Protein and PTM Identification

IDeA National Resource Proteomics Workshop
May 24, 2017

Samuel G. Mackintosh, Ph.D.
Dept. of Biochemistry and Molecular Biology
University of Arkansas for Medical Sciences

sgmackintosh@uams.edu
Protein identification by LC-MS/MS:

- in-gel trypsin digestion
- reverse-phase HPLC separation
- electrospray ionization
- peptide ion detection by MS
- peptide fragmentation by MS/MS
- protein identification by database searching

\[
\begin{align*}
+\text{NH}_3\text{-TYAAGDFPILTNMCR-COO}^- \\
+\text{NH}_3\text{-GTHLAAECVPK-COO}^- \\
+\text{NH}_3\text{-AAGTYCAGDVR-COO}^- \\
+\text{NH}_3\text{-ILESTCVYLAGR-COO}^- \\
\end{align*}
\]

trypsin

\[
\begin{align*}
\text{RT:} & \quad 10.41 & \text{AV:} & \quad 1 & \text{NL:} & \quad 3.02\times10^5 \\
\text{T:} & \quad \text{ITMS + c ESI} & \text{Full ms} & \quad [375.00-1500.00] \\
\end{align*}
\]

\[
\begin{align*}
5-20\ \mu\text{M sample} & \text{in 10-1000 mM aqueous HEPES buffer} \\
\text{Nanoflow capillary} & \text{tip (1-10 \mu m)} \\
\text{Taylor cone} & \text{Atmospheric pressure} \\
\text{Droplet formation} & \text{Vacuum stages} \\
\text{Droplet fusion} & \text{To mass analyser} \\
\end{align*}
\]
Tryptic digestion of BSA:
• cleaves C-terminal of K or R

MW range of peptide products:
146.19 – 2435.82

• not all peptides will be ionized and detected
• there will be gaps in sequence coverage
LC-MS of 20 fmol of BSA

H$_2$O/acetonitrile gradient

RT: 0.00 - 30.08

NL: 2.89E6
Base Peak F: ITMS + c ESI Full ms
[375.00-1500.00] MS
BSA_060908_01
MS scan – determines masses of intact peptides

Ecoli_041708_01 #1310  RT: 14.94  AV: 1  NL: 8.31E4
F: ITMS + c ESI Full ms [375.00-1500.00]
Data-Dependent Acquisition (DDA)

most intense peptide ions are selected for fragmentation
MS/MS scan – determines masses of peptide fragments
**MS/MS fragmentation methods:**

Collision-induced dissociation (CID) is most likely to occur at peptide bonds

- **b ions**: charge is retained on N-terminus
- **y ions**: charge is retained on C-terminus

Electron-transfer dissociation (ETD) is more likely to occur at N-Cα bonds

- **c ions**: charge is retained on N-terminus
- **z ions**: charge is retained on C-terminus

High-energy collisional dissociation (HCD) produces b and y ions
CID – collision-induced dissociation

takes place in ion trap in presence of inert collision gas (He or Ar)
CID fragmentation of peptide YICDNQDTISSK

b ion series:

b1: Y-
b2: YI-
b3: YIC-
b4: YICD-
b5: YICDN-
b6: YICDNQ-
b7: YICDNQD-
b8: YICDNQDT-
b9: YICDNQDTI-
b10: YICDNQDTIS-
b11: YICDNQDTISS-

y ion series:

y1: K-
y2: KS-
y3: KSS-
y4: KSSI-
y5: KSSIT-
y6: KSSITD-
y7: KSSITDQ-
y8: KSSITDQN-
y9: KSSITDQND-
y10: KSSITDQNDDC-
y11: KSSITDQNDCl-
Peptide sequence determination by MS/MS fragmentation

\[ +\text{NH}_3-\text{YICDNQDTISSK-COO}^- \]
Peptide/protein identification by database searching

- perform a theoretical trypsin digestion of all known proteins
- determine a theoretical MS/MS fragmentation for each peptide
- compare experimental spectrum to each theoretical spectrum
### Mascot peptide scoring

#### Monoisotopic mass of neutral peptide Mr(calc): 1442.6347
Fixed modifications: Carbamidomethyl (C)

Ions Score: 71  Expect: 0.000021

<table>
<thead>
<tr>
<th>#</th>
<th>b</th>
<th>b++</th>
<th>b*</th>
<th>b++</th>
<th>b0</th>
<th>b0++</th>
<th>Seq</th>
<th>y</th>
<th>y++</th>
<th>y*</th>
<th>y++</th>
<th>y0</th>
<th>y0++</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>164.0706</td>
<td>82.5389</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>277.1547</td>
<td>139.0810</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>437.1853</td>
<td>219.0963</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>552.2123</td>
<td>376.6089</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>666.2552</td>
<td>333.6312</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>794.3138</td>
<td>397.6605</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>909.3407</td>
<td>455.1740</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1010.3884</td>
<td>505.6978</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1113.4275</td>
<td>562.2509</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>10</td>
<td>1210.5045</td>
<td>605.7559</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>11</td>
<td>1297.5365</td>
<td>649.2719</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>
</tbody>
</table>

#### NCBI BLAST search of YICDNDTITSSK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)

Other BLAST web gateways

### All matches to this query

<table>
<thead>
<tr>
<th>Score</th>
<th>Mr(calc): 1442.6347</th>
<th>Delta</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>70.8</td>
<td>1.0839</td>
<td>YICDNDTITSSK</td>
<td></td>
</tr>
<tr>
<td>22.2</td>
<td>1.0839</td>
<td>YICDNDTITSSK</td>
<td></td>
</tr>
<tr>
<td>16.2</td>
<td>0.9629</td>
<td>YICDNDTITSSK</td>
<td></td>
</tr>
<tr>
<td>14.8</td>
<td>1.9224</td>
<td>ANFAYNNELLMKK</td>
<td></td>
</tr>
</tbody>
</table>
Compilation of database search results
Challenges in mapping post-translational modifications

• Stoichiometry is often unfavorable

• Sequence coverage gaps

• Relatively unstable modifications are often lost during peptide fragmentation
RLSVVGPPNR

chromatogram

Time (min)
0 10 20 30 40 50 60 70 80 90 100
Relative Abundance

NL Bas FTM ms [37 150 Cue
phosphorylated peptide unmodified peptide
Protease digestion of proteins for MS analysis

Commonly used alternate proteases
• AspN – cleaves on N-terminal side of Asp residues
• GluC – cleaves on C-terminal side of Glu residues
• LysC – cleaves on C-terminal side of Lys residues
• ArgC – cleaves on C-terminal side of Arg residues
• chymotrypsin – cleaves on C-terminal side of hydrophobic residues
• proteinase K – not sequence specific; used for limited digestion
“Typical” phosphopeptide spectrum:

\[^{+\text{NH}_3}\text{KSESTSSSYNYR-COO}^-\]
Comparison of ETD and CID for phosphopeptide MS/MS

Mapping of post-translational modifications requires

1) Identification of modified peptides

AND

2) Localization of modified sites
PTM site localization using Ascore