Biophysics 101: Genomics & Computational Biology

Section 9: Mass Spectrometry and Proteomics
Michael Jones
Nov. 18th 2003

Mass Spectrometry

• Macromolecular MS only became feasible with past ~15 yrs with advent of new milder ionization methods:
  • Electrospray ionization (ESI)
  • Matrix-assisted laser desorption/ionization (MALDI)
• Molecular weight precision ~ ± 0.01%

Instrumentation

• Purpose – to measure the mass of molecules
• Three Components
  – Ion Source – Convert molecules to ion’s in the gas phase – Nebulize and charge
  – Mass Analyzer – Separate or filter the mixture of ions by their mass to charge ratio (m/z)
  – Detector – Detect ions and abundances and convert to electrical signal to produce a digital response

Instrumentation

• Ion-Source
  – API (Atmospheric Pressure Ionization) or Electrospray (ESI)
  – High pressure gradient, heat and electric field aids ionization and nebulization
  – Charges in solvent collapse onto analyte
  – One of the first methods used to ionize large biomolecules

Instrumentation

• Ion-Source
  – MALDI (Matrix Assisted Laser Desorption Ionization)
  – Analyte, matrix and cation (H+) are co-crystalized
  – Matrix has absorbance at the wave length of the laser
  – Laser excites matrix, energy is transferred to analyte as matrix evaporates leaving bare analyte molecules
    • (At some point the cation is transferred from solution to matrix and then to the analyte)
  – Charged analyte ions are accelerated through a voltage gradient to the mass analyzer

Instrumentation

• Ion-Source
  – MALDI (Matrix Assisted Laser Desorption Ionization)
**Instrumentation**

- **Mass Analyzers**
  - Quadrupole mass spectrometer (Most popular method)
    - DC and alternating RF voltages applied to four parallel rods creates electric field that focuses ions of a particular m/z towards the detector
  - Relatively easy to build
  - Limited Mass Range
  - Need multiple analyzers for ms/ms

- **Tandem Mass Spectrometry**
  - Time of Flight (TOF)
    - Ions are accelerated through a voltage gradient into a field free tube towards a detector
    - Velocities resulting from kinetic energy generated at gradient determines "time of flight"
    - Lighter ions – Lower TOF
  - Unlimited mass range
  - Fast scan times
  - High quality spectra

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Application</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESI-Triple QP</td>
<td>LCMS, tandem MS</td>
<td>Limited Mass Range, relatively easy to make</td>
</tr>
<tr>
<td>ESI-QP Ion-Trap</td>
<td>LCMS, tandem MS</td>
<td>Limited mass range, trap improves sensitivity and automated analysis</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Analysis of purified samples</td>
<td>Low sample consumption, tolerant of buffer and salt contaminants (Ex. 2D Gel samples)</td>
</tr>
<tr>
<td>QP-TOF</td>
<td>LCMS, Tandem MS</td>
<td>High mass accuracy, fast scan times.</td>
</tr>
<tr>
<td>Fourier (transform ion cyclotron resonance)</td>
<td>LCMS, Tandem MS, Top Down MS (Fragment large biomolecules)</td>
<td>Very High mass accuracy, ability to fragment very large molecules</td>
</tr>
</tbody>
</table>

Tandem Mass Spectrometry

- High peptide with high energy gas and fragment
- Trade off between identification and quantification
  - Mass Spec will not be collecting quantitative data when it is fragmenting peptide

**Tandem Mass Spectrometry (ms/ms)**

<table>
<thead>
<tr>
<th>Peptide Fragmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>x, y, z</td>
</tr>
<tr>
<td>x, b, c</td>
</tr>
<tr>
<td>Neutral Loss, NH$_3$ and H$_2$O loss</td>
</tr>
<tr>
<td>Internal Fragments</td>
</tr>
<tr>
<td>Immonium Ions</td>
</tr>
</tbody>
</table>

Peptide Fragmentation Nomenclature

- Integers indicate amino acid closest to charged terminus
- NH$_3$ loss
- NH$_3$ contains one residue with ammonia neutral loss
- b++ a double charge b ion
- Fragments generated dependent on peptide sequence
**Tandem Mass Spectrometry (ms/ms)**

- Spectral Interpretation
  - Manual interpretation
  - de novo sequencing: Graph Algorithms, dynamic programming
  - Sequence spectrum correlation: Match experimental spectral pattern to pattern predicted from known sequence databases

**Standard proteome analysis by 2DE and MS**

- Mass Spectrometry
- Trypsin digestion of complex mixture
- Select mass from protein of interest
- Sequence peptide of interest

**Complex Mixture Analysis with Data Dependent Scanning and Dynamic Exclusion on Ion Trap**

- 2 hr gradient
- Ion Chromatogram
- HPLC
- 100’s of proteins
- Repeat Process
- 10 sec cycle time
- Generate 1000’s of MS/MS Spectra

**LC-MS Technology**

- Data Dependent MS/MS
  - Quadrupole Ion Trap
  - Fill the trap, scan for most intense, refill trap with desired mass, fragment
  - Allows for dynamic fractionation of peptides
  - Higher Throughput
  - Complex Mixtures
    - Complexes
    - Body Fluids
    - Media

**Detection of Post Translational Modifications (PTM’s)**

- MALTI-TOF Mass Spectrum of tryptic digestion of purified protein (GCSF)
- All of the masses of the detected peptides match except 3227.8 and 3243.8
- These are 16 and 32 amu more then the native N-terminal peptide MKLMVLQLLLWHSALWTVHEATPLGPAR
- Delta amu of 16 indicates an oxidation. Met is known to be susceptible to oxidation
- If you want to know which amino acid is modified in the 3227.8 amu peptide you need to do ms/ms

**Peptide Sequencing by LC/MS/MS**

- Mass Spectrometry
- Trypsin digestion of complex mixture
- Select mass from peptide of interest
- Sequence peptide of interest
Post-translational Modifications

- Phosphorylation
- Ubiquitination
- etc

Detection of Post Translational Modifications (PTM’s)

Complete characterization is difficult in purified samples and nearly impossible in complex mixtures.

Some modifications won’t be seen since peptide is not detected or not identified.

FTMS and top down proteomics is enhancing the ability to analyze PTM’s in Complex Mixtures.

2D Gel analysis, antibody identification and affinity purification helps.

Proteomic analysis of post-translational modifications


Quantization

Stable-isotope protein labeling

Quantization

Isotope-coded affinity tag (ICAT)

Quantization

ApproxQuant

- Spectrum Mill feature
- Absolute intensity determined by intensity data collected for each parent ion
- Compare intensity data from identical peptides across samples
- Approximate method since data lost in duty cycle

Parent XIC (Extracted Ion Chromatogram)

Retention Time

Collected Data

Quant Data lost in ms/ms duty cycle

Spot Detection

XIC Current

Quantization

MS-Only

- Avoid any ms/ms analysis, create ms scans only-Perform ms/ms after quantization and data analysis
- Component Detection
  - generating XIC’s for all mass ranges or do spot detection
  - Need to align components since elution times on LC may vary
### Quantization

<table>
<thead>
<tr>
<th>Method</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stable-isotope protein labeling</td>
<td>Need to synthesize proteins in presence of isotope, experimental design issues of two-color system</td>
</tr>
<tr>
<td>ICAT</td>
<td>Enriches (and limited) for peptides with particular amino acids (Ex. Cys), experimental design issues of two-color system</td>
</tr>
<tr>
<td>AproxiQuant</td>
<td>Very approximate absolute quantization, easy to implement and use</td>
</tr>
<tr>
<td>LC-MS Only</td>
<td>Very good absolute quantization, computationally challenging (Component detection, alignment, noise), identification requires additional steps</td>
</tr>
</tbody>
</table>

### Computational Methods for Identification of proteins

#### Peptide Mass Fingerprint
- Compare spectra of peptides from experimental digestion of purified protein to theoretical digestion of sequence database
  - The better the experimental digest corresponds to the theoretical digest the higher the score
  - Mass Accuracy, incomplete digestion, contaminations, PTM’s all affect the result
- PeptIdent – includes pI (SwissProt)
- Mascot (MatrixScience)
- MS Tag SpectrumMill (Agilent)

#### MS/MS Ion or MS-Tag Search
- Compare spectra of experimentally fragmented peptide (ms/ms) to theoretical fragmentation of predicted sequence database peptides
  - Uses parent mass and fragment ion-masses
  - Does not require a purified protein
- Not always easy to predict how a peptide will fragment
  - Incomplete fragmentation along peptide bonds, neutral losses, differing intensities
- MS Tag SpectrumMill (Agilent)
- Sequest
- Mascot (MatrixScience)

### Mixture Functional Classification

- Pie Charts often show subjective representation of complex mixture
- Use statistical categorization method

Sample – Synovial fluid of arthritis patients
Category Set (n) – GO Process
Scoring Method – Fishers Exact
Universe (N) – All Human LocusLink
Resources

- Base Peak (http://www.spectroscopynow.com)
- http://www-methods.ch.cam.ac.uk/meth/ms/theon/malsk.html
- http://www-methods.ch.cam.ac.uk/meth/ms/theon/quadrupole.html
- http://www.spectroscopynow.com/Spy/basehtml/SpyH/1,1181,4-1-2-0-0-news_detail-472977884-842,00.html
- http://david.niaid.nih.gov/david/lease.htm (Categorization)