Summary

- Omics and Mass spectrometry
- Why study the proteome?
- Bottom-up proteomics: an LC-MS story
- DIA and HRM
- Quantification
The interface between Chemistry/Biology is essential

Two of the three most cited papers of all time report analytical chemistry techniques to study biological systems


To understand a biological process we often need a better comprehension of the associated chemical environment
Omics or Omix

‘Omic’ sciences are perhaps the best example of a successful integration of chemistry and biology.


The ‘omic’ revolution has been driven by advances in analytical chemistry, from DNA microarray technology to mass spectrometry.

2 divergent lines of enquiry with regard to ‘omic’ sciences and systems biology:

(1) how can analytical chemistry be improved to better answer key biological questions?

(2) are the right questions being asked that take advantage of the tools at our disposal?
The example of Mass Spec

1. How much?
2. How deep?
3. How fast?
4. What flexibility?

Part of the Calutron mass spectrometer first used for preparative MS


Summary

• Omics and Mass spectrometry
• **Why study the proteome?**
• Bottom-up proteomics: an LC-MS story
• DIA and HRM
• Quantification
Proteomics

**Proteome**: The pool of PROteins isoforms expressed by one genOME of a cell or a tissue at one given time in one given environment. (Wilkins et al., Biotechnol Gene Eng Rev, 1995)

**Objective 1**: Exhaustivity

**Proteomic Analysis**: Dynamic and quantitative analysis of the regulation of expression of genes products for a given biological process in order to decipher molecular interaction mechanisms. (Anderson et Anderson, Electrophoresis, 1998)

**Objective 2**: Quantitative Analysis
From DNA to proteins

1 gene = 1 protein ??

1 gene = several proteins

phosphorylation, acetylation, oxidation, glycosylation, lipoylation

No apparent correlation between the size of the genome and the complexity of the organism

<table>
<thead>
<tr>
<th>Organism</th>
<th># Chromosom</th>
<th>Genome size (pb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>2x23</td>
<td>3.2 E9</td>
</tr>
<tr>
<td>Mus musculus (mouse)</td>
<td>2x20</td>
<td>2.7 E9</td>
</tr>
<tr>
<td>Drosophila melanogaster (Fly)</td>
<td>2x4</td>
<td>1.4 E8</td>
</tr>
<tr>
<td>Tobacco</td>
<td>2x24</td>
<td>4.8 E9</td>
</tr>
<tr>
<td>Zea mays (corn)</td>
<td>2x10</td>
<td>3.9 E9</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae (yeast)</td>
<td>17</td>
<td>1.5 E7</td>
</tr>
<tr>
<td>E Coli (bacteria)</td>
<td>1</td>
<td>4.64 E5</td>
</tr>
<tr>
<td>Phage λ</td>
<td>1</td>
<td>4.85 E4</td>
</tr>
<tr>
<td>Allium cepa (onion)</td>
<td>2x8</td>
<td>1.5 E10</td>
</tr>
<tr>
<td>Amoeba dubia (amibae)</td>
<td>2x6</td>
<td>6.7 E11</td>
</tr>
<tr>
<td>Lysandra atlantica (blue lycaenid butterfly)</td>
<td>250</td>
<td>N/A</td>
</tr>
<tr>
<td>Ophioglossum reticulatum</td>
<td>2x630</td>
<td>1.6 E11</td>
</tr>
</tbody>
</table>

**Protein–coding Human genes**

- 2001: 30000
- 2004: 25000
- 2012: 21065
- 2013: 20774
- 2014: 19000

---

**Total Number of Protein-Coding Genes**

- Drosophila melanogaster (fruitfly): 13,917
- Pan troglodytes (chimpanzee): 18,746
- Canis familiaris (dog): 19,856
- Bos taurus (cow): 19,994
- Caenorhabditis elegans (nematode): 20,517
- Homo sapiens (human): 20,774
- Arabidopsis thaliana (mustard weed): 27,416
- Physcomitrella patens (moss): 35,938
- Oryza sativa (rice): 40,577
- Populus trichocarpa (poplar): 41,377
- Manihot esculenta (cassava): 47,164
- Malus domestica (apple): 57,386
- Triticum aestivum (bread wheat): >94,000

---

Correlation with mRNA

• Lower stability of mRNA (lower half-life)
• Larger distribution of abundance of protein

Figure 2 | mRNA and protein levels and half-lives. a, b, Histograms of mRNA (blue) and protein (red) half-lives (a) and levels (b). Proteins were on average 5 times more stable and 2,800 times more abundant than mRNAs and spanned a higher dynamic range. c, d, Although mRNA and protein levels correlated significantly, correlation of half-lives was virtually absent.

SAME GENOME BUT DIFFERENT PROTEOME

Why study the proteome?

Genome is static but
Structures change just as the caterpillar develops into a butterfly
**Taxonomy of Omics**

**CHALLENGES OF PROTEIN PTM ANALYSIS BY MS**

- **Genomics**
  - Gene sequences

- **Transcriptomomics**
  - mRNA sequences & expression level

- **Proteomics**
  - Proteins isoforms sequences & expression levels

- **Metabolomics**
  - Metabolites as products or substrate of enzymes

Let’s focus

- **PTMomics**
  - Proteoforms sequences & expression levels

**Proteins and Proteoforms: New Separation**
Challenges Regnier FE and Kim JH, Anal Chem. 2018

How many human proteoforms are there? Aebersold R et al Nat Chem Biol. 2018
Summary

- Omics and Mass spectrometry
- Why study the proteome?
- **Bottom-up proteomics: an LC-MS story**
- DIA and HRM
- Quantification
1: Proteins
Extraction, purification of proteins
Proteolysis to generate peptides

2: Peptides
Complex mixtures
Separation and purification

3: Mass spectrometry
Peptide masses
Peptide sequences

4: Data processing
Raw data analysis

MS-based untargeted proteomics
802 tools
Shotgun Proteomics

Unknown organization

Spare parts inventory

Identification of original composition

False positive identification

Potential modifications

Artifacts
Endoproteolysis

Example: trypsin, which cleaves after the carboxyl moiety
  . of lysine (K)
  . of arginine (R) (except when followed by proline)

Generate tryptic peptides whose sequencing in tandem mass spectrometry is favored by the basic residue at C-ter

1 missed cleavage:
  R,
  EMINDER,
  ATTEACTIVITYK,
  FTR,
  YPSINK,
  PEPTIDEBNDS,
  REMINDER,
  EMINDER ATTEACTIVITYK,
  ATTEACTIVITYK FTR,
  FTRYPSINK,
  YPSINK PEPTIDEBNDS
Separation and purification

- Charge → ion exchange
- Polarity → HILIC
- Hydrophobicity → reverse phase chromatography

Elution: organic solvent gradient (acetonitrile, methanol, ...)
LC-MS is a Gold standard for bottom-up proteomics

- Liquid chromatography aims at the separation of molecules according to their physico-chemical properties using various stationary and mobile phases.

- From people in the separation sciences we need a detector and

  > A mass spectrometer is one of them, that allows the m/z measurement of eluted species.

- Separation of proteins or peptides?
- Which phase, which columns?
- Which ion source?
Chromatographic tools

Mix of proteins

Separation: IEC, SEC, affinity

Fractions collection

Digestion

Mix of peptides

Separation: IEC, RP

1D Gel

MS and MS/MS
Chromatographic separation mode

Properties of proteins

- Size (MW) → Size/steric exclusion chromatography

Chromatographic methods

Elution Order: large molecules first

Porous material?
Chromatographic separation mode

Properties of proteins

- Size (MW)
- Activity (affinity)

Chromatographic methods

- Size/steric exclusion chromatography
- Affinity chromatography

Elution: Ionic strength gradient
pH gradient
competitive elution
Chromatographic separation mode

Properties of proteins

- Size (MW)
- Activity (affinity)
- Charge (pI)

Chromatographic methods

- Size/steric exclusion chromatography
- Affinity chromatography
- Ion exchange chromatography

Elution: Ionic strength gradient
pH gradient

Anions exchange: amino moiety
Cations exchange: sulfonic/phospho/carboxylic moiety

pH > pI: -
pH < pI: +
Chromatographic separation mode

Properties of proteins

- Size (MW)
- Activity (affinity)
- Charge (pI)
- Polarity

Chromatographic methods

- Size/steric exclusion chromatography
- Affinity chromatography
- Ion exchange chromatography
- Normal phase chromatography

Elution: organic solvent gradient (Hexane, toluene,...)
Chromatographic separation mode

Properties of proteins

- Size (MW)
- Activity (affinity)
- Charge (pI)
- Polarity

Chromatographic methods

- Size/steric exclusion chromatography
- Affinity chromatography
- Ion exchange chromatography
- Normal phase chromatography
- Hydrophilic interaction chromatography (HILIC)

Elution: aqueous phase gradient

Easy LC-MS interface
Chromatographic separation mode

Properties of proteins

- Size (MW)
- Activity (affinity)
- Charge (pI)
- Polarity
- Hydrophobicity

Chromatographic methods

- Size/steric exclusion chromatography
- Affinity chromatography
- Ion exchange chromatography
- Normal phase chromatography
- Hydrophilic interaction chromatography (HILIC)
- Reversed phase chromatography (RP)

Elution: organic solvent gradient (acetonitrile, methanol, ...)

Easy LC-MS interface
Selection of column dimension

- Column selection is a function of the analytes of interest, **nature**, **complexity** and **abundance**.

- **Size and nature of particules**

  - **Van Deemter equation**

    \[ H = 2\lambda_d d_p + 2\frac{\gamma D_m}{v} + \frac{\omega d_p^2}{D_m} \]

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

  - **A: Convection diffusion** (\( \varphi_{\text{stationnaire}} \))

  - **B: Longitudinal diffusion** (\( \varphi_{\text{mobile}} \))

  - **C: Mass transfer resistance** (\( \varphi_{\text{stationnaire}} / \varphi_{\text{mobile}} \))
Selection of column dimension

- Column selection is a function of the analytes of interest, **nature**, **complexity** and **abundance**.

- **Size and nature of particules**

\[ H = 2\lambda d_p + 2\gamma \frac{D_m}{v} + \frac{\omega d_p^2}{D_m} \]

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

\[ \Delta P = \frac{\phi \eta L u}{d_p^2} \]
Selection of column dimension

- Column selection is a function of the analytes of interest, **nature**, **complexity** and **abundance**.

- Pore size/ grafted alkyle chain

  peptides

  proteins

  $d_{\text{pore}} > 500 \text{ Å}$ macroporous

  $20 \text{ Å} < d_{\text{pore}} < 500 \text{ Å}$ mesoporous

  $d_{\text{pore}} < 20 \text{ Å}$ microporous
Selection of column dimension

- Column selection is a function of the analytes of interest, nature, complexity and abundance.

- Column length

\[ R = 2 \frac{(t_{r2} - t_{r1})}{(\omega_1 + \omega_2)} = \sqrt{\frac{N_2}{N}} \frac{\alpha - 1}{\alpha} \frac{k_2}{1 + k_2} \]

\[ N = \frac{L}{H} \quad (eq.2) \]

Subplots:

- Sample: Kidney Biopsy (S. Liu et al)

Bar charts:
- 3h with 50cm length
  - \( \omega_{1/2} = 0.28 \) min

- 3h with 15cm length
  - \( \omega_{1/2} = 0.47 \) min

Graphs:
- LTQ - FT Ultra (ThermoFisher Scientific)
Selection of column dimension

- Column selection is a function of the analytes of interest, nature, complexity and abundance.

➢ Gradient duration/slope

![Graph showing peptide and protein count over gradient duration](image)

- # MS/MS
- # Peptides (FDR<1%)
- # Proteins (3 pep./prot.)

Sample: Kidney Biopsy (S. Liuu et al)

LTQ - FT Ultra (ThermoFisher Scientific)
Selection of column dimension

• Column selection is a function of the analytes of interest, nature, complexity and abundance.

➤ Column internal diameter and flow rate

<table>
<thead>
<tr>
<th>Diamètre interne</th>
<th>Nomenclature</th>
<th>Débit</th>
<th>Quantité</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 mm</td>
<td>Conventionnel LC</td>
<td>1ml/min</td>
<td>1-200 μg (10pmol)</td>
</tr>
<tr>
<td>2 mm</td>
<td>Narrowbore LC</td>
<td>200μl/min</td>
<td>2-50 μg (5pmol)</td>
</tr>
<tr>
<td>1 mm</td>
<td>Micro LC</td>
<td>40μl/min</td>
<td>0,05-10 μg (1pmol)</td>
</tr>
<tr>
<td>300 μm</td>
<td>Capillaire LC</td>
<td>4μl/min</td>
<td>1 ng-1μg (500 fmol)</td>
</tr>
<tr>
<td>75 μm</td>
<td>NanoLC</td>
<td>300nl/min</td>
<td>0,02-0,05 ng (1 fmol)</td>
</tr>
</tbody>
</table>

Increased sensitivity  (ESI signal concentration dependant)
Advantage of low flow rates

• The conventional ESI interfaces operate at high flow rates: typically, 1-1000μL.min⁻¹ is a typical concentration-sensitive technique. Within this range, increasing the flow rate does not normally increase the signal.

• NanoESI-MS, at low flow rates (typically, 300-50 nL.min⁻¹), exhibits superior mass sensitivity, with a high ionization efficiency.

\[ C_{\text{max}} = \frac{mN^{1/2}}{(2\pi)^{1/2}V_0(1 + k)} \]

- \( C_{\text{max}} \): eluted analyte concentration
- \( m \): absolute abundance
- \( N \): column efficiency
- \( V_0 \): column volume
- \( k \): retention factor

• Max Concentration = \( K/(\text{Volume colonne}) \)

→ decrease the column diameter
→ optimization of the flow rate for nanoESI mode
Which ion source?

- **ESI**: Compatibility with liquid phase

- **MALDI**: Fraction collection

- **Complementarity:**
  - Ion mode => different ion nature
  - Advantages MALDI: Off-line analysis
  - Advantages ESI: Speed
TriVersa NanoMate (Advion)

Combination Autosampler/ion source
- Fraction collector
- Sample cooling
- “Spray sensing”

4 operation modes
- direct infusion
- LC Coupling
- Fraction collector
- LESA (Liquid Extraction Surface Analysis)
Other interfaces nano-ESI

- EASY-Spray
- Chip Cube Interface

- Reduction of dead volumes (extra-column, leaks, etc.)
- Robustness ease of use
- Costs
Following the samples ... and others

loading pump 100% solvent A

syringe sample

waste

micro pump ACN gradient

Trap col.

Analytical column (RP)

ESI-FTMS

Pics des contaminants de PEG

578.877
722.818
728.838
599.382
593.350
636.311
662.411
698.437
740.463
764.490
767.480
818.888
828.888

NL: 3.17E6
Following the samples ... and others

<table>
<thead>
<tr>
<th>PROTEINE</th>
<th>SANS</th>
<th>AVEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADH1 YEAST</td>
<td>139 +/- 19</td>
<td>2 +/- 0</td>
</tr>
<tr>
<td>ALBU BOVIN</td>
<td>401 +/- 20</td>
<td>7 +/- 1</td>
</tr>
<tr>
<td>BGAL ECOLI</td>
<td>149 +/- 21</td>
<td>3 +/- 1</td>
</tr>
<tr>
<td>CYC BOVIN</td>
<td>49 +/- 43</td>
<td>1 +/- 1</td>
</tr>
<tr>
<td>LYS C CHICK</td>
<td>74 +/- 3</td>
<td>1 +/- 1</td>
</tr>
<tr>
<td>TRFE BOVIN</td>
<td>236 +/- 148</td>
<td>4 +/- 2</td>
</tr>
</tbody>
</table>

Gain in sensitivité

Optimization of peptides analysis in NanoLC-MS/MS

- In all cases:
  - Duration of gradient according to sample complexity
  - Duration of dynamic exclusion according to peak width

- ESI:
  - duty cycle
  - Scan time (best ratio quality/time)
  - # of MS/MS per MS

- MALDI: time is not (?) an issue ...
  - Fraction collection
  - MS/MS (energy, # laser shots)
  - # de MS/MS per spot
NanoLC-MS with sequential MS/MS and MS

Elution gradient
0 => 50 % B in 180 min then 90% B for 15 min @300nl/min
A = 98% water/ 2% ACN/ 0,1% AF (v:v:v) ; B = 10% water/ 90% ACN/ 0,1% AF (v:v:v)

Automated dynamic exclusion

Automatic edition of a precursor list with their fragments

Instruments : Ultimate 3000 RSLC (Dionex), Qexactive (ThermoFisher Scientific)
Example of LC MS profile
Thyroid biopsy

Condition LC: RSLC, column C18, 5µ, 300A, 75µm id, 50 cm, 35°C, 300nL/min, 200bar
Condition MS: Qexactive, DDA scan top 10, res_{m/z300} MS 50000, res_{m/z300} MS/MS 17500
Zoom 50-55min, m/z 500-900, peak width=30s, sampling time MS 0,6s and MS/MS < 120ms
Raw data processing

Full scan MS

MS/MS
Peptide sequencing: Example of one MS/MS spectra

Two ladders are superimposed in opposite direction

Identification of sequence HYQLNQQWER
High accuracy/high resolution effect on a peptidic mix

F040223_0127_i.pgm  #1-17  RT: 0.00-0.06  AV: 17  NL: 9.86E5
T: FTMS + p NSI Full ms [300.00-1500.00]
High accuracy/high resolution effect on a peptidic mix

**Exp 684.3800**  
**Th 684.3777**  
**Exp 684.3898**  
**Th 684.3870**

1st isotope: **VLDTGGPISVPVG**  
2nd isotope: **INVIGEPIDER**  
R : **400 000**
NanoLC-MS with parallel IT MS/MS and FT MS

Elution Gradient
0 => 50 % B in 35 min then 100% B for 10 min
A = 98% water/ 2% ACN/ 0,1% AF (v:v) ; B = 10% water/ 90% ACN/ 0,085% AF (v:v)

Automatic edition of a precursor list with their fragments

Instruments : Ultimate 3000 (Dionex), LTQ FT (ThermoFisher Scientific)
Hela digest in 30 min gradient by nanoLC LTQ MS/MS FTICR MS

At m/z 400
Hela digest in 30 min gradient by nanoLC LTQ MS/MS FTICR MS
Hela digest in 30 min gradient by nanoLC LTQ MS/MS FTICR MS

Accuracy increases with resolution
Hela digest in 30 min gradient by nanoLC LTQ MS/MS FTICR MS

Peptide inferences

<table>
<thead>
<tr>
<th>Reso</th>
<th>Peptides</th>
<th>decoy</th>
<th>FDR</th>
<th>Proteins</th>
<th>Peptides</th>
<th>decoy</th>
<th>FDR</th>
<th>Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>12500</td>
<td>4014</td>
<td>87</td>
<td>2.17</td>
<td>825</td>
<td>3636</td>
<td>115</td>
<td>3.16</td>
<td>530</td>
</tr>
<tr>
<td>25000</td>
<td>4168</td>
<td>95</td>
<td>2.28</td>
<td>821</td>
<td>4143</td>
<td>108</td>
<td>2.61</td>
<td>531</td>
</tr>
<tr>
<td>50000</td>
<td>3856</td>
<td>88</td>
<td>2.28</td>
<td>852</td>
<td>3887</td>
<td>123</td>
<td>3.16</td>
<td>556</td>
</tr>
<tr>
<td>100000</td>
<td>3863</td>
<td>110</td>
<td>2.85</td>
<td>844</td>
<td>3918</td>
<td>138</td>
<td>3.52</td>
<td>572</td>
</tr>
<tr>
<td>200000</td>
<td>3801</td>
<td>105</td>
<td>2.76</td>
<td>870</td>
<td>3822</td>
<td>125</td>
<td>3.27</td>
<td>600</td>
</tr>
</tbody>
</table>

Protein inferences

<table>
<thead>
<tr>
<th>Reso</th>
<th>Peptides</th>
<th>decoy</th>
<th>FDR</th>
<th>Proteins</th>
<th>Peptides</th>
<th>decoy</th>
<th>FDR</th>
<th>Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>12500</td>
<td>4014</td>
<td>87</td>
<td>2.17</td>
<td>825</td>
<td>3636</td>
<td>115</td>
<td>3.16</td>
<td>530</td>
</tr>
<tr>
<td>25000</td>
<td>4168</td>
<td>95</td>
<td>2.28</td>
<td>821</td>
<td>4143</td>
<td>108</td>
<td>2.61</td>
<td>531</td>
</tr>
<tr>
<td>50000</td>
<td>3856</td>
<td>88</td>
<td>2.28</td>
<td>852</td>
<td>3887</td>
<td>123</td>
<td>3.16</td>
<td>556</td>
</tr>
<tr>
<td>100000</td>
<td>3863</td>
<td>110</td>
<td>2.85</td>
<td>844</td>
<td>3918</td>
<td>138</td>
<td>3.52</td>
<td>572</td>
</tr>
<tr>
<td>200000</td>
<td>3801</td>
<td>105</td>
<td>2.76</td>
<td>870</td>
<td>3822</td>
<td>125</td>
<td>3.27</td>
<td>600</td>
</tr>
</tbody>
</table>

For the records: 3740 proteins, 19587 peptides in 120 min gradient on a Q Exactive HF
Challenges in LC-MS

- A biological sample is always complex!!
  - Thousands of analytes to identify within a large dynamic range....
- A biological sample is always contaminated!!
  - Constitutive endogenous contaminations
  - Exogenous contaminations
- MS: simple mix

Solution: Increase peak capacity in MS and in LC
Peak capacity

Max number of resolved compounds using a given analytical method in given analytical conditions. 

Giddings, Anal. Chem. 39 (1967)


Multidimensional separation
If 2 orthogonal dimensions
Golden rule: $^{2D}P_c = ^{1P_c} x ^{2P_c}$

- peak capacity
- dynamic range

Increased separation of peptides
Increased sequenced peptides
Increased sequence coverage
Increased proteome coverage
on-line 2DLC

SMBP CNRS/ESPCI USR 3149

on-line 2DLC

External valve

Fractionation column (SCX)

Injection valve

Purification cartridge 1

Collection valve V2 = sampler

loading pump 1

Waste

trap col. 1

jumper

Elution valve V1

loading pump 2

micro pump 2

ACN gradient

Elution of peptides from trap column 2 towards MS

Fractionation of eluate from 1st on trap column 1

Waste

Trap col. 2

Purification cartridge 2

Syringe

Waste

 loading pump 1

micro pump 1

salt gradient

Sample

SMBP CNRS/ESPCI USR 3149

External valve

Injection valve

Purification cartridge 1

Collection valve V2 = sampler

Elution valve V1

loading pump 2

micro pump 2

ACN gradient

Analytical column (RP)

ESI-FTMS
on-line 2DLC

No relationship between retention times of 1\textsuperscript{st} and 2d dimension

**Orthogonality**

Peak capacity 2DLC x MS/MS

\[ 2\text{DMS/MS}P_c = 2P_c \times MS/MS P_c \]
\[ 2\text{DMS/MS}P_c = 1P_c \times 2P_c \times MS/MS P_c \]
\[ 2\text{DMS/MS}P_c = 12 \times 80 \times 20 = 18000 \]
off-line 2DLC

Fractionation of 1ère dimension eluate

Elution of peptides From fractions towards MS
2DLC : off-line vs on-line

Bias against hydrophobic peptides

Example: Skin Proteome

Bias against hydrophilic peptides

2D Off-line

33% of Keratins

2D On-line

25% of Keratins

2D Off-line

60% of Keratins

2D On-line

1/3 of peptides communs

1D

Proteins inferences

SMBP CNRS/ESPCI USR 3149
Alternatives

Gradient 1DLC plus LOD detection MALDI-MS

- Simple setup
- Different ionization mode
- Robust separation
- Large peaks
- Longer acquisition time

Complementarity of analytical strategies

RP/RP MS/MS is also a good solution
Summary

• Omics and Mass spectrometry
• Why study the proteome?
• Bottom-up proteomics: an LC-MS story
• **DIA and HRM**
• Quantification
Co-eluted peptides


In most cases the most intense peptides are selected in DDA MS/MS

Minor co-eluted are ignored and information is lost

Example of 9393 proteins from Saccharomyces Cerevisiae after tryptic digestion (2 miscleav. 771753 peptides)
Other analytical strategies

DIA: Data Independant Acquisition
  SWATH™
  MS^E
  AIF (All ion Fragmentation)

HRM: Hyper Reaction Monitoring
  Multiplex SRM (Selected Reaction Monitoring)
  Pseudo SRM
  PRM (Parallel reaction monitoring)
Data Independent Acquisition (DIA)

precursor ions  precursor range  collision cell selection

high resolution detector (i.e. TOF, orbitrap)

assay matching  time

quantification

spectral library
Fourier transform-all ion reaction monitoring FT ARM

- Based on accurate peptide fragment mass measurements
- All ions fragmented in every scan

A) TIC  B) Complex fragmentation spectrum of all ions.  C) Hypothetical fragmentation spectrum of one peptide  
D) Dot product of the 2 spectra, and Score Chromatogram.

Fourier transform-all ion reaction monitoring FT ARM


http://brucelab.gs.washington.edu/
Spectral library: 2D-LC-MSMS, enrichment, cellular fractionation, long LC gradient

Quantitative analysis with a large number of conditions: untreated samples
SIM/tMS² for Data Independent Protein Quantification

**Orbitrap**
R=240,000

**Linear Ion Trap**

<table>
<thead>
<tr>
<th>SIM 400 - 600 amu</th>
<th>SIM 600 - 800 amu</th>
<th>SIM 800 -1000 amu</th>
</tr>
</thead>
<tbody>
<tr>
<td>406 m/z</td>
<td>574 m/z</td>
<td>806 m/z</td>
</tr>
<tr>
<td>418 m/z</td>
<td>586 m/z</td>
<td>818 m/z</td>
</tr>
<tr>
<td>598 m/z</td>
<td>774 m/z</td>
<td>974 m/z</td>
</tr>
<tr>
<td>606 m/z</td>
<td>786 m/z</td>
<td>986 m/z</td>
</tr>
<tr>
<td>618 m/z</td>
<td>798 m/z</td>
<td>998 m/z</td>
</tr>
</tbody>
</table>

17 sequential CID MS/MS scans with 12 Da isolation Windows

3.6 s cycles, covering 1000 m/z
DIA Workflow on Orbitrap Fusion Tribrid

Quantitation using SIM at 240K
XIC of 613.317+613.818 (Heavy GISNEGQNASIJK)

Peptide sequence confirmation using spectral library
MS/MS of 12 amu wide window
PRM vs SRM

SRM « Selected Reaction Monitoring » (QqQ)

Peptides Mix → Selection → Fragmentation → Q1 → q2 → Q3 → 1 → 2 → 3 → 4 → 5 → Peak

Sequential detection of fragments → Superposition of transitions

PRM « Parallel Reaction Monitoring » (QqOrbitrap)

Peptides Mix → Selection → Fragmentation → Orbitrap → Simultaneous fragment detection at high resolution → Superposition post-acquisition of transitions

**Targeted analysis: sample sampling**

minimum 8 - 10 points per peak

![Graph showing relative intensity vs. time with cycle time and 10% H markers.](image)

\[
\text{#peptides} = \frac{\text{temps de cycle}}{\text{temps d'acquisition}}
\]

⇒ Up to 31 peptides / cycle
(R=17 500 at m/z 200, 64ms/acquisition)

Q Exactive (Thermo Scientific)
The case of amyloidoses

Group of diseases
- Extracellular deposit (aggregate) of insoluble misfolded proteins
- Hereditary (genetic mutation) or senile (aging)
- Life-threatening organ failure, e.g. myocardial wall thickening

Complex mechanisms

Native precursor protein
- Ex: Serum amyloid A (SAA)

β-sheets rich structure

Amyloid fibrills

Amyloid deposits

Common components:
- Glycosaminoglycans
- Serum amyloid P-component (SAMP)
- Apolipoprotein E (APOE)

Sophie Liuu, Emmanuelle Demey
Gilles Grateau, David Buob (Hop Tenon)

DDA MS/MS mode: automated hierarchical clustering

Patients

Controls

Intensity scale of peptides (area)

Data processing
Maxquant/Perseus

3 technical replicates
### Targeted vs non-targeted analysis

#### Marqueurs amylose

<table>
<thead>
<tr>
<th>Proteine</th>
<th>Peptide</th>
<th>Aire SRM</th>
<th>Aire PRM</th>
<th>Aire TOP N PD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAMP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYSC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAA1/SAA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAA4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGKC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTHY</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOA4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIBA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Classificateurs amylose

<table>
<thead>
<tr>
<th>Proteine</th>
<th>Peptide</th>
<th>Aire SRM</th>
<th>Aire PRM</th>
<th>Aire TOP N PD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGHG1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGHG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGHM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGHM/MUCB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRFL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BGH3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GELS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFGM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Intensity scale of peptides (area)**
Summary

- Omics and Mass spectrometry
- Why study the proteome?
- Bottom-up proteomics: an LC-MS story
- DIA and HRM
- **Quantification**
Why Quantitative Proteomics?

Most biological issues cannot be solved by the identification of one protein.

In most cases the variation of the abundance of a protein together with its modification state is required.

Mass spectrometry is not quantitative

- Due to non uniform instrumental response mass spectrometry cannot be used directly for quantitation of macrobiomolecules

- The intensity of a peptide is not a direct measure of its abundance

- Comparison of chemically identical species:
  - Two peptides with a different isotopic composition within one single run: Isotopic labelling
  - The same peptide detected in 2 different runs: Label free quantification
Quantification bottom-up

- Technological limitations to detect and differentiate the isotopic labels
- Bioinformatics and statistics limitation to process the data with a correct normalization

S-E Ong & M Mann; Nature Chemical Biology 1, 252 - 262 (2005)
M Bantscheff, M Schirle, G Sweetman, J Rick & B Kuster; Anal Bioanal Chem 389(4), 1017-1031 (2007)
In vivo Metabolic Labeling Stable Isotope Labeling by Amino Acid in Cell Culture (SILAC)

Figure 1. SILAC with $^{13}$C6-Arg. One cell population is cultured in $^{13}$C6-Arg labeling medium, whereas the other population is grown in normal arginine. In a typical SILAC strategy, the cells are combined and then lysed; however, they can also be lysed separately and mixed at a later stage of the experiment. Proteins are digested with trypsin and analyzed by LC−MS/MS.
Reporter ion relative quantification

Possibilité d’enrichissement des peptides marqués

TMT (Tandem Mass Tag)
iTraq
10 and 11-plex TMT: high resolution MS/MS required
Overview of AQUA Strategy (Developed by Gygi et al)

1. Select peptide from protein of interest
2. Synthesize analogous peptide incorporating stable isotope ($^{13}$C, $^{15}$N) labeled amino acid
3. Proteolysis with enzyme to generate peptide of interest
4. Optimize separation and/or analysis scheme for quantitation by mass spectrometry
5. Analyze using mass spectrometry and quantitate by comparison of native peptide signal to AQUA peptide signal

VPQVSTPTLVEVSR

VPQVSTPTLVEVSR*
Intensity-measurement uncertainty correlation

Corrélation avec le rapport signal sur bruit (S/B)

Codage des Arg et Lys

Mélange 1:1

~1/3 des ions à S/B < 5

Targeted analysis improves the S/N

J Proteome Res 2008, 7, 4756-65, Bakalarski et al.
CHALLENGES OF PROTEIN PTM ANALYSIS BY MS

PTMomics and relative quantification

What do we want to know?

- Sample 1 vs. Sample 2
- Modified vs. unmodified
- PTM1 vs. PTM2
- Absolute titration
Cysteine redoxome: a case study for quantification

The case of reversible oxidation of Cys

Reactive oxygen and nitrogen species (ROS/RNS)

- Low concentrations
  - regulation physiological processes,
  - reversible

- High concentrations:
  - Deleterious process (~burning, rusting); e.g. ageing
  - irreversible damage to DNA, proteins and lipids

Different redox states of protein cysteines

From Antiox. & Redox Signal., 26(7), 2017


**Challenges of Protein PTM analysis by MS**

**Redox proteomics strategies**

**Differential labelling of Cys residues**

**Enrichment of oxidized Cys residues**

**OxICAT and ICAT based strategies** (Leichert PNAS 2008, Fu MCP 2009, Garcia-Santamarina Nat Protoc 2014)

**CysTMT, iodoTMT and OxiTMT** (Behring FASEB 2014, Murray MCP 2012, Shakir Anal Bio Chem 2017)

**OcSILAC** (Chiappetta HUPO 2011, Shakir submitted)
Challenges of protein PTM analysis by MS

OxiTMT: Label-based relative quantification

TMT1/TMT3: oxidized Cys fold change
TMT2/TMT4: protein fold change

Shakir Anal Bio Chem 2017
Challenges of protein PTM analysis by MS

PTMs quantification specificities

1 single peptide

\[
\frac{I_{\text{mod peptide}^2}}{I_{\text{mod peptide}^1}} > \frac{I_{\text{protein}^2}}{I_{\text{protein}^1}}
\]

TMT1/TMT3: oxidized Cys fold change

All proteolytic peptides

\[
\frac{I_{\text{protein}^2}}{I_{\text{protein}^1}}
\]

TMT2/TMT4: protein fold change

Modification or expression level?

\[
\begin{align*}
\frac{I_{\text{mod peptide}^2}}{I_{\text{mod peptide}^1}} &= 2 \\
\frac{I_{\text{protein}^2}}{I_{\text{protein}^1}} &= 2 \\
\frac{I_{\text{mod peptide}^2}}{I_{\text{mod peptide}^1}} \times \frac{I_{\text{protein}^1}}{I_{\text{protein}^2}} &= 1
\end{align*}
\]

1% to 3% or 30% to 90% ????
**Challenges of protein PTM analysis by MS**

**Label based site occupancy**

**Site Occupancy**

\[
\frac{N \text{ modified pept}}{N \text{ total pept}}
\]

- **TMT1/TMT3**: oxidized Cys fold change
- **TMT2/TMT4**: protein fold change
- **TMT1/TMT2**: oxidized CYS site occupancy  
  **SAMPLE 1**
- **TMT3/TMT4**: oxidized CYS site occupancy  
  **SAMPLE 2**

Olsen J.V. & Mann M., Sci. Signal. 2010
Shakir Anal Bio Chem 2017
**E. Coli** redox proteomics analysis by OxiTMT

E. Coli model before (1) and after (2) 30min 1mM H2O2

Quantification: 1229 iodoTMT Cys, 580 proteins

- bound fraction: 886 peptides (1019 Cys), 487 proteins (172 specific).
- unbound fraction: 834 peptides (893 Cys), 408 proteins (93 specific)

25 up-regulated proteins in H2O2 treated cells
- oxidoreduction and generation of precursor
- metabolites and energy pathways

18 down-regulated oxidized fraction

Araki K. J. Proteome Res 2016
Considering some risks...

Where do the biases come from?

- Sample preparation (biological replicates)
- Sampling
- LC-MS (separation, ionization, matrix effects)
- MS sampling and scan events
- Quantification of each peptide from runs
- Quantification of proteins from peptides
Bias from peptide to peptide
**Challenges of Protein PTM Analysis by MS**

**Mini signals, mini results**

Co-isolation of peptides is less frequent

Specific challenge of ion reporter PTM quantification compared to protein quantification

Error increase for low abundance
Shift of experimental ratios towards higher values

Biased quantification of low abundance species.

Oxidized unabundant cysteine reporter ions are more affected than total cysteine reporter ions.
Protein coverage: Sampling and specificity

Bottom up partial characterization of the species of interest might lead to misinterpretation...
Two parallel and complementary techniques

Top-down or Bottom-up?

We known how to purify, fragment and sequence peptides with high sensitivity and high throughput (like « pros »!)

- **Bottom-up approach**: sequential processes each providing necessary information to describe a whole.
  - identification of proteins, sequencing and PTM characterization from proteolytic peptides

- **Top-down approach**: processes starting from the raw material aiming to transform and modify it to simplify its description.
  - identification of proteins and global structural analysis (sequence, PTMs) without proteolysis


For intact proteins, we are still beginners...
Acknowledgments

- SMBP / ESPCI
  - Giovanni Chiappetta
  - Shakir Shakir
  - Massamba Ndiaye
  - Martha Zoupoulaki
  - Nicolas Eskenazi
  - Emmanuelle Demey-Thomas
  - Yann Verdier
  - Iman Haddad
  - Sophie Liuu
  - Anne Marie Hesse
  - Sega Ndiaye

- Oxidative Stress & Cancer Lab CEA
  - Michel TOLEDANO
  - Alise PONSERO

- Applied Statistics
  - ESPCI
  - Isabelle RIVALS