Tandem Mass Spectrometry and Protein Sequencing
Genomic and Proteomic Research

• Genomic research:
  – Research on genomes.
  – Defines potential contributors to cellular functions

• Proteomic Research:
  – Research on the comprehensive group of proteins expressed by a given cell or tissue.
  – Expressed genome defines actual contributors.
Central tools in proteome research

• SDS- PAGE
  – 1D and 2D gel electrophoresis

• Protein sequencing
  – Edman degradation
  – Mass Spectrometric technique (Tandemmm and Maldi)

• Identification techniques
  – Immunologic technique (Western blotting)
Western blotting

• Identifies a protein based on pattern of antibody recognition.
• Presumptive and require the availability of a suitable antibody
• The confidence of the identification is limited by problems with the specificity of the antibody
Tandem Mass-Spectrometry

• Cut bands of interest directly from the gel, digest, analyze and identify

• Advantages:
  – High sensitivity: femtomole level
  – Rapid speed of analysis
  – Large amounts of information generated in each experiment
  – Ability to characterize post-transitional modifications
Mass Spectrometers

• Separate and measures ions based on their mass-to-charge (m/z) ratio.

• Operate under high vacuum (keeps ions from bumping into gas molecules)

• Key specifications are resolution, mass measurement accuracy, and sensitivity.

• Several kinds exist: for bioanalysis, quadrupole, time-of-flight (TOF) and ion traps are most used.
Mass Spectrometer Schematic

High Vacuum System

Inlet
- Sample Plate
- Target
- HPLC
- GC
- Solids probe

Ion Source
- MALDI
- ESI
- IonSpray
- FAB
- LSIMS
- EI/CI

Mass Filter
- TOF
- Quadrupole
- Ion Trap
- Mag. Sector
- FTMS

Detector
- Microch plate
- Electron Mult.
- Hybrid Detec.

Data System
- PC's
- UNIX
- Mac

Turbo pumps
- Diffusion pumps
- Rough pumps
- Rotary pumps
Different Ionization Methods

• Electron Impact (EI - Hard method)
  – small molecules, 1-1000 Daltons, structure
• Fast Atom Bombardment (FAB – Semi-hard)
  – peptides, sugars, up to 6000 Daltons
• Electrospray Ionization (ESI - Soft)
  – peptides, proteins, up to 200,000 Daltons
• Matrix Assisted Laser Desorption (MALDI-Soft)
  – peptides, proteins, DNA, up to 500 kD
Comparison of Ionisation Techniques

Molec. Weight

Non Polar

Highly Polar

EI

APCI

TSP

ESI
Soft Ionization

- Soft ionization techniques keep the molecule of interest fully intact
- Electro-spray ionization first conceived in 1960’s by Malcolm Dole but put into practice in 1980’s by John Fenn (Yale)
- MALDI first introduced in 1985 by Franz Hillenkamp and Michael Karas (Frankfurt)
- Made it possible to analyze large molecules via inexpensive mass analyzers such as quadrupole, ion trap and TOF
Ion Sources

- Electrospray ionization
- Atmospheric pressure chemical ionization
- Atmospheric pressure photoionization
- Matrix-assisted laser desorption ionization (MALDI)
Electrospray

• Softest ionisation technique
• Best for polar non-volatile compounds (proteins, peptides, nucleic acids, Pharmaceuticals, natural products)
• Coupled to LC at a flow range of 2-1000 ul/min, nanospray (10 nL/min – 2 uL/min)
• Ions are ejected from charged vapour droplets to gas phase producing M+H+ or M - H- ions
• Can produce multiply charged ions allowing determination of high molecular weight proteins
• Not very tolerant of non-volatile salts
Electrospray Ionization

- Sample dissolved in polar, volatile buffer (no salts) and pumped through a stainless steel capillary (70 - 150 µm) at a rate of 10-100 µL/min
- Coupled to LC at a flow range of 2-1000 ul/min, nanospray (10 nL/min – 2 uL/min)
- Strong voltage (3-4 kV) applied at tip along with flow of nebulizing gas causes the sample to “nebulize” or aerosolize
- Aerosol is directed through regions of higher vacuum until droplets evaporate to near atomic size (still carrying charges)
Electrospray ion source

- Nebulizer gas
- Solvent spray
- Dielectric capillary entrance
- Ions
- Heated nitrogen drying gas
Ion Sources make ions from sample molecules (Ions are easier to detect than neutral molecules.)

Electrospray ionization:

- High voltage applied to metal sheath (~4 kV)
- Sample Inlet Nozzle (Lower Voltage)
- Partial vacuum
- Pressure = 1 atm
  Inner tube diam. = 100 um
- Sample In solution
- $\text{N}_2$ gas
- Charged droplets
- $\text{MH}^+$
- $\text{MH}_2^+$
- $\text{MH}_3^+$

Electrospray ionization makes ions from sample molecules by applying high voltage to the metal sheath, creating charged droplets.
Atmospheric Pressure Ionisation

- Most important and widely used LC / MS technique
- API two types
  - Electrospray
  - Atmospheric Pressure Chemical Ionisation
- > 99% new LC/MS use API source
- Ionisation takes place outside vacuum region
Atmospheric Pressure Ionization (API)

General Schematic of an API interface

Atmospheric Pressure Interface

Intermediate Vacuum Interface

High Vacuum Interface
Atmospheric Pressure Chemical Ionization (APCI)

• Used for wide range polarity of compounds
• HPLC eluent (up to 2ml/min flow rate) is vaporized at up to 600 C
• The Corona discharge needle ionizes solvent molecules.
• A combination of collisions and charge transfer reactions between the solvent and the analyte results in the transfer of a proton to form either M+H+ or M-H- ions
• Compounds can thermally degrade
• Multiply charged ions rare
• More tolerant to salts
APCI Ion Source

- HPLC inlet
- Nebulizing gas
- Nebulizer (sprayer)
- Vaporizer (heater)
- Corona discharge needle
- Drying gas
- Capillary
APPI Ion Source
Advantages of API

- Soft ionisation (gives the molecular weight)
- Sensitive (low pg amounts routinely)
- Robust, simple, run routinely 24 hr/day
- Wide range of flow rates (nanospray to analytical)
- Wide range of applications (drugs, proteins)
- Wide range of industries
MALDI Ionization

• Absorption of UV radiation by chromophoric matrix and ionization of matrix

• Dissociation of matrix, phase change to super-compressed gas, charge transfer to analyte molecule

• Expansion of matrix at supersonic velocity, analyte trapped in expanding matrix plume (explosion/”popping”)

![Diagram of MALDI Ionization process]
Matrix Assisted Laser Desorption Ionisation

- Sample is ionized by bombarding sample with laser light
- Sample is mixed with a UV absorbant matrix (sinapinic acid for proteins, 4-hydroxycinnaminic acid for peptides)
- Light wavelength matches that of absorbance maximum of matrix so that the matrix transfers some of its energy to the analyte (leads to ion sputtering)
- Coupled to Time of Flight MS
- Not coupled to LC
- High mass range achievable
- Calibrants may be external or included in sample
- Reproducibility issues
Resolution & Resolving Power

• Width of peak indicates the resolution of the MS instrument
• The better the resolution or resolving power, the better the instrument and the better the mass accuracy
• Resolving power is defined as: $\frac{\Delta M}{M}$
• $M$ is the mass number of the observed ($\Delta M$) is the difference between two masses that can be separated
How is mass resolution calculated?

\[ R = \frac{M}{\Delta M} \]

FWHM = \Delta M
What if the resolution is not so good?

At lower resolution, the mass measured is the average mass.
What is the advantage of using high resolution mass spectrometry?

The advantage of high resolution measurement is to eliminate chemical background of the same nominal mass but different accurate mass and, therefore to increase the signal to noise ratio and the sensitivity respectively.

Better selectivity is obtained by applying high resolution in case of isobaric compounds, i.e. two compounds of same nominal mass but different accurate mass. With low resolution only a combined Spectral result is obtained under Product ion conditions. With high resolution separate detection and therefore separate isolation and MS/MS spectra are obtained.

Example:

<table>
<thead>
<tr>
<th></th>
<th>R= 1000</th>
<th>R= 5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound mass1: 372.351 Da</td>
<td>OVERLAPPING</td>
<td>SEPARATION</td>
</tr>
<tr>
<td>Compound mass2: 372.421 Da</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How is mass defined?

Assigning numerical value to the intrinsic property of “mass” is based on using carbon-12, $^{12}\text{C}$, as a reference point.

One unit of mass is defined as a Dalton (Da).

One Dalton is defined as $1/12$ the mass of a single carbon-12 atom.

Thus, one $^{12}\text{C}$ atom has a mass of 12.0000 Da.
Isotopes

- Most elements have more than one stable isotope.

For example, most carbon atoms have a mass of 12 Da, but in nature, 1.1% of C atoms have an extra neutron, making their mass 13 Da.

- Why do we care?

Mass spectrometers can “see” isotope peaks if their resolution is high enough.

If an MS instrument has resolution high enough to resolve these isotopes, better mass accuracy is achieved.
Masses in MS

- Monoisotopic mass is the mass determined using the masses of the most abundant isotopes.
- Average mass is the abundance weighted mass of all isotopic components.
Stable isotopes of most abundant elements of peptides

<table>
<thead>
<tr>
<th>Element</th>
<th>Mass</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>1.0078</td>
<td>99.985%</td>
</tr>
<tr>
<td></td>
<td>2.0141</td>
<td>0.015</td>
</tr>
<tr>
<td>C</td>
<td>12.0000</td>
<td>98.89</td>
</tr>
<tr>
<td></td>
<td>13.0034</td>
<td>1.11</td>
</tr>
<tr>
<td>N</td>
<td>14.0031</td>
<td>99.64</td>
</tr>
<tr>
<td></td>
<td>15.0001</td>
<td>0.36</td>
</tr>
<tr>
<td>O</td>
<td>15.9949</td>
<td>99.76</td>
</tr>
<tr>
<td></td>
<td>16.9991</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>17.9992</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Mass spectrum of peptide with 94 C-atoms (19 amino acid residues)

“Monoisotopic mass”

1981.84

No $^{13}$C atoms (all $^{12}$C)

1982.84

One $^{13}$C atom

1983.84

Two $^{13}$C atoms
When the isotopes are clearly resolved the **monoisotopic mass** is used as it is the most accurate measurement.
Average mass corresponds to the centroid of the unresolved peak cluster.

When the isotopes are not resolved, the centroid of the envelope corresponds to the weighted average of all the isotope peaks in the cluster, which is the same as the average or chemical mass.
Different Mass Analyzers

• Quadrupole Analyzer (Q)
  – Low (1 amu) resolution, fast, cheap

• Time-of-Flight Analyzer (TOF)
  – No upper m/z limit, high throughput

• Ion Trap Mass Analyzer (TRAP)
  – Good resolution, all-in-one mass analyzer

• Ion Cyclotron Resonance (FT-ICR)
  – Highest resolution, exact mass, costly
Quadrupole Mass Analyzer

- Consists of four parallel rods arranged in a square.
- The analyte ions are directed down the center of the square.
- Voltages applied to the rods generate electromagnetic fields.
- These fields determine which mass-to-charge ratio of ions can pass through the filter at a given time.
- The simplest and least expensive mass analyzers
Figure 3. Time-of-flight analysis of ions of various masses, each with a single charge. For clarity and simplicity, this shown in a linear time-of-flight mass spectrometer that does not have an ion mirror.
Time-of-flight (TOF) mass analyzer

- A uniform electromagnetic force is applied to all ions at the same time, causing them to accelerate down a flight tube.
- Lighter ions travel faster and arrive at the detector first.
- So the mass-to-charge ratios of the ions are determined by their arrival times.
- Have a wide mass range and can be very accurate in their mass measurements.
Mass Spec Equation (TOF)

\[ \frac{m}{z} = \frac{2Vt^2}{L^2} \]

- \( m \) = mass of ion
- \( z \) = charge of ion
- \( L \) = drift tube length
- \( t \) = time of travel
- \( V \) = voltage
**Ion Trap**

- Consists of a ring electrode and two end caps.
- Ions entering the chamber are “trapped” there by electromagnetic fields.
- Another field can be applied to selectively eject ions from the trap.
- Able to perform multiple stages of mass spectrometry without additional mass analyzers.
- Ions stored by RF & DC fields.
- Scanning field can eject ions of specific m/z.
- Advantages
  - MS/MS/MS.....
  - High sensitivity full scan MS/MS.
Fourier transform-ion cyclotron resonance (FT-ICR) Mass Analyzer

• Ions entering a chamber are trapped in circular orbits by powerful electrical and magnetic fields.

• When excited by a radio-frequency (RF) electrical field, the ions generate a time-dependent current.

• This current is converted by Fourier transform into orbital frequencies of the ions which correspond to their mass-to-charge ratios.

• Can perform multiple stages of mass spectrometry without additional mass analyzers.

• Have a wide mass range and excellent mass resolution.

• Very expensive mass analyzers.
Tandem Mass Spectrometry
Mass Spectrometry

Steps: 1. Mass Analysis
      2. Collision (Fragmentation)
      3. Mass Analysis

Collisional Activation-
1. Impart kinetic energy to an ion by collision with an inert gas.
2. Kinetic energy is converted to internal energy in the ion.
3. Fragmentation of the unstable ion.

Precursor Ion + Inert Gas \[\rightarrow\] Product Ions
\[(N_2, Ar, He)\]
Collision-Induced Dissociation

- CID (collosion-induced) or CAD (chemically activated)
- Generate
  - Molecular ions $M^+$ or $M^-$
  - Protonated molecules $[M + H]^+$
  - Simple adduct ions $[M + Na]^+$
  - Ions representing simple losses such as the loss of a water $[M + H - H_2O]^+$
MS vs. MS/MS

MS

Separation
Identification

GC
HPLC
CE

MS/MS

Inlet
Ionize
Mass Analyze
Detect

Inlet
Ionize
Mass Analyze
Fragment
Mass Analyze
Detect

Collision Cell
MS1
MS2
Mass Spectrometry

Sample Inlet

Ionization & Adsorption of Excess Energy

Fragmentation (Dissociation)

Mass Analysis

Detection
What is Tandem MS?

- Uses 2 (or more) mass analyzers in a single instrument
  - One purifies the analyte ion from a mixture using a magnetic field.
  - The other analyzes fragments of the analyte ion for identification and quantification.
Tandem in Space

Mass Analyzer 1 - Collision Cell - Mass Analyzer 2

Ex: Quadrupole - Collision Cell - Quadrupole
Quadrupole - Collision Cell - Time of Flight

Collision Cells: RF-only quadrupoles, hexapoles, or octapoles

Collision Cell Functions:
1. Fragment selected ion
2. Contain all product ions
   *i.e.* all m/z
3. Transmit product ions to 2nd mass analyzer
Tandem in Space MS

- Triple Quatrupole
- Hybrid Instruments
  - ESI-QTOF
    - Electrospray ionization source + quadrupole mass filter + time-of-flight mass analyzer
  - MALDI-QTOF
    - Matrix-assisted laser desorption ionization + quadrupole + time-of-flight mass analyzer
A Triple Stage Quadrupole Mass Analyzer

Finnigan TSQ

ESI Probe

Square Rod Ion Transmission to Analytical Quads

Q0  Q1  Q2  Q3

Q2 is Non-Linear Collision Cell

Hyperbolic, high precision quadrupoles

Electron Multiplier, Detection System
Q-TOF Mass Analyzer
Inside of Mass Spectrometer

Figure 1. Ion source, ion optics, and mass filter from the Agilent LC/MSD TOF, an API oa-TOF mass spectrometer
Tandem in Time - Ion Trap

- Ion traps are ion trapping devices that make use of a three-dimensional quadrupole field to trap and mass-analyze ions
- invented by Wolfgang Paul (Nobel Prize 1989)
- Offer good mass resolving power, and even MS\textsuperscript{n} capability.
Ion Trap Mass Analyzer

Single Ion Trap
1. Trap all m/z ions.
2. RF scan to eject all m/z except the targeted m/z.
3. Apply RF pulse to accelerate trapped ions and fragment ions via gas collisions.
4. Perform m/z scan of product ions.
Ion-Source CID with a Single and a Triple Quadrupole MS
Tandem MS

1. Tandem in Space- >1 mass analyzer
2. Tandem in Time-
   a. 1 mass analyzer only
   b. sequentially trap ions

Precursor ion mass analysis

Collision-induced fragmentation

Precursor ion selection by m/z

Product ion mass analysis
What is MS/MS?

Peptide mixture

1 peptide selected for MS/MS

Have only masses to start

MS/MS

The masses of all the pieces give an MS/MS spectrum
Interpretation of an MSMS spectrum to derive structural information is analogous to solving a puzzle.

Use the fragment ion masses as specific pieces of the puzzle to help piece the intact molecule back together.
Data Dependent Experiments

- Ion Traps can run MS and data dependent product ion scans
- TSQ has three different data dependent modes
  - Full scan MS triggered
  - Neutral loss triggered
  - Precursor ion triggered
MS-MS Scan Modes

A. Mass spectrum scan

Measure m/z:
No Collisions

B. Product ion scan

Product Ion Scan:
Peptide Sequencing

C. Precursor ion scan

Precursor Ion Scan:
Phosphorylated Peptides
(PO₃⁻ m/z = 79)
Product ion scans

• There are two types of product ion scans:

• Full Scan Product ion are used for qualitative applications to obtain structural information.

• Selected Reaction Monitoring Product ion scans are used for Quantitative target analysis.
What are product ion scans?

• Product ion scans also know as daughter ion scans
• Q1 is set to allow only the transmission of one m/z
• The parent ion collides with Argon gas in Q2 to create fragment or product ions
• Product ions are scanned through Q3
What are precursor ion scans?

- Precursor ion scans also known as parent ion scans
- Q3 is set to allow only a fragment ion of one m/z to pass
- Q1 is scanned
- The precursor ions collide with Argon gas in Q2 to create fragment or product ions
- Only those compounds which give that specific fragment ion are detected
Precursor ion scans are used for screening experiments where a group of compounds all give the same fragment ion.
What are neutral loss scans?

- Both Q1 and Q3 are scanned together.
- Q3 is offset by the neutral loss under investigation.
- The precursor ions collide with Argon gas in Q2 to create fragment ions.
- Only those compounds which give a fragment having that specific loss are detected.
Neutral loss scans

Neutral loss scans are used for screening experiments where a group of compounds all give the same loss.
Protein Sequencing

"Bottom-Up" Sequencing- (most common)
   a. Cleave protein into peptides.
   b. Send peptides into MS for sequencing

"Top-Down" Sequencing- (difficult but fast)
   a. Send intact protein into mass spec.
   b. Fragment & sequence

Why peptides instead of proteins?
   1. Increased stability
   2. Better solubility
   3. Greater sensitivity
   4. Easier to sequence if < 20 amino acids
   5. Fewer (usually <1) translational modifications/peptide
   5. Cheaper instrumentation (proteins require an FTICR for sequencing)
Protein Identification

• 2D-GE + MALDI-MS
  – Peptide Mass Fingerprinting (PMF)

• 2D-GE + MS-MS
  – MS Peptide Sequencing/Fragment Ion Searching

• Multidimensional LC + MS-MS
  – ICAT Methods (isotope labelling)
  – MudPIT (Multidimensional Protein Ident. Tech.)

• 1D-GE + LC + MS-MS

• De Novo Peptide Sequencing
Protein Sequencing By MS

Sample preparation/fractionation
- SDS-PAGE
- 2D-gel electrophoresis

Protein digestion
- Trypsin
- Lys-C
- Asp-N
- Glu-C

Peptide separation
- HPLC
- Ion exchange

Sample ionization
- Electrospray ionization
- MALDI

Mass spectrometry
- Quadrupole
- Time of flight
- Quadrupole ion traps
- FTICR

Mass spectrum

Data analysis
- PeptideSearch
- Sequest
- Mascot
Breaking Proteins into Peptides

protein → trypsin → peptides

MPSER GTDIMR PAKID

HPLC → To MS/MS
Proteolyzed Proteins Need Separation

Cleaved proteins yield a complex peptide mixture & must be separated prior to MS.

Separation Characteristics:
1. Typically reverse phase (hydrophobicity).
   May need multi-dimensional separation.
2. Remove contaminants *i.e.* detergents, salts
3. Reduce complexity but overlapping peaks OK
4. Couple directly to ESI/MS
   a. Elute in smallest possible volume
   b. Peak width of 10-60 s

Ex: μscale- HPLC, capillary electrophoresis, microfluidic chips
Tandem Mass Spectrometry

Scan 1707

Scan 1708
Peptide Mass Fingerprinting (PMF)
Protein Identification by Tandem Mass Spectrometry

MS/MS instrument

Database search
- Sequest
- de Novo interpretation
- Sherenga
Tandem Mass Spectrum

- Tandem Mass Spectrometry (MS/MS): mainly generates partial N- and C-terminal peptides
- Spectrum consists of different ion types because peptides can be broken in several places.
- Chemical noise often complicates the spectrum.
- Represented in 2-D: mass/charge axis vs. intensity axis
MS/MS (tandem MS)
Amino Acid Sequencing and Protein Identification
$\text{MS}^n$

- Ion source
- Mass selection of precursor ion (MS)
  - Fragmentation of precursor ion and analysis of fragment ions ($\text{MS}^2$)
  - Fragmentation of a selected fragment ion and analysis of the second generation fragment ions ($\text{MS}^3$)
- Second generation fragment ion
Peptide sequencing by MS/MS

(Standing 2003 *Curr. Opin. Struct. Biol.* 13, 595-601)
# Amino acid abbreviations

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Three-letter abbreviation</th>
<th>One-letter abbreviation</th>
<th>Amino acid</th>
<th>Three-letter abbreviation</th>
<th>One-letter abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>Asp</td>
<td>D</td>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
</tr>
<tr>
<td>Glutamic Acid</td>
<td>Glu</td>
<td>E</td>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>Valine</td>
<td>Val</td>
<td>V</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>Asparagine or</td>
<td>Asx</td>
<td>B</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>aspartic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>Glutamine or</td>
<td>Glx</td>
<td>Z</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>glutamic acid</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Peptide bonds

\[
\begin{align*}
R^1 & \quad H_3N-\text{CH-CH}-C-\text{OH} + H-N-\text{CH-CH}-\text{COO}^- \\
& \quad H_2O \xrightarrow{\text{H}_2O} H_3N-\text{CH-CH}-C-N-\text{CH-CH}-\text{COO}^-
\end{align*}
\]
Protein Backbone

\[ \text{Protein Backbone:} \quad \text{H...-HN-CH-CO-NH-CH-CO-NH-CH-CO...OH} \]

\[ \text{R}_{i-1} \quad \text{R}_i \quad \text{R}_{i+1} \]

\[ \text{N-terminus} \quad \text{AA residue}_{i-1} \quad \text{AA residue}_i \quad \text{AA residue}_{i+1} \quad \text{C-terminus} \]
Peptide Fragmentation by MS

- Peptides tend to fragment along the backbone.
- Fragments can also lose neutral chemical groups like NH$_3$ and H$_2$O.
Breaking Protein into Peptides and Peptides into Fragment Ions

- Proteases, e.g. trypsin, break protein into peptides.
- A Tandem Mass Spectrometer further breaks the peptides down into fragment ions and measures the mass of each piece.
- Mass Spectrometer accelerates the fragmented ions; heavier ions accelerate slower than lighter ones.
- Mass Spectrometer measures mass/charge ratio of an ion.
N- and C-terminal Peptides

G P F N A

G P F N A

G P F N A

G P F N A

N-terminal peptides

C-terminal peptides
Terminal peptides and ion types

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Mass (D) 57 + 97 + 147 + 114 = 415</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPFN</td>
<td></td>
</tr>
<tr>
<td>H2O</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Mass (D) 57 + 97 + 147 + 114 - 18 = 397</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPFN</td>
<td></td>
</tr>
<tr>
<td>without</td>
<td></td>
</tr>
<tr>
<td>H2O</td>
<td></td>
</tr>
</tbody>
</table>
N- and C-terminal Peptides

N-terminal peptides

C-terminal peptides

486 GPAFNNA

415 GPFPNA

301 GPFPNA

154 GPFPNA

57 GPFPNA

71 GPAFNNA

185 GPAFNNA

332 GPAFNNA

429 GPAFNNA
N- and C-terminal Peptides

486

N-terminal peptides
C-terminal peptides

415

301

154

57

71

185

332

429
Fragmentation in MS/MS produces sequence-specific fragment ions

- Peptides fragment along the amide backbone to produce sequence-specific fragment ions.
- Different patterns of sequence ions are produced from the different masses of the side chain groups.
- Ions containing the N-terminus are typically of type ‘a’ or ‘b’.
- Ions containing the C-terminus are known as ‘y’ ions.
- Tryptic peptides predominantly produce ions of type ‘y’ due to the presence of Arg or Lys at the C-terminus.

Use database search routines to identify or profile known proteins.

De novo sequencing requires interpretation of spectra either manually or by computer algorithm.
Residue masses of the amino acids

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>One-letter code</th>
<th>Residue mass (Da)</th>
<th>Immonium ion (m/z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>G</td>
<td>57.02</td>
<td>30</td>
</tr>
<tr>
<td>Alanine</td>
<td>A</td>
<td>71.04</td>
<td>44</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
<td>87.03</td>
<td>60</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
<td>97.05</td>
<td>70</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
<td>99.07</td>
<td>72</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
<td>101.05</td>
<td>74</td>
</tr>
<tr>
<td>Cystine</td>
<td>C</td>
<td>103.01</td>
<td>76</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
<td>113.08</td>
<td>86</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
<td>113.08</td>
<td>86</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>114.04</td>
<td>87</td>
</tr>
<tr>
<td>Aspartate</td>
<td>D</td>
<td>115.03</td>
<td>88</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
<td>128.06</td>
<td>101</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
<td>128.09</td>
<td>101</td>
</tr>
<tr>
<td>Glutamate</td>
<td>E</td>
<td>129.04</td>
<td>102</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
<td>131.04</td>
<td>104</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
<td>137.06</td>
<td>110</td>
</tr>
<tr>
<td>Oxidized Methionine</td>
<td>Mo</td>
<td>147.04</td>
<td>120</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
<td>147.07</td>
<td>120</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
<td>156.10</td>
<td>129</td>
</tr>
<tr>
<td>Carbamidomethylcysteine</td>
<td>C*</td>
<td>160.03</td>
<td>133</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
<td>163.06</td>
<td>136</td>
</tr>
<tr>
<td>Acrylocystein</td>
<td>C*</td>
<td>174.04</td>
<td>147</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
<td>186.08</td>
<td>159</td>
</tr>
</tbody>
</table>

*This table also includes the one-letter abbreviations commonly used when writing peptide sequences and the m/z of the immonium ions with the form NH$_2$=CHR$^+$*
Amino acids combinations that are equal to a single amino acid residue mass*

<table>
<thead>
<tr>
<th>Amino acid combination</th>
<th>Residue mass (Da)</th>
<th>Equivalent amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td>114</td>
<td>N</td>
</tr>
<tr>
<td>GA</td>
<td>128</td>
<td>Q, K</td>
</tr>
<tr>
<td>GV</td>
<td>156</td>
<td>R</td>
</tr>
<tr>
<td>GE</td>
<td>186</td>
<td>W</td>
</tr>
<tr>
<td>AD</td>
<td>186</td>
<td>W</td>
</tr>
<tr>
<td>SV</td>
<td>186</td>
<td>W</td>
</tr>
<tr>
<td>SS</td>
<td>174</td>
<td>C^a</td>
</tr>
</tbody>
</table>

*The single-letter amino acid codes are used in this table.
Neutral losses observed from ions with different amino acid compositions*

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Neutral loss</th>
<th>Amino acid</th>
<th>Neutral loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>loss</td>
<td></td>
<td>K</td>
<td>17</td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>E</td>
<td>18</td>
</tr>
<tr>
<td>G</td>
<td>-</td>
<td>M</td>
<td>48</td>
</tr>
<tr>
<td>S</td>
<td>18</td>
<td>H</td>
<td>-</td>
</tr>
<tr>
<td>P</td>
<td>-</td>
<td>Mo</td>
<td>64</td>
</tr>
<tr>
<td>V</td>
<td>-</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>T</td>
<td>18</td>
<td>R</td>
<td>17</td>
</tr>
<tr>
<td>C</td>
<td>34</td>
<td>C*</td>
<td>92</td>
</tr>
<tr>
<td>L/I</td>
<td>-</td>
<td>Y</td>
<td>-</td>
</tr>
<tr>
<td>N</td>
<td>17</td>
<td>C*</td>
<td>106</td>
</tr>
<tr>
<td>D</td>
<td>18</td>
<td>W</td>
<td>-</td>
</tr>
</tbody>
</table>

The consecutive loss of small neutral molecules in an energetically favored process in collisionally induced dissociation. The nature of the neutral that is lost is dependent on the amino acid composition of the product ion. In this table, the one-letter amino acid codes are used. The – designates that no neutral losses occur for that amino acid.
Look-up table for the m/z of the b$_2$-ion*

<table>
<thead>
<tr>
<th>G</th>
<th>A</th>
<th>S</th>
<th>P</th>
<th>V</th>
<th>T</th>
<th>C</th>
<th>L/I</th>
<th>N</th>
<th>D</th>
<th>Q/K</th>
<th>E</th>
<th>M</th>
<th>H</th>
<th>Mo/F</th>
<th>R</th>
<th>C*</th>
<th>Y</th>
<th>C$^a$</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>129</td>
<td>143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>145</td>
<td>159</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>155</td>
<td>169</td>
<td>185</td>
<td>195</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>157</td>
<td>171</td>
<td>187</td>
<td>197</td>
<td>199</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>159</td>
<td>173</td>
<td>189</td>
<td>199</td>
<td>201</td>
<td>203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>161</td>
<td>175</td>
<td>191</td>
<td>201</td>
<td>203</td>
<td>205</td>
<td>207</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L/I</td>
<td>171</td>
<td>185</td>
<td>201</td>
<td>211</td>
<td>213</td>
<td>215</td>
<td>217</td>
<td>227</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>172</td>
<td>186</td>
<td>202</td>
<td>212</td>
<td>214</td>
<td>216</td>
<td>218</td>
<td>228</td>
<td>229</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>173</td>
<td>187</td>
<td>203</td>
<td>213</td>
<td>215</td>
<td>217</td>
<td>219</td>
<td>229</td>
<td>230</td>
<td>231</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q/K</td>
<td>186</td>
<td>200</td>
<td>216</td>
<td>226</td>
<td>228</td>
<td>230</td>
<td>232</td>
<td>242</td>
<td>243</td>
<td>244</td>
<td>257</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>187</td>
<td>201</td>
<td>217</td>
<td>227</td>
<td>229</td>
<td>231</td>
<td>233</td>
<td>243</td>
<td>244</td>
<td>245</td>
<td>258</td>
<td>259</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>189</td>
<td>203</td>
<td>219</td>
<td>229</td>
<td>231</td>
<td>233</td>
<td>235</td>
<td>245</td>
<td>246</td>
<td>247</td>
<td>260</td>
<td>261</td>
<td>263</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>195</td>
<td>209</td>
<td>225</td>
<td>235</td>
<td>237</td>
<td>239</td>
<td>241</td>
<td>251</td>
<td>252</td>
<td>253</td>
<td>266</td>
<td>267</td>
<td>269</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mo/F</td>
<td>205</td>
<td>219</td>
<td>235</td>
<td>245</td>
<td>247</td>
<td>249</td>
<td>251</td>
<td>261</td>
<td>262</td>
<td>263</td>
<td>276</td>
<td>277</td>
<td>279</td>
<td>285</td>
<td>295</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>214</td>
<td>228</td>
<td>244</td>
<td>254</td>
<td>256</td>
<td>258</td>
<td>260</td>
<td>270</td>
<td>271</td>
<td>272</td>
<td>285</td>
<td>286</td>
<td>288</td>
<td>294</td>
<td>304</td>
<td>313</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C$^*$</td>
<td>218</td>
<td>232</td>
<td>248</td>
<td>258</td>
<td>260</td>
<td>262</td>
<td>264</td>
<td>274</td>
<td>275</td>
<td>276</td>
<td>289</td>
<td>290</td>
<td>292</td>
<td>298</td>
<td>308</td>
<td>317</td>
<td>321</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>221</td>
<td>235</td>
<td>251</td>
<td>261</td>
<td>263</td>
<td>265</td>
<td>267</td>
<td>277</td>
<td>278</td>
<td>279</td>
<td>292</td>
<td>293</td>
<td>295</td>
<td>301</td>
<td>311</td>
<td>320</td>
<td>324</td>
<td>327</td>
<td></td>
</tr>
<tr>
<td>C$^a$</td>
<td>232</td>
<td>246</td>
<td>262</td>
<td>272</td>
<td>274</td>
<td>276</td>
<td>278</td>
<td>288</td>
<td>289</td>
<td>290</td>
<td>303</td>
<td>304</td>
<td>306</td>
<td>312</td>
<td>322</td>
<td>331</td>
<td>335</td>
<td>338</td>
<td>349</td>
</tr>
<tr>
<td>W</td>
<td>244</td>
<td>258</td>
<td>274</td>
<td>284</td>
<td>286</td>
<td>288</td>
<td>290</td>
<td>300</td>
<td>301</td>
<td>302</td>
<td>315</td>
<td>316</td>
<td>318</td>
<td>324</td>
<td>334</td>
<td>343</td>
<td>347</td>
<td>350</td>
<td>361</td>
</tr>
</tbody>
</table>

*The m/z of all possible b2-ions for combinations of the amino acids residue masses. Mo to designate oxidized methionine, C$^*$ to designate carbamidomethylcysteine, and C$^a$ to designate acrylocystein. Single entries are made for the isobaric amino acid pairs L and I, Q and K, and Mo and F.
Peptide Fragmentation

H -- N --- C --- C --- N --- C --- C --- N --- C --- C --- N --- C -- COOH

\( b_2 - \text{H}_2\text{O} \)
\( a_2 \)
\( b_2 \)
\( a_3 \)
\( b_3 \)

\( y_3 \)
\( y_2 \)
\( y_1 \)

\( y_3 - \text{H}_2\text{O} \)
\( y_2 - \text{NH}_3 \)
• The peaks in the mass spectrum:
  – **Prefix** and **Suffix** Fragments.
  – Fragments with **neutral losses** (-H₂O, -NH₃)
  – Noise and missing peaks.
Protein Identification with MS/MS

Peptide Identification:

MS/MS

Intensity

mass

G V D L K
Peptide Fragmentation in a Collision Cell

1. Due to collisions with gas.
2. Mobile proton from the amino terminus promotes cleavage.
4. At low energies, get mostly b- and y-ions:
   
   **b-ions**: amino terminal fragment if it retains H+ (+1 charge)
   
   **y-ions**: carboxy terminal fragment (+1 or +2 charge)
Peptide Fragmentation

E=Glu
G=Gly
S=Ser
F=Phe
N=Asn
P=Pro
V=Val
A=Ala
R=Arg

175.10
246.14
345.21
459.25
556.30
670.35
799.39
928.43
985.45
1132.52
1279.59
1366.62
1423.64
1552.69
1. A series of b- and y-ions are produced due to the fragmentation of different amide bonds.

2. Subscript refers to the number of R groups on the fragment.

3. y-ions are more common and more stable than b-ions.
CID fragmentation

$y_1 = 147$ Da

$y_2 = 310$ Da

$y_3 = 457$ Da

$y_4 = 544$ Da

$y_5 = 645$ Da

$\Delta = 87$ Da

$\Delta = 101$ Da

$\Delta = 147$ Da

$\Delta = 163$ Da

$(M+2H)^{2+} = 717$ Da

$\Delta = 72$ Da
Peptides Can Fragment At Other Sites

1. Amino Terminal Fragments:
   \(a_m, b_m, c_m\)

2. Carboxy Terminal Fragments:
   \(x_n, y_n, z_n\)

3. The fragments can also fragment espec if have a mobile \(H^+\)

4. Various side chain reactions
CID fragmentation

b-ion production
Sequencing From A y-Ion Series

1st MS Analysis

Select Ions at 617.28 & Send to Collision Cell

2nd MS Analysis of the Fragments (mostly y-ions)
Sequencing From A y-Ion Series
Masses in MS

- Monoisotopic mass is the mass determined using the masses of the most abundant isotopes.
- Average mass is the abundance weighted mass of all isotopic components.
# Amino Acid Residue Masses

**Monoisotopic Mass**

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Monoisotopic Mass</th>
<th>Amino Acid</th>
<th>Monoisotopic Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>57.02147</td>
<td>Aspartic acid</td>
<td>115.02695</td>
</tr>
<tr>
<td>Alanine</td>
<td>71.03712</td>
<td>Glutamine</td>
<td>128.05858</td>
</tr>
<tr>
<td>Serine</td>
<td>87.03203</td>
<td>Lysine</td>
<td>128.09497</td>
</tr>
<tr>
<td>Proline</td>
<td>97.05277</td>
<td>Glutamic acid</td>
<td>129.0426</td>
</tr>
<tr>
<td>Valine</td>
<td>99.06842</td>
<td>Methionine</td>
<td>131.04049</td>
</tr>
<tr>
<td>Threonine</td>
<td>101.04768</td>
<td>Histidine</td>
<td>137.05891</td>
</tr>
<tr>
<td>Cysteine</td>
<td>103.00919</td>
<td>Phenylalanine</td>
<td>147.06842</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>113.08407</td>
<td>Arginine</td>
<td>156.10112</td>
</tr>
<tr>
<td>Leucine</td>
<td>113.08407</td>
<td>Tyrosine</td>
<td>163.06333</td>
</tr>
<tr>
<td>Asparagine</td>
<td>114.04293</td>
<td>Tryptophan</td>
<td>186.07932</td>
</tr>
</tbody>
</table>
Mass Calculation (Glycine)

\[
\text{NH}_2\text{—CH}_2\text{—COOH}
\]

Amino acid

\[
\text{R}_1\text{—NH—CH}_2\text{—CO—R}_3
\]

Residue

**Monoisotopic Mass**

\[
\begin{align*}
^{1}H & = 1.007825 \\
^{12}C & = 12.00000 \\
^{14}N & = 14.00307 \\
^{16}O & = 15.99491
\end{align*}
\]

**Glycine Amino Acid Mass**

\[
5xH + 2xC + 2xO + 1xN = 75.032015 \text{ amu}
\]

**Glycine Residue Mass**

\[
3xH + 2xC + 1xO + 1xN = 57.021455 \text{ amu}
\]
MS/MS Spectra Can Be Complex

1. Many types of fragments (Some expected ones will be absent)
2. Amino acid isomers- Leucine & Isoleucine, m = 113.08
3. Amino acid isobars- Glutamine (m = 128.06) Lysine (m = 128.09)

4. **Table 4.3. Amino acids combinations that are equal to a single amino acid residue mass.*

<table>
<thead>
<tr>
<th>Amino acid combination</th>
<th>Residue mass (Da)</th>
<th>Equivalent amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td>114</td>
<td>N</td>
</tr>
<tr>
<td>GA</td>
<td>128</td>
<td>Q, K</td>
</tr>
<tr>
<td>GV</td>
<td>156</td>
<td>R</td>
</tr>
<tr>
<td>GE</td>
<td>186</td>
<td>W</td>
</tr>
<tr>
<td>AD</td>
<td>186</td>
<td>W</td>
</tr>
<tr>
<td>SV</td>
<td>186</td>
<td>W</td>
</tr>
<tr>
<td>SS</td>
<td>174</td>
<td>C&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
Calculating Peptide Masses

• Sum the monoisotopic residue masses
  
  **Monoisotopic Mass**: the sum of the exact or accurate masses of the lightest stable isotope of the atoms in a molecule
  
  • Add mass of H₂O (18.01056)
  • Add mass of H⁺ (1.00785 to get M+H)
  • If Met is oxidized add 15.99491
  • If Cys has acrylamide adduct add 71.0371
  • If Cys is iodoacetylated add 58.0071
  • Other modifications are listed at
    
    – [http://prowl.rockefeller.edu/aainfo/deltamassv2.html](http://prowl.rockefeller.edu/aainfo/deltamassv2.html)

  \[ \text{^1H-1.007828503 amu} \quad \text{^12C-12} \]
  
  \[ \text{^2H-2.014017780 amu} \quad \text{^13C-13.00335, ^14C-14.00324} \]
Peptide Mass Fingerprinting (PMF)

• Used to identify protein spots on gels or protein peaks from an HPLC run

• Depends on the fact that if a peptide is cut up or fragmented in a known way, the resulting fragments (and resulting masses) are unique enough to identify the protein

• Requires a database of known sequences

• Uses software to compare observed masses with masses calculated from database
## Principles of Fingerprinting

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Mass (M+H)</th>
<th>Tryptic Fragments</th>
</tr>
</thead>
</table>
| >Protein 1  
acedfhsakdfqea  
sdfpkivtmeeeewe  
ndadnfekqwfe | 4842.05 | acedfhsak  
dfgeasdfpk  
ivtmeeeewendadnfek  
gwfe |
| >Protein 2  
acekdfhsadfqea  
sdfpkivtmeeeewe  
nkdadnfqeqwfe | 4842.05 | acek  
dfhsadfqgeasdfpk  
ivtmeeeewenk  
dadnfqeqwfe |
| >Protein 3  
acedfhsadfqeeka  
sdfpkivtmeeeewe  
nda kdnfqeqwfe | 4842.05 | acedfhsadfqgek  
asdfpk  
ivtmeeeewendak  
dnfegwfe |
# Principles of Fingerprinting

## Sequence

<table>
<thead>
<tr>
<th>Protein 1</th>
<th>Mass (M+H)</th>
<th>Mass Spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td>acedfhsakdfqea</td>
<td>4842.05</td>
<td></td>
</tr>
<tr>
<td>sdfpkivtmeewenfdanfekqwe</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein 2</th>
<th>Mass (M+H)</th>
<th>Mass Spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td>acekdhsadfqea</td>
<td>4842.05</td>
<td></td>
</tr>
<tr>
<td>sdfpkivtmeewenkdadnfeqwe</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein 3</th>
<th>Mass (M+H)</th>
<th>Mass Spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td>acedfhsadfqeka</td>
<td>4842.05</td>
<td></td>
</tr>
<tr>
<td>sdfpkivtmeewennda kdqwe</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Preparing a Peptide Mass Fingerprint (PMF) Database

- Take a protein sequence database (Swiss-Prot or nr-GenBank)
- Determine cleavage sites and identify resulting peptides for each protein entry
- Calculate the mass (M+H) for each peptide
- Sort the masses from lowest to highest
- Have a pointer for each calculated mass to each protein accession number in databank
# Building A PMF Database

<table>
<thead>
<tr>
<th>Sequence DB</th>
<th>Calc. Tryptic Frags</th>
<th>Mass List</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;P12345</td>
<td>acedfhsak</td>
<td>450.2017 (P21234)</td>
</tr>
<tr>
<td>acedfhsakdfqea</td>
<td>dfgeasdfpk</td>
<td>609.2667 (P12345)</td>
</tr>
<tr>
<td>sdfpkivtmeewe</td>
<td>ivtmeewendadnfek</td>
<td>664.3300 (P89212)</td>
</tr>
<tr>
<td>ndadnfekqwfe</td>
<td>gwfe</td>
<td>1007.4251 (P12345)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1114.4416 (P89212)</td>
</tr>
<tr>
<td>&gt;P21234</td>
<td>acek</td>
<td>1183.5266 (P12345)</td>
</tr>
<tr>
<td>acekdfhsadfqea</td>
<td>dfhsadfgeasdfpk</td>
<td>1300.5116 (P21234)</td>
</tr>
<tr>
<td>sdfpkivtmeewe</td>
<td>ivtmeewenk</td>
<td>1407.6462 (P21234)</td>
</tr>
<tr>
<td>nkdadnfeqwfe</td>
<td>dadnfeqwfe</td>
<td>1526.6211 (P89212)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1593.7101 (P89212)</td>
</tr>
<tr>
<td>&gt;P89212</td>
<td>acedfhsadfqek</td>
<td>1740.7501 (P21234)</td>
</tr>
<tr>
<td>acedfhsadfqeka</td>
<td>asdfpk</td>
<td>2098.8909 (P12345)</td>
</tr>
<tr>
<td>sdfpkivtmeewe</td>
<td>ivtmeewendak</td>
<td></td>
</tr>
<tr>
<td>ndakdnfeqwfe</td>
<td>dnfegwfe</td>
<td></td>
</tr>
</tbody>
</table>
The Fingerprint (PMF) Algorithm

• Take a mass spectrum of a trypsin-cleaved protein (from gel or HPLC peak)

• Identify as many masses as possible in spectrum (avoid autolysis peaks of trypsin)

• Compare query masses with database masses and calculate # of matches or matching score (based on length and mass difference)

• Rank hits and return top scoring entry – this is the protein of interest
MALDI Fingerprinting

1. Purify protein.

2. Digest with trypsin.

3. Perform MALDI-MS (NOT tandem MS).

4. Obtain a signature for that protein composed of the peptide masses.

5. Compare peptide masses to a database of expected peptide masses from each known protein for that species.

6. Frequently this identifies the protein and its amino acid sequence unambiguously.
# Query vs. Database

<table>
<thead>
<tr>
<th>Query Masses</th>
<th>Database Mass List</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>450.2201</td>
<td>450.2017 (P21234)</td>
<td>2 Unknown masses</td>
</tr>
<tr>
<td>609.3667</td>
<td>609.2667 (P12345)</td>
<td>1 hit on P21234</td>
</tr>
<tr>
<td>698.3100</td>
<td>664.3300 (P89212)</td>
<td>3 hits on P12345</td>
</tr>
<tr>
<td>1007.5391</td>
<td>1007.4251 (P12345)</td>
<td></td>
</tr>
<tr>
<td>1199.4916</td>
<td>1114.4416 (P89212)</td>
<td>Conclude the query</td>
</tr>
<tr>
<td>2098.9909</td>
<td>1183.5266 (P12345)</td>
<td>protein is P12345</td>
</tr>
<tr>
<td></td>
<td>1300.5116 (P21234)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1407.6462 (P21234)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1526.6211 (P89212)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1593.7101 (P89212)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1740.7501 (P21234)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2098.8909 (P12345)</td>
<td></td>
</tr>
</tbody>
</table>
Database search

PeptIdent (ExPasy)
Mascot (Matrix Science)
MS-Fit (Prospector; UCSF)
ProFound (Proteometrics)
MOWSE (HGMP)

Human Genome Mapping Project

Theoretical vs Experimental MS spectra.
Uninterpreted MS/MS Database Search

Theoretical spectra for given precursor mass

MS/MS Data

m/z – Int precursor mass

Sequence Database

* Assign scores to overlaps
* (Normalize scores)
* Keep best match
What You Need To Do PMF

• A list of query masses (as many as possible)

• Protease(s) used or cleavage reagents

• Databases to search (SWProt, Organism)

• Estimated mass and pl of protein spot (opt)

• Cysteine (or other) modifications

• Minimum number of hits for significance

• Mass tolerance (100 ppm = 1000.0 ± 0.1 Da)

• A PMF website (Prowl, ProFound, Mascot, etc.)
PMF on the Web

- **ProFound**
- **MOWSE**
  - [http://srs.hgmp.mrc.ac.uk/cgi-bin/mowse](http://srs.hgmp.mrc.ac.uk/cgi-bin/mowse)
- **PeptideSearch**
  - [http://www.narrador.embl-heidelberg.de/GroupPages/Homepage.html](http://www.narrador.embl-heidelberg.de/GroupPages/Homepage.html)
- **Mascot**
  - [www.matrixscience.com](http://www.matrixscience.com)
- **PeptIdent**
# ProFound Results

## ProFound - Search Result Summary

<table>
<thead>
<tr>
<th>Rank</th>
<th>Probability</th>
<th>Est'd Z</th>
<th>Protein Information and Sequence Analyse Tools (T)</th>
<th>%</th>
<th>p/</th>
<th>kDa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.2e-001</td>
<td>0.12</td>
<td>gi</td>
<td>1522204</td>
<td>ref</td>
<td>NP</td>
</tr>
<tr>
<td>2</td>
<td>2.2e-001</td>
<td>0.12</td>
<td>gi</td>
<td>17547403</td>
<td>ref</td>
<td>NP</td>
</tr>
<tr>
<td>3</td>
<td>7.6e-002</td>
<td></td>
<td>gi</td>
<td>23054472</td>
<td>gb</td>
<td>ZP</td>
</tr>
<tr>
<td>4</td>
<td>7.6e-002</td>
<td></td>
<td>gi</td>
<td>19920902</td>
<td>ref</td>
<td>NP</td>
</tr>
<tr>
<td>5</td>
<td>2.6e-002</td>
<td></td>
<td>gi</td>
<td>19572314</td>
<td>emb</td>
<td>CAD19081.1</td>
</tr>
<tr>
<td>+6</td>
<td>2.5e-002</td>
<td></td>
<td>gi</td>
<td>2133779</td>
<td>psi</td>
<td>S63985</td>
</tr>
<tr>
<td>7</td>
<td>2.3e-002</td>
<td></td>
<td>gi</td>
<td>15450423</td>
<td>gb</td>
<td>A4K96505.1</td>
</tr>
<tr>
<td>+8</td>
<td>2.0e-002</td>
<td></td>
<td>gi</td>
<td>7495844</td>
<td>psi</td>
<td>T25534</td>
</tr>
<tr>
<td>9</td>
<td>1.9e-002</td>
<td></td>
<td>gi</td>
<td>21293583</td>
<td>gb</td>
<td>EAA05728.1</td>
</tr>
<tr>
<td>10</td>
<td>1.6e-002</td>
<td></td>
<td>gi</td>
<td>16112031</td>
<td>ref</td>
<td>NP</td>
</tr>
</tbody>
</table>
You can use this page to submit an MOWSE database search.

MOWSE will search the *uniprot protein sequence database* with protein fragment information, and return the protein(s) which most likely correspond to your peptide-data.

### MOWSE Parameters

You must supply various parameters in order to run MOWSE successfully. We have provided default values for some of these parameters, information to help you select appropriate alternatives can be found by following the links associated with the various parameter fields.

- **Sample name**: test_data
- **Select the reagent used**: Trypsin
- **Tolerance**: 0.1
- **Whole sequence molecular weight**: 0.0
- **Molecular weight**:
  - **Factor**: 25
  - **Value**: 0.2

**Peptide mass data**
PeptIdent

Peptide Mass Fingerprinting

- Name of the unknown protein: unknown
- Database: Swiss-Prot
- Species to be searched: HOMO SAPIENS (HUMAN)
- Enter a list of peptide masses (separated by spaces or newlines) that correspond to the unknown protein.
- All peptide masses are: [M+H]^+ or [M+D] or [M-H], and monocisotopic or average.
- The peptide masses are: [with cysteines treated with: nothing (in reduced form)]
  [with acrylamide adducts on cysteines]
  [with methionine oxidized].
- Mass tolerance: ± 0.2 Dalton
- Send the result by e-mail

- pl: within pl range: 1.00
- Mw: (in Dalton, not kDa) within Mw range (in percent): 20
- Enzyme: Trypsin
- Allow for 0 missed cleavage sites (M.C.)
- Report only proteins with at least 4 peptide hits
- Display a maximum of 20 matching proteins.
- Choose information about sequence portion covered by the matching peptides.

Or upload a file in one of the supported formats from your computer. The peptide masses will be extracted automatically from this file:
Mascot Scoring

- The statistics of peptide fragment matching in MS (or PMF) is very similar to the statistics used in BLAST.

- The scoring probability follows an extreme value distribution.

- High scoring segment pairs (in BLAST) are analogous to high scoring mass matches in Mascot.

- Mascot scoring is much more robust than arbitrary match cutoffs (like % ID).
Extreme Value Distribution

it is the limit distribution of the maxima of a sequence of independent and identically distributed random variables. Because of this, the EVD is used as an approximation to model the maxima of long (finite) sequences of random variables.

Scores greater than 72 are significant
Probability Based Mowse Score

Score is $-10\times \log(P)$, where $P$ is the probability that the observed match is a random event. Protein scores greater than 72 are significant ($p<0.05$).
Mascot/Mowse Scoring

• The Mascot Score is given as $S = -10\times\log(P)$, where $P$ is the probability that the observed match is a random event.

• Try to aim for probabilities where $P<0.05$ (less than a 5% chance the peptide mass match is random).

• Mascot scores greater than 72 are significant ($p<0.05$).
Advantages of PMF

• Uses a “robust” & inexpensive form of MS (MALDI)

• Doesn’t require too much sample optimization

• Can be done by a moderately skilled operator (don’t need to be an MS expert)

• Widely supported by web servers

• Improves as DB’s get larger & instrumentation gets better

• Very amenable to high throughput robotics (up to 500 samples a day)
Limitations With PMF

• Requires that the protein of interest already be in a sequence database

• Spurious or missing critical mass peaks always lead to problems

• Mass resolution/accuracy is critical, best to have <20 ppm mass resolution

• Generally found to only be about 40% effective in positively identifying gel spots