Proteomics and Mass Spectrometry

Ron Bose, MD PhD
Biochemistry and Molecular Cell Biology Programs
Lab: 4515 McKinley Research Building, 3rd floor
Washington University School of Medicine
Introduction

Definition of Proteomics:
The large scale identification and characterization of proteins in a cell, tissue, or organism.

Traditional Biochemistry

Proteomics

http://www.chem.purdue.edu/people/faculty/Images/Tao%20proteomics-cartoon.jpg
Introduction

Definition of Proteomics:
The large scale identification and characterization of proteins in a cell, tissue, or organism.

Well Established Methods for Proteomics
1. 2D-gels
2. Mass Spectrometry

Methods still under development
1. Protein Arrays
2. Antibody Arrays
3. Proteome-wide coverage with Antibodies
2 Dimensional Gel Electrophoresis

First Dimension: pI by Isoelectric Focusing
Second Dimension: MW by standard SDS-PAGE

- First Published in 1975 by Pat O’Farrell
- Can separate at least 1,000 proteins
- Problems with run to run reproducibility limits the ability to easily compare multiple samples.

- Solution to this problem: DIGE (Difference Imaging Gel Electrophoresis)
DIGE experiment

Samples
- control
- treated
- pooled internal standard

Labeled proteins
- label with Cy3
- label with Cy5
- label with Cy2

Co-migration in 2-D electrophoresis
- IPG strip
- SDS gel

Image analysis
- Scan at different wavelengths
- Cy3 image
- Cy5 image
- Cy2 image

Slide courtesy of Tracy Andacht
DIGE experiment

Data from the labs of Tim Ley and Reid Townsend

Bredemeyer et al., PNAS 101:11785, 2004
Limitations of DIGE

1. Protein solubility during Isoelectric Focusing.
   - Membrane proteins often lost.

2. Size Limits – difficulty with proteins >100 kD.

3. Identification of the proteins in each spot is tedious and slow.
   - Use of robotics

4. Individual spots typically contain several proteins.
   - Intensity change is therefore the sum of the changes of each individual protein.
Principles of Mass Spectrometry

The Importance of Mass:
1. The mass of a molecule is a fundamental physical property of a molecule.
2. Mass can be used to identify the molecule.

Fragmentation provides Chemical Structure:
If you fragment a molecule in a predictable manner and make measurements on the individual fragments, you can determine the chemical structure of the molecule.
Biological Applications of Mass Spectrometry

1. Peptides and Proteins
2. Lipids
3. Oligosaccharides

MALDI-TOF spectrum of a synthesized 25mer peptide.

Measured mass = 2740.6 Da

Calculated mass = 2741.1 Da
Biological Applications of Mass Spectrometry

1. Peptides and Proteins
2. Lipids
3. Oligosaccharides

Methodology to identify lipids by mass spectrometry.


Figure 1. Schematic diagram of shotgun lipidomics based on intrasource separation and multidimensional MS analysis for global lipidome profiling and quantitation directly from a crude extract of a biologic sample.
Biological Applications of Mass Spectrometry

1. Peptides and Proteins
2. Lipids
3. Oligosaccharides: Analysis of Milk

Applications of Mass Spectrometry in the Physical Sciences

Widely used in Analytical Chemistry and Organic Chemistry.

Examples:

- Analyzing of drugs during chemical synthesis
- Identifying chemicals molecules or checking for contaminants.
- Environmental
  - Measuring toxins such as PCB and Heavy Metals
- Geology
  - Analyzing petroleum or petrochemicals
Applications of Mass Spectrometry in the Physical Sciences

Space Exploration: Mars Curiosity Rover

Sources: www.nasa.gov and Los Alamos National Laboratory
Applications of Mass Spectrometry in the Physical Sciences

Space Exploration: Mars Curiosity Rover

Sample Analysis at Mars (SAM) Instrument Suite

1. Mass Spectrometer
2. Gas Chromatograph
3. Laser Spectrometer

Sources: www.nasa.gov and Los Alamos National Laboratory
Applications of Mass Spectrometry in the Physical Sciences

Undersea Exploration: Deep Water Horizon Spill
Applications of Mass Spectrometry in the Physical Sciences

Undersea Exploration: Deep Water Horizon Spill

Scientific instruments used to measure the oil spill, including Mass Spectrometers for chemical analysis.
Applications of Mass Spectrometry in the Physical Sciences

Anti – Terrorism and Civil Defense:

 IonScan Mass Spectrometry Used at Airports and other facilities for the detection of Explosives and Narcotics.

Manufacturer: Smiths Detection
Trypsin – a protease that cleaves after basic residues (R or K).

Protein of Interest:
Identifying a Protein by Mass Spectrometry on Its Tryptic Peptides

Products from Trypsin digest.

Average length of tryptic peptides = 10 aa residues

Slide courtesy of Andrew Link
Identifying a Protein by Mass Spectrometry on Its Tryptic Peptides

Select an Individual Peptide in the Mass Spectrometer

Performed by adjusting the electrical fields in the mass spectrometer.

Slide courtesy of Andrew Link
Identifying a Protein by Mass Spectrometry on Its Tryptic Peptides

Impart energy to the peptide by colliding it with an inert gas (Argon or Helium).
Identifying a Protein by Mass Spectrometry on Its Tryptic Peptides

Measure the masses of the fragment ions.

Slide courtesy of Andrew Link
Identifying a Protein by Mass Spectrometry on Its Tryptic Peptides

The mass difference between the peaks corresponds directly to the amino acid sequence.

B-ions contain the N-terminus.
Identifying a Protein by Mass Spectrometry on Its Tryptic Peptides

$Y$-ions contain the C-terminus

Slide courtesy of Andrew Link
Identifying a Protein by Mass Spectrometry on Its Tryptic Peptides

The entire spectrum contains $B$-ions, $Y$-ions, and other fragment ions.
Identifying a Protein by Mass Spectrometry on Its Tryptic Peptides

The puzzle: The B, Y, and other ions occur together and we cannot distinguish them just by simple inspection of the spectrum.
Identifying a Protein by Mass Spectrometry on Its Tryptic Peptides

Actual spectra also have noise (either chemical noise or electrical noise).
Identifying a Protein by Mass Spectrometry on Its Tryptic Peptides

The final spectrum: the interpretation requires experience and aid by software algorithms.
Software for Interpreting Peptide Mass Spectra

**Statistical Matching**
Work by statistically matching the measured spectra with the theoretical spectra of all possible tryptic peptides from an organism.

1. SeQuest
2. MASCOT
3. X! Tandem
4. OMSSA

Requires a fully sequenced genome.

**De novo sequencing** (determines a peptide sequence based on the spacings of the fragment ions).

1. PepNovo
Example of an Actual Spectrum

Peptide 326-334 with phosphorylation on Y326

pY Imm.
The Hardware for Peptide Mass Spectrometry

Different Types:
- Electrospray
- MALDI

Ionization Source ➔ Mass Analyzer ➔ Detector

Time of Flight (TOF)
- Quadropole
- Ion Trap
- OrbiTrap
- Ion Cyclotron Resonance (ICR)

Output: Spectra

Liquid Chromatography

Vacuum Pump

Pump
The Hardware for Peptide Mass Spectrometry

- Liquid Chromatography
- Ionization Source
- Mass Analyzer and Detector
- Vacuum Pumps
Movie of MALDI – TOF mass spectrometer.

http://www.youtube.com/watch?v=OKxRx0ctrl0

Movie of FT-ICR mass spectrometer.

http://www.youtube.com/watch?v=a5aLIm9q-Xc&feature=related
Limitations and Cautions of Proteomics: The Range of Protein Concentrations In Yeast

Picotti et al., Cell – Aug 21, 2009
Limitations and Cautions of Proteomics:

The Range of Protein Concentrations In Human Plasma

**Classical Plasma Proteins**

3 - 4 log range of Mass Spectrometers

<table>
<thead>
<tr>
<th>Protein</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albumin</td>
<td>40 g/l</td>
</tr>
<tr>
<td>C4 Complement Factor B</td>
<td>0.1 g/l</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>&lt; 100 µg/l</td>
</tr>
<tr>
<td>TNFα</td>
<td>&lt; 1 ng/l</td>
</tr>
</tbody>
</table>

Anderson & Anderson, MCP 1:845, 2002
Limitations and Cautions of Proteomics: The Range of Protein Concentrations In Human Plasma

Depletion
Remove abundant proteins that are not of interest to your experiment. Methods: Antibody based depletion, selective lysis technique, subcellular fractionation, etc.

Enrichment
Enrich for the proteins of interest.
Methods – Lysis techniques or subcellular fractionation, affinity-based enrichment (antibodies, resins, etc).

Fractionation
Reduce the complexity of your sample by separating the proteins into different fractions and running these fractions separately.
Examples of Proteomic Experiments

1. Identification of Single Proteins
2. Identification of Proteins in the Nuclear Pore Complex
3. Identification of Proteins in the Secretory Pathway
4. Quantitative Measurement of Signal Transduction Pathways
Mary Olanich, a graduate student in Jason Weber lab, wanted to identify proteins binding to the untranslated regions (UTR) of the NPM mRNA.

She performed a pull-down assay with biotinylated NPM mRNA. Protein bands obtained were visualized with a fluorescent protein stain.

Single bands were cut from the gel and proteins ID’ed by MS.

Olanich et al., Oncogene 30(1):77-86, 2011.
Yeast Nuclear Pore Complexes are 50 MDa in size.
Contain approximately 30 different proteins.
Total number of proteins in the NPC is at least 456.

Strategy to Identify NPC Proteins

1. Make a highly pure NPC preparation
2. Extensive fractionation and Mass Spec protein identification.
3. Validate results with:
   a. Immunofluorescence
   b. Epitope tagging
   c. Immuno-electron microscopy

Rout et al., J Cell Bio 148:635-651, 2000
Strategy to Identify NPC Proteins

Blue = Known NPC associating proteins
Red = Proteins believed not to be NPC associated

Rout et al., J Cell Bio 148:635-651, 2000
Strategy to Identify NPC Proteins

Each band was cut out and digested with trypsin.

Mass Spec analysis was done by looking at the MS spectra and the MS/MS spectra.

MS spectrum of a mixture of 3 yeast proteins, all about 120 kD size, and trypsin auto-digestion peptides (marked by T).

Each peak can be isolated in the Mass Spectrometer and then fragmented to give MS/MS spectra and peptide sequence information.

Rout et al., J Cell Bio 148:635-651, 2000
• Started with a high quality preparation of Rough Microsomes (RM), Smooth Microsomes (SM), and Golgi apparatus (G).

• Fractionate the proteins on SDS-PAGE, cut thin slices of gel, digest with trypsin and run on Mass Spec.

Gilchrist et al., Cell 127:1265-81, 2006
Identification of Secretory Pathway Proteins

They identified over 1400 proteins and divided them into 23 functional categories.

Semi-quantitative measurements of protein abundance were made by spectral counting (ie – the number of observed spectra for a protein correlates with its abundance).

Gilchrist et al., Cell 127:1265-81, 2006
Protein Quantitation with Mass Spectrometry

- In Western blots, each antigen-antibody pair has a different affinity and “response characteristics.”
- Therefore, we can make comparison protein A in sample 1 vs. 2 vs. 3, but not protein A vs. protein B in the same sample.
- Similarly, in Mass Spec, every peptide has its own ionization and detection characteristics.
1. Stable Isotope Labels based Quantitation

   Examples of Stable Isotopes: $^{13}$C, $^{15}$N, $^{2}$H, $^{18}$O

   Advantage of Stable Isotopes: They are easy separated and distinguished in the Mass Spec.

   Approach: An internal comparison within one Mass Spec run. Different samples can be “labeled” with different isotopes.

   Advantages: Precision of quantitation, less susceptible to artifacts in Mass Spec runs.

   Limitations: Cost of isotopes. Limited number of isotope combinations are feasible.

2. Label-free Quantitation – No isotopes used.
Please, Consider the Following:

Isotopes of Carbon

<table>
<thead>
<tr>
<th>Isotope</th>
<th>Mass</th>
<th>Abundance in Nature</th>
<th>Half-life</th>
<th>Radioactivity release</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{12}\text{C}$</td>
<td>12 exactly</td>
<td>98.9%</td>
<td>Stable</td>
<td>None</td>
</tr>
<tr>
<td>$^{13}\text{C}$</td>
<td>13.003</td>
<td>1.07%</td>
<td>Stable</td>
<td>None</td>
</tr>
<tr>
<td>$^{14}\text{C}$</td>
<td>14.003</td>
<td>Trace</td>
<td>5,700 years</td>
<td>$\beta$ particle</td>
</tr>
<tr>
<td>$^{11}\text{C}$</td>
<td>11.011</td>
<td>Non-natural</td>
<td>20 min</td>
<td>positron</td>
</tr>
</tbody>
</table>
Please, Consider the Following:

Isotopes of Carbon

<table>
<thead>
<tr>
<th>Isotope</th>
<th>Mass</th>
<th>Abundance in Nature</th>
<th>Half-life</th>
<th>Radioactivity release</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{12}\text{C}$</td>
<td>12 exactly</td>
<td>98.9%</td>
<td>Stable</td>
<td>None</td>
</tr>
<tr>
<td>$^{13}\text{C}$</td>
<td>13.003</td>
<td>1.07%</td>
<td>Stable</td>
<td>None</td>
</tr>
<tr>
<td>$^{14}\text{C}$</td>
<td>14.003</td>
<td>Trace</td>
<td>5,700 years</td>
<td>$\beta$ particle</td>
</tr>
<tr>
<td>$^{11}\text{C}$</td>
<td>11.011</td>
<td>Non-natural</td>
<td>20 min</td>
<td>positron</td>
</tr>
</tbody>
</table>

Commonly used in Mass Spectrometry for Quantitative Measurements

DO NOT USE IN MASS SPEC.
Protein Quantitation with Mass Spectrometry

Introduce Stable Isotope by Metabolic Labeling

Control

- Normal arginine medium

Treatment 1

- $^{13}\text{C}_6$- arginine medium (+ 6 Da)

Treatment 2

- $^{13}\text{C}_6^{15}\text{N}_4$- arginine medium (+ 10 Da)

Mix Lysates

Fractionate Proteins on SDS-PAGE

Digest Bands with Trypsin

Identify and Quantify Proteins by Mass Spec

Bose et al., PNAS 103: 9773-8, 2006
Protein Quantitation with Mass Spectrometry

Protein 1  
+0  500.304
+6  503.309
+10 VGQAQDILR  505.303

Protein 2  
+0  638.930
+6  640.925
+10 VAGQSSPSGIQSR  642.405

Protein 3  
+0  635.924
+6  637.405
+10 FFEILSPVYR  640.412

Protein 4  
+0  459.781
+6  462.788
+10 HDGAFLIR  464.782

Key
+0  Control  $^{12}$C-Arginine
+6  Treatment 1  $^{13}$C$_6$-Arginine
+10 Treatment 2  $^{13}$C$_6^{15}$N$_4$-Arginine

Bose et al., PNAS 103: 9773-8, 2006
Protein Quantification with Mass Spectrometry

Introduce Stable Isotope by Chemical Labeling

- Amine reactive tags – iTRAQ (Ross et al., MCP 3:1154, 2003)

![Diagram of iTRAQ tag](image)

- Cys reactive tags - ICAT
- Incorporating $^{18}$O during Trypsin digestion
Studying EGFR Signal Transduction with Quantitative Proteomics

Introduce Stable Isotope by Chemical Labeling

Zhang et al., MCP 4: 1240-50, 2005
Mapping Her2/neu Tyrosine Kinase Signaling using Quantitative Proteomics

A. Empty Vector  Her2/neu

250 kD  150 kD  100 kD  75 kD

B. Vehicle

<table>
<thead>
<tr>
<th>Her2 inhibitor (µM)</th>
<th>0.01</th>
<th>0.1</th>
<th>1</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gefitinib 1 µM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

WB: Anti-pTyr

Bose et al., PNAS 103:9773, 2006
Mapping Her2/neu Tyrosine Kinase Signaling using Quantitative Proteomics

1. **Empty vector cells**
   - Normal arginine medium

2. **Her2/neu cells**
   - $^{13}\text{C}_6$-arginine medium (+ 6 Da)

3. **Her2/neu cells**
   - $^{13}\text{C}_6^{15}\text{N}_4$-arginine medium (+ 10 Da)

Mix Lysates

Imunoaffinity Purify with Antiphosphotyrosine Antibodies

Resolve on SDS-PAGE

Digest Bands with Trypsin

Identify and Quantify Proteins by LC-MS/MS

Bose et al., PNAS 103:9773, 2006
Mapping Her2/neu Tyrosine Kinase Signaling using Quantitative Proteomics

A. Fold Change with Her2/neu

B. Fold Change with Her2 kinase inhibitor

Bose et al., PNAS 103:9773, 2006
Mapping Her2/neu Tyrosine Kinase Signaling using Quantitative Proteomics

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Fold Change with Her2</th>
<th>Fold Inhibition with 100 nM PD168393</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Increased Phosphorylation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Known Her2/neu Signaling Proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Her2/neu</td>
<td>24.5</td>
<td>3.1</td>
</tr>
<tr>
<td>HSP90α</td>
<td>4.1</td>
<td>2.2</td>
</tr>
<tr>
<td>HSP90β</td>
<td>3.0</td>
<td>1.7</td>
</tr>
<tr>
<td>PLCγ1</td>
<td>2.6</td>
<td>2.0</td>
</tr>
<tr>
<td>PI 3-kinase p85β</td>
<td>2.2</td>
<td>0.8</td>
</tr>
<tr>
<td>PI 3-kinase p85α</td>
<td>1.9</td>
<td>1.1</td>
</tr>
<tr>
<td>PI 3-kinase p110β</td>
<td>1.7</td>
<td>1.0</td>
</tr>
<tr>
<td>Fyn</td>
<td>1.7</td>
<td>0.8</td>
</tr>
<tr>
<td>Known EGFR Signaling Proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOK1</td>
<td>4.2</td>
<td>2.0</td>
</tr>
<tr>
<td>RIN1</td>
<td>3.8</td>
<td>2.0</td>
</tr>
<tr>
<td>STAT1</td>
<td>3.5</td>
<td>1.6</td>
</tr>
<tr>
<td>Delta-Catenin</td>
<td>3.5</td>
<td>2.1</td>
</tr>
<tr>
<td>RAP1α</td>
<td>2.8</td>
<td>0.8</td>
</tr>
<tr>
<td>RhoGD1α</td>
<td>2.2</td>
<td>0.7</td>
</tr>
<tr>
<td>Cortactin</td>
<td>2.1</td>
<td>1.1</td>
</tr>
<tr>
<td>p120-rasGAP</td>
<td>1.7</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Bose et al., PNAS 103:9773, 2006
Mapping Her2/neu Tyrosine Kinase Signaling using Quantitative Proteomics

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Fold Change with Her2</th>
<th>Fold Inhibition with 100 nM PD168393</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Increased Phosphorylation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emerin</td>
<td>9.9</td>
<td>5.8</td>
</tr>
<tr>
<td>SKAP55 homologue</td>
<td>8.0</td>
<td>2.1</td>
</tr>
<tr>
<td>FYB (Fyn binding protein)</td>
<td>5.3</td>
<td>1.1</td>
</tr>
<tr>
<td>Semaphorin 7A</td>
<td>5.8</td>
<td>0.7</td>
</tr>
<tr>
<td>Transmembrane protein 33</td>
<td>4.0</td>
<td>1.6</td>
</tr>
<tr>
<td>17β hydroxysteroid dehydrogenase 12</td>
<td>3.8</td>
<td>1.9</td>
</tr>
<tr>
<td>RAB18</td>
<td>3.2</td>
<td>1.3</td>
</tr>
<tr>
<td>AXL receptor tyrosine kinase</td>
<td>2.8</td>
<td>1.4</td>
</tr>
<tr>
<td>RIKEN cDNA 2310079N02</td>
<td>2.6</td>
<td>2.2</td>
</tr>
<tr>
<td>Moesin</td>
<td>2.5</td>
<td>1.2</td>
</tr>
<tr>
<td>PDCD6/ALG-2</td>
<td>2.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Calcineurin B homologous protein 1</td>
<td>1.8</td>
<td>1.4</td>
</tr>
<tr>
<td>Poly(A) binding protein, cytoplasmic 1 (PABPC1)</td>
<td>1.6</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Bose et al., PNAS 103:9773, 2006
Mapping Her2/neu Tyrosine Kinase Signaling using Quantitative Proteomics

Bayesian Network Analysis of Proteomic Results
Results

Identified 6900 phosphorylation sites on 1850 proteins.

Changes with LPS:
24% of sites increased.
9% of sites decreased.

Measured the phosphorylation of 187 proteins annotated as transcriptional regulators.
They linked proteomics measurements with changes in gene expression.

Weintz et al., MSB 6:371, 2010
Studying Toll-Like Receptor Signaling in Macrophages using Quantitative Proteomics

Weintz et al., MSB 6:371, 2010
Limitations and Cautions: Sizes of Proteomic Experiments

A Medium sized Proteomic Experiment:
  Several hundred proteins – time required: Months

A Large Proteomic Experiment:
  A few thousand proteins – time required: 1-3 YEARS.

Proteomics cannot currently analyze as many genes as DNA microarray technology can!

Proteomics is also highly technically demanding and often requires a lot of optimization and small scale testing before performing a large experiment.
Mass Spectrometry at Washington University

• Wash U receives NIH funding for the Biological and Biomedical Mass Spectrometry Research Resource.

• At least 8 labs at Wash U. perform biological mass spectrometry experiments.

• Available instruments on the Wash U medical campus, Wash U Danforth campus, and the Danforth Plant Science Center include:
  – At least 30 mass spectrometers.
  – 5 LTQ-OrbiTrap mass spectrometers (some of the latest and highest performance instruments).
Summary (Part 1)

1. There is widespread use of mass spectrometry in both the biological and physical sciences.

2. Proteins are usually digested into peptides. Peptide sequence is determined by fragmentation in the Mass Spectrometer.

3. Protein abundance and enrichment or fractionation methods are critical to consider in the planning of proteomic experiments.

4. Proteomics can identify proteins and map their post-translational modifications. Components of protein complexes and intracellular pathways can be analyzed by proteomics.
5. Quantitative proteomics can be performed by incorporating stable isotopes into proteins or by using label-free quantitation methods.

6. Proteomics cannot analyze as many genes as DNA microarray technology. Further, proteomics is highly technically demanding and often requires a lot of optimization.

7. Many labs at Wash U. use mass spec and proteomics. Wash U. has a lot of the necessary equipment and expertise to conduct mass spectrometry experiments.