Mass Spectrometry
Outline

• Motivation
• Mass of molecules
• Mass spectrometer
• Bottom-up proteomics experiment
Technology Overview

Biology sample → Separated Proteins → Mass Spectrometry

What’re in the sample?

Mass Spectral Data
Why Bother? – An Example

HER2-positive breast cancer is a breast cancer that tests positive for a protein called human epidermal growth factor receptor 2 (HER2), which promotes the growth of cancer cells.

“HER2-positive breast cancers tend to be more aggressive than other types of breast cancer. They're also less responsive to hormone treatment. However, treatments that specifically target HER2 are very effective.”
• Personalized
• Predictive
• Preventative
• Participatory
The Mission of Our Research

Biology sample → Separated Proteins

Proteins

What’s in the sample?

Answer this question.

Mass Spectrometry

Data

Answer this question.
Outline

• Motivation
• Mass of molecules
• Mass spectrometer
• Bottom-up proteomics experiment
Chemical Composition of Living Matter
27 of 92 natural elements are essential.

Elements in biomolecules (organic matter):

H, C, N, O, P, S

These elements represent approximately 92% of dry weight.

Organic Matter
Organized in "building blocks"

- amino acids
- polypeptides (proteins)
- monosaccharides
- starch, glycogen
- nucleic acids
- DNA, RNA
### Mass (Weights) of Atoms and Molecules

<table>
<thead>
<tr>
<th>Nominal mass (Exact mass)</th>
<th>Nominal mass (Exact mass)</th>
<th>Percent abundance</th>
<th>Average mass</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C</strong> 12 (12.00000)</td>
<td></td>
<td>98.9%</td>
<td>12.00115</td>
</tr>
<tr>
<td>13 (13.00335)</td>
<td></td>
<td>1.1%</td>
<td></td>
</tr>
<tr>
<td><strong>H</strong> 1 (1.00783)</td>
<td></td>
<td>99.98%</td>
<td>1.008665</td>
</tr>
<tr>
<td>2 (2.0140)</td>
<td></td>
<td>0.02%</td>
<td></td>
</tr>
<tr>
<td><strong>O</strong> 16 (15.99491)</td>
<td></td>
<td>99.8%</td>
<td>15.994</td>
</tr>
<tr>
<td>18 (17.9992)</td>
<td></td>
<td>0.02%</td>
<td></td>
</tr>
<tr>
<td><strong>N</strong> 14 (14.00307)</td>
<td></td>
<td>99.63%</td>
<td>14.0067</td>
</tr>
<tr>
<td>15 (15.00011)</td>
<td></td>
<td>0.37%</td>
<td></td>
</tr>
</tbody>
</table>

**Monoisotopic mass**

**Isotopes**
Amino Acids

• There are 20 amino acids. All have the same basic structure but with different side chains:

```
H-N-C-C-OH
```

• Examples:

  - Glycine, or Gly, or G
  - Arginine, or Arg, or R
Peptides and Proteins

Glycine, or Gly, or G

Arginine, or Arg, or R

Atom mass:
C=12.000
H=1.008
N=14.003
O=15.995

N-terminal

C-terminal

residue

peptide bonds
# Amino Acid Mass Table

<table>
<thead>
<tr>
<th>Name</th>
<th>3-letter code</th>
<th>1-letter code</th>
<th>Monoisotopic Mass</th>
<th>Average Mass</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>71.03711</td>
<td>71.08</td>
<td>C₃H₅NO</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>156.10111</td>
<td>156.2</td>
<td>C₆H₁₂N₄O</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>114.04293</td>
<td>114.1</td>
<td>C₄H₆N₂O₂</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>Asp</td>
<td>D</td>
<td>115.02694</td>
<td>115.1</td>
<td>C₄H₅NO₃</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>103.00919</td>
<td>103.1</td>
<td>C₃H₅NOS</td>
</tr>
<tr>
<td>Glutamic Acid</td>
<td>Glu</td>
<td>E</td>
<td>129.04259</td>
<td>129.1</td>
<td>C₅H₇NO₃</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
<td>128.05858</td>
<td>128.1</td>
<td>C₅H₈N₂O₂</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>57.02146</td>
<td>57.05</td>
<td>C₂H₄NO</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>137.05891</td>
<td>137.1</td>
<td>C₆H₇N₃O</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>113.08406</td>
<td>113.2</td>
<td>C₆H₁₁NO</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>113.08406</td>
<td>113.2</td>
<td>C₆H₁₁NO</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>128.09496</td>
<td>128.2</td>
<td>C₆H₁₂N₂O</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
<td>131.04049</td>
<td>131.2</td>
<td>C₅H₉NOS</td>
</tr>
<tr>
<td>Phenyalanine</td>
<td>Phe</td>
<td>F</td>
<td>147.06841</td>
<td>147.2</td>
<td>C₉H₉NO</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
<td>97.05276</td>
<td>97.12</td>
<td>C₅H₇NO</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
<td>87.03203</td>
<td>87.08</td>
<td>C₃H₅NO₂</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
<td>101.04768</td>
<td>101.1</td>
<td>C₄H₇NO₂</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
<td>186.07931</td>
<td>186.2</td>
<td>C₁₁H₁₆N₂O</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
<td>163.06333</td>
<td>163.2</td>
<td>C₉H₉NO₃</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
<td>99.06841</td>
<td>99.13</td>
<td>C₅H₉NO</td>
</tr>
</tbody>
</table>
Peptide Mass

\[ m(A_1) + m(A_2) + m(A_3) + m(A_4) + m(H_2O) \]
Outline

• Motivation
• Mass of molecules
• Mass spectrometer
• Bottom-up proteomics experiment
Mass Spectrometer

• Given molecules (small or large), ionize them into ions (molecules with electrical charge), measure the mass to charge ratios (m/z).

• Some ionization techniques let $z=1$. Others can derive $z$ from data computationally.
A Mass Spectrum

peak list

... 2789.22  3597.0
2790.22  5018.0
2791.23  4406.0
2792.23  2868.0
2793.23  1234.0
...
...
Three Basic Components

• Ionizer
  – provide electric charges to the molecules
  – MALDI, **ESI**

• Mass Analyzer
  – Separates the ions according to m/z
  – Magnetic Sector, Iontrap, **TOF**, Quadrupole, FT, **Orbitrap**

• Detector
  – Detect the separated ions
  – Electron multiplier, FT

• In addition, how to fragment the parent ion is also important for today’s proteomics
Ionization

- MALDI (Matrix Assisted Laser Desorption/Ionization)
- ESI (Electrospray Ionization)

Koichi Tanaka  
John Fenn

Nobel prize in Chemistry, 2002
Electrospray Ionization: Formation of Charged Droplets

Formation of multiply charged ions
ESI History

- A.B. from Berea College in his new hometown.
- 1940: Ph.D. from Yale University.
- 1962: He joined the Yale University faculty.
- 1987: he reached the mandatory retirement age (70).
- University-mandated move to smaller laboratory space.
- Started to work on ESI.
- 1994: Fenn joined Virginia Commonwealth University.
- The patent rights to ESI became the subject of a legal case between Yale University and Fenn.
- 2005: Yale was awarded over one million dollars and partial patent rights to the technique.
• Adjusting DC voltage allows different m/z ions to pass. (Mass filter)
• The complete spectrum is obtained by scanning whole range.

Mass range 10-4,000 Da
Mass Analyzer (2) – TOF

- Time of Flight.

Time of flight is proportional to $\sqrt{m/z}$
Mass Analyzer (3) – Orbitrap

Moving ions are trapped around an electrode.

Fourier transform to convert the time-field signal to frequencies.

By shaping the electrode appropriately, ions also move left and right. Left-right frequency proportional to $\sqrt{m/z}$.

Notes:
- All-mass detection
- Noise equiv. to 20 ions (1 sec)
Photos

Sciex Tripletof

Thermo Orbitrap
Each peak corresponds to a type of (peptide) ions in the sample. This tells us the mass-to-charge of those peptides, but not the sequence.
Tandem Mass Spectrometry

• Can combine different mass analyzers. E.g. Q-Tof.
MS/MS of A Peptide

\begin{align*}
&b_1 \quad A \text{ NELLLNVK} \\
&b_2 \quad A N \text{ ELLLNVK} \\
&b_3 \quad A N E \text{ LLLNVK} \\
&b_4 \quad A N E L \text{ LLNVK} \\
&b_5 \quad A N E L L \text{ LLNVK} \\
&b_6 \quad A N E L L L \text{ NVK} \\
&b_7 \quad A N E L L L N \text{ VK} \\
&b_8 \quad A N E L L L L N V \text{ K}
\end{align*}

\begin{align*}
&\gamma_1 \quad 147.1 \\
&\gamma_2 \quad 246.2 \\
&\gamma_3 \quad 360.2 \\
&\gamma_4 \quad 473.3 \\
&\gamma_5 \quad 586.4 \\
&\gamma_6 \quad 699.5 \\
&\gamma_7 \quad 828.5
\end{align*}

\begin{align*}
&b_{2-NH3} \quad 298.1 \\
&b_{3-NH3} \quad 330.2 \\
&b_{4-NH3} \quad 411.2 \\
&b_{5-NH3} \quad 524.3 \\
&b_{7-NH3-H2O} \quad 810.5
\end{align*}
Possible Ways to Interpret MS/MS Data

1. Database search
   - protein DB
   - peptides

2. de novo sequencing
   - peptides

MS/MS Spectra
Outline

• Motivation
• Mass of molecules
• Mass spectrometer
• Bottom-up proteomics experiment
Intact Proteins v.s. Peptides

• An intact protein is usually too large for MS/MS.
  – Although some people are doing this now.
• There are enzymes that digest proteins into short peptides.
Protease

• A **protease** is any **enzyme** that conducts **proteolysis**.
• Or: a protease breaks protein in water.
• Trypsin digests at site [KR] \[^{^P}\].
Liquid Chromatography

• LC or HPLC (High Performance LC) is used to further separate the complex peptide mixture, before they enter MS.

• In proteomics, LC separates peptides according to their hydrophobicity.
  – (from the Attic Greek hydro, meaning water, and phobos, meaning fear)
HPLC

• Read wikipedia or a biochemistry book if you’re interested. But for the purpose of this course:

• Given a mixture of peptides, HPLC will elute different peptides at different time.
  – Elution time or retention time (RT).
  – And hence separate the peptides.

• Separation isn’t perfect.
  – Each peptide elutes at a time window (as an LC peak).
  – Multiple peptides’ LC peaks can overlap.
LC-MS/MS

MS or survey scan

![LC-MS/MS Diagram](image-url)
(1) A mixture of protein is digested into peptides with added enzyme (usually trypsin).

(2) The resulting peptides are separated with liquid chromatography (LC). Different peptides elute at different time. The separation may not be perfect.

(3) The spectrometer scans the peptide ions at a particular time and obtains a profile MS scan. Each peak in the MS spectrum supposedly corresponds to a peptide.

(4) The spectrometer selects a peak (a peptide ion) with the first mass analyzer, fragments it and produces an MS/MS scan with the second mass analyzer.
Proteins

Protein A
PAKGTIRHIHGCDKRGAPWPAS...

Protein B
MSERNHLREIIGNERV......

Protein C
LSIMQDKDYSASFIS......

Peptides

PAK
MSER
LSIMQDK
HIHGCDK
EIIGNERV
SIMQMDYSASFIS
......

MS/MS

Peptide sequencing