Sample Preparation Methods for MS Based Proteomics

Lecture 2
31 March, 2010

Shenheng Guan
Department of Pharmaceutical Chemistry
University of California, San Francisco
sguan@cgl.ucsf.edu
Lecture 2 Outline

• Digestion Enzymes

• Protein and Peptide Separation Methods

• HPLC

• Multidimensional LC

• Microfluidics
MS Plays a Key Role

Functional Diversity/System Complexity

Genomics → Proteomics → Metabolomics

Functional Proteomics/Genomics

Systems Biology
Proteins

Pure Protein

Peptides

Pure Peptide

Gel/AC

Digest

LC

Topdown

Shotgun

Bottomup

ESI

Molecular Ions

Single m/z Ions

Fragments

MS Spectra

1st MS

Dissociation

2nd MS

Protein ID

Protein Quantification

PTM Sites

Other Software Tools

Search Engine

MS Operation

Computing Operation

Proteomics Approaches
Protein complex analysis by mass spectrometry

Affinity purification/SDS gel separation/in-gel digest/LCMSMS mass spectrometry analysis/Protein sequence database search for protein identification is the commonly used bioanalytical methodology.

Large scale experiments are possible (Nature2006v440p637)
Isolating Interacting Proteins by Affinity chromatography

*epitope tags*

<table>
<thead>
<tr>
<th>epitope tags</th>
<th>composition</th>
<th>affinity matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLAG</td>
<td>DYKDDDDDK</td>
<td>FLAG antibody</td>
</tr>
<tr>
<td>HA</td>
<td>YPYDVPDYA</td>
<td>HA antibody</td>
</tr>
<tr>
<td>C-MYC</td>
<td>EQKLISEEDL</td>
<td>c-MYC antibody</td>
</tr>
<tr>
<td>6XHIS</td>
<td>HHHHHHH</td>
<td>Immobilized metal affinity (IMAC)</td>
</tr>
<tr>
<td>Biotinylation signal</td>
<td>78 amino acids</td>
<td>avidin/streptavidin</td>
</tr>
<tr>
<td>Strep binding</td>
<td>10-50 amino acids</td>
<td>avidin/streptavidin</td>
</tr>
<tr>
<td>Protein A</td>
<td>137 amino acids</td>
<td>IgG</td>
</tr>
<tr>
<td>Calmodulin binding</td>
<td>26 amino acids</td>
<td>Calmodulin</td>
</tr>
<tr>
<td>peptide</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Tandem Affinity Purification
*Increased Specificity and Decreased Contaminations*

The original yeast TAP tag

![Diagram of the original yeast TAP tag](image)

The optimised GS-TAP tag

![Diagram of the optimised GS-TAP tag](image)

Small tag for fast purification

![Diagram of the small tag](image)

The tag is used to purify cross-linked complexes in denaturing conditions

![Diagram of the cross-linked complexes](image)
Preparation of Proteins for Digestion

Or for Topdown Analysis

Denature proteins

- Make whole chain accessible
  - Detergents: SDS, acid-cleavable (AnalChem2003v75p6642)
  - 8M urea – heated may cause carbamylation (AnalBiochem1999v267p57)
  - 6M Guanidine HCl; ACN or MeOH, etc

Reduce disulfide bonds

- General in basic conditions
  - Dithiothreitol (DTT, 2-SH), TCEP – wide pH range (NatBiotech2001v19p379)

Alkylate thiol groups (in dark)

- Iodoacetic acid, +58 Da (historical)
- Iodoacetamide, +57 Da
- May have opportunity to introducing labels (e.g., ICAT)

S-S bonds are special PTM and Alkylation is an artificial modification
# Commonly Used Protein Digest Reagents

<table>
<thead>
<tr>
<th>enzymes</th>
<th>cleavage</th>
<th>don't cut</th>
<th>side</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsin</td>
<td>KR</td>
<td>P</td>
<td>c</td>
<td>aspartic protease</td>
</tr>
<tr>
<td>Lys-C</td>
<td>K</td>
<td>P</td>
<td>c</td>
<td>serine protease</td>
</tr>
<tr>
<td>Glu-C</td>
<td>E</td>
<td>P</td>
<td>c</td>
<td>serine protease</td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td>FYWL</td>
<td>P</td>
<td>c</td>
<td>serine protease</td>
</tr>
<tr>
<td>Arg-C</td>
<td>R</td>
<td>P</td>
<td>c</td>
<td>serine protease</td>
</tr>
<tr>
<td>PepsinA</td>
<td>FLE?</td>
<td>c?</td>
<td>aspartic protease</td>
<td></td>
</tr>
<tr>
<td>Asp-N</td>
<td>D</td>
<td>n</td>
<td>Metalloendopeptidase</td>
<td></td>
</tr>
<tr>
<td>Lys-N</td>
<td>K</td>
<td>n</td>
<td>Metalloendopeptidase</td>
<td></td>
</tr>
<tr>
<td>CNBr</td>
<td>M</td>
<td>c</td>
<td>chemical</td>
<td></td>
</tr>
<tr>
<td>Formic_acid</td>
<td>D</td>
<td>c</td>
<td>chemical</td>
<td></td>
</tr>
</tbody>
</table>
How to Choose A Suitable Enzyme

How specific the cleavage is
  • Glu-C: Cuts after E, some D.
  • Chymotrypsin: Cuts after F, W, Y, some L, occasionally V…

Generate suitable peptide size – 10-20 amino acid residues
  • Conveniently analyzed by LCMSMS

Digest condition is suitable to the sample
  • Pepsin works at low pH and is good for H/D exchange

Autolysis and ligation (?)
  • Don’t want any peptides from enzyme itself
Yeast 20S Proteasome Peptides

- **Tryptic**: 11 peptides
- **Asp-N**: 11 peptides
- **Glu-C**: MW distribution
- **CNBr**: MW distribution
Trypsin is the Most Popular Enzyme

• High Enzyme Specificity – cuts all Lys and Arg (to lesser extent followed by Pro).
• Produces peptides with basic C-terminus – give good (CID) fragmentation series
• Majority of peptides 7 -20 amino acids in length
• Relatively inexpensive compared to other proteases

New England BioLab’s modified Trypsin is treated with L-(tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK) to inactivate chymotryptic activity (Biochemistry 1963v2p 252). It is modified by acetylation of the ε-amino groups of lysine residues to prevent autolysis.

Modified Trypsin Protein Sequence:

```
  1 IVGGYTCaENSVPYQSLNAGYHFGGSLINDQWVSAAHCYQHYIQRGEYNID
  61 VLEGQFIDASKIKRHPKYSSWTLDNILLIKLSTPAVINAIVSTLLLPSACASA
 121 GTECLISGWNLTSSGVNYPDLLQCLVAPLLSHADCEASYPGQITNNMICAGFLEG
 181 GKDSCQGDSGPGPACNGQLQGIVSWGVCAGGKGPVYTKCNYVDWIQETIAANS
```
## Tryptic digest of hemoglobin b chain

<table>
<thead>
<tr>
<th>m/z</th>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>246.1812</td>
<td>60</td>
<td>61</td>
<td>(K)VK(A)</td>
</tr>
<tr>
<td>319.1401</td>
<td>145</td>
<td>146</td>
<td>(K)YH(-)</td>
</tr>
<tr>
<td>412.2303</td>
<td>62</td>
<td>65</td>
<td>(K)AHGK(K)</td>
</tr>
<tr>
<td>932.5200</td>
<td>9</td>
<td>17</td>
<td>(K)SAVTALWGK(V)</td>
</tr>
<tr>
<td>952.5098</td>
<td>1</td>
<td>8</td>
<td>(-)VHLTPEEK(S)</td>
</tr>
<tr>
<td>1126.5640</td>
<td>96</td>
<td>104</td>
<td>(K)LHVDPENFR(L)</td>
</tr>
<tr>
<td>1149.6739</td>
<td>133</td>
<td>144</td>
<td>(K)VVAGVANALAHK(Y)</td>
</tr>
<tr>
<td>1274.7256</td>
<td>31</td>
<td>40</td>
<td>(R)LLVVYYPWTQR(F)</td>
</tr>
<tr>
<td>1314.6648</td>
<td>18</td>
<td>30</td>
<td>(K)VNVDEVGGEALGR(L)</td>
</tr>
<tr>
<td>1378.7001</td>
<td>121</td>
<td>132</td>
<td>(K)EFTPPVQAAYQK(V)</td>
</tr>
<tr>
<td>1421.6729</td>
<td>83</td>
<td>95</td>
<td>(K)GTFATLSELHCDK(L)</td>
</tr>
<tr>
<td>1669.8908</td>
<td>67</td>
<td>82</td>
<td>(K)VLGAFSDGLAHLDNLK(G)</td>
</tr>
<tr>
<td>1719.9727</td>
<td>105</td>
<td>120</td>
<td>(R)LLGNVLVCVLAHHFGK(E)</td>
</tr>
<tr>
<td>2058.9477</td>
<td>41</td>
<td>59</td>
<td>(R)FFESFGDLSTPDAVMGNPK(V)</td>
</tr>
</tbody>
</table>
Trypsin May Not Be The Best Choice

Rhodopsin (Membrane Protein)

\[
\text{MNGTEGPNFYVPFSNATGVVRSPFEYPQYLYLAEPWQFSMLAAYMFLLILVGLFPINFLTLVTVQHKKLRTTPLNYINLNAVSDLVMVGLGGFTSTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVLALIERYYVVKPMSNRFGENHAIMGVATFWMLALACAAPPLAGWSRYP}\]

\[
\text{EGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMIIIFFCYGQLVFTVKEAAQQQESISATTQKAEEKVTRMVIMVIAFLICWVPYASVAFYIFTHQGSNFGPIFMTPIAFFAKSAAIYNPVIIIMMNKQFRNCMLTTICCGKNPLGDEASATVSKTETSQVAPA}\]

<table>
<thead>
<tr>
<th>Mass</th>
<th>Residues</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>903.4424</td>
<td>340 – 348</td>
<td>(K)TETSQVAPA (-)</td>
</tr>
<tr>
<td>1357.574</td>
<td>315 - 325</td>
<td>(R)NCMLTTICCGK (N)</td>
</tr>
<tr>
<td>1403.665</td>
<td>326 - 339</td>
<td>(K)NPLGDEASATVSK (T)</td>
</tr>
<tr>
<td>1490.709</td>
<td>232 - 245</td>
<td>(K)EAAQQQQESATTQK (A)</td>
</tr>
<tr>
<td>1499.75</td>
<td>136 - 147</td>
<td>(R)YVVVCKPMSNFR (F)</td>
</tr>
<tr>
<td>1727.887</td>
<td>297 - 311</td>
<td>(K)SAAIYNPVIIIMMNK (Q)</td>
</tr>
<tr>
<td>2257.071</td>
<td>1 - 21</td>
<td>-( )MNGTEGPNFYVPFSNATGVVR (S)</td>
</tr>
<tr>
<td>3231.569</td>
<td>148 - 177</td>
<td>(R)FGENHAIMGVAFTWWMALACAAPPLAGWSR (Y)</td>
</tr>
<tr>
<td>5058.603</td>
<td>253 - 296</td>
<td>(R)MVIIMVIAFLICWVPYASVAFYIFTHQGSNFGPIFMTPIAFFAK (S)</td>
</tr>
<tr>
<td>5388.76</td>
<td>22 - 66</td>
<td>(R)SPFEYPQYLYLAEPWQFSMLAAYMFLLILVGLFPINFLTLVTVQHK (K)</td>
</tr>
<tr>
<td>6503.206</td>
<td>178 - 231</td>
<td>(R)YIPPLGQSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMIIIFFCYGQLVFTVK (E)</td>
</tr>
<tr>
<td>7182.745</td>
<td>70 - 135</td>
<td>(R)TPLNYILLNLAVDLFMVGLGGFTSTLYTSLHGYFVFGPTGCNLEGFFATLGGGEIALWSLVLALIER (Y)</td>
</tr>
</tbody>
</table>

• Also, protein solubility may be an issue for membrane proteins
In-gel Digestion

If possible, alkylate proteins before running gel

1. Cut gel
2. **Wash** with 25mM ABC in 70% ACN
3. Reduce with 10mM DTT in 25mM ABC Alkylate with 50mM IAA in 25mM ABC
4. **Wash** with 25mM ABC in 70% CAN (dehydrate)
5. Digest with trypsin in 25mM ABC over night
6. Extract peptides with 5% FA in 50% ACN
7. Remove ACN before LCMSMS analysis

   ABC (ammonium bicarbonate) solution, pH=7.8

*In-gel digest is very robust, but hydrophobic peptides may not be easily extracted*
In-Solution Digestion

• Remove impurity and dissolve in 25 mM ammonium bicarbonate (ABC)
  Make sure digest mixture is slightly basic (tryptic digest)
  Remove any inhibitors

• Protein may need to be denatured in order to get efficient digestion
  ACN: most enzymes can tolerate up to 20-30% ACN
  Urea or guanidine HCl: as much as tolerated by the enzyme:
    Trypsin and LysC works in 2M urea or 1M guanidine HCl
  Add a detergent only if necessary

• 1-10% enzyme (w/w), ~4 h, @37°C
Other Useful Info on Digestion

1. Enzymes isolated from different sources may display very different activity (Roche vs WAKO Lys-C) and have different contaminants.

2. Asp-N:
   - has not been sequenced, so you will not identify Asp-N peptides in database searching.
   - is a metallo-protease, thus chelating agents will inactivate it.

3. Covalent modifications on the target protein may slow down or prevent cleavages: Lys(Me), Lys(Me2), Lys(Ac).

4. Glycosylation may sterically hinder proteolysis:
   - N-linked sugars can be removed by peptide-N-glycosidase F (PNGase F).

5. Endoproteases are poor exoproteases (cut at ends of peptides).

6. Enzymes may act as ligases – moving a few residues from one terminus to the other → transpeptidation.

7. Proteases also tend to suffer from substrate inhibition.
Other Useful Info on Digestion

Lys-C and trypsin cleave at lysine analogs. These can be generated by beta-elimination of phosphate, glycopeptides or sulfur groups, primarily on serines. The resulting dehydroalanine can be reacted with 2-aminoethanethiol or cysteamine HCl.


Knight, Z et al. (2003) Nat Biotech. 21, 1047-1054
Digestion with Chemicals

- CNBr in 0.1M HCl, neat formic acid, or 70% TFA for 2 days
  - Cleave at C-terminal after Met
  - Works with precipitated proteins
  - Can alkylate after digest
  - Methionine becomes homoserine AND homoserine lactone

- Asn↓Gly - 2M hydroxylamine, 2M Guanidine HCl (pH 9), 45°C, 4h

- Asp↓Xxx: 10 mM HCl, 108°C, 2 h.

- Effective for membrane proteins.
  - Will work in-gel; on the blot; in a slurry
  - May access sites when protease cannot.

<table>
<thead>
<tr>
<th>Homoserine lactone</th>
<th>Homoserine</th>
</tr>
</thead>
</table>

J. Trinidad
Inhibitors

Your sample can continue to change after you isolate it.

• Proteases may still be active.
  In the MS analysis of peptides you are assuming peptides are formed by the protease you add.

• Identifying ‘non-specific’ cleaved peptides is more difficult.

• If you are interested in phosphorylation, you should add phosphatase inhibitors.
Salt and Mass Spectrometry

• Proteins and Peptides in mass spectrometry are typically analyzed in a protonated state; i.e. [M+H]+

• If metal salts are present, then metal adducts can be formed; e.g. [M+Na]+ or [M+K]⁺.
  • Having protonated and metal adducts makes the spectrum more complicated to interpret.
  • Metal adducted peptides do not fragment as readily as protonated, making identification by fragmentation analysis difficult.

• If a salt crystallizes in the electrospray capillary it can block flow, meaning it has to be replaced and sample is lost. It can also block the orifice into the mass spectrometer.

• Some salts form clusters (most notably phosphate), and these can drown out the signal from all other components in the sample.
Phosphate Clusters

\[ [\text{H}_2\text{O}+4\text{HPO}_3]^+ \]
Other Common Contaminants

• Detergents
  Co-elute with peptides
  Suppress ionization (surface tension)
  Acid cleavable (PPS Silent Surfactant)

  ![PPS Silent Surfactant](http://www.proteindiscovery.com/pages/products/ppsl.png)

• Glycerol
  Stabilizing agent for protein storage. Viscous, involatile.
  As with detergents, it changes ESI behavior
  and can prevent spray by forming a large droplet

• Trifluoroacetic acid (TFA)
  Reduce charges of peptide ions
  Forms adducts with peptide, especially with proteins
Contaminants in the Air

Polydimethylcyclosiloxane ions can be used as mass calibrants (MCP2005v4p2010)
The Needs for Protein/Peptide Fractionation

Protein Concentrations in Human Plasma
Immunodepletion of 12 Human Serum Highly Abundant Proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_1)-Acid Glycoprotein</td>
<td>Fibrinogen</td>
</tr>
<tr>
<td>(\alpha_1)-Antitrypsin</td>
<td>Haptoglobin</td>
</tr>
<tr>
<td>(\alpha_2)-Macroglobulin</td>
<td>IgA</td>
</tr>
<tr>
<td>Albumin</td>
<td>IgG</td>
</tr>
<tr>
<td>Apolipoprotein A-I</td>
<td>IgM</td>
</tr>
<tr>
<td>Apolipoprotein A-II</td>
<td>Transferrin</td>
</tr>
</tbody>
</table>
Protein Separation Methods

**Electrophoresis**
- SDS PAGE; Blue native gel, complexes *(MCP2004v8p176)*
- IEF
- Capillary electrophoresis (CE)

**Affinity Separation**
- Beads: avidin-coated beads
- Nanoparticles, goldNP *(Cheng,MCP_Oct4, 2009)*

**Physical Separation**
- MW cutoff filter, Dialysis membrane

**Protein Separation in MS**
- Ion mobility – separate by conformation *(AnalBioanalChem2008v391p905)*
  - Millipore’s Ultrafiltration membrane

**HPLC**
HPLC Systems

**Autosampler**
- Automatically load samples into the system
- Keep samples at low temperature (~10°C)

**Binary pumps**
- Deliver solvent gradient
- UPLC can deliver <100nl/min at ~10,000psi pressure

**Detector**
- MS - online
- UV is most common, 208nm for peptides
- Conductivity is useful in ion exchange

**Sample collector**
- Useful for manual 2D LC

**Online LC**
- Low flow rate: ~300nl/min with nanospray
- Reverse phase (RP) and MS compatible mobile phase
Van Deemter equation

\[ H = A + \frac{B}{u} + C \cdot u \]

- **H** = Height equivalent of theoretical plates (HETP)
  (smaller, the better separation)
- **A** = Eddy-diffusion
- **B** = Longitudinal diffusion
- **C** = mass transfer kinetics of
  the analyte between mobile and stationary phase
- **u** = Linear Velocity.

Assumption: analyte is in equilibrium between stationary and mobile phases
(it never happens!)

HETP is a measurement of the column efficiency, determined by particle
disperse flow rate etc. as shown in the figure

UPLC needs high back pressure to operate!
(Small particles and higher linear velocity)
HPLC
Separation Theory

**retention factor**, \(k'_A = \frac{(t_R - t_M)}{t_M}\), is proportional to partition constant \((K=C_m/C_s)\) between mobile and stationary phases

**selectivity**, \(\alpha\), for the separation of two species (A and B)

\[\alpha = \frac{k'_B}{k'_A}\]

**number of theoretical plates** is defined with length of column, \(L\), as

\[N = \frac{L}{\text{HETP}}\]

and can be calculated from chromatogram

\[N = \frac{5.55 t_R^2}{w_{1/2}^2}\]

**resolution** \(R\) of two species, A and B, is defined as

\[R = \frac{(t_R)_B - (t_R)_A}{[(w_{1/2})_A + (w_{1/2})_B]/2}\]

and is related to efficiency \((N)\), selectivity \((\alpha)\), and retention factor \((k'_B)\) as

\[R = \frac{\sqrt{N} \left(\frac{\alpha - 1}{\alpha}\right) k'_B}{4 \left(1 + k'_B\right)}\]

Master Resolution Equation

### Liquid Chromatography

#### Types

<table>
<thead>
<tr>
<th></th>
<th>Proteins</th>
<th>Peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse Phase (RP)</td>
<td>Low pH</td>
<td>Low pH and High pH</td>
</tr>
<tr>
<td>Ion Exchange</td>
<td>WCX and WAX</td>
<td>SCX</td>
</tr>
<tr>
<td>Size Exclusion (SEC)</td>
<td>Yes</td>
<td>-</td>
</tr>
</tbody>
</table>

Hydrophilic Interaction Chromatography (HILIC) is of weak cation exchange (WCX) type.
# HPLC Column Formats

<table>
<thead>
<tr>
<th>Column Diameter (mm)</th>
<th>Typical Flow Rate</th>
<th>Sample Capacity</th>
<th>Maximum Practical Sample Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capillary</td>
<td>0.075</td>
<td>0.25 μL/min</td>
<td>0.05 μg</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>1 μL/min</td>
<td>0.2 μg</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>5 μL/min</td>
<td>1 μg</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>10 μL/min</td>
<td>2 μg</td>
</tr>
<tr>
<td>Microbore</td>
<td>1.0</td>
<td>25–50 μL/min</td>
<td>0.05–10 μg</td>
</tr>
<tr>
<td>Narrowbore</td>
<td>2.1</td>
<td>100–300 μL/min</td>
<td>0.2–50 μg</td>
</tr>
<tr>
<td>Analytical</td>
<td>4.6</td>
<td>0.5–1.5 mL/min</td>
<td>1–200 μg</td>
</tr>
<tr>
<td>Semi-preparative</td>
<td>10</td>
<td>2.5–7.5 mL/min</td>
<td>1,000 μg</td>
</tr>
<tr>
<td>Preparative</td>
<td>22</td>
<td>10–30 mL/min</td>
<td>5 mg</td>
</tr>
<tr>
<td>Process</td>
<td>50</td>
<td>50–100 mL/min</td>
<td>25 mg</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>150–300 mL/min</td>
<td>125 mg</td>
</tr>
</tbody>
</table>

Online LCMSMS

General purpose separation
HPLC
Columns

Chromatographic Support
• Silica beads: 5μm diameter, 300Å pore size
  Can be packed in lab
• ethyl-bridged hybrid support can stand high pH
  (AnalChem2003v75p6781, Water’s Xbridge columns)
• Monolith: can be made with Sol-gel process (JSeparationSci2005v28p1628)
  Low back pressure

Cross section of a monolith column
AnalChem2000v72p1275
HPLC

Reverse Phase

• Organic solvent – Acetonitrile (CH3CN, ACN)
  Low viscosity, UV cut off at 190nm
• Peptides **online**:
  Stationary phase: C18
  Solvent A: 0.1% formic acid (FA) in water
  Solvent B: 0.1% formic acid in acetonitrile
  Gradient: 0% B – 42% B in 40mins
• Proteins **offline**:
  Stationary phase: C4
  Solvent A: 0.1% trifluoroacetic acid (TFA) in water
  Solvent B: 0.1% TFA in acetonitrile (ACN)
  Gradient: 0% B – 100% B in 60mins

Functionize silica support with C18 stationary phase
Why the ion pairing reagent is needed?

- Lower pH to 2
  All but –COOH protonated
- Reduce charge interaction
  By masking +ionic groups with –ions
- TFA is an excellent ion pairing reagent
  Flurocompounds are hydrophobic
  TFA is very acidic (pKa, 0.3)
  But it forms adducts with peptide ions
- If TFA is not an option, use formic acid or acetic acid
  Formic acid (pKa, 3.75) works better for peptide separation than acetic acid (pKa, 4.75)

<table>
<thead>
<tr>
<th>Group on AA</th>
<th>pKa</th>
</tr>
</thead>
<tbody>
<tr>
<td>carboxylates</td>
<td>2 to 4</td>
</tr>
<tr>
<td>thiols</td>
<td>~ 8</td>
</tr>
<tr>
<td>amines</td>
<td>6 to 11</td>
</tr>
<tr>
<td>guanidines</td>
<td>12 to 13</td>
</tr>
</tbody>
</table>

It is historical reason to have an ion pairing reagent, because silica beads is not stable when pH>8.
LC Retention Time Prediction

Quantitative structure-retention relationships (QSRR) (ChemRev2007v107p3212)
\[ t_r = (t_r \text{ for AAs}) + (t_r \text{ from Van der Waals volume}) + (t_r \text{ from partition}) \]

Relative hydrophobicity (MCP2004v3p908)
Retention coefficients of amino acid residues
http://hs2.proteome.ca/SSRCalc/SSRCalc.html

Use known data to train the network
Stationary phase

- **Strong Cation Exchange (SCX)**
  PolySULFOETHYL Aspartamide
  Salt gradient, pH=3
  Used as 1\textsuperscript{st} dimension of 2DLC

- **Hydrophilic Interaction Chromatography (HILIC)**
  WCX (PolyCAT A): Poly Aspartic Acid
  Salt or pH gradient (histone PTM variants, MCP2009v8p2266)
HPLC Columns


- Zwitterionic stationary phase (http://www.nestgrp.com/protocols/polylc/erlic/erlic.shtml)

- Excellent separation behavior according to peptide charge states over a wide pH range

- “Ideal” phosphopeptide separation (JProteomeRes2008v7p4869)
HPLC

Packing your own nanoLC columns

• Use a blank column with frit at the end or frit at the spray tip
  (http://www.newobjective.com/products/cols_index.html)

• Choose desired resin (packing material, buck media,…)
  Type: RP, SCX, HILIC, etc
  Particle size: 3µm, 5µm
  Pore size: 300Å, 1000Å
  Can buy hlic resin from:
  (http://www.nestgrp.com/prices/PolyLC.shtml#polycatb)

• Pack with a “bomb”
  10% resin + 90% MeOH in a glass vial
  Add a Teflon coated stir bar
  Increase N2 pressure while stirring

A 100µm X 10cm column only uses ~3 mg of resin!
Multi-Dimensional Liquid Chromatography

*Case Studies*

• LCs at different dimensions should be **orthogonal**
  Separation based on different properties

• SCX-RP
  General proteomics strategy with tryptic digest
  Lys-N digest may be better

• High pH RP - low pH RP

• RP – HILIC
  Histone analysis
MudPIT (Multidimensional Protein Identification Technology)

1) Load Acidified Digest
2) Equilibration
3) Salt Step
4) Wash
5) RP Gradient
6) Go to # 2

SCX: Partisphere

50 μm Split Capillary

50 μm Fused Silica Capillary from HPLC

PEEK Micro Cross

Gold Lead for Electrical Contact with Liquid (1.8 kV)

100 μm Packed Capillary Column

SCX

RP

Capillary Opening into Mass Spectrometer
2D LC Identification of Phosphopeptides
SCX Separation of Tryptic Peptides

Column: 5X115-mm polysulfoethyl A
Gradient: 0 to 350 mM KCl in 30% acetonitrile, 5 mM KH2PO4, pH 2.7 in 90mins
Lys-N vs Trypsin Digestion

1. Lys-N only incorporates a single O into peptide  
   Good for 18O labeling (MCP2005v4p1550)  
   Trypsin produces a mixture of two oxygen atoms  
   (Electrophoresis1996v17p945)

2. Charge distributions are different  
   MSMS fragmentation patterns

3. Lys-N cleaves –X↓KP- and trypsin does not cleave –KP-
SCX Separation of Lys-N Peptides

(B)
CID

Number of unique peptides

1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49

I. II. III. IV.

N-Acetylated
3 Basic residues
Phosphorylated
4 Basic residues
Single lysine
5 Basic residues
2 Basic residues
6 Basic residues

MolCellProteomics2009v8p190
SCX Separation of Lys-N Phosphopeptides
Low pH RPLC vs High pH RPLC

2DLC based on High pH RPLC and Low pH RPLC
LC Separations for Histone Analysis

Acid Extracted Core Histone from 293 Cells

For online HILIC with volatile salt and pH gradient, see (MCP2009v8p2266)
Microfluidics

Small channel dimensions
>1mm, Laminar flow
High surface to volume ratio

Device integration
Separation channels
  Packed or CE
Reactors
Mixers
Valves

Base materials
Polydimethylsiloxane (PDMS)
Glass, quartz
Silicon
Polycarbonate

Fabrication techniques
molding/bonding
NaOH etching
Plasma etching
Injection molding

Integrated circuits → Microelectromechanical systems (MEMS)
→ Microfluidics → Biological Applications

Nature2006v442p374, p394
Microfluidics
*Interface with MS*

**Nanomate**

- ESI spary tip array
- Si deep etching

- Automated sample loading for infusion
- LC sample collection and reanalysis
- Top-down
- System optimization

www.advion.com
Microfluidics in Proteomics

**Examples**

*Immobilized microfluidic enzymatic reactors (IMERs)*

- 10s digestion time
- Reduction of autolysis
- Easy separation interface
- Utilize monolithic material
  - High surface area
  - High mechanical strength
- Integrate nano-materials
  - High and homogeneous surface

*Glass chip for electrophoretic separation and ionization of peptides*
(AnalChem2008v80p6881)

BSA tryptic digest

Electrophoresis2004v25p3550
AnalChimActa2009v632p1
Microfluidics
Integrated Enrichment and NanoLC System

**Agilent HPLC-chip**

Functionality
- Graphitized carbon column for oligo-saccharide separation
- TiO2 column for phosphopeptide enrichment
- Trap column
- NanoRP analytical LC column
- Spray tip

Fabrication
- Polyimide
- Inkjet printhead processes
Microfluidics

Free-Flow Zone Electrophoresis

Zone
- Const. pH
- Const. E

IEF
- pH gradient
- Const. E

Isotachophoresis
- Two buffers
- Const. E

Field step
- Const. pH
- Field gradient

Different buffer
Microfluidics
Free-Flow Zone Electrophoresis

Fraction from cathodic side

Fraction from middle outlet

Fraction from anodic side

AnalChem2010_Song
Increase Charge States in ESI

carbonic anhydrase II  
(+ 0.1\% m-NBA)

phosphorylase b  
(+ 1.0\% m-NBA + 0.1\% m-NBA)
Soft-landing of Mass Selected Ions

Source protein  PKAc (MW 40,856)  Landed protein

m/z (Th)

2800  3000  3200  3400  3600  3800

+13

3146.9

+13

3140.9

2934.1  3413.0

2921.0  3404.2

+14

+12
Direct MS Analysis of Intact Protein Complex

• adenosine 5´-triphosphate (ATP)–binding cassette transporter (BtuC<sub>2</sub>D<sub>2</sub>)
• nanoelectrospray of membrane protein complex in a micellar solution
• CID of protein complex into components
• Found α-N-Dgluconyl-His tag as a modification on BtC
Conclusions

• 1D SDS gel – in gel digest is the most commonly used method

• 2D LC is widely used for shotgun proteomics

• Tryptic peptides are compatible with C18 RPLC and CID

• AspN and LysC are common alternatives; LysN becomes more popular

• Microfluidics allows for automation and fast analysis
Amino Acids

<table>
<thead>
<tr>
<th></th>
<th>Acronym</th>
<th>Letter</th>
<th>C</th>
<th>H</th>
<th>N</th>
<th>O</th>
<th>S</th>
<th>Monoisotopic Mass (Da)</th>
<th>Occurrence(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>57.02146</td>
<td>6.89</td>
</tr>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>71.03711</td>
<td>7.36</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>87.03203</td>
<td>7.35</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>97.05276</td>
<td>5.01</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>99.06841</td>
<td>6.46</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>101.04768</td>
<td>5.92</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>103.00918</td>
<td>1.77</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>6</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>113.08406</td>
<td>5.69</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>6</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>113.08406</td>
<td>9.25</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>114.04293</td>
<td>4.59</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>Asp</td>
<td>D</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>115.02694</td>
<td>5.15</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
<td>5</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>128.05858</td>
<td>4.04</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>6</td>
<td>12</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>128.09496</td>
<td>5.8</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>Glu</td>
<td>E</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>129.04259</td>
<td>6.18</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>131.04048</td>
<td>2.32</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>137.05891</td>
<td>2.25</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
<td>9</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>147.06841</td>
<td>4.08</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>6</td>
<td>12</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>156.10111</td>
<td>5.18</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
<td>9</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>163.06333</td>
<td>3.24</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
<td>11</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>186.07931</td>
<td>1.34</td>
</tr>
</tbody>
</table>