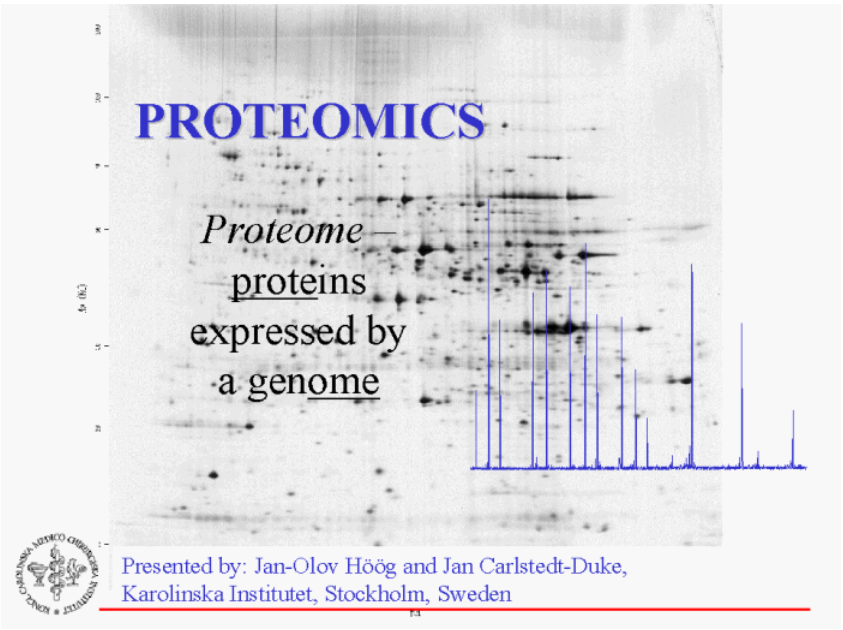
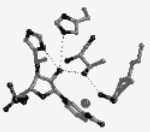
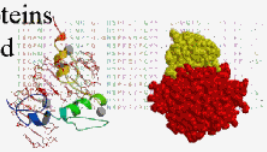


Presented by: Jan-Olov Höög and Jan Carlstedt-Duke,
Karolinska Institutet, Stockholm, Sweden

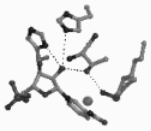
The tools of life



- All cellular processes in biological organisms involve proteins, the tools of life.
- Chemical reactions are catalyzed and controlled by enzymes, molecules are transported, signals are transduced, gene expression is controlled, and energy is harvested through the multifaceted activities of proteins.
- Information underlying all these activities is coded by the genomes.
- A cornerstone for functional genomics/proteomics is therefore the expression of the encoded proteins and the characterization of their functional and structural properties.

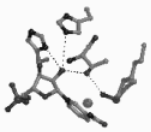


Some genomes completed



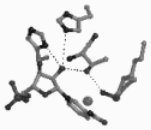
Organism	Genome size (Mb)	Number of genes	Year
Mycoplasma genitalium	0.58	470	1995
Haemophilus influenze	1.8	1,743	1991
Escherichia coli	4.6	4,285	1996
Saccharomyces cerevisiae	12.1	6,266	1996
Caenorhabditis elegans	97	19,000	1998
Drosophila melanogaster	180	13,600	2000
Homo sapiens	3,000	40,000	2001

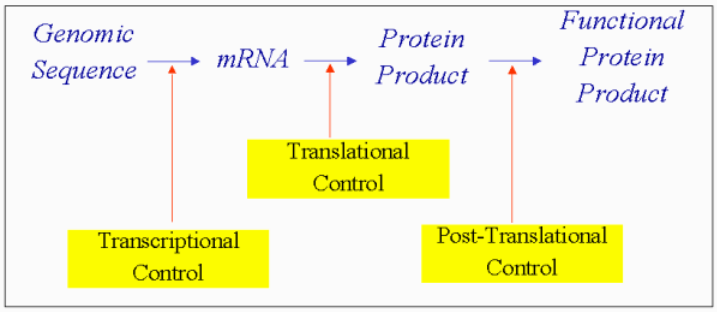
Proteomics – usually Analysis of a large number of proteins with a combination of 2D-gels and masspectrometry



Today it includes both the indentification of a large number of expressed proteins by a genome as well as the characterization of functional and structural relationship

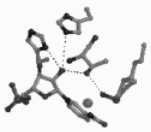
Protein production pathway





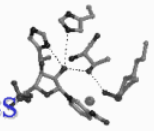
mRNA level ≠ expressed protein level nor does it indicate the nature of the functional protein product

Large scale protein analysis



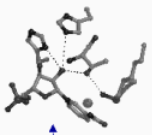
- The Proteome is the protein complement of the genome
- Proteomics is the use of qualitative and quantitative protein level measurements to characterize biological processes (e.g. diseases)
- Diseases are treated at the protein level
- Proteins are drug targets

Challenges: Genomis *vs* Proteomics

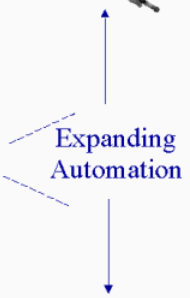


DNA	Protein
• Static	• Very dynamic
• Can be amplified	• Can not be amplified
• Little complexity: Single component	• Post-translationally modified (very complex)
• Good solubility characteristics	• Variable solubility

Key technologies in proteomics

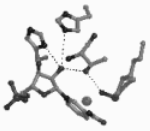


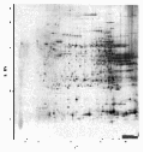
- Reproducible 2D Gel Technology
- Staining and Scanning Technology
- Mass Spectrometry for Identification
- Databases (protein and genome)
- Database Searching Algorithms

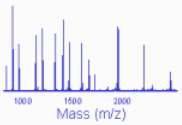


The key challenge in Proteome research is the automation and integration of these technologies

Experimental approach







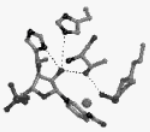
Run gel; stain; scan

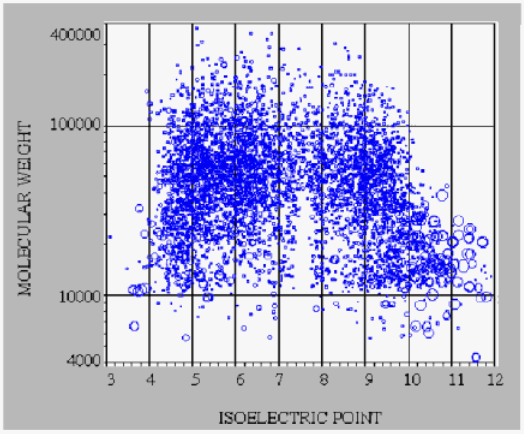
Extract peptides; mass analyze

Excise spot; wash; digest

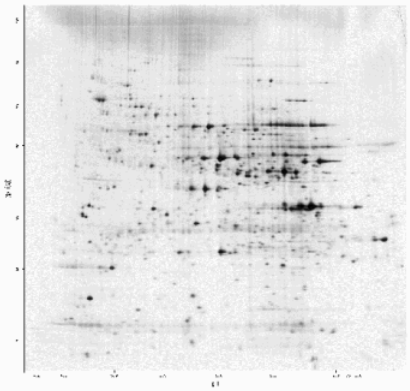
Database search

Theoretical 2D gel of yeasts





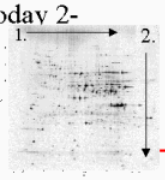
Experimental 2D gel of yeasts



From EXPASY (http://www.expasy.ch)

Reproducible 2D gels (1)

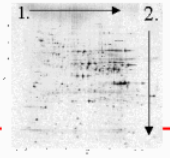
- Electrophoresis is today the central technique for the analysis of a large number of proteins at the same time.
- The methodology as such was invented by Tiselius almost half a centuary ago. The first "2-D gel electrophoresis" was published in 1956 by Smithies and Poulik describing the use of paper and starch gel electrophoresis. 1975 O'Farrell optimised the 2-D separation procedure that is the basis for today 2-D PAGE.



Key technologies in proteomics

Reproducible 2D gels (2)

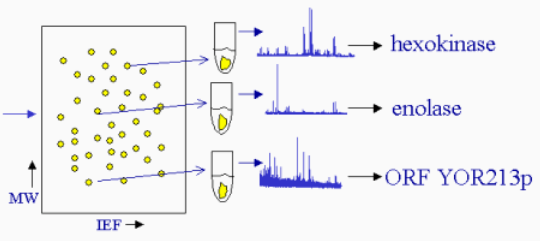
- First gel dimension – Isoelectric focusing (IEF) with immobilised pH gradients (IPG)
- Second gel dimension – SDS polyacrylamide gel electrophoresis (PAGE)
- In both cases precast gels of high quality is a basic requirement.



Need for scaleable proteomics

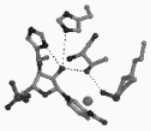
Samples → Separation → Digestion → Identification

- normal samples
- diseased samples
- serum, CSF…
- ± drug
- time-course



1000s of samples =1000s of gels → 1000s of proteins/gel → 10^6-10^{12} proteins

Sample preparation



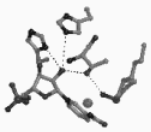
Pretreatment of samples before running gel

Solubilisation → **Denaturation** → **Reduction**

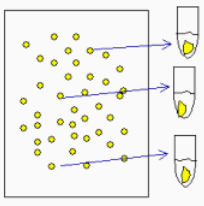
- Non-protein components have to be removed, e.g. nucleic acids.
- Typical solubilisation buffer: 8 M urea, 4% CHAPS, 50 mM DTT and 40 mM Tris.

This step is very often the most important for a successful result.

In-gel digestion

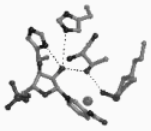


- Dice each gel slice into small pieces and add 25mM NH4HCO3/50% ACN
- Extract the supernatant and transfer to a separate tube (to be discarded).
- Repeat.
- Speed Vac the gel pieces to complete dryness.
- Add trypsin solution (trypsin in 25mM NH4HCO3) and incubate at 4°C for 30 min.
- Spin briefly and incubate at 37°C overnight.



http://donatello.ucsf.edu/ingel.html

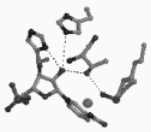
Identification of peptide fragments with masspectrometry



- Direct MALDI-TOF MS[1] of protein immobilised on membrane or in gel matrix (low yield from directly eluted proteins)
- MALDI or ESI-MS[2] analysis of peptides following digestion of proteins from gels or immobilised on membranes

[1]Matrix assisted laser desorption/inoisation - time of flight masspectrometry
[2]Electrospray ionisation-masspectrometry

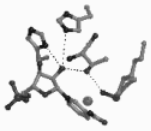
Proteome diversity

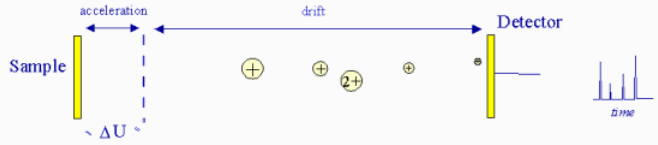


co- and post-translational modifications of the proteins

From one gene sequence to one predicted protein sequence to many protein products, the latter to be identified with masspectrometry

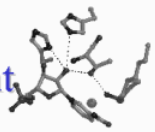
The Time-of-Flight (TOF) method

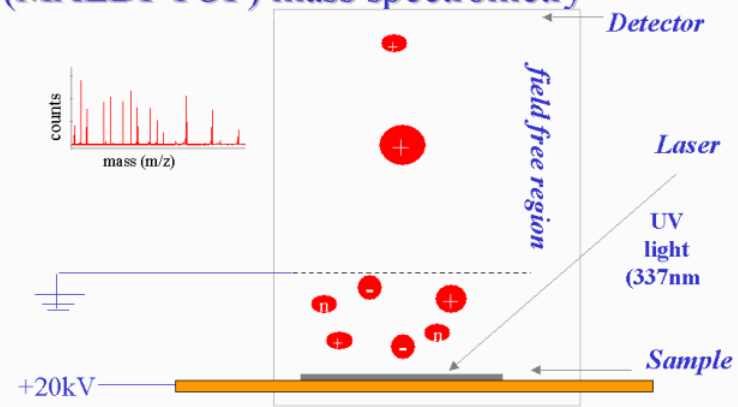




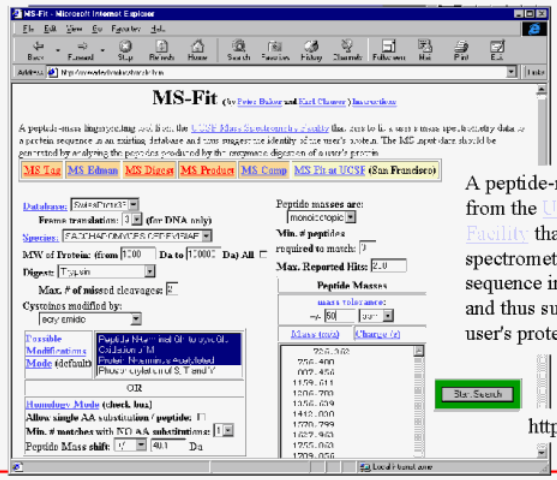
$$TOF = A\ (m/z)^{1/2} + B$$

Matrix Assisted Laser Desorption/Ionization Time-of Flight (MALDI-TOF) mass spectrometry





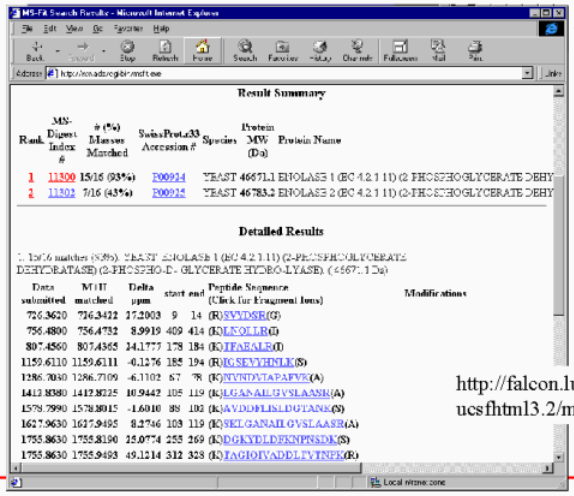
Database searching



A peptide-mass fingerprinting tool from the UCSF Mass Spectrometry Facility that tries to fit a user's mass spectrometry data to a protein sequence in an existing database and thus suggest the identity of the user's protein.

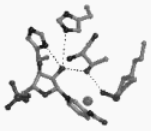
http://donatello.ucsf.edu/

Database searching: hit list



http://falcon.ludwig.ucl.ac.uk/ucsfhtml3.2/msfit.htm

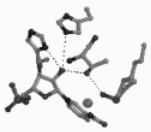
Mass accuracy in database searching



Molecular Ion	Allowed error (ppm)	# Matched Proteins	
		All Species	Yeast
1975.954	1000	754	173
	100	200	76
	10	44	6
1526.884	1000	982	120
	100	197	27
	10	19	2
1055.541	1000	1300	173
	100	514	76
	10	87	6

* SWISSPROT r.33, total proteins all species= **52205**, **yeast = 3653**

Mass accuracy in database searching



# of peptides	error (ppm)	# Matched Proteins	
		All Species	Yeast
2 of 3	1000	149	20
3 of 3		2	1
2 of 3	100	10	1
3 of 3		1	1
2 of 3	10	2	1
3 of 3		1	1

Masses searched = 1288.7113, 1420.7549, 2447.0978

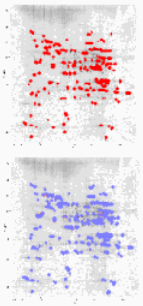
* SWISSPROT r.33, total proteins = **52205**

Developments

Difference gel electrophoresis (DIGE)

Dye-labelled proteins co-separated on the same gel

- First protein extract is labelled with e.g. Cy2 and second protein extract is labelled with Cy3
- Co-separation by 2-D PAGE
- Image gel at two different wavelengths
- Differential analysis
- Inducible proteins can be detected with only one gel run



Developments

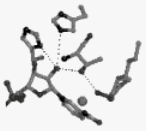
Multidimensional protein identification technology (MudPIT)

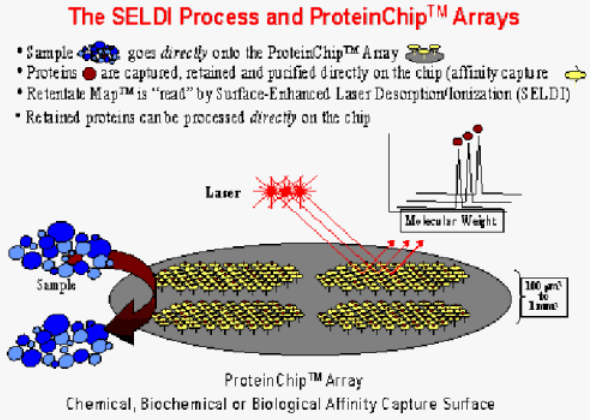
- Multidimensional liquid chromatography
- Tandem mass spectrometry
- Database searching

MudPIT has been applied to the proteome of the *Saccharomyces cerevisiae* and yielded the largest proteome analysis to date. A total of 1,484 proteins were detected and identified. Categorization of these hits demonstrated the ability to detect and identify proteins rarely seen in proteome analysis, including low-abundance proteins like transcription factors and protein kinases.

Developments

Protein-array chips (1)

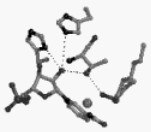




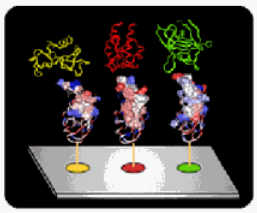
• http://www.ciphergen.com/

Developments

Protein-array chips (2)

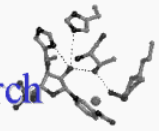


The HIP™ chip, the first high-throughput addressable array of capture proteins, which will revolutionize drug discovery and functional proteomics. With this chip it is possible to rapidly assess the protein expression profile of a biological sample so as to measure the relative level of hundreds and ultimately thousands of different proteins simultaneously. Differential analysis of these profiles can then be used to identify the proteins in human tissues that are affected as a result of a particular disease or in response to a specific therapeutic drug.

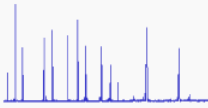


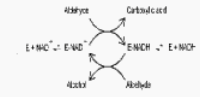
• http://www.phylos.com/

Useful links for proteomics research



- http://base-peak.wiley.co.uk/links/
- http://donatello.ucsf.edu/
- http://prospector.ucsf.edu/
- http://ai.sri.com/ecocyc/ecocyc.html
- http://www.ebi.ac.uk/
- http://www.embnet.org/
- http://www.expasy.ch/
- http://www.hon.ch/
- http://www.pdb.bnl.gov/
- http://www.protocol-online.net/index.htm
- http://www.nature.com/genomics/





To obtain the complete picture of Proteomics/Functional genomics