Shotgun proteomics

Niels Lion
Service Régional Vaudois de Transfusion Sanguine
Lausanne
Spring 2011
niels.lion@mavietonsang.ch
Outline

- Workflow
- Electrospray ionisation interface for LC-MS/MS
- 1D-LC-MS/MS
- Protein identification strategy (and problems)
- 2D-LC-MS/MS
- Multidimensional Protein Identification Technology
General LC-MS/MS workflow

- Protein mixture
- Peptide mixture
- SCX
- RP
- 1 MS spectrum
- n MS/MS spectra

Generation of hundreds of MS spectra, thousands of MS/MS spectra

Importance of automatic identification tools
Challenges in LC and MS hyphenation

<table>
<thead>
<tr>
<th>LC</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liquid phase</td>
<td>Gas phase</td>
</tr>
<tr>
<td>High flow rate</td>
<td>Moderate to high vacuum</td>
</tr>
<tr>
<td>Analytes are non ionised</td>
<td>Analytes are ionised</td>
</tr>
</tbody>
</table>
Electrospray ionization

$E = 0$

$0 < E < E_{\text{onset}}$

$E > E_{\text{onset}}$

Solvent evaporation

Droplet fission

Droplet fission

Solvent evaporation

Ion evaporation
Each peak corresponds to one protonation state (n protons) \[ \Rightarrow \text{calculation of the molecular weight} \]
# Molecular weight calculation

| 606.4186 | 998.1549 |
| 628.8412 | 1060.4766 |
| 652.9886 | 1131.1078 |
| 679.0679 | 1211.8293 |
| 707.3204 | 1304.9694 |
| 738.0296 | 1413.6328 |
| 771.5306 | 1542.0533 |
| 808.2221 | 1696.1578 |
| 848.5828 | 1884.5078 |
| 893.1921 | 2119.9453 |
| 942.7577 | 2422.6506 |
Molecular weight calculation

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>606.4186</td>
<td>998.1549</td>
</tr>
<tr>
<td>628.8412</td>
<td>1060.4766</td>
</tr>
<tr>
<td>652.9886</td>
<td>1131.1078</td>
</tr>
<tr>
<td>679.0679</td>
<td>1211.8293</td>
</tr>
<tr>
<td>707.3204</td>
<td>1304.9694</td>
</tr>
<tr>
<td>738.0296</td>
<td>1413.6328</td>
</tr>
<tr>
<td>771.5306</td>
<td>1542.0533</td>
</tr>
<tr>
<td>808.2221</td>
<td>1696.1578</td>
</tr>
<tr>
<td>848.5828</td>
<td>1884.5078</td>
</tr>
<tr>
<td>893.1921</td>
<td>2119.9453</td>
</tr>
<tr>
<td>942.7577</td>
<td>2422.6506</td>
</tr>
</tbody>
</table>

1. Charge state determination:
2. Molecular weight evaluation

\[ m/z = \frac{M + zH}{z} \]
Molecular weight calculation

<table>
<thead>
<tr>
<th>m/z</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>606.4186</td>
<td>998.1549</td>
</tr>
<tr>
<td>628.8412</td>
<td>1060.4766</td>
</tr>
<tr>
<td>652.9886</td>
<td>1131.1078</td>
</tr>
<tr>
<td>679.0679</td>
<td>1211.8293</td>
</tr>
<tr>
<td>707.3204</td>
<td>1304.9694</td>
</tr>
<tr>
<td>738.0296</td>
<td>1413.6328</td>
</tr>
<tr>
<td>771.5306</td>
<td>1542.0533</td>
</tr>
<tr>
<td>808.2221</td>
<td>1696.1578</td>
</tr>
<tr>
<td>848.5828</td>
<td>1884.5078</td>
</tr>
<tr>
<td>893.1921</td>
<td>2119.9453</td>
</tr>
<tr>
<td>942.7577</td>
<td>2422.6506</td>
</tr>
</tbody>
</table>

\[
m/z = \frac{(M + zH)}{z}
\]

\[
M = z \times m/z - zH
\]

If \( z = 10 \):

- \( M = 14126.328 \)
- \( M = 13869.4797 \)
- \( M = 13561.2624 \)

If \( z = 11 \):

- \( M = 15538.960 \)
- \( M = 15410.533 \)
- \( M = 15256.4202 \)

If \( z = 12 \):

- \( M = 16951.5936 \)
- \( M = 16951.5863 \)
- \( M = 16951.578 \)
Molecular weight calculation

<table>
<thead>
<tr>
<th>z</th>
<th>m/z</th>
<th>M</th>
<th>ΔM</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>606.4186</td>
<td>998.1549</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>628.8412</td>
<td>1060.4766</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>652.9886</td>
<td>1131.1078</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>679.0679</td>
<td>1211.8293</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>707.3204</td>
<td>1304.9694</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>738.0296</td>
<td>1413.6328</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>771.5306</td>
<td>1542.0533</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>808.2221</td>
<td>1696.1578</td>
<td></td>
</tr>
<tr>
<td>09</td>
<td>848.5828</td>
<td>1884.5078</td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>893.1921</td>
<td>2119.9453</td>
<td></td>
</tr>
<tr>
<td>07</td>
<td>942.7577</td>
<td>2422.6506</td>
<td></td>
</tr>
</tbody>
</table>

\[
m/z = (M + zH)/z
\]

\[
M = z^* (m/z - zH)
\]

\[
z = 12:
M = 16951.5936
M = 16951.5863
M = 16951.578
M = 16951.6373 \pm 0.0513 \text{ Da}
\]

Thursday, May 5, 2011
ESI features

+ Multiply charged analytes

+ Good for on-line coupling with liquid phase separations

− Multiply charged analytes (need for deconvolution)

− Ion suppression

− Requires good desolvation (presence of organic solvent)

− Intolerant to salts
Flow rate versus sensitivity

nanoelectrospray has three orders of magnitude better sensitivity than electrospray
## Different flow rates

<table>
<thead>
<tr>
<th>Flow rate</th>
<th>LC</th>
<th>ESI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 mL/min</td>
<td>microbore, analytical (4.16 mm columns)</td>
<td>NO (split)</td>
</tr>
<tr>
<td>1-10 µL/min</td>
<td>capillary (300µm columns)</td>
<td>microspray</td>
</tr>
<tr>
<td>0.05-0.3 µL/min</td>
<td>nano (75µm columns)</td>
<td>nanospray</td>
</tr>
</tbody>
</table>
1D-LC-MS

As electrospray ionisation is not tolerant to non-volatile salts (as commonly found in biological buffers), 1D-LC-MS/MS is always based on reversed-phase chromatography.
Peptide fragmentation
Low energy fragmentation

Charge on the N-ter part

\[
\begin{align*}
\text{a}_2 & & R_1 & \oplus & R_2 \\
H_2N-CH-CO-NH=CH \\
\text{b}_2 & & R_1 & \oplus & R_2 \\
H_2N-CH-CO-NH-CH-C=O \\
\text{c}_2 & & R_1 & \oplus & R_2 \\
H_2N-CH-CO-NH-CH-CO-NH_3
\end{align*}
\]

Charge on the C-ter part

\[
\begin{align*}
\oplus & & R_3 \\
CO-NH-CH-COOH \\
\oplus & & R_3 \\
H_3N-CH-COOH \\
\oplus & & R_3 \\
CH-COOH
\end{align*}
\]

Immonium ions

\[
\begin{align*}
\oplus & & R_2 \\
H_2N=CH
\end{align*}
\]

Observable on all MS/MS instruments
High energy fragmentation

Charge on the N-ter part

Charge on the C-ter part

Observable only on TOF/TOF instruments
Different mass analysers

- TOF
- TOF-TOF
- QqQ ou Q-Trap
- Q-TOF
- FT-ICR
- IT

Resolution > 100,000
Various kind of MS are available
MS/MS combinations

- TOF/TOF
- ion trap
- triple quadrupoles
- C-trap-orbitrap
- Q-Trap....
Peptide ions are trapped in the oscillatory electric field
Procedure for MS/MS in ion traps

- Inject ions in the trap

- Ramp the ejection voltage to sequentially destabilize ion trajectories to the trap outlet to the detector

- Inject the same ions

- Eject all ions except the one to be fragmented

- Increase kinetic energy to induce peptide / He collisions and peptide fragmentation

- Ramp the ejection voltage to sequentially destabilize ion trajectories to the trap outlet to the detector
Extreme example

4µg E. Coli protein digest injected on a 350 cm, 100 µm ID C18 column; peptides were separated over 41 hours at 500 nL/min.

Identification of 2602 E. Coli proteins (830 membrane proteins) in single LC-MS/MS experiment.

Anal. Chem. 2010, 82, 2616–2620
HPLC Chip interface

In shotgun proteomics, sequence coverages are very low!!!
Shotgun, bottom-up: poor MS/MS identification

Distribution of XCorr scores over 10000 MS/MS spectra of human proteins.

Example from RBC proteomics

During RBC storage, RBC loose microvesicles by a controlled (yet unknown) process. These microvesicles have a pro-coagulant activity that might be detrimental for the patient.

A: RBC membranes
B: RBC microvesicles
Example from RBC proteomics

- 395 identified proteins with an average of 7 peptides per protein, 24% sequence coverage

- See Scaffold example....
How to increase the confidence in MS/MS identifications?
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name and composition</th>
<th>Structure</th>
<th>Isotopic mass</th>
<th>Average mass</th>
<th>pKA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala A</td>
<td>Alanine C₃H₅NO</td>
<td></td>
<td>71.03711</td>
<td>71.0788</td>
<td></td>
</tr>
<tr>
<td>Arg R</td>
<td>Arginine C₅H₁₂N₄O</td>
<td></td>
<td>156.1011</td>
<td>156.1876</td>
<td>13.2</td>
</tr>
<tr>
<td>Asn N</td>
<td>Asparagine C₄H₆N₂O₂</td>
<td></td>
<td>114.04293</td>
<td>114.1039</td>
<td></td>
</tr>
<tr>
<td>Asp D</td>
<td>Aspartic acid C₄H₅NO₃</td>
<td></td>
<td>115.02694</td>
<td>115.0886</td>
<td>3.9</td>
</tr>
<tr>
<td>Cys C</td>
<td>Cysteine C₃H₅NOS</td>
<td></td>
<td>103.00919</td>
<td>103.1448</td>
<td>8.4</td>
</tr>
<tr>
<td>Gln Q</td>
<td>Glutamine C₅H₈N₂O₂</td>
<td></td>
<td>128.05858</td>
<td>128.1308</td>
<td></td>
</tr>
<tr>
<td>glu E</td>
<td>Glutamic acid C₅H₇NO₃</td>
<td></td>
<td>129.04259</td>
<td>129.1155</td>
<td>4.3</td>
</tr>
<tr>
<td>Gly G</td>
<td>Glycine C₂H₃NO</td>
<td></td>
<td>57.02146</td>
<td>57.0520</td>
<td></td>
</tr>
<tr>
<td>His H</td>
<td>Histidine C₆H₇N₃O</td>
<td></td>
<td>137.05891</td>
<td>137.1412</td>
<td>6.1</td>
</tr>
<tr>
<td>Ile I</td>
<td>Isoleucine C₆H₁₁NΟ</td>
<td></td>
<td>113.08406</td>
<td>113.1595</td>
<td></td>
</tr>
<tr>
<td>Leu L</td>
<td>Leucine C₆H₁₁NΟ</td>
<td></td>
<td>113.08406</td>
<td>113.1595</td>
<td></td>
</tr>
<tr>
<td>Lys K</td>
<td>Lysine C₆H₁₂N₂O</td>
<td></td>
<td>128.09496</td>
<td>128.1742</td>
<td>10.8</td>
</tr>
<tr>
<td>Met M</td>
<td>Methionine C₆H₉NOS</td>
<td></td>
<td>131.04049</td>
<td>131.1986</td>
<td></td>
</tr>
<tr>
<td>Phe F</td>
<td>Phenylalanine C₉H₉NO</td>
<td></td>
<td>147.06841</td>
<td>147.1766</td>
<td></td>
</tr>
<tr>
<td>Pro P</td>
<td>Proline C₅H₉NO</td>
<td></td>
<td>97.05276</td>
<td>97.1167</td>
<td></td>
</tr>
<tr>
<td>Ser S</td>
<td>Serine C₃H₅NO₂</td>
<td></td>
<td>87.03203</td>
<td>87.0782</td>
<td></td>
</tr>
<tr>
<td>Thr T</td>
<td>Threonine C₄H₇NO₂</td>
<td></td>
<td>101.04768</td>
<td>101.1051</td>
<td></td>
</tr>
<tr>
<td>Trp W</td>
<td>Tryptophane C₁₁H₁₀N₂O</td>
<td></td>
<td>186.07931</td>
<td>186.2133</td>
<td></td>
</tr>
<tr>
<td>Tyr Y</td>
<td>Tyrosine C₉H₉NO₂</td>
<td></td>
<td>163.06333</td>
<td>163.1760</td>
<td>10.1</td>
</tr>
<tr>
<td>Val V</td>
<td>Valine C₅H₉NO</td>
<td></td>
<td>99.06841</td>
<td>99.1326</td>
<td></td>
</tr>
<tr>
<td>Nter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.7</td>
</tr>
<tr>
<td>Cter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.5</td>
</tr>
</tbody>
</table>

*Note: Additional rows for Nter and Cter are not included in the table*
Positively charged amino acids

Histidine

\[
\begin{align*}
\text{C}_2\text{H}_2\text{N}^+ & \overset{\leftrightarrow}{\longrightarrow} \text{C}_2\text{H}_2\text{N} + \text{H}^+ \\
6.1
\end{align*}
\]

Lysine

\[
\begin{align*}
\text{C}_2\text{H}_2\text{C}\text{C}\text{C}\text{C}\text{NH}_3^+ & \overset{\leftrightarrow}{\longrightarrow} \text{C}_2\text{H}_2\text{C}\text{C}\text{C}\text{NH}_2 + \text{H}^+ \\
10.8
\end{align*}
\]

Arginine

\[
\begin{align*}
\text{C}_2\text{H}_2\text{C}\text{C}\text{N}\text{C}\text{NH}_2^+ & \overset{\leftrightarrow}{\longrightarrow} \text{C}_2\text{H}_2\text{C}\text{C}\text{N}\text{C}\text{NH}_2 + \text{H}^+ \\
13.2
\end{align*}
\]

N-ter

\[
\begin{align*}
\text{NH}_3^+ & \overset{\leftrightarrow}{\longrightarrow} \text{NH}_2 + \text{H}^+ \\
9.7
\end{align*}
\]
Negatively charged amino acids

Cysteine $\xrightarrow{\text{H}_2}\text{C} - \text{SH} \xrightarrow{\text{H}^+} \text{C} - \text{S}^- + \text{H}^+$ 8.4

Tyrosine $\xrightarrow{\text{H}_2}\text{C} - \text{H}_2\text{O} - \text{OH} \xrightarrow{\text{H}^+} \text{C} - \text{H}_2\text{O} - \text{O}^- + \text{H}^+$ 10.8

Aspartic acid
$\xrightarrow{\text{H}_2}\text{C} - \text{C} - \text{O} - \text{OH} \xrightarrow{\text{H}^+} \text{C} - \text{C} - \text{O}^- + \text{H}^+$ 3.9

Glutamic acid
$\xrightarrow{\text{H}_2}\text{C} - \text{C} - \text{C} - \text{O} - \text{OH} \xrightarrow{\text{H}^+} \text{C} - \text{C} - \text{C} - \text{O}^- + \text{H}^+$ 4.3

C-ter $\text{COOH} \xrightarrow{\text{H}^+} \text{COO}^- + \text{H}^+$ 2.5
Charge of a polypeptide

\[
\text{net charge} = \sum_{\text{positively charged}} \frac{n_i}{K_i 10^{-pH}} + 1 - \sum_{\text{negatively charged}} \frac{n_j}{10^{-pH} K_j} + 1
\]

Well-defined pI

Poorly-defined pI
Isoelectric focalisation

Diffusion-migration flux (linear pH gradient):

\[ \mathbf{J}_i = -c_i \tilde{u}_i \text{grad} \tilde{\mu}_i = -c_i \tilde{u}_i \text{grad} \mu_i - c_i \tilde{u}_i z_i F \text{grad} \phi \]

Flux conservation:

\[ \frac{\partial c_i}{\partial pH} = -\text{div} \mathbf{J}_i = 0 \]

\[- \frac{\partial}{\partial pH} \left[ - \frac{c_i \tilde{u}_i RT}{c_i} \frac{\partial c_i}{\partial pH} - c_i \tilde{u}_i z_i F \frac{\partial \phi}{\partial pH} \right] = 0 \]

\[ RT \frac{\partial c_i}{\partial pH} = c_i z_i FE \]
Resolution

If the charge is proportional to the pH:

\[ z_i = -p_i \times pH \]

\[ RT \frac{\partial c_i}{\partial pH} = -c_i p_i FE \times pH \]

Which then gives the concentration:

\[ c_i = c_i^{\text{max}} \exp \left( \frac{-p_i FE (pH - pI)^2}{2 RT} \right) \]
Concentration profile

RDCEHKY

Net charge

pH

YGGFL

Net charge

pH

Normalised concentration

pH

Normalised concentration

pH
Off-Gel run
Off-Gel run
Off-Gel run
Off-Gel run
Global strategy

OGE1 of plasma proteins from 50 µg up to 60 mg

- anode
- 15 fractions – 0.2 U pH / well
- cathode

pH 4.0

Trypsin digestion of fraction 6 – 0.2 pH U (pH 5.2)

OGE2 of the peptide solution

- anode
- 225 fractions – 0.5 U pH / well
- cathode

pH 3.0

LC-MS/MS analysis of the peptide fractions
and sequence database search (Phenyx and SpectrumMill)

pH 10.0
Theoretical pI of the peptides identified in protein fraction 6 of human plasma

104 Proteins with 5<pI<5.15
675 Peptides

Heller et al. Electrophoresis 2005, 26, 1174–1188
Peptide pI and score correlation

- True positives
- False positives
- False negatives
- True negatives
Validation of bioinformatics hits thanks to peptide pI analysis

Peptide distribution of human plasma protein

- Identified peptides
- % of peptide in the right fraction

Bioinformatic Score (SpectruMil)

7 8 9 10 11 12 13

peptide Nbre

0 500 1000 1500 2000 2500 3000 3500

% in correct fraction

96 94 92 90 88 86 84 82 80 78 76
Drosophila melanogaster proteome,

Intensity (x 10^4)

m/z

Added Value for Tandem Mass Spectrometry Shotgun Proteomics Data Validation through Isoelectric Focusing of Peptides

Manfred Heller#, §, Mingliang Ye&, Philippe E. Michel#, Patrick Morier#, Daniel Stalder§, Martin A. Jünger*, Ruedi Aebersold&, °, Frédéric Reymond#, Joël S. Rossier#, *

J. Proteome Research, in press
Retention time calculation

Figure 3. Linear MS scan number to $R_T$ curve fitting with the combined PeptideProphet and high scoring PHENYX (p-value < $10^{-7}$) peptide identifications. Linear curve fitting was done with the least squares method applying a 95% confidence interval (thin lines) using Matlab (lower acceptance boundary: $y = 0.01991^*x - 4.203$; upper acceptance boundary: $y = 0.01991^*x + 33.175$).
How to increase the depth of analysis?
Alternative strategy: offline selection

1st LC-MS run

Then selection of interesting spots for targeted MS/MS
DDA versus targeted proteomics

Analysis of the tryptic digest of beta-lactoglobulin

2D-LC-MS/MS
Principle of 2D-LC-MS/MS

Sample is loaded on the SCX column, fractions are eluted by salt steps to the RP enrichment column, the RP column is rinsed by the nanopump.
Principle of 2D-LC-MS/MS

First valved is switched to the waste, 2nd valved is switched to inject on fraction eluted from the first dimension into the RP column.
10 protein digest

Different peptides are eluted from the SCX column by different KCl concentrations

RP profile of SCX fractions
2D-LC-MS/MS

Total yeast extract (proteome)

~ 20,000 protéines

Trypsin digestion

Total yeast peptide mixture

Strong cation exchange chromatography

80 peptide fractions

Automated reverse-phase chromatography with tandem mass spectrometry

A total of ~162,000 MS/MS spectra

Database searching & data processing

Identification of 26,815 peptides (7,537 unique peptides, 1,504 proteins)

Peng J. et al. J. Proteome Res. 2003, 2, 43-50
MudPit

Sequential elution of peptides from the SCX part into the RP part by 12 salt steps.

MudPit resulted in a dynamic range of 10'000 (most abundant/lowest abundant peptide)

Anal. Chem. 2001, 73, 5683-5690
Take-home message

• Shotgun proteomics is the current workhorse of proteomic research.

• However, it still suffers from a number of drawbacks (low sequence coverages, reproducibility, quantitation)

• The most critical step is the interpretation of the results (list of proteins into biological information)