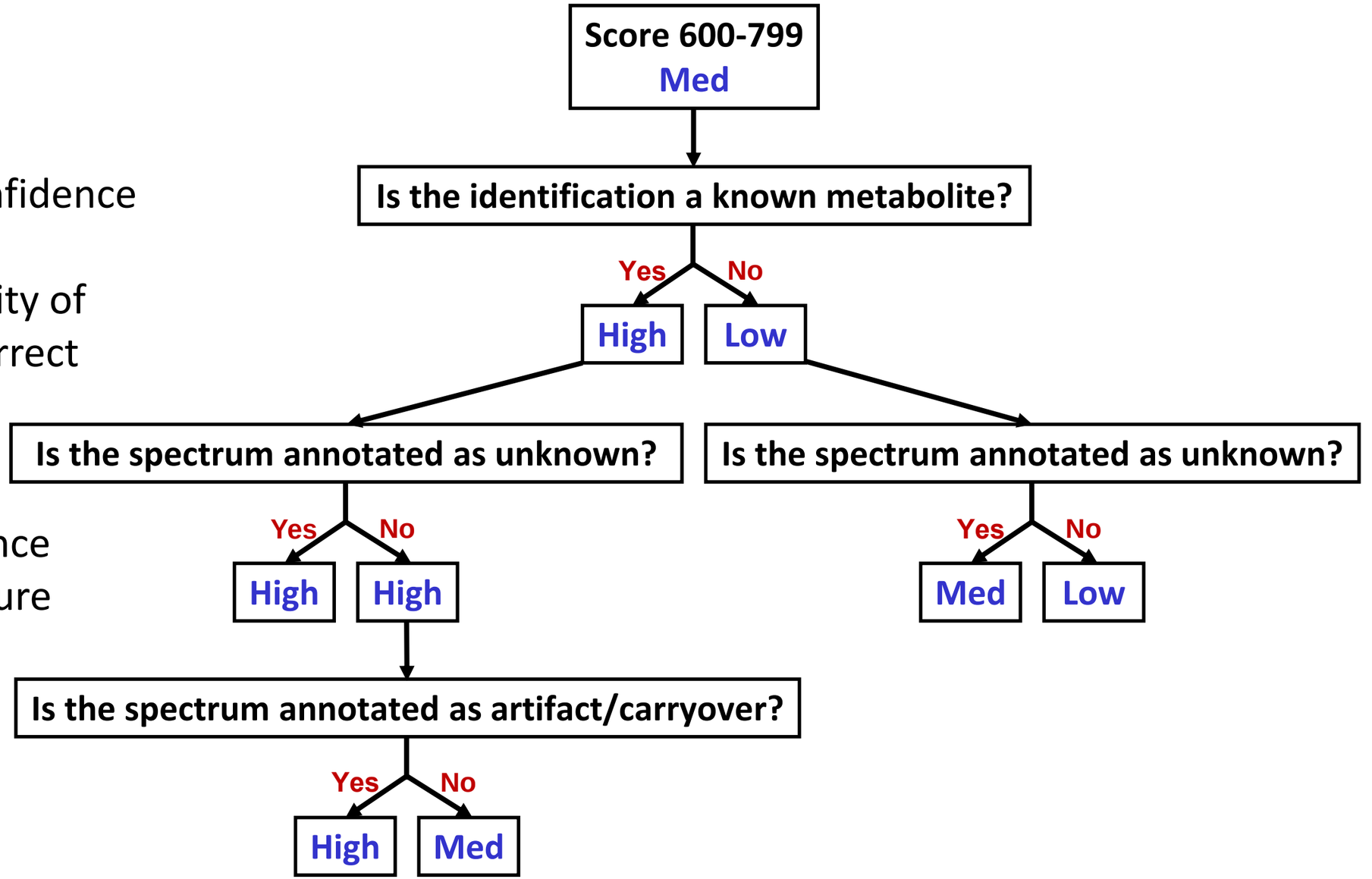


Examples of Annotations- EICs



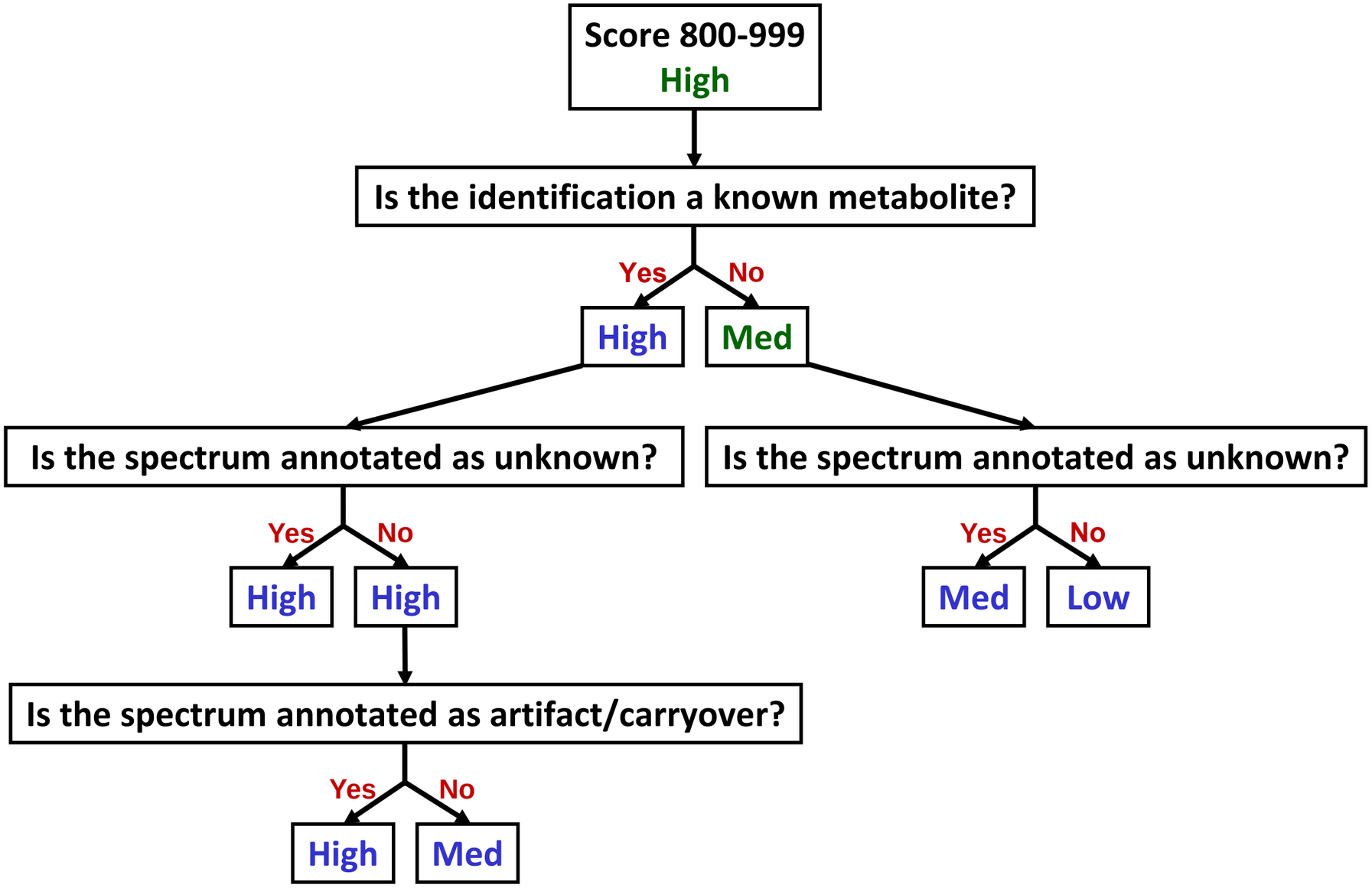
Assigning Confidence – Med Initial Confidence

- Developed a qualitative confidence level to provide further information about probability of the MS/MS match being correct
- Complementary to confidence levels in proposed in literature

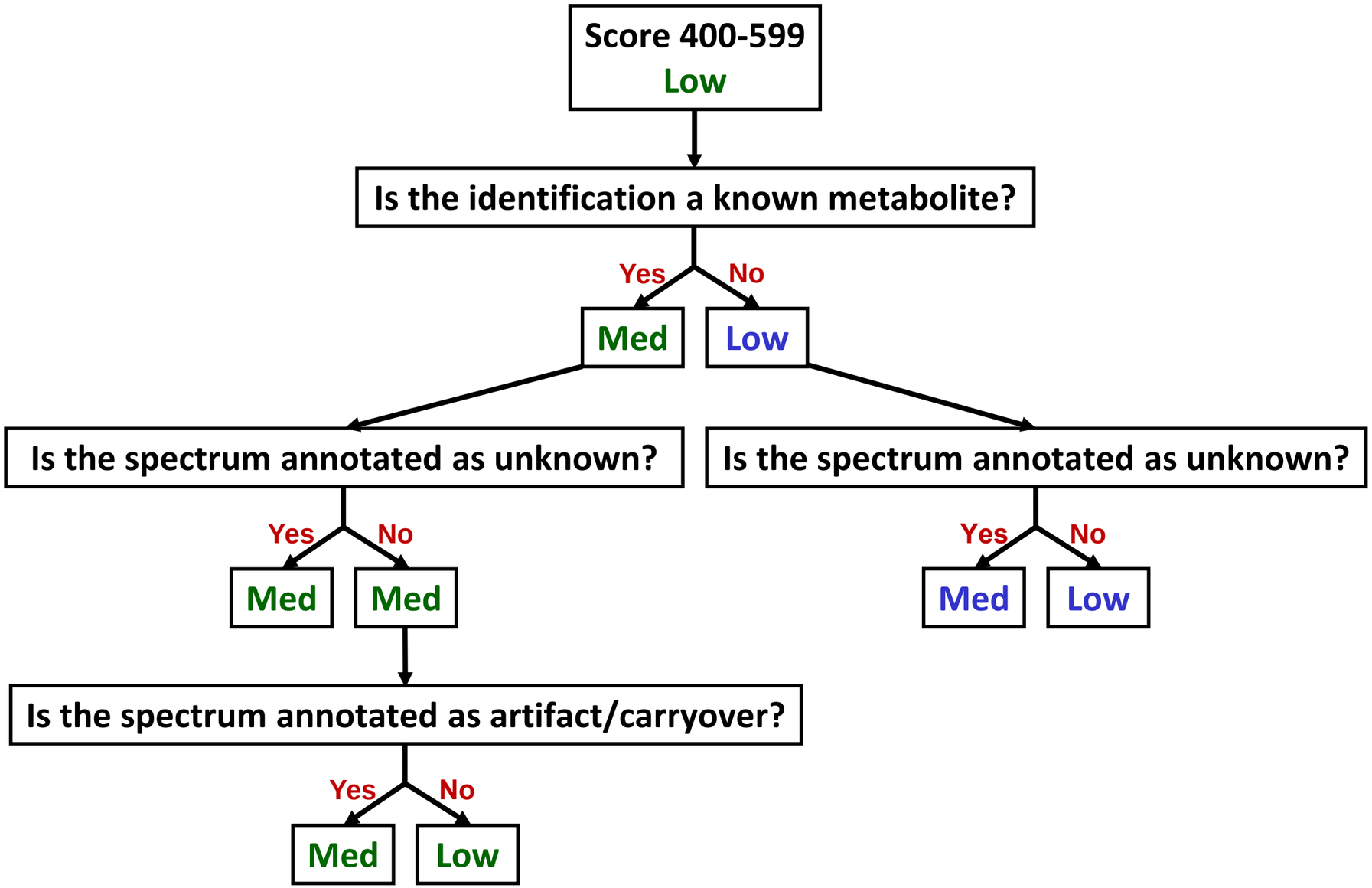


Confidence Levels in IDs- Schymanski, *et al. Environmental science & technology* 2014, 48 (4), 2097-8

Workflow for High Initial Confidence



Workflow for Low Initial Confidence



Summary

- Created a recurrent spectral library for CHO cell metabolites and media
- Improved MS/MS identification accuracy
- Developed a strategy to annotate unidentified spectra
- Developed strategy to assign confidence to MS/MS IDs
- Next Steps:
  - Make library available at https://chemdata.nist.gov/
  - Automate the annotation and assignment of confidence
  - Apply library to analysis of samples
- Manuscript is being prepared

Thanks for your Attention

National Institute of Standards and Technology
U.S. Department of Commerce

Mass Spectrometry Data Center

CHO Cell Growth:

Biomolecular Labeling Lab (BL2) at the Institute for Bioscience and Biotechnology Research (IBBR)

- Zvi Kelman
- Renae Preston
- Lila Kashi

Have questions or want to talk?

BOOTH 616 – Today from 2:30-5:00 pm

Tytus Mak – WOA 3:50 pm, here