Basic Proteomic Workflow

IDeA National Resource for Proteomics
Workshop for Graduate Students and Post-docs
Renny Lan
05/16/2017
Outline

- Concepts of proteomics
- Proteomic strategies
- Overview of the Proteomics Approaches
- General workflow of shotgun proteomics
- Technologies of proteomics
- Keys to success
- The proteomic workflow in our core
What is Proteomics?

- The study of the proteome
- **Proteome** - the entire set of proteins expressed by a genome, cell, tissue or organism at a certain time, under defined conditions

Understanding the proteome allows for:
- ✔ Complementing other functional genomics approaches
- ✔ Characterization of proteins and relative abundance
- ✔ Understanding the post-translational modification (PTM)
- ✔ Understanding protein-protein interactions
- ✔ Identification of disease biomarkers
- ✔ Generation of new hypothesis
20,000 genes

~100,000 transcripts

>1,000,000 proteins
The level of any protein is governed by many complex biochemical events.

Most proteins are at levels of $10 - 100$ copies per cell.

Some proteins are at levels of $10,000 - 1,000,000$ copies per cell.

Favored research drug targets (signal proteins) are actually low in abundance.
Post-Translational Modifications (PTM)

- Phosphorylation cascades are involved in many signalling pathways.
- Various modifications regulate microtubule function.
- Plasma-membrane proteins can be linked to the membrane by a GPI anchor.
- Plasma-membrane proteins can carry N-glycans.
- The histone code controls many nuclear processes.
- Nuclear and cytoplasmic proteins can carry O-glycans.
- Polyubiquitylation can induce protein degradation.
- 26S proteasome.
- Target protein.
# Challenges of Proteins vs. DNA

<table>
<thead>
<tr>
<th>DNA</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Static</td>
<td>• Very dynamic: Serum proteins</td>
</tr>
<tr>
<td>• Can be amplified</td>
<td>• Cannot be amplified</td>
</tr>
<tr>
<td>• Little complexity: Single component</td>
<td>• Very complex: Post-translational modification</td>
</tr>
<tr>
<td>• Good solubility characteristics</td>
<td>• Variable solubility: Membrane proteins</td>
</tr>
</tbody>
</table>

Data analysis!!
Proteomic strategies
“How much” do you want to know

- Peak and valleys: Area under the curve
- Exploiting nature: Stable isotopes (SILAC)
- Isobaric labeling: iTRAQ and TMT
- Label free: Spectral counting
- Putting it together: Peptides to proteins
Peak and valleys: Area under the curve

- Classical LC: retention time & light absorbance

- Amount of material is directly proportional to area under the curve (trace)

<table>
<thead>
<tr>
<th>Absorbance unit (nm)</th>
<th>Retention time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-260 nm = DNA</td>
<td>-280 nm = Protein</td>
</tr>
<tr>
<td>-214 nm = Peptide</td>
<td></td>
</tr>
</tbody>
</table>
Peak and valleys: Area under the curve

- Classical LC: retention time & light absorbance

  - Peptide digestion
  - Pump → Detector

- Amount of material is directly proportional to area under the curve (trace)
Peak and valleys: Area under the curve

- LC/MS: retention time, m/z and ion intensity

- Peptide digestion
- Pump
- Mass spectrometer

- LC/MS: retention time, m/z and ion intensity

- Mass-to-charge ratio (m/z) - m is the mass in Daltons and z is the charge
UHPLC peak chosen at 26.47 min

Mass at 571.36 chosen for MS/MS
Overview of the Proteomics Approaches

Top-down proteomics
- Protein mixture (Mass ≤ 50 kDa)
- Separation of proteins
- MS analysis of intact protein (≤ 50 kDa)
- Intensity vs. m/z

Middle-down proteomics
- Protein (mixture) (No mass limit)
- Digestion
- Separation of peptides
- MS analysis of peptides (~2000 – 2000 Da)
- Intensity vs. Time vs. m/z

Bottom-up proteomics
- Protein (mixture) (No mass limit)
- Digestion
- Separation of peptides
- MS analysis of peptides (~500 – 3000 Da)
- Intensity vs. Time vs. m/z
Top-down approach
– Full coverage of identified modifications which is used to ID and quantify the whole proteins
– Does not handle large proteins well
– Does not handle complex samples well

Bottom-up approach
– Most common approach
– Easily adaptable to high-throughput analysis
– Handles complex samples well
– Many peptides are not detected
Shot-gun Proteomic

- Trypsin digestion:
  - In solution or in gel
  - Protein denaturation
  - Alkylation and reduction

- Enrichment:
  - UHPLC Fractionation
  - Affinity resins or IPs

- UHPLC separation
- MS1 for quantification
- MS2 for identification
- Data analysis
- Pathway analysis
Identification vs. Quantitation

- What’s there? How much of it is there?
- How sure are you about the ID?
- How sure are you about the abundance?
- Not there versus not detectable
Technology of Proteomics

- Sample clean-up or enrichment
- Protein digestion
- Separation of proteins or peptides
- Sample clean-up
- Quantitative labeling
- Analysis of proteins
- Database utilization
- Data analysis
### Types of Enrichment Technologies in Proteomics

<table>
<thead>
<tr>
<th>Method</th>
<th>Purification based on</th>
<th>Purification done using</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affinity purification</td>
<td>Antibody binding</td>
<td>IP, affinity resin or Ab-magnetic beads</td>
<td>Phosphopeptide enrichment (PolyMAC-Ti)</td>
</tr>
<tr>
<td>Ultrafiltration</td>
<td>Size exclusion</td>
<td>Centrifugal filters</td>
<td>Native peptide enrichment</td>
</tr>
<tr>
<td>Electrophoresis</td>
<td>charge, charge/mass ratio, size, or shape</td>
<td>1DE</td>
<td></td>
</tr>
<tr>
<td>Liquid Chromatography</td>
<td>charge, hydrophobicity, size, or specificity</td>
<td>Ion-exchange, reverse phase, size-exclusion, affinity chromatography</td>
<td></td>
</tr>
<tr>
<td>Salt-out</td>
<td>Protein interaction/ electrolyte-non electrolyte interaction</td>
<td>Different salt concentrations</td>
<td>Ammonium sulfate precipitation</td>
</tr>
<tr>
<td>Depletion kit</td>
<td>Removing high abundant proteins</td>
<td>Immobilized antibody</td>
<td>Top 12 Abundant Protein Depletion Spin Columns</td>
</tr>
</tbody>
</table>
Protease Digestion

• Reducing agent: DTT or TCEP
• Alkylating reagent: iodoacetamide
• Trypsin: most commonly used, cleaves on C-terminal side of Arg and Lys unless next residue is Pro
• Other 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
<table>
<thead>
<tr>
<th>Method</th>
<th>Separation based on</th>
<th>Separation done using</th>
<th>Further steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gel Electrophoresis (1D)</td>
<td>Molecular mass</td>
<td>Gel (which acts like a molecular sieve)</td>
<td>In-gel digestion of proteins → LC-MS/MS or MALDI-TOF-MS</td>
</tr>
<tr>
<td>Gel Electrophoresis (2D)</td>
<td>Isoelectric point (pI; IEF) &amp; Molecular mass</td>
<td>Gel, potential and ampholytes</td>
<td></td>
</tr>
<tr>
<td>Reverse Phase (C8 or C4) chromatography</td>
<td>hydrophobicity and molecular weight</td>
<td>HPLC</td>
<td>Protein(s) → Digest to peptides → LC-MS/MS or MALDI-TOF-MS</td>
</tr>
<tr>
<td>Gel Filtration</td>
<td>Molecular weight</td>
<td>HPLC</td>
<td></td>
</tr>
<tr>
<td>Ion Exchange</td>
<td>Cation or Anion affinity</td>
<td>FPLC</td>
<td></td>
</tr>
<tr>
<td>Affinity Chromatography</td>
<td>DNA, RNA, Anti-body, peptides etc</td>
<td>HPLC</td>
<td></td>
</tr>
<tr>
<td>Mud-PIT (Multidimensional Protein Identification Technology)</td>
<td>Cation Exchange &amp; hydrophobicity (used for peptides; not for proteins)</td>
<td>HPLC</td>
<td>Online MS/MS analysis</td>
</tr>
</tbody>
</table>
# Types of Clean-Up Technologies in Proteomics

<table>
<thead>
<tr>
<th>Method</th>
<th>Purification based on</th>
<th>Purification done using</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein Precipitation</td>
<td>Protein solubilization</td>
<td>Organic solvents</td>
<td>TCA / Acetone</td>
</tr>
<tr>
<td>Liquid-Liquid extraction</td>
<td>Protein solubilization</td>
<td>Chloroform/methanol</td>
<td>Lipid removal</td>
</tr>
<tr>
<td>Solid phase extraction</td>
<td>hydrophobicity</td>
<td>C18 cartridge</td>
<td>Sep-Pak SPE</td>
</tr>
<tr>
<td>Gel electrophoresis</td>
<td>Isoelectric point &amp; Molecular mass</td>
<td>Gel, potential and ampholytes</td>
<td>1D and 2D SDS PAGE</td>
</tr>
<tr>
<td>Tandem Affinity purification tag (TAP tag)</td>
<td>Affinity to the bead or resin tag</td>
<td>Fusion protein with TAP tag</td>
<td>Study of protein-protein interactions</td>
</tr>
<tr>
<td>Ultrafiltration</td>
<td>Size exclusion</td>
<td>Centrifugal filters</td>
<td>Buffer exchange</td>
</tr>
<tr>
<td>Low Pressure LC</td>
<td>Affinity to the columns</td>
<td>Liquid Chromatography</td>
<td>RP, ion-exchange, affinity columns</td>
</tr>
</tbody>
</table>
Separation methods: 2D differential gel electrophoresis

Dimension 1:
Isoelectric focusing
PI based separation

Dimension 2:
Gel electrophoresis
MW based separation
Difference Gel Electrophoresis (DIGE)

- Use Cy Dyes to label samples so they can be run together on the same gel
- Robotic excision of gel spots showing differential fluorescence for mass spectrometric analysis
- Protein purification: not necessary
- Protein ID: up to thousands per sample
- Often good for visualizing PTMs

![DIGE Image](image_url)
DIGE-Overview

Fig 1-1. Outline of Ettan DIGE system (when used with three CyDye DIGE Fluor minimal dyes separated in a single gel).

Courtesy of GE Healthcare
Virtual elimination of gel-to-gel variation reveals induced biological change with statistical accuracy capable of revealing differences in abundance of less than 10% between samples.

Conclusion:

Expression decrease in sample 3.
DIGE-Pros and Cons

- High sensitivity
- Linearity of the dyes utilized—quantitate over 5 orders of magnitude
- Straightforward significant reduction of experimental error
- High reproducibility

- Requires high resolution 2D gels
- Not ideal for membrane proteins
- Not ideal for serum type samples
- Some protein spots identify more than one protein or do not have enough protein to identify the spot
- Labor Intensive
Key to Success

- **Good separations: reduce the complexity**
  - ✓ Multiple dimensions: 1D or 2D gel, immunochemistry, reverse phase chromatography, ionization, basic fractionation..

- **High resolution: distinguish the peak from the baseline**
  - ✓ Background clean-up: salts, detergent, polymers...
  - ✓ Maintain the quantity: Start with high concentration and be aware of the lost of peptides in each step
  - ✓ State of art instrumentation

- **High sensitivity: detect low abundant peptides**

- **High reproducibility:**
  - ✓ Quantitative labeling
  - ✓ Good pipetting and bench work skills
  - ✓ Monitoring the concentration in different steps
Proteomics Workflow in UAMS Proteomics Core

Sample preparation
- Serum depletion

Enrichment
- 12 to 24 gel fractions
- Off-line high pH Fractionation (13)
- Peptide enrichment

UHPLC-MS/MS
- UHPLC separation
- MS1: full spectra
- MS/MS: fragments

Data analysis
- Comparative proteomics
- Scaffold Persus R....
- TMT labeling
- iBAQ
- SILAC Label Free
- MaxQuant
- Raw data
- Quantification
- Protein quant
- MaxQuant
- Raw data
- Identification
- Mascot
- Protein IDs
- DB
Peptide Quant to Protein Quant: End of the day

<table>
<thead>
<tr>
<th>Protein (or peptide)</th>
<th>Genes/Accessions</th>
<th>Log2 SILAC Ratio</th>
<th>Q-value (FDR) of change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein A</td>
<td>ProA/P10299A</td>
<td>0.12</td>
<td>0.94</td>
</tr>
<tr>
<td>Protein B</td>
<td>ProB/P76281M</td>
<td>0.36</td>
<td>0.66</td>
</tr>
<tr>
<td>Protein C</td>
<td>ProC/P36233W</td>
<td>1.2</td>
<td>0.02</td>
</tr>
<tr>
<td>Protein D</td>
<td>ProD/P09736Q</td>
<td>-2.4</td>
<td>0.52</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

➢ Table of values organized around proteins (or peptides)
➢ A ratio that indicates a fold-change vs a control condition
➢ A false discovery rate statistic for each protein (peptide) that the ratio is different from the null hypothesis (unchanged)
➢ A prioritized list of candidates for follow-up studies
Question?