Stemina
BIOMARKER DISCOVERY

From Cell Culture to Patients

De Novo Metabolite Chemical Structure Determination

Paul R. West Ph.D.

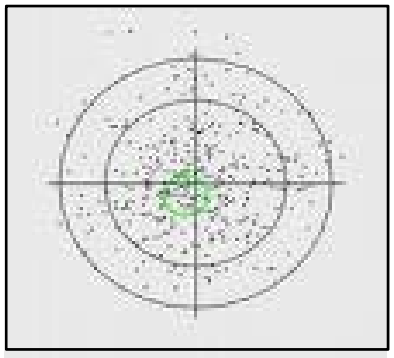
Stemina Biomarker Discovery, Inc.

INTRODUCTION

As a part of non-targeted metabolomic analysis, when peaks in the mass spectra are thought to be from a potential metabolite and no structure annotations arise after exhaustive database searches, how do you determine and validate its chemical structure?

The primary focus of this presentation will be metabolite chemical structure determination using LC-MS, accurate mass and MS-MS.

It is VERY important that as many false positive results as possible from the metabolite discovery process are removed!



Shotgun Metabolomics (nontargeted)

...after you have discovered a potential metabolite by accurate mass LC-MS and you have the following data for the "mass feature":

1. Retention time
2. Accurate mass
3. Abundance

NEXT STEPS

Stemina BIOMARKER DISCOVERY

METABOLOMICS WORKFLOW

LC-MS Steps toward Structure Determination:

1. Utilize the accurate mass of the feature (re-measure if needed).
2. Calculate possible formulae (eliminate unrealistic formulae).
3. Search a metabolite database for mass and formulae.
4. Search a chemical structure database (ChemSpider etc.).
5. Acquire and interpret (time consuming!) an MS-MS spectrum.
6. Obtain reference standards – perform MS, MS-MS and compare.
7. Perform other types of analysis:

LC-MS Data Acquisition

Other LC-MS Analyses

- Run the samples using ESI(-) (AVOID TFA!) or a different ionization (i.e. APCI).
- Use a diode array (UV-Vis) detector in-line between the HPLC and the MS, a tool for structure functional groups, i.e. aromaticity. Use lambda max data.
- Don't be afraid to try pseudo-MSMS (in source fragmentation), particularly on a TOF. This can work very well, but ONLY if the chromatographic separation is sufficient to isolate the compound and the base peak is from that compound. It works in ESI(-) too.
- During validation experiment, inject 3 different concentrations of the reference standard. The overlapped exact mass EICs at that retention time will then show peaks with different abundances and EIC peaks from other isobaric species will not.

SAMPLE ANALYSIS TIPS AND TRICKS

Orthogonal Analytical Techniques

Use ALL of the Tools in your Toolbox!

1. Other types of MS and MS-MS (i.e. GC-MS – low MW, derivatize).
2. Use Diode Array UV-Vis in-line on LC-MS.
3. Collect fractions (can use mass-directed fraction collection) then
   a. Perform NMR analysis OR even better -
   b. Perform LC-NMR (if there's enough compound).

      LC-NMR Sensitivity is now μg with cryo-probes.
4. Perform IR. Yes, this can be diagnostic for functional groups and there are A LOT of spectra in IR libraries.

ANALYTICAL TECHNIQUES

Technique	Sensitivity/ reproducibility	Structure & Formula Accuracy	Analyte Versatility	Speed	Notes
GC-LRMS	Excellent	Good	Fair	Fast	Only volatile or derivatized cpds. MW < 500
GC-HRMS	Excellent	Very Good	Fair	Fast	
LC-LRMS	Very Good/Good	Good	Good	Medium	Poor mass accuracy
LC-HRMS	Very Good/ Good	Very Good	Good	Medium	Accurate mass
NMR	Fair/Very Good	Excellent	Very Good	Slow	Collect mg fractions and purify sample
LC-NMR	Fair/Good	Excellent	Very Good	Very Slow	Need mg of sample
IR	Fair/Good	Fair	Good	Slow	Slow prep. Functional groups only
UV Vis	Variable	Poor	Fair	Fast	Good inline with LC-MS

Mass Spectrometers Capable of MS-MS

Technique	Sensitivity	Mass Accuracy	Notes
FT-MS^n	Very Good	1 ppm	MS^n High res (exact mass) precursor selection but can have insufficient collision energy
QTOF-MS-MS	Very Good	10 ppm	High res (exact mass) precursor selection but can have insufficient collision energy
Orbitrap-MS-MS	Very Good	5 ppm	Some m/z limitations, depend on model
Q-Trap MS^n	Very Good	0.2 Da	MS^n but product ions only @ *m/z* > 1/3 Precursor
Q3-MS-MS	Excellent	0.2 Da	MRM, SRM quantitation

MS-MS Data Acquisition

- For unknowns, allow a wider window for the precursor to capture isotopes.
  However, must narrow the precursor window if other isobaric species are co-eluting that are more abundant than ion of interest otherwise you will get fragment ions for the wrong precursor molecule.

- $(M+Na)^+$ ions are often more stable than $(M+H)^+$ and do not fragment well, or show a lot of non-diagnostic MS-MS fragment ions.

- If you are not sure which adduct a spectral peak is, $(M+H)^+$, $(M+Na)^+$ etc., run ESI(-). This will sometimes give $(M-H)^-$ for a compound that only shows $(M+Na)^+$ in ESI(+). Compare the exact masses of the adducts to be sure. Then perform ESI(-) MS-MS.

LC-MS Data Interpretation

Know your chromatography!

Understanding the chromatography can help to eliminate certain compounds.

If there are known compounds in the data – examine their structures to get an idea of what types of compounds elute when.

Some types of compounds such as quaternary ammonium species, keto-enols etc. can show a wider elution profile

Example: For reverse-phase C18 chromatography, a hydrophilic compound will usually elute early in the gradient, so if the metabolite elutes late, it is may be more hydrophobic.

Exact Mass Analysis

The "exact mass" of a compound is usually determined by using a mass spectrometer capable of “high resolution” mass measurements (HRMS), measuring the ions to the 4th or 5th decimal place. Mass accuracy is usually defined in terms of relative mass error (RME) in units of ppm (parts per million).

$$\text{Relative Mass Error (ppm)} = \frac{\text{Theoretical } m/z - \text{Observed } m/z}{\text{Theoretical } m/z} \times 10^6 \text{ ppm}$$

Exact mass measurements can significantly narrow down the possible number of molecular formulae. The more accurate the measurement, the fewer possible number of formulae.

Exact Mass Analysis – Reality Check

ppm error is a sliding scale and is molecular weight dependent

Correlating Formulae with ppm error

When valence rules and typical limits on the numbers of C,H,N,O are considered:

- At nominal mass 118; no formulae are closer together than 34 ppm.
- At nominal mass 500, there are only five formulae that have a neighbor candidate less than 5 ppm away.
- At an ion mass of 750.4, there are 626 formulae that have a neighbor formula less than 5 ppm away.

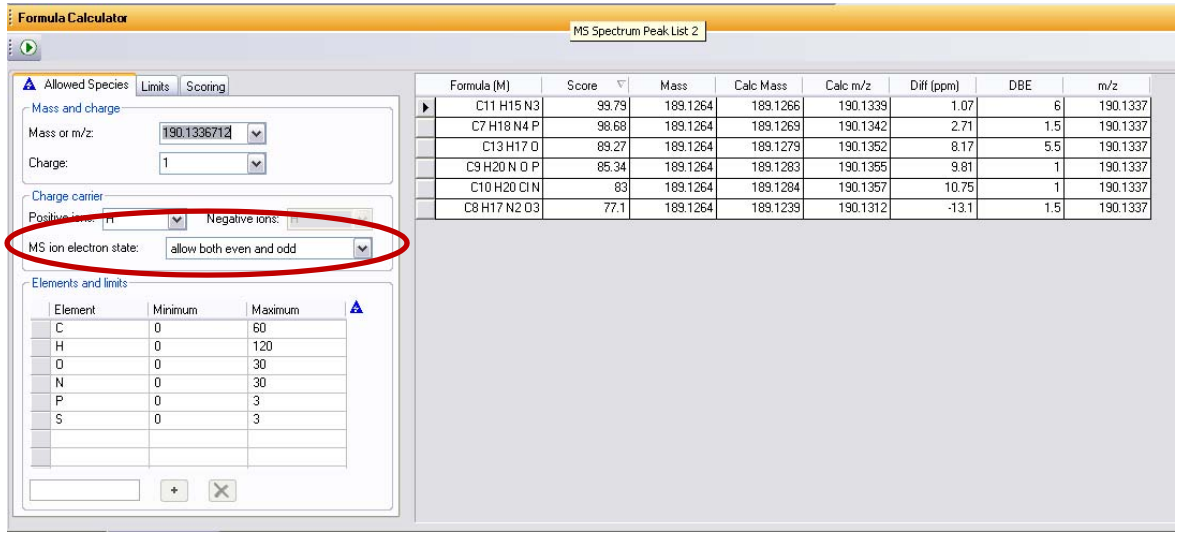
MS Data Interpretation – Odd and Even e^- Ions

Knowing which type is present can be diagnostic!

- Molecular Ions: EI produces odd electron ions. ESI and APCI usually produce even electron ions such as $(M+H)^+$.
- If you see an odd electron ion in an ESI spectrum, it could be a fragment ion or pre-charged (i.e. quaternary nitrogen compound).

$C_xH_yN_zO_n$

$X - 0.5y - 0.5z + 1$
= a whole number for odd electron ions.



MS data interpretation

- **Nitrogen Rule:**

  For neutral organic compounds containing C,H,N,O,S,P:

  Odd number of nitrogens = odd MW.

  Even nitrogens = even MW.

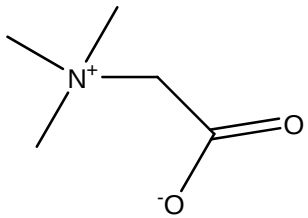
- **Ring Double Bond Equivalents** (RDBE) or
- **Double Bond Equivalents** (DBE) are calculated from valence values of elements contained in a calculated formula and is a useful tool for structure determination but MUST BE USED WITH CAUTION! IT CAN BE WRONG!

MS data interpretation – Isobaric species

Be careful with annotated compounds from databases. Even just the molecular ion can be confusing depending on how structure is drawn. Most databases show neutral molecules:

Same formula – different structures.
For positive ion analysis, a proton must be added to each of these to create a cation.

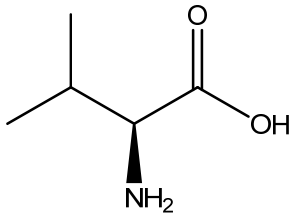
Betaine



Chemical Formula: $C_5H_{11}NO_2$
Exact Mass: 117.0790

Quaternary nitrogen (neutral as drawn)

Valine



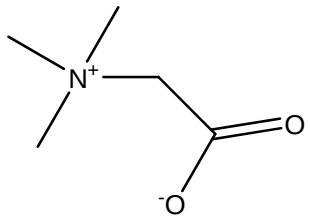
Chemical Formula: $C_5H_{11}NO_2$
Exact Mass: 117.0790

Neutral

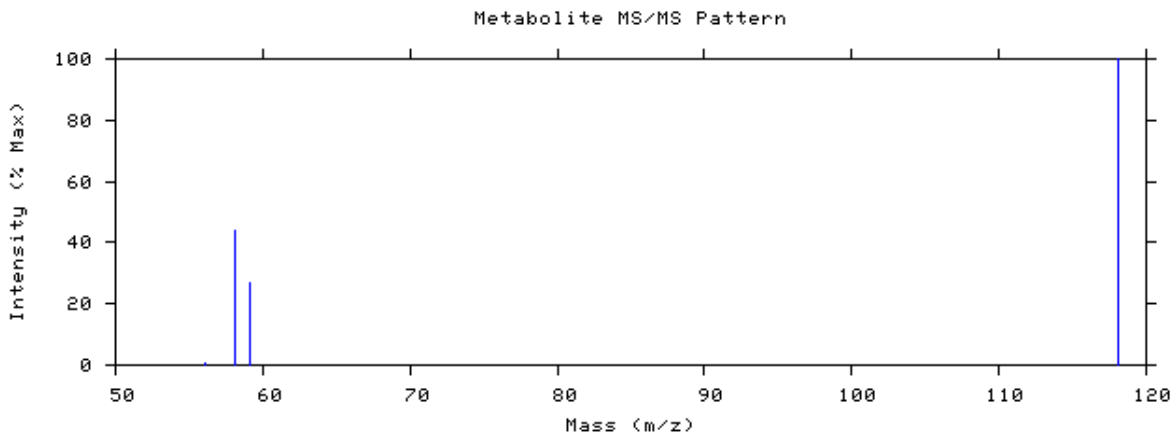
MS-MS data interpretation – Isobaric species (same mass)

MS cannot distinguish the compounds, but MS-MS can:

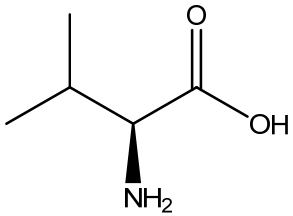
Betaine



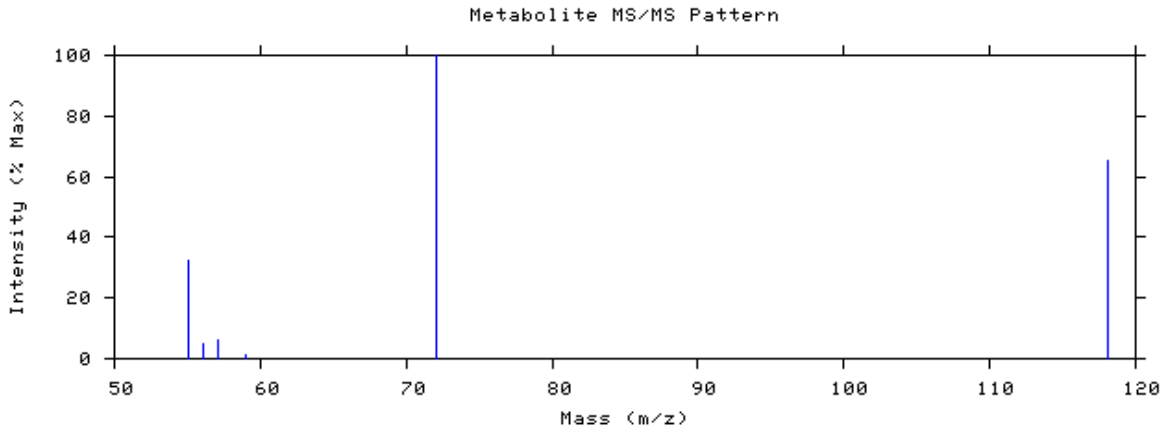
Chemical Formula: $C_5H_{11}NO_2$
Exact Mass: 117.0790



Valine



Chemical Formula: $C_5H_{11}NO_2$
Exact Mass: 117.0790



MS data interpretation

Molecular Ion Determination – Mass Differences

- between two peaks suspected of being molecular ions.

Positive ions

	M	M+	$(M+H)^+$	$(M+NH_4)^+$	$(M+Na)^+$	$(M+K)^+$
M		.0005	1.0073	18.0338	22.9893	38.9632
M+	.0005		1.0068	18.0333	22.9888	38.9627
$(M+H)^+$	1.0073	1.0068		17.0260	21.9820	37.9559
$(M+NH_4)^+$	18.0338	18.0333	17.0260		4.9555	20.9294
$(M+Na)^+$	22.9893	22.9888	21.9820	4.9555		15.9739
$(M+K)^+$	38.9632	38.9627	37.9559	20.9294	15.9739	

Negative ions

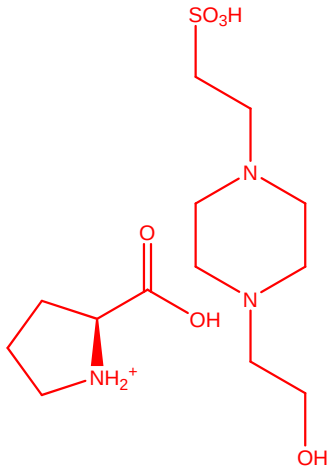
	M	$(M-H)^-$	$(M+Na-2H)^-$	$(M+HCOO)^-$	$(M+CH_3COO)^-$
M		-1.0073	20.9747	44.9982	59.0139
$(M-H)^-$	-1.0073		23.9966	21.9820	60.0392
$(M+Na-2H)^-$	20.9747	23.9966		24.0235	38.0392
$(M+HCOO)^-$	44.9982	46.0055	24.0235		14.0157
$(M+CH_3COO)^-$	59.0139	60.0392	38.0392	14.0157	

MS Data Interpretation

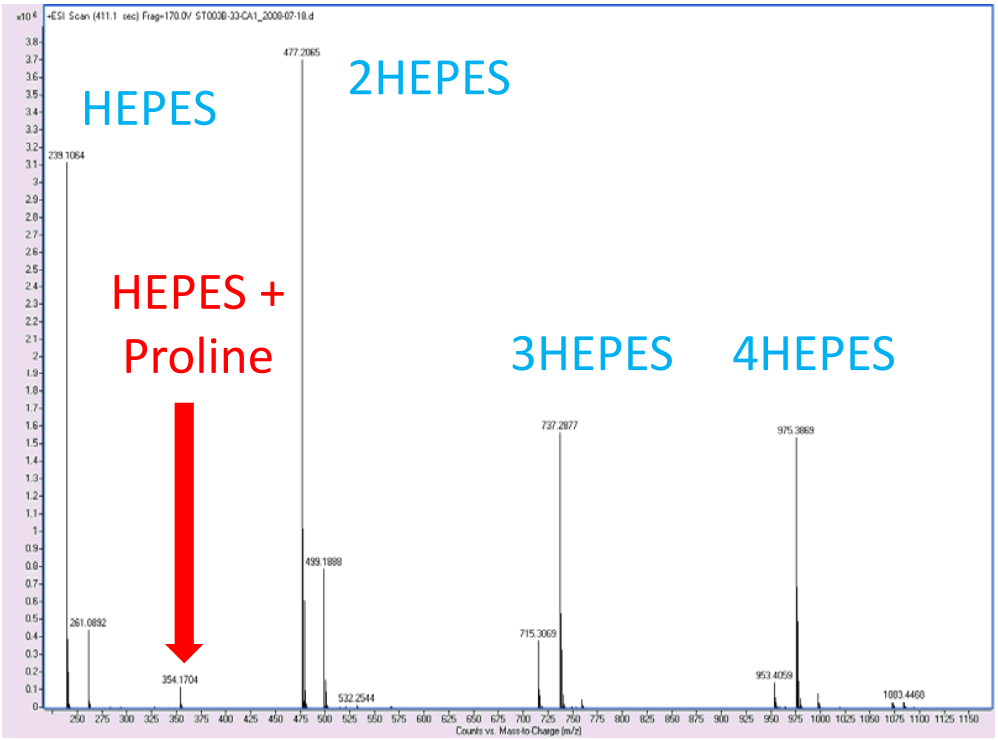
Molecular Ion Determination –Adducts

Strong impurities such as buffers (such as HEPES below) and polymers can form dimers, trimers etc. and also adducts with small molecules.

HEPES-Pro Cluster



Chemical Formula: $C_{13}H_{28}N_3O_6S^+$
Exact Mass: 354.1693



Stemina BIOMARKER DISCOVERY

MS data interpretation

Molecular Ion Determination - Mass Defect

Mass defect - difference between the measured *m/z* and integer mass.

Examples: Relative to integer mass 396.0000:

m/z 396.2734 shows positive mass defect

m/z 395.9223 shows a negative mass defect

Generally more useful for MW < ~800.

The majority of organic compounds show a positive mass defect.

Negative mass defects:	Positive mass defects:
S = 31.972	N = 14.003
O = 15.995	H = 1.0073
P = 30.974	

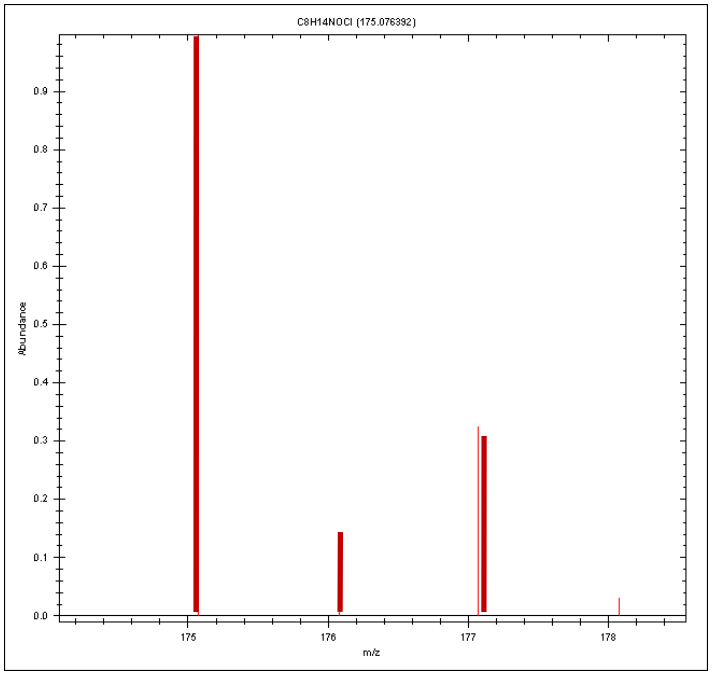
MS Data Interpretation

Molecular Ion Determination – Isotope Pattern

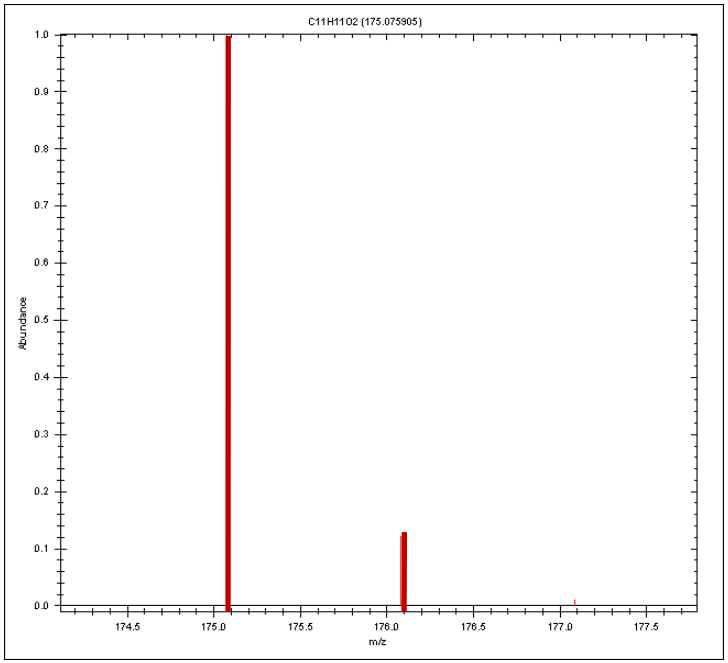
Isotope patterns can be very useful for formula determination.

Two formulae close in mass can show very different isotope patterns:

$C_8H_{14}NOCl$ = 175.0764



$C_{11}H_{11}O_2$ = 175.0759



MS Data Interpretation

Molecular Ion Determination – Common Contaminants

Know your LC-MS samples and systems.
You may see these peaks in the mass spectra.

Compile tables and/or a database of contaminants:

Plasticizers ..... Phthalates, BPA

Silicone..... Poly-dimethylsiloxane – often used in vial septa

Polymers..... PEG, PPG

Buffers..... HEPES, TWEEN 20, TWEEN 80

Example Contaminant Table

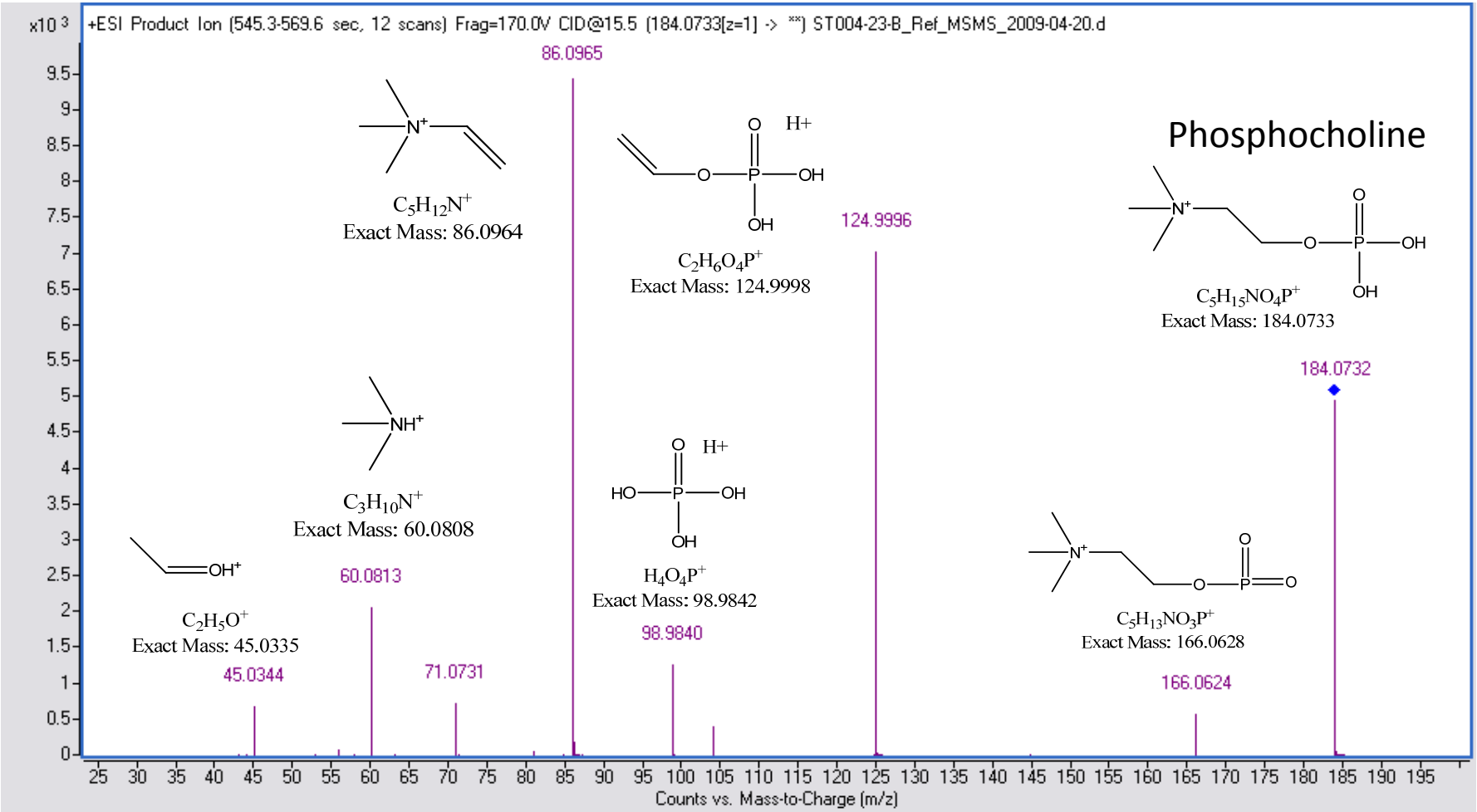
Compound	M	n	Monomer Unit	$(M+H)^+$	$(M+NH_4)^+$	$(M+Na)^+$	$(M+K)^+$	$(M-H)^-$	$(M+Na-2H)^-$	$(M+HCOO)^-$
HEPES (buffer)	238.0987	1	238.0987	239.106	256.1325	261.088	277.0619	237.0914	259.0734	283.0969
	476.1974	2		477.2047	494.2312	499.1867	515.1606	475.1901	497.1721	521.1956
	714.2961	3		715.3034	732.3299	737.2854	753.2593	713.2888	735.2708	759.2943
Polydimethyl-siloxane (silicone – material used in vial septa)	236.109	1	74.0188	237.1163	254.1428	259.0983	275.0722	235.1017	257.0837	281.1072
	310.1278	2		311.1351	328.1616	333.1171	349.091	309.1205	331.1025	355.126
	384.1466	3		385.1539	402.1804	407.1359	423.1098	383.1393	405.1213	429.1448
	458.1654	4		459.1727	476.1992	481.1547	497.1286	457.1581	479.1401	503.1636
	532.1842	5		533.1915	550.218	555.1735	571.1474	531.1769	553.1589	577.1824

MS-MS Data Interpretation

- Use exact mass differences between peaks in the MS-MS spectrum to determine formulae of the corresponding neutral losses to help figure out fragments of the structure.

- Can usually assume that precursor adduct will be the same adduct as its product ions, BUT NOT ALWAYS. Occasionally a $(M+Na)^+$ precursor can fragment and give a $(M+H)^+$ product ion.

MS-MS Data Interpretation - Example Spectrum



Stemina BIOMARKER DISCOVERY

MS and MS-MS Data Interpretation - Software

ChemDraw – calculates exact masses for structures and fragments. I use it as a tool for figuring out fragmentation of a structure or building structures from MS-MS product ion exact masses. Certain versions can also show MS fragments for a given structure.

MathSpec (www.mathspec.com)– Dan Sweeney's new software for de novo structure determination from exact mass fragmentation.

Mass Frontier (Thermo) - is designed to aid in the interpretation of MSn spectra. Given a molecular structure, Mass Frontier will predict fragmentation patterns and pathways, but structure must be known.

ACD MS Manager (ACDLabs) – provides tools for MS-MS interpretation and a spectral database.

MS and MS-MS Data Interpretation - References

"Interpretation of Mass Spectra" by Fred McLafferty.

This is THE BIBLE (and almost as old) on interpretation of MS data. Primarily written for the interpretation of EI data, many of the rules hold true for ESI MS-MS data as well.

Metabolite Databases on the Web

Search the metabolite databases for mass, formulae and structure.

- METLIN http://metlin.scripps.edu/
- Fiehn GC-MS Metabolomics Library http://fiehnlab.ucdavis.edu/
- HMDB http://www.hmdb.ca/search/spectra?type=ms_search
- KEGG http://www.genome.jp/kegg/ligand.html#rdb
- MassBank http://www.massbank.jp/index.html

MS-MS Spectral Libraries on the Web

Search for MS-MS spectra matches. These can also provide invaluable insights for correlating fragmentation with functional group types.

- METLIN http://metlin.scripps.edu/
- HMDB http://www.hmdb.ca/search/spectra?type=ms_search
- MassBank http://www.massbank.jp/index.html

Search accurate mass MS-MS spectra of metabolites.

Search by neutral loss or m/z difference!

Detailed experimental conditions for the original spectra.

NOTE: The manuals are in Japanese.

CONCLUSIONS

De Novo metabolite structure determination can be a very time consuming and difficult process – requiring days to even months to accomplish!

It is VERY important that false positive results in the metabolite discovery process are detected and removed as early and often as possible to avoid spending significant instrumentation and time resources attempting to discern the structure of a meaningless component!

GOOD LUCK
AND THANK YOU!