Analytical Mass Spectrometry and Proteomics
Analytical Mass Spectrometry and Proteomics

- Genome: ~20-25,000 genes
- Transcriptome: ~100,000 transcripts
- Proteome: >1,000,000 proteins

Steps:
- Alternative promoters
- Alternative splicing
- mRNA editing
- Post-translational modifications
- Proteome Complexity
Analytical Mass Spectrometry and Proteomics

**Transcriptomics**
- RNA
- Preprocessing for Microarray
- Hybridization
- Scanning
- Normalization
- Differentially Expressed Genes
  - CTR
  - GI
  - CP

**Proteomics**
- Proteins
  - Tryptic Digestion
    - iTRAQ Labeling of tryptic peptides
    - Combine iTRAQ labeled peptides from 8 samples
    - 2D Fractionation of peptides using SCX and RP
    - Loading fractions on MALDI Plate
    - ABI 4800 MALDI TOF/TOF
    - Preprocessing and analysis using Protein Pilot
  - Normalization
  - Differentially Expressed Proteins
  - CTR
  - GI
  - CP
Analytical Mass Spectrometry and Proteomics

Analytical Realities

MS & Proteomics is not the PCR reaction.
- Not template driven
- No primers
- No amplification

Orders of Magnitude matter
*i.e.* concentration
Which of these is more abundant?
- Ribosomes
- Heat shock proteins
- Tubulin
- Albumin
- Sox-2 pluripotent stem cell transcription factor

MS analysis is not very biased

What will ionize?
- Proteins
- Lipids
- Carbohydrates
- Small molecules
- Salts

MS analysis is very sensitive, accurate, and fast.

What do these overcome?
- complexity
- isomers
- isobars
- Bad experimental design
- Keratin and other contaminants
Targeted proteomics

Analysis of a preselected group of proteins delivers more precise, quantitative, sensitive data to more biologists. Vivien Marx reports.

Although the number and identity of protein-coding genes in humans and many other organisms are known to a certain level of approximation, the numbers of proteins produced by each of these genes remains a mystery. Further complicating matters, given the many possible splice forms and post-translational modifications, the potential number of proteins is “staggering,” says Arizona State University researcher Josh LaBaer, who is also president-elect of the US Human Proteome Organization. A protein is also dynamic. “It’s phosphorylated this minute; it’s not phosphorylated the next minute,” he says. This is fascinating science, but it makes proteins in a complex, dynamic sample hard to precisely measure.

Understanding disease-related changes, for example, calls for reliable, quantitative ways of assessing protein levels, and mass spectrometers are instruments able to nail that task. But the data from so-called discovery proteomics experiments in which mass spectrometry is used to identify a

“I personally can’t wait until we stop hearing about someone describing how big of a list of proteins, peptides or phosphopeptides they detected,” says one researcher critical of discovery proteomics who did not wish to be identified. Proteomics has been doing “my list is bigger than your list” for far too long. “It is way more important to measure the one right protein than 10,000 wrong ones.”

Scientists wanting to follow well-founded hunches about dozens or hundreds of proteins seek a focused, reproducible, quantitative view of a small subset of the whole proteome in their lab vials. High-throughput biology experiments, which include DNA sequencing, genome analysis and gene expression analysis, are generating massive

Targeted proteomics detects proteins of interest with high sensitivity, quantitative accuracy and reproducibility.

aka – USE your Brain & Your skills as a cellular biologist, molecular biologist, protein biochemist, etc…even as a clinical physician.

Untargeted MS by blind faith is time consuming.

• Generates a lot of data
• Produces lots of possibilities
• Produces even better pictures
• Leaves you overwhelmed
• Without true direction to move in
• Unless you have an army
Analytical Mass Spectrometry and Proteomics

Analytical Realities

**Direct measurement** of proteins and other molecules
- Often closest to reality
- Controls Still Matter!

**Orders of Magnitude matter**
*i.e. concentration*

**Enrichment Strategies**
- Again Separation (tissue, cellular, molecular, protein class levels)
- Antibody enrichment (assuming)
- Affinity tags

**MS analysis is not very biased**
- Mass accuracy and chemical formulas do not lie
- You can separate based on
  - Class of molecules
  - Location
  - Separation (chromatography)

**MS analysis is very sensitive, accurate, and fast.**

**Targeting Strategies**
- Put the blinders on (focus)
- Create unique tags between samples
- Deplete abundant things
- Apply what is already known
- Ask specific questions
Mass Resolution = $m / \Delta m$ 50%

**FT-ICR MS**

RP = 491.2594 / 0.0055 amu
= 89,319

**LTQ MS**

RP = 491.45 / 0.69 amu
= 712
How is mass resolution calculated?

\[ R = \frac{M}{\Delta M} \]

\[ \text{FWHM} = \Delta M \]
Stable isotopes of most abundant elements of peptides

<table>
<thead>
<tr>
<th>Element</th>
<th>Mass</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>1.0078</td>
<td>99.985%</td>
</tr>
<tr>
<td></td>
<td>2.0141</td>
<td>0.015</td>
</tr>
<tr>
<td>C</td>
<td>12.0000</td>
<td>98.89</td>
</tr>
<tr>
<td></td>
<td>13.0034</td>
<td>1.11</td>
</tr>
<tr>
<td>N</td>
<td>14.0031</td>
<td>99.64</td>
</tr>
<tr>
<td></td>
<td>15.0001</td>
<td>0.36</td>
</tr>
<tr>
<td>O</td>
<td>15.9949</td>
<td>99.76</td>
</tr>
<tr>
<td></td>
<td>16.9991</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>17.9992</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Mass Resolution = $m / \Delta m$ 50%

RT: 0.16 AV: 1 NL: 6.39E5

McE_Trp_PepB03_FT #15

T: FTMS + p NSI Full ms [200.00-1200.00]

FT-ICR MS

RP = 491.2594 / 0.0055 amu
= 89,319

LTQ MS

RP = 491.45 / 0.69 amu
= 712
Mass measurement accuracy depends on resolution.

High resolution means better mass accuracy.

- 15 ppm error, Resolution = 18100
- 24 ppm error, Resolution = 14200
- 55 ppm error, Resolution = 4500
Two peptides - same nominal mass - simulation

Peptide mixture: [Val$^5$]-Angiotensin II
Sequence: DRVYVHPF
Formula: $C_{49}H_{69}N_{13}O_{12}$
Exact mass: $[M+2H]^{2+} = 516.76671$

Lys-des-Arg$^9$-Bradykinin
Sequence: KRPPGFSPF
Formula: $C_{50}H_{73}N_{13}O_{11}$
Exact mass: $[M+2H]^{2+} = 516.78490$

$\Delta m$ (mmu): RP = 18,000

18.2 mmu RP = 56,700
### Is Mass Accuracy Important?

Results for error limit up to 5 ppm

<table>
<thead>
<tr>
<th>Theoretical Mass</th>
<th>Delta ppm</th>
<th>Delta mmu</th>
<th>RDB</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ppm (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>516.76671</td>
<td>0.0</td>
<td>0.0</td>
<td>21.0</td>
<td>C\textsubscript{49}H\textsubscript{71}O\textsubscript{12}N\textsubscript{13}</td>
</tr>
<tr>
<td>516.76647</td>
<td>0.5</td>
<td>0.2</td>
<td>15.0</td>
<td>C\textsubscript{49}H\textsubscript{79}O\textsubscript{11}N\textsubscript{9}S\textsubscript{2}</td>
</tr>
<tr>
<td>2 ppm (10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>516.76638</td>
<td>0.6</td>
<td>0.3</td>
<td>12.0</td>
<td>C\textsubscript{41}H\textsubscript{75}O\textsubscript{14}N\textsubscript{15}S\textsubscript{1}</td>
</tr>
<tr>
<td>516.76705</td>
<td>-0.7</td>
<td>-0.3</td>
<td>11.5</td>
<td>C\textsubscript{43}H\textsubscript{77}O\textsubscript{15}N\textsubscript{12}S\textsubscript{1}</td>
</tr>
<tr>
<td>516.76604</td>
<td>1.3</td>
<td>0.7</td>
<td>16.0</td>
<td>C\textsubscript{48}H\textsubscript{75}O\textsubscript{16}N\textsubscript{9}</td>
</tr>
<tr>
<td>516.76738</td>
<td>-1.3</td>
<td>-0.7</td>
<td>20.5</td>
<td>C\textsubscript{51}H\textsubscript{73}O\textsubscript{13}N\textsubscript{10}</td>
</tr>
<tr>
<td>516.76604</td>
<td>1.3</td>
<td>0.7</td>
<td>21.5</td>
<td>C\textsubscript{47}H\textsubscript{69}O\textsubscript{11}N\textsubscript{10}</td>
</tr>
<tr>
<td>5 ppm (23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>516.76580</td>
<td>1.8</td>
<td>0.9</td>
<td>15.5</td>
<td>C\textsubscript{47}H\textsubscript{77}O\textsubscript{10}N\textsubscript{12}S\textsubscript{2}</td>
</tr>
<tr>
<td>516.76772</td>
<td>-2.0</td>
<td>-1.0</td>
<td>16.5</td>
<td>C\textsubscript{44}H\textsubscript{73}O\textsubscript{11}N\textsubscript{16}S\textsubscript{1}</td>
</tr>
<tr>
<td>516.76773</td>
<td>-2.0</td>
<td>-1.0</td>
<td>11.0</td>
<td>C\textsubscript{45}H\textsubscript{79}O\textsubscript{16}N\textsubscript{9}S\textsubscript{1}</td>
</tr>
<tr>
<td>516.76805</td>
<td>-2.6</td>
<td>-1.3</td>
<td>25.5</td>
<td>C\textsubscript{52}H\textsubscript{69}O\textsubscript{9}N\textsubscript{14}</td>
</tr>
<tr>
<td>516.76537</td>
<td>2.6</td>
<td>1.3</td>
<td>16.5</td>
<td>C\textsubscript{46}H\textsubscript{73}O\textsubscript{15}N\textsubscript{12}</td>
</tr>
<tr>
<td>516.76807</td>
<td>-2.6</td>
<td>-1.4</td>
<td>7.0</td>
<td>C\textsubscript{38}H\textsubscript{79}O\textsubscript{14}N\textsubscript{15}S\textsubscript{2}</td>
</tr>
<tr>
<td>516.76513</td>
<td>3.0</td>
<td>1.6</td>
<td>10.5</td>
<td>C\textsubscript{46}H\textsubscript{81}O\textsubscript{14}N\textsubscript{8}S\textsubscript{2}</td>
</tr>
<tr>
<td>516.76839</td>
<td>-3.3</td>
<td>-1.7</td>
<td>16.0</td>
<td>C\textsubscript{45}H\textsubscript{75}O\textsubscript{9}N\textsubscript{15}S\textsubscript{2}</td>
</tr>
<tr>
<td>516.76479</td>
<td>3.7</td>
<td>1.9</td>
<td>20.0</td>
<td>C\textsubscript{52}H\textsubscript{75}O\textsubscript{11}N\textsubscript{9}S\textsubscript{1}</td>
</tr>
<tr>
<td>516.76872</td>
<td>-3.9</td>
<td>-2.0</td>
<td>25.0</td>
<td>C\textsubscript{54}H\textsubscript{71}O\textsubscript{10}N\textsubscript{11}</td>
</tr>
</tbody>
</table>

Is Mass Accuracy Important?
IgAN Pathogenesis: Multi-hit Hypothesis

Autoimmune features
- Autoantigen (Gal-deficient IgA1)
- Autoantibody (binds the autoantigen)

Lai AJP 2008,
Novak Semin Immunopathol 2012

IgA1 O-glycosylation

IgA1

Hinge-region
O-linked glycans

V L
V H Cα1
Cα1 V L

N-linked glycans

Gal-deficient glycans

Ser/Thr Ser/Thr
Ser/Thr Ser/Thr
Ser/Thr Ser/Thr
Ser/Thr Ser/Thr

O-glycans

IgA1 O-glycoforms

V P S T P T P S P S T P S C
(225) (228) (230) (232) (233) (236)

V P S T P T P S P S T P S C
225 228 230 232 233 236

V P S T P T P S P S T P S C
225 228 230 232 233 236

V P S T P T P S P S T P S C
225 228 230 232 233 236

V P S T P T P S P S T P S C
225 228 230 232 233 236

GalNAc
5
Gal
3

H Y T N P S Q D V T V P C P V P S T P P T P S P S T P C

GalNAc
5
Gal
2

V P S T P T P S P S T P S C

GalNAc
4
Gal
4

IgA1 O-glycosylation

O-linked glycosylation
IgA1

Renfrow JBC 2005
Takahashi MCP 2010
Takahashi JPR 2012
LC-FT-ICR MS
Analysis of IgA1 O-glycosylation

STPPTPSPSCCHPR
232
732.9785
245

GalNAc
Gal

611.267
678.9612
732.9785
746.6541
800.6724
854.6904

m/z
650 700 750 800 850

B3 (28-30 min) B4 (30-32 min) B3 + B4
Measuring molecular mass

PeptideAA’s
+ Glycan residues
+ H⁺ (protons)
Measuring Formula Mass + protons

Peptide AA's + Glycan residues + H⁺ (protons) 1.00727 (not H 1.00782)

Elemental Composition: $\text{C60 H96 N19 O20 S2}$

**STPPTPSPSCCHPR**

<table>
<thead>
<tr>
<th>MH⁺(av)</th>
<th>MH⁺(mono)</th>
<th>MH⁺³(av)</th>
<th>MH⁺³(mono)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1467.6756</td>
<td>1466.6515</td>
<td>489.8968</td>
<td>489.5553</td>
</tr>
</tbody>
</table>

Elemental Composition: $\text{C82 H132 N21 O35 S2}$

**STPPTPSPSCCHPR**

+ 2 $\text{N}$-acetylgalactosamine + 1 galactose

<table>
<thead>
<tr>
<th>MH⁺(av)</th>
<th>MH⁺(mono)</th>
<th>MH⁺³(av)</th>
<th>MH⁺³(mono)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2035.7710</td>
<td>2034.8631</td>
<td>679.2620</td>
<td><strong>678.9592</strong></td>
</tr>
</tbody>
</table>
$[\text{STPPTPSPSCHPR} + 3\text{H}]^{3+}$

$678.9592 - 678.9612 = 0.0020 \quad 2.96 \text{ ppm mass error}$
Tandem Mass Spectrometry

Terminal GalNAc at Ser^{230} Thr^{236}
IgAN Pathogenesis: Multi-hit Hypothesis

Autoimmune features
- Autoantigen (Gal-deficient IgA1)
- Autoantibody (binds the autoantigen)
Relative abundance of individual site microheterogeneity (%)

<table>
<thead>
<tr>
<th></th>
<th>225T</th>
<th>228T</th>
<th>230S</th>
<th>232S</th>
<th>233T</th>
<th>236T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>0.4</td>
<td>N.D.</td>
<td>10.4</td>
<td>N.D.</td>
<td>64.0</td>
<td>65.4</td>
</tr>
<tr>
<td>Gal-deficient</td>
<td>0.3</td>
<td>N.D.</td>
<td>27.0</td>
<td>0.3</td>
<td>11.8</td>
<td>24.2</td>
</tr>
<tr>
<td>Disaccharides</td>
<td>99.3</td>
<td>100.0</td>
<td>62.6</td>
<td>99.7</td>
<td>24.2</td>
<td>10.3</td>
</tr>
</tbody>
</table>

V P S T P T P S T P S S T P S S T P S S

Defining IgA1 O-glycan Heterogeneity by High Resolution Mass Spectrometry

- Range of O-glycoforms
- Sites of attachment
- Sites of Gal-deficiency
- Isomers
- Quantitative value for each O-glycosylated form
- Highly reproducible

<table>
<thead>
<tr>
<th></th>
<th>225T</th>
<th>228T</th>
<th>230S</th>
<th>232S</th>
<th>233T</th>
<th>236T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>0.4</td>
<td>N.D.</td>
<td>10.4</td>
<td>N.D.</td>
<td>64.0</td>
<td>65.4</td>
</tr>
<tr>
<td>Gal-deficient</td>
<td>0.3</td>
<td>N.D.</td>
<td>27.0</td>
<td>0.3</td>
<td>11.8</td>
<td>24.2</td>
</tr>
<tr>
<td>Disaccharides</td>
<td>99.3</td>
<td>100.0</td>
<td>62.6</td>
<td>99.7</td>
<td>24.2</td>
<td>10.3</td>
</tr>
</tbody>
</table>

Relative abundance of individual site microheterogeneity (%)

<table>
<thead>
<tr>
<th>Glycopeptide</th>
<th>(n=19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPSSTPTPSPSPS</td>
<td></td>
</tr>
</tbody>
</table>

Profiled IgAN patient sera
Hit #1
Increased
Gal-deficient IgA1

Hit #2
Production of unique anti-glycan antibodies

Hit #3
Formation of pathogenic IgA1-containing circulating immune complexes

Hit #4
Glomerular
Injury

Increased ECM production
Cytokines
Growth factors

IgA1 complexes
Mesangial cell
Podocyte

Mesangial deposition
Gal-deficient IgA1 : Gal-complete IgA1
(IgAN patients vs. age / race-matched Healthy Controls)

Gal-deficient : Gal-complete ratio of a subset of IgA1 O-glycoforms

(n=19)

Gal Deficient : Gal Complete

p < 0.0001

controls

IgAN patients
IgAN Pathogenesis: Multi-hit Hypothesis

**Hit #1**
- Increased Gal-deficient IgA1

**Hit #2**
- Production of unique anti-glycan antibodies

**Hit #3**
- Formation of pathogenic IgA1-containing circulating immune complexes

**Hit #4**
- Glomerular injury
  - Proliferation
  - ECM production
  - Cytokines
  - Growth factors

Mesangial deposition
- IgA1 complexes
- Mesangial cell
- Podocyte

- Cytokines
IgAN Pathogenesis: Multi-hit Hypothesis

- **Hit #1**: Increased Gal-deficient IgA1
- **Hit #2**: Production of unique anti-glycan antibodies
- **Hit #3**: Formation of pathogenic IgA1-containing circulating immune complexes
- **Hit #4**: Glomerular injury
IgAN Pathogenesis: Multi-hit Hypothesis

Hit #1: Increased Gal-deficient IgA1

Hit #2: Production of unique anti-glycan antibodies

Hit #3: Formation of pathogenic IgA1-containing circulating immune complexes

Hit #4: Glomerular injury

- Proliferation
- ECM production
- Cytokines
- Growth factors
IgAN Pathogenesis: Multi-hit Hypothesis

- **Hit #1**: Increased Gal-deficient IgA1
- **Hit #2**: Production of unique anti-glycan antibodies
- **Hit #3**: Formation of pathogenic IgA1-containing circulating immune complexes
- **Hit #4**: Mesangial deposition

**IgA1 complexes formation**

- **Podocyte**
- **Mesangial deposition**

**Antibody production**

- **Podocytes**
- **Mesangial deposits**

**Immune complexes**

- Negative Control
- Immune Complexes
- IgA Depleted Immune Complexes

**Graph**

- Normalized spectral count
- Bars for different proteins

**UAB**

- 'Knowledge that will change your world'

**References**

- AEDTAVYYCAK
- AEDTAVYYCAR
- AEDTAVYYCSR
- SEDTAVYYCAR
- AEDTAIYYCAR
- AEDTAIYYCSR
- ADDTAVYYCAR
- ADDTAVYYCSR
- LSSVTAADTAVYYCAR