Application of mass spectrometry to the analysis and identification of peptides, proteins and other biological molecules

Matthew Renfrow, PhD
6-4681
renfrow@uab.edu

S. Barnes & M. Renfrow-UAB 1/12/11
Overview

• MALDI-TOF MS
  – Peptide mass fingerprinting

• Electrospray MS
  – Analysis of intact proteins
  – Molecular weight calculations
  – Max Entropy for MW estimation

• Peptide analysis
  – Purity - ESI-MS is a revelation

• Integration of MS with LC and CE
  – Multidimensional LC of peptides

• Tandem MS
  – Identifying peptide amino acid sequences

S. Barnes & M. Renfrow-UAB 1/12/11
Peptide = chain of amino acids

N terminus

polypeptide chain

C terminus
N O N P O L A R
Glycine (Gly)  Alanine (Ala)  Valine (Val)  Leucine (Leu)  Isoleucine (Ile)  Methionine (Met)  Tryptophan (Trp)  Phenylalanine (Phe)  Proline (Pro)

P O L A R
Serine (Ser)  Threonine (Thr)  Cysteine (Cys)  Tyrosine (Tyr)  Asparagine (Asn)  Glutamine (Gln)

E l e c t r i c a l l y C h a r g e d

A c i d i c
Aspartic Acid (Asp)  Glutamic Acid (Glu)

B a s i c
Lysine (Lys)  Arginine (Arg)  Histidine (His)
Benefit of removing salt from tryptic digest

Salt Contamination

Trypsin autolysis peak

ZipTip (desalting)

S. Barnes & M. Renfrow-UAB 1/12/11
Factors from conventional experiments that impact MALDI-TOF analysis

• Tolerance of buffers/chemicals used in sample preparation
  – NaCl up to 150 mM
  – Urea up to 2-3 M (carbamoylation can occur!)
  – Guanidinium-HCl up to 2 M

• Tolerance of detergents
  – SDS up to 0.05%

• Staining Protocols
  – Whole proteins form adducts with Coomassie
  – Silver staining modifies selected peptides

S. Barnes & M. Renfrow-UAB 1/12/11
Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS)

• Advantages of MALDI-TOF
  – More tolerant to common buffers than ESI, but…
  – High degree of sensitivity, moderate mass accuracy, and mass resolution
  – High mass compounds, i.e. proteins, PEG…

• Common Applications of MALDI-TOF
  – Masses of large proteins and other compounds
  – Enzymatic digestion profiles of proteins to establish their identity
  – Peptide sequencing (TOF-TOF)
  – In situ protein/peptide imaging

S. Barnes & M. Renfrow-UAB 1/12/11
Schematic of Matrix Assisted Laser Desorption/Ionization time-of-flight mass spectrometry.
Electrospray ionization

• ESI-MS is very sensitive to the presence of electrolyte species -
  – these ionize more easily than solutes and may also form adducts with solutes

• In ESI-MS, multiple charge states are possible
  – These lead to more accurate MWs

• This is a softer ionization than MALDI where the UV laser at 337 nm alters the chemistry of modifications such as Tyr-NO₂ and Cys-SNO
Guarantees of purity based on observation of “a single peak by reverse-phase HPLC” and by “it gave the correct sequence when analyzed by Edman degradation” are hollow. The lower spectrum was of a “pure” HPLC peak. The method of purification was amended and the upper spectrum was obtained.

S. Barnes & M. Renfrow-UAB 1/12/11
Ionizing proteins and peptides

- $^{+\text{H}_3\text{NCHR}_1\text{CO(NHCHR}_n\text{CO))}_n\text{NHCHR}_2\text{COOH}$ is the ion that’s found in dilute acid solution.

- If there are internal basic residues, then the ions will be of the form $[\text{M+nH}]^{n^+}$, where $n = 1, 2, \text{etc.}$

- A tryptic peptide will have a N-terminal amino group and an amino group from Arg or Lys.
  - If the peptide has a mol. wt. of 1000 Da, then the singly charged ion will have a $m/z$ of 1001, whereas the doubly charged ion has a $m/z$ of 501.
Peptide = chain of amino acids

N terminus

polypeptide chain

C terminus

S. Barnes & M. Renfrow-UAB 1/12/11
Nonpolar:
- Glycine (Gly)
- Alanine (Ala)
- Valine (Val)
- Leucine (Leu)
- Isoleucine (Ile)
- Methionine (Met)
- Tryptophan (Trp)
- Phenylalanine (Phe)
- Proline (Pro)

Polar:
- Serine (Ser)
- Threonine (Thr)
- Cysteine (Cys)
- Tyrosine (Tyr)
- Asparagine (Asn)
- Glutamine (Gln)

Electrically Charged:
- Aspartic Acid (Asp)
- Glutamic Acid (Glu)
- Lysine (Lys)
- Arginine (Arg)
- Histidine (His)

Dept. Biol. Penn State ©2002
Peptide mass fingerprinting

• This method was developed because of the availability of predicted protein sequences from genome sequencing

• Proteins did not have to have been previously sequenced - only that the open reading frame in the gene is known - the rest is a virtual exercise in the hands of statisticians, bioinformaticists and computers

• However, remember the matching is only as good as the database content - this can change
Protein analysis by MALDI 2010

Eppendorf tube → destain → Speed-Vac

Water Bath 37°C → Incubate overnight → trypsin 1:20

Peptide extraction → MALDI plate

Desalting Ziptip

Mass (m/z)

S. Barnes & M. Renfrow-UAB 1/12/11
Proteolytic enzymes used to hydrolyze proteins

The choice of enzyme largely depends on the nature of the amino acid sequence and the specific issue that is being addressed

- Trypsin - *cleaves at arginine and lysine residues*
- Chymotrypsin - *cleaves hydrophobic residues*
- Arg-C - *cleaves at arginine residues*
- Glu-C - *cleaves at aspartate/glutamic acid residues*
- Lys-C - *cleaves at lysine residues*
- V8-protease - *cleaves at glutamic acid residues*
- Pepsin - *cleaves randomly but consistently, at acid pH*

See [http://www.abrf.org/JBT/1998/September98/sep98m_r.html](http://www.abrf.org/JBT/1998/September98/sep98m_r.html)

S. Barnes & M. Renfrow-UAB 1/12/11
Searching databases with peptide masses to identify proteins

Best site is at www.matrixscience.com

The program (MASCOT) can search the OWL or NCBI databases using a set of tryptic peptide masses, or the fragment ions (specified or unspecified) of peptides

Presents the expected set of tryptic peptides for each matched protein

S. Barnes & M. Renfrow-UAB 1/12/11
Choice of peptidase

• Analogous to DNA restriction enzymes

• Tryptic peptide fingerprinting may identify, not one, but several highly related protein candidates (e.g., actins)

• Inspection of the sequences may reveal that there is a difference at one residue that distinguishes between two candidates.

• If for instance it is a glutamate, then use of Glu-C or V8-protease may enable the two proteins to be correctly identified

• INSPECT sequences carefully
Sequence of β-lactoglobulin

MKCLLLALAL TCGAQLIVT QTMKGLDIQK
VAGTWWYSLAM AASDISLLDA QSAPLRVYVE
ELKPTPEGDL EILLQKWENG ECAQKKIIAE
KTKIPAVFKI DALNENKVLV LDTDYKKYLL
FCMENSAEPE QSLACQCLVR TPEVDDEALE
KFDKALKALP MHIRLSSFNPT QLEEQCHI

S. Barnes & M. Renfrow-UAB 1/12/11
Peptides from digestion with Glu-C

MKCLLLALALALTCAQALIVTQTMKGLD
IQKVAGTWYSLEMAASD ISLLD AQSAPLRVYVE
E LKPTPE GD LE ILLQKWE NGE CAQKKIIAE
KTKIPAVFKID ALNE NKVLVLD TD YKKYLLFCME
NSAE PE QSLACQCLVVRTPE VD D E ALE KFD
KALKALPMHIRLSFNPTQLE E QCHI

S. Barnes & M. Renfrow-UAB 1/12/11
# Amino acid residue masses

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Mass</th>
<th>Amino Acid</th>
<th>Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>71.037</td>
<td>Leucine</td>
<td>113.084</td>
</tr>
<tr>
<td>Arginine</td>
<td>156.101</td>
<td>Lysine</td>
<td>128.094</td>
</tr>
<tr>
<td>Asparagine</td>
<td>114.043</td>
<td>Methionine</td>
<td>131.040</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>115.027</td>
<td>Phenylalanine</td>
<td>147.068</td>
</tr>
<tr>
<td>Cysteine</td>
<td>103.009</td>
<td>Proline</td>
<td>97.053</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>129.043</td>
<td>Serine</td>
<td>87.032</td>
</tr>
<tr>
<td>Glutamine</td>
<td>128.058</td>
<td>Threonine</td>
<td>101.048</td>
</tr>
<tr>
<td>Glycine</td>
<td>57.021</td>
<td>Tryptophan</td>
<td>186.079</td>
</tr>
<tr>
<td>Histidine</td>
<td>137.059</td>
<td>Tyrosine</td>
<td>163.063</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>113.084</td>
<td>Valine</td>
<td>99.068</td>
</tr>
</tbody>
</table>

The m/z value of a peptide [M+H]^+ is the sum of the residue masses plus 18.015 for H₂O plus 1.008. So, what is it for ISLLD?

\[
113.084 + 87.032 + 113.084 + 113.084 + 115.027 + 18.015 + 1.008 = 560.334
\]

S. Barnes & M. Renfrow-UAB 1/12/11
**Expected peptides from trypsin and Glu-C digestion of bovine β-lactoglobulin**

| 837.4764  | 800.4876  |
| 916.4734  | 929.5455  |
| 1064.4466 | 1003.5605 |
| 1065.5827 | 1232.6634 |
| 1245.5845 | 1259.7722 |
| 1658.7843 | 1337.6632 |
| 2275.2586 | 1447.7032 |
| 2313.2588 | 1811.8996 |
| 2647.2023 | 2307.3006 |
| 2707.3760 | 2819.5265 |

Assumes all cuts are complete, there is no oxidation of Met residues, and Cys residues are unmodified

S. Barnes & M. Renfrow-UAB 1/12/11
**MASCOT Peptide Mass Fingerprint**

<table>
<thead>
<tr>
<th><strong>Your name</strong></th>
<th>Stephen Barnes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Email</strong></td>
<td><a href="mailto:sbarnes.uab@gmail.com">sbarnes.uab@gmail.com</a></td>
</tr>
<tr>
<td><strong>Search title</strong></td>
<td>beta globulin test</td>
</tr>
<tr>
<td><strong>Database</strong></td>
<td>NCBIrr</td>
</tr>
<tr>
<td><strong>Taxonomy</strong></td>
<td>All entries</td>
</tr>
<tr>
<td><strong>Enzyme</strong></td>
<td>Trypsin</td>
</tr>
<tr>
<td><strong>Allow up to</strong></td>
<td>0 missed cleavages</td>
</tr>
</tbody>
</table>
| **Fixed modifications** | Acetyl (K)  
Acetyl (N-term)  
Acetyl (Protein N-term)  
Aminated (Protein N-term)  
Aminated (C-term) |
| **Variable modifications** | Acetyl (K)  
Acetyl (N-term)  
Aminated (Protein N-term)  
Aminated (C-term) |
| **Protein mass** | kDa |
| **Peptide tol. ±** | 1.0 Da |
| **Mass values** | MH⁺ / Mᵦ / M-H⁻ |
| **Monoisotopic** | Average |
| **Data file** | Choose File |

Enter the ions here in this box

<table>
<thead>
<tr>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>837.4764</td>
</tr>
<tr>
<td>916.4734</td>
</tr>
<tr>
<td>1064.4466</td>
</tr>
<tr>
<td>1065.5827</td>
</tr>
<tr>
<td>1245.5845</td>
</tr>
<tr>
<td>1658.7843</td>
</tr>
</tbody>
</table>
Mascot Search Results

User: Stephen Barnes
Email: sbarnes.uab@gmail.com
Search title: beta globulin test
Database: NCBInr 20061230 (4378862 sequences; 1508892933 residues)
Timestamp: 1 Jan 2007 at 02:17:51 GMT
Top Score: 210 for gi|87196497, lactoglobulin, beta [Bos taurus]

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event. Protein scores greater than 79 are significant (p<0.05).

Protein Summary Report

<table>
<thead>
<tr>
<th>Format As</th>
<th>Protein Summary</th>
<th>Help</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Significance threshold p&lt; 0.05</td>
<td>Max. number of hits 20</td>
</tr>
</tbody>
</table>
Protein records provided by MASCOT search

<table>
<thead>
<tr>
<th>Accession</th>
<th>Mass</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>87196497</td>
<td>19870</td>
<td>210 lactoglobulin, beta [Bos taurus]</td>
</tr>
<tr>
<td>gi</td>
<td>4388846</td>
<td>18269</td>
<td>179 Chain, Bovine Beta-Lactoglobulin Complexed With Palmitate, Lattice Z</td>
</tr>
<tr>
<td>gi</td>
<td>223780</td>
<td>18165</td>
<td>152 lactoglobulin beta</td>
</tr>
<tr>
<td>gi</td>
<td>72079</td>
<td>18255</td>
<td>152 beta-lactoglobulin - water buffalo</td>
</tr>
<tr>
<td>gi</td>
<td>520</td>
<td>19908</td>
<td>150 beta-lactoglobulin [Bos taurus]</td>
</tr>
<tr>
<td>gi</td>
<td>20178290</td>
<td>20010</td>
<td>148 Beta-lactoglobulin precursor (Beta-LG)</td>
</tr>
<tr>
<td>gi</td>
<td>165839</td>
<td>19934</td>
<td>148 beta-lactoglobulin</td>
</tr>
<tr>
<td>gi</td>
<td>2194088</td>
<td>18297</td>
<td>147 Chain A, Bovine Beta-Lactoglobulin, Lattice X</td>
</tr>
<tr>
<td>gi</td>
<td>110612608</td>
<td>19891</td>
<td>144 beta-lactoglobulin [Bubalus bubalis]</td>
</tr>
<tr>
<td>gi</td>
<td>162748</td>
<td>17156</td>
<td>126 beta-lactoglobulin</td>
</tr>
<tr>
<td>gi</td>
<td>125912</td>
<td>19962</td>
<td>125 Beta-lactoglobulin precursor (Beta-LG)</td>
</tr>
<tr>
<td>gi</td>
<td>7245834</td>
<td>18363</td>
<td>124 Chain A, Structural Changes Accompanying Ph-Induced Dissociation Of The</td>
</tr>
<tr>
<td>gi</td>
<td>229460</td>
<td>18355</td>
<td>124 lactoglobulin beta</td>
</tr>
<tr>
<td>gi</td>
<td>4388939</td>
<td>18355</td>
<td>124 Chain, Structural Basis Of The Tanford Transition Of Bovine Beta-Lacto</td>
</tr>
<tr>
<td>gi</td>
<td>49259423</td>
<td>18339</td>
<td>124 Chain X, The Cys121ser Mutant Of Beta-Lactoglobulin</td>
</tr>
<tr>
<td>gi</td>
<td>54037712</td>
<td>18139</td>
<td>120 Beta-lactoglobulin (Beta-LG)</td>
</tr>
<tr>
<td>gi</td>
<td>57164367</td>
<td>19908</td>
<td>117 beta-lactoglobulin [Ovis aries]</td>
</tr>
<tr>
<td>gi</td>
<td>90108547</td>
<td>18264</td>
<td>82 Chain A, Reindeer Beta-Lactoglobulin</td>
</tr>
<tr>
<td>gi</td>
<td>71980384</td>
<td>20035</td>
<td>80 beta-lactoglobulin [Rangifer tarandus tarandus]</td>
</tr>
<tr>
<td>gi</td>
<td>26352113</td>
<td>13020</td>
<td>60 unnamed protein product [Mus musculus]</td>
</tr>
</tbody>
</table>

S. Barnes & M. Renfrow-UAB 1/12/11
Comparison of observed and predicted tryptic peptides

<table>
<thead>
<tr>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>Delta</th>
<th>Start</th>
<th>End</th>
<th>Miss</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>837.4764</td>
<td>836.4691</td>
<td>836.4691</td>
<td>0.0001</td>
<td>158</td>
<td>164</td>
<td>0</td>
<td>K.ALPMHIL.R</td>
</tr>
<tr>
<td>916.4734</td>
<td>915.4661</td>
<td>915.4661</td>
<td>-0.0000</td>
<td>100</td>
<td>107</td>
<td>0</td>
<td>K.IDALNENK.V</td>
</tr>
<tr>
<td>1064.4466</td>
<td>1063.4393</td>
<td>1063.4393</td>
<td>0.0001</td>
<td>77</td>
<td>85</td>
<td>0</td>
<td>K.WENGECAQK.K</td>
</tr>
<tr>
<td>1065.5827</td>
<td>1064.5754</td>
<td>1064.5753</td>
<td>0.0001</td>
<td>108</td>
<td>116</td>
<td>0</td>
<td>K.VLVALTDYK.K</td>
</tr>
<tr>
<td>1245.5845</td>
<td>1244.5772</td>
<td>1244.5772</td>
<td>0.0000</td>
<td>141</td>
<td>151</td>
<td>0</td>
<td>R.TPEVDDEALEK.F</td>
</tr>
<tr>
<td>1658.7843</td>
<td>1657.7770</td>
<td>1657.7770</td>
<td>0.0000</td>
<td>165</td>
<td>178</td>
<td>0</td>
<td>R.LSFNPQDEEQCHI.-</td>
</tr>
<tr>
<td>2275.2586</td>
<td>2274.2513</td>
<td>2274.2513</td>
<td>0.0000</td>
<td>3</td>
<td>24</td>
<td>0</td>
<td>K.CLLLALALTCAQALIVTQTMK.G</td>
</tr>
<tr>
<td>2313.2588</td>
<td>2312.2515</td>
<td>2312.2515</td>
<td>0.0001</td>
<td>57</td>
<td>76</td>
<td>0</td>
<td>R.VYVEELKPTPEGDEILLQK.W</td>
</tr>
<tr>
<td>2647.2023</td>
<td>2646.1950</td>
<td>2646.1950</td>
<td>0.0001</td>
<td>118</td>
<td>140</td>
<td>0</td>
<td>K.YLLFCMENSAEPEQSLACQCLVR.T</td>
</tr>
<tr>
<td>2707.3760</td>
<td>2706.3687</td>
<td>2706.3686</td>
<td>0.0001</td>
<td>31</td>
<td>56</td>
<td>0</td>
<td>K.VAGTWYSLAMAASDISLLDAQSAPLR.V</td>
</tr>
</tbody>
</table>
Search against SwissProt database

<table>
<thead>
<tr>
<th>Accession</th>
<th>Mass</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LACB_BOVIN</td>
<td>19870</td>
<td>271</td>
<td>Beta-lactoglobulin precursor (Beta-LG) (Allergen Bos d 5) - Bos taurus (Bovine)</td>
</tr>
<tr>
<td>LACB_BUBBU</td>
<td>20010</td>
<td>197</td>
<td>Beta-lactoglobulin precursor (Beta-LG) - Bubalus bubalis (Domestic water buffalo)</td>
</tr>
<tr>
<td>LACB_CAPHI</td>
<td>19962</td>
<td>135</td>
<td>Beta-lactoglobulin precursor (Beta-LG) - Capra hircus (Goat)</td>
</tr>
<tr>
<td>LACB_OVIMU</td>
<td>18139</td>
<td>104</td>
<td>Beta-lactoglobulin (Beta-LG) - Ovis orientalis musimon (Mouflon)</td>
</tr>
<tr>
<td>LACB_SHEEP</td>
<td>19908</td>
<td>102</td>
<td>Beta-lactoglobulin-1/3 precursor (Beta-LG) - Ovis aries (Sheep)</td>
</tr>
<tr>
<td>YKH5_YEAST</td>
<td>52055</td>
<td>41</td>
<td>Hypothetical 52.1 kDa protein in SMY1-MUD2 intergenic region - Saccharomyces cerevisiae</td>
</tr>
<tr>
<td>TCPE_VIBCH</td>
<td>16247</td>
<td>40</td>
<td>Toxin coregulated pilus biosynthesis protein H (TCP pilus biosynthesis protein tcpH)</td>
</tr>
<tr>
<td>POLG_DEN22</td>
<td>54291</td>
<td>39</td>
<td>Genome polyprotein [Contains: Envelope protein E] (Fragment) - Dengue virus type 2</td>
</tr>
<tr>
<td>PFP1_PYRKO</td>
<td>18404</td>
<td>38</td>
<td>Intracellular protease 1 (EC 3.2.---) (Intracellular protease 1) - Pyrococcus koda</td>
</tr>
<tr>
<td>VIRB9_AGRT5</td>
<td>32181</td>
<td>38</td>
<td>Protein virB9 precursor - Agrobacterium tumefaciens (strain C58 / ATCC 33970)</td>
</tr>
<tr>
<td>VPS11_YEAST</td>
<td>32001</td>
<td>36</td>
<td>Vacuolar protein sorting-associated protein 71 (SWR complex protein 6) - Saccharomyces</td>
</tr>
<tr>
<td>YHR7_YEAST</td>
<td>71812</td>
<td>34</td>
<td>TPR repeat-containing protein YHR117W - Saccharomyces cerevisiae (Baker's yeast)</td>
</tr>
<tr>
<td>LGB_LOTJA</td>
<td>15745</td>
<td>34</td>
<td>Leghemoglobin - Lotus japonicus</td>
</tr>
<tr>
<td>YH13_VACCV</td>
<td>13615</td>
<td>34</td>
<td>Hypothetical 13.6 kDa HindIII-C protein - Vaccinia virus (strain Western Reserve /</td>
</tr>
<tr>
<td>MKT1_YEAST</td>
<td>94415</td>
<td>34</td>
<td>Protein MKT1 - Saccharomyces cerevisiae (Baker's yeast)</td>
</tr>
<tr>
<td>H2A1_LILLO</td>
<td>12165</td>
<td>34</td>
<td>Histone H2A.1 (GcH2A) - Lilium longiflorum (Trumpet lily)</td>
</tr>
<tr>
<td>RS12_BRUA2</td>
<td>13863</td>
<td>33</td>
<td>30S ribosomal protein S12 - Brucella abortus (strain 2308)</td>
</tr>
<tr>
<td>RS12_BRUAB</td>
<td>13863</td>
<td>33</td>
<td>30S ribosomal protein S12 - Brucella abortus</td>
</tr>
<tr>
<td>RS12_BRUME</td>
<td>13863</td>
<td>33</td>
<td>30S ribosomal protein S12 - Brucella melitensis</td>
</tr>
<tr>
<td>RS12_BRUSU</td>
<td>13863</td>
<td>33</td>
<td>30S ribosomal protein S12 - Brucella suis</td>
</tr>
</tbody>
</table>
Things to consider when doing peptide mass fingerprinting

- Proteins can be oxidized both biologically (real data) and during the workup.
- Treat the protein or the peptide digest with a reagent that reacts with Cys sulphydryl groups - e.g., iodoacetamide, iodoacetic acid, N-ethylmaleimide or 4-vinylpyrididine. Cysteines may also have reacted with acrylamide in the gel.
- Set the options in the fixed or variable modification boxes before searching.
- Allow for at least one missed cleavage - trypsin does not cut when Lys or Arg are followed by a Pro residue.

S. Barnes & M. Renfrow-UAB 1/12/11
Other web sites for peptide analysis

- [http://prowl.rockefeller.edu/](http://prowl.rockefeller.edu/)
  - Choose ProFound

- [http://prospector.ucsf.edu/](http://prospector.ucsf.edu/)
  - Choose MS-fit
Further information on identified protein

• Take the protein identifier number:
  – For bovine β-lactoglobulin it is gi|520
  – Under protein, paste in the gi number
  – A link to the protein will appear
  – Click on Blink - this is similar to BLAST, but better
  – Scroll down the list and select 1CJ5
  – Click on image of protein structure
  – To view a 3D-image of the protein, first download Cn3D from the NCBI site or RasMol
    • Bring a picture of beta-lactoglobulin to the next class
Blurring of protein space

- Identification using MALDI-TOF with MASCOT depends on:
  - Number of peptides recognized as being part of the protein
  - The mass accuracy of the peptides that are recognized
  - Pre-2000, an accuracy of better than 0.05 Da in a 1000 Da peptide (i.e., 50 ppm) was sufficient to distinguish the unknown protein from the other proteins in the databases at that time
  - Now, the protein information space has become more dense and MALDI-TOF is no longer adequate
  - Previously identified proteins may not be correct
ESI mass spectrum of ribonuclease

Cumulative MW estimate = 13,680.29

SD = 2.94

<table>
<thead>
<tr>
<th>Peak (m/z)</th>
<th>Intensity</th>
<th>Charge (est.)</th>
<th>Mol. Wt. (Est.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>978.00</td>
<td>7,778</td>
<td>14.00000</td>
<td>13,677.89</td>
</tr>
<tr>
<td>1,053.00</td>
<td>18,532</td>
<td>13.02656</td>
<td>13,675.90</td>
</tr>
<tr>
<td>1,141.00</td>
<td>59,087</td>
<td>11.95446</td>
<td>13,679.91</td>
</tr>
<tr>
<td>1,245.00</td>
<td>33,275</td>
<td>10.96146</td>
<td>13,683.91</td>
</tr>
<tr>
<td>1,369.00</td>
<td>32,390</td>
<td>10.03219</td>
<td>13,679.92</td>
</tr>
<tr>
<td>1,521.00</td>
<td>35,668</td>
<td>8.99995</td>
<td>13,679.93</td>
</tr>
<tr>
<td>1,711.00</td>
<td>16,624</td>
<td>7.99996</td>
<td>13,679.94</td>
</tr>
<tr>
<td>1,956.00</td>
<td>3,333</td>
<td>6.97955</td>
<td>13,684.94</td>
</tr>
</tbody>
</table>

S. Barnes & M. Renfrow-UAB 1/12/11
Calculation of molecular weights and ion states

- For two ions in a series for a peptide of molecular weight M, the lower \( m/z \) value (x) will be for the n+1 ion state and the larger \( m/z \) value (y) will be for the n+ ion state.
  - (1) \( \frac{(M+n)}{n} = y \)
  - (2) \( \frac{(M+n+1)}{(n+1)} = x \)

- Hence
  - (3) \( M+n = ny \) and \( M = ny-n \)
  - (4) \( M+n+1 = (n+1)x \) and \( M = (n+1)x-(n+1) \)

- Hence
  - \( ny-n = (n+1)x - (n+1) \)
  - \( ny-n-xn+n = x-1 \)
  - \( n(y-x) = x-1 \)
  - \( n = \frac{(x-1)}{(y-x)} \)

- The value of n can then be substituted in equation (1) to obtain the molecular weight of the peptide.

S. Barnes & M. Renfrow-UAB 1/12/11
Aprotinin
Lysozyme
Myoglobin

S. Barnes & M. Renfrow-UAB 1/12/11
Myoglobin
Probing of Mammalian Cells

A

B

H2B

H2B.A

H2B.K/Q

H2B.F

H2B.B

H2B.B+1Ac

H4

H4K20

Me2

H2A-1

H2A.Q

H2A.O

H2A.O+1Ac

H2A-2

H2A,C,D,I,N,P

H2A.L

H2A.C*+1Ac

H2A.A

H3.1

H3.2

H3.3

Time (minutes)

m/z

Kelleher, Northwestern Univ.
http://groups.molbiosci.northwestern.edu/kelleher/index.html
Deconvolution of MS data

• When several proteins are present, then their multiply charged ion clusters overlap

• Can this be overcome? - yes, use the MaxEntropy program provided by Micromass
Deconvolution of oxidized forms of $\beta$-lactoglobulin

S. Barnes & M. Renfrow-UAB 1/12/11

Junlong Shao
Each ion is $[\text{M}+n\text{H}]^{n+}$

For 50+ charge state of a 50 kDa protein,

$$m/z = \frac{[50,000 + 50]}{50} = 1,001$$
MaxEnt deconvolution of MWs

- 57,195.2 ± 12.9 Da
  E. coli GRoEL
- 50,011.3 ± 18.3 Da
  6xHis-tag BAT
- 15,157.6 ± 0.9 Da

S. Barnes & M. Renfrow-UAB 1/12/11

Courtesy of Mindan Sfakianos
ESI-MS of 4HNE-Modified Cytochrome C

Native Cytochrome C

Cytochrome C + One 4HNE Michael Addition (+156)

Cytochrome C + One 4HNE Schiff Base (+138)

Cytochrome C + Two 4HNE Michael Addition (+312)

Cytochrome C + Three 4HNE Michael Addition (+467)

S. Barnes & M. Renfrow-UAB 1/12/11  Courtesy of Amanda Isom (d. 2005)
Summary of determining MW by ESI

• The multiple charge states of a protein allow:
  – Mol Wt of large proteins to be estimated
  – It’s a super SDS-PAGE gel
• Important to remember that the protein sample must be free of salt
  – Typically, a sample is cleaned up on a short reverse-phase column prior to electrospray
  – Alternative, use ammonium acetate as buffer
Studying high molecular weight complexes by ESI

- Most instrument ESI interfaces have a limited $m/z$ range - up to 3,000
- In protein complexes water, and hence H+ ions, is “squeezed” out, thereby substantially increasing observed $m/z$ values
- Interfaces that pass ions with $m/z$ values above 10,000 have been designed
nanoESI-MS of HMW complexes of small heat shock proteins

Note the large \( m/z \) values (6,000-7,000) for the observed ions.

The ESI data were deconvoluted to reveal the distribution of the masses of the complexes.

Sobbott et al., J Biol Chem 277:38921

S. Barnes & M. Renfrow-UAB 1/12/11
Studying intact membrane protein complexes by gas-phase mass spectrometry (top-down analysis)

- Electrosprayed in neutral detergent
  - Dodecyl maltoside (DDM)
- Carried out on Waters Qtof II with modification of the ESI interface
- Requires high voltage (Δ200 V) to be applied across the interface and collision cell

S. Barnes & M. Renfrow-UAB 1/12/11
Effect of increasing voltage on BtuC₂D₂ membrane complexes observed in nanoMS

2 µl of 5 µM solution of the complex subjected to nanoelectrospray
Use of FT-MS in ESI of proteins

• The very high resolving power of FT-MS enables a direct measure of charge state of an individual ion since each peptide or polypeptide will have several/many isotope peaks.

• The distance in Da between successive isotope peaks of a multiply charged ion is the reciprocal of the number of charges.
Bovine Serum Albumin (66 kDa)
4.7 T Actively Shielded Magnet

ESI: BSA

S. Barnes & M. Renfrow-UAB 1/12/11

Bruker Daltonics
LC-MS of peptide mixtures

Pre-column for desalting

Load sample

Analytical reverse phase column
75 μm i.d. x 15 cm

Flow rate 200 nl/min

Acetonitrile gradient

Collision gas

Q1

Q2

Electrostatic reflector

TOF detector

S. Barnes & M. Renfrow-UAB 1/12/11
The MUDPIT approach

MUlti-Dimensional Protein Identification Technology

Digest with trypsin without any protein separation

20-50 tryptic peptides per protein (100-250 peptides in this example)

Can they be resolved?

S. Barnes & M. Renfrow-UAB 1/12/11
MUDPIT - MUlti-Dimensional Protein Identification Technology

- Hydrolyze everything!
- For a cell expressing 5,000 proteins, this leads to >100,000 peptides
- Can be fractionated, but still 10,000-20,000 to differentiate
- Enormous bioinformatics problem

S. Barnes & M. Renfrow-UAB 1/12/11

John Yates
Cation exchange of peptides

Electrostatic capture onto resin bead in H⁺ form

Stepwise elution with NH₄⁺

S. Barnes & M. Renfrow-UAB 1/12/11
Connecting CE and LC to MALDI analysis

- CE analysis
- nanoLC analysis

Creates 20 mm wide tracks that can be scanned by MALDI laser for MS analysis

Parallel capture of effluents of 8 nanoLC separations on Mylar - can be scanned simultaneously by fast laser

S. Barnes & M. Renfrow-UAB 1/12/11
Pros/Cons of laying down LC or EC separations on matrix plate

- Allows off-line analysis both in real time and then in a retrospective mode
- MALDI-TOF analysis is very fast
- Can also do TOF-TOF MS-MS analysis
- BUT what happens chemically on the acidic environment on the surface of the plate during storage?
- Also, can the laser beam cause chemical changes?

S. Barnes & M. Renfrow-UAB 1/12/11
Sequencing of peptides

• Using tandem mass spectrometry in a triple quadrupole, Q-tof, or ion trap instrument, the parent ion is first selected in the first quadrupole.

• The parent ion is collided with argon gas and it breaks into fragments (daughter ions).

• By identifying the daughter ions, the peptide amino acid sequence is inferred.
Tandem mass spectrometry on a triple quadrupole instrument

- Daughter ion spectra
- Parent ion spectra
- Multiple reaction ion scanning

S. Barnes & M. Renfrow-UAB 1/12/11
The quadrupole analyzer (Q3) is slow and insensitive - it’s a filter - thus throws away large amounts of data.

TOF detector collects all ions generated and yields fmol rather than pmol sensitivity.

Also gives far greater mass accuracy - from 1000 ppm on the triple quad to <20 ppm on the Q-tof.

Crucially important for automated interpretation of MS-MS spectra to yield amino acid sequence.

S. Barnes & M. Renfrow-UAB 1/12/11