“Omics” Methods in Redox Biology

Jiri Adamec
What is the “OMICS”

Latin – *ome*, mass or body of

Middle English – *some*, a group of

-**ome** Complete set of molecules in an organism or cell at given time

-**omics** The global analysis of molecules in biological systems.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Genomics</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA</td>
<td>Transcriptomics</td>
<td>Transcriptome</td>
</tr>
<tr>
<td>Proteins</td>
<td>Proteomics</td>
<td>Proteome</td>
</tr>
<tr>
<td>Ions</td>
<td>Ionomics</td>
<td>Ionome</td>
</tr>
</tbody>
</table>
The Most Important “OMICS”

ECONOMICS!!!
Targeted vs. OMICS Approach

Targeted Analysis:

- Specific to a single molecule (or a small set of the molecules)
- In depth characterization of biophysical, biochemical and structural properties
- Very detailed however doesn’t reflect cell/organism complexity
- Research is often isolated and knowledge assembly is difficult (too many different experimental conditions from various labs)
- May lead to simplistic conclusions (e.g. new drug design vs. side effects)
Targeted vs. OMICS Approach

OMICS Analysis:

- Attempt to determine as many molecules as possible
- Great concept however current technology is limiting factor
- Except of transcriptome we don’t know how many molecules are in cell/organism at given time
- Difficult to sort very complex data - bioinformatics
Solution

Combination of OMICS and Targeted Analysis:

- OMICS is very useful to
  - setup matrix (base) for targeted analysis
  - place individual findings in context
  - suggest new targets and/or hypotheses
Our goal – High resolution picture
But be careful with data interpretation
Genes, mRNAs, Proteins and Metabolites - Relationship

- Transcription
- Translation
- PTM
- Protein complex
- Misfolding

Metabolites
OMICS and Systems Biology

Genomics → Proteomics → Metabolomics

Transcriptomics

Functional Proteomics/Genomics

Systems Biology
Is a field in biology that aims at system level understanding of biological systems, where a bunch of parts that are connected to one another and work together. It attempts to create predictive models of cells, organs, biochemical processes and complete organisms.

The ultimate goal of system biology is to develop **in-silico biosystems**

Remember? Hegel developed complex philosophical system concept and System Science
High through-put analysis
Transcriptome
Metabolome
Proteome

Statistical analysis
$t$-test (Volcano plot)

Fold change

95% significance

COSA Clustering

Principal Component Analysis

Pathway analysis
Metabolic Pathways
Physiological Pathways
Correlation Network
Proteomics

- Proteome may be thought of and studied in three ways:
  - One is in terms of the proteins present. Efforts to define or catalog all the components in a proteome are referred as “Cataloging Proteomics”
  - The other way to think of a proteome is in terms of its quantitative features and changes in the relative concentration of proteins as a function of a stimulus or genetic change. This is referred to as “Comparative Proteomics”
  - Identification of proteins in the cell and the determination of their role in physiological and pathophysiological functions is called “Functional Proteomics”
Approaches to Proteomics

**“Bottom-Up” based proteomics**
- Sample preparation
- Proteolysis
- Separation (e.g. 2D LC)
- Quantification (MS)
- Statistical analysis
- Peptides selection
- Peptide ID (MS/MS)

**“Top-down” based proteomics**
- Sample preparation
- Separation (e.g. 2D gel)
- Quantification
- Statistical analysis
- Proteins selection
- Proteolysis
- Protein ID (MS/MS)
“Bottom-Up” based proteomics – Workflow

Sample collection → Delipidation → Abundant protein removal → Trypsin digestion → Isotopic labeling → Sample clean up → Separation and Detection → MudPIT

HPLC column → SDS-based → GIST → iTRAQ → C-18 column → Reversed phase C-18 → ESI ion trap

MALDI TOF-TOF → ESI Q-TOF

Peak alignment → Normalization → Pattern recognition

Data mining and analysis
One-step “Shotgun” approach

One-step (shotgun) approach – quantification and identification is done in single run through cycling steps: 1/ MS survey scan; 2/ identification of 3-5 most abundant peaks; 3/ MS/MS of those peaks.

- It is random, time limited, process capable of ~15-20 MS/MS scans per 20 s (elution time of individual peptide)
- Targets abundant peaks only
- Vast majority identified peptides is not up/down regulated
- Many important low abundant peaks with significant up/down regulation is missed
- Great for quick screening

Two-step approach – samples are first run in MS only mode for quantification. Following statistical analysis only up/down regulated peaks are subjected to targeted MS/MS for identification.

- Much more accurate then shotgun
- Capable to identify and quantify low abundant proteins
- Requires two runs of same sample

Options and considerations when selecting a quantitative proteomics strategy

Bruno Domon & Ruedi Aebersold Affiliations

Nature Biotechnology 28, 710–721 (2010) doi:10.1038/nbt.1661 Published online 09 July 2010
Shot Gun Problem – too many peptides
Sample complexity, the Achilles heel of “Bottom-up” - LC/MS based proteomics

• The average protein will yield 50 tryptic peptides.
• This means a tryptic digest of the average proteome can have 300,000 to 1,500,000 peptides.
• When a sample of this complexity is run on an RPC column with a peak capacity of 300 each fraction will have roughly 1,000 peptides.
• MS instruments can not handle this level of complexity.
These numbers are huge, what’s the solution?

- Reduce sample complexity
  - Subcellular fractionation
  - Add more LC dimensions
  - Affinity selecting part of the digest through PTMs (Functional Proteomics)
1 Dimensional LC-MS Approach

Reversed phase C-18 LC/MS (RPC/MS): 5 proteins mix

- Peak capacity is ~ 5,000
- Great for preliminary studies (identify ~ 300 proteins)
Multi dimensional chromatography - MudPIT

1st Dimension

SCX

2nd Dimension

NaCl Fractions

50 mM → C18-RPC
75 mM → C18-RPC
100 mM → C18-RPC
125 mM → C18-RPC
150 mM → C18-RPC
175 mM → C18-RPC
250 mM → C18-RPC
500 mM → C18-RPC

• Peak capacity is ~ 30,000 (depends on fractions number
• Identify ~ 1,300 proteins)
Issue #2 – Label-free Quantification

- Works well but has some potential issues
  - Strongly depends on LC-MS system reproducibility
  - Intensity of any peak is not only function of peptide concentration. It also depends on analyte composition (ion suppression)
- Some of these potential issues could be overcome by using isotopic labeling strategies
Stable isotope labeling strategies

• **Post-biosynthetic labeling.**
  – Labeling amino groups (GIST and iTRAQ)
  – Labeling cysteine residues (ICAT)

• **Biosynthetic labeling**
  – *In vivo* incorporation of isotopic labeled species - growing cells in media enriched in $^{12}\text{C}$ vs. $^{13}\text{C}$ (SILAC)
  – Synthetic peptides with incorporated $^{15}\text{N}$ (AQUA)
Global Internal Standard Technology (GIST)

Basic principle of GIST is that peptides from control and experimental samples are modified by two chemically identical but isotopically distinct labeling reagents. After mixing these two samples, each peptide from the control serves as an internal standard in experimental sample for determining the relative concentration of this peptide.
GIST – Chemical Structure of Labeling Reagents

Acetate-based reagent

\[
\begin{align*}
\text{CH}_3 & \quad \text{C} \quad \text{O} \quad \text{N} \\
\text{O} & \quad \text{O} & \quad \text{N} \\
\text{CH}_3 & \quad \text{C} \quad \text{O} \quad \text{N}
\end{align*}
\]

\[H = ^1\text{H}\]

Light forms

\[C = ^{12}\text{C}\]

Heavy forms

\[H = ^2\text{H}\]

\[C = ^{13}\text{C}\]

Propionate-based reagent

\[
\begin{align*}
\text{CH}_3 & \quad \text{CH}_2 \quad \text{C} \quad \text{O} \quad \text{N} \\
\text{O} & \quad \text{O} & \quad \text{N} \\
\text{CH}_3 & \quad \text{CH}_2 \quad \text{C} \quad \text{O} \quad \text{N}
\end{align*}
\]

[diagram]
Amino Groups Are Easily Alkylated

All primary amino groups are labeled.

Note that Arg is not acetylated.

Trypsin cleaves polypeptides C-terminal to lysine and arginine
Labeling Samples from Two Different Sources

After derivatization these samples are mixed.

An internal standard is created for each peptide.

Sample 1 – peptide A
peptide-NH$_2$ + NHS-CO-CH$_3$
MW 500

Sample 2 – peptide A
peptide-NH$_2$ + NHS-CO-CD$_3$
MW 500

peptide from control sample labeled with light form

peptide from experimental sample labeled with heavy form

peptide-NH$_2$ + NHS-CO-CH$_3$ peptide-NH$_2$ + NHS-CO-CD$_3$
MW 542 MW 545
Multiplex example of the iTRAQ protocol

Quantitative analysis of protein expression using iTRAQ and mass spectrometry
Ry Y Tweedie-Cullen & Magdalena Livingstone-Zatchej
Isabelle M. Mansuy Group (HIFO)

The derivatized peptides are indistinguishable in MS, but exhibit intense low-mass MS/MS signature ions that support quantitation
ICAT (Isotope-Coded Affinity Tag) labeling strategy


- ICAT, designed in Prof. Ruedi Aebersold’s lab, consists of alkylating the cysteines of proteins with either a light or a heavy tag, which respectively contains protons or deuteriums (H or D), or alternatively $^{12}$C or $^{13}$C

- ICAT allows for selection of cysteine containing peptides
Determining the oxidation state of protein thiol using ICAT technology (OxICAT).

Leichert L I et al. PNAS 2008;105:8197-8202
SILAC (Stable Isotope Labeling with Amino acids in Cell culture)

SILAC was developed in CEBI (Ong et. al. MCP 2002) as a simple and accurate approach for mass spectrometric (MS)-based quantitative proteomics. The method relies on the incorporation of amino acids with substituted stable isotopic nuclei (in this case deuterium $^2$H, $^{13}$C, $^{15}$N).

SILAC-labeling of tissue samples are not possible!

Incorporation of the labeled amino acid (d3) over five timepoints, showing the eventual replacement of unlabeled d0 form of the peptide by the d3 form after 5 days of adaptation. (from Ong et. al 2002, MCP ) Asterix (*) marks an unrelated peptide.
From the Sigma AQUA Peptide Library, select and order an optimal tryptic peptide corresponding to your protein of interest.

Optimize your LC-MS/MS protocol to resolve and monitor native and corresponding AQUA peptides.

Add a known amount of AQUA Peptide to your biological protein sample.

Digest with trypsin.

Analyze by LC-MS/MS. Generate extracted ion chromatograms for native and AQUA peptides, and calculate the amount of native peptide present.

This method was developed by Dr. Steve Gygi and colleagues at Harvard Medical School [Stemmman O, Zou H, Gerber SA, Gygi SP, Kirschner MW; Dual inhibition of sister chromatid separation at metaphase, Cell 2001, Dec 14, 107: 715-726]
“Top-down” based proteomics

- Sample preparation
- Separation (e.g. 2D gel)
- Quantification
- Statistical analysis
- Proteins selection
- Proteolysis
- Protein ID (MS/MS)
Serum sample preparation and 2D-ELFO

Fractions contain low abundant proteins

Protein assay

Database

Search

MALDI digestion

Trypsin

Spots of interest

PDQuest SYPRO Ruby

Top-Down Proteomics Separation

Agilent Spin Cartridge

Desalt

Concentrate

Aliquoting

IEF

pI 3 10

12% acrylamide gel

Stain

SDS-PAGE

100 µg protein

Database

MALDI digestion

Trypsin

Spots of interest

PDQuest SYPRO Ruby

Top-Down Proteomics Separation

Agilent Spin Cartridge

Desalt

Concentrate

Aliquoting

IEF

pI 3 10

12% acrylamide gel

Stain

SDS-PAGE

100 µg protein

Database

MALDI digestion

Trypsin

Spots of interest

PDQuest SYPRO Ruby

Top-Down Proteomics Separation

Agilent Spin Cartridge

Desalt

Concentrate

Aliquoting

IEF

pI 3 10

12% acrylamide gel

Stain

SDS-PAGE

100 µg protein

Database

MALDI digestion

Trypsin

Spots of interest

PDQuest SYPRO Ruby

Top-Down Proteomics Separation

Agilent Spin Cartridge

Desalt

Concentrate

Aliquoting

IEF

pI 3 10

12% acrylamide gel

Stain
Abundant protein removal (APR)

"Top-Down" Proteomics Separation

Crude serum (210 µg)

Serum after APR (100 µg)
Points to consider about 2-D GE

- Post-translational modification (PTM) and ragged ends created variants that can differ in electrophoretic mobility, i.e. the same protein will appear in multiple spots. (phosphorylation and sialic acid variants of glycoproteins)
- PTM variants may not vary in charge and not be recognized. (neutral sugar glycoprotein variants, methylation, disulfide)
- Quantification by staining is not linear.
- Proteins must be soluble to be separated in the 2-D gel system. Membrane proteins are difficult to get into solution.
- It can take days to complete an analysis in the 2-D gel mode.
- Reproducibility is an issue.
- 2-D GE is often said to be an art. Everyone can get data, but getting great, reproducible data requires experience.
Functional Proteomics – Oxidative Stress Study
Protein Damage under Oxidative Stress

Goal is to determine stressor specific protein damage and to establish spatial models of cellular response to various forms of stress.
Frataxin – Protein Responsible for Friedreich Ataxia

Mitochondrial protein frataxin encoded by FRDA gene.

- Number of GAA triplet repeats is critical for FRDA development (60-1800 in patients)
- GAA extension interferes with transcription

Friedreich ataxia
- Neurodegenerative diseases
- Autosomal recessive (1:50,000)
- Onset around 15 years of age
- Terminated with death
- Progressive gait & limb ataxia
- Hypertrophic cardiomyopathy
- Loss of vibratory sense
- Muscle weakness
- Diabetes mellitus

Exon 1  Exon 2  Exon 3  Exon 4  Exon 5  Exon 6  Exon 7

\[(GAA)_n\]  Frataxin deficiency

Campuzano et al, 1996
Proposed Frataxin Function

- **Heme**
- **Frataxin**
- **Rust**
- **Fe/S**
- **Fe$^{2+}$ in bio-available form**
- **Cytoplasm**
- **Fe$^{3+}$**
- **OH$^-$**
- **H$_2$O$_2$**

No Frataxin
How to Pick a Few Figs from Thistles (Without Breaking the Bank)
Look for Unique Features
Carbonyl contents and strategy for LC-MS based Identification of carbonylated proteins

Sample preparation

Biotinylation of carbonylated proteins

Affinity selection with avidin

Trypsin digestion

nanoRPLC-MS/MS (ion trap)

Peptide/Protein ID

Wild type

\[ \Delta yfh1 \]

\[
\text{NH}_2 + \text{O} = \text{C} - \text{protein}
\]

\[
\text{N} = \text{CH} - \text{protein}
\]
### Functional and Spatial Distribution of Identified Proteins

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Wild</th>
<th>ΔYFH1</th>
<th>Localization^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC1, Acetyl-CoA carboxylase</td>
<td>✓</td>
<td>✓</td>
<td>C, ER, N, M</td>
</tr>
<tr>
<td>SPE3, Spermidine synthase</td>
<td>✓</td>
<td></td>
<td>C, N, M</td>
</tr>
<tr>
<td>VPS13, YLL040c</td>
<td>✓</td>
<td></td>
<td>C, M, E</td>
</tr>
<tr>
<td>TOP2, Essential type II topoisomerase</td>
<td>✓</td>
<td>✓</td>
<td>N, M</td>
</tr>
<tr>
<td>MIS1, Mitochondrial C1-tetrahydrofolate synthase</td>
<td>✓</td>
<td></td>
<td>M</td>
</tr>
<tr>
<td>ILV6, Regulatory subunit of acetolactate synthase</td>
<td>✓</td>
<td></td>
<td>M</td>
</tr>
<tr>
<td>MRE11, Subunit of a complex with Rad50p and Xrs2p</td>
<td>✓</td>
<td>✓</td>
<td>C, N, M</td>
</tr>
<tr>
<td>PET127, Protein with a role in mitochondrial RNA stability and/or processing</td>
<td>✓</td>
<td></td>
<td>M</td>
</tr>
<tr>
<td>RIB3, DHBP synthase</td>
<td>✓</td>
<td></td>
<td>C, N, M</td>
</tr>
<tr>
<td>GDH2, NAD(+) dependent glutamate dehydrogenase</td>
<td>✓</td>
<td></td>
<td>C, M</td>
</tr>
<tr>
<td>MTC1, YIL123C</td>
<td>✓</td>
<td>✓</td>
<td>C, G, ER, M</td>
</tr>
<tr>
<td>UTP10, Nucleolar protein</td>
<td>✓</td>
<td></td>
<td>N, M</td>
</tr>
<tr>
<td>CFD1, Iron-sulfur cluster binding protein</td>
<td>✓</td>
<td></td>
<td>C, M</td>
</tr>
<tr>
<td>SYG1, Plasma membrane protein</td>
<td>✓</td>
<td></td>
<td>MEM, C, M, V</td>
</tr>
<tr>
<td>SSE2, HSP70 family</td>
<td>✓</td>
<td></td>
<td>C, M</td>
</tr>
<tr>
<td>SAC1, Phosphatidylinositol (PI) phosphatase</td>
<td>✓</td>
<td></td>
<td>CP, C, ER, G, M, V</td>
</tr>
<tr>
<td>MRPL10, Mitochondrial ribosomal protein of the large subunit</td>
<td>✓</td>
<td></td>
<td>M</td>
</tr>
<tr>
<td>TIM23, Essential protein of the mitochondrial inner membrane</td>
<td>✓</td>
<td></td>
<td>M</td>
</tr>
</tbody>
</table>

- Overall we found 53 and 57 carbonylated proteins in wild type and the Δyfh1 strain, respectively.
- 14 proteins were found in both strains - half of them were carbonylated glycolytic enzymes,
- Unique damage in WT: transcription in the nucleus,
- Unique damage in the Δyfh1 strain: protein synthesis within the cytosol and mitochondria.
Affected Pathways
Identification of Oxidized Amino Acids

Peroxiredoxin

Thioredoxin