Post-translational Modification Identification by Mass Spectrometry

Liwen Zhang
Mass Spectrometry and Proteomics Facility
The Ohio State University

Summer Workshop 2015
PTM: Chemical Modifications on Proteins
Increases the Functional Diversity of the Proteome
Regulate Activity, Localization and Interaction

Image from https://www.lifetechnologies.com/overview-post-translational-modification.html
Histone Modifications

- Lysine Methylation
- Arginine Methylation
- Acetylation
- Phosphorylation
- Citrullination
- Sumoylation
- Ubiquinylation

**Chemical Structures:**
- Lysine
- Acetyl-Lysine
- Phospho Serine, Threonine and Tyrosine amino acids
- K-GGXXX
- K-GG

**Acetylation by KATs**
**Deacetylation by HDACs**

**KGGXX:**
Peptide bond  Sumoylated  iso peptide bond

**Ubiquitin:**
Peptide bond  Ubiquitin iso peptide bond

**K-GGXXX:**
# Post-translational Modifications

<table>
<thead>
<tr>
<th>Modification</th>
<th>Modified Residue</th>
<th>Mass Shift (Da)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disulphide bond formation</td>
<td>C</td>
<td>-2.0157</td>
<td>protein stability</td>
</tr>
<tr>
<td>Methylation</td>
<td>K/R</td>
<td>14.0157</td>
<td>regulation of gene expression</td>
</tr>
<tr>
<td>Hydroxylation</td>
<td>P</td>
<td>15.9949</td>
<td>Protein stability and protein-ligand interaction</td>
</tr>
<tr>
<td>Oxidation of Met</td>
<td>M</td>
<td>15.9949</td>
<td>usually introduced during digestion</td>
</tr>
<tr>
<td>Acetylation</td>
<td>K</td>
<td>42.0106</td>
<td>Protein stability, protection of N terminus, regulation of protein-DNA interactions</td>
</tr>
<tr>
<td>S-Glutathionylation</td>
<td>C</td>
<td>305.0682</td>
<td>oxidative stress; preventing irreversible oxidation of protein thiol; control of cell-signaling pathways by modulating protein function</td>
</tr>
<tr>
<td>Phosphorylation</td>
<td>S/T/Y</td>
<td>79.9663</td>
<td>Signaling, activation/inactivation of enzyme activities, modulation of molecular interactions</td>
</tr>
<tr>
<td>Nitration</td>
<td>Y</td>
<td>44.9851</td>
<td>oxidative damage during inflammation</td>
</tr>
<tr>
<td>Ubiquitination</td>
<td>K</td>
<td>114.0429</td>
<td>destruction signal</td>
</tr>
<tr>
<td>Oxidation of Cys</td>
<td>C</td>
<td>31.9721/47.9847</td>
<td>oxidative/nitrosative stress</td>
</tr>
<tr>
<td>Nitrosylation</td>
<td>C</td>
<td>28.9902</td>
<td>oxidative/nitrosative stress</td>
</tr>
<tr>
<td>Glycosylation</td>
<td>N/S/T</td>
<td>........</td>
<td>Excreted proteins, cell-cell recognition/signaling O-GlcNAc, reversible, regulatory functions</td>
</tr>
</tbody>
</table>
Method for PTM Detections

- **Specific Fluorescence Dye:** Pro-Q Diamond --- Phosphorylation
  Pro-Q Emerald --- Glycosylation
- **Specific Antibodies:** Nitration; Phosphorylation; Methylation;
  Ubiquitination; Acetylation
- **Mass Shift/PI Change:** Gel Approaches

**Pros:**
High sensitivity, Favor the modification

**Cons:**
Lack of accuracy
Interference from other residues
Lack the ability to locate the actual modification sites

Mass Spectrometry!!!!!
Peptide Modifications Identification by MS/MS

- Virtually anything that shifts the mass can be determined by MS
- Using MS/MS allows for identification of the modification AND location
- You must have an idea of what modification you are looking for
- First use protein stain to examine various modifications
- Western analysis VS MS analysis

**Cons:** Detection based on ionization efficiency and abundance of the peptides, thus may not favor modified peptides.

- LC-MS/MS allows for the identification of low abundant modifications
**MSMS Fragments and Different Fragmentation Technique**

CID: a, b and y ion, loss of H2O, NH3, side-chain

PSD: a, b and y ion

IRMPD: b and y ion, loss of side-chain

ECD: c, y and z ions

ETD: c, y and z ions
Sequence Nomenclature for Mass Ladder

\[ \text{b}_2 - \text{b}_1 = 56\text{Da} + R_2 \quad \text{b}_3 - \text{b}_2 = 56\text{Da} + R_2 \]

\[ \text{y}_3 - \text{y}_2 = 56\text{Da} + R_2 \quad \text{y}_2 - \text{y}_1 = 56\text{Da} + R_2 \]

Mass = 56Da + R
GLLEDLGYDVVK
G L L E D L G Y D V V V K

m/z

b12

1273.72

y2

246.00

y2

m/z
G L L E D L G Y D V V V K

246.00 y2
345.32 y3

99.04 Val

1174.68 b11
1273.72 b12

99.32 Val
GLLEDLGYD

246.00 y2
345.32 y3
444.29 y4

1075.51 b10
1174.68 b11
1273.72 b12

99.17 Val
99.04 Val

G L L E D L G Y D V V V V K

99.32 Val
98.97 Val
GLLEDLYDYDVVVK

246.00  y2
345.32  y3
444.29  y4
559.27  y5

976.45  b9
1075.51  b10
1174.68  b11
1273.72  b12

99.06  Val
99.17  Val
99.04  Val

99.32  Val
98.97  Val
114.98  Asp
Database Searching

*Mascot*
• 16 node cluster for high-speed data processing

**Protein sequences are digested and fragmented *In Silico* which produces an enormous peak list**

**Raw MS/MS data is converted to a peak list and compared against the *In Silico* peak list.**

**Critical point of database searching**
• The enzyme must work properly (no non-specific cleavage and missed cuts)
• The mass accuracy limits must be set appropriately
• Any modifications must be accounted for (modified cysteine)
• The database must contain the protein
Mascot Database Search Engine

User Name: Kari
Search Title: Trypsin full
Taxonomy: Homo sapiens (human)
Database: Sabre_All, salmonella, Smansionsequence
Fixed Modifications: Carbamidomethyl (C)
Variable Modifications: Acetyl (K), Deamidated (NQ), GlyGly (K), Oxidation (M)
Peptide Mass: 4+ (Charge)
Peptide Tolerance: ±20 ppm
MS/MS Tolerance: ±0.8 Da
MS/MS Ions Search: Mascot generic
Quantitation: None
Instrument: ESI-TRAP
Report Top: AUTO hits
Enzyme Used: Trypsin/P

20 ppm for orbitrap
1.5 Da for LTQ
### MASCOT Results

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Protein Mascot</th>
<th>Mowse Score</th>
<th>Protein Mol. Weight</th>
<th># of spectra matched</th>
<th># of peptides matched</th>
<th>The Exponentially Modified Protein Abundance Index (emPAI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYG_EQUBU</td>
<td>2474</td>
<td>Myoglobin</td>
<td>Equus burchelli</td>
<td>0</td>
<td>0</td>
<td>186.7</td>
</tr>
<tr>
<td>MYG_KOGR</td>
<td>492</td>
<td>Myoglobin</td>
<td>Koga breviceps</td>
<td>0</td>
<td>0</td>
<td>86.7</td>
</tr>
<tr>
<td>MYG_PHYCD</td>
<td>317</td>
<td>Myoglobin</td>
<td>Physeler zedon</td>
<td>0</td>
<td>0</td>
<td>3.16</td>
</tr>
</tbody>
</table>

The Exponentially Modified Protein Abundance Index (emPAI) is calculated by

$$emPAI = 10^{\frac{\text{Observed}}{\text{Expected}}} - 1$$
Example 1 — Phosphorylation Identification by MS/MS

```
41T S S[DN] N N V S(PO4) 48 V T N V S V A K

m/z=850.542+
```

```
<table>
<thead>
<tr>
<th>391.12</th>
<th>505.27</th>
<th>619.29</th>
<th>718.37</th>
<th>802.43</th>
<th>M-H3PO4</th>
</tr>
</thead>
<tbody>
<tr>
<td>b4</td>
<td>b5</td>
<td>b6</td>
<td>b7</td>
<td>886.36</td>
<td>b9*</td>
</tr>
<tr>
<td>Asn</td>
<td>Asn</td>
<td>Val</td>
<td></td>
<td></td>
<td>Thr</td>
</tr>
<tr>
<td>886.36</td>
<td>b9*</td>
<td>1101.51</td>
<td>b11*</td>
<td></td>
<td>1287.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

```
| 886.36 | b9* |
| 1101.51 | b11* |
| 1287.76 | b13* |
| 1457.78 | b15* |
```

```
<table>
<thead>
<tr>
<th>317.28</th>
<th>404.24</th>
<th>503.32</th>
<th>617.36</th>
<th>718.37</th>
<th>817.37</th>
<th>984.39</th>
<th>1083.45</th>
<th>1197.51</th>
<th>1311.59</th>
<th>1426.56</th>
</tr>
</thead>
<tbody>
<tr>
<td>y3</td>
<td>y4</td>
<td>y5</td>
<td>y6</td>
<td>y7</td>
<td>y8</td>
<td>y9</td>
<td>y10</td>
<td>y11</td>
<td>y12</td>
<td>y13</td>
</tr>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Asn</td>
<td>Thr</td>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>886.36</td>
<td>y9*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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>
```

```
<table>
<thead>
<tr>
<th>317.28</th>
<th>404.24</th>
<th>503.32</th>
<th>617.36</th>
<th>718.37</th>
<th>817.37</th>
<th>984.39</th>
<th>1083.45</th>
<th>1197.51</th>
<th>1311.59</th>
<th>1426.56</th>
</tr>
</thead>
<tbody>
<tr>
<td>y3</td>
<td>y4</td>
<td>y5</td>
<td>y6</td>
<td>y7</td>
<td>y8</td>
<td>y9</td>
<td>y10</td>
<td>y11</td>
<td>y12</td>
<td>y13</td>
</tr>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Asn</td>
<td>Thr</td>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>886.36</td>
<td>y9*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Example 1 — Phosphorylation Identification by MS/MS

```
850.542+ (87+80+99) Ser(PO4)-Val
```

```
41T S S[DN] N N V S(PO4) 48 V T N V S V A K

bn*=bn-H3PO4
yn*=yn-H3PO4
```
Example 2 — Nitration Identification by MS/MS

Measured $m/z = 775.61^{2+}$
Theoretical $m/z = 775.37^{2+}$
Example 3 — Identification of Cysteine Oxidation by MS/MS
Example 4 — Identification of DMPO Adduct by MS/MS

Theoretical $m/z=859.94^{2+}$
Observed $m/z=859.89^{2+}$

DMPO=$C_6H_9NO$ (111.0684 Da)
PTM Identification: when modification is on the terminus of the peptide
PTM Identification: when modification is on the terminus of the peptide

$^{44}X$ K E G V V H G V A T V A E $^{60}K$

Mass of X = (M+H) - y14 = 826.45*2-1-1423.55 = 228.35
Mass of X = b2 - Mass of Lysine - H = 357.19 - 128.02 - 1.01 = 228.16

Mass of X (b2): 357.19
Mass of Lysine-H: 128.02
Mass of H: 1.01
PTM Identification: when modification is on the terminus of the peptide

Modification occurs on lysine residue
Mass of modification = Mass of X - Mass of Lysine = 228.16 - 128.09 = 100.05 Da
Succinylation: Addition of –COCH2CH2CO- to lysine residue (+100.0115 Da)

High mass accuracy instrument needed
Use high mass accuracy for MS/MS too
MSn can be performed to resolve the structure of modification if necessary
Identification of PTMs:

Know the Modification
(Stability, Mass Shift, possible modification site)

High Samples Purity

High Concentration

High Mass Accuracy/Resolution
Difficulties for PTMs Identification

- Generating Peptides containing modified residue
  --Different Enzymes/Digestion Condition

- Low population of modifications (usually less than 1%)
  --Enrichment
  --Targeted Method

- Modification not stable
  --Experiment Condition (reducing reagents, enzymes)
  --Fresh sample

- Modification labile in MS analysis
  --Neutral Loss Method
  --Different Fragmentation Method
# Enzymes for Proteome Research

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Cleave Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsin</td>
<td>K-X and R-X</td>
</tr>
<tr>
<td>Endoprotease Lys-C</td>
<td>K-X except when X = P</td>
</tr>
<tr>
<td>Endoprotease Arg-C</td>
<td>R-X except when X = P</td>
</tr>
<tr>
<td>Endoprotease Asp-N</td>
<td>X-D</td>
</tr>
<tr>
<td>Endoprotease Glu-C</td>
<td>E-X except when X = P</td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td>L-X, F-X, Y-X and W-X</td>
</tr>
<tr>
<td>Cyanogen Bromide</td>
<td>X-M</td>
</tr>
</tbody>
</table>
Using the CORRECT Enzyme

Trypsin: \textsubscript{309}AAL\textsubscript{356}EFEDGMYANLGI\textsubscript{332}GLIPLL\textsubscript{344}NSF\textsubscript{332}ISP\textsubscript{344}NM\textsubscript{356}TVH\textsubscript{356}LQ\textsubscript{356}SEN\textsubscript{344}G\textsubscript{356}VLG\textsubscript{356}GL\textsubscript{356}PYP\textsubscript{356}KL

Chymotrypsin: \textsubscript{329}A\textsubscript{332}SNF

\textit{GluC}: \textsubscript{315}DGMYANLG\textsubscript{344}GIPLL\textsubscript{344}NSF\textsubscript{344}ISP\textsubscript{344}NM\textsubscript{344}TV\textsubscript{344}LQ\textsubscript{344}SE\textsubscript{344}G\textsubscript{344}L

\textbf{Phosphorylation}
### Optimize Digestion Conditions

Sequence coverage: 97.2%

<table>
<thead>
<tr>
<th>Start</th>
<th>Observed</th>
<th>Mr(calc)</th>
<th>ppm</th>
<th>M</th>
<th>Score</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA2-20</td>
<td>932.9369</td>
<td>1863.8592</td>
<td>1863.9</td>
<td>-0.34</td>
<td>0</td>
<td>104 M.ADESSDAAGEPQPAPAPVRR.R</td>
</tr>
<tr>
<td>AA2-21</td>
<td>674.3265</td>
<td>2019.9576</td>
<td>2020</td>
<td>-1.67</td>
<td>1</td>
<td>95 M.ADESSDAAGEPQPAPAPVRR.R</td>
</tr>
<tr>
<td>AA2-22</td>
<td>726.3605</td>
<td>2176.0598</td>
<td>2176.1</td>
<td>-1.06</td>
<td>2</td>
<td>66 M.ADESSDAAGEPQPAPAPVRR.R</td>
</tr>
<tr>
<td>AA23-37</td>
<td>555.9385</td>
<td>1664.7936</td>
<td>1664.8</td>
<td>1.78</td>
<td>1</td>
<td>55 R.SSANYRAYATEPHAK.K</td>
</tr>
<tr>
<td>AA29-37</td>
<td>494.2502</td>
<td>986.4858</td>
<td>986.48</td>
<td>3.69</td>
<td>0</td>
<td>26 R.AYATEPHAK.K</td>
</tr>
<tr>
<td>AA29-39</td>
<td>622.3464</td>
<td>1242.6782</td>
<td>1242.7</td>
<td>4.97</td>
<td>2</td>
<td>44 R.AYATEPHAK.K</td>
</tr>
<tr>
<td>AA29-41</td>
<td>486.9404</td>
<td>1457.7994</td>
<td>1457.8</td>
<td>0.25</td>
<td>3</td>
<td>35 R.AYATEPHAKKSK.K</td>
</tr>
<tr>
<td>AA29-46</td>
<td>658.37</td>
<td>1972.0882</td>
<td>1972.1</td>
<td>1.43</td>
<td>4</td>
<td>34 R.AYATEPHAKKSKSISASR.K</td>
</tr>
<tr>
<td>AA47-59</td>
<td>382.7495</td>
<td>1526.9691</td>
<td>1527</td>
<td>7.14</td>
<td>2</td>
<td>35 R.KLQLQIAMQIAKQ</td>
</tr>
<tr>
<td>AA52-64</td>
<td>530.952</td>
<td>1589.8341</td>
<td>1589.8</td>
<td>4.51</td>
<td>1</td>
<td>85 K.TLMLQIAKQ</td>
</tr>
<tr>
<td>AA52-69</td>
<td>735.7083</td>
<td>2204.1029</td>
<td>2204.1</td>
<td>4.57</td>
<td>2</td>
<td>49 K.TLMLQIAKQ</td>
</tr>
<tr>
<td>AA60-69</td>
<td>436.1964</td>
<td>1305.5675</td>
<td>1305.6</td>
<td>4.31</td>
<td>1</td>
<td>25 K.QEMEREAEAER.R</td>
</tr>
<tr>
<td>AA76-99</td>
<td>730.1263</td>
<td>2916.476</td>
<td>2916.5</td>
<td>5.98</td>
<td>2</td>
<td>31 R.VLRLTCQPLELDGFEELQDLCR.Q</td>
</tr>
<tr>
<td>AA79-99</td>
<td>850.4134</td>
<td>2548.2183</td>
<td>2548.2</td>
<td>5.24</td>
<td>1</td>
<td>95 R.TRCQPLELDGFEELQDLCR.Q</td>
</tr>
<tr>
<td>AA81-99</td>
<td>1146.5381</td>
<td>2921.0616</td>
<td>2921.1</td>
<td>2.36</td>
<td>0</td>
<td>115 R.CQPLELDGFEELQDLCR.Q</td>
</tr>
<tr>
<td>AA81-104</td>
<td>725.1068</td>
<td>2896.3981</td>
<td>2896.4</td>
<td>0.75</td>
<td>1</td>
<td>69 R.CQPLELDGFEELQDLCR.Q</td>
</tr>
<tr>
<td>AA105-118</td>
<td>565.6145</td>
<td>1693.8217</td>
<td>1693.8</td>
<td>3.43</td>
<td>2</td>
<td>41 R.VDKVDEERYVEAK.V</td>
</tr>
<tr>
<td>AA105-121</td>
<td>1012.028</td>
<td>2022.0415</td>
<td>2022</td>
<td>7.19</td>
<td>3</td>
<td>39 R.VDKVDEERYVEAKVT.K</td>
</tr>
<tr>
<td>AA119-132</td>
<td>787.4489</td>
<td>1572.8833</td>
<td>1572.9</td>
<td>7</td>
<td>1</td>
<td>63 K.VTKNITEIALTQK.I</td>
</tr>
<tr>
<td>AA119-137</td>
<td>745.4208</td>
<td>2233.2405</td>
<td>2233.2</td>
<td>3.91</td>
<td>2</td>
<td>65 K.VTKNITEIALTQKIDYLGR.F</td>
</tr>
<tr>
<td>AA119-139</td>
<td>605.5953</td>
<td>2418.3522</td>
<td>2418.3</td>
<td>1.68</td>
<td>3</td>
<td>54 K.VTKNITEIALTQKIDYLGR.K</td>
</tr>
<tr>
<td>AA122-132</td>
<td>623.3372</td>
<td>1244.6599</td>
<td>1244.7</td>
<td>-1.06</td>
<td>0</td>
<td>79 K.NITEIALTQK.I</td>
</tr>
<tr>
<td>AA122-137</td>
<td>953.5199</td>
<td>1905.0252</td>
<td>1905</td>
<td>2.38</td>
<td>1</td>
<td>66 K.NITEIALTQKIDYLDR.G</td>
</tr>
<tr>
<td>AA122-139</td>
<td>697.7211</td>
<td>2090.1414</td>
<td>2090.1</td>
<td>2.03</td>
<td>2</td>
<td>48 K.NITEIALTQKIDYLGR.K</td>
</tr>
<tr>
<td>AA122-139</td>
<td>697.7211</td>
<td>2090.1414</td>
<td>2090.1</td>
<td>2.03</td>
<td>2</td>
<td>48 K.NITEIALTQKIDYLGR.K</td>
</tr>
<tr>
<td>AA122-139</td>
<td>697.7211</td>
<td>2090.1414</td>
<td>2090.1</td>
<td>2.03</td>
<td>2</td>
<td>48 K.NITEIALTQKIDYLGR.K</td>
</tr>
<tr>
<td>AA138-147</td>
<td>420.2665</td>
<td>1257.7776</td>
<td>1257.8</td>
<td>0.46</td>
<td>3</td>
<td>26 R.GFKRPTLR.R</td>
</tr>
<tr>
<td>AA148-163</td>
<td>866.964</td>
<td>1731.9134</td>
<td>1731.9</td>
<td>0.61</td>
<td>1</td>
<td>79 R.VRISADAMMQALLGTR.A</td>
</tr>
<tr>
<td>AA150-163</td>
<td>739.3765</td>
<td>1476.7385</td>
<td>1476.7</td>
<td>-2.94</td>
<td>0</td>
<td>107 R.ISADAMMQALLGTR.A</td>
</tr>
<tr>
<td>AA164-171</td>
<td>466.2661</td>
<td>930.5176</td>
<td>930.51</td>
<td>4.52</td>
<td>1</td>
<td>58 R.AKESDLIR.A</td>
</tr>
<tr>
<td>AA176-187</td>
<td>505.9445</td>
<td>1514.8118</td>
<td>1514.8</td>
<td>4.36</td>
<td>3</td>
<td>59 K.QVKKEDIEKEN.R</td>
</tr>
<tr>
<td>AA188-205</td>
<td>678.3385</td>
<td>2031.9937</td>
<td>2032</td>
<td>6.93</td>
<td>2</td>
<td>43 R.EVGDKRNIKSDMGMEGR.K</td>
</tr>
<tr>
<td>AA194-205</td>
<td>645.8306</td>
<td>1289.6466</td>
<td>1289.6</td>
<td>5.29</td>
<td>1</td>
<td>78 R.KDIALSGMEGR.K</td>
</tr>
<tr>
<td>AA195-205</td>
<td>581.7822</td>
<td>1161.5499</td>
<td>1161.5</td>
<td>4.39</td>
<td>0</td>
<td>95 K.NIDALSGMEGR.K</td>
</tr>
<tr>
<td>AA195-206</td>
<td>645.8309</td>
<td>1289.6472</td>
<td>1289.6</td>
<td>5.77</td>
<td>1</td>
<td>54 K.NIDALSGMEGR.K</td>
</tr>
<tr>
<td>AA195-207</td>
<td>709.8787</td>
<td>1417.7414</td>
<td>1417.7</td>
<td>4.74</td>
<td>2</td>
<td>46 K.NIDALSGMEGR.K</td>
</tr>
<tr>
<td>AA195-211</td>
<td>627.329</td>
<td>1878.9651</td>
<td>1879</td>
<td>1.58</td>
<td>4</td>
<td>29 K.NIDALSGMEGRKKFEG.-</td>
</tr>
</tbody>
</table>

Trypsin; 1:200; RT; 20 min Digestion
Difficulties for PTMs Identification

- Generating Peptides containing modified residue
  --Different Enzymes/Digestion Condition

- Low population of modifications
  --Enrichment
  --Targeted Method (For better MSMS)
PTM Enrichment Techniques (Start with mg of samples)

Acetylation: Anti-acetyl lysine polyclonal antibody

Phosphopeptides: Immobilized Metal Affinity Chromatography (Fe3+)
Metal Oxide Affinity Chromatography (TiO2, ZrO2)
Reversible Covalent Binding
(Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. Mass Spectrom Rev. 2010 Jan-Feb;29(1):29-54.)
Phospho-Threonine Antibody (P-Thr-Polyclonal)

Nitration: Anti-Nitrotyrosine polyclonal antibody

Ubiquitination: K-ε-GG–specific antibody (enrichment has to be done prior to other treatment on lysine)

Glycopeptides: Lectin affinity enrichment
Covalent Interactions
Chromatographic separation
(Glycopeptide enrichment and separation for protein glycosylation analysis., J Sep Sci. 2012 Sep;35(18):2341-72.)

Methylation: Anti-methyl lysine/araginine antibody
Enrichment for Phosphorylation

Cell Culture

extract cells

TiO$_2$ beads

peptides

In-solution digestion

protein

cell lysis

2D LC-MS/MS Phosphopeptides/proteins Identification

Slide courtesy of Nilini S. Ranbaduge
Example---Targeted MS Scan (for Known Modifications)

Full Scan
100:1

SIC of 516.7793²⁺ (SLVLC(CAM)TPSR)

SIC of 479.2730²⁺ (SLVLfGlyTPSR)

MS/MS of 516.7793²⁺ (SLVLC(CAM)TPSR)

MS/MS of 479.2730²⁺ (SLVLfGlyTPSR)

NOT Observed!!!!

NL:
4.05E7
Base Peak
m/z = 516.7690-516.7896  MS 17625_1_rerun

NL:
5.98E5
Base Peak
m/z = 479.2634-479.2826  MS 17625_1_rerun

MS/MS of 516.7793²⁺ (SLVLC(CAM)TPSR)

MS/MS of 479.2730²⁺ (SLVLfGlyTPSR)

17625_1_rerun
#2715-9589
RT: 18.73-51.92
AV: 99
NL: 5.71E3

T: Average spectrum MS2 516.79 (2715-9589)
Example---Targeted MS Scan (for Known Modifications)

Full Scan

Targeted @ 479.2730^2+(SLVLfGlyTPSR)/516.7793^2+(SLVL(CAM)TPSR)

MS/MS of 516.7793^2+(SLVL(CAM)TPSR)

MS/MS of 479.2730^2+(SLVLfGlyTPSR)
Difficulties for PTMs Identification

- Generating Peptides containing modified residue
  --Different Enzymes/Digestion Condition

- Low population of modifications
  --Enrichment
  --Targeted Method

- Modification not stable
  --Experiment Condition (reducing reagents, enzymes)
  --Fresh sample
Adjust Experiment Condition to Preserve Modifications

```
185N A C187 G S G Y D F D V F V V R199
```

```
138L V E G C142 L V G G R146
```

**A**

```
274.25 373.31 520.42 619.48 734.51 996.57 1159.62 1216.57 1303.64 1360.66
y2A y3A y4A y5A y6A y8A y9A y10A y11A y12A
V F V D FD Y G S G
```

**B**

```
408.17 465.20 552.31 609.30 772.30 1034.35 1149.39 1248.41 1395.53 1494.50 1593.59
y3B y4B y5B y6B y7B y8A y9A y10A y11A y12A
G S G Y FD D V F V V
```

m/z
Using Different Enzyme to Protect Modifications

Theoretical m/z = 783.40892+, Observed m/z = 783.40802+, Mass Error = 1.15ppm

O=Pyrrolysine
### LTQ DATA—Fresh Sample

**Sequence Coverage 39%; T95 was phosphorylated**

<table>
<thead>
<tr>
<th>Start</th>
<th>Observed (kDa)</th>
<th>Mr(expt) (kDa)</th>
<th>Mr(calc) (kDa)</th>
<th>Delta</th>
<th>Score</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>976.3</td>
<td>1950.5854</td>
<td>1949.856</td>
<td>0.729</td>
<td></td>
<td>KSVSTPSEAGSQDSGDAGVGR.T</td>
</tr>
<tr>
<td>92-112</td>
<td>976.44</td>
<td>1950.8654</td>
<td>1950.84</td>
<td>0.025</td>
<td></td>
<td>94 KSVSTPSEAGSQDSGDAGVGR.T + Deamidated (NQ)</td>
</tr>
<tr>
<td>92-112</td>
<td>651.53</td>
<td>1951.5682</td>
<td>1950.84</td>
<td>0.728</td>
<td></td>
<td>66 KSVSTPSEAGSQDSGDAGVGR.T + Deamidated (NQ)</td>
</tr>
<tr>
<td>92-112</td>
<td>651.62</td>
<td>1951.8382</td>
<td>1950.84</td>
<td>0.998</td>
<td></td>
<td>81 KSVSTPSEAGSQDSGDAGVGR.T + Deamidated (NQ)</td>
</tr>
<tr>
<td>92-112</td>
<td>1016.28</td>
<td>2030.5454</td>
<td>2030.807</td>
<td>-0.261</td>
<td></td>
<td>86 KSVSTPSEAGSQDSGDAGVGR.T + Deamidated (NQ); Phospho (ST)</td>
</tr>
<tr>
<td>92-112</td>
<td>678</td>
<td>2030.9782</td>
<td>2030.807</td>
<td>0.172</td>
<td></td>
<td>59 KSVSTPSEAGSQDSGDAGVGR.T + Deamidated (NQ); Phospho (ST)</td>
</tr>
</tbody>
</table>
Orbitrap XL Data---Two Days old, Stored at 4C

Sequence Coverage 53%; No Phosphorylation Detected

<table>
<thead>
<tr>
<th>Start</th>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>ppm</th>
<th>Score</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>32-112</td>
<td>975.9351</td>
<td>1949.8556</td>
<td>1949.8563</td>
<td>-0.36</td>
<td>149</td>
<td>K.SVSTPSEAGSQDSGDGAVGSR.T</td>
</tr>
<tr>
<td>32-112</td>
<td>975.9354</td>
<td>1949.8562</td>
<td>1949.8563</td>
<td>-0.052</td>
<td>149</td>
<td>K.SVSTPSEAGSQDSGDGAVGSR.T</td>
</tr>
<tr>
<td>32-112</td>
<td>975.9377</td>
<td>1949.8608</td>
<td>1949.8563</td>
<td>2.33</td>
<td>143</td>
<td>K.SVSTPSEAGSQDSGDGAVGSR.T</td>
</tr>
<tr>
<td>32-112</td>
<td>975.9377</td>
<td>1949.8609</td>
<td>1949.8563</td>
<td>2.39</td>
<td>130</td>
<td>K.SVSTPSEAGSQDSGDGAVGSR.T</td>
</tr>
<tr>
<td>32-112</td>
<td>975.9386</td>
<td>1949.8626</td>
<td>1949.8563</td>
<td>3.27</td>
<td>125</td>
<td>K.SVSTPSEAGSQDSGDGAVGSR.T</td>
</tr>
<tr>
<td>32-112</td>
<td>975.9403</td>
<td>1949.8661</td>
<td>1949.8563</td>
<td>5.02</td>
<td>142</td>
<td>K.SVSTPSEAGSQDSGDGAVGSR.T</td>
</tr>
<tr>
<td>32-112</td>
<td>975.9407</td>
<td>1949.8668</td>
<td>1949.8563</td>
<td>5.39</td>
<td>135</td>
<td>K.SVSTPSEAGSQDSGDGAVGSR.T</td>
</tr>
</tbody>
</table>

Detected 30 times!!!
Difficulties for PTMs Identification

- Generating Peptides containing modified residue
  --Different Enzymes/Digestion Condition

- Low population of modifications
  --Enrichment
  --Targeted Method

- Modification not stable
  --Experiment Condition (reducing reagents, enzymes)
  --Fresh sample

- Modification labile in MS analysis
  --Neutral Loss Method
  --Different Fragmentation Method
Neutral Loss Scan

Scan Dissociate Scan

A

PTM-specific ion signal

Stable PTM

MS/MS

m/z

Δm (PTM, position)

B

PTM-specific neutral loss

Labile PTM

MS/MS

m/z

Δm (PTM = PTM)

(aa)

C - C-terminal amino acid residue

△ - Amino acid residue modified with stable PTM

N - N-terminal amino acid residue

Hydrothermal synthesis of α-Fe₂O₃@SnO₂ core–shell nanotubes for highly selective enrichment of phosphopeptides for mass spectrometry analysis. *Nanoscale*, 2010, 2, 1892-1900

**Alpha Casein**

Most abundant peak represents a loss of phosphoric acid (98 Daltons/2 = 49 due to the fact that the peptide has a +2 charge state).

**NL of Phosphorylation:** 98 (+1), 49 (+2)...

- Help Identify the Sequence but don’t Give Information about the Location of the Modification if Multiple Sites Coexisting.
- Sensitive not Ideal

 Fragmentation of the peak resulting from the neutral loss of phosphoric acid resulted in an information-rich spectrum that correctly identified a phosphorylation site on a protein.

Precursor mass selected for MS/MS fragmentation.
ECD/ETD: Keep the Phosphorylation Intact!!!

H$_3$PO$_4$ Group may Leave upon CID Dissociation; Actual Amino Acid Location of H$_3$PO$_4$ Group May not be Clear if >2 Sites in the Sequence

CID: a, b and y ion, loss of H2O, NH3, side-chain
ECD/ETD: c, y and z ions

Electron Capture Dissociation (ECD, ICR)/ Electron Transfer Dissociation (ETD, Traps) will Generate Peptide Fragments without Losing the Sidechain Phosphorylation

Sweet, Anal. Chem. 2006, 78, 7563-7569
Leann M. Mikesh et al. Biochimica et Biophysica Acta 1764 (2006) 1811–1822
Choose a Good Enzyme to Generate The GG Tag — the Identification of Ubiquitination by MS/MS

MQIFVKTLTG KTITLEVEPS DTIENVKAKI
QDKEGIPPDO QRLIFAGKQL EDGRTLSDYN
IQKESTLHLV LRLRGG

Trypsin!!!
Choose a Good Enzyme to Generate Signature Modification Group—the Identification of Ubiquitination by MS/MS
Unknown Truncation Sites Identification

Separate Suspected Truncation Product(s) from Intact Protein

Intact MS Measurement

Top down Sequencing

Separation on 1D SDS PAGE
Cut the band (truncated protein)
Digestion
LC/MSMS

Peptide Sequencing
MASCOT

Protein sequence

Suggested truncation sites
Search against new sequences

Validation
Example---Unknown Truncation Sites Identification

MASCOT RESULT:

1  MVLWILWRPF  GFSGRFLKLE  SHSITESSKL  IPVAWTSLTQ  MLLEAPGIFL
51  LGQRKRFSVM  PETETHERET  ELSFPPSDVR  GMTKLDRTAF  KKTNVIPVLK
101  VRKEIVSKL M  RSLKRAALQR  PGIRRVIEDP  EDKESRILML  DPYKIFTHDS
151  FEKAELSVLE  QINVSPQISK  YNLETYEHF  KSEEILRAVL  PEGQDVTSGF
201  SRIGHIAHNL  LRDHQLPFKH  LIGQVMIDKN  PGITSAVNIK  NNIDNMYRNF
251  QMEVLSEQVN  MMKVRENNY  TYLEDFSKVY  WNPRLSTEHS  RITLELLKPGD
301  VLFDVFAGVG  PFAIPVAKKN  CTVFANDLNP  ES HKWLYNC  KLNKVDQKV K
351  VFNLKDGEFL  QGPKVEELMQ  LLGLSERKRP  SVHVVMNLPA  KAIIEFLS AFK
401  WLLDGQPCSS  EFLP1VHCYS  FSDKANPAED  VRQAGAVLG  ISLEACSSVH
451  LVRRNAPNKE  MLCTFQIPA  SVLYKNQTRN  PENHEDPPLK  RQRTAEAFSD
501  EKTQIVSNT

~260AA = ~27-28kDa

Conservation of structure and mechanism by Trm5 enzymes RNA 2013 Sep; 19(9): 1192–1199
Example---Unknown Truncation Sites Identification

Constructed new sequences (a total of 24) and searched the data against these sequences:

>16502 Sequence 1 (241-509)
NNIDNMYRFQMEVLSGEQNMNMTKVRENNTYEFDFSKVYWNPLSTEHSRITELLKPGDVLFDVFAGVGPGFAIPVAKKNCTVFANDLPESHKWL
LKYCNKLKNVDQKVKNLDBGKDFLQGPKVEELMQLLGLSHERKPSVHVVMNLPACKAEFLSFKWLDDQGQPCSEFLPIVHCYSFKDANPAEDVRQR
AGAVLGISLEACSSVHLVRNVAPNKEMLCITFQIPASVLYKNQTRNPENHEDPLKRQRTAEAFSDEKTQIVSNT <

>16502 Sequence 2 (242-509)
NIDNMYRFQMEVLSGEQNMNMTKVRENNTYEFDFSKVYWNPLSTEHSRITELLKPGDVLFDVFAGVGPGFAIPVAKKNCTVFANDLPESHKWL
LKYCNKLKNVDQKVKNLDBGKDFLQGPKVEELMQLLGLSHERKPSVHVVMNLPACKAEFLSFKWLDDQGQPCSEFLPIVHCYSFKDANPAEDVRQR
AGAVLGISLEACSSVHLVRNVAPNKEMLCITFQIPASVLYKNQTRNPENHEDPLKRQRTAEAFSDEKTQIVSNT <

>16502 Sequence 23 (263-509)
TKVRENNTYEFDFSKVYWNPLSTEHSRITELLKPGDVLFDVFAGVGPGFAIPVAKKNCTVFANDLPESHKWL
LKYCNKLKNVDQKVKNLDBGKDFLQGPKVEELMQLLGLSHERKPSVHVVMNLPACKAEFLSFKWLDDQGQPCSEFLPIVHCYSFKDANPAEDVRQR
AGAVLGISLEACSSVHLVRNVAPNKEMLCITFQIPASVLYKNQTRNPENHEDPLKRQRTAEAFSDEKTQIVSNT <

>16502 Sequence 24 (264-509)
KVRENNTYEFDFSKVYWNPLSTEHSRITELLKPGDVLFDVFAGVGPGFAIPVAKKNCTVFANDLPESHKWL
LNCYCNKLKNVDQKVKNLDBGKDFLQGPKVEELMQLLGLSHERKPSVHVVMNLPACKAEFLSFKWLDDQGQPCSEFLPIVHCYSFKDANPAEDVRQR
AGAVLGISLEACSSVHLVRNVAPNKEMLCITFQIPASVLYKNQTRNPENHEDPLKRQRTAEAFSDEKTQIVSNT <

MASSMATRIX Results:

<table>
<thead>
<tr>
<th>hit#</th>
<th>score</th>
<th>decoy%</th>
<th>protein description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hit1</td>
<td>3533</td>
<td>0.00%</td>
<td>16502 Sequence 12 (252-509)</td>
</tr>
<tr>
<td>hit2</td>
<td>3522</td>
<td>0.00%</td>
<td>16502 Sequence 23 (263-509)</td>
</tr>
<tr>
<td>hit3</td>
<td>3519</td>
<td>0.00%</td>
<td>16502 Sequence 14 (254-509)</td>
</tr>
<tr>
<td>hit4</td>
<td>3512</td>
<td>0.00%</td>
<td>16502 Sequence 21 (261-509)</td>
</tr>
</tbody>
</table>
Observed $m/z = 749.3441^{2+}$
Theoretical $m/z = 749.3409^{2+}$
Mass Error $= 4.27$ ppm

Conservation of structure and mechanism by Trm5 enzymes RNA. 2013 Sep; 19(9): 1192–1199
Observed $m/z = 619.3016^{2+}$
Theoretical $m/z = 619.2994^{2+}$
Mass Error=3.55ppm

Conservation of structure and mechanism by Trm5 enzymes RNA. 2013 Sep; 19(9): 1192–1199
Conservation of structure and mechanism by Trm5 enzymes RNA. 2013 Sep; 19(9): 1192–1199
Conservation of structure and mechanism by Trm5 enzymes RNA. 2013 Sep; 19(9): 1192–1199
Glycosylation—A Different Story

“The cell surface landscape is richly decorated with oligosaccharides anchored to proteins or lipids within the plasma membrane. Cell surface oligosaccharides mediate the interactions of cells with each other and with extracellular matrix components.” *Science* 291:2337

Why Glycosylation Investigation is SO HARD???

- Only Protein Modification Requiring Detailed Structural Characterization (sugar chain/branched)
- Poor Ionization efficiency (large mass increase, sometimes has negative charge, hydrophobicity..)
- Heterogeneity
- Large Size
Enrichment (Lectin Affinity Binding)
Derivatization (Permethylation)
LC Separation (PGC Column for Glycans, HILIC Column for Glycopeptides)
Both MALDI and ESI will Work
High Mass Accuracy Desired
Negative Mode Works Better
ETD Improve the MSMS Identification