PAF introduction
The PAF (also called UNIL Proteomics Platform) is a research and service core facility devoted to the high-performance analysis of proteins.

**Service:**
- Project discussion
- Identification of gel-separated proteins by MS
- Identification and quantitation of proteins in complex samples by multi-dimensional separation techniques
- Targeted measurements of protein and phosphoprotein levels with antibody detection

**Research:**
- Development of proteomics methods
- Applications to biological problems

**Teaching:**
- Techniques, possibilities & limitations of proteomics approaches
New and old tools

- Protein separation techniques
  - Liquid chromatography
  - Electrophoresis
  - ...
- Protein identification techniques
  - Mass Spectrometry
  - Antibody-based techniques
- Protein quantification techniques
  - Antibody based techniques
  - Dye-binding techniques
  - Mass Spectrometry

- Genome sequence databases
- Protein sequence databases
- Biological knowledgebases:
  - functions
  - pathways
  - seq. motifs
  - 3D structures
Challenges of *in vivo* proteomics

- **Complexity**:  
  - 20’400 genes in *H. sapiens*,  
  - 15’000 (?) in a single cell  
  - >100’000 chemically distinct protein species?

- **Dynamic range**:  
  - $10^5$ x to $10^{10}$ x between low and high-abundance proteins

- **Plasticity**:  
  - continuous variation in expression levels, PTM’s, degradation,…

- Proteins…  
  - have vastly different physico-chemical properties  
    - *(acidic, basic, hydrophilic, hydrophobic, …)*  
  - cannot be amplified ….
PAF technology and instrumentation - 1

1D-electrophoresis  2D-electrophoresis

Liquid Chromatography

Thermo LTQ-Orbitrap Velos Pro  Thermo Q-Exactive Plus  Thermo Orbitrap Fusion

Nano-HPLC-Orbitrap
Mass spectrometers
PAF technology and instrumentation - 2

RPPA slide spotter: Nano-Plotter 2.1 (GeSiM)

RPPA slide incubations: Biomek 3000 (Beckman-Coulter)

RPPA slide detection: InnoScan 710 (Innopsys)
Main proteomics workflows @ PAF

• Protein identification by nanoLC-MS/MS

• Shotgun analysis of complex samples

• Protein quantification:
  – SILAC
  – Multiplex quantitation (iTRAQ)
  – Label free (LFQ)

• PTM analysis:
  – Phosphoproteomics

• Targeted analysis (RPPA)
  – Levels of proteins and phosphoproteins
WORKFLOW 1:

Protein identification by nanoLC-MS/MS
Protein ID by 2D-gel/SDS-PAGE and nanoLC-MS/MS

Biological question

Output:
- Protein identification in simple/complex mixtures
- Extensive sequence coverage and peptide mapping
- Analysis of modified peptides possible

Database matches
- DHX9_HUMAN
- NFM_HUMAN
- Q9BQG0
- MY06_HUMAN
- TP2A_PIG
- Q7Z5Y2
- FLIH_HUMAN
- TP2B_MOUSE
- S3B1_HUMAN
- Q8VCW5
- Q8CHF9

ATP-dependent RNA helicase A
Neurofilament triplet M protein
Hypothetical protein
Myosin VI.
DNA topoisomerase II, alpha isozyme
Rho-interacting protein 3.
Flightless-I protein homolog.
DNA topoisomerase II, beta isozyme
Splicing factor 3B subunit
Similar to alpha internexin neuronal
MKIAA0376 protein (Fragment).

Q7Z5Y2
- Mass: 118789
- Total score: 178
- Peptides matched: 6
- Rho-interacting protein 3.

Peptide sequence:
- Q8VCW5
- Similar to alpha internexin neuronal MKIAA0376 protein (Fragment).

Output:
- Protein identification in simple/complex mixtures
- Extensive sequence coverage and peptide mapping
- Analysis of modified peptides possible
WORKFLOW 2:

Shotgun analysis of complex samples
### Shotgun sequencing from complex mixtures

1. **Multiprotein complex**
2. **Denaturation, Proteolytic digestion**
3. **Complex peptide mixture (1000-20000 species)**
4. **Nano rp-LC-MSMS**
5. **Db search**

### List of identified proteins

1. P45218
2. P21543
3. Q12588
4. P32651
5. Q01245
6. ....
Fractionation to reduce complexity (1)

GeLC-MS workflow

Complex mixture

Trypsin digestion

LC-MS/MS

Db search

Protein IDs
1. P45218
2. P21543
3. Q12588
4. P32651
5. Q01245
....
....
1545. Q34258
Ex: protein-protein interactions analysis by affinity purification
Fractionation to reduce complexity (2)

Peptide-based fractionation

Complex mixture

Trypsin digestion

Peptide fractionation

LC-MS/MS

Db search

Protein IDs
1. P45218
2. P21543
3. Q12588
4. P32651
5. Q01245
...
...
1545. Q34258

SCX LC or Peptide IEF
Data analysis and distribution software: Scaffold

Export to Excel

Quantitative analysis

Filtering parameters

Protein ID probability

Percent Coverage

Percentage of Total Spectra

Exclusive Unique Peptide Count

Exclusive Unique Spectrum Count

Exclusive Spectrum Count

Total Spectrum Count

Quantitative Value

Link to UniProt annotations

Free Scaffold viewer: www.proteomesoftware.com
WORKFLOW 3:

Protein quantification
Techniques for large scale quantitative proteomics

2D Electrophoresis

- Classical (Label-free)
- DIGE (Labelling)

MS-based methods

- Label-free methods
  - MS-based
  - Spectral counting
  - Targeted SRM/MRM

Labelling methods

- Chemical Labelling
  - ICAT
  - iTRAQ
  - ICPL
  - TMT
  - Di-ME
  - $^{18}\text{O}$
  - $^{15}\text{N}$
  - ...

- Metabolic Labelling
  - SILAC
  - ISIS
  - ...

➢ quantification at protein level

➢ quantification at peptide level
**Metabolic labelling**

- Sample A: Cells grown in light isotope-containing medium
- Sample B: Cells grown in heavy isotope-containing medium
- Harvest cells
- Mix cell lysates
- Trypsin digestion
- LC-MS/MS

**Relative intensity**

- Light: Sample A — Light, Sample B — Heavy
- Heavy: Sample A — Heavy, Sample B — Heavy

**Chemical labelling**

- Sample A: Protein extraction and trypsin digestion
- Sample B: Protein extraction and trypsin digestion
- Light labeling
- Mix A and B
- LC-MS/MS

**Label free**

- Sample A: Protein extraction and trypsin digestion
- Sample B: Protein extraction and trypsin digestion
- LC-MS/MS

**Labeling**

- Analytical variability minimized
- Number of samples limited (2-8)

**Label free**

- Number of samples unlimited
- Simpler sample preparation
- Analytical variability
- Computationally heavy (XIC)
Relative quantification by stable isotope labelling

Sample A
Light

Sample B
Heavy

Mix
analyse

Labelling strategies:

- Chemical (side chains: C, K, N-term)
  ICAT, iTRAQ, ICPLP,…

- Metabolic (K, R, all)
  SILAC,…

- Enzymatic
  Trypsin + 18O, …

Co-analyse
Eliminate analytical variability

Δm
SILAC experiment workflow

A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC).

Data analysis software!

MaxQuant
Chemical labelling:
Isobaric Tags (iTRAQ)- multiplex quantification

Control

Treated

Harvest cells

Trypsin digestion

< in vitro chemical labeling

< Same peptide from 4 samples has same mass (isobaric)

< quantification by tags in MS/MS spectra at fixed M/Z

Multiplexed Protein Quantitation in *Saccharomyces cerevisiae*
Using *Amine-reactive Isobaric Tagging Reagents*

Figure 1. The concept of iTRAQ™ Reagent chemistry (example of a 4-plex experiment) Each sample is labeled with one of the four iTRAQ Reagents and then pooled prior to MS analysis.
How to label? Pros and cons

- **Metabolically** (during protein synthesis)
  - Incorporation of one or more labelled amino acid
    - (+) “native” proteins
    - (+) compatible w. purifications
    - (+) accurate
    - (-) need cultivatable organism
    - (-) limited multiplexing (max. 3)

- **Chemically** (post protein synthesis)
  - “specific” chemical modification of AA side chain
    - (+) any sample can be done
    - (+) higher multiplex (iTRAQ max 8-plex)
    - (-) side (or incomplete) reactions
    - (-) separate purifications
    - (-) less accurate
## Quantitation summary

<table>
<thead>
<tr>
<th>Application</th>
<th>Multiplexing</th>
<th>Accuracy (process)</th>
<th>Quantitative proteome coverage</th>
<th>Linear dynamic range&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Ease of use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolic protein labeling</td>
<td>2-3</td>
<td>+++</td>
<td>++</td>
<td>1–2 logs</td>
<td>+</td>
</tr>
<tr>
<td>• Complex biochemical workflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Cell culture systems only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical protein labeling (MS)</td>
<td>2-3</td>
<td>+++</td>
<td>++</td>
<td>1–2 logs</td>
<td>+</td>
</tr>
<tr>
<td>• Medium to complex biochemical workflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical peptide labeling (MS)</td>
<td>2-3</td>
<td>++</td>
<td>++</td>
<td>2 logs</td>
<td>+</td>
</tr>
<tr>
<td>• Medium complexity biochemical workflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical peptide labeling (MS/MS)</td>
<td>2-8</td>
<td>++</td>
<td>++</td>
<td>2 logs</td>
<td>+</td>
</tr>
<tr>
<td>• Medium complexity biochemical workflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enzymatic labeling (MS)</td>
<td>2</td>
<td>++</td>
<td>++</td>
<td>1–2 logs</td>
<td>++</td>
</tr>
<tr>
<td>• Medium complexity biochemical workflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spiked peptides</td>
<td>multiple</td>
<td>++</td>
<td>+</td>
<td>2 logs</td>
<td>++</td>
</tr>
<tr>
<td>• Medium complexity biochemical workflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Label free (ion intensity)</td>
<td>multiple</td>
<td>+</td>
<td>+++</td>
<td>2–3 logs</td>
<td>++</td>
</tr>
<tr>
<td>• Simple biochemical workflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Whole proteome analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Label free (spectrum counting)</td>
<td>multiple</td>
<td>+</td>
<td>+++</td>
<td>2–3 logs</td>
<td>+++</td>
</tr>
<tr>
<td>• Simple biochemical workflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Whole proteome analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> In MRM mode, dynamic range may be extended to 4–5 logs

WORKFLOW 4:

PTM analysis
f.ex. Phosphoproteomics
Caveat

Protein identification

IS NOT

protein characterisation

Two peptides are enough to identify a protein but we are still identifying two peptides, not the entire protein

Highly similar sequences cannot be distinguished

For finding PTMs extensive sequence coverage is essential !!!
PTM characterization by MS

\[ \text{ETYGDMADCCEK} \]

\[ \text{ETYGD} + \text{MoxADCCEK} \]
Phosphoproteomics by TiO$_2$ P-peptide enrichment

Variants / options

• SILAC label
• $\alpha$-P-Tyr protein enrichment
• Asp, Glu esterification

ID: $P$-proteins $P$-sites
WORKFLOW 5:

Targeted protein quantification
RPPA: Reversed Phase Protein Array

- Samples spotted on slide
- Antibodies applied for detection
RPPA pipeline

1. Spotting
   - Non contact piezoelectric spotting
   - 0.8 nL sample volume
   → 2-10 cell equivalents per spot
   - 3D nitrocellulose surface
   → optimal protein adsorption
   - Up to 2000 spots/slide

2. Detection
   - Blocking
   - 1ry and 2ry antibodies
   - Signal amplification

3. Imaging

4. Analysis
   • SuperCurve

Near Infra Red
RPPA results

RPPA

Samples

Dilutions

300 µm

0 30s 1' 5' 15' 1h 3h

Pixel intensity

65'000 0

1/16

Comparison with Western blot