Beyond Trypsin: Strategies to Improve Mass Spec Sequence Coverage and PTM Analysis

Michael M. Rosenblatt
mike.rosenblatt@promega.com

July 10, 2012
Presentation Outline

I. Introduction to Bottom-Up Proteomics

II. The Role of Trypsin

III. Trypsin Enhancers

IV. When Trypsin is not enough...
   1) Lys-C – a tool for proteolysis under denaturing conditions
   2) Other proteases to increase coverage
   3) Membrane protein analysis
   4) Post-Translational Modification (PTM) analysis

V. Glycosidases
Applications of Mass Spec in Biology

Mass spec can answer multiple questions in biology

Protein Structure (HDX Mass Spec)

Protein Interactions

Biomarker Discovery

Subcellular Localization

Protein Expression

Drug Discovery

Drug Binding Studies (Chemical Proteomics)
**Why is Mass Spec a Powerful Tool for Biology?**

**Western Blot: Traditional Approach**

Antibody Required

**Mass Spec: Unbiased Analysis**

Chromatogram

MS/MSpectra

Protein ID based on matches to Spectra

**Mass Spec Advantages:**

- No pre-existing knowledge required - unbiased
- Ability to identify 1000’s of proteins in single run
- Quantitative
- Highly reproducible
- High dynamic range
- Automatable
- Eliminates need for Antibody
Bottom-Up Proteomics: Unbiased Profiling of Complex Protein Extracts

Proteases convert protein to peptides to be analysed by the Mass Spec

**Step #1:** Sample Preparation (Pure Protein, Protein Complex, IP, Cells, Tissue, Serum etc.)

**Step #2:** Reduction of Sample Complexity (SDS-PAGE, SCX/HILIC, affinity enrichment, subcellular fractionation etc.)

**Step #3:** Protein Isolation (excise gel band, collect fractions etc.)

**Step #4:** Protein digestion (trypsin or other proteases)

**Step #5:** LC-MS/MS Analysis

**Step #6:** Software assisted protein identification
The **range of masses** is limited in the typical mass spec (250-4000 daltons). This is not a problem for small molecule measurements, but large molecules are challenging...

*The problem is solved by using site-specific endoproteinases (proteases).*
Why Trypsin?

Protein → Proteolysis → Peptides → Mass spec analysis → Peptide Sequence

Why is Trypsin viewed as the “Gold Standard” for bottom up proteomics?
**Why Trypsin....Multiple Reasons!**

- Average size of peptides is between 700-1500 daltons (ideal for MS analysis)
- All peptides have a C-terminal charge (due to K/R)
- Highly active
- Highly specific
- Autolysis can be controlled by lysine/arginine modification
How Does Mass Spec “Sequence” Proteins?

Steps in the Process:
1. Protein sequences are stored in a database (i.e. UNIPROT)
2. Sequences are digested in silico (based on the appropriate protease)
3. Peptide masses are measured and MS/MS spectra recorded
4. Peptide sequence candidates which match the mass of the peptides measured (within a certain tolerance) are selected and a theoretical MS/MS spectrum is generated (i.e. a bar-code)
5. Bar-code is then matched to the MS/MS spectrum.

Sequence: CGQESEIIYTDKEKK

- b-type ions: N-terminal origination
- y-type ions: C-terminal origination

The spacing between the ions correspond to the masses of the amino acids.
What are the Challenges With Proteolysis?

**Challenges:**
- Protein folded too tightly – Protease can’t access
- Protein is insoluble and require additives

**Solutions:**
- Urea, Guanidine HCl and organic solvents (e.g., acetonitrile) – denaturing agents
- Detergents such as Triton X-100 and SDS

**Limitations:**
- Protease Inhibition
- Unwanted side effects (i.e., protein modification).
- Detergents hurt the hardware (LC and Mass Spec)
**ProteaseMAX™ Surfactant**

Cleavable bonds

\[ \text{Acid or temperature} \]

ProteaseMAX™ surfactant $\rightleftharpoons$ Hydrophobic tail + Zwitterionic head

- Solubilizing and denaturing properties of surfactant – no effect on LC-MS/MS
- Decomposes within 8 hours at Room Temperature (in solution)
- Improves digestion efficiency for compact/tightly folded proteins
- Solubilizes Membrane Proteins
Myoglobin (Proteolytically resistant protein) is rapidly digested in presence of ProteaseMAX™

30 min digestion at 37 °C

without surfactant

Intact myoglobin

30 min digestion at 37 °C

with surfactant

Peptides
ProteaseMAX™ readily solubilizes the membrane protein bacteriorhodopsin

- Bacteriorhodopsin is insoluble in aqueous solutions and resistant to proteolysis (panel A)
- With the surfactant, this protein solubilizes within 1-2 minutes at room temperature and easily digested (panel B)
ProteaseMAX™ Surfactant: Conclusions

- Improved digestion of Stably folded proteins
- Improved solubility of membrane proteins
- Faster overall digestion kinetics
- Higher throughput
- No cleanup required
But, When is Trypsin Not Enough?

- Substantial number of tryptic peptides are too long or too short for Mass Spec analysis
- Tryptic cleavage sites might not be accessible due to PTM’s (i.e. phosphorylation, glycosylation, histone methylation/acetylation)
- Certain proteins are not efficiently digested by trypsin (i.e. membrane proteins and proteins in tight conformation)
- Additional proteases allow for more sequence coverage and protein ID’s (especially in complex samples)
Alternative Proteases Overcome the Limitations of Trypsin

Examples to be discussed:

- Using other proteases to increase protein identifications
- Improve digestion efficiency of compact proteins with Lys-C
- “Sequence” membrane proteins with Elastase, Pepsin, and Thermolysin
- Complementation of trypsin in Histone analysis with Arg-C
- Identification of phosphorylation sites on MAPK requires Trypsin, Chymotrypsin, and Elastase
# Alternative Proteases For Mass Spec

Examples of the Most Commonly Used Proteases (in addition to trypsin)

<table>
<thead>
<tr>
<th>Protease</th>
<th>Cleavage site</th>
<th>Example of use</th>
</tr>
</thead>
</table>
| Lys-C     | NNNN[K]NNN (K is lysine) | • Active under denaturing conditions  
• Produces larger peptides than trypsin  
• Useful for ETD applications |
| Glu-C     | NNNN[E]NNN (E is glutamate)  
Glu-C can also cleave at aspartate residue also depending on the pH | • Alternative to trypsin when trypsin produces peptides outside of required mass window (too small or larger)  
• E cleavage in Phosphate buffer (pH = 7.8)  
• E and D in ammonium buffers. |
| Asp-N     | NNNN[D]NNN (D is aspartate) | Alternative to trypsin when trypsin produces peptides outside of required mass window (too small or larger) |
| Chymotrypsin | NNNN(F/Y/W)[D]NNN (F, Y and W are aromatic residues phenylalanine, tyrosine, and tryptophan) | Digests hydrophobic proteins (i.e., membrane proteins) |
| Arg-C     | NNNN[R]NNN (R is arginine)  
(can also cleave c-terminal side of K) | Analysis of histone posttranslational modifications |
| Pepsin    | Nonspecific protease (advantage – digestion at low pH) | Hydrogen-deuterium Exchange Mass Spec |
| Thermolysin | Nonspecific protease (advantage – digestion at high temperature) | • Digestion of proteolytically difficult proteins  
• Structural Studies |
| Elastase  | Nonspecific protease | • Used to increase protein coverage  
• Protein structural studies |
Alternative Proteases Increase Protein Identifications by 20 %

Identified proteins increased from 3313 to 3908 (20%), upon applying alternative proteases.

Sequence coverage increased by 172%.

Swaney et al. (2010) J. of Proteome Res. 9:1323-1329
Lys-C: A Valuable Tool for Proteomics Studies

- As **highly active and specific** as trypsin

- **Unlike trypsin, lys-C remains active under highly denaturing conditions** (8M Urea) allowing for digestion of unfolded proteins (which don’t digest when folded)

- **Generates larger peptides** than trypsin (useful for ETD studies)
An Alternative Workflow: **Sequential Digestion with Lys-C and Trypsin**

- Lys-C can digest the unfolded protein without loss of activity.
- Sequential digestion with Lys-C followed by trypsin is often used to ensure high digestion efficiency.
- Many proteins require denaturation prior to digestion.

**Steps:**

1. Digest with Lys-C in 8M Urea
2. Dilute to 1.5 M Urea and digest with trypsin
3. Analyze using a mass spectrometer
**A Recombinant Lys-C: A Low-Cost Alternative to the Native Enzyme**

- Promega has developed a novel recombinant Lys-C ($8/\mu g$ versus $30/\mu g$)
- rLys-C is as active and robust as the native enzyme with performance comparable to other suppliers
- rLys-C is also active under denaturing conditions

**Total Protein Coverage**
*Yeast Extract*

<table>
<thead>
<tr>
<th>Identified peptides</th>
<th>Identified proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>rLys-C</td>
<td>Lys-C, Supplier A</td>
</tr>
<tr>
<td>rLys-C</td>
<td>Lys-C, Supplier A</td>
</tr>
</tbody>
</table>
**rLys-C Can Tolerate Denaturation**

Recombinant Lys-C retains activity under strongly denaturing conditions (8M Urea) digesting a proteolytically resistant protein.
A Trypsin/Lys-C Mixture: Improved Proteolytic Efficiency

Supplementing trypsin with Lys-C dramatically improves digestion efficiency

Product available later this year: e-mail mike.rosenblatt@promega.com or gary.kobs@promega.com for details
Robust and Consistent Proteolysis of Complex Protein Mixtures with Trypsin/Lys-C

Digestion of yeast and mouse protein extracts with Trypsin/rLys-C mix

<table>
<thead>
<tr>
<th></th>
<th>Yeast extract #1</th>
<th>Yeast extract #2</th>
<th>Mouse extract #1</th>
<th>Mouse extract #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completely digested</td>
<td>94.8%</td>
<td>95.4%</td>
<td>96.7%</td>
<td>95.6%</td>
</tr>
<tr>
<td>peptides, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The remaining missed cleavages represent a few percent of total peptide population and predominantly occur at (R/K)/(D/E) sites or at peptide N-termini.
Membrane Proteins: A Proteolytic Challenge

- Membrane Proteins contain a large number of hydrophobic residues and therefore are challenging for trypsin to digest
- They also require solubilizing agents for proteolysis, another problem for trypsin
- Additional proteases, like Pepsin, elastase, and thermolysin, which will cut around hydrophobic sequences are beneficial

**Bacteriorhodopsin:**

- Prototypical membrane protein (7-TM GPCR)
- Trypsin does not cleave any of the 7 helices
Multiple Proteases Improves Coverage of Membrane Proteins

Bacteriorhodopsin:

**MLELLPTAVEGVSQAQITGRPEW|VLALGTALMLGLGTLYFLVKGGMGVS|PD|AKKFAITTLVPAIAFTMYL|SMLLG|YGLTM|VPFGGGEQNP|IY|W|AR|Y|ADWLFTTP|LLL|LALLVDAD|QGTILALVGA|G|MIITG|L|VGA|LTKVY|SY|FVW|WAI|STAAMLYILYVL|F|FGFTSKAESMR|PE|VASTFKL|RNVTVVLW|SAYPV|VVL|GSEGAG|IVPLNIE|T|LM|F|MVLDVSA|KV|G|F|GL|L|LRSRAIFGEA|E|APEPSAGDGA|AATS**

- Red and Blue are alternating tryptic peptides
- Underlined sequences correspond to embedded TM regions

Trypsin 27% Coverage

**tr|B0R5N9|B0R5N9_HALS3 (100%), 28,257.6 Da**

Bacteriorhodopsin OS=Halobacterium salinarum (strain ATCC 29341 / DSM 671 / R1) GN=bop PE=4 SV=1

14 unique peptides, 15 unique spectra, 22 total spectra, 72/262 amino acids (27% coverage)

Combination of coverage from Elastase, Thermolsyin and Pepsin 90% Coverage

**tr|B0R5N9|B0R5N9_HALS3 (100%), 28,257.6 Da**

Bacteriorhodopsin OS=Halobacterium salinarum (strain ATCC 29341 / DSM 671 / R1) GN=bop PE=4 SV=1

222 unique peptides, 236 unique spectra, 579 total spectra, 236/262 amino acids (90% coverage)

The combination of 3 additional enzymes increased coverage by over 60% and identified 7/8 trans-membrane domains
The Principle of PTM analysis by MS/MS

The spacing between the fragment ions will have a specific mass shift that is PTM specific

Methylation (K)
VPMVDIkGPK

Methylation (R)
GLGTDESLIELLLcSr

Ubiquitination (K)
LIFAGkQLEDGR

Phosphorylation (S/T/Y)
QVAEQGGDLSPAANR

Acetylation (K)
GLGkGGAkR

Nitration (Y)
VVLAyEPVWAIGTGK

Modified residues are in blue

Data courtesy of MS Bioworks LLC
Cell Signaling: A Post-Translational Modification (PTM) Driven Process

- Signal Transduction is central to cell growth and plays a central role in multiple diseases including cancer and diabetes.
- Much of the signalling is governed by PTM’s like phosphorylation, acetylation, ubiquitination, and O-GlcNac to name a few.
- Mass Spec is a powerful tool for both site-specific identification of not only PTM modified proteins but the specific site of modification.

Some Major PTM’s

<table>
<thead>
<tr>
<th>PTM</th>
<th>Affected Residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetylation</td>
<td>Lysine</td>
</tr>
<tr>
<td>Methylation</td>
<td>Lysine (up to 2 methyl groups) /Arginine (up to 3 methyl groups)</td>
</tr>
<tr>
<td>Ubiquitination</td>
<td>Lysine</td>
</tr>
<tr>
<td>Phosphorylation</td>
<td>Serine /Threonine /Tyrosine</td>
</tr>
<tr>
<td>O-GlcNac</td>
<td>Serine /Threonine /Tyrosine</td>
</tr>
<tr>
<td>O-glycans</td>
<td>Serine /Threonine</td>
</tr>
<tr>
<td>N-glycans</td>
<td>Asparagine</td>
</tr>
<tr>
<td>Nitration</td>
<td>Cysteine/tyrosine</td>
</tr>
</tbody>
</table>
Arg-C Complements Trypsin for Analysis of Histone Modifications

N-terminus of Histone H4:

MSGRGKGGKG  LGKGGAKRHR  KVLRDNIQGI  TKPAIRRLAR  RGGVKRISGL

- Methylated site identified with trypsin
- Di-methylated site identified with trypsin
- Acetylated sites identified with trypsin
- Additional methylated site identified with Arg-C
- Additional di-methylated sites identified with Arg-C
- Additional acetylated site identified with Arg-C
Protease Combinations Increase Identification of Phospho Sites

Identification of Tyrosine Phosphorylation in Erk1/2

Trypsin – IADPEHDHTGFLTE(pY)VATR

Chymotrypsin – LTE(pY)VATRW

Elastase – GFLTE(pY)VAT

MS/MS spectrum is of poor quality

Note complimentary y- and b-ion pairs around pY

Note complimentary y- and b-ion pairs around pY

Data courtesy of MS Bioworks LLC
Glycosidases Have Multiple Roles in Proteomics

Glycosidases are useful for:

- Glycomics
- Glycoproteomics
- Improving protease coverage by unmasking protease sites

Cleavage site is inaccessible for trypsin

[Diagram showing the structure of proteins and glycans with cleavage sites and protease sites.]
## Some Common Glycosidases

<table>
<thead>
<tr>
<th>Glycosidase</th>
<th>Substrate</th>
<th>Product</th>
<th>Change Detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNGase F</td>
<td>XXNX(S/T)XX</td>
<td>XXDX(S/T)XX</td>
<td>1 dalton Mass Shift – N to D conversion</td>
</tr>
<tr>
<td></td>
<td>N = Asn (glycan containing)</td>
<td>D = Asp</td>
<td></td>
</tr>
<tr>
<td>Endo-H</td>
<td>XXNX(S/T)XX</td>
<td>XXN(GlcNac)X(S/T)XX</td>
<td>203 daltons – Mass of one GlcNac</td>
</tr>
<tr>
<td></td>
<td>N = Asn (glycan containing)</td>
<td>N(GlcNac) = Asparagine bonded to a single GlcNac</td>
<td></td>
</tr>
<tr>
<td>Protein Deglycosylation Mixture</td>
<td>All glycoproteins</td>
<td>XXDX(S/T)XX</td>
<td>• 1 dalton for N-linked glycans</td>
</tr>
<tr>
<td>(contains both N and O-glycosidases)</td>
<td></td>
<td>D = Asp</td>
<td>• No change for O-linked, but no glycan attached, so peptide is unmodified.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(O-linked amino acids are unchanged)</td>
<td></td>
</tr>
</tbody>
</table>

*Glycosidases will launch in Fall 2012*
**PNGase F /Endo H Cleavage Specificity**

**High Mannose Structure**
- N,N’-diacetylchitobiose core
- \( \alpha_1,6 \) and \( \alpha_1,3 \) mannoses
- \( \beta_1,4 \) galactoses

**Hybrid Structure**
- Endo H cleavage Site
- PNGase F cleavage Site
- \( \alpha_1,6 \) and \( \alpha_1,3 \) mannoses
- \( \beta_1,4 \) galactoses

**Complex Glycan Structure**
- Sialic Acid

*Endo H is unable to cleave N-linked complex-type glycans*
A Recombinant PNGase F

- Protein is based on the 34 kDa protein secreted by *flavobacterium meningosepticum*
- Total Activity is comparable to the endogenous enzyme
- Active under native and mildly denaturing conditions and therefore very effective for proteomic workflows

*Note:* After removal of the glycan, the Asparagine (N) residue is converted to Aspartic Acid (D)
Characterization of PNGase F Treated Proteins

PNGase F - + - + - +

250 150 100 75 50 37 25 20 15

IgG Ovalbumin RNAse B

Protein Sequence:
Alpha-1-acid glycoprotein OS=Bos taurus GN=ORM1 PE=2 SV=1
16 unique peptides, 22 unique spectra, 32 total spectra, 132/202 amino acids (65% coverage)

Identification of Glycosylation Sites:
Band shifted to lower MW indicates removal of N-glycan chains
**PNGase F Treatment Improves Sequence Coverage**

- Majority of sequence covered without PNGase F treatment
- However, highlighted glycan was not observed due to attached glycan
- Treatment with PNGase F gave sites of attachment ((N 207 and N211) were identified)
- Site of attachment confirmed by MS/MS analysis (see above spectrum)
An Example of a Typical Serum Glycopeptide Identification Workflow

**Step #1:** Obtain Serum Sample

**Step #2:** Deplete Serum of IgG/Albumin

**Step #3:** Trypsin digestion

**Optional Step #4:** Enrich Glycopeptides with Lectins (ConA/WGA)

**Step #5:** Glycosidase (PNGase F) Treatment

**Step #5:** LC-MS/MS Analysis

**Step #6:** Software assisted Protein Identification

Glycan
PNGase F Increases Identification of Glycopeptides

**Total Deamidated SpC – Serum Sample**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Depleted/Enriched</td>
</tr>
<tr>
<td>B</td>
<td>Depleted/Enriched</td>
</tr>
<tr>
<td>C</td>
<td>Depleted</td>
</tr>
<tr>
<td>D</td>
<td>Depleted</td>
</tr>
<tr>
<td>E</td>
<td>Enriched</td>
</tr>
<tr>
<td>F</td>
<td>Enriched</td>
</tr>
<tr>
<td>G</td>
<td>None</td>
</tr>
<tr>
<td>H</td>
<td>None</td>
</tr>
</tbody>
</table>

- Albumin/IgG was removed from Serum
- WGA and ConA lectins were used to enrich samples
- Depletion alone appears to yield the greatest number of glycopeptides identified
**Endo H Treatment is Effective in Deglycosylating Glycoproteins**

<table>
<thead>
<tr>
<th>Endo H</th>
<th>-</th>
<th>+</th>
<th>-</th>
<th>+</th>
<th>-</th>
<th>+</th>
<th>-</th>
<th>+</th>
<th>-</th>
<th>+</th>
<th>-</th>
<th>+</th>
</tr>
</thead>
</table>

| IgG | Ovalbumin | RNAse B | Lactoperox | Acid Glycoprot | Fetuin |

Endo H is reactive toward some, but not all, glycoprotein substrates. Thus, a powerful biochemical tool to determine glycan compositions.
Endo H Treatment of Lactoperoxidase Identifies Glycosylation Sites

Using a combination of Trypsin and Endo H, followed by Mass Spec analysis, the site of glycan attachment (N203) could be determined, precisely.
Conclusions

- Trypsin is the best protease to start with when preparing Mass Spec samples.
- The use of a Mass Spec compatible solubilizing agent (i.e. ProteaseMax™) can dramatically improve the efficiency of proteolysis.
- Using alternative proteases will increase the number of protein ID’s.
- Alternative proteases may also increase your identification and confidence in assignment of PTM’s.
- For identification of glycosylation sites, the use of either PNGase F or Endo H may be required.
# Promega’s Protease Portfolio

<table>
<thead>
<tr>
<th>Product</th>
<th>Catalog #</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsin Gold</td>
<td>V5280</td>
<td>100 μg</td>
</tr>
<tr>
<td>Trypsin, Sequencing Grade</td>
<td>V5111, V5113 (Frozen)</td>
<td>100 μg</td>
</tr>
<tr>
<td>Immobilized Trypsin</td>
<td>V9012, V9013</td>
<td>2, 4 mL</td>
</tr>
<tr>
<td>Lys-c</td>
<td>V1071</td>
<td>5 μg</td>
</tr>
<tr>
<td>Recombinant Lys-c</td>
<td>V1671</td>
<td>15 μg</td>
</tr>
<tr>
<td>Arg-C</td>
<td>V1881</td>
<td>10 μg</td>
</tr>
<tr>
<td>Asp-N</td>
<td>V1621</td>
<td>2 μg</td>
</tr>
<tr>
<td>Glu-C</td>
<td>V1651</td>
<td>50 μg</td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td>V1061, V1062</td>
<td>25 μg, 100 μg</td>
</tr>
<tr>
<td>Thermolysin</td>
<td>V4001</td>
<td>25 mg</td>
</tr>
<tr>
<td>Pepsin</td>
<td>V1959</td>
<td>250 mg</td>
</tr>
<tr>
<td>Elastase</td>
<td>V1891</td>
<td>5 mg</td>
</tr>
<tr>
<td>Protease Max</td>
<td>V2071, V2072</td>
<td>1 mg, 5 mg</td>
</tr>
</tbody>
</table>
# New Glycosidases from Promega

<table>
<thead>
<tr>
<th>Product</th>
<th>Catalog #</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNGase F</td>
<td>V4831</td>
<td>500 mIU</td>
</tr>
<tr>
<td>Endo H</td>
<td>V4871, V4875</td>
<td>10,000 and 50,000 units</td>
</tr>
<tr>
<td>Deglycosylation Mixture</td>
<td>V4931</td>
<td>100 µL</td>
</tr>
<tr>
<td>Fetuin</td>
<td>V4961</td>
<td>500 µg</td>
</tr>
</tbody>
</table>

*These products should be available some time in Q4, 2012*