Welcome!

Mass Spectrometry meets Cheminformatics
Tobias Kind and Julie Leary
UC Davis

Course 7: Concepts for LC-MS

Class website: CHE 241 - Spring 2008 - CRN 16583
Slides: http://fiehnlab.ucdavis.edu/staff/kind/Teaching/
PPT is hyperlinked – please change to Slide Show Mode
General LC-MS data processing for small molecules

Confirm with MS/MS or MS^n fragmentation
Deconvolution and evaluation of LC-MS data

LC-MS run 40 minutes
C8 column, Agilent-UPLC
Chlamydomonas extract

FT-ICR-MS mass spectrum
MS1 @ 50,000 resolving power
Check charge state = 1
756.57 represents [M+H]^+
Peak Picking with ACD/SpecManager 9.0
Processing of LC-MS data - use of MassFrontier
Deconvolution and evaluation of LC-MS data

Example with HighChem Mass Frontier

LC-MS detected compound
Marked with blue triangle

LC-MS detected compound
Marked with blue triangle
141 peaks extracted

Extracted MS1 peak
Library search useless
(only single peak)
Adduct removal and detection during ESI-LC-MS runs

\[ [M+H]^+ = 756.577 \]
\[ M = 755.5627 \]

### Ion name | Ion mass | Charge | Mult | Mass | Result: | Reverse:
---|---|---|---|---|---|---
**1. Positive ion mode**
M+2H | M/2 + 1.007276 | 2+ | 0.5 | 1.007276 | 378.788638 | 377.277724
M+H+NH4 | M/2 + 9.520550 | 2+ | 0.5 | 9.520550 | 387.301912 | 368.764450
M+H+Na | M/2 + 11.998247 | 2+ | 0.5 | 11.998247 | 389.779609 | 366.286753
**M+H** | M + 1.007276 | 1+ | 1 | 1.007276 | 756.570000 | 755.562724
M+NH4 | M + 18.033823 | 1+ | 1 | 18.033823 | 773.596547 | 738.536177
M+Na | M + 22.989218 | 1+ | 1 | 22.989218 | 778.551942 | 733.580782

**2. Negative ion mode**
M-3H | M/3 - 1.007276 | 3- | 0.33 | -1.007276 | 250.846965 | 253.197276
M-2H | M/2 - 1.007276 | 2- | 0.5 | -1.007276 | 376.774086 | 379.292276
M-H | M - 1.007276 | 1- | 1 | -1.007276 | 754.555448 | 757.577276
M+Na-2H | M + 20.974666 | 1- | 1 | 20.974666 | 776.537390 | 735.595334
M+Cl | M + 34.969402 | 1- | 1 | 34.969402 | 790.532126 | 721.600598

Download Adduct-Calculator [Link]
**Problem:** Is this the pure mass spectrum or from overlapping peaks? Is it M+H or M+Na or any of the 40 other adducts?

Example data from crocin standard mixture (expected MW: 976.965)
Adduct removal and detection during LC-MS runs

<table>
<thead>
<tr>
<th>Ion name</th>
<th>Ion mass</th>
<th>Charge</th>
<th>Mult</th>
<th>Mass</th>
<th>Result:</th>
<th>Reverse:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Positive ion mode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M+2H</td>
<td>M/2 + 1.007276</td>
<td>2+</td>
<td>0.5</td>
<td>1.007276</td>
<td>427.672721</td>
<td>437.152724</td>
</tr>
<tr>
<td>M+H+NH4</td>
<td>M/2 + 9.520550</td>
<td>2+</td>
<td>0.5</td>
<td>9.520550</td>
<td>436.185995</td>
<td>428.639450</td>
</tr>
<tr>
<td>M+H+Na</td>
<td>M/2 + 11.998247</td>
<td>2+</td>
<td>0.5</td>
<td>11.998247</td>
<td>438.663692</td>
<td>426.161753</td>
</tr>
<tr>
<td>M+H</td>
<td>M + 1.007276</td>
<td>1+</td>
<td>1</td>
<td>1.007276</td>
<td>854.338166</td>
<td>875.312724</td>
</tr>
<tr>
<td>M+NH4</td>
<td>M + 18.033823</td>
<td>1+</td>
<td>1</td>
<td>18.033823</td>
<td>871.364713</td>
<td>858.286177</td>
</tr>
<tr>
<td>M+Na</td>
<td>M + 22.989218</td>
<td>1+</td>
<td>1</td>
<td>22.989218</td>
<td>876.320108</td>
<td>853.330782</td>
</tr>
<tr>
<td>2. Negative ion mode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-3H</td>
<td>M/3 - 1.007276</td>
<td>3-</td>
<td>0.33</td>
<td>-1.007276</td>
<td>251.182724</td>
<td>253.197276</td>
</tr>
<tr>
<td>M-2H</td>
<td>M/2 - 1.007276</td>
<td>2-</td>
<td>0.5</td>
<td>-1.007276</td>
<td>377.277724</td>
<td>379.292276</td>
</tr>
<tr>
<td>M-H</td>
<td>M - 1.007276</td>
<td>1-</td>
<td>1</td>
<td>-1.007276</td>
<td>755.562724</td>
<td>757.577276</td>
</tr>
<tr>
<td>M+Na-2H</td>
<td>M + 20.974666</td>
<td>1-</td>
<td>1</td>
<td>20.974666</td>
<td>777.544666</td>
<td>735.595334</td>
</tr>
<tr>
<td>M+Cl</td>
<td>M + 34.969402</td>
<td>1-</td>
<td>1</td>
<td>34.969402</td>
<td>791.539402</td>
<td>721.600598</td>
</tr>
</tbody>
</table>

Adduct can be removed before or after formula generation.
For good isotopic pattern matching remove adduct after formula generation.
7 Golden Rules apply LEWIS and SENIOR check (adduct needs to be removed)
Formula Generation from accurate mass measurement

<table>
<thead>
<tr>
<th>Compound</th>
<th>MW</th>
<th>PPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHN$<em>5$O$</em>{13}$P$_{15}$</td>
<td>755.563348</td>
<td>1.1 ppm</td>
</tr>
<tr>
<td>CHN$<em>{11}$P$</em>{19}$</td>
<td>755.543156</td>
<td>25.9 ppm</td>
</tr>
</tbody>
</table>

Apply **Seven Golden Rules** for correct molecular formulas
Apply heuristic and mathematical and chemometric rules for filtering elemental compositions
Isotopic Pattern generator from formula

Example from MWTWIN
Is usually included in every LC-MS software
Isotopic pattern equally important as accurate mass

Experimental result

\[
A+1 = 47.44\% \\
A+2 = 10.92\%
\]

Abundances for all molecular formulae

We can discard all other results outside the error box.
Current box reflect +/- 10% error.
Problems during LC-MS peak detection and MS deconvolution

Multiple peaks detected – Solution: adjust deconvolution settings

Mass spectra not clean – Solution: manual peak extraction

Not enough peaks detected – Solution: increase signal noise (S/N) settings

Finding optimum settings is:
• non-trivial and can change in different matrices
• can be evaluated on standards and quality check mixtures
• can be obtained by self-sharpening algorithms
UPLC-FT-MS data extraction with MassFrontier

Approach: generate molecular formula using **Seven Golden Rules**;
find matching isomers in molecular databases;
confirm possible matches by in-silico fragmentation (usually impossible);
Seven Golden Rules – generate possible molecular formulas

5 formula candidates left with 30 ppm mass accuracy and 10% isotopic abundances. These are candidates with good isotopic pattern match. These 5 were found in PubChem.

- **C42H78NO8P** - 1 isomer hit
- **C42H77NO10** - 1 isomer hit
- **C39H73N5O9** - 0 isomer hit
- **C43H82NO7P** - 2 isomers found
- **C43H73N5O6** - 2 isomers found
- **C45H77N3O6** - 1 hit found
- **C45H69N7O3** - 1 hit found

**Scan speed problem:**
Due to poor ion statistics only few scans are collected.
Mass accuracy and isotopic abundance accuracy are bad.
Structural isomer lookup example in ChemSpider
In-silico fragmentation with MassFrontier using fragmentation library of 20,000 mechanisms from literature
In-silico fragmentation with MassFrontier

Experimental peaks m/z 478.45 and 496.46 were detected in MS/MS spectrum
In-silico fragmentation should match the experimental fragmentation.
In-silico - using a computer library of 20,000 fragmentation rules from the MS literature

Possible solution (2 fragments match)
**Example discussion:**

Fragment at m/z 236 not explained; molecular ion may be wrong; Substance can be potential new compound
Must be confirmed by NMR or external standard

**General problems:**

Best approach is to generate MS and MS^n and MS^e mass spectral libraries
Adduct removal is a problem
Building target lists is always good (know what to expect)
Focus on certain substance classes only
Focus on single compound only

Substance must be known for in-silico approach
Fragmentation rules must be captured for in-silico approach
In-silico approaches work best for peptides, carbohydrates, lipids (due to known and stable fragmentations)
The Last Page - What is important to remember:

Always use peak picking and mass spectral deconvolution for LC-MS data

Apply accurate mass, accurate isotopic abundances together for formula generation
Make use of high resolving power whenever possible

Use MS/MS data and mass spectra from different ionization voltages
Use existing MS/MS libraries or create your own MSn tree libraries

Use molecular isomer databases for obtaining possible structure candidates
Confirm if possible with MSn data or other possible filter constraints
Tasks (36 min):

(1) Download and install one tool from
http://www.ms-utils.org/wiki/pmwiki.php/Main/SoftwareList

(2) Download sample data from Open Data repositories and
report your experience with the selected software (1) on the sample data
Reading List (44 min):

Hardware and Software Challenges for the Near Future: Structure Elucidation Concepts via Hyphenated Chromatographic Techniques

QUALITATIVE AND QUANTITATIVE MASS SPECTRAL ANALYSIS

Congruent Strategies for Carbohydrate Sequencing. 3. OSCAR: An Algorithm for Assigning Oligosaccharide Topology from MSn Data

Data Reduction of Isotope-resolved LC-MS Spectra

Comparative LC-MS: A landscape of peaks and valleys
Links:

Used for research: (right click – open hyperlink)

Mass Spectrometry in Life Science: Technology and Data-Evaluation

Managing and Exploring Large Data Sets Generated by Liquid Separation - Mass Spectrometry

Metabolomics Technologies applied to the Identification of Compounds in Plants

Of general importance for this course:

http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/Structure_Elucidation/