New Additions to the Stanford “Proteomic” Community

Prof. Mike Snyder

Prof. Josh Elias

Instrumental Upgrades
Bottom Up and Top Down

Top Down

Protein Centric

LC

Bottom Up

Peptide

Centric
Data Analysis and Distribution of Results

MS

Raw data extract

Database Search

Results

Sequest

Mascot

Scaffold

Stanford University

Mass Spectrometry
MS Instrumentation: High Mass Accuracy, Resolution, Sensitivity and Speed

Proteomics 2.0, Precision Proteomics

Improving Mass Accuracy in Proteomics

Better certainty of protein identifications
Ability to detect polymorphisms, post-translational modifications

<table>
<thead>
<tr>
<th>Low Resolution</th>
<th>Medium Resolution</th>
<th>High-Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 0.1 Da accuracy</td>
<td>0.1-0.01 Da accuracy</td>
<td>0.01-0.001 Da accuracy</td>
</tr>
<tr>
<td>Ion Traps, Quadrupoles, triple quadrupoles</td>
<td>Time-of-Flight, hybrids with quadrupoles</td>
<td>FT ICR MS, FT-Orbitraps, hybrids with ion traps</td>
</tr>
</tbody>
</table>
Accurate Mass Measurement Significantly Aids ID and More

A
Unit Resolution (Ion Trap)
Average mass:
Expt: 3373.2 Da
Theo: 3371.5 Da

B
High Resolution (FTMS)
Monoisotopic mass:
Expt: 3369.508 Da
Theo: 3369.510 Da
VRSE TENSALGLQVTEREEVRGRELGLK
(AA 50-79 in Lamin-B1)

A
Current MS/MS Fragmentation (mass accuracy ~30-300 ppm)
Expect Score ~10^-4

B
Next-Generation MS/MS Fragmentation (mass accuracy <1-5 ppm)
Expect Score ~10^-14
1. Anecdotal - report of a single measurement

2. Statistical - accuracy estimated from a statistical distribution of mass errors

3. Max. allowed mass deviation (MMD) - mass accuracy cutoff value when database searching
Need for Sensitivity Over 10 Orders of Magnitude

Human Proteome Organization
Depth of Proteome Coverage

Avg. Sequence coverage 30%

Figure 2 | Proteome coverage. a, Comparison of coverage of MS-based proteomics with GFP- and TAP-tagging methods. Numbers are the identified proteins by each method and, in parentheses, the number of dubious open reading frames (ORFs). b, Identified proteins per copy number bin for MS-based proteomics and the two tagging approaches. Copy numbers were estimated by correlation between summed peptide intensity per protein and the quantitative western blotting data (Methods).
Quantification

• Stable isotope labeling by amino acids cell culture (SILAC)

• Isobaric labeling for relative and absolute quantification (iTRAQ)

• Stable isotope dimethyl labeling

• Label free
The "Histome"
Modifications the Need for Enrichment

Fig. 4. Theoretical distribution of unmodified tryptic peptide concentrations in a complex biological sample (solid line) and the resulting distribution of modified peptide concentrations (dashed line) assuming 10 modifications per peptide at a substoichiometric range of 1:2 to 1:100.

Fig. 5. The average number of modified peptides per single unmodified peptide at a given concentration. The distributions of modified and unmodified peptides are shown in Fig. 4.

Phosphopeptide Analysis
Phosphopeptide Analysis using ETD and/or HCD

11,955 Phosphopeptides
Evaluating Data for Accuracy

Identification at the peptide level:
Probability at the protein level:
### Protein Sequence

<table>
<thead>
<tr>
<th>MAPFLRIAFN</th>
<th>SYELGSLOQE</th>
<th>DEANQPFCAV</th>
<th>KMKEALSTER</th>
<th>GKTLVQQKKTPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYPEWKTSTFD</td>
<td>AHIYEGVRVIQ</td>
<td>IYLMR</td>
<td>AEEEP</td>
<td>VSVETVGVSV</td>
</tr>
<tr>
<td>KAETFWDLOQP</td>
<td>QAKVLMSVQY</td>
<td>FLEDVDCKQS</td>
<td>MRSDEAKFP</td>
<td>TMNRRGAIKQ</td>
</tr>
<tr>
<td>AKIHYIKHEF</td>
<td>IATFFGQPT</td>
<td>FCSVCKDFVW</td>
<td>GLNKQGYKCR</td>
<td>QCNAAIHKKC</td>
</tr>
<tr>
<td>IDKIIIRCTGT</td>
<td>TAANSRDTIF</td>
<td>QKERFNIIDMP</td>
<td>HRFKVKHYNMS</td>
<td>PTFCDHCGLS</td>
</tr>
<tr>
<td>LWGLVQGLKL</td>
<td>CEDCGMNVHH</td>
<td>KCREKVANLC</td>
<td>GINQLLLAEAA</td>
<td>LNOVTQRASR</td>
</tr>
<tr>
<td>RPSDSASSEPV</td>
<td>GIYQGFKEKT</td>
<td>GVAGEDMQDN</td>
<td>SGTYGKIWEGB</td>
<td>SSKCNINNF</td>
</tr>
<tr>
<td>FHKVLGKGSGF</td>
<td>GKVLLGELKG</td>
<td>RGEYFAIKAL</td>
<td>KKDVAVLDIDD</td>
<td>VECMTVEKRV</td>
</tr>
<tr>
<td>TLTAANPFPL</td>
<td>THLICTFQTK</td>
<td>DHLFFVMEFL</td>
<td>NGGDLMYHIQ</td>
<td>DKGKRFELYRA</td>
</tr>
<tr>
<td>TFLYAAEIMCG</td>
<td>LQFLHSHKII</td>
<td>YRLDKLQDNVL</td>
<td>LDRDCHIKIA</td>
<td>DFGMCKENIF</td>
</tr>
<tr>
<td>GESRATASTFG</td>
<td>TPDYIAPEIL</td>
<td>QGLKYTFSVD</td>
<td>WWSFGVLLYE</td>
<td>MIGQSPFHG</td>
</tr>
<tr>
<td>DDEDELFESEI</td>
<td>RVDTPHYPRW</td>
<td>ITKESKDILE</td>
<td>KLFEREPKT</td>
<td>LGVTGNIKH</td>
</tr>
<tr>
<td>PFFKTINWTL</td>
<td>LEKRRLEPPF</td>
<td>RPKVKSPRDP</td>
<td>SNFDQEFUNE</td>
<td>KARLSDKSN</td>
</tr>
<tr>
<td>LIDSMDQSAF</td>
<td>AGFSFVNPKF</td>
<td>EHLLED</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FDR: False Discovery Rate

For a number $m$ of MS/MS spectra (probability within dataset)

FPR: False Positive Rate

A single spectrum
Sample Prep: Selective to Desired Outcome

Full characterization of a single molecule-including PTMs?

Global proteome study?

Specific for phosphorylation, acetylation, methylation, ubiquitination.....?

Sample

GeLCMS  Solution phase
GeLCMS Works Better Than Ever..

<table>
<thead>
<tr>
<th></th>
<th>Gel Region</th>
<th>1*</th>
<th>2</th>
<th>3</th>
<th>4*</th>
<th>5*</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9*</th>
<th>Total</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overnight</strong></td>
<td>Peptides</td>
<td>135</td>
<td>210</td>
<td>271</td>
<td>108</td>
<td>96</td>
<td>243</td>
<td>163</td>
<td>130</td>
<td>51</td>
<td>1407</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteins</td>
<td>23</td>
<td>28</td>
<td>36</td>
<td>18</td>
<td>13</td>
<td>45</td>
<td>36</td>
<td>28</td>
<td>6</td>
<td>233</td>
<td></td>
</tr>
<tr>
<td><strong>Pmax_1hr</strong></td>
<td>Peptides</td>
<td>210</td>
<td>198</td>
<td>253</td>
<td>253</td>
<td>204</td>
<td>266</td>
<td>149</td>
<td>147</td>
<td>98</td>
<td>1778</td>
<td>&gt;371</td>
</tr>
<tr>
<td></td>
<td>Proteins</td>
<td>31</td>
<td>28</td>
<td>36</td>
<td>29</td>
<td>24</td>
<td>45</td>
<td>37</td>
<td>31</td>
<td>12</td>
<td>273</td>
<td>&gt;40</td>
</tr>
</tbody>
</table>
GeLCMS Works Better Than Ever.. But

Coomassie

Detection Limits

Brilliant Blue 50 ng
Colloidal 10-20 ng

Silver

Mass Spec
Compatible* 1-5 ng

*No fixing/staining steps involving formaldehyde/glutaraldehyde
The Compromise

Sypro Ruby

Detection Limit

5-10 ng
Size Matters? Why

- Detection limit of protein staining is on a weight basis.
- Detection limit of protein with the mass spectrometer is on a molar basis.
- Higher the molecular weight, at the same mass, the higher the detection limit will be for the mass spectrometer.
- 1.0 ng of a 15kd protein is 67 fmol, while 1.0 ng of a 250kd protein is only 4 fmol.
- Both proteins will have similar stain intensities, but there is 15 times less protein on a molar basis from the 250kd protein.
- Protein stains detect total protein, mass spectrometer detects proteins individually.
Most all surfactants and detergents are detrimental

Triton X-100

SDS
Figure 4 | Overlap between phosphopeptide isolation methods on the level of identified phosphorylation sites. dbbTiO₂ is not shown, as 95% of the phosphopeptides identified from the dbbTiO₂ samples were also identified in the pTiO₂ samples.

B. Bodenmiller et al, 234 | VOL.4 NO.3 | MARCH 2007 | NATURE METHODS
Solution: FASP Solubilizing the Proteome

1. **Solubilize in SDS**
2. **MW cutoff filter**
3. **Re-solubilize in 8M Urea**
4. **Fractionate**
   - OGE
   - SCX
5. **Digest Lys-C, Trypsin**
Solution: FASP Solubilizing the Proteome Results
Acknowledgements

SUMS
• Allis Chien
• Karolina Krasinska
• Pavel Aronov
• Theresa McLaughlin

• Vincent and Stella Coates Foundation