Mass Spectrometry and Proteomics

Professor Xudong Yao
Bioanalytical Chemistry
Spring 2007
• Proteomics and “-omics”

• Roles of mass spectrometry

• Comparative proteomics

• Chemical proteomics
Protein, Proteome and Systems Biology

DNA → RNA → Protein

Replication

Genome → Transcriptome → Proteome

Proteomics

Analytical Definition of Proteomics
Identity, Quantity and Function of All the Proteins in a Mixture

Mass Spectrometry
Objectives of Proteomics

Function
- Interaction
- Activity

Identity

Quantity

Time

Time-Dependence

Proteome
Analytical Challenges in Post-genome Research

- Sample complexity
  - “Peak capacity”
  - Multi-dimensional separation

- Collective analysis
  - Not traditional, one-by-one analysis
  - Sensitive, specific and quantitative
  - and the answers is …

- Data treatment, analysis and achieving
  - Hardware
  - Software

- Researchers of multi-disciplinary training
A Typical Proteomics Experiment

Sample Preparation
- Protein/Peptide Chemistry
- Enzymology
- Separations:
  - Advanced HPLC and Electrophoresis

Mass Spectrometry
- New Ionization techniques
- Contemporary & New Mass Spectrometric Methods

Bioinformatics
- Applications of Mass Spectrometry and Database Searching Tools

Biology
- Protein Posttranslational Modification
  - Protein Phosphorylation
  - Proteolysis
Roles of Mass Spectrometry
The mass analysis process as compared to the dispersion of light by a prism
Mass Spectrometry

Ion Source (ESI, MALDI) → Mass Analyzer → Ion Detector

Sample Introduction → Mass Spectrum

Vacuum

Data System

Intensity vs. m/z
Mass Analyzer

• Separate ions by mass/charge

• Common types
  – Quadrupole mass filter (Q), Time-of-Flight (TOF), Ion Trap (IT), Fourier Transform Ion Cyclotron Resonance (FTICR)

• Tandem mass spectrometry
  – Spatial, such as Q-q-TOF, TOF-TOF, Q-q-Q
  – Temporal, such as IT, FTICR
  – Spatial and Temporal, such as IT-FTICR, Q-q-FTICR, IT-TOF
Soft Ionizations of Biomolecules

1. Matrix-Assisted Laser Desorption/Ionization (MALDI)
2. Electrospray Ionization (ESI)

Nobel Prize in 2002
Matrix-Assisted Laser Desorption/Ionization (MALDI) Generates Singly-Charged Ions

- Sample probe
- Laser beam
- Light-absorbing matrix
- Analyte molecule
- Matrix ion
- Analyte ion $\text{MH}^+$

Diagram showing the process of MALDI where a laser beam generates ions from a light-absorbing matrix.
Electrospray Ionization (ESI) Generates Multiply-Charged Ions
Tandem MS: Basic Concepts

MS-1

Mass-selected precursor

Fragmentation Chamber

Ion Selection

Mass Analysis

MS-2
Mass Spectrometry for Proteins and Peptides

- Structural elucidation
- Quantitative analysis
- Sensitive and automatic mixture analysis
- Sequencing proteins
- Identification of post-translational modifications
- High-order structures and dynamics of proteins and protein dynamics
Comparative Proteomics

Gel or Non-gel
Label or Non-label
Comparative Proteomics

Relative Changes in proteins including concentration and composition

Proteome 1

Stable Isotope Coding

Separation

Relative Quantitation

Constant Ratio

Abnormal Ratios

Identification of Proteins

Proteome 2

Sampling/Cell Culture Fractionation/Separation

Mass Spectrometry Bioinformatics
Reduce sample complexity at the protein level: DDF (Differential Detergent Fractionation)

1. Tween/DOC Extraction (Nuclear Fraction) (Prolactin)
2. Tween/DOC Extraction (Nuclear Fraction) (Control)
3. Triton X-100 Extraction (Membrane/Organelle Fraction) (Prolactin)
4. Triton X-100 Extraction (Membrane/Organelle Fraction) (Control)
5. Digitonin Extraction (Cytosolic Fraction) (Prolactin)
6. Digitonin Extraction (Cytosolic Fraction) (Control)
7. Protein Ladder
GEL: 2-Dimensional Electrophoresis (2-DE)

- Difficulties with: extreme pI proteins, low abundance, proteins, hydrophobic proteins
- Inefficient in-gel digestion of proteins for MS analysis
- Labor extensive

NON-GEL: MudPIT Analysis of Protein Complex

- Easier separation
- Easier automation
- “Problematic” proteins
  - Small, large, hydrophobic, low abundance
- Easier sample preparation for MS analysis
- Computational capacity
- **Quantitative capability**

A Glance at Proteomic Bioinformatics
Shotgun² Collision Induced Dissociation Mass Spectrometry

1. In-source Shotgun(CID)
2. In-collision-cell Shotgun(CID)
3. Shotgun²(CID) = 1 + 2
   - CID of [M+3H]³⁺
   - CID of [M+2H]²⁺

In-Source
Comparative Proteomics

Relative Changes in proteins including concentration and composition

Proteome 1
- Stable Isotope Coding
- Separation
- Relative Quantitation
- Constant Ratio
- Abnormal Ratios
- Identification of Proteins

Proteome 2
- Sampling/Cell Culture Fractionation/Separation
- Mass Spectrometry Bioinformatics
Label: MS-Based Relative Quantitation

• Large differences in concentration
  – Direct ESI/MALDI MS

• Small differences in protein concentration
  – Stable isotope dilution: inherent choice
Introduction of Stable Isotopes

• Criteria for isotope internal standards
  – Ideally behaving the same before mass analysis
• Metabolic labeling during biosynthesis/bioprocess
• Post-biosynthesis/bioprocess labeling: chemical and enzymatic
  – Functional groups on side chains: -SH, -OPO$_3$H$_3$
  – Termini: N-terminal, C-terminal
  – Active/Binding sites: Chemical Proteomics

\[ H_2N \quad (A_1A_2A_3\ldots A_n)_i \quad COOH \]

Coding Module
Advantages of Modular Design

- Isotope Coding
  - Universal
    - Important to small proteins
  - Specific
  - Efficient
  - Minimal Structural Modification
    - Chromatographic co-elution
    - Stable during separation

- Separation
  - Portable to all separation platforms, including affinity separation
Tryptic Incorporation of Two $^{18}\text{O}$ Atoms into Peptide C-Terminus
Adenovirus as Model System

- Minimal real system
  - DNA + Proteins
- Known genome/proteome
  - Predictable proteolytic peptides
- Defined architecture
  - Predictable protein expression
    - Comparative quantitation of Ad2/Ad5 proteins
    - Dynamic range of 600-fold
  - Capsid protein modeling membrane/hydrophobic proteins
- Mutations modeling post-translational modifications

Shenk T. In *Fundamental Virology*, 1996
 Combined Isotope-Coded Peptides

Ad2 Virion

$\text{H}_2^{18}\text{O}$

Trypsin

Lys-C

Ad5 Virion

$\text{H}_2^{16}\text{O}$

Combined Isotope-Coded Peptides

MALDI-FTICR MS

Most of peptides
Ad2:Ad5 = 3:1

Theoretical MW of Ad Tryptic Peptides

Protein Identification and Quantitation
MALDI-FTICR Mass Spectrum of Combined Digests

- 1597.78, Protein V
- TSTEVQTDPWFR
- 218O
- 1603.83, Protein VII
- TTVDDAIDAVVEEAR
- 218O
- 1628.96, Protein V
- VLRPGTTVVFTPGER
- 218O
Controversies and Challenges in Proteolytic Labeling

- Reported controversies in tryptic $^{18}$O labeling
  - One $^{18}$O incorporation for K-terminated peptides
  - Low efficient incorporation of two $^{18}$O for short peptides
  - Two $^{18}$O in each new peptides

- Capabilities of endoprotease for $^{18}$O labeling
  - Two $^{18}$O incorporation by trypsin, Lys-C, and Glu-C only
  - One $^{18}$O incorporation by chymotrypsin…

- Challenges for automated, large-scale application
  - Amount and cost of $H_2^{18}$O
Decoupling Proteolytic $^{18}$O Labeling from Protein Digestion

Proteins in solution prior to digestion
Peptide labeling in small volume of $\text{H}_2^{18}\text{O}$
Separate optimization of digestion and labeling
Automatic, high-throughput, large-scale applications
Dissection of Proteolytic Incorporation of Two $^{18}$O

Amide Bond Cleavage

Carboxyl Oxygen Exchange

Molecular Basis for Cleavage and Exchange

Protease catalyzes exchange

TWO $^{18}$O INCORPORATION
$^{16}\text{O}$-to-$^{18}\text{O}$ Exchange Studied by MALDI-FTICR MS

- YGGFMR($^{16}\text{O}_2$)
- YGGFMR($^{18}\text{O}_2$)
- YGGFMR($^{16}\text{O}^{18}\text{O}$)

![Graph showing the exchange of $^{16}\text{O}$ and $^{18}\text{O}$ with MALDI-FTICR MS data.]
Determination of Reaction Initial Rates

\[ [R] = [R_o] \frac{I_o}{I_{total}(I_o, I_2, I_4, M_o, M_2, M_4)} \]
Kinetics Comparison in R- and K-Peptides

<table>
<thead>
<tr>
<th></th>
<th>YGGFMR</th>
<th>YGGFMK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{cat}$ (min$^{-1}$)</td>
<td>3500±500</td>
<td>2800±300</td>
</tr>
<tr>
<td>$K_M$ (µM)</td>
<td>1300±300</td>
<td>4400±700</td>
</tr>
<tr>
<td>$k_{cat}/K_M$ (µM$^{-1}$min$^{-1}$)</td>
<td>2.6±0.9</td>
<td>0.64±0.14</td>
</tr>
</tbody>
</table>
Simultaneous Mass Spectrometric Determination of Kinetics for Trypsin-Catalyzed $^{16}$O-to-$^{18}$O Exchange

Complete Exchange for Mixture
Enzymatic $^{18}$O Labeling

- Universal two $^{18}$O labeling of proteolytic peptides by protease-catalyzed exchange
  - Both K- and R-terminated peptides
  - Chymotrypsin and pepsin for two $^{18}$O labeling, in addition to trypsin, Lys-C, Glu-C, ...
  - Both short and long peptides

- 4 Da mass increase at the C-terminus of proteolytic peptides to be differentiated in mass spectrometry
Mass Spectrometry of Peptide-$^{16}\text{O}_2/^{18}\text{O}_2$ Pairs

- $^{18}$O-labeling enabled mass spectrometric quantitation
- Effect of peak resolution on quantitation
- Analysis on different mass analyzer configurations
- More than relative quantitation from differential oxygen labeling
Relative Quantitation
Using Paired Isotope Clusters

Observed isotopic distribution of 1:1 mixture of $^{18}$O and $^{16}$O samples

Theoretical natural isotopic distribution

\[
\text{Ratio} = \frac{I_4}{I_0} + \left( 1 - \frac{M_2}{M_0} \right) \frac{I_2}{I_0} + \left[ \left( \frac{M_2}{M_0} \right)^2 - \frac{M_2}{M_0} - \frac{M_4}{M_0} \right]
\]

\[
\frac{I_5}{I_1}
\]
Correlation of ESI Quantitation with Peptide UV Quantitation

Peptide FVNQHLCGSHELVE

slope: 0.94±0.03
r²: 0.997

## Labeling Consistency

<table>
<thead>
<tr>
<th>BSA Peptide Sequence</th>
<th>Sequence Position</th>
<th>Unlabeled ($I_0$) Elution Time (minutes)$^a$</th>
<th>Labeled ($I_4$) Elution Time (minutes)$^a$</th>
<th>Ratio $1^b$</th>
<th>$I_4/I_0^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACFAVE</td>
<td>589-594</td>
<td>46.48</td>
<td>46.48</td>
<td>0.95</td>
<td>0.87</td>
</tr>
<tr>
<td>KKFWGKLYE</td>
<td>155-164</td>
<td>52.17</td>
<td>52.09</td>
<td>0.84</td>
<td>0.85</td>
</tr>
<tr>
<td>TYVPKAFDE</td>
<td>519-527</td>
<td>52.93</td>
<td>52.93</td>
<td>0.99</td>
<td>0.96</td>
</tr>
<tr>
<td>DKDVCKNYQE</td>
<td>335-344</td>
<td>58.43</td>
<td>58.43</td>
<td>0.99</td>
<td>0.97</td>
</tr>
<tr>
<td>DKGACLPPKIE</td>
<td>196-206</td>
<td>57.36</td>
<td>57.36</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>KQIKKQTALVE</td>
<td>544-554</td>
<td>67.70</td>
<td>67.70</td>
<td>0.90</td>
<td>0.88</td>
</tr>
<tr>
<td>LLYYANKYNGVFQE</td>
<td>177-190</td>
<td>75.32</td>
<td>75.32</td>
<td>0.99</td>
<td>1.07</td>
</tr>
<tr>
<td>YAVSVLLRLAKE</td>
<td>364-375</td>
<td>77.56</td>
<td>77.48</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>AKDAFLGSFLYE</td>
<td>345-356</td>
<td>88.78</td>
<td>88.78</td>
<td>0.86</td>
<td>0.81</td>
</tr>
<tr>
<td>DYLSLILNRLCVLHE</td>
<td>474-488</td>
<td>109.11</td>
<td>109.03</td>
<td>0.85</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Average 0.92 0.92

Standard Deviation 0.06 0.08

Correlation between MS and Western-Blot Quantification of Thioredoxin in Doxorubicin-Treated HeLa Cells

ESI-IT-MS Analysis of Anion-X Fraction 7: (Labeling: Dox Treated $^{16}$O, Control $^{18}$O)

Zoom Scan: Relative Quantification

$^{16}$O$_2$ : $^{18}$O$_2 \cong 5:1$

SDS-PAGE Analysis of Whole Cell-Extracts:

Coomassie-Stain

Western-Blot (Anti-Thioredoxin Ab)

11 kDa

Mr

Dox Control

Dox Control

Courtesy of Dan Clark of Stratagene
Effect of Peak Resolution on $^{18}\text{O}/^{16}\text{O}$ Ratio (I)

MALDI-FTICR

Effect of Peak Resolution on $^{18}$O/$^{16}$O Ratio (II)

MALDI-TOF

- APC2_Human: 0.91
- APC3_Human: 0.94
- A1AH_Human: 0.94

ESI-IT

- APC2_Human: 1.2
- APC3_Human: 0.87
- A1AH_Human: 1.3

Heller, Mattou, Menzel, Yao. JASMS 2003, 14, 704.
Effect of Isolation Window Width on Quantitation Using $^{16}$O/$^{18}$O y-Ion Pairs on IT MS

Isolation of $I_0 = 689.93$

Isolation of $I_4 = 691.93$

Protein Sequence Ions
Generated by Tandem Mass Spectrometry

Mass Analyzer 1 ➔ Fragmenting the Selection Ion ➔ Mass Analyzer 2
## Masses of Amino Acid Residues

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Symbol</th>
<th>Residue Mass (Da)</th>
<th>Immonium Ion Mass (Da)</th>
<th>Side-Chain Mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala A</td>
<td>71.04</td>
<td>44</td>
<td>15</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg R</td>
<td>156.10</td>
<td>129</td>
<td>100</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn N</td>
<td>114.04</td>
<td>87</td>
<td>58</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>Asp D</td>
<td>115.03</td>
<td>88</td>
<td>59</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys C</td>
<td>103.01</td>
<td>76</td>
<td>47</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>Glu E</td>
<td>129.04</td>
<td>102</td>
<td>73</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln Q</td>
<td>128.06</td>
<td>101</td>
<td>72</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly G</td>
<td>57.0</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Histidine</td>
<td>His H</td>
<td>137.06</td>
<td>110</td>
<td>81</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile I</td>
<td>113.08</td>
<td>86</td>
<td>57</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu L</td>
<td>113.08</td>
<td>86</td>
<td>57</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys K</td>
<td>128.10</td>
<td>101</td>
<td>72</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met M</td>
<td>131.04</td>
<td>104</td>
<td>75</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe F</td>
<td>147.07</td>
<td>120</td>
<td>91</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro P</td>
<td>97.05</td>
<td>70</td>
<td>–</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser S</td>
<td>87.03</td>
<td>60</td>
<td>31</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr T</td>
<td>101.05</td>
<td>74</td>
<td>45</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp W</td>
<td>186.08</td>
<td>159</td>
<td>30</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr Y</td>
<td>163.06</td>
<td>136</td>
<td>107</td>
</tr>
<tr>
<td>Valine</td>
<td>Val V</td>
<td>99.07</td>
<td>72</td>
<td>43</td>
</tr>
</tbody>
</table>
Quantitation Based on MS/MS Spectrum (y-Ions)

MTVEPGLEPEVR

$^{16}\text{O}/^{18}\text{O}$ Paired Peptides Facilitate and Validate Peptide Sequencing

- **QqTOF MS/MS**

- **IT MS/MS**

- C-terminal mass 575.3 Da

- N-terminal mass 470.1 Da

- Relative Intensity

- m/z 400 500 600 700 800 900 1000 1100 1200

- Peptide sequence: QD AY VS
Advantages of Modular Design

- Isotope Coding
  - Universal
    - Important to small proteins
  - Specific
  - Efficient
  - Minimal Structural Modification
    - Chromatographic co-elution
    - Stable during separation

- Separation
  - Portable to all separation platforms:
    LC/LC, solution IEF, affinity separation
Protein Pools of Digitonin Extract of MCF-7 Cells

Digitonin Extraction

- Digitonin fraction
  - cytosolic
  - soluble cytoskeletal proteins

- Properties
  - functional proteins
  - soluble proteins

Cancer Cells

Digitonin Extraction

5,000xg

Supernatant

120,000xg

Pellet

Digitonin Extract

C4 Fractionation

Modified according to “Ramsby, Makowski Method Mol. Biol., 112, 1999.”
LC-MS of Peptides from MCF-7 Digitonin Fraction

Base peak chromatography of total peptide mixture

- Protein Ratio (MelR/WT) = 1.1
- $I_0/I_4$ Ratio (MelR/WT) = 1.1 ± 0.3
- 83% (184/223) peptides in 1.1 ± 0.3

Yao, Fenselau. ASMS Annual Conference, 2001
Protein Expression Changes in MCF-7 Cells Upon Acquisition of Melphalan Resistance

[Most Proteins in a Ratio (MelR/WT) of 1.1]

(K)LLPQLTYLDGYDR(E)
PHAPI2b /April protein
pl = 4.0
MelR/WT = 2.0

(R)GIVTNWDDMEK(I)
602308605F1
NIH_MGC_88 Homo sapiens cDNA clone
MelR/WT = 2.6

Yao, Fenselau. ASMS Annual Conference, 2001
Analysis of Human Plasma Sample: Example 1

Human Plasma

Abundance Protein Depletion

Denaturing SEC

Small Proteins

Large Proteins

Fraction X

RP

Tryptic Digestion

1x Peptide Pool 1

1x Peptide Pool 2

Enzymatic $^{16}$O-Labeling

Enzymatic $^{18}$O-Labeling

2x MG $^{16}$O-Peptides

1x MG $^{18}$O-Peptides

Combine Differentially-Labeled Peptides

SCX

MALDI-MS/MS

$\mu$-RP-HPLC

ESI-MS/MS

Heller, Mattou, Menzel, Yao. JASMS 2003, 14, 704.
LC-MALDI & LC-ESI MS Analysis of Differentially $^{18}$O/$^{16}$O-Labeled Peptides Present in Human Plasma

LC-MALDI-MS (TOF)

- $m/z$ 748.3
- $m/z$ 752.3
- $m/z$ 1708.5
- $m/z$ 1712.5

LC-nanoESI-MS (QTOF)

- $m/z$ 570.29
- $m/z$ 571.63
- $m/z$ 748.43
- $m/z$ 752.47

LC-ESI-MS (IT)

- $m/z$ 570.32
- $m/z$ 571.65
- $m/z$ 748.42
- $m/z$ 752.42

Heller, Mattou, Menzel, Yao. JASMS 2003, 14, 704.
Analysis of Human Plasma Sample: Example 2

Qian et. al., Smith MCP 2005, March 7
Coupling $^{16}\text{O}/^{18}\text{O}$-Labeling and Solution Isoelectric Focusing for Peptide Analysis

- Comparative investigation of insoluble nuclear subproteome
- Sample separation after combining differentially-labeled peptides

Isotope-Coded Affinity Tag (ICAT)

- Unique Chemistry for -SH
- Affinity Tag
- Isotope-Coded Linker

Assembling Separation Module and Coding Module

Separation Module

Coding Module

Protein → H₂¹⁸/¹⁶O → Glu-C → FVNQHLCGSHLVE

Biotin

Mass/Charge

Relative Intensity
$^{16}$O/$^{18}$O-Labeling and Affinity Enrichment to Quantitate Proteins