Principles and Applications of Proteomics
Overview

• Why Proteomics?

• 2-DE
  – Sample preparation
  – 1\textsuperscript{st} & 2\textsuperscript{nd} dimension separation
  – Data Analysis
  – Sample preparation for Mass Spectrometry

• Mass Spectrometry
  – MALDI-TOF, TANDEM MS
  – Identification of MS spectra

• Applications
  – ICAT, Phosphoproteomics, etc.
Roles of Proteins

- Proteins are the instruments through which the genetic potential of an organism are expressed = active biological agents in cells

- Proteins are involved in almost all cellular processes and fulfill many functions

- Some functions of Proteins
  - enzyme catalysis, transport, mechanical support, organelle constituents, storage reserves, metabolic control, protection mechanisms, toxins, and osmotic pressure
The Virtue of the Proteome

- Proteome = protein compliment of the genome
- Proteomics = study of the proteome
- Protein world = study of less abundant proteins
- Transcriptomics – often insufficient to study functional aspects of genomics
Why Proteomics?

• Whole Genome Sequence – complete, but does not show how proteins function or biological processes occur
• Post-translational modification – proteins sometimes chemically modified or regulated after synthesis
• Proteins fold into specific 3-D structures which determine function
• Gain insight into alternative splicing
• Aids in genome annotation
## Some Covalent Post-Translational Modifications

<table>
<thead>
<tr>
<th>Modification</th>
<th>Residues</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleavage</td>
<td>Various</td>
<td>Activation of proenzymes and precursors</td>
</tr>
<tr>
<td>Glycosylation</td>
<td>Asn, Ser, Thr</td>
<td>Molecular targeting, cell-cell recognition etc</td>
</tr>
<tr>
<td>Phosphorylation</td>
<td>Ser, Thr, Tyr</td>
<td>Control metabolic processes &amp; signalling</td>
</tr>
<tr>
<td>Hydroxylation</td>
<td>Pro, Lys</td>
<td>Increase H-bonding &amp; glycosylation sites</td>
</tr>
<tr>
<td>Acetylation</td>
<td>Lys</td>
<td>Alter charge &amp; weaken interactions with DNA</td>
</tr>
<tr>
<td>Methylation</td>
<td>Lys</td>
<td>Alter interactions with other molecules</td>
</tr>
<tr>
<td>Carboxylation</td>
<td>Glu</td>
<td>More negative charge, e.g. to bind Ca</td>
</tr>
<tr>
<td>Transamidation</td>
<td>Gln, Lys</td>
<td>Formation of crosslinks in fibrin</td>
</tr>
</tbody>
</table>
Different Approaches for Proteome Purification and Protein Separation for Identification by MS

- A. Separation of individual proteins by 2-DE
- B. Separation of protein complexes by non-denaturing 2-DE
- C. Purification of protein complexes by affinity chromatography + SDS-PAGE
- D. Multidimensional chromatography.
- E. Fractionate by Organic Solvent – separate complex protein mix, hydrophobic membrane proteins

(van Wijk, 2001, Plant Physiology 126, 501-508)
2-Dimensional Protein Electrophoresis (2-DE)

Purify Proteins from desired organelle, cell, or tissue

Separate Protein mixture in 1-D by pI

Separate Protein Mixture in 2-D by MW

Stain Gel, Data Analysis

Protein Identification by MS
Plant Protein Extraction and Fractionation

1. Leaf Tissue + Homogenization Buffer
2. Phenol Extraction
3. Back-Extraction
4. NH₄OAc in MeOH and Acetone ppt

Low Stringency
- 1-D IEF
- 2-D SDS PAGE

High Stringency
- 1-D IEF
- 2-D SDS PAGE
First Dimension IEF: Immobilized pH Gradients

**IPG principle:**

pH gradient is generated by a number (6-8) of well-defined chemicals (immobilines) which are co-polymerized with the acrylamide matrix.

- IPG allows the generation of pH gradients of any desired range between pH 3 and 12.
- Sample loading capacity is much higher.

The method of choice for micropreparative separation or spot identification.
Components of IEF Buffer

- **Chatotropes**
  - 8M Urea
  - OR…7M Urea/2M Thiourea

- **Surfactants**
  - 4% CHAPS
  - OR….2% CHAPS / 2% SB-14

- **Reducing Agents**
  - 65mM Dithioerythritol
  - OR….100mM Dithiothretiol
  - OR….2mM tributyl phosphine

- **Ampholytes: 2%**
First Dimension IEF: Procedure

- Individual Strips: 24, 18, 11-13, 7cm long; 0.5mm thick

Procedure:

1. Rehydrate dry IPG strips (12h)
2. Apply Sample (during or after rehydration)
3. Run IPG Strips (high V, low current, 20°C 4h)
Second Dimension Separation: SDS-PAGE

1. Pour linear or gradient standard SDS-PAGE gel (std = 12%)
2. Equilibrate 1-D Gel for SDS-PAGE
3. Load 1-D Gel onto SDS-PAGE gel
4. Apply Protein Ladder with Application Strips
5. Seal 1-D Gel with 0.5% LMP Agarose
6. Run Gel constant mA
7. Stain Gel : Coomassie Blue, Colloidal Coomassie Blue, Silver Stain
8. Visualize Gel & Record Image by Scanning or CCD Camera
2-DE With Immobilized pH Gradients

Gorg, A. 2000, Proteome Research, ch4. Springer
Image Analysis

Commonly Used Software:
- ImageMaster™
- Melanie III™
- PDQuest™
- ALL EXPENSIVE - $5-10k

Software Functions:
- Quantification
- Detection
- Alignment
- Comparison
- Matching
- Synthetic Gaussian Image from Image of Sample used in all phases
Differential Protein Expression
From Protein To Gene

In Gel Trypsin Digest

Reduce and Extract Peptides

MALDI-TOF m/z

Database search of m/z
Genomics Solutions
Prospector (MS-FIT)

SDS
pl 4

LC-ESI MS/MS m/z

Database search ion fragments from individual peptides
Matrix Science (MASCOT)
Prowl (ProFound)
Spot Picking

Pick Protein Spot From Gel

- Manual or Automatic

Prepare Sample for MS

- Wash Sample
- Dehydrate Sample
- Dry Sample
- In-gel digestion with trypsin (30ng trypsin, 37C, 16h)
- Extract tryptic peptides from gel
- Desalt and concentrate sample
Basic Components of a Mass Spectrometer

- Inlet
- Ion Source
- Mass Analyzer
- Detector
- Instrument control system
- Vacuum system
- Data System

Variations of instrument components typically used in protein sequencing and identification experiments:

**Instrument Component**

**Sample Inlet**
1. Direct probe or stage
2. Capillary column liquid chromatography

**Ion Source**
1. Electrospray
2. Matrix-assisted laser desorption

**Mass Analyzer**
1. Quadrupole mass filter
2. Ion trap mass analyzer
3. Time-of-flight mass analyzer

Types of Mass Spectrometers

- MALDI-TOF

- ESI TANDEM MASS SPEC INSTRUMENTS
  1. Quadropole Mass Analyzers
  2. Ion Trap Mass Analyzers
  3. TOF Mass Analyzers
MALDI-TOF: How the MALDI Source Works

- Tryptic peptides co-crystallized with matrix compound on sample stage
- Irradiation with UV-laser
- Matrix compound vaporized and included peptide ions moved to gas phase
- Protonated peptide ions enter MS

A. MALDI ionization process

B. MALDI-TOF in linear mode

C. MALDI-TOF with reflectron

ELECTROSPRAY IONIZATION (ESI)

TANDEM MS- TRIPLE QUADROPOLE MS

A. Quadropole Mass Analyzer

B. Trajectories of ion with selected $m/z$ verses ion without selected $m/z$

C. Full-Scan Mode

D. Tandem MS-MS Mode

TANDEM MS: TRIPLE QUADRUPOLe MS

Electrospray ion source → Octapole lens → First quadrupole mass filter → Second quadrupole mass filter → Octapole collision cell → Detector

Product ion scan

<table>
<thead>
<tr>
<th>MS I</th>
<th>Collision cell</th>
<th>MS II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass select a single m/z</td>
<td>CID</td>
<td>Scan to acquire spectrum</td>
</tr>
</tbody>
</table>

CID
TANDEM MS: ION TRAP MS

A. Ion Trap – Ions collected in trap maintained in orbits by combination of DC and radiofrequency voltages

B. Radiofrequency voltages on selected ions scanned to eject ions based on m/z and select particular ion m/z

C. Collision-Induced Dissociation

D. Scan out of product ions according to m/z

Ion Trap - MS^n

TANDEM MS: QUADRUPOLE TIME OF FLIGHT MS (Q-TOF)

## Comparison of MALDI-TOF and MS/MS

### MALDI-TOF
- Sample on a slide
- Spectra generate masses of peptide ions
- Protein Id by peptide mass fingerprinting
- Expensive
- Good for sequenced genomes

### TANDEM MS
- Sample in solution
- MS-MS spectra reveal fragmentation patterns – amino acid sequence data possible
- Protein Id by cross-correlation algorithms
- Very Expensive
- Good for unsequenced genomes
Protein Identification Using Peptide Mass Fingerprinting (MALDI-TOF Data)

2-DE Gel → Intact Protein → Experimental Proteolytic Peptides → Experimental MS

DNA Sequence Database → Protein Sequence Database → Theoretical Proteolytic Peptides → Theoretical MS

Computer Search
Databases Available for Id of MS Spectra

- **SWISS-PROT** – nr database of annotated protein sequences. Contains additional information on protein function, protein domains, known post-translational modifications, etc. ([http://us.expasy.org/sprot](http://us.expasy.org/sprot))

- **TrEMBL** - computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

- **PIR-International** – nr annotated database of protein sequences. ([http://www-nbrf.georgetown.edu/](http://www-nbrf.georgetown.edu/))

- **NCBIInr** – translated GenBank DNA sequences, Swiss-Prot, PIR.

- **ESTdb** – expressed sequence tag database (NIH/NSF)

- **UniProt** – proposed new database. Will joint Swiss-Prot, TrEMBL, PIR. [http://pir.georgetown.edu/uniprot/](http://pir.georgetown.edu/uniprot/)
Programs Used to Identify Mass Spectra

- 3 main types programs available

1. Use proteolytic peptide fingerprint for protein Id (ie MALDI-TOF data).
   - PeptIdent, MultiIdent, ProFound

2. Programs that operate with MALDI-TOF or MS-MS spectra or combination of both
   - PepSea, MASCOT, MS-Fit, MOWSE

3. Programs that operate with MS-MS spectra only
   - SEQUEST, PepFrag, MS-Tag, Sherpa
Mass Spec Algorithms for Protein Id (MS-MS only)

- More perfect algorithms use additional information such as pI, MW, amino acid composition, etc (example: MOWSE algorithm).

**Mascot Search Results**

Probability Based Mowse Score
Score is -10*log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 52 indicate identity or extensive homology (p<0.05).

Peptide Summary Report

1. gi|6066418 Mass: 27447 Total score: 99 Peptides matched: 2
   ascorbate peroxidase [Lycopersicon esculentum]

<table>
<thead>
<tr>
<th>Query</th>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>Delta</th>
<th>Miss Score</th>
<th>Rank</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>683.05</td>
<td>2046.13</td>
<td>2045.90</td>
<td>0.23</td>
<td>0</td>
<td>38</td>
<td>YADEDAPFFADYAEAHLK</td>
</tr>
<tr>
<td>11</td>
<td>690.44</td>
<td>2068.30</td>
<td>2068.04</td>
<td>0.26</td>
<td>0</td>
<td>61</td>
<td>QMGLSDQDIVALSGAHTLGR</td>
</tr>
</tbody>
</table>

2. gi|99375 Mass: 29040 Total score: 61 Peptides matched: 1
   L-ascorbate peroxidase (EC 1.11.1.11) precursor - Arabidopsis thaliana (fragment)

<table>
<thead>
<tr>
<th>Query</th>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>Delta</th>
<th>Miss Score</th>
<th>Rank</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>690.44</td>
<td>2068.30</td>
<td>2068.07</td>
<td>0.22</td>
<td>1</td>
<td>61</td>
<td>QMGLSDKDIVALSGAHTLGR</td>
</tr>
</tbody>
</table>
Proteomics Applications

• Differential Display Proteomics
  – DIGE – Difference gel electrophoresis
  – MP – multiplexed proteomics
  – ICAT – isotope coded affinity tagging
Protein Expression Profile Analysis

- Control sample
- Diseased sample
- Composite image construction
- Proteome analysis
- Excision of protein feature
- Digest to fragments
- Mass spectrum
- Database query

Drug Discovery Today
Difference Gel Electrophoresis (2D-DIGE)

Sample 1
- Derivatize with Propyl Cy3
- Cy3-labeled Sample

Sample 2
- Derivatize with Methyl Cy5
- Cy5-labeled Sample

Mix and run samples on a single 2-D gel

(isoelectric point)

(spot too dim to view by eye)

Acquire images from the gel

(Cy3 image)

(Cy5 image)

Overlay images

Total Protein Differential Display Map

(Unlu, 1997, electrophoresis 18, 2071)
Multiplexed Proteomics (MP)

(Steinberg, 2001, Proteomics 1,841, 2071)
Isotope-Coded Affinity Tagging (ICAT)

(Smolka, 2002, Mol Cell Proteomics 1, 19-29)
Conclusions

• 2-DE is a powerful technique to separate complex protein mixtures and analyze proteomes.

• Mass Spectrometry microsequencing can identify proteins from 2-DE gels and other samples.

• There are multiple databases and computer programs available to analyze MS data for protein Identification.

• Proteomics approach can be used to identify all proteins in a particular sample, elucidate additional components of biochemical pathway(s), or analyze post-translational modifications at a small or large scale.