An Introduction to
Mass-Spectrometry-Based Proteomics

Nick Till
MacMillan Group Meeting 07/29/2020
**The Broader Context for MS-Based Proteomics**

### Protein-Interaction Networks

- Core-complex protein
- Bait protein
- Weak-interactor protein

### Subcellular Localization

- Mitochondrion
- Cytosol
- Nucleus
- Lysosome
- Endoplasmic reticulum

### Activity-based Profiling

- Reactive group
- Reporter group
- ABPP probe

### Biomarker Discovery

- Patients with signs and symptoms of the same cancer
- Measure biomarkers
- Patients with high and low levels
- Patients with low and high levels
proteomics: the study of proteomes and their functions (or the large scale study of proteins)

protein measurements can be direct read-outs of biological activity
The Broader Context for MS-Based Proteomics

DNA

RNA

RNAseq: transcript levels with good coverage

transcript levels can be misleading in predicting protein levels

protein

- translational efficiency
- post-translational modifications
- protein degradation kinetics can vary

**p53 regulation:** mediated by phosphorylation and proteasomal degradation

*p53 mRNA transcript levels are a poor indicator of p53 levels and activity*
**The Broader Context for MS-Based Proteomics**

**Large-Scale Yeast 2-Hybrid**

- Large-scale identification of PPIs

**Protein Microarrays**

- Large-scale protein-binding studies

**MS-based proteomics:** often the tool of choice for large-scale analysis of protein levels and interactions


An Overview of Topics Covered

- Part 1: basic workflow and technology for discovery proteomics
  - How proteins are handled and analyzed
  - Data Peptide assignment and protein inference

- Part 2: methods for (relative) quantitative proteomics
  - Label-free methods
  - Whole-cell isotopic labeling strategies
  - Chemical mass tags

- Part 3: targeted proteomics and its application to biomarker discovery
Technology Development and Workflow for MS Proteomics

cells or tissue sample → proteins → samples

proteins

proteome analysis
Technology Development and Workflow for MS Proteomics

~20,000 proteins
(not counting PTMS, alternative splicing)

Intensity

m/z

window ≈ 1500 m/z

direct injection results in high complexity MS

separation step necessary prior to MS analysis
Technology Development and Workflow for MS Proteomics

image analysis identifies up/down regulation

isolate spot from gel

protein ID by MS

Technology Development and Workflow for MS Proteomics

- Protein MW > 10,000 Da
- Poor recovery by LC
- Peptide MWs < 4,000 Da
- Good recovery across peptides

"Bottom-up proteomics" data analysis involves reconstructing protein identity.

Peptide fragments (incomplete) → Protein identity

Data analysis involves reconstructing protein (protein inference)
**Technology Development and Workflow for MS Proteomics**

**trypsin cleavage:** high-specificity serine protease cleaves after K (lysine) or R (arginine) residues

Technology Development and Workflow for MS Proteomics

N-terminus—AVTKWGSRAGPAVTKKEIGAASTQVRAGDSLQPKGTVALER
N-terminus—AVTKWGSRAGPAVTKKEIGAASTQVRAGDSLQPKGTVALER

large mixture of tryptic peptides are then subjected to LC/MS² analysis

<table>
<thead>
<tr>
<th>protease</th>
<th>cleavage specificity¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>trypsin</td>
<td>-K,R,t1-Z; not -K,R,t1-P;</td>
</tr>
<tr>
<td>endoproteinase Lys-C</td>
<td>-K,t1-Z;</td>
</tr>
<tr>
<td>chymotrypsin</td>
<td>-W,F,Y,t1-Z; and -L,M,A,D,E,t1-Z; at a slower rate</td>
</tr>
<tr>
<td>subtilisin</td>
<td>broad specificity to native and denatured proteins</td>
</tr>
<tr>
<td>elastase</td>
<td>-B,t1-Z;</td>
</tr>
<tr>
<td>endoproteinase Lys-N</td>
<td>-Z,t1-K;</td>
</tr>
<tr>
<td>endoproteinase Glu-C</td>
<td>-E,t1-Z; and 3000 times slower at -D,t1-Z;</td>
</tr>
<tr>
<td>endoproteinase Arg-C</td>
<td>-R,t1-Z;</td>
</tr>
<tr>
<td>endoproteinase Asp-N</td>
<td>-Z,t1-D; and -Z,t1-cysteic acid; but not -Z,t1-C;</td>
</tr>
<tr>
<td>proteinase K</td>
<td>-X,t1-Y;</td>
</tr>
<tr>
<td>OmpT</td>
<td>-K, R,t1-K, R;</td>
</tr>
</tbody>
</table>

¹B = uncharged, nonaromatic amino acids (i.e., A, V, L, I, G, S)

X = aliphatic, aromatic, or hydrophobic amino acids; and Z = any amino acid.

Technology Development and Workflow for MS Proteomics

MS-proteomics sample preparation and workflow

Cells or tissue → Proteins → Peptides → UHPLC → Mass spectrometer

Step 1: Extraction
Step 2*: Digestion
Step 3*: Enrichment for PTMs

common context-specific modifications

subcellular fractionation
protein interactions
enrichment for PTMs

Meissner, F.; Mann, M. Nature Immunology 2014, 15, 112–117.
Technology Development and Workflow for MS Proteomics

MS/MS (MS²) analysis: peptides are further fragmented into ions for sequence identification

- multiple methods for MS/MS analysis
  - QqQ (triple quadrupole)
  - Q-TOF (quadrupole time-of-flight)
  - Q Exactive (quadrupole orbitrap)

- multiple modes of fragmentation
  - CID (collision-induced dissociation)
  - ECD (electron-capture-dissociation)
  - observed ions depend on method

Q Exactive setup (Thermo Fisher)

- collision cell (fragmentation)
- quadrupole (mass filter)
- orbitrap (mass detector)
- nanospray (ionization)
Technology Development and Workflow for MS Proteomics

This is an idealized picture, in reality fragmentation is incomplete and requires more analysis for peptide ID.
observed MS$^2$ spectrum

database candidate 1

missing peaks penalized

database candidate 2

low intensity
low score

database candidate 3

perfect match

observed peptide MS$^2$ spectra are scored against database MS$^2$ spectra to identify parent ion

- SEQUEST
- MASCOT
- OMSSA
- X!Tandem
- MaxQuant

Technology Development and Workflow for MS Proteomics

In bottom-up MS-proteomics, peptides (not proteins), are directly measured.

Protein inference is the process of extrapolating protein information from peptide measurements.

Technology Development and Workflow for MS Proteomics

A and B can be identified

B cannot be identified

sequence homology complicates analysis
- protein families
- alternative splicing
- protein isoforms (and point mutations)

peptide signal poses additional challenge
- low intensity ions (especially with DDA)
- small proteins (few tryptic sites)
- PTMs suppress signal

>30% of protein assignments are made based on a single peptide ID

Technology Development and Workflow for MS Proteomics

peptide coverage information and much more available on proteomics databases

- jPOST
- MassIVE
- ProteomicsDB
- PeptideAtlas
- MaxQB
An Overview of Topics Covered

- **Part 1: basic workflow and technology for discovery proteomics**
  - How proteins are handled and analyzed
  - Data Peptide assignment and protein inference

- **Part 2: methods for (relative) quantitative proteomics**
  - Label-free methods
  - Whole-cell isotopic labeling strategies
  - Chemical mass tags

- **Part 3: targeted proteomics and its application to biomarker discovery**
Digestion efficiency is protein-dependent:
  relative amounts of A and B may not
  be reflected by [peptide]

Ionization efficiencies vary by >100-fold

Matrix effects can change signal intensity (as for all MS)
Spectral counting or XIC used to compare abundances of a protein

Housekeeping proteins are typically utilized for concentration normalization

Many experimental arms can be compared, but many replicates (high $n$) often needed
Methods and Applications of Quantitative MS-Proteomics

Method 1: label-free quantitative proteomics

Data-dependent acquisition (DDA): only the highest intensity MS$^1$ precursor ions are selected for MS$^2$ analysis

10-20 most intense ions pass to CID cell per MS$^1$ scan (previously 3-8)

Spectral counting: instances of peptide MS$^1$ ion observation (verified by MS$^2$) summed up for relative quantitation

or

Extracted ion chromatogram (XIC): integrate ion intensity vs. time plot to quantitate peptide (MS$^1$)
νₔ: gas flow directs ions into TIMS (trapped ion mobility spectrometer)

E: electric field opposes gas flow, causing ion trapping

mobility depends on collisional cross section (CCS)

ions eluted by ramping down electric field strength
Methods and Applications of Quantitative MS-Proteomics

**Advantages of TIMS-TOF technology**

- Added dimension of separation
- Faster scanning speed
- Deeper protein coverage
  (can ID 2000 proteins from 10 cells of material)

Separation based on CCS is impressive


Method 2: stable isotope labeling with amino acids in cell culture (SILAC)

controls for ionization efficiency
differences and matrix effects

trypsin cut sites

...AVTKWGSRAGPAVTKEIGAASTQVRAGDSLQPKGTVALER...

\(^{13}\text{C}_6\text{Lys and }^{13}\text{C}_6^{15}\text{N}_2\text{Arg are optimal}\)
SILAC is well-suited to measure protein turnover kinetics

- heavy peptide signal decreases with time
- light peptide signal increases with time

MHC class I peptide display

- important for self/non-self recognition, tumor immunity
- peptide display requires proteasomal degradation


SILAC is well-suited to measure protein turnover kinetics

<table>
<thead>
<tr>
<th>No.</th>
<th>Mass</th>
<th>Sequence</th>
<th>Protein</th>
<th>Complete turnover time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1210.7</td>
<td>LLDVPTAAVQA</td>
<td>γ-Interferon-inducible protein IP-30 precursor</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>1011.5</td>
<td>LLDVPTAAV</td>
<td>γ-Interferon-inducible protein IP-30 precursor</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>951.6</td>
<td>LLGPRLVLA</td>
<td>TMP21; transmembrane trafficking protein</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>965.6</td>
<td>ALATUHQV</td>
<td>COP9 complex subunit 7a</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>913.6</td>
<td>GILLTLVQL</td>
<td>Catenin β1</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>885.8</td>
<td>LLLPQLATA</td>
<td>NADH dehydrogenase (ubiquinone) 1α subcomplex</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>861.5</td>
<td>ILQFTTL</td>
<td>3-Hydroxy-3-methylglutaryl-coenzyme A reductase</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>1037.7</td>
<td>KLLLEPVLL</td>
<td>Similar to 40 S ribosomal protein S16</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>1034.6</td>
<td>FVFPHELIL</td>
<td>Solute carrier family 1 (neutral amino acid transporter), member 5</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>910.5</td>
<td>ALPPVLT</td>
<td>Unnamed protein product</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>946.5</td>
<td>SLVEEADLA</td>
<td>Hypothetical protein FLJ30668</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>1079.7</td>
<td>VLLKARLVPFA</td>
<td>NPD019</td>
<td>9</td>
</tr>
<tr>
<td>13</td>
<td>947.5</td>
<td>ALYYAVNNV</td>
<td>Seven-transmembrane domain protein</td>
<td>9</td>
</tr>
<tr>
<td>14</td>
<td>1121.5</td>
<td>TLWVPVEV</td>
<td>B-cell translocation protein 1</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>959.5</td>
<td>SLFGGVVI</td>
<td>Polymerase (DNA-directed), α (70 kDa)</td>
<td>9</td>
</tr>
<tr>
<td>16</td>
<td>974.6</td>
<td>AILPTSIFL</td>
<td>SKB1 homolog</td>
<td>9</td>
</tr>
<tr>
<td>17</td>
<td>846.5</td>
<td>ALRSITSV</td>
<td>Unknown (protein for MGC:14124)</td>
<td>9</td>
</tr>
<tr>
<td>18</td>
<td>885.5</td>
<td>ALVGLNV</td>
<td>Progestin and adipQ receptor family member IV</td>
<td>9</td>
</tr>
<tr>
<td>19</td>
<td>900.5</td>
<td>ALFGVALL</td>
<td>Protein-disulfide isomerase mER80 precursor</td>
<td>12</td>
</tr>
<tr>
<td>20</td>
<td>1015.5</td>
<td>FDVPVLTV</td>
<td>Transcription intermediary factor 1</td>
<td>18</td>
</tr>
<tr>
<td>21</td>
<td>1032.5</td>
<td>ALPEFTEL</td>
<td>Similar to eukaryotic translation initiation factor 2, 26 subunit 3γ, 52 kDa</td>
<td>24</td>
</tr>
<tr>
<td>22</td>
<td>1094.6</td>
<td>SLPPDIALVG</td>
<td>SEC23 protein</td>
<td>24</td>
</tr>
<tr>
<td>23</td>
<td>1306.7</td>
<td>ALWDETQGTV</td>
<td>Guanine nucleotide-binding protein, β-2 subunit</td>
<td>24</td>
</tr>
<tr>
<td>24</td>
<td>1115.5</td>
<td>SLFEGTWL</td>
<td>Hydroxymethylglutaryl-CoA synthase</td>
<td>24</td>
</tr>
<tr>
<td>25</td>
<td>999.6</td>
<td>VIAELRGV</td>
<td>Nucleolar protein 5A</td>
<td>24</td>
</tr>
<tr>
<td>26</td>
<td>984.5</td>
<td>ALMPVLNOV</td>
<td>Homolog of yeast mRNA transport regulator 3</td>
<td>48</td>
</tr>
<tr>
<td>27</td>
<td>908.5</td>
<td>NLTSVFI</td>
<td>Similar to RIKEN cDNA 2610003J06</td>
<td>48</td>
</tr>
<tr>
<td>28</td>
<td>1258.6</td>
<td>FLFGSPTYVL</td>
<td>Fatty-acid synthase</td>
<td>48</td>
</tr>
<tr>
<td>29</td>
<td>999.5</td>
<td>ILQGSSFLG</td>
<td>ET putative translation product</td>
<td>48</td>
</tr>
<tr>
<td>30</td>
<td>968.6</td>
<td>SLDPVEV</td>
<td>Similar to RIKEN cDNA 431004K08</td>
<td>48</td>
</tr>
<tr>
<td>31</td>
<td>1020.8</td>
<td>FLRSIVQNL</td>
<td>Proteasome 26 S non-ATPase subunit 1</td>
<td>96</td>
</tr>
<tr>
<td>32</td>
<td>1038.6</td>
<td>YLLPAVHI</td>
<td>DEAD box polypeptide 17 isoform p82; probable</td>
<td>96</td>
</tr>
<tr>
<td>33</td>
<td>929.5</td>
<td>SLDDKIGA</td>
<td>RNA-dependent helicase p72</td>
<td>96</td>
</tr>
<tr>
<td>34</td>
<td>1049.6</td>
<td>VLQDLAFFL</td>
<td>Polymerase I and transcript release factor</td>
<td>96</td>
</tr>
<tr>
<td>35</td>
<td>786.4</td>
<td>SLAGGLGV</td>
<td>Protein similar to heterogeneous nuclear ribonucleoprotein K</td>
<td>24</td>
</tr>
</tbody>
</table>

Methods and Applications of Quantitative MS-Proteomics

Milner, E.; Barnea, E.; Beer, I.; Admon, A. Molecular & Cellular Proteomics 2006, 5, 357–365
Methods and Applications of Quantitative MS-Proteomics

Method 3: chemical labelling with isotopically-labeled tags

1. combine samples
2. compare heavy/light conjugates (MS)

Current state-of-the-art: isobaric mass tags (iTRAQ, TMT)

\[ \text{reporter ion} + \text{mass normalizer} = \text{constant} \]

- sites of $^{15}$N and $^{13}$C labels

Methods and Applications of Quantitative MS-Proteomics

Method 3: chemical labelling with isotopically-labeled tags

- Up to 11-plex possible today
- $^{13}$C and $^{15}$N, no $^2$H used
- High resolution MS$^2$ needed

Rauniyar, N.; Yates, J. R. J. Proteome Res. 2014, 13, 5293–5309
Methods and Applications of Quantitative MS-Proteomics

PROTACs cause changes in expression levels at the proteome level

How can these changes be assigned to degradation or transcriptional regulation?


Savitski, M. M.; Zinn, N.; Faelt-Savitski, M.;…Bantscheff, M. Cell 2018, 173, 260-274.e25
Combining SILAC and TMT to study protein regulation mechanisms

Methods and Applications of Quantitative MS-Proteomics
FYTDD1 critical for mRNA nuclear export, hence protein synthesis

off target degradation by JQ1-VHL
Methods and Applications of Quantitative MS-Proteomics

MS-based protein stability/small molecule binding assay

direct interaction between JQ1-VHL and FYTDD1 confirmed,
new PROTAC prepared without off-target activity
**Reductive Dimethylation**: alternative chemical labeling strategy with stable isotopes

---

**Methods and Applications of Quantitative MS-Proteomics**

- **13CD2H or 12CH3**

**Image Description**:

- T cells are labeled with SILAC light.
- (1) Vehicle change to SILAC heavy.
- (2) Vehicle SILAC light.
- (3) HSP90 inhibitor change to SILAC heavy.

**Chemical Structures**:

- 

**Graphs**:

- **Protein turnover, vehicle control**
  - Signal in MS^1
  - Degradation
  - New synthesis

- **Protein turnover, HSP90 inhibitor**
  - Signal in MS^1
  - Degradation
  - New synthesis

**Log2 protein FC**

- **Mature**
  - Treatment vs. vehicle control (donor 1)
- **Nascent**
  - Treatment vs. vehicle control (donor 2)

**References**:

- Savitski, M. M.; Zinn, N.; Faelth-Savitski, M.;...Bantscheff, M. *Cell* 2018, 173, 260-274.e25
An Overview of Topics Covered

- Part 1: basic workflow and technology for discovery proteomics
  - How proteins are handled and analyzed
  - Data Peptide assignment and protein inference

- Part 2: methods for (relative) quantitative proteomics
  - Label-free methods
  - Whole-cell isotopic labeling strategies
  - Chemical mass tags

- Part 3: targeted proteomics and its application to biomarker discovery
Beyond DDA-based Shotgun Proteomics

**DDA:** most common approach to discovery proteomics, hypothesis-free

**SRM:** accurate, reproducible quantitation of up to ~500 peptides (chosen)

**DIA:** accurate, reproducible, deep coverage, but requires a DDA measurement first

**Technology Development and Workflow for MS Proteomics**

Biomarkers “generate clinically useful information that could be used to change the course of the disease for a patient”

- Biomarker signature development utilizes targeted proteomics