Proteomics – basics, methods, applications

Basics and „omics": Genomics

- Determination of human genome: February, 2001
- Genome: Winkler, 1920; „GENes and chromosOMEs"

The complete set of chromosomal and extrachromosomal genes of an organism, a cell, an organelle or a virus; the complete DNA component of an organism. [IUPAC Biotech]

→ Genomics: Roderick, 1986 (Genomics, journal)

- Science of using DNA- and RNA-based technologies to demonstrate alterations in gene expression.
- Sequencing data
- Discovery of new genes
- Gene mapping
- Comparison of different genomes (species, evolution)
- New genetic technologies

Basics and „omics": Proteomics

Proteome: M. Wilkins (1994)
Conference on Genome and Protein Maps, Siena, Italy
Protein complement expressed by a genome.

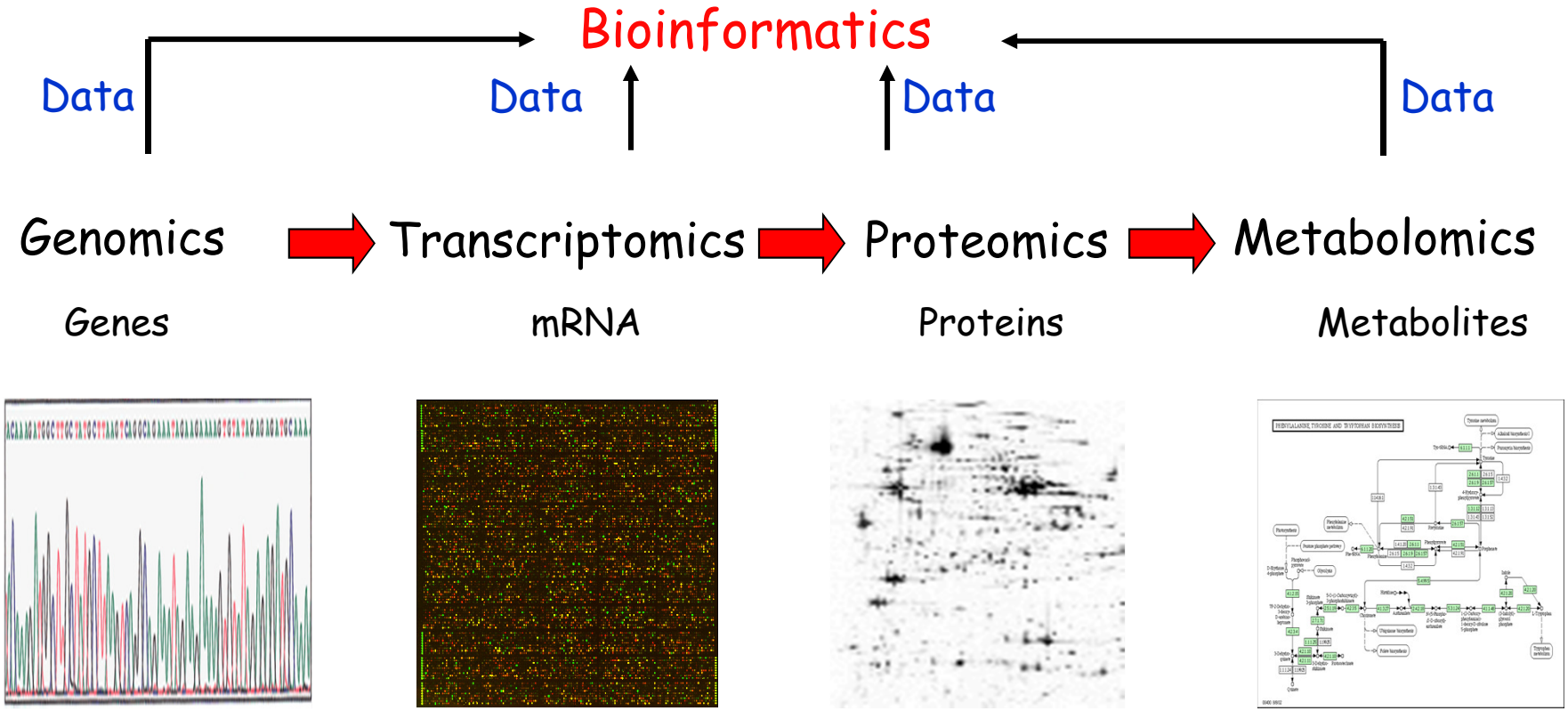
All of the proteins in an organism, organelle or virus. The set of expressed proteins in a given type of cell or organism, at a given time, under defined conditions.

> If the genome is a list of the instruments in an orchestra, the proteome is the orchestra playing a symphony. (R. Simpson)

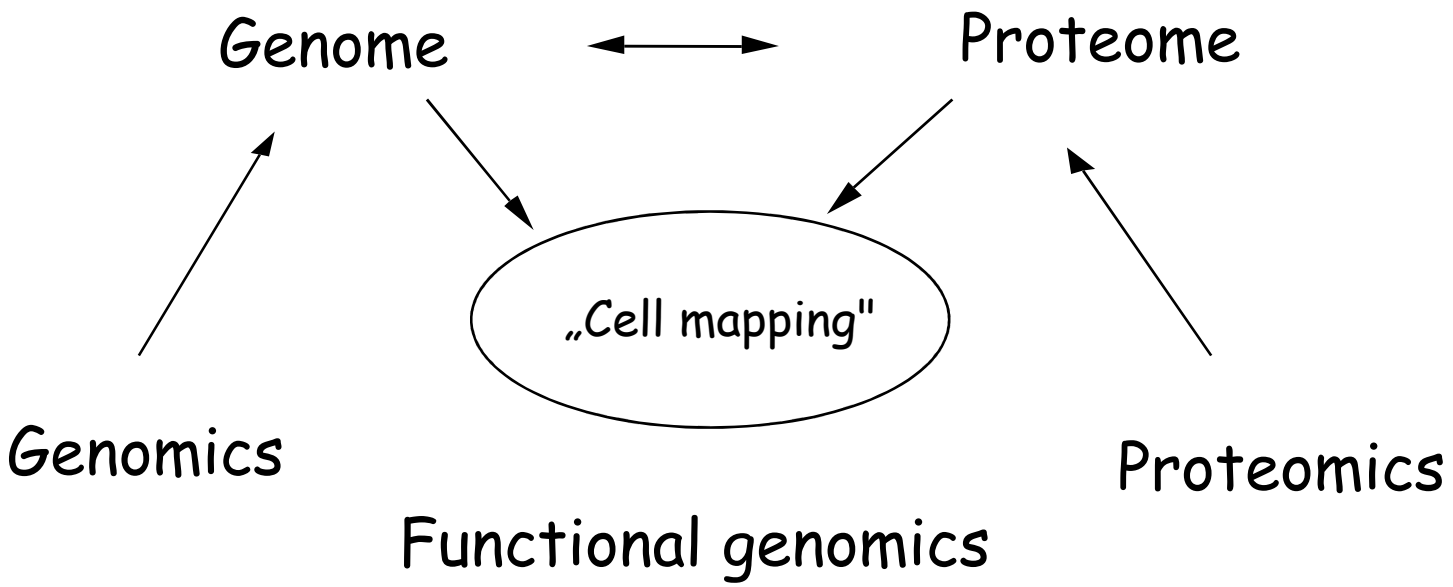
Proteomics:

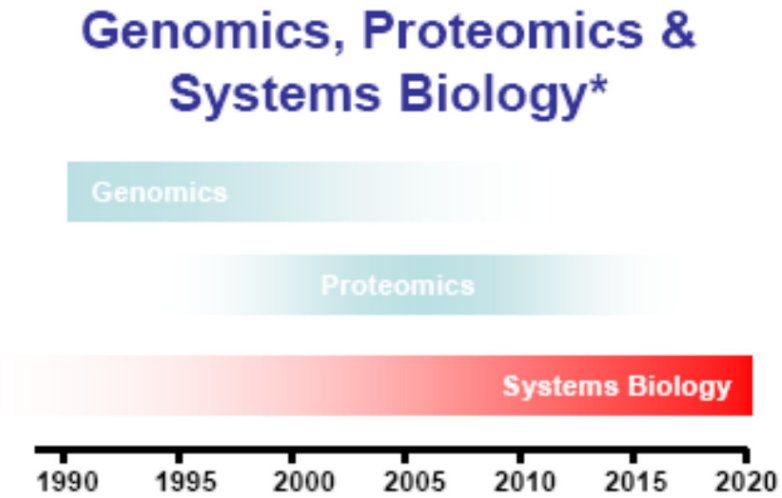
1. Qualitative or quantitative analysis and comparison of the proteomes under different conditions for better understanding of biological processes (e.g. disease vs healthy sample)
2. Identification (sequence, 3D structure) and protein modifications
3. Collection of experimental data in databases
4. Determination of the following protein properties:
   - localization
   - post-translational modifications
   - interactions (e.g. metabolics and signalization)
   - activity
   - function (based on predicitions and experimental data)
5. Better understanding of diseases, diagnostics, drug development

Further „omics"

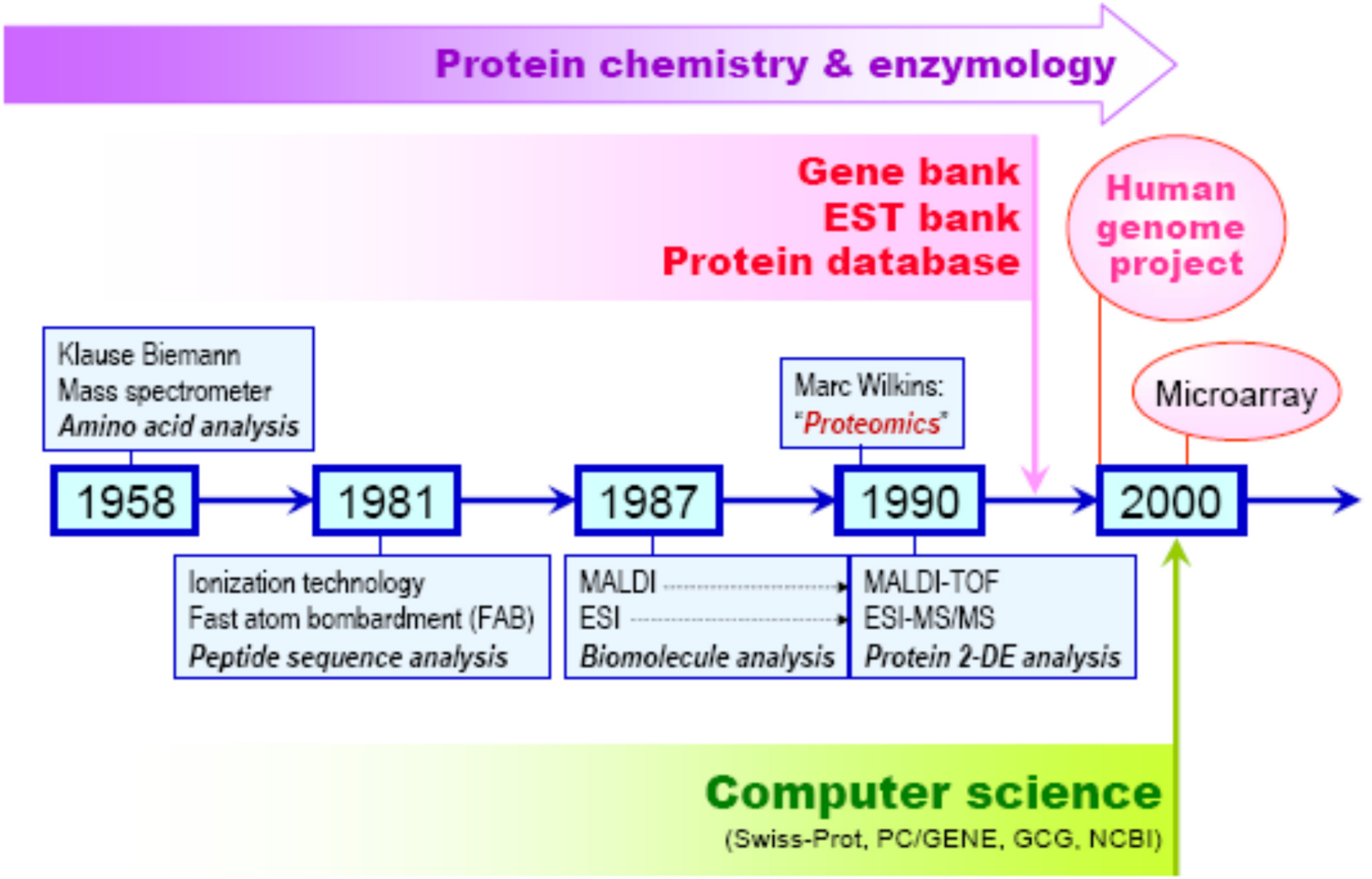


Outline

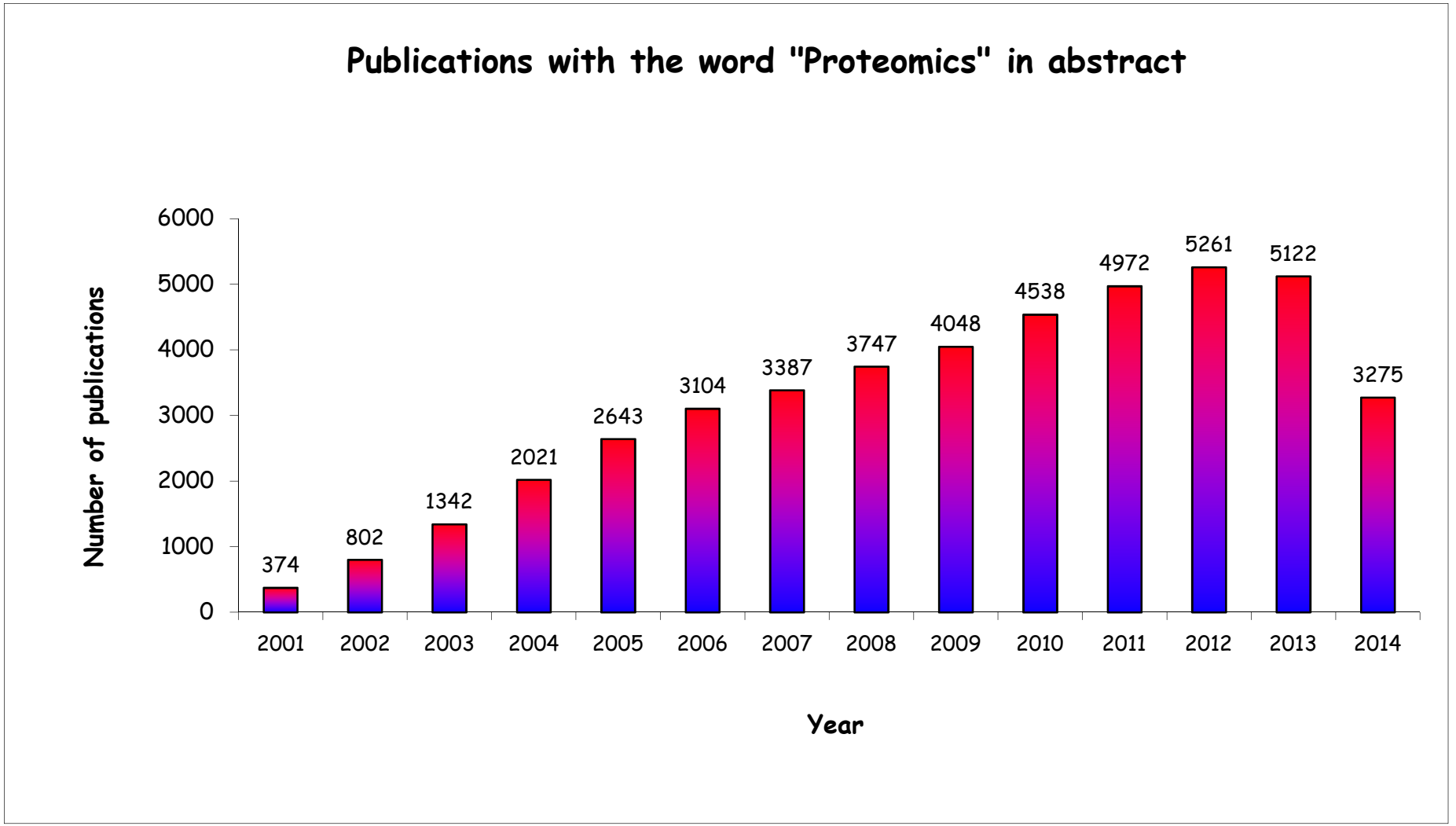

Genome
Proteome
„Cell mapping"
Genomics
Proteomics
Functional genomics


Genomics, Proteomics &
Systems Biology*
Genomics
Proteomics
Systems Biology
1990
1995
2000
2005
2010
2015
2020

Historical review for proteomics

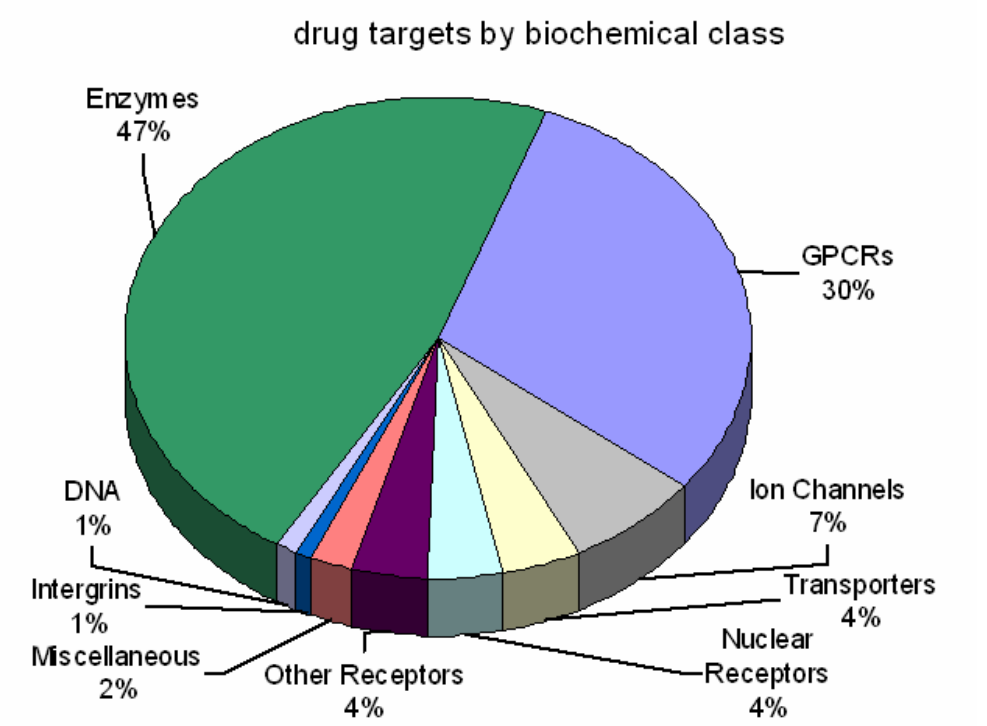


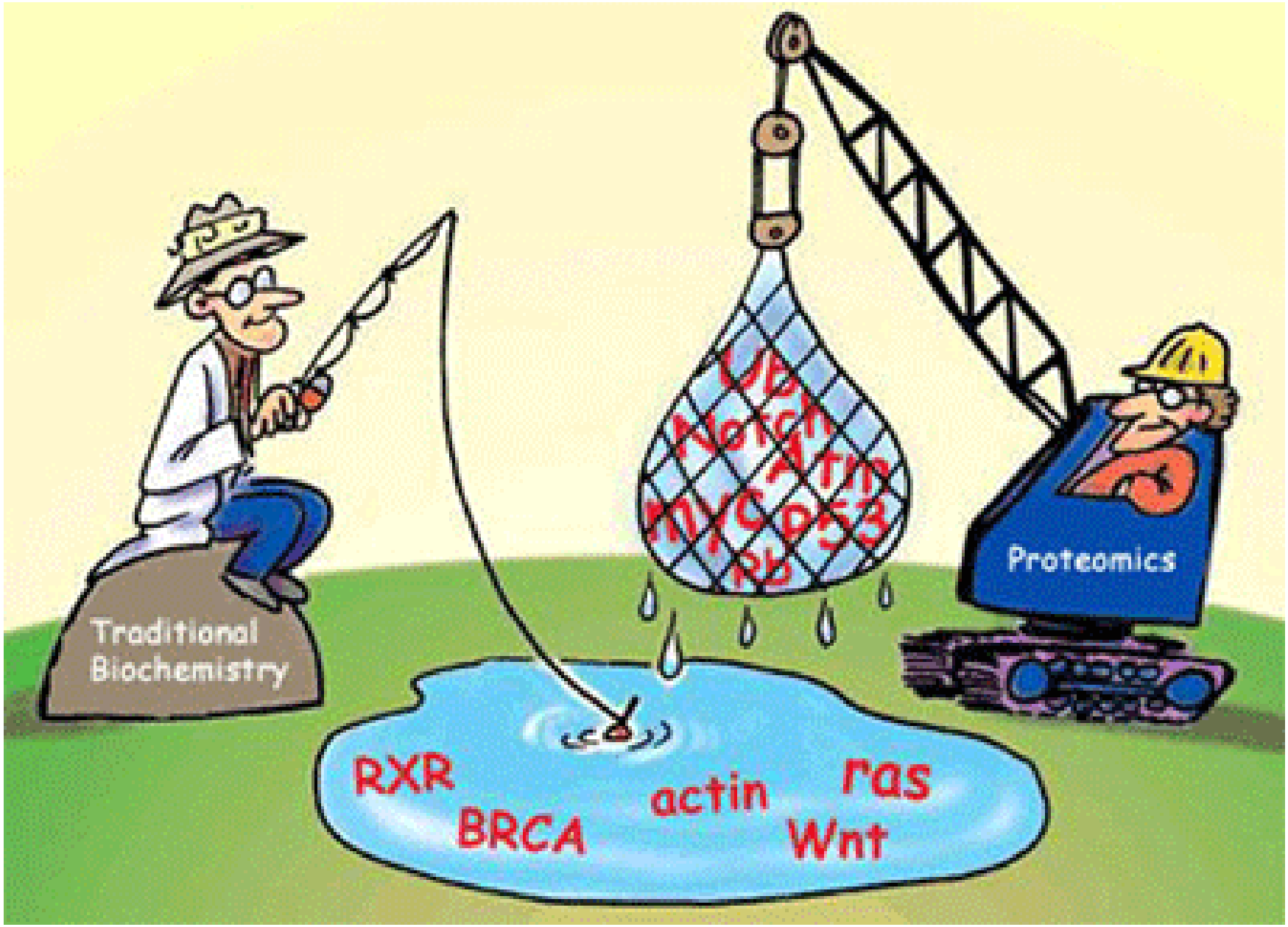
Adapted from Liebler (2002) Introduction to Proteomics, Foreword


Publications with the word "Proteomics" in abstract
Number of publications
6000
5000
4000
3000
2000
1000
0
374
802
1342
2021
2643
3104
3387
3747
4048
4538
4972
5261
5122
3275
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
Year

Why do we need Proteomics?

- The mRNA expression is not always in correlation with the protein expression
- mRNA expression of biological samples (e.g. serum, urine) can not be easily analyzed
- Active agents in cells, tissues, organisms are proteins and not the genes
- Modifications which can not be predicted from the DNA sequence (post-translational modifications)
- Localization and interaction studies could be carried out with proteomical approaches




Traditional
Biochemistry
Proteomics
RXR
BRCA
actin
ras
Wnt

Proteomics

Functional

Activity-based proteomics:
Find connections between functional and gene-expression data and/or protein-protein interaction

Chemical proteomics:
Proteom-based examinations
Determination of enzyme activity,
Chemical libraries

Toxicoproteomics:
Determination of specific proteins and study of their concentration alteration

Integrated

Microbial proteomics

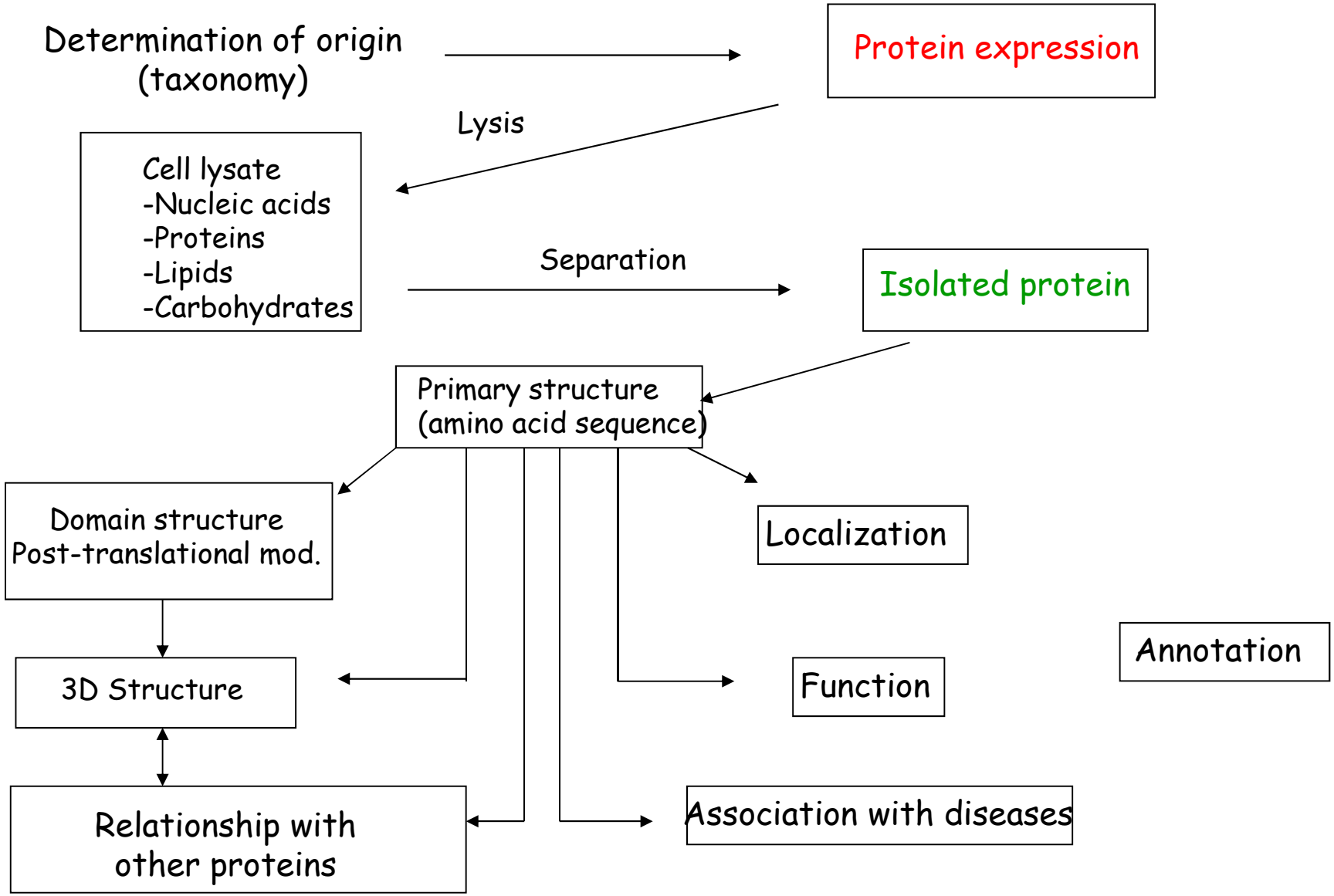
Targeted proteomics:
Tissue proteomics initiative (National Cancer Institute)

Structural

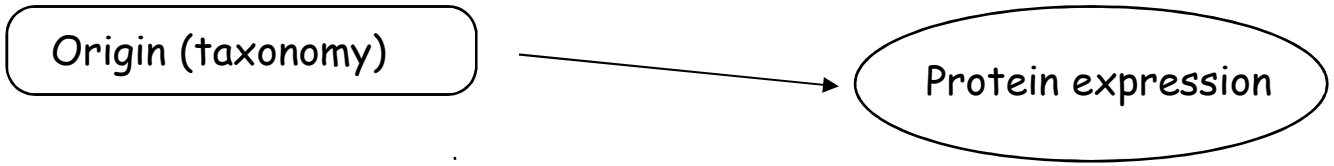
Filoproteomics:
Identification of unknown bacterial isolates based on biomarker similarities in databases

Computational proteomics:
High throughput 3D/4D structural analysis and application in every life sciences

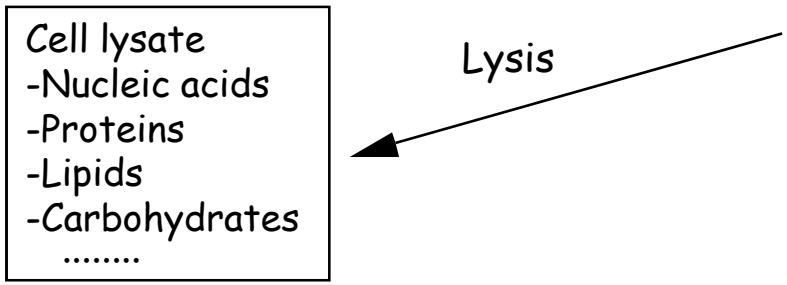
From the expression towards identification



From the expression towards identification... Difficulties

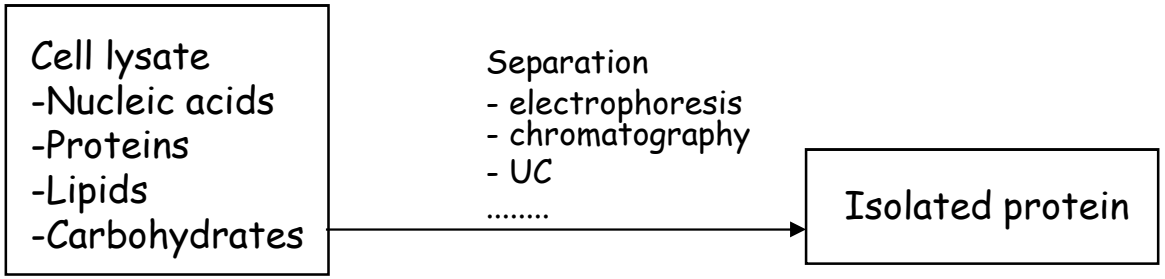


1. 30 000 genes - 500 000 proteins
2. Different expression profile: quantity, quality:
   - time
   - temperature
   - light
   - chemical signal -- „natural"
     -- „environmental"
     (drug, impurities etc.)



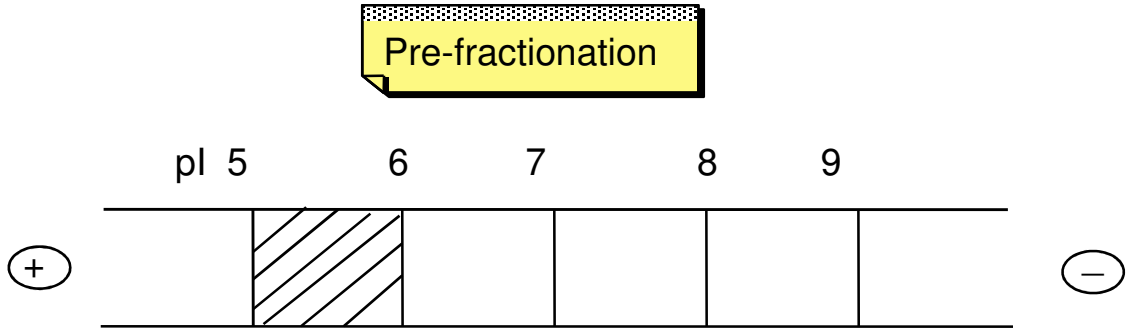
1. Incomplete, molecular desintegration.
2. Solubility (hydrophobe proteins).
3. Enzyme activty, degradation.

From the expression towards identification... Difficulties



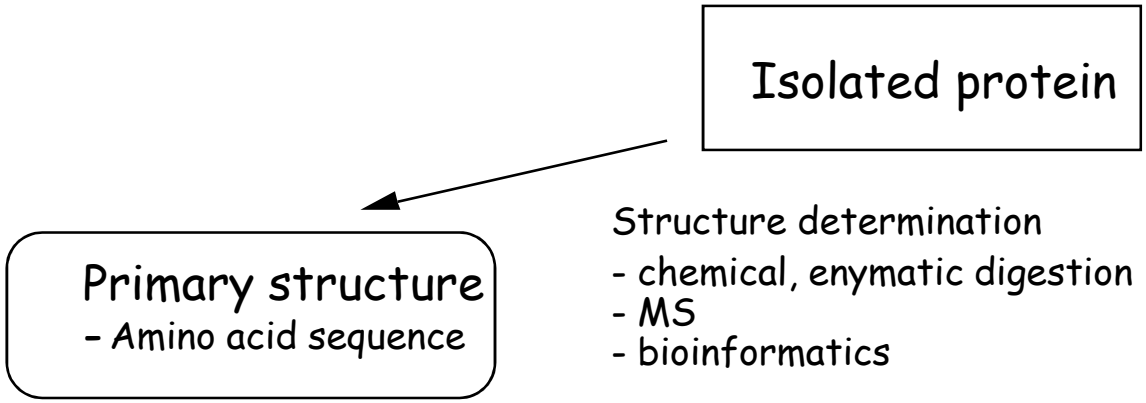
1. 500 000 protein; ca. 10000 protein/sample, but only about 2000 could be separated by 2D gel electrophoresis

2. Small and big
3. Small amount („rare protein").
4. Low concentration.
5. High amount ("abundant protein").
6. High concentration.
7. Membrane protein/hydrophobe protein (solubility problems)



E.g. human serum: 600 spots, 205 in the pI range of 5-6

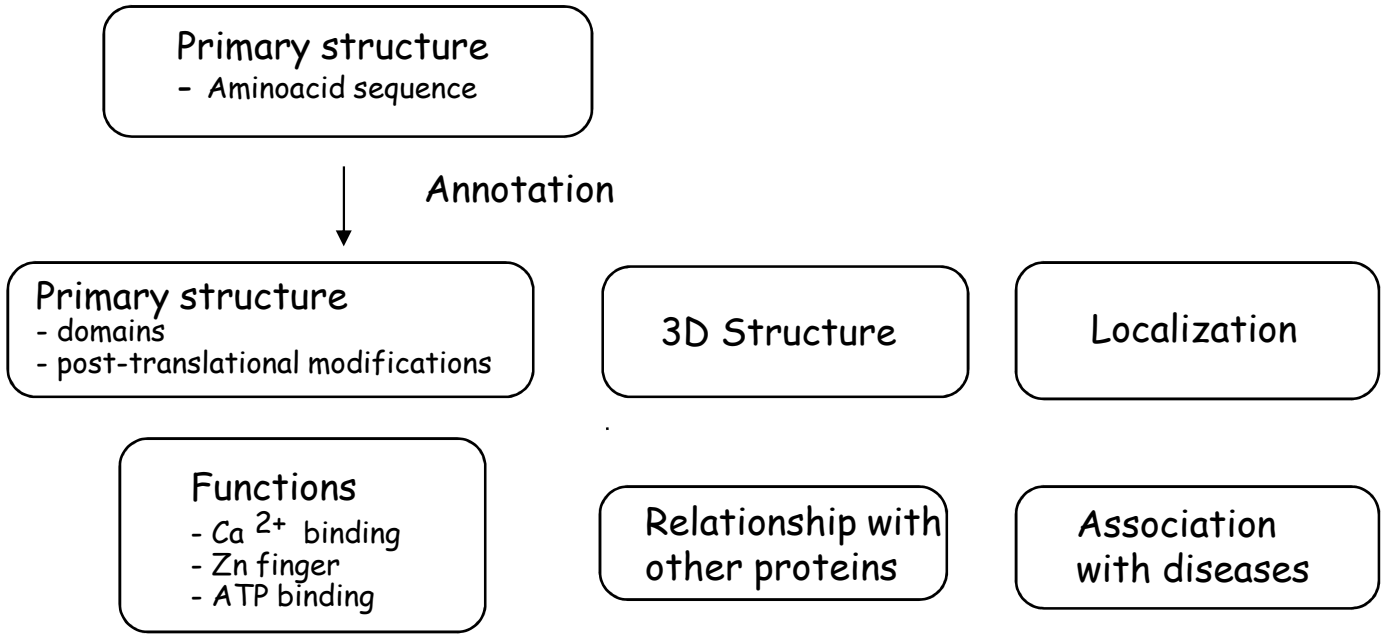
From the expression towards identification... Difficulties



1. Disulfide bridge- domain structure.
2. Side chain amide bond.
3. Non-covalent complex.
4. Post-translational modification.
5. Enyzme ‚resistant' protein (e.g. prion).
6. Protein missing from the database

E.g. human serum: 600 spots, 205 in the pI range of 5-6, only 105 could be found in the database.

From the expression towards identification... Difficulties



1. Unknown 3D structure
   - isolation, synthesis, X-ray/NMR
   - prediction
2. Unknown localization
   - cell fractionation
   - radiolabeled biosynthesis
3. Unknown function
   - binding
   - measurement of biological function
4. Comparative studies
   - species
   - "normal" vs „treated"

NiceProt View of SWISS-PROT: P16070

Entry name	CD44 HUMAN
Primary accession number	P16070
Entered in SWISS-PROT in	Release 14, April 1990
Sequence was last modified in	Release 35, November 1997
Annotations were last modified in	Release 41, March 2002
Protein name	CD44 antigen [Precursor]
Synonyms	Phagocytic glycoprotein I
	PGP-1
	HUTCH-I
	Extracellular matrix receptor-III
	ECMR-III
	GP90 lymphocyte homing/adhesion receptor
	Hermes antigen
	Hyaluronate receptor
	Heparan sulfate proteoglycan
	Epican
	CDw44
Gene name	CD44 or LHR
From	Homo sapiens (Human) [TaxID: 9606]

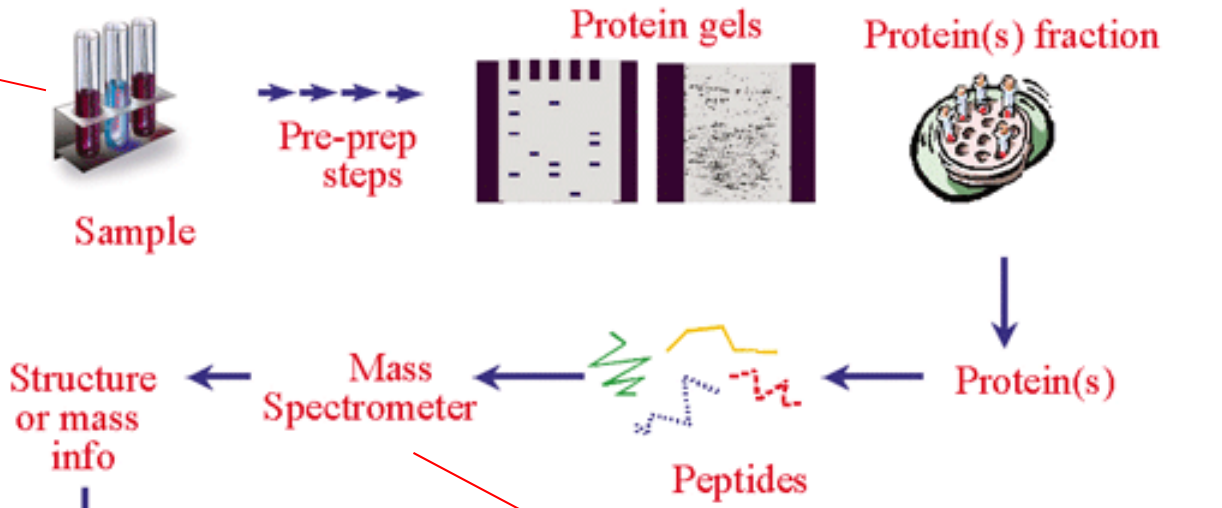
Taxonomy Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

References

[1] SEQUENCE FROM NUCLEIC ACID (VARIOUS ISOFORMS). TISSUE=Lymphoblast; MEDLINE=93101687; PubMed=1465456; [NCBI, ExPASy, EBI, Israel, Japan]
Screaton G.R., Bell M.V., Jackson D.G., Cornelis F.B., Gerth U., Bell J.I.; "Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons." Proc. Natl. Acad. Sci. U.S.A. 89:12160-12164 (1992).

Proteomical workflow

Sample preparation: easy, reproducible, cold, protease inhibitors, (optional: organelle isolation, purification with antibody)



ESI-MS, MALDI-TOF…

Separation:
- 1 and 2D gel electrophoresis
- Reverse Phase HPLC
- SCX HPLC (strong cation exchange)
- 2D LC/MS

SWISS-PROT
TrEMBL
RefSeq XPs
Ensembl

Quantitative proteomics: ICAT (Isotope Coded Affinity Tag) or SILAC (Stable Isotope Labeling by Amino acids in Cell culture)

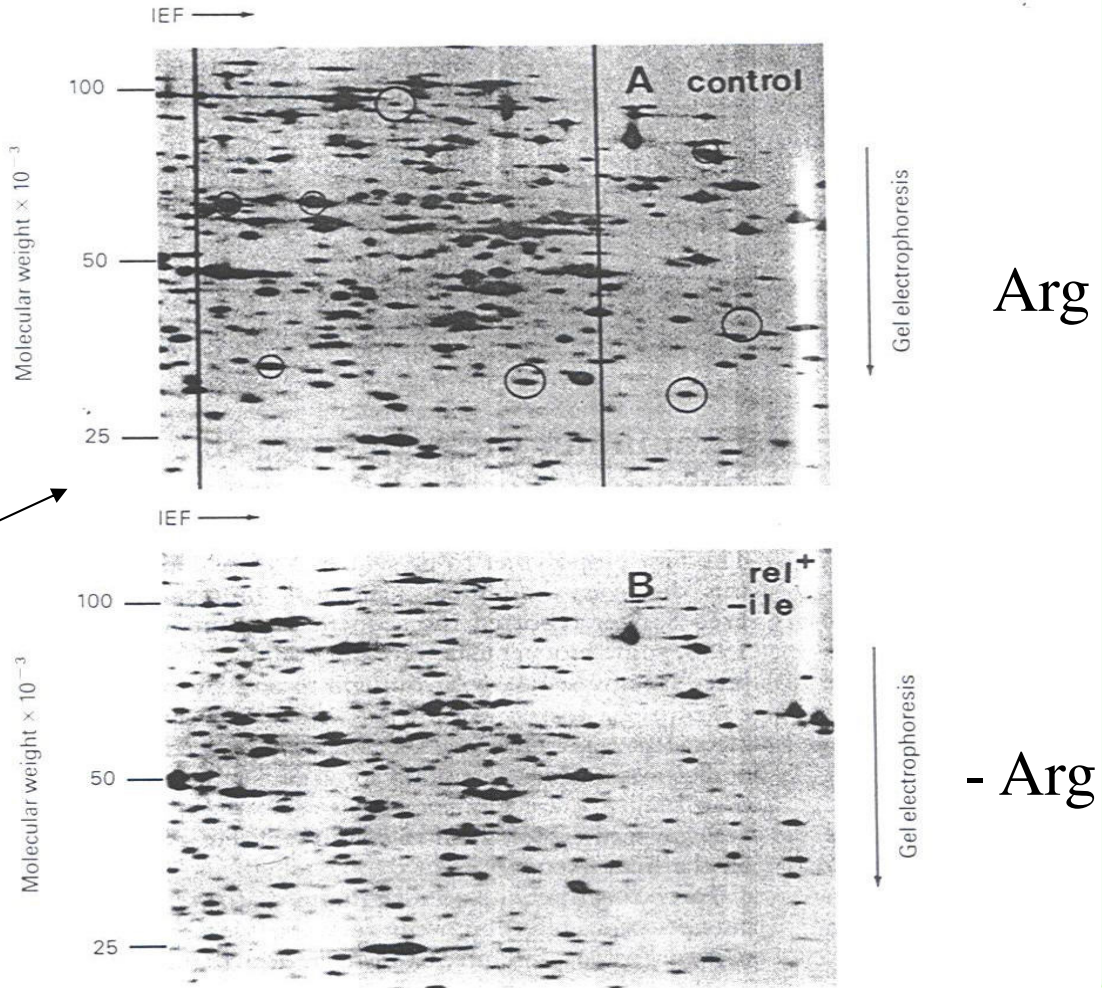
Sample preparation

(Mainly for 2D-gels)

- Any type of cell could be analyzed (plant-, microorganism-, animal cells, cell culture, organs etc.)
- Proteins with isolelectric point between 2 and 11
- Proteins between 8 and 200 kDa
- Post-translational modifications
- Cells: 6 x 10^6 cell/gel
- Tissues: 2,5-50 mg blood-free tissues
- Body fluids: 10^{10} cells
- Liophylized protein sample: 100-500 µg/gel (protein concentration: 5-30 mg/ml, buffer <50 mM salt, detergents: only non-ionic detergents)
- Sample buffer (7M Urea, 4% CHAPS, 50mM DTT, 2M thiourea)

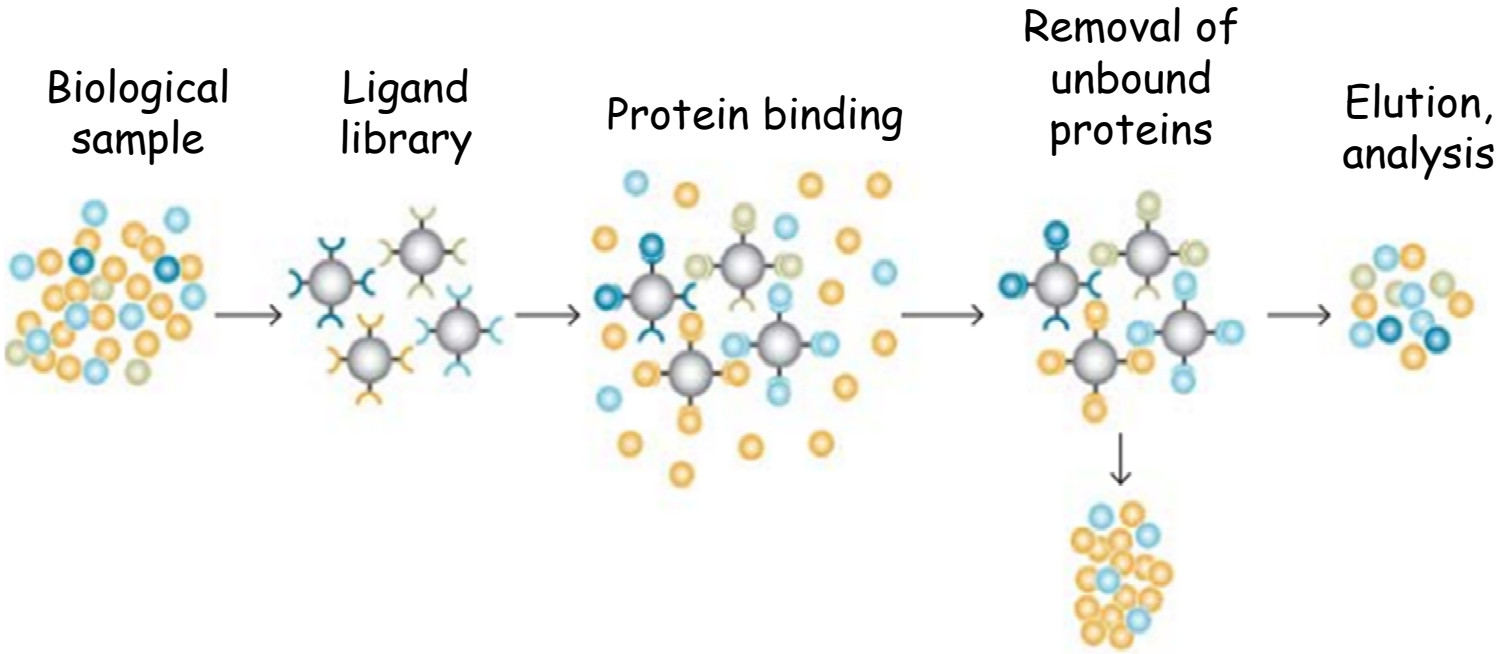
Difficulties in sample preparation

- Expression could be organ-, tissue- or time-specific
- Proteins in small amount
- Abundant proteins
- Conditions are very important (compositi of the cell culture media)
- Membrane proteins



The effect of medium composition on E-coli protein expression
Farrell PHO (1978)

ProteoMiner (Bio-Rad)



Solubilization of membrane proteins for 2D gel electrophoresis

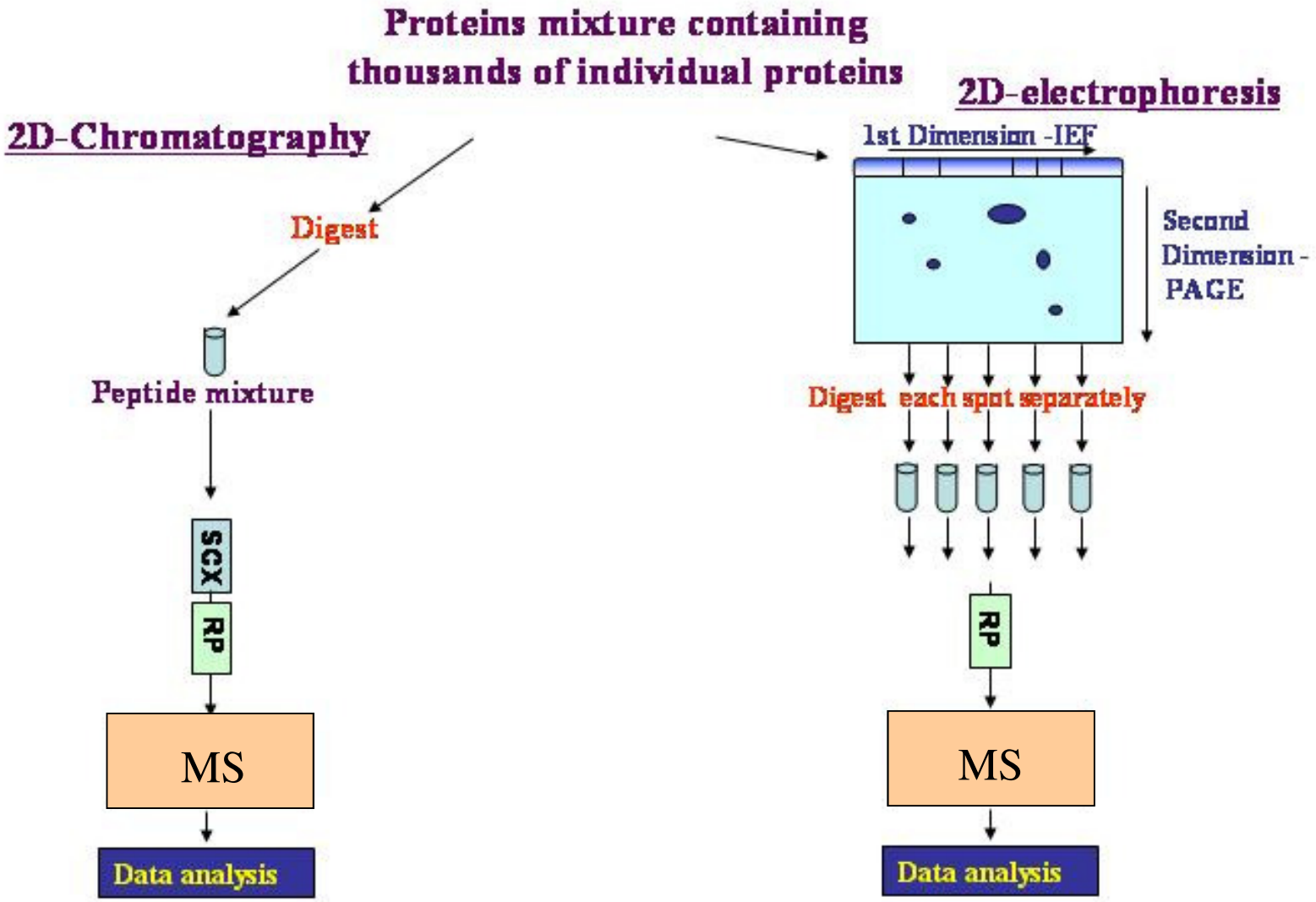
Why membrane proteins represent only 1% of the identified proteins?

1. Small amount
2. High pI
3. Poor solubility in aqueous solutions

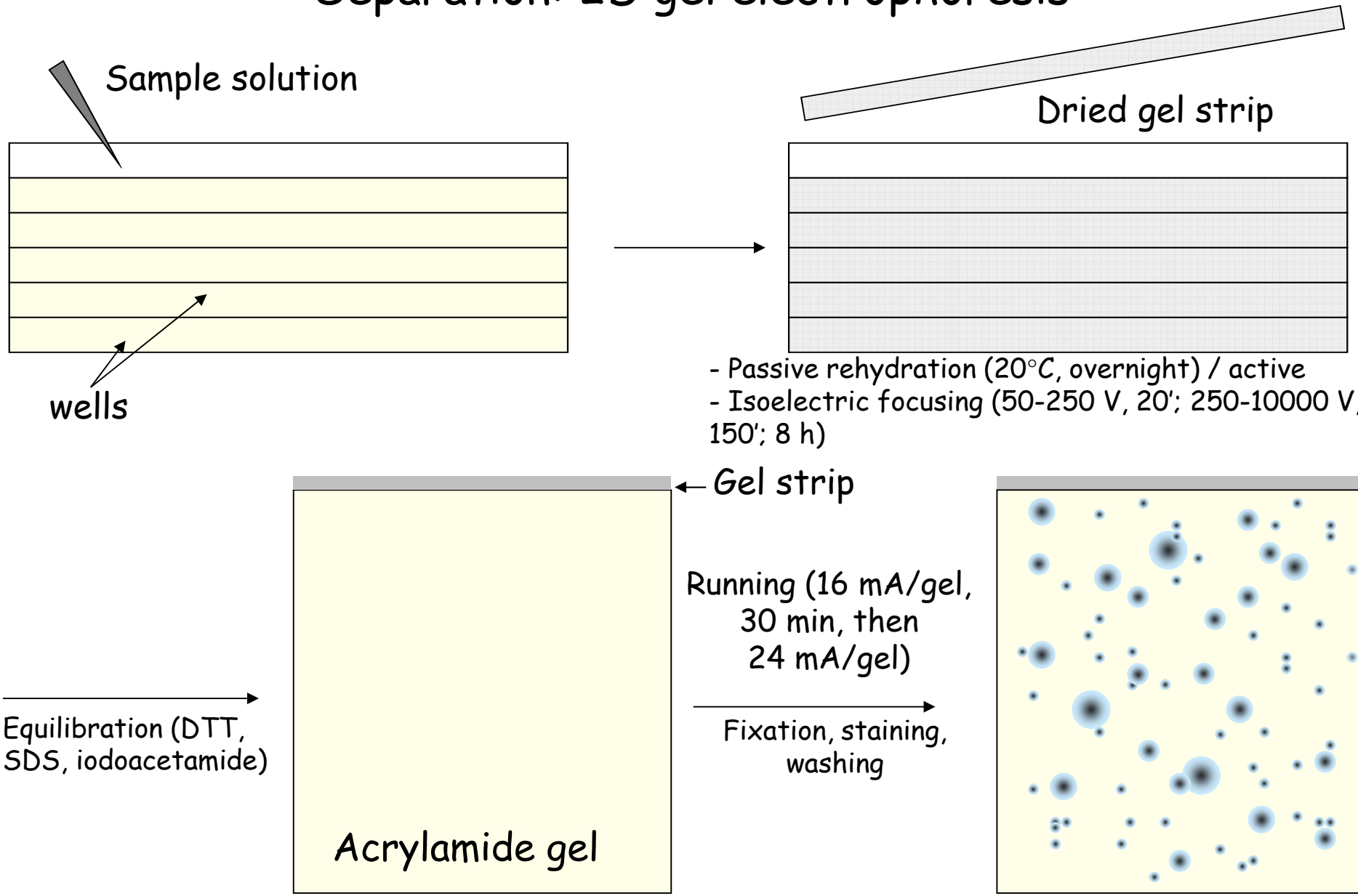
Solution

1. Isolation from the lipid environment
2. They should be solubilized and
3. Remain in solution during IEF

Methods

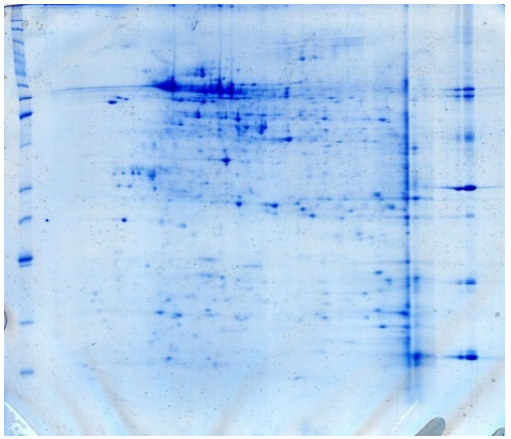

Proteins mixture containing
thousands of individual proteins
2D-Chromatography
Digest
Peptide mixture
SCX
RP
MS
Data analysis
2D-electrophoresis
1st Dimension -IEF
Second
Dimension -
PAGE
Digest each spot separately
RP
MS
Data analysis

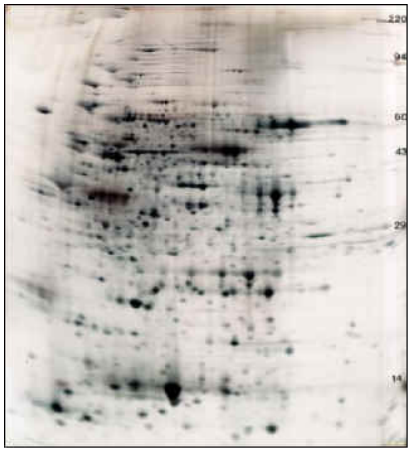
Separation: 2D gel electrophoresis

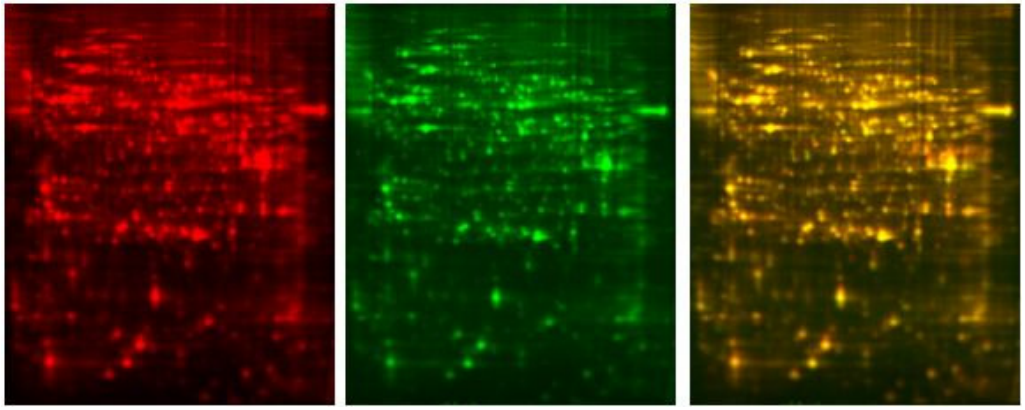


Detection: dyes, softwares

Detection limit: 1 ng
10000 copies/cell (but the amount of regulator proteins are less than 10000 copies/cell)
Coomassie G250 or R250 →
Fluorescent dyes for DIGE (Cy3, 5)
Silver staining

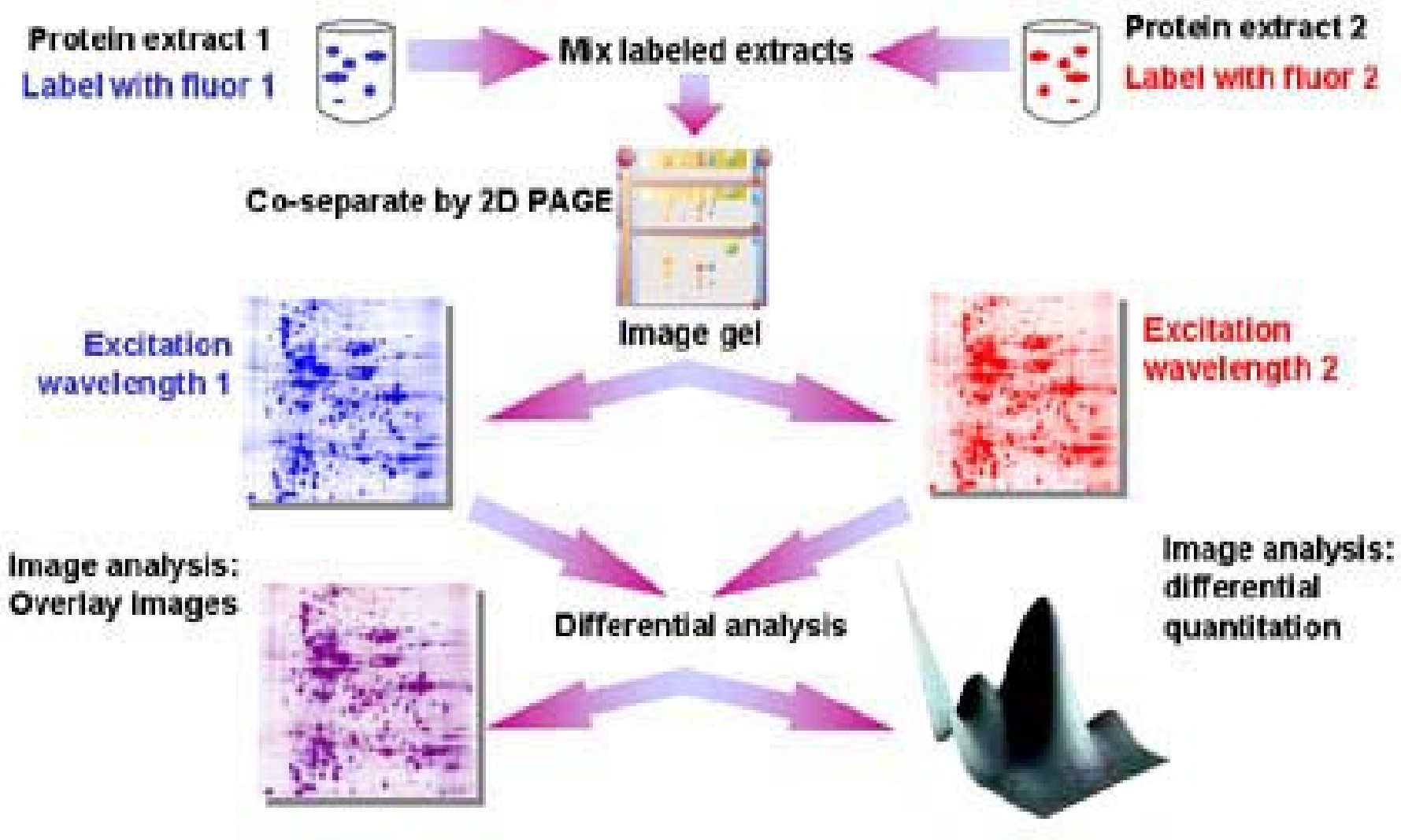




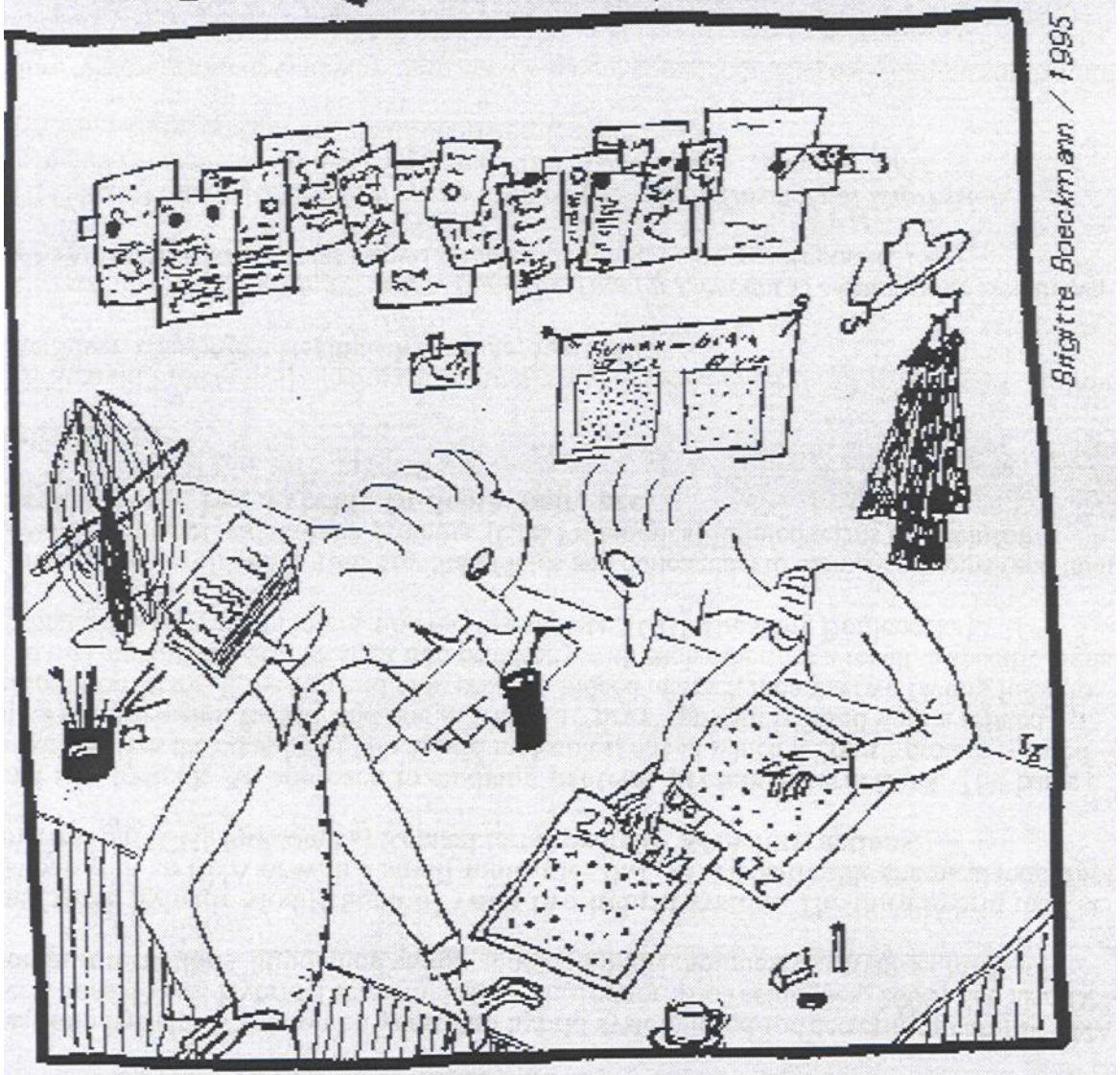


Softwares for analysis: Progenesis, DeCyder, PDQuest…

Differential gel electrophoresis (DIGE)

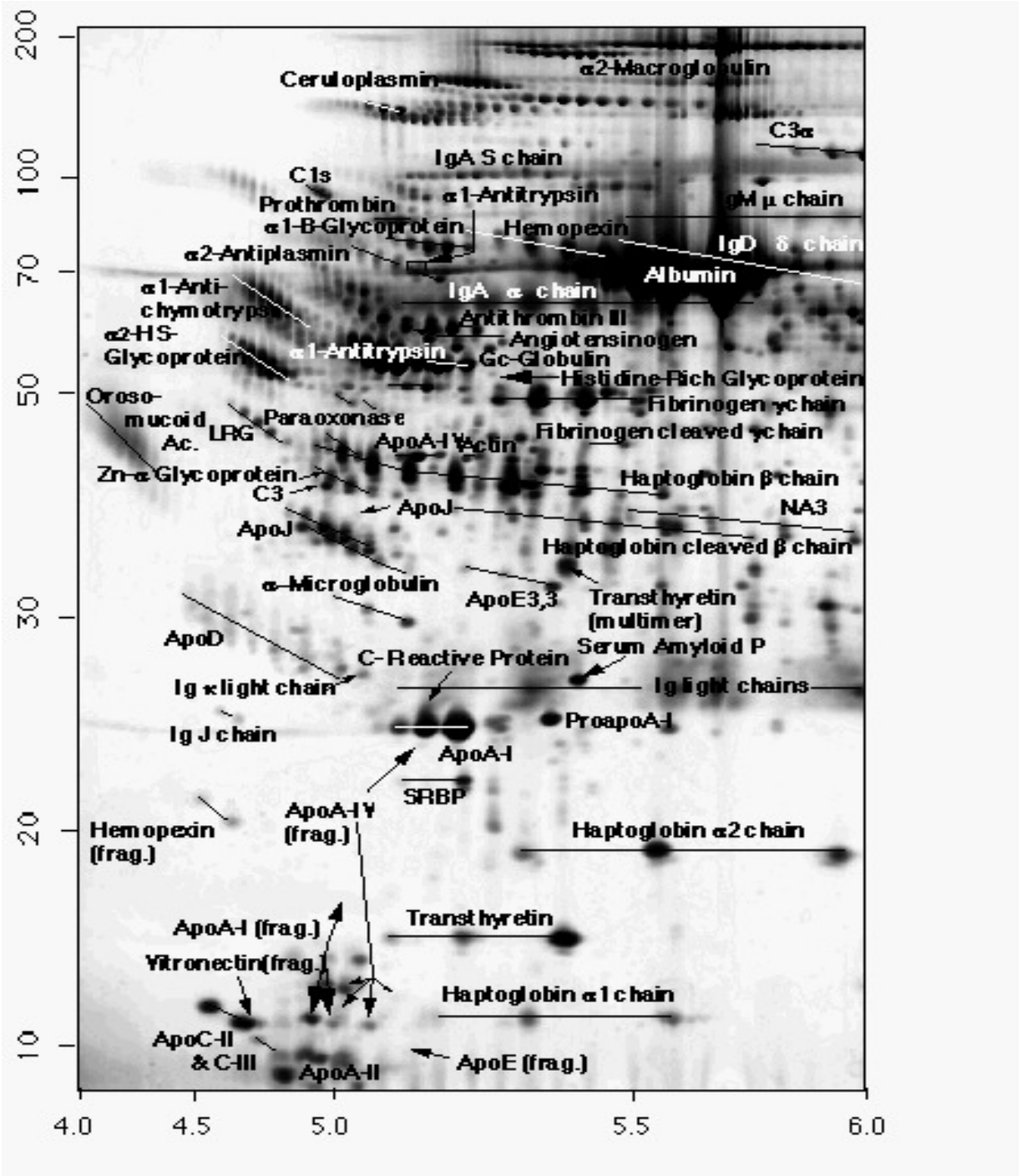


2D gels



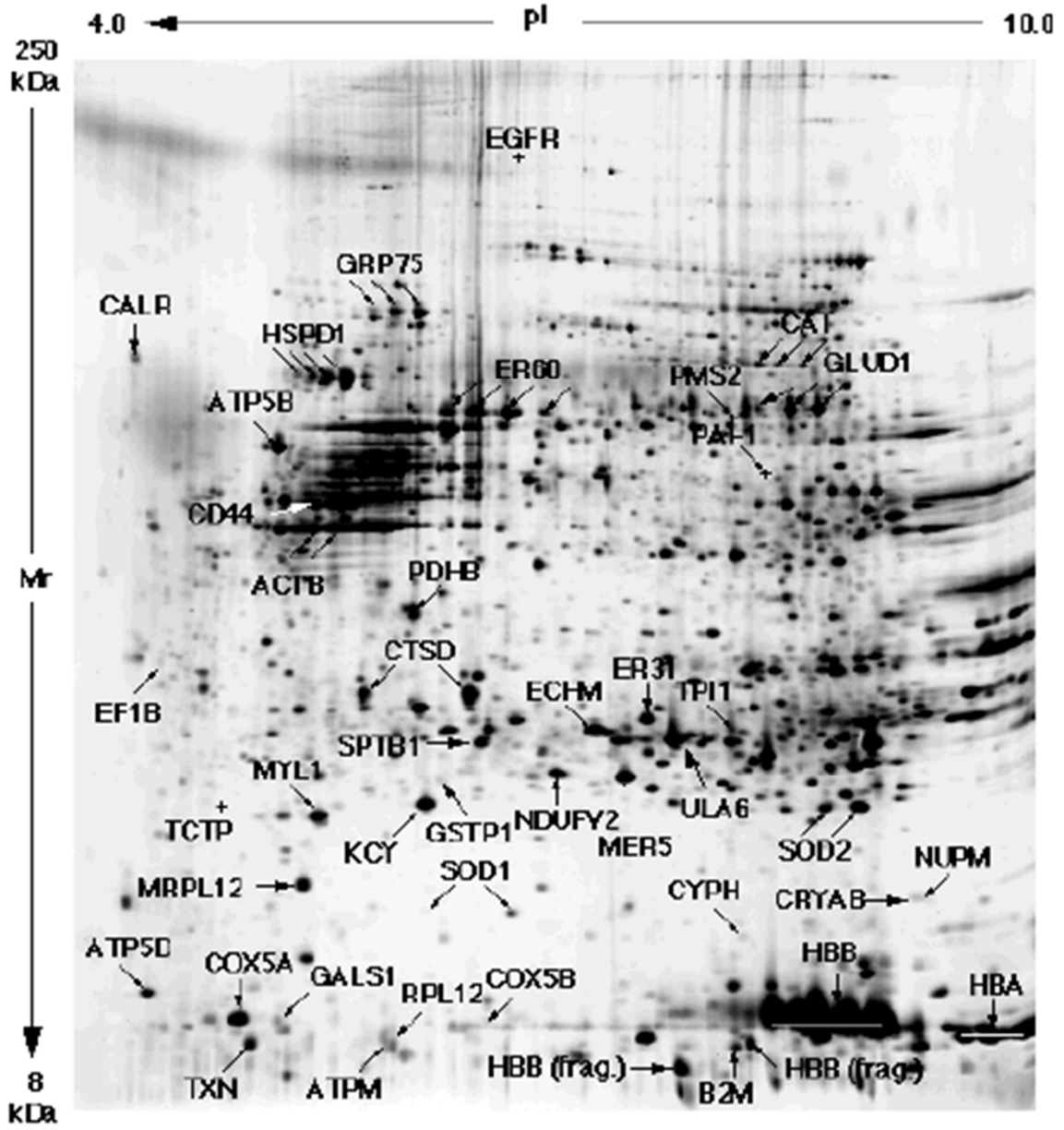
"You've got one protein missing..."
"No, you've one extra protein!"

Protein map of human plasma
(2D gel electrophoresis)
(JC Sanchez et al. 1995).

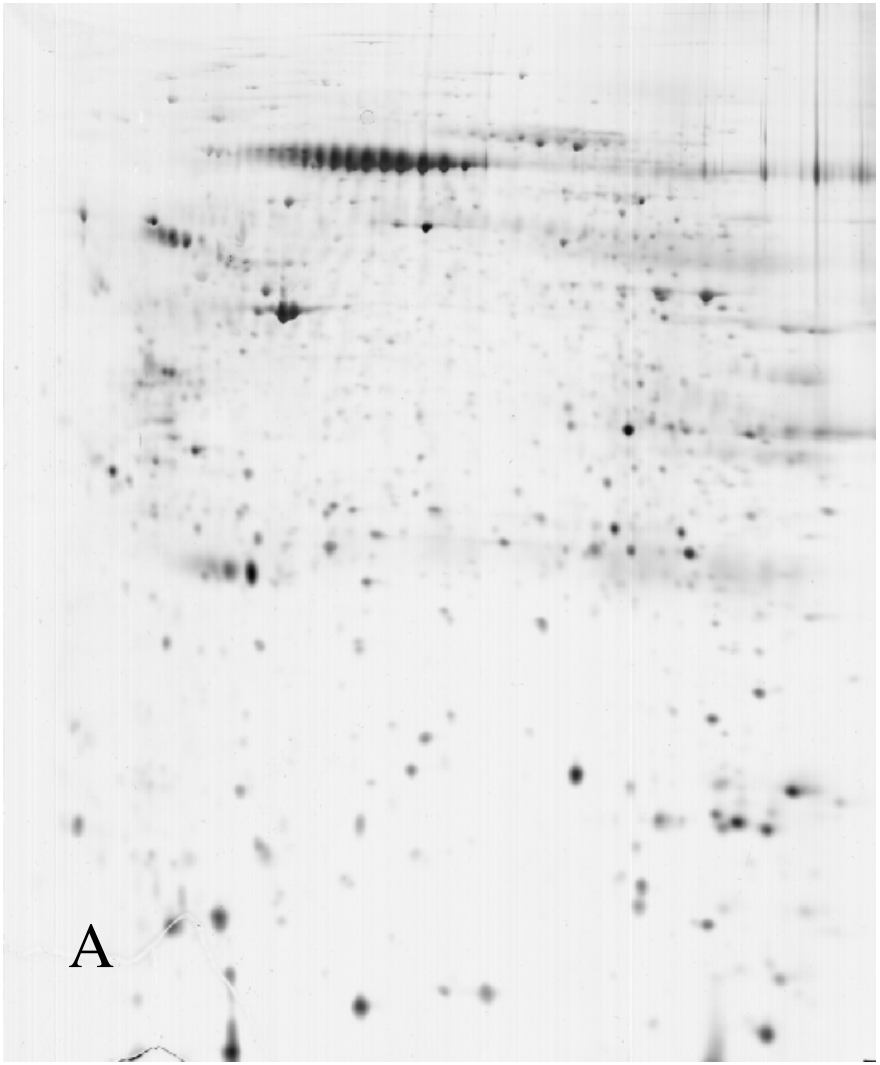


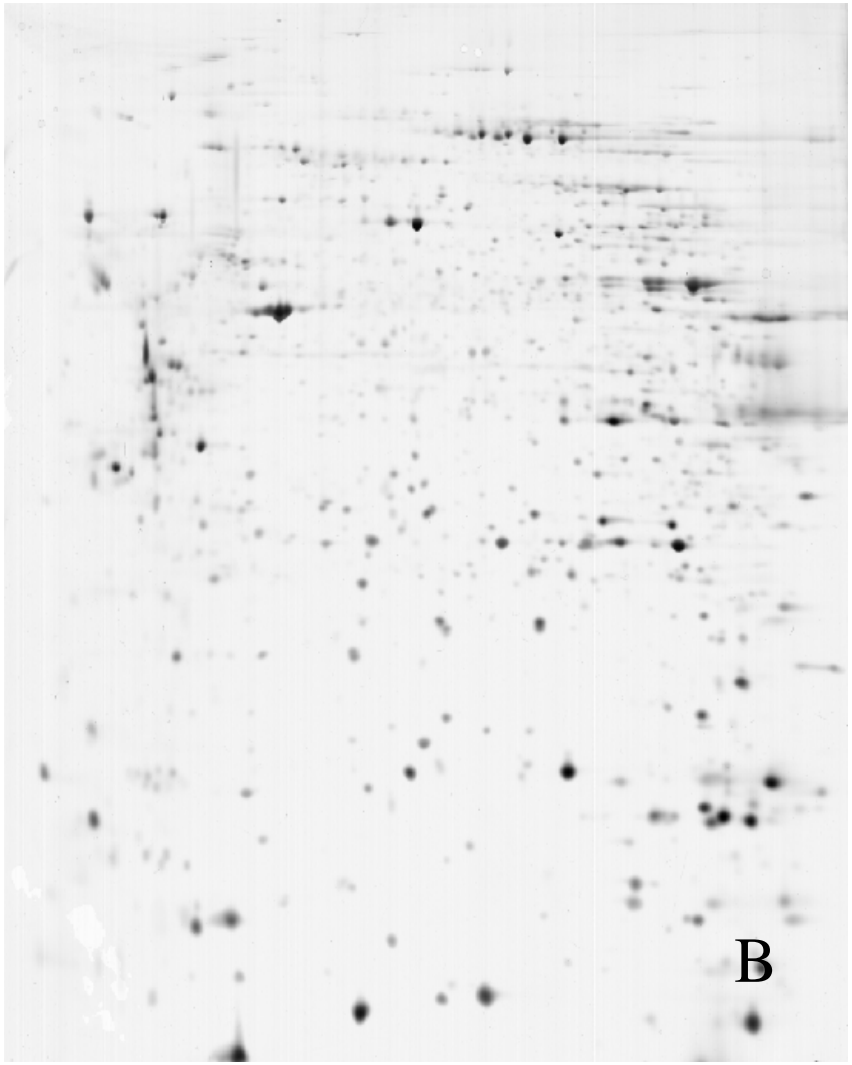
8 - 200 kDa
pI = 4,0 – 6,0

Human colon epithel protein map (2D gel electrophoresis of the cell lysate)
(MA Raymond et al. 1997).



Fibroblast lysate 2D gel electrophoresis
A: healthy, B: Down-syndrome

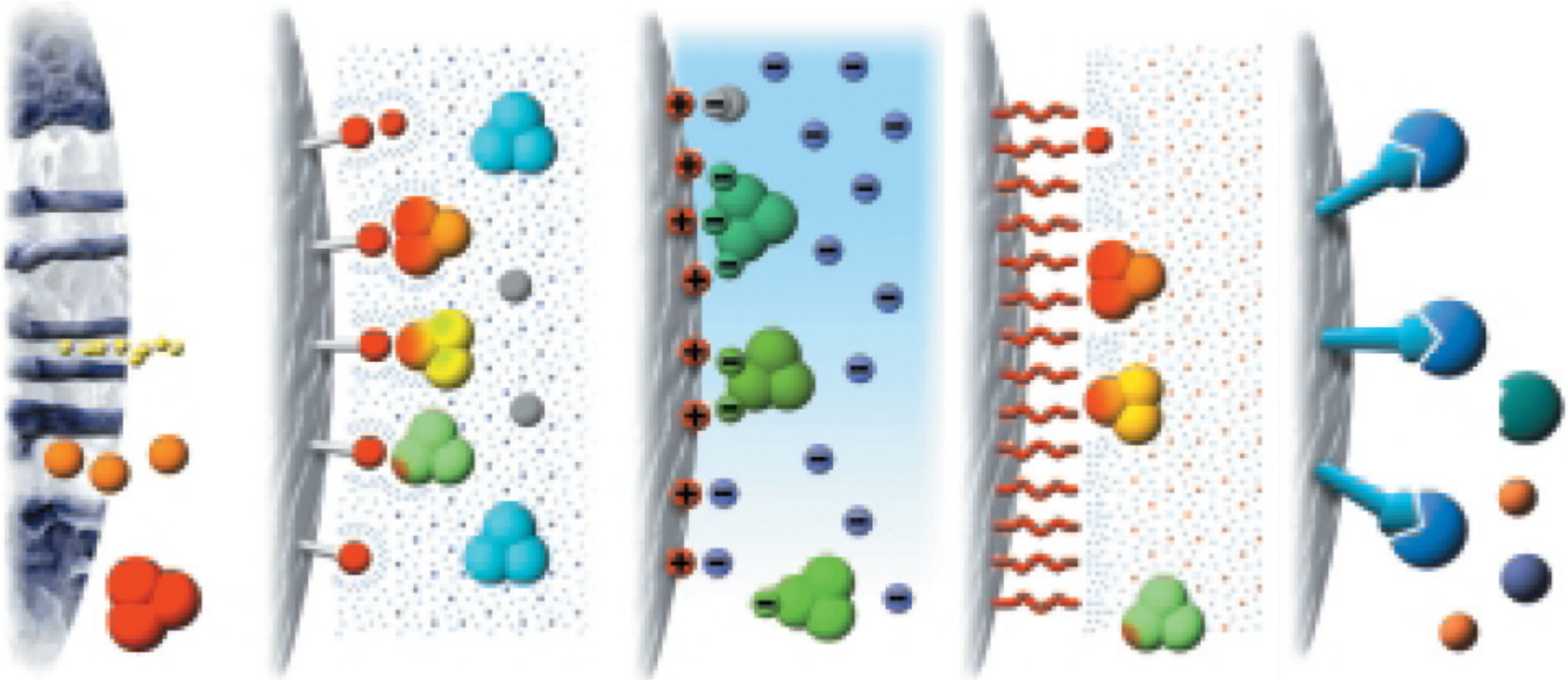




After separation…

- Cut the spots from the gel, wash (water, acetonitrile)
- Reduction: DTT (without this step the enzymatic sites might be hidden)
- Alkylation: iodoacetamide, wash with acetonitrile and dry (speedvac)
- In-gel digestion: enzymatic or chemical
- Reagents could be used for digestion:
  - Trypsin: Lys, Arg (protein: trypsin 50:1)
  - Thermolizin: Leu, Ile, Val, Phe, Met, Ala
  - BrCN: Met
- Extraction (ACN, formic acid; after evaporation, solve the sample in eluent used for HPLC-MS)

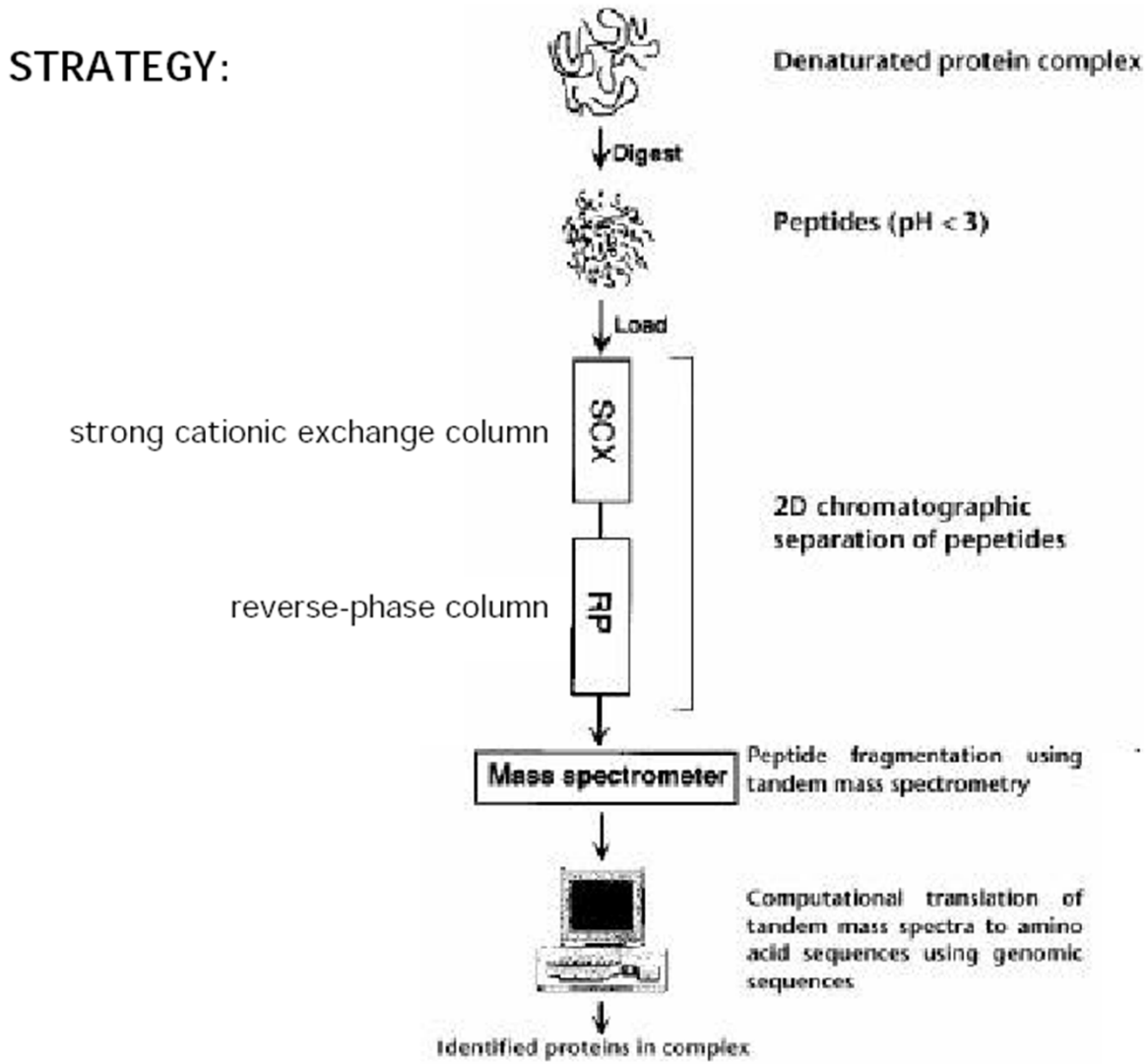
Non-gel based separation methods: chromatography for the analysis of peptides and proteins



Gelfiltration | Hydrophobic interaction | Ion-exchange | Reverse phase | Affinity

MudPIT

Multidimensional Protein Identification Technology → 2D chromatography



ICAT: Isotope Coded Affinity Tag

1. Protein isolation, isotope labeling

   Labeling with two different stable isotope (biotin is required for the isolation of the peptides)

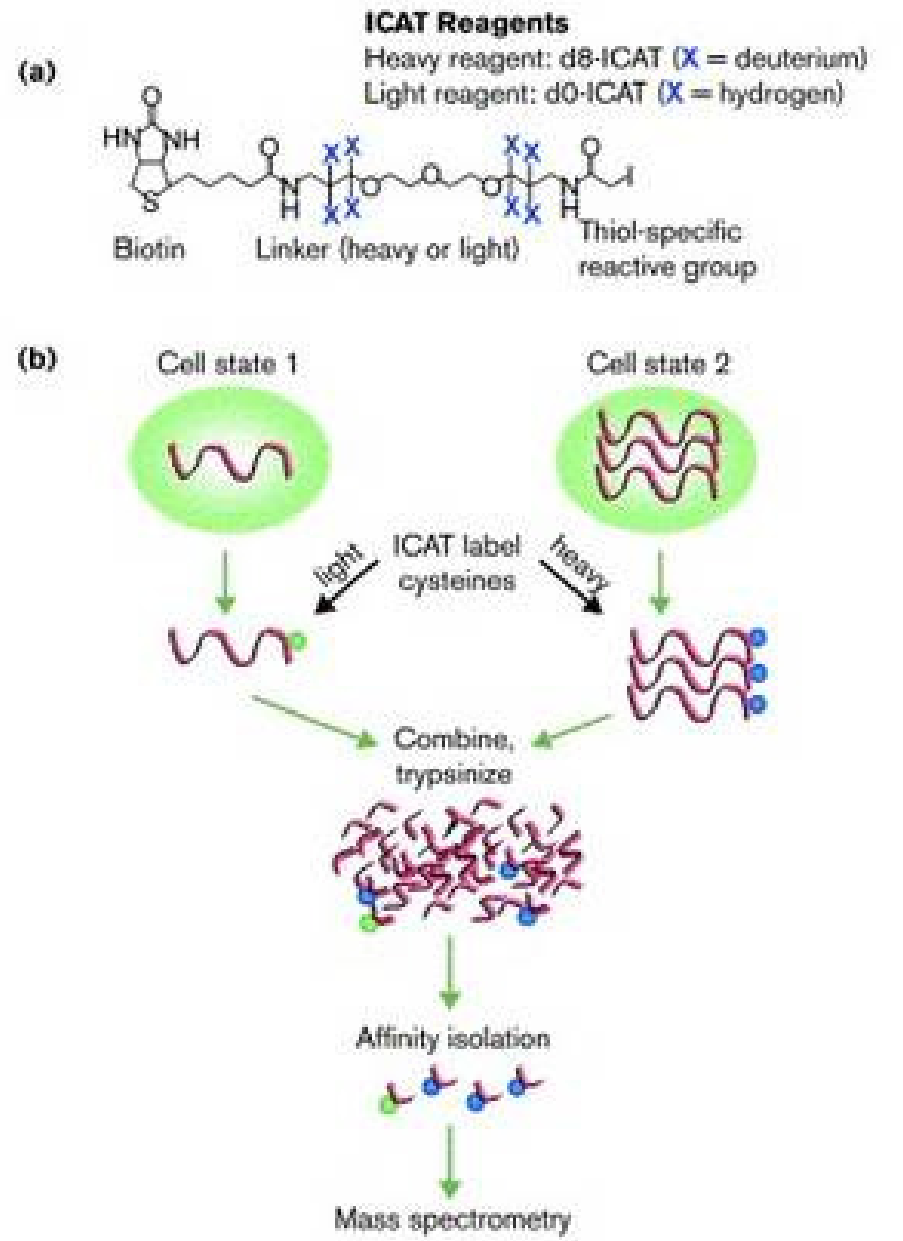
2. Mix the two samples labeled with different isotopes

3. Digestion

4. Separation

5. Mass spectrometry

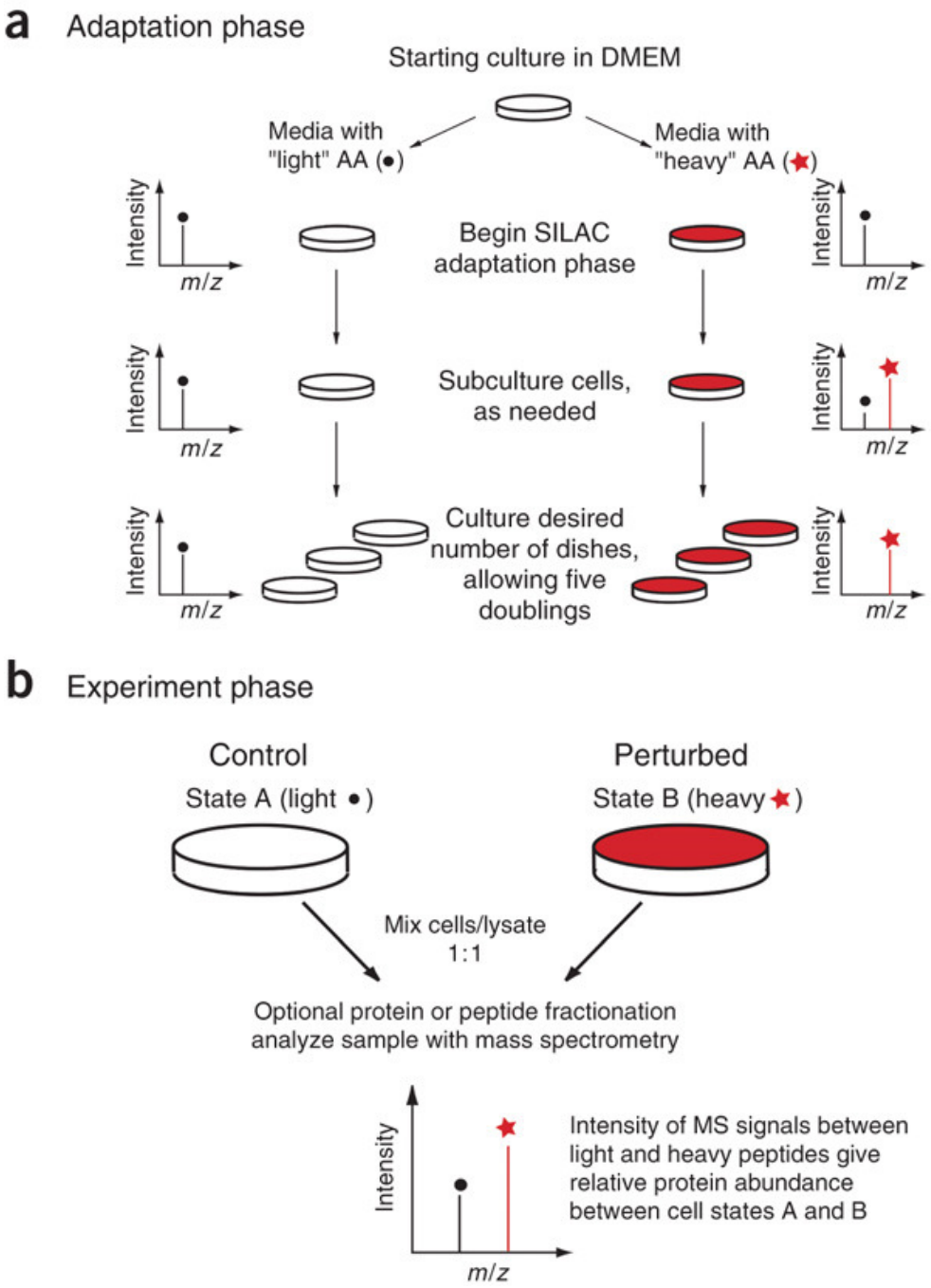
6. Identification from database



SILAC: Stable Isotope Labeling by Amino acids in Cell culture

Principle: incorporation of amino acids into the proteins of the cell.

Before treatment and analysis some cell division cycle are needed for the incorporation.

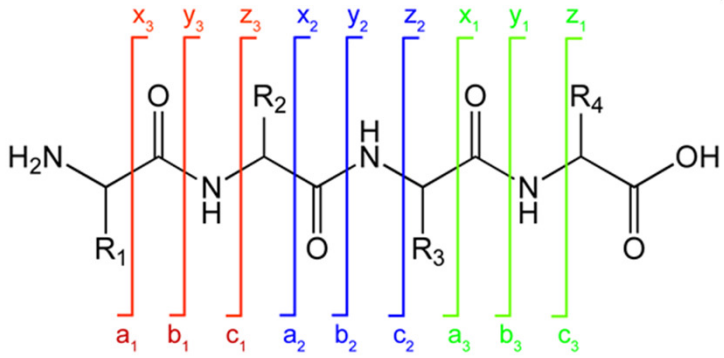


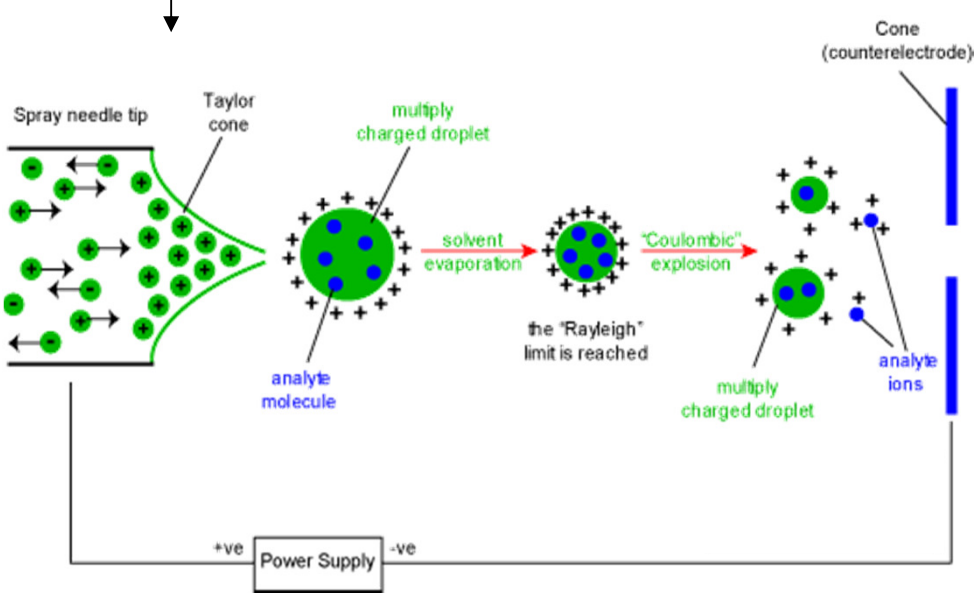
Mass spectrometry

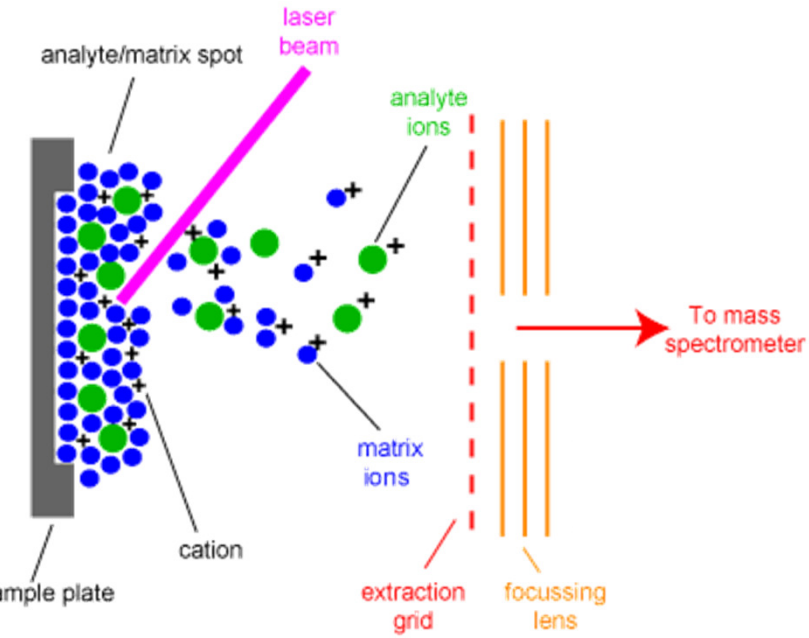
Ion source → Analyzer → Detector

MALDI
ESI

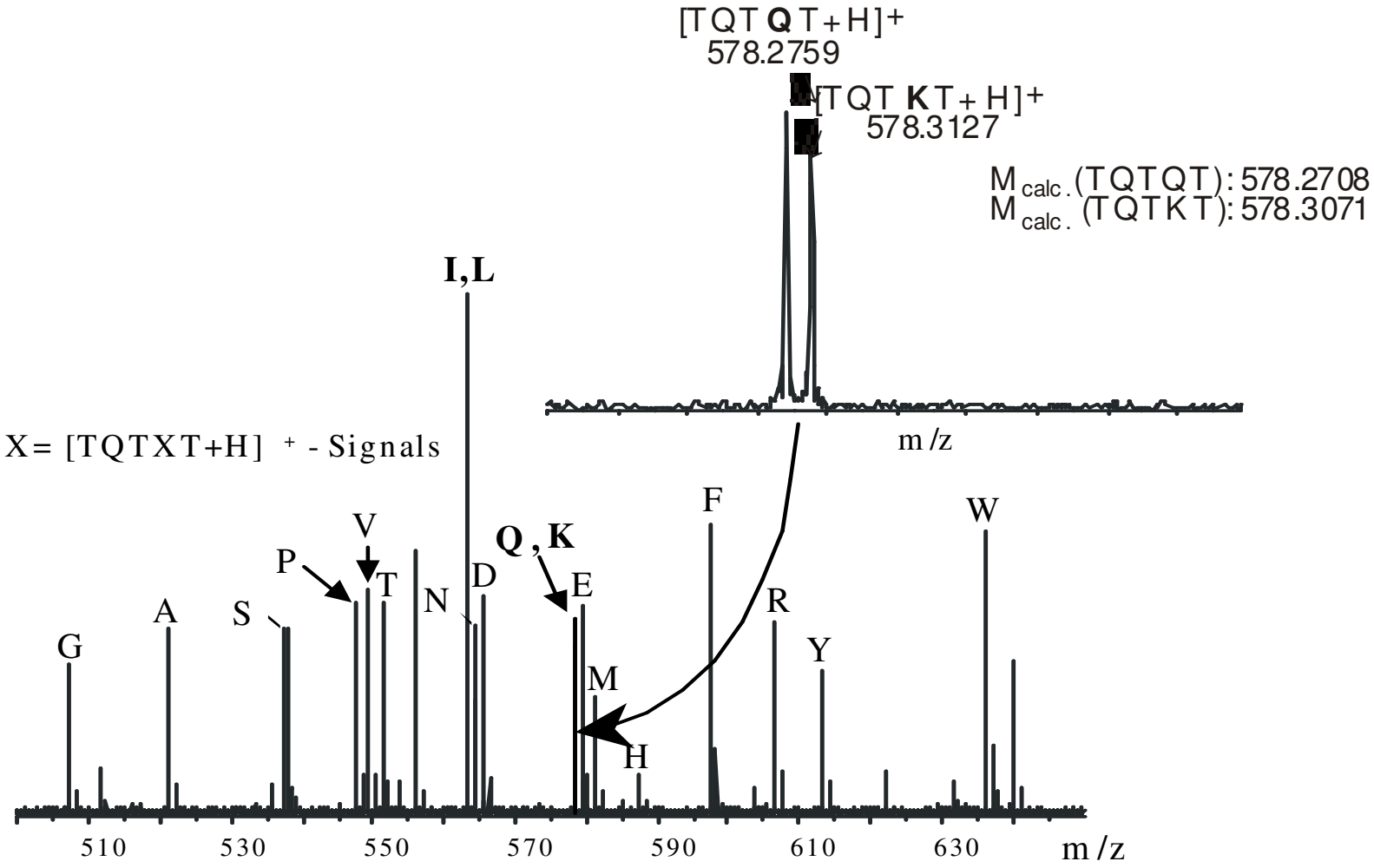
TOF, iontrap,
FT-ICR…







Study of TQTXT (X = 19 aminoacid) peptide library with ESI-FTICR mass spectrometry



TQTQT/TQTKT Ms difference: 0,368

Search in database

MS based search

Mass spectrum

↓

Peptide Mass Fingerprint (PMF)

↓

Protein identification based on molecular mass

MS-MS based search

MS/MS spectrum

↓

Sequence-specific information for every peptides

↓

Protein identification based on the sequence information of the peptides

For identification

At least 5 peptide mass should be identical

At least 2 fragmentation patterns should be identical

Searching engines

MS-based search

- Theoretical digestion of the protein
- Looking for the highest similarity between the experimental and the theoretical peptide masses

MS-MS based search

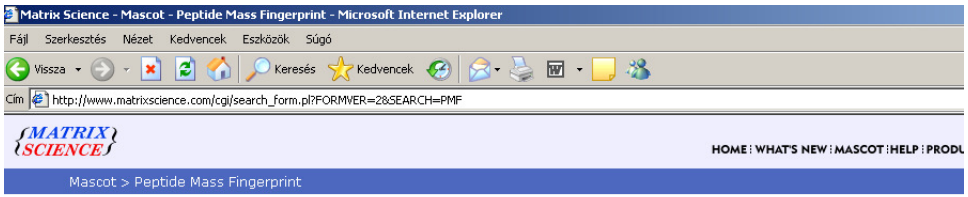
- Theoretical digestion of the protein
- Production of theoretical fragment ions for every possible tryptic peptides derived from the digestion
- Looking for the highest similarity for peptide identification
- Protein identification based on the peptide sequences

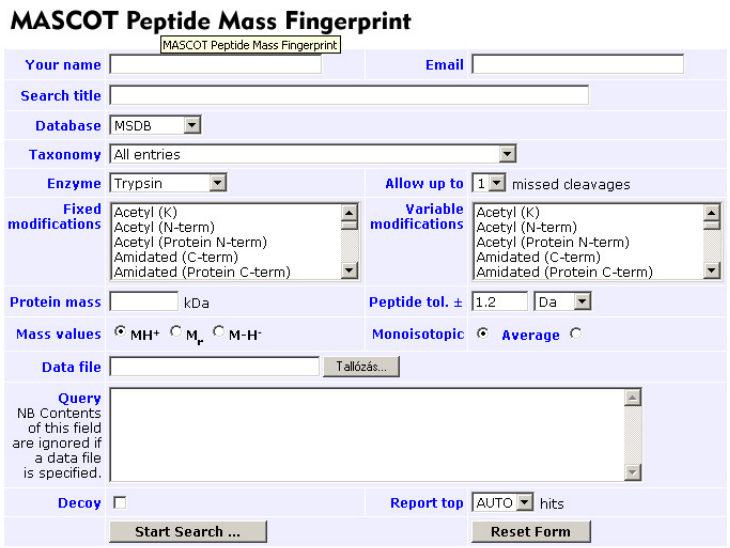
MASCOT: both search

Mascot

- Combined search of MS/MS, sequence and mass spectrometrical data

- Searching in MSDB, SwissProt and in NCBI database





ExPASy (Expert Protein Analysis System)

www.expasy.org

-2D gel electrophoresis database

-Technical information

-Searching based on protein name and identification number

Databases

International Protein Index (IPI)

http://www.ebi.ac.uk/IPI/IPIhelp.html

(October 2001)

Human Proteom, Collection of human protein sequence information from 5 databases

SWISS-PROT	Swiss-Prot is a manually curated biological database of protein sequences started by Amos Bairoch and developed by the Swiss Institute of Bioinformatics and the European Bioinformatics Institute
TrEMBL	TrEMBL, (Translated EMBL) is a very large protein database in SwissProt format generated by computer translation of the genetic information from the EMBL Nucleotide Sequence Database.
RefSeq	The Reference Sequence (RefSeq) database is an open access, annotated and curated collection of publicly available nucleotide sequences (DNA, RNA) and their protein products. This database is built by National Center for Biotechnology Information (NCBI), and, unlike GenBank, provides only a single record for each natural biological molecule (i.e. DNA, RNA or protein) for major organisms ranging from viruses to bacteria to eukaryotes.
Ensembl	Ensembl is a joint scientific project between the European Bioinformatics Institute and the Wellcome Trust Sanger Institute, which was launched in 1999 in response to the imminent completion of the Human Genome Project.

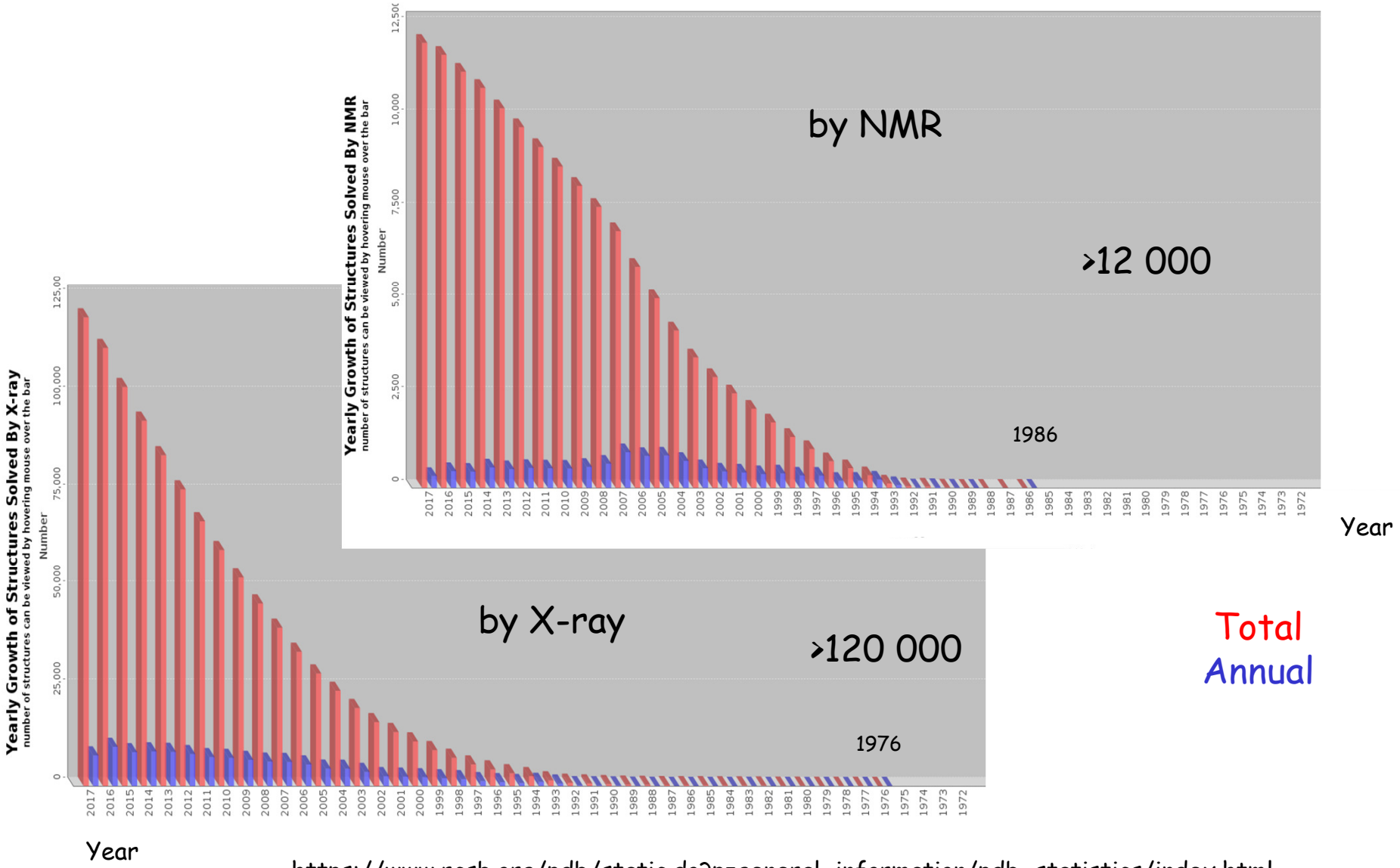
Protein

- Protein Sequence Databases: PIR, PRF, RefSeq, SwissProt, TrEMBL
- Protein Family Databases: InterPro, iProClass, Pfam, PIR-ALN, ProDom
- Protein Structure Databases: CSD, MMDB,NDH, PDB,
- Protein Structural Classification Databases: CATH, DSSP, FSSP, HSSP, SCOP,
- Post-translational Modification Databases: O-GlyBase, RESID,

Protein-protein interactions

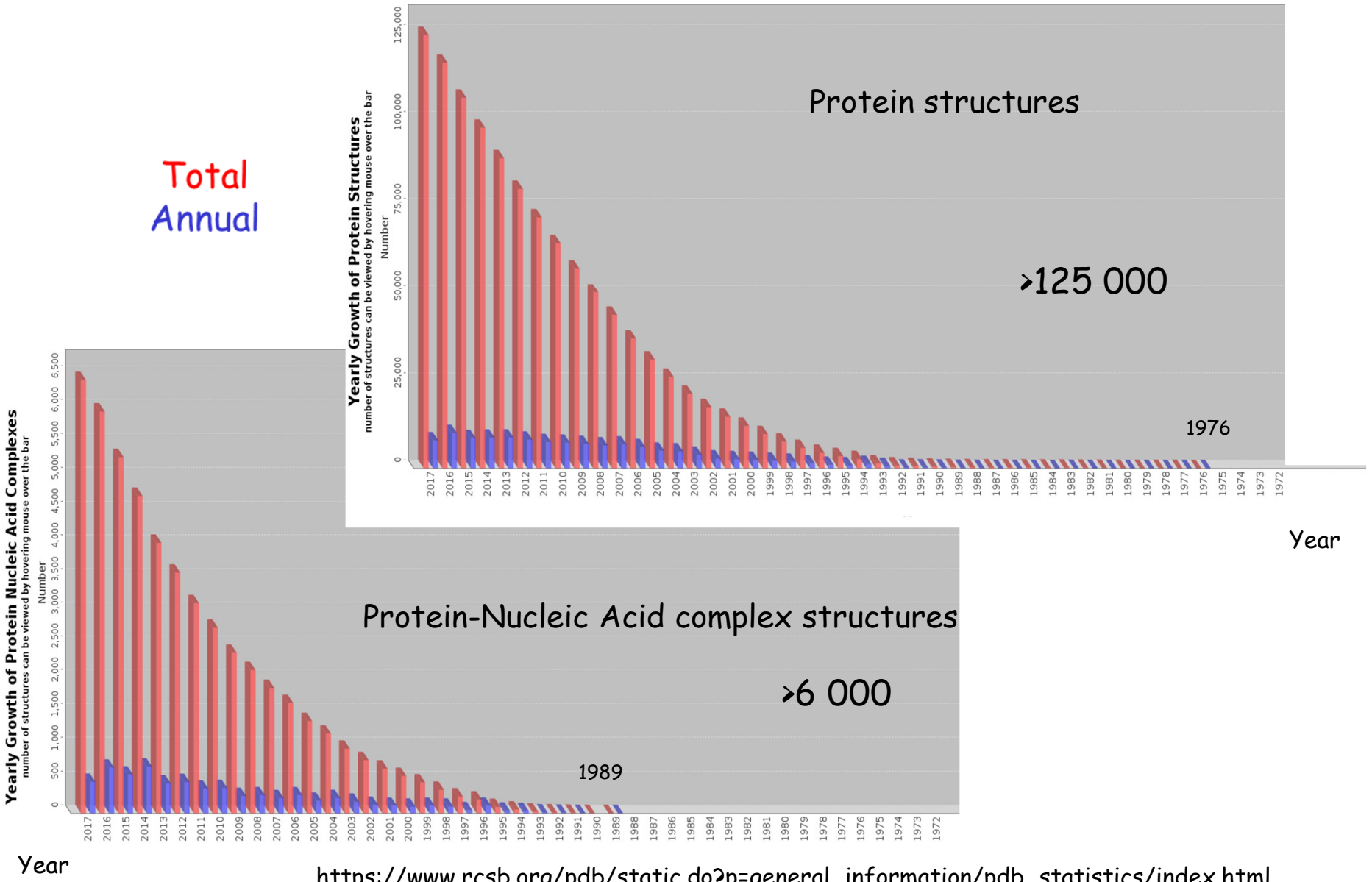
- **CuraGen:** Portal.curagen.com
- **DIP:** Dipode-mbi.ucla.edu
- **Interact:** Bioinf.man.ac.uk/interactso.htm
- **MIPS:** mips.biochem.mpg.de
- **ProNet:** Pronet.doublewist.com

Growth of released structures per year by experimental method



https://www.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/index.html

Growth of released structures per year by molecular type

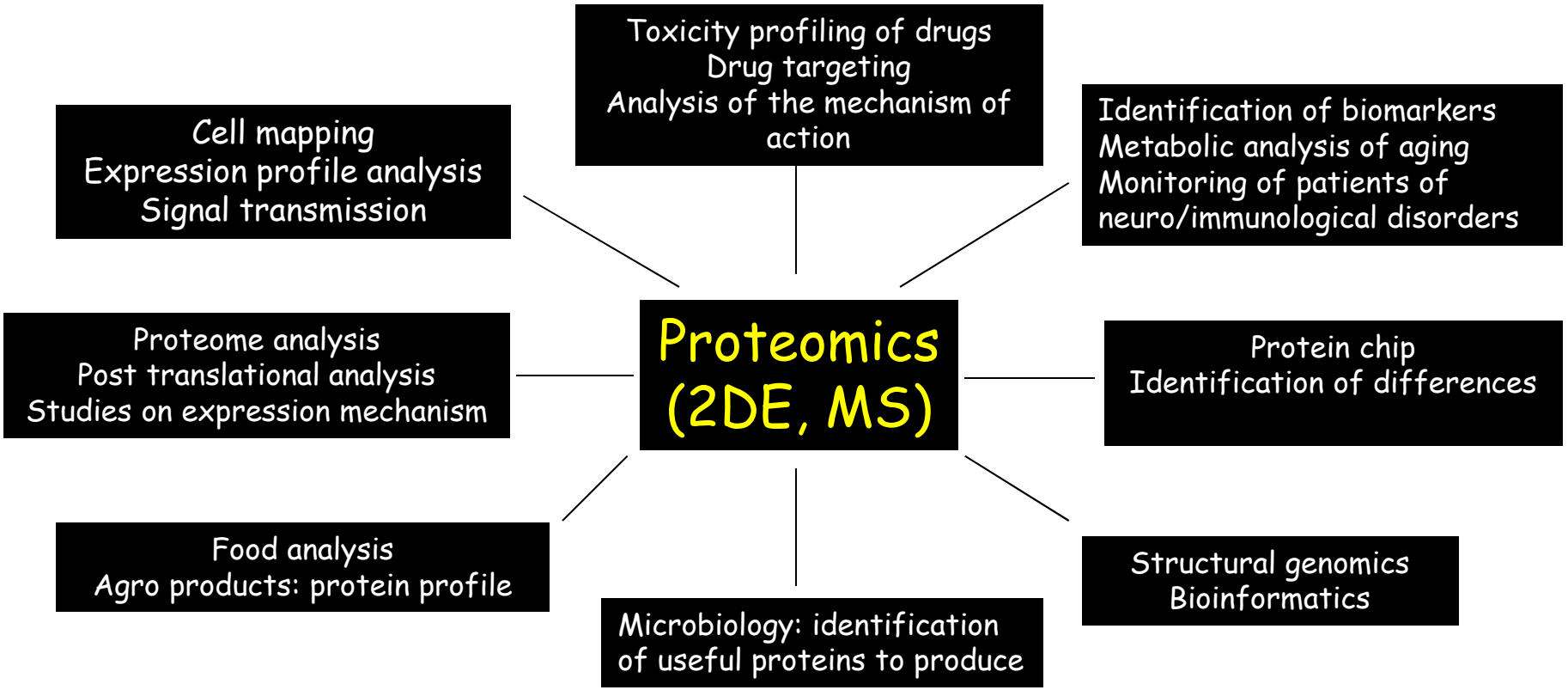


https://www.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/index.html

Advantages of the method

- Mass Spectrometry: high sensitivity and precise molecular mass deterimantion, user-friendly instruments
- Protein sequence databases: a lot of information
- Bioinformatics: searching engines → fast results

Application of proteomics



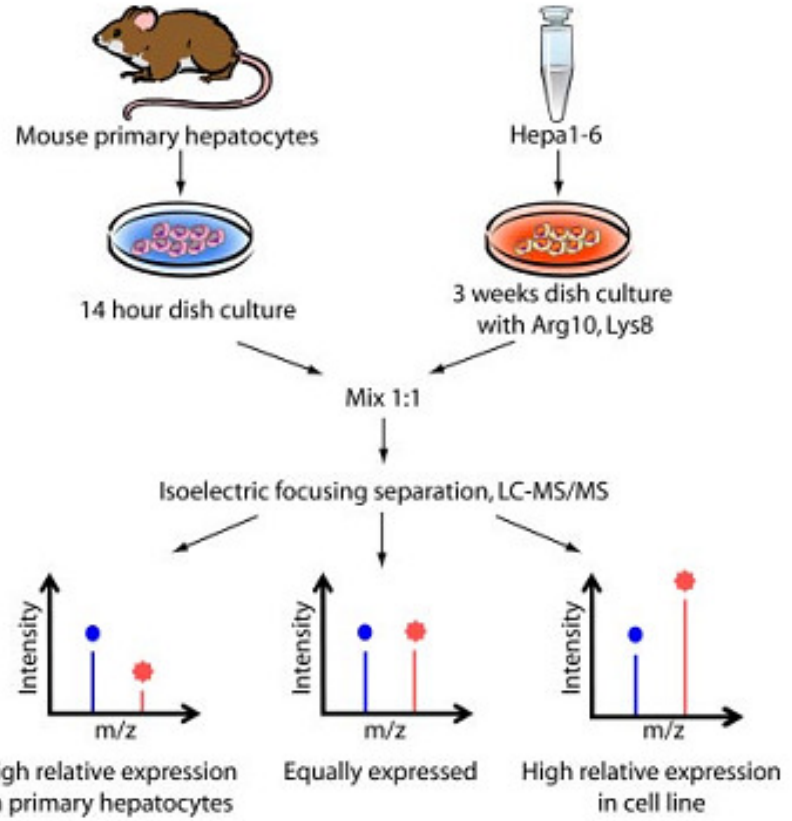
Example: Comparative Proteomic Phenotyping of Cell Lines and Primary Cells to Assess Preservation of Cell Type-specific functions (Pan et al., Molecular & Cellular Proteomics, 2009)

Aim: Comparison of tumor cells (Hepa 1-6) and normal hepatocytes

Method: SILAC

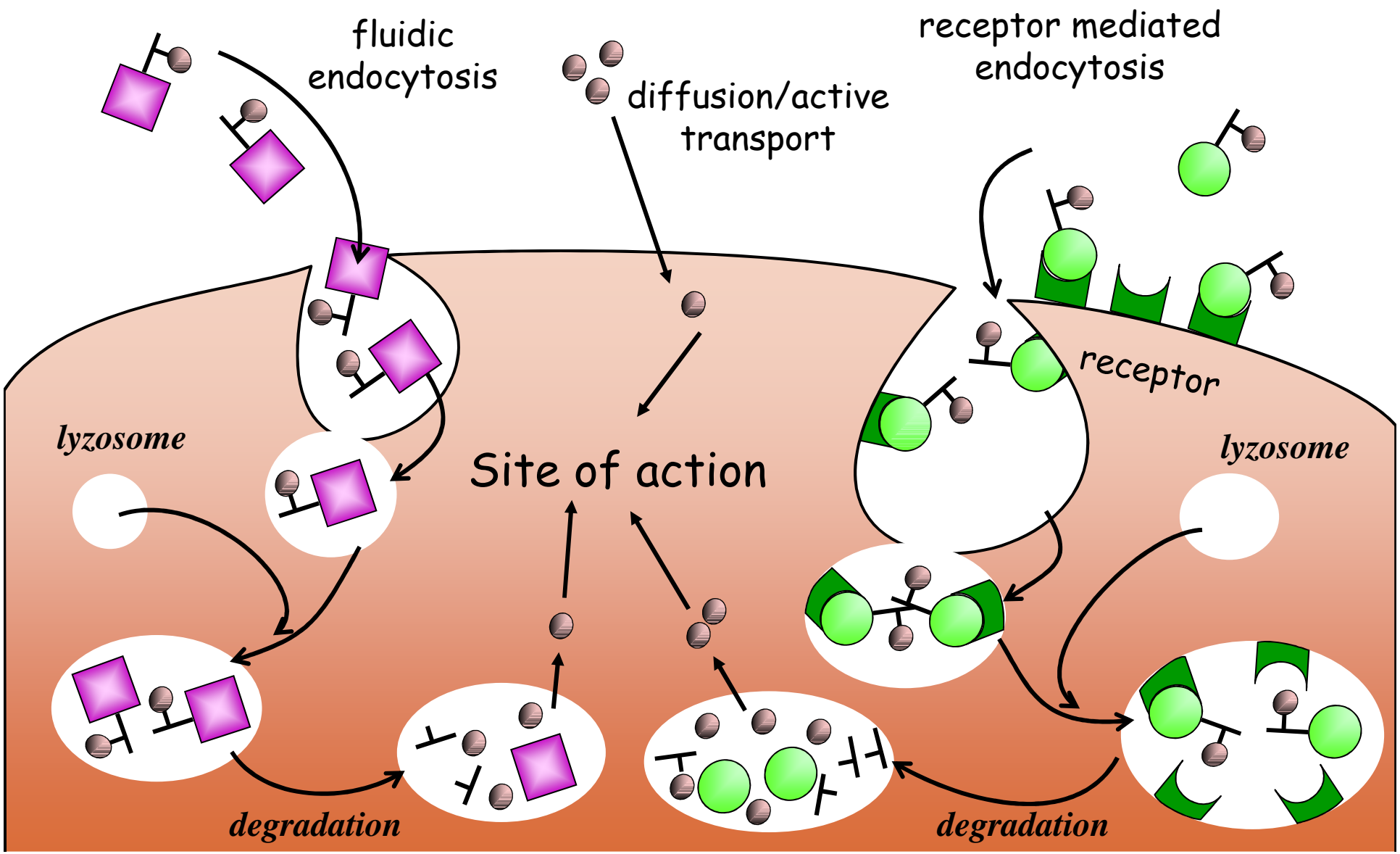
Results: Hepa 1-6 cells had less mitochondrial proteins, higher expression of cell cycle proteins was observed, less drug metabolism enzyme

Identification of 3350 proteins

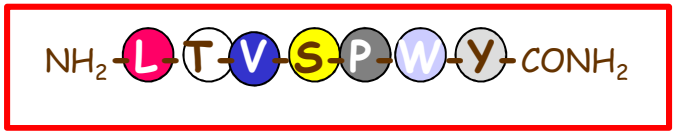


An example from our lab

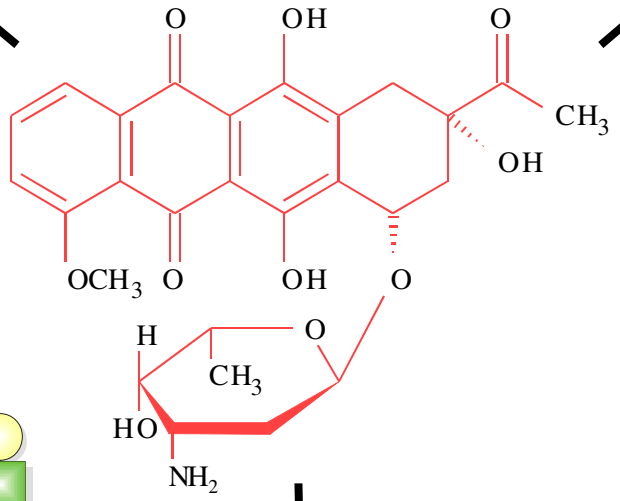
Uptake and liberation of bioactive entities



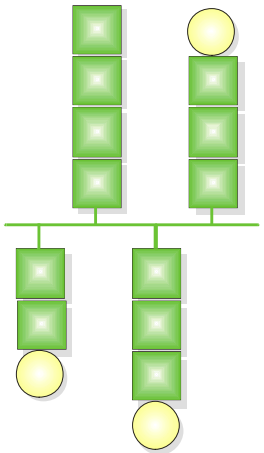
Daunomycin conjugates with oligo- or polypeptide



Orbán E. et al.:
Bioconjugate Chem.
92: 489-499 (2011)



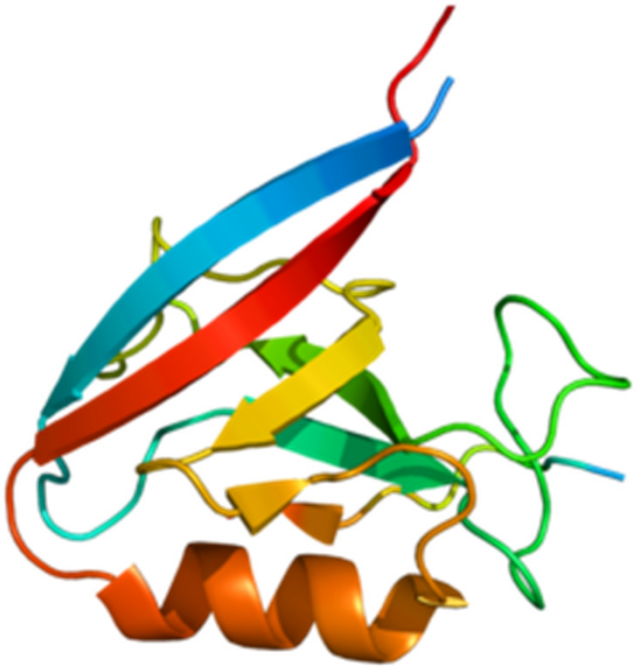
Sztaricskai F. et al.: *J Antibiotics (Tokyo),*
58: 704 (2005)
Bánóczi Z. et al. *Archivoc* **140**, (2008)
Miklán Zs. et al. *Biopolymers* **92**: 489 (2009)



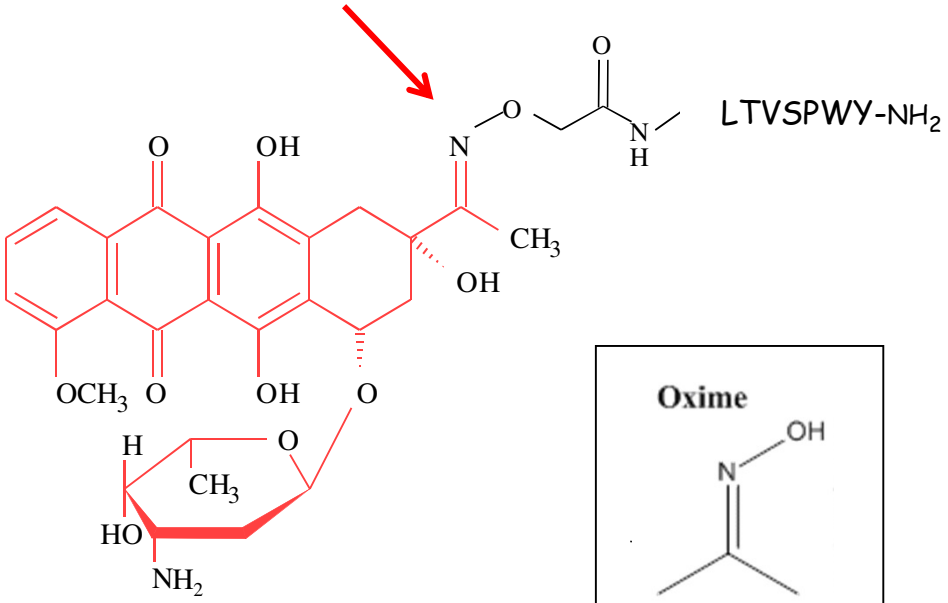
Hudecz F. et al. *Bioconjugate Chem.* **3**: 49 (1992)
Gaál D., Hudecz F. *Eur.J.Cancer.* **34**: 155 (1998)
Szabó R. et al. *Bioconjugate Chem.* **19**: 1078 (2008)
Reményi, J. et al. *Biochim. Biophys. Acta* **1798**: 2209 (2010)

ErbB2 receptor and peptide LTVSPWY

- ErbB2 receptor is overexpressed on several tumor cells, a potential target in cancer chemotherapy
- Peptide LTVSPWY derived from a phage library as an ErbB2 binding ligand

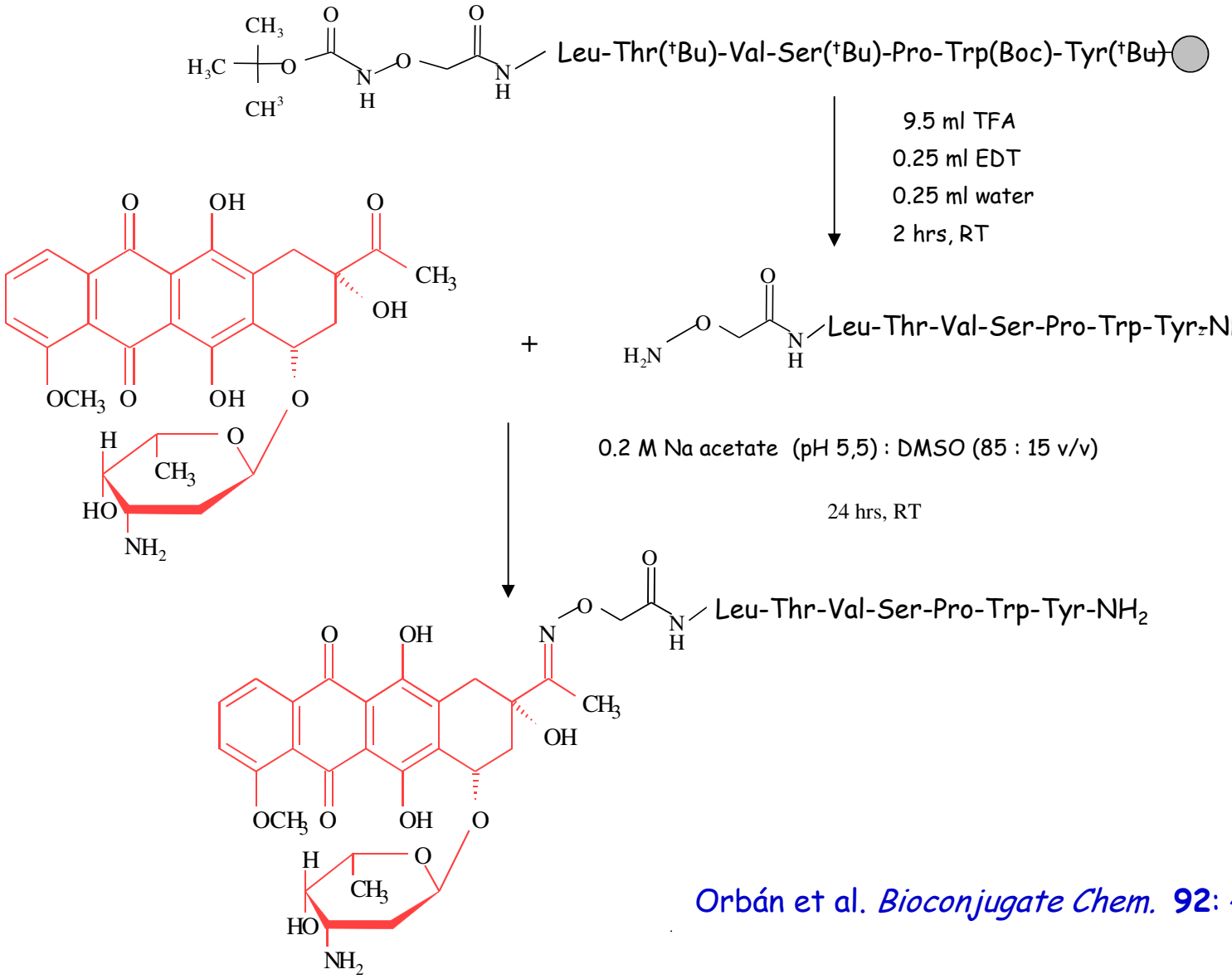


http://www.genenames.org



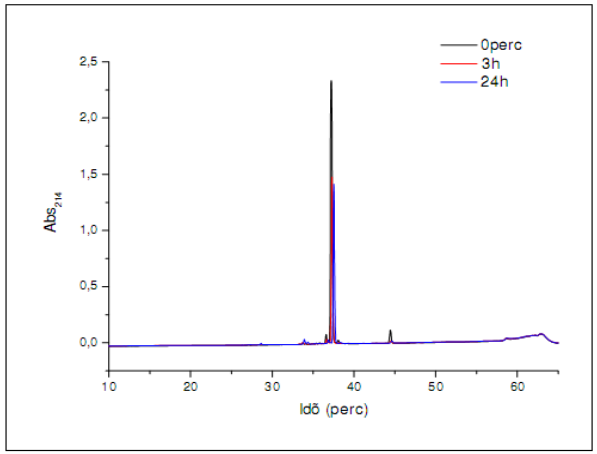
[1]Wang et al., *Cancer Res 67*(7), 3337-44, 2007.
[2]Seoane et al., *J Natl Cancer Inst 102* (18), 1432-46, 2010.
[3]Shadidi et al., *Drug Resist Updat 6* (6), 363-71, 2003.

Synthesis of Dau-heptapeptide conjugate with oxime bond

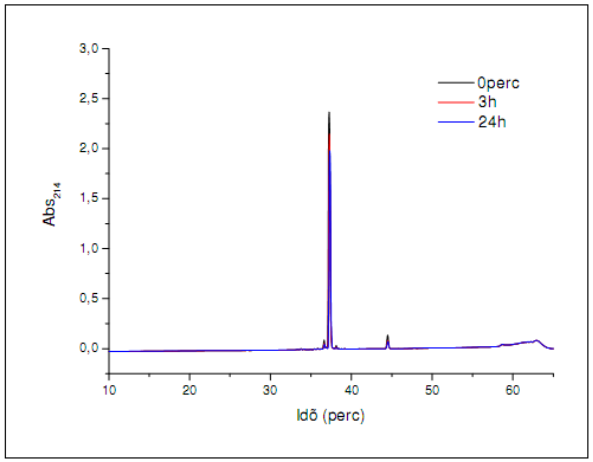


Orbán et al. *Bioconjugate Chem.* **92**: 489 (2011)

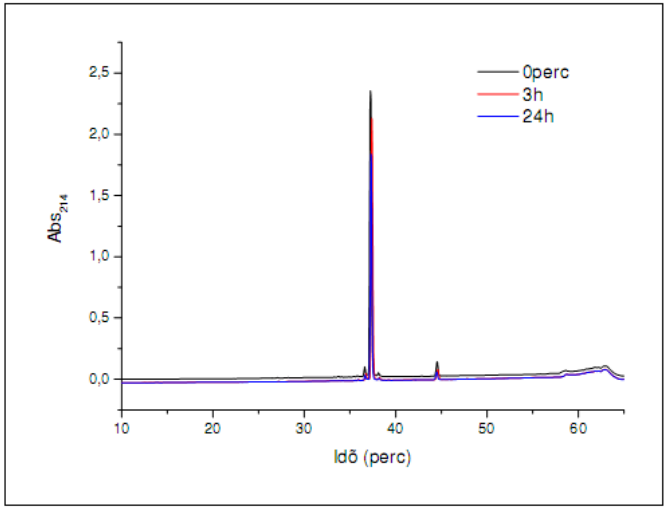
pH stability of of Dau=Aoa-LTVSPWY-amide conjugate with oxime bond



Analytical RP-HPLC chromatogram
in 0.1 M Na citrate buffer, pH 2.5



Analytical RP-HPLC chromatogram
in 0.1 M Na citrate buffer, pH 5.0



Analytical RP-HPLC chromatogram
in 0.1 M Na citrate buffer, pH 7.0

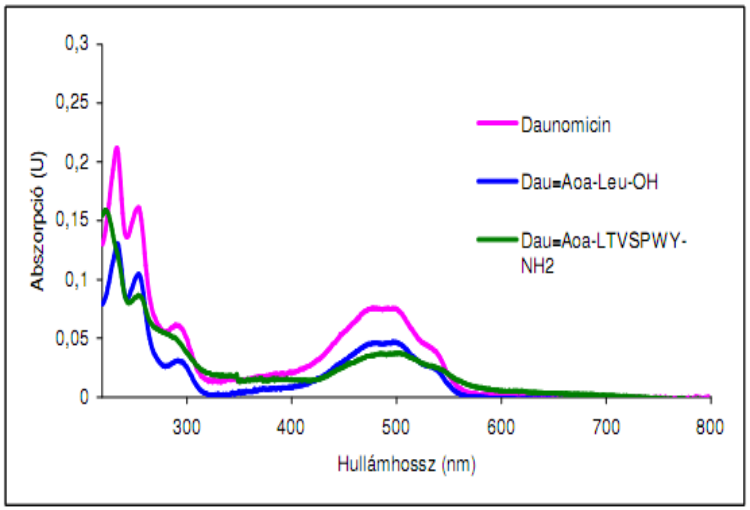
Characteristics of Dau=Aoa-LTVSPWY-amide conjugate with oxime bond

Compound	MS[a] [M]		R_t[b] (min)	IC_{50} (μM)	Uptake[c]
	Calc.	Measd.			
Dau=Aoc-LTVSPWY-NH_2	1224,3	1224,3	27,0	2,07	3179
Dau	527,5	n.a.	34,9	0,05	48621

[a] SELDI-MS

[b] HPLC, Column: Supelcoil LC-18-DB (C18, 120 A, 5 μm, 4,6 x 250 mm), gradient elution: 0-5 min 5% eluent B, 5-50 min 90% eluent B, where eluent A: 0.1 % TFA in water, eluent B: 0.1% TFA in AcN-water (80-20 v/v %)

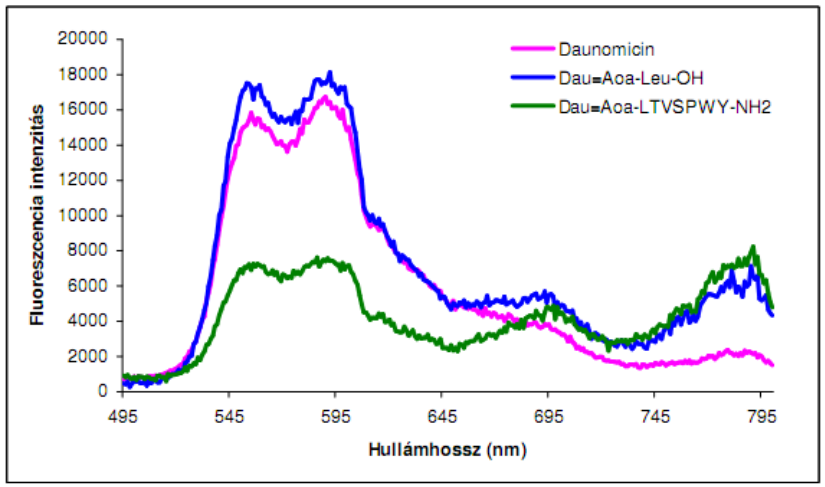
Absorbtion spectra



in 0.1 M Tris buffer, pH 7.4
c = 1,8 x 10^{-5} M (Dau)

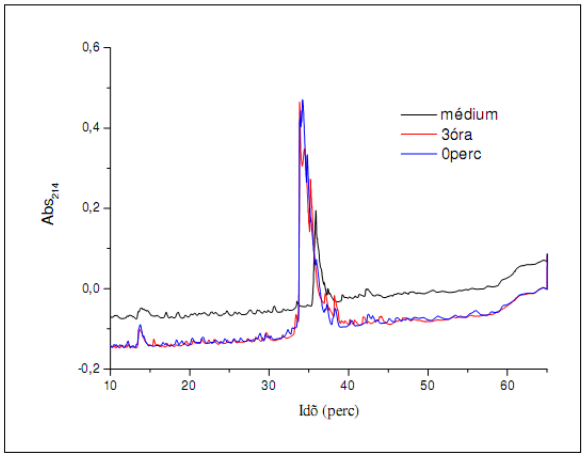
Emission spectra

λ_{ex} = 473 nm



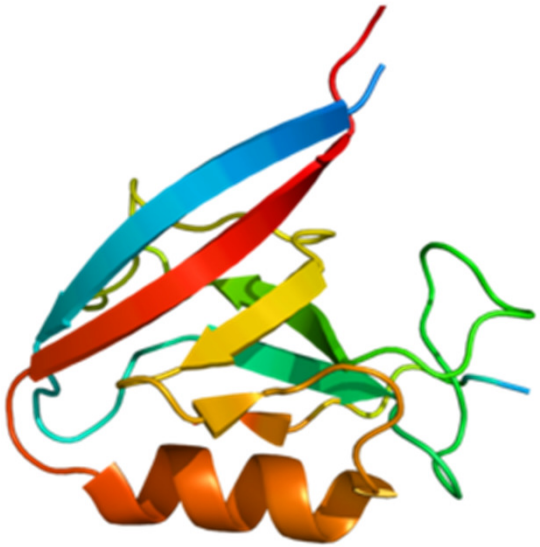
in 0.1 M Tris buffer, pH 7.4
c = 1,8 x 10^{-5} M (Dau)

In vitro cytotoxicity and uptake of Dau=Aoa-LTVSPWY-amide conjugate

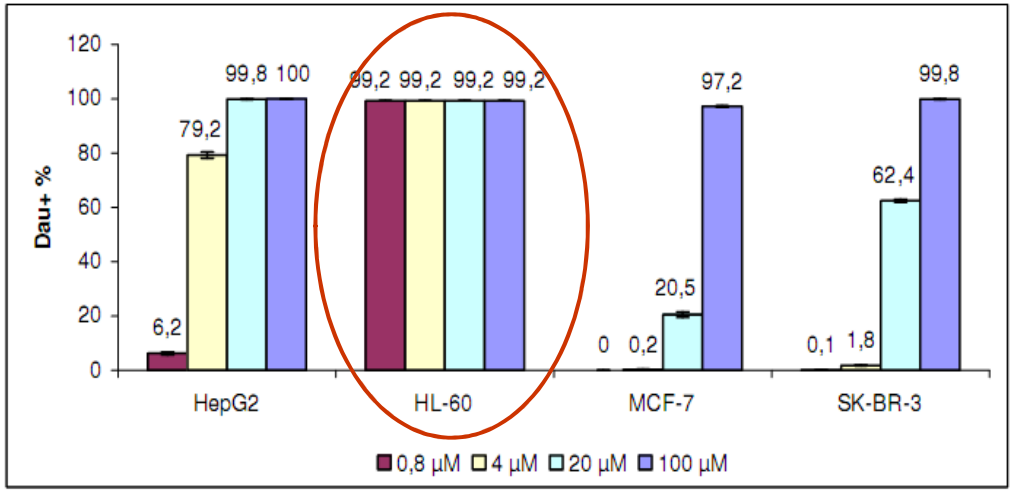


Stability in RPMI-1640 medium with 10% sera

In vitro cytotoxicity		
Cell line	IC_{50} ± s.d. (μM)	
	Conjugate	Dau
HepG2	3.07 ± 0,02	0.66 ± 0.21
HL-60	0.53 ± 0.12	0.05 ± 0,03
MCF-7	7.42 ± 0.5	0.18 ± 0.09
SK-BR-3	37.9 ± 2.64	3.64 ± 0.52

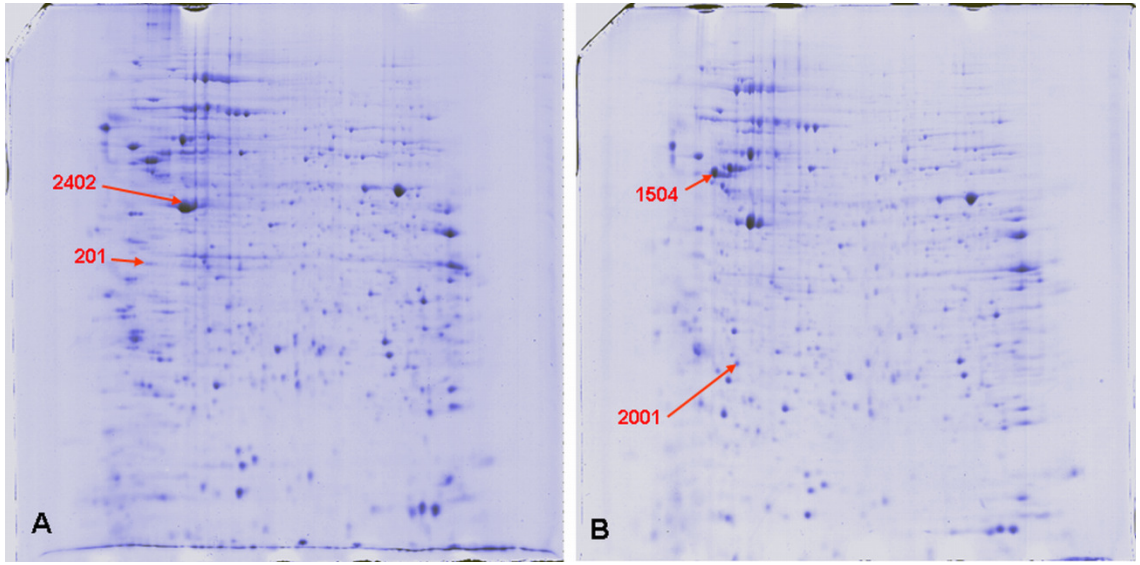


http://www.genenames.org



Cellular uptake c = 0.8 – 100 μM, 90 min

Protein expression profile of non-treated (A) and Dau-treated HL-60 cells (B).



t = 24 hr

Identified proteins	Average protein expression level in	
	Controll	Dau
Actin, cytoplasmic 1 (Beta-actin)	11178.5	1615.8
Proliferating cell nuclear antigen (PCNA) (Cyclin)	1440.7	171.5
Ran-specific GTPase-activating protein (Ran-binding protein 1)	789.7	1648.5
Tubulin beta chain (Tubulin beta-5 chain)	1337.6	9713.9

Arrows and spot numbers show the significantly different spots on the gel where expression level was higher.

Protein expression profile of non-treated (A) and Dau-treated HL-60 cells (B).

Identified Protein	Spot number	Average level in		Fold-change	Mascot score	M_r(Da)	pI
		Control	Dau				
Proliferating cell nuclear antigen (PCNA) (Cyclin)	201	1440.7	171.5	0.12	2111	28768.78	4.57
Tubulin beta chain (Tubulin beta-5 chain)	1504	1337.6	9713.9	7.26	11510	49670.82	4.78
Ran-specific GTPase-activating protein (Ran-binding protein 1) (RanBP1)	2001	789.7	1648.5	2.09	560	23310.12	5.19
Actin, cytoplasmic 1 (Beta-actin)	2402	11178.5	1615.8	0.14	17877	41736.73	5.29

Spot number: for the identification on the gel.

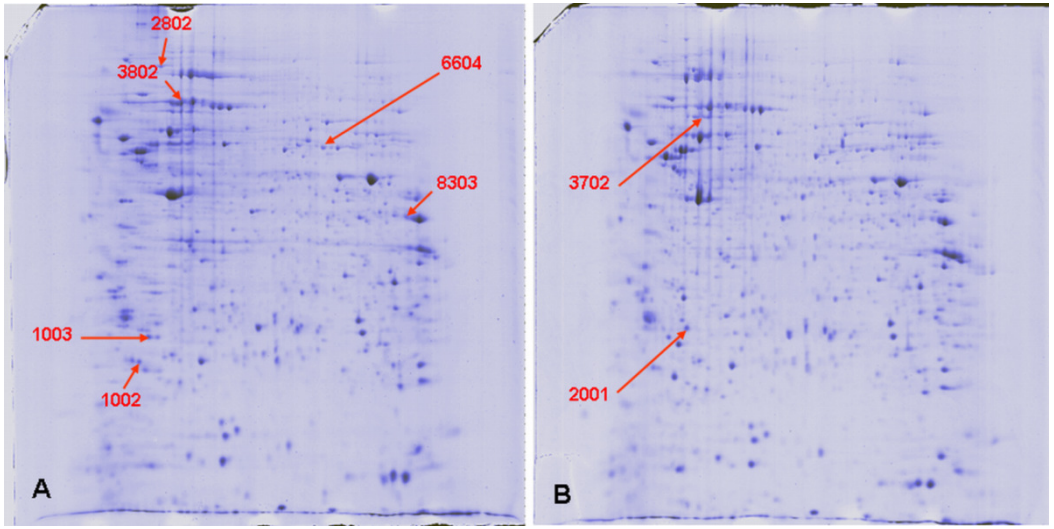
Proteins with different expression level were identified after tryptic in-gel digestion using OrbiTrap nano-LC-MS/MS mass spectrometry and MASCOT database. **Average levels of the protein**: calculated by PDQuest 8.0 software.

Fold change: the ratio of the average protein expression level in the conjugate and Dau-treated samples.

M_r : the theoretical molecular weight

pI : the theoretical isoelectric point of the identified protein.

Protein expression profile of non-treated (A) and Dau=Aoa-LTVSPWY-NH_2 conjugate-treated HL-60 cells (B)



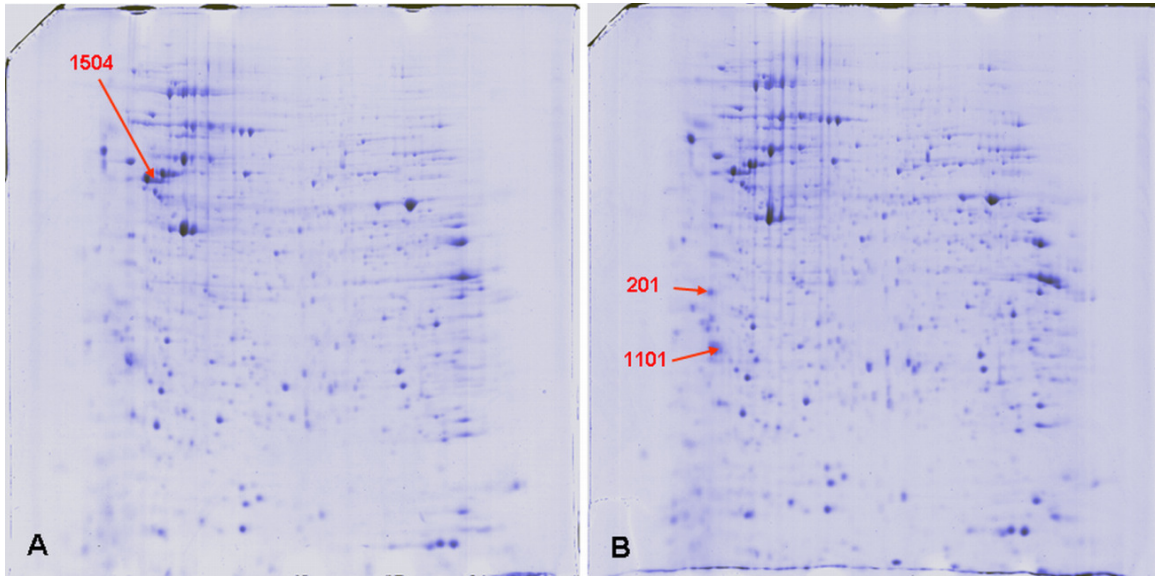
t = 24 hr

Identified proteins	Average protein expression level in	
	Control	Dau=Aoa-LTVSPWY-NH_2
D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95) (3-PGDH)	1067.9	526.0
Fructose-bisphosphate aldolase A (EC 4.1.2.13) (Muscle-type aldolase)	2105.1	999.3
Heat shock cognate 71 kDa protein (Heat shock 70 kDa protein 8)	6507.8	5.6
Plastin-2 (L-plastin) (Lymphocyte cytosolic protein 1) (LCP-1)	482.3	1065.6
Ran-specific GTPase-activating protein (Ran-binding protein 1)	789.7	1805.0
Rho GDP-dissociation inhibitor 2 (Rho GDI 2) (Rho-GDI beta) (Ly-GDI)	2253.9	5.6
Transitional endoplasmic reticulum ATPase (Valosin-containing protein) (VCP)	994.4	271.7
Translationally-controlled tumor protein (TCTP) (p23) (Histamine-releasing factor)	1406.2	5.6

Protein expression profile of non-treated (A) and Dau=Aoa-LTVSPWY-NH_2 conjugate-treated HL-60 cells (B)

Identified Protein	Spot number	Average level in Control	Average level in Dau=Aoa-LTVSPWY-NH_2	Fold-change	Mascot score	M_r(Da)	pI
Translationally-controlled tumor protein (TCTP) (p23) (Histamine-releasing factor) (HRF)	1002	1406.2	5.6	0.004	3665	19595.34	4.84
Rho GDP-dissociation inhibitor 2 (Rho GDI 2) (Rho-GDI beta) (Ly-GDI)	1003	2253.9	5.6	0.002	3288	22988.01	5.10
Ran-specific GTPase-activating protein (Ran-binding protein 1) (RanBP1)	2001	789.7	1805.0	2.29	560	23310.12	5.19
Transitional endoplasmic reticulum ATPase (TER ATPase) (15S Mg(2+)-ATPase p97 subunit) (Valosin-containing protein) (VCP)]	2802	994.4	271.7	0.27	2657	89321.80	5.14
Plastin-2 (L-plastin) (Lymphocyte cytosolic protein 1) (LCP-1) (LC64P)	3702	482.3	1065.6	2.21	4924	70288.39	5.29
Heat shock cognate 71 kDa protein (Heat shock 70 kDa protein 8)	3802	6507.8	5.6	0.001	10972	70898.09	5.37
D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95) (3-PGDH)	6604	1067.9	526.0	0.49	4131	56650.5	6.29
Fructose-bisphosphate aldolase A (EC 4.1.2.13) (Muscle-type aldolase) (Lung cancer antigen NY-LU-1]	8303	2105.1	999.3	0.47	1143	39420.02	8.30

Protein expression profile of Dau-treated (A) and Dau=Aoa-LTVSPWY-NH_2 conjugate-treated HL-60 cells (B)



t = 24 hr

Identified proteins	Average protein expression level in	
Proliferating cell nuclear antigen (PCNA) (Cyclin	171.5	2165.8
14-3-3 protein gamma (Protein kinase C inhibitor protein 1) (KCIP-1)	157.7	1814.0
Tubulin beta chain (Tubulin beta-5 chain	9713.9	1981.3

A = Dau, c = 0.024 μM

B = Dau=Aoa-LTVSPWY-NH_2 conjugate c = 9 μM

Arrows and spot numbers show the significantly different spots on the gel where expression level was higher.

Protein expression profile of Dau-treated (A) and Dau=Aoa-LTVSPWY-NH_2 conjugate-treated HL-60 cells (B)

		Average level in					
Identified Protein	Spot number	Dau	Dau=Aoa-LTVSPWY-NH_2	Fold-change	Mascot score	Mr(Da)	pI
Proliferating cell nuclear antigen (PCNA) (Cyclin)	201	171.5	2165.8	12.6	2111	28768.78	4.57
14-3-3 protein gamma (Protein kinase C inhibitor protein 1) (KCIP-1)	1101	157.7	1814.0	11.5	3116	28302.59	4.80
Tubulin beta chain (Tubulin beta-5 chain)	1504	9713.9	1981.3	0.2	11510	49670.82	4.78

Spot number: for the identification on the gel.

Proteins with different expression level were identified after tryptic in-gel digestion using OrbiTrap nano-LC-MS/MS mass spectrometry and MASCOT database. **Average levels of the protein**: calculated by PDQuest 8.0 software.

Fold change: the ratio of the average protein expression level in the conjugate and Dau-treated samples.

M_r : the theoretical molecular weight

pI : the theoretical isoelectric point of the identified protein.

Comparison of protein expression profiles of Dau- and Dau-peptide conjugate- and non-treated HL-60 cells: an interpretation

		Average level in			
Identified Protein	Control	Dau	Fold-change	Dau=Aoa-LTVSPWY-NH_2	Fold-change
Proliferating cell nuclear antigen (PCNA) (Cyclin)	1440.7	171.5 ↓	0.12	2165.8 ↑	12.6
Tubulin beta chain (Tubulin beta-5 chain)	1337.6	9713.9 ↑	7.26	1981.3 ↓	0.2
Ran-specific GTPase-activating protein (Ran-binding protein 1)	789.7	1648.5	2.09	No change	
Actin, cytoplasmic 1 (Beta-actin)	11178.5	1615.8	0.14	No change	
14-3-3 protein gamma (Protein kinase C inhibitor protein 1) (KCIP-1)		157.7	No change	1814.0	11.5

1. Cyclin and tubulin beta-5 are involved in both processes.
2. Ran-binding protein 1 and actin are involved in Dau action.
3. 14-3-3 protein gamma is involved in Dau-conjugate action.

Conclusions

The expression level of several proteins altered due to the treatment with the free drug (Dau) or its conjugate in comparison with proteins from untreated cells.

After treatment with Dau for 24 h, the expression levels of cytoskeletal as well as cell-cycle regulatory proteins (four) have been changed.

Three proteins were identified, whose expression was lower (tubulin beta chain) or markedly higher (proliferating cell nuclear antigen and protein kinase C inhibitor protein 1) after administration of HL-60 cells with Dau-peptide conjugate vs free drug. These proteins are cytoskeletal proteins or involved in signalisation or metabolism.