

**A.Y. 2018/19**

***BIOANALYTICAL PROTEOMICS  
AND INTERACTOMICS***

**Elisa Michelini**

Dip. Chimica “G. Ciamician”, Università di Bologna

Via Selmi 2

Policlinico S. Orsola-Malpighi, Via Massarenti 9, Pad 11

[elisa.michelini8@unibo.it](mailto:elisa.michelini8@unibo.it)

Tel 051/ 343398

Cell. 338/6860955

# Timetable

November 2018	December 2018	January 2019	February 2019
1 Th	1 Sa	1 Tu New Year's Day	1 Fr
2 Fr	2 Su	2 We	2 Sa
3 Sa	3 Mo BDB: 9-13 ; LB1: 14-18	3 Th	3 Su
4 Su	4 Tu PB: 10-13 ; PB: 14-17	4 Fr	4 Mo LB1 (A. Via)
5 Mo	5 We BPI: 9-13; AG: 14-17	5 Sa	5 Tu LB1 (A. Via)
6 Tu	6 Th BPI: 9-13; LB1: 14-18	6 Su	6 We LB1 (A. Via)
7 We	7 Fr PB: 10-13 ; AG: 14-17	7 Mo PB: 10-13 ; BDB: 14-18	7 Th LB1 (A. Via)
8 Th	8 Sa	8 Tu BDB: 9-13 ; LB1: 14-18	8 Fr LB1 (A. Via)
9 Fr	9 Su	9 We BPI: 9-13; AG: 14-17	9 Sa
10 Sa	10 Mo BDB: 9-13 ; LB1: 14-18	10 Th BPI: 9-13; PB: 14-17	10 Su
11 Su	11 Tu PB: 10-13 ; PB: 14-17	11 Fr BPI: 9-13 ; AG: 14-17	11 Mo Bologna Winter School
12 Mo BDB: 9-13 ; AG: 14-17	12 We BPI: 9-13; AG: 14-17	12 Sa	12 Tu Bologna Winter School
13 Tu PB: 10-13 ; PB: 14-17	13 Th BPI: 9-13; LB1: 14-18	13 Su	13 We Bologna Winter School
14 We BPI: 9-13; LB1: 14-18	14 Fr PB: 10-13 ; AG: 14-17	14 Mo PB: 10-13 ; BDB: 14-18	14 Th Bologna Winter School
15 Th BPI: 9-13; LB1: 14-18	15 Sa	15 Tu BDB: 9-13 ; PB: 14-17	15 Fr Bologna Winter School
16 Fr PB: 10-13 ; AG: 14-17	16 Su	16 We BPI: 9-13; AG: 14-17	16 Sa
17 Sa	17 Mo BDB: 9-13 ; LB1: 14-18	17 Th BPI: 9-13; PB: 14-17	17 Su
18 Su	18 Tu PB: 10-13 ; PB: 14-17	18 Fr PB: 10-13 ; AG: 14-17	18 Mo
19 Mo BDB: 9-13 ; LB1: 14-18	19 We BPI: 9-13; AG: 14-17	19 Sa	19 Tu
20 Tu PB: 10-13 ; PB: 14-17	20 Th BPI: 9-13; LB1: 14-18	20 Su	20 We
21 We BPI: 9-13; AG: 14-17	21 Fr PB: 10-13 ; AG: 14-17	21 Mo PB: 10-13 ; BDB: 14-18	21 Th
22 Th BPI: 9-13; LB1: 14-18	22 Sa	22 Tu BDB: 9-13 ; PB: 14-17	22 Fr
23 Fr PB: 10-13 ; AG: 14-17	23 Su	23 We BPI: 9-13; PB: 14-17	23 Sa
24 Sa	24 Mo	24 Th BPI: 9-13; PB: 14-18	24 Su
25 Su	25 Tu Christmas Day	25 Fr BPI: 9-13 ; PB: 14-17	25 Mo
26 Mo BDB: 9-13 ; AG: 14-17	26 We	26 Sa	26 Tu
27 Tu PB: 10-13 ; PB: 14-17	27 Th	27 Su	27 We
28 We BPI: 9-13; PB: 14-17	28 Fr	28 Mo	28 Th
29 Th BPI: 9-13; AG: 14-17	29 Sa	29 Tu	
30 Fr LB1: 10-13 ; LB1: 14-18	30 Su	30 We	
	31 Mo	31 Th	

# Oral examinations

- Poster presentation (24-25 Jan 2019)
- Oral examination on February, data to be defined

# Readings/Bibliography:

- Slides available on UNIBO repository
- on line, selected papers and books
- Introduction to Proteomics: Principles and Applications (Methods of Biochemical Analysis). Nawin C. Mishra, Günter Blobel, 2010. Wiley
- LC-MS/MS in Proteomics: Methods and Applications (Methods in Molecular Biology). Pedro R. Cutillas, John F. Timms, 2010. Humana press
- Integrative Proteomics. Ed.: Tsz-Kwong Man and Ricardo J. Flores , ISBN 978-953-51-0070-6 [open access  
<http://www.intechopen.com/books/integrative-proteomics> ]

# PROTEOMICS

# Proteomics

## BIG DATA ISSUE

Major bottleneck is effectively handling and storing the enormous amounts of data that will be generated (the big data problem).

Size of 'Big Data' in the US health-care system is increasing at a rate of between 1.2 and 2.4 exabytes per year, and will rise faster as additional new data intensive technologies are adopted (e.g. microarrays, whole genome sequencing, and imaging).

**Countries with large populations and emerging economies, such as China and India, will be generating zettabyte ( $10^{21}$ ) to yottabyte ( $10^{24}$ ) amounts of health-related data each year.**

*Chen R, Mias GI, Li-Pook-Than J, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell. 2012;148:1293–1307*

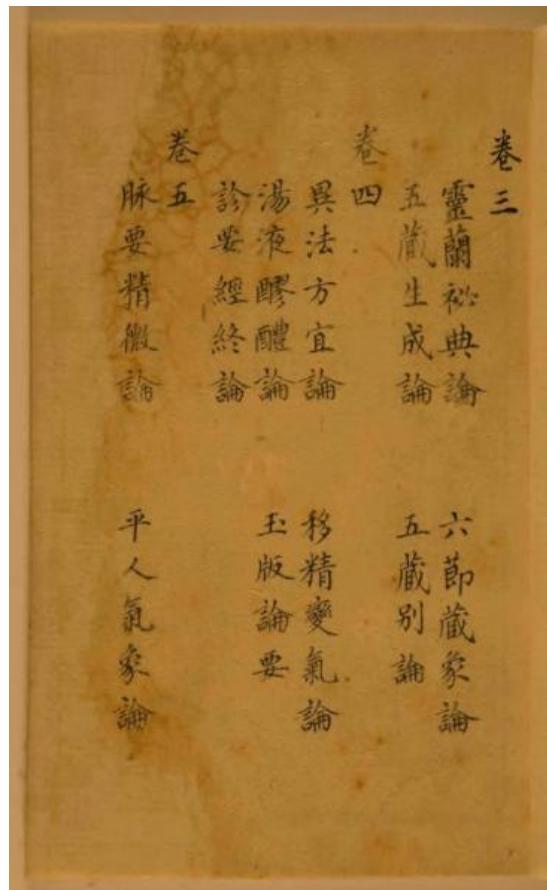
*Roda A, Michelini E, Caliceti C, Guardigli M, Mirasoli M, Simoni P. Advanced bioanalytics for precision medicine. Anal Bioanal Chem. 2018 Jan;410(3):669-677.*

## PERSONALIZED MEDICINE

The emergence of the 'omics' platforms (genomics, proteomics, metabolomics, transcriptomics, and interactomics) now gives us a pipeline around which to develop the infrastructure required for personalized precision, preventive, and participatory medicine.

**More than 25% of all new drugs approved by the US FDA in 2015 relating to personalized Medicine.**

# Personalized and precision medicine, a new concept?



First described in The Yellow Emperors Canon of Internal Medicine over 2000 years ago

The Systems Biology group in Seattle coined the term P4 medicine

**4P: Predictive, Preventive,  
Personalized and Participatory**

# General classification for proteomics

## Structural Proteomics

-Goal is to map the 3D structure of protein and protein complex. eg X-ray crystallography and NMR spectroscopy

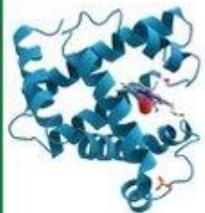
## Functional Proteomics

To study protein protein interaction, 3D structure cellular localization and PTMS in order to understand the physiological function of the whole set of proteome.

## Expression proteomics -

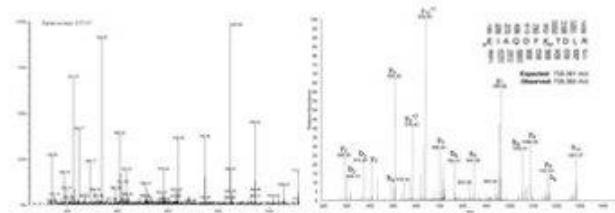
Quantitative study of protein expression between samples that differ by some variable

## Top-down



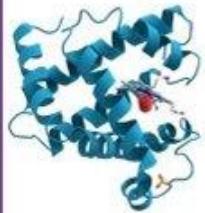
Protein(s)

(LC)  
MS  
MSMS



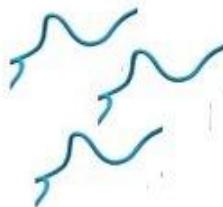
MS      MSMS

## Bottom-up



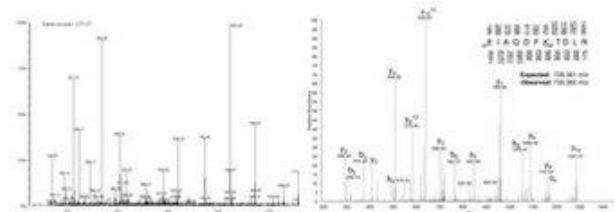
Protein(s)

Enzymatic  
digestion



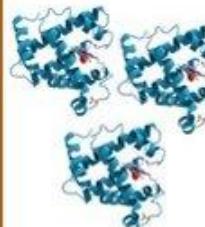
Peptides

(LC)  
MS  
MSMS



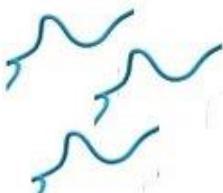
MS      MSMS

## Shotgun



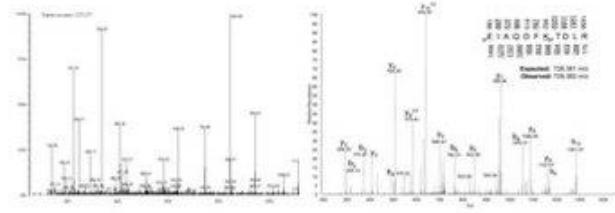
Protein  
mixture

Enzymatic  
digestion



Peptides

2D-LC/MS  
MS  
MSMS



MS      MSMS

# PROGRAM (I)

- **Introduction to proteomics:** structural and functional proteomics, strategies for proteomic analysis.
- **Separative techniques in proteomics:** gel electrophoresis (SDS PAGE, 2D PAGE), non-denaturing gel electrophoresis, capillary electrophoresis, HPLC (reverse phase, ion exchange, size exclusion), multidimensional HPLC.
- **Mass spectrometry techniques in proteomics:** soft ionization techniques, MALDI and ESI ionization sources, quadrupole mass analyzers, time-of-flight mass analyzers (TOF), hybrid analyzers (Q-TOF), mass spectrometers (MALDI/TOF and ESI/Q-TOF), SELDI/TOF analysis.
- **Protein identification:** analysis of intact proteins by ESI/MS and MALDI/MS, in-gel digestion, in-gel digest analysis by MALDI/MS and ESI/MS, database search for protein identification.
- **Quantitative proteomics:** label-free approaches and stable isotope labelling
- **Applications of proteomics:** proteomics in clinical analysis, “shotgun” proteomics, peptide fingerprint mapping, study of protein complexes, detection and identification of post-transcriptional protein modifications.

# PROGRAM (II)

- **BIG DATA analysis in proteomics:** recent advances in data integration-based methods to uncover personalized information from big data produced by proteomics and other omics studies
- **Interactomics:** Methods to study protein-protein interactions: co-immunoprecipitation, yeast-two hybrid system, FRET, BRET
- **Metabolomics:** the “end point of cellular regulation”
- **Ultrasensitive and fast methods based on molecular recognition** combined with selective and highly detectable probes: Immunoassay, hybridization reactions, biosensors
- **Luminescence Molecular Imaging:** localization and quantification of analytes in single cells and tissues (combined with microscopy), whole organ and living organisms using nanoparticles and recombinant bioluminescent cells”

# Current state of proteomics databases and repositories

A large proportion of the proteomics community was reticent to openly share the data they produced. However, the sharing of not only the knowledge obtained through proteomics experiments (through scientific publications) and the underlying data, has increasingly become standard practice, and is now even mandatory or strongly advised in many of the relevant scientific journals.

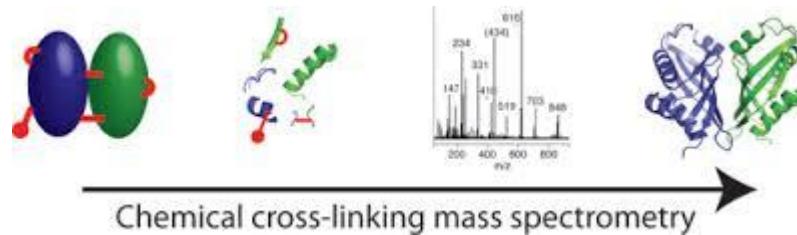
A number of funders (e.g. the Wellcome Trust, EC, NIH) enforce the public deposition of data from projects they fund as a way to maximize the value of the funds provided

- Global Proteome Machine Database (GPMDB <https://thegpm.org/> )
- PeptideAtlas <http://www.peptideatlas.org/#>
- <http://proteomecentral.proteomexchange.org/cgi/GetDataset>
- PRIDE database <https://www.ebi.ac.uk/pride>

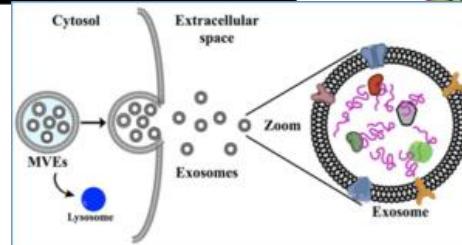
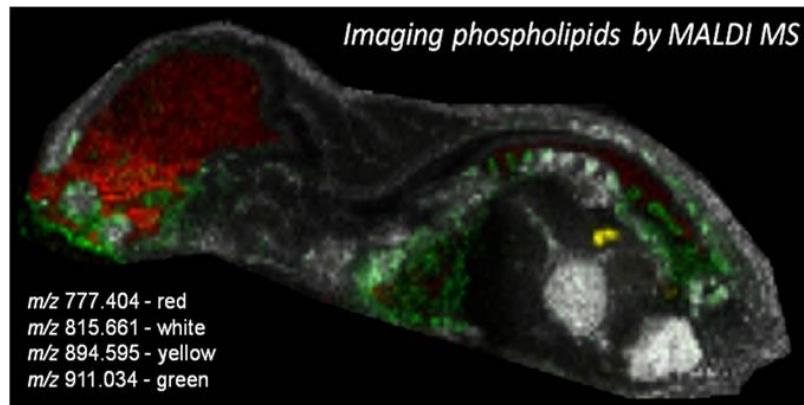
Others: ProteomicsDB, Mass Spectrometry Interactive Virtual Environment (MassIVE), Chorus, MaxQB, PeptideAtlas SRM Experiment Library (PASSEL), Human Proteinpedia

# A glance on advanced tools

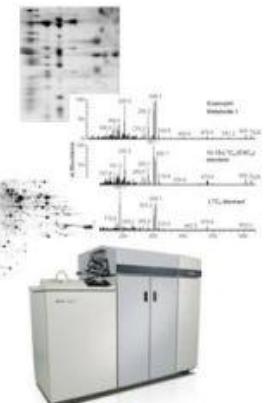
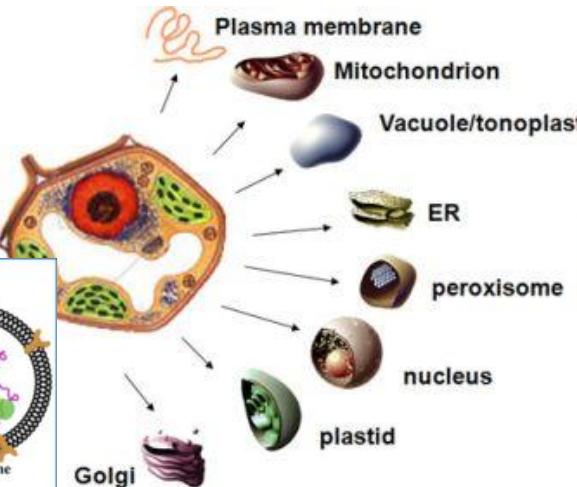
Chemical cross-linking mass spectrometry (CXMS) and bioinformatics: probing 3D structures and conformational changes



## Mass Spectrometry Imaging for Clinical Research



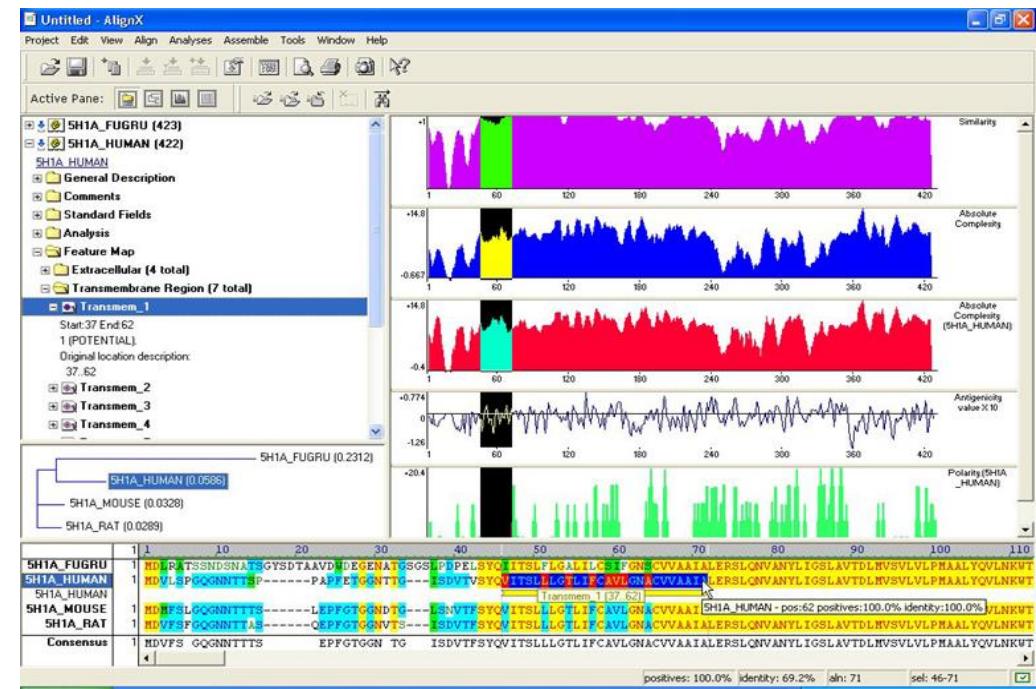
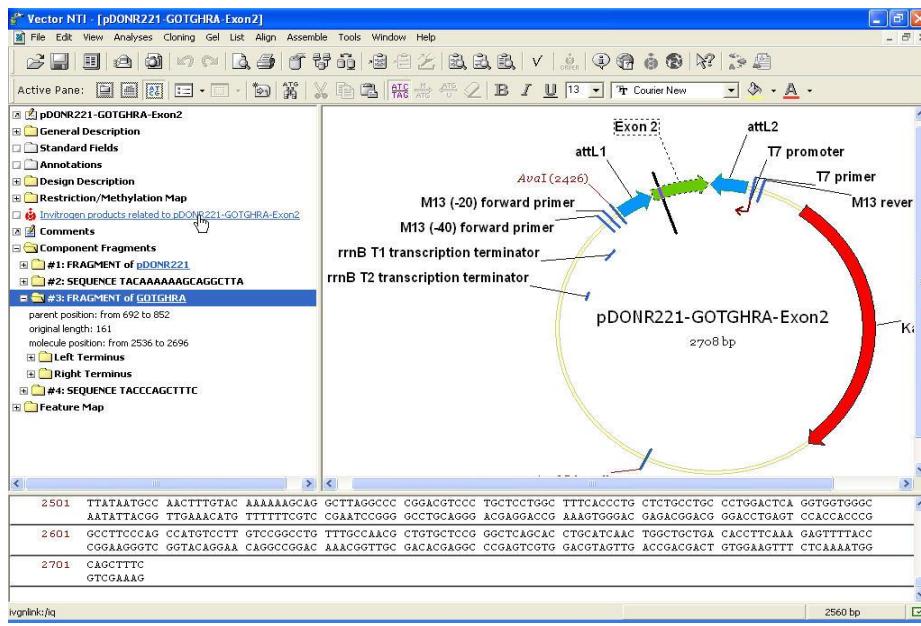
## Subcellular and extracellular/exosome proteomics



# Tutorial exercises and in silico labs

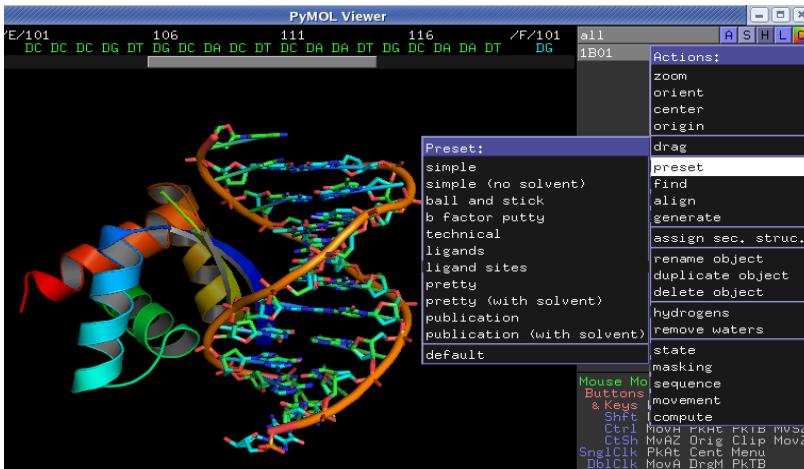
# In silico strategy for cloning, in vitro mutagenesis and recombinant protein expression:

# From PCR of cDNA to proteins alignment



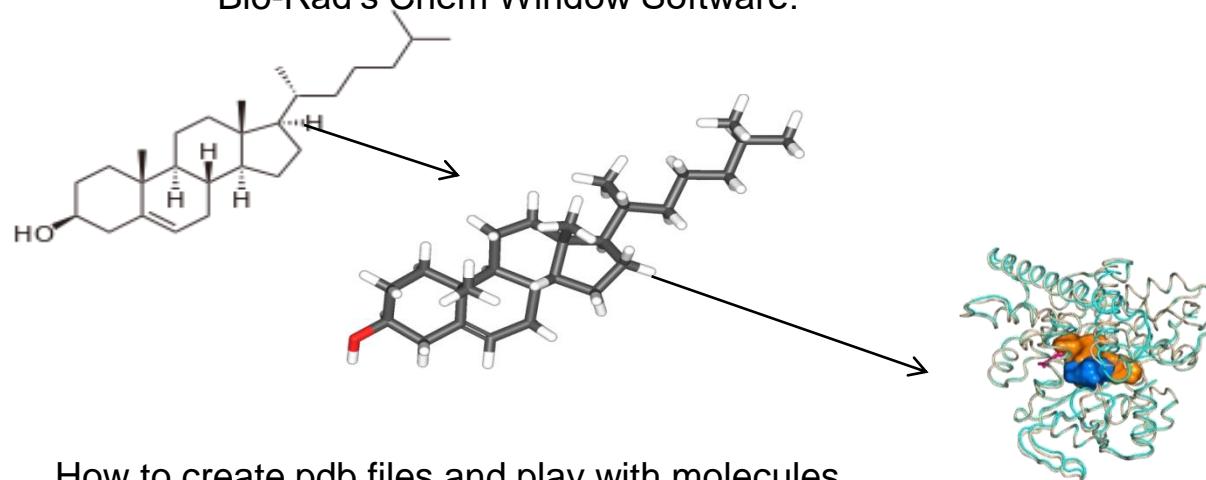
## Pymol

A molecular visualization system.



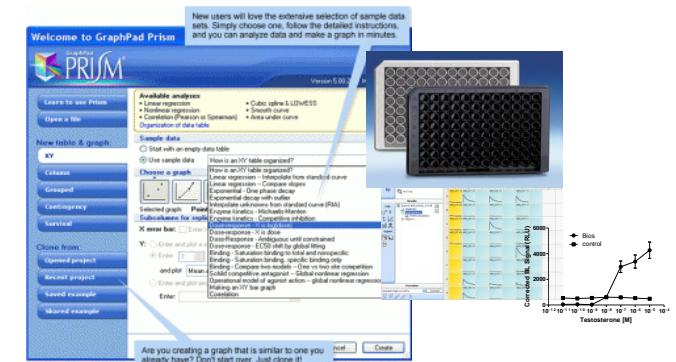
## DrawIt (KnowItAll-Academic Edition)

It is 2-D chemical structure drawing tool developed using Bio-Rad's Chem Window Software.



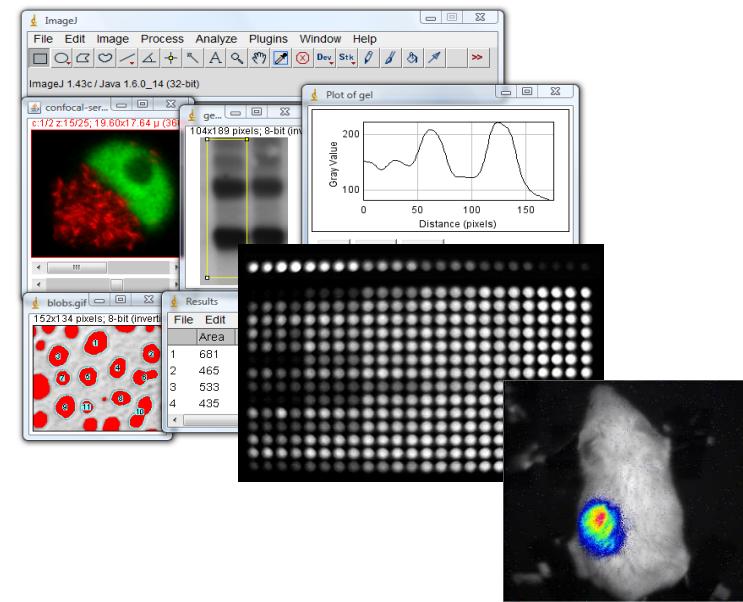
**GraphPad Prism**

It combines scientific graphing, comprehensive curve fitting (nonlinear regression), understandable statistics, and data organization.

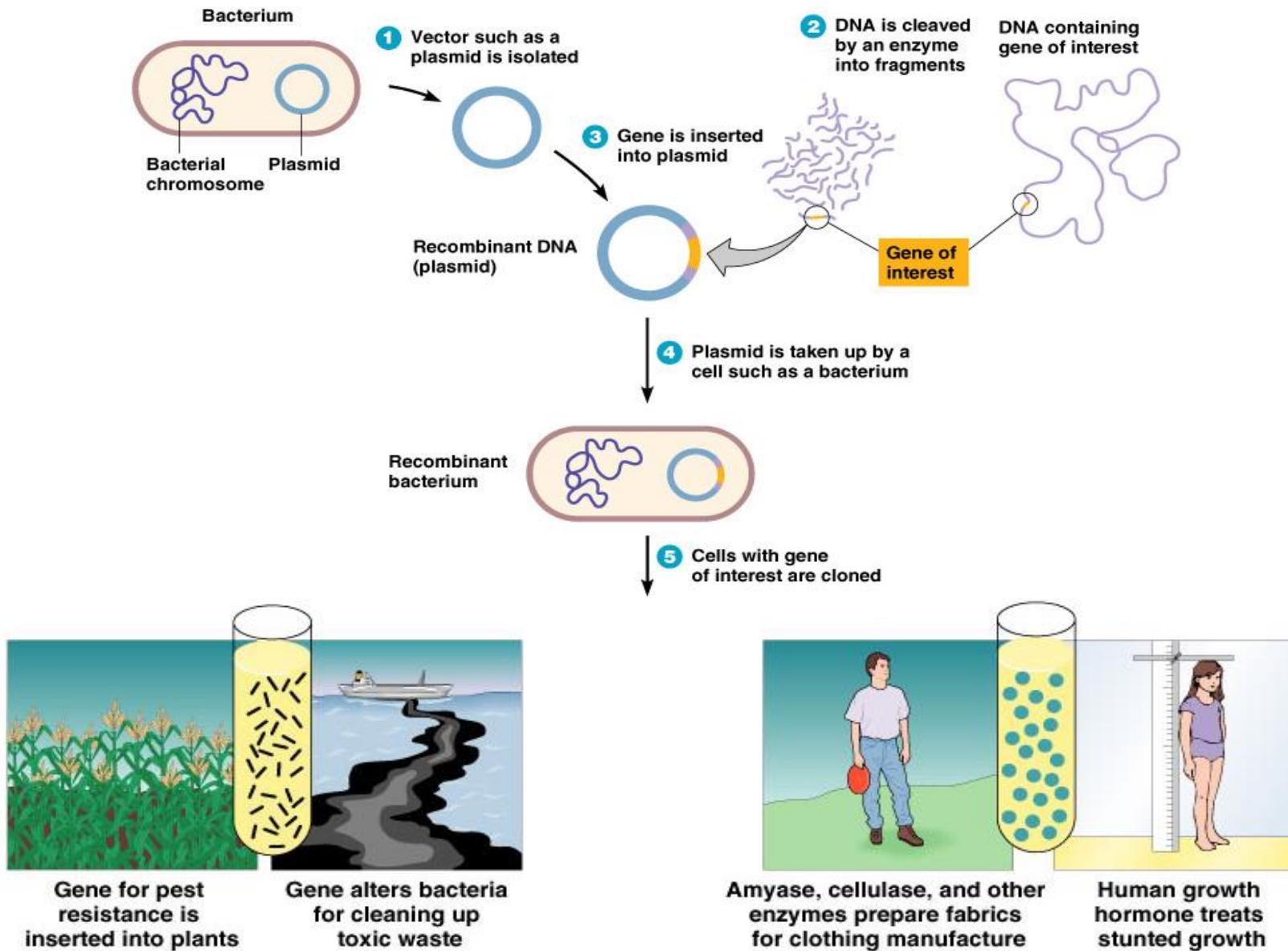


## ImageJ

It is an open source image processing program designed for scientific multidimensional images.



# Production of recombinant proteins at a glance



# *What is proteomics*

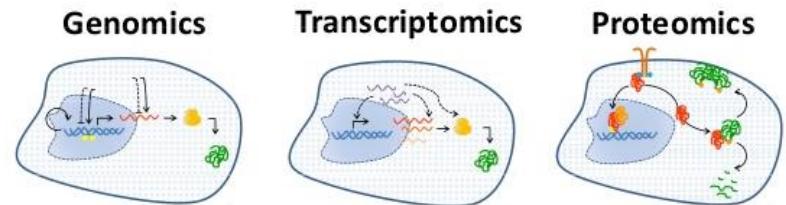
Proteomics is the large-scale study of proteins, particularly their structures and functions.

**Proteome** = Proteins encoded by the genome

This term was coined in analogy with genomics, and while it is often viewed as the "next step", proteomics is much more complicated than genomics.

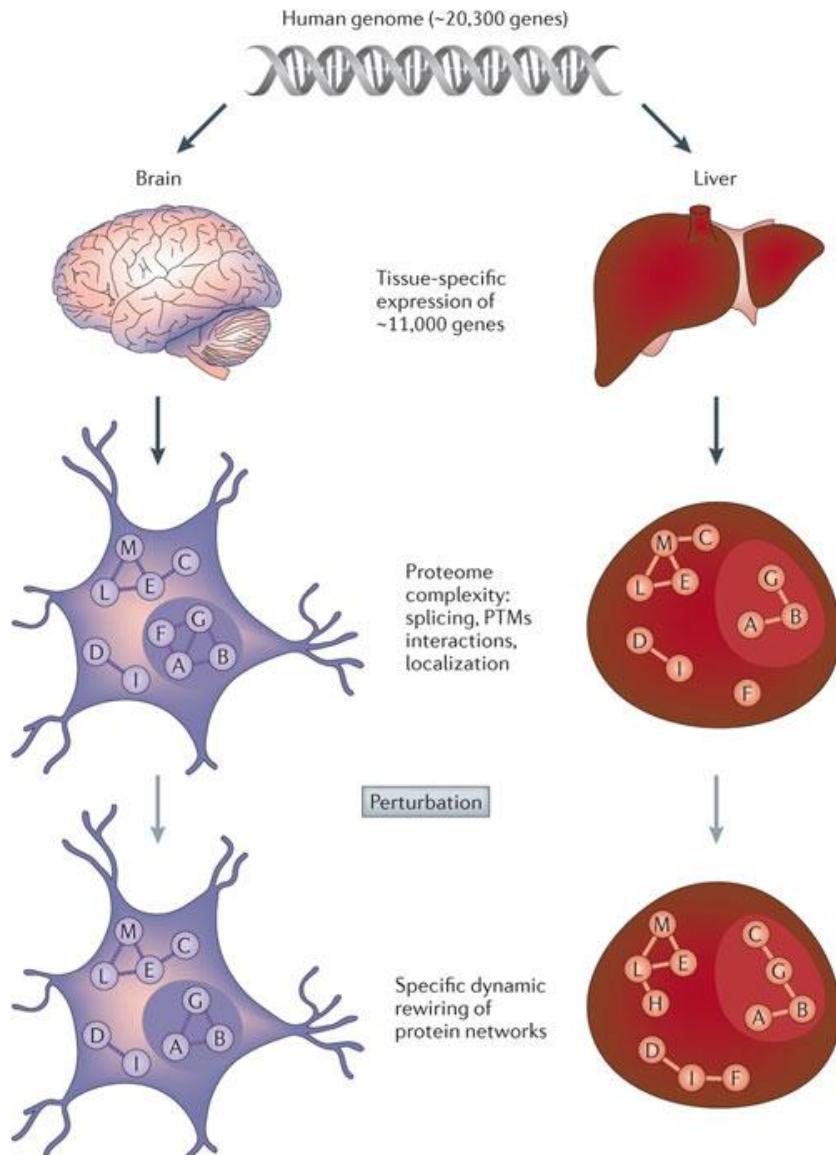
Most importantly, while the genome is a rather constant entity, the proteome differs from cell to cell and is constantly changing through its biochemical interactions with the genome and the environment.

One organism will have radically different protein expression in different parts of its body, in different stages of its life cycle and in different environmental conditions.



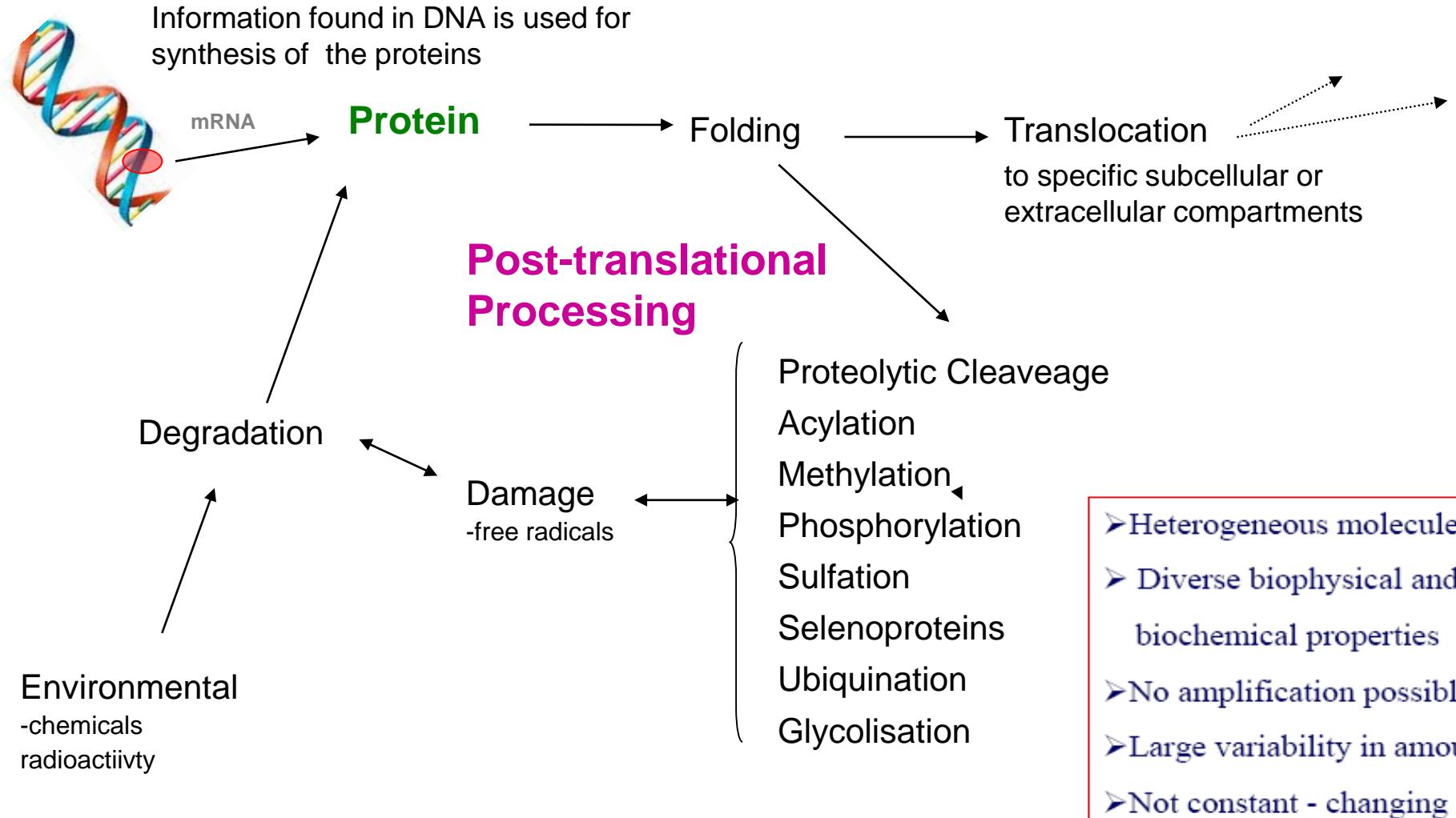
Location, Time, Interaction partners, PTMs, Dynamics, Turnover

# From genome to proteome

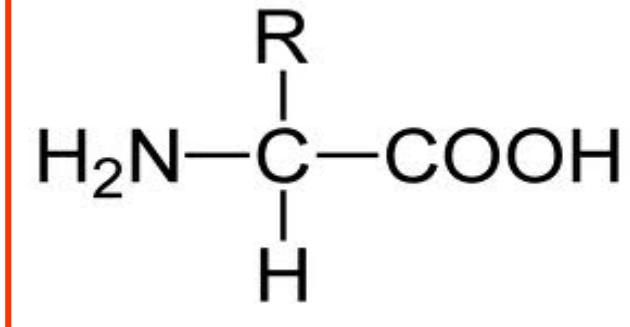


The human genome contains approximately 20,300 genes. The specific expression of a subset of the genome (~11,000 genes) determines the molecular backbone of the cellular phenotype (that is, the tissue cell types). A much higher order of complexity is achieved by the intricate mechanisms of protein regulation, including splicing variants, post-translational modifications (PTMs), protein–protein interactions (PPIs) and subcellular localization. This creates time-dependent tissue- and organelle-specific protein networks that respond differently to perturbations (for example, ageing or drug treatment).

# Life cycle of a protein



# Amino Acids



- Amino acids are building blocks for proteins
- They have a central  $\alpha$ -carbon and  $\alpha$ -amino and  $\alpha$ -carboxyl groups
- 20 different amino acids
- Same core structure, but different side group (R)
- The  $\alpha$ -C is chiral (except glycine); proteins contain only L-isoforms.
- Amino acids are amphotytes,  $\text{pK}_a$  of  $\alpha$ -COOH is  $\sim 2$  and of  $\alpha$ -NH<sub>2</sub> is  $\sim 9$
- At physiological pH most aa occur as zwitterions.

$\text{pK}_a$ : the negative logarithm of the ionization constant of an acid, a measure of the strength of an acid. The lower the  $\text{pK}_a$ , the stronger the acid.

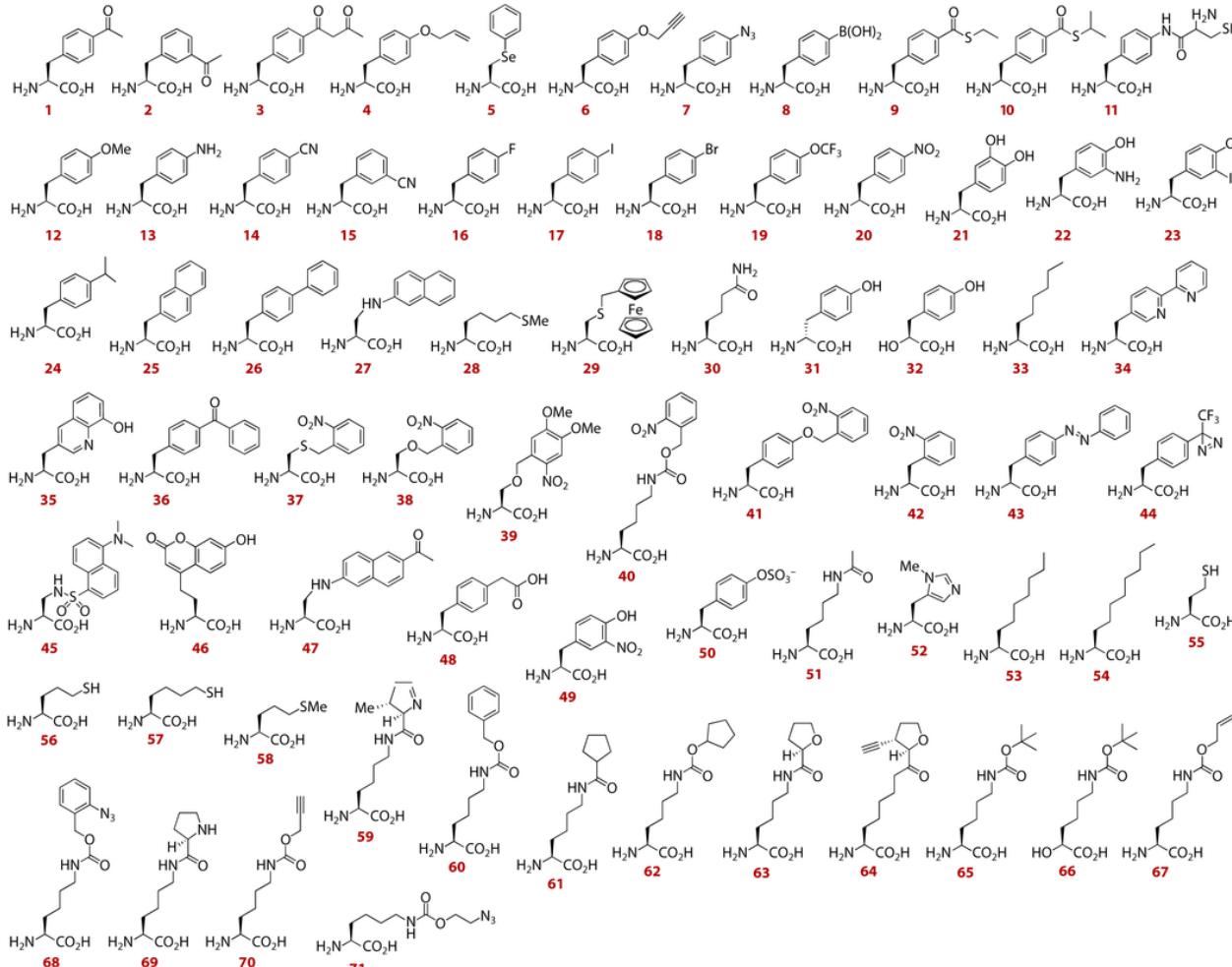
# Classification of Amino Acids (Based on R-group)

- Aliphatic: gly (G), ala (A) , val (V), leu (L), ile (I)
- Aromatic: Trp (W), Phe (F), Tyr (Y), His (H),
- Sulphur : Met (M), Cys (C)
- Hydroxyl: Ser (S), Thr (T), Tyr (Y)
- Cyclic: pro (P)
- Carboxyl: asp (D), glu (E)
- Amine: lys (K), arg (R)
- Amide: asn (N), gln (Q)

# Classification of Amino Acids (based on polarity)

- **Hydrophobic / non-polar R group:** glycine, alanine, valine, leucine, isoleucine, methionine, proline, phenylalanine, tryptophan
- **Polar R group (net charge 0 at pH 7.4):** Serine, threonine, cysteine, tyrosine, asparagine, glutamine, histidine
- **Polar R group (charged ion at pH 7.4):** aspartate, glutamate, lysine, arginine

# Unnatural aminoacids



# Proteins

- Linear polymers of aa via amide linkages form
  - peptides (1-10)
  - polypeptides (11-100)
  - proteins (>100)
- Eg: glutathione (GSH), vasopressin (9), insulin (51)
- Proteins have an amino-end and carboxyl-end
- In the lab, proteins can be hydrolyzed (to aa) by strong acid treatment
- Physiologic hydrolysis by peptidases and proteases

# Protein Structure: 4 levels

Protein structure is organized hierarchically from so-called *primary structure* to *quaternary structure*. Higher-level structures are *motifs* and *domains*.

- **Primary** structure: aa sequence
- **Secondary** structure: local regularly occurring structure in proteins and is mainly formed through hydrogen bonds between backbone atoms
- **Tertiary** structure: 3D complex folding (describes the packing of alpha-helices, beta-sheets and random coils with respect to each other on the level of one whole polypeptide chain)
- **Quaternary** structure: describes the spatial organization of the chains (if there is more than one polypeptide chain present in a complex protein. Then quaternary structure)

# Molecular Structures

**Primary structure** a chain of amino acids

**Secondary structure** three dimensional form, formally defined by the hydrogen bonds of the polymer

Amino acids vary in their ability to form the various secondary structure elements.

$\alpha$ -helices



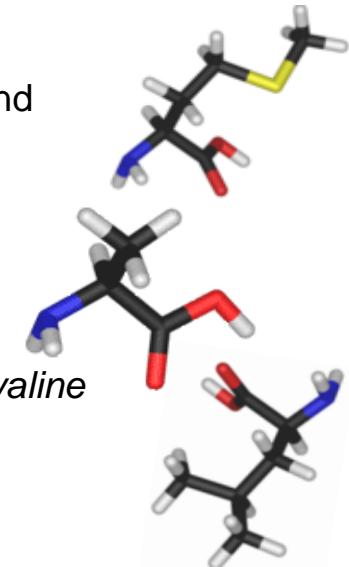
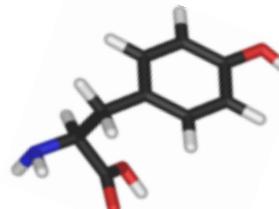
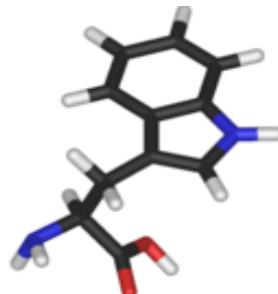
Amino acids that prefer to adopt helical conformations in proteins include *methionine*, *alanine*, *leucine*, *glutamate* and *lysine* ("MALEK" in amino acid 1-letter codes)

$\beta$ -sheets



The large aromatic residues (*tryptophan*, *tyrosine* and *phenylalanine*) and C $\beta$ -branched amino acids (*isoleucine*, *valine* and *threonine*) prefer to adopt  $\beta$ -strand conformations.

Confer similar properties or functions when they occur in a variety of proteins



# Sequence alignment

Sequence alignment is a way of arranging primary sequences (of DNA, RNA, or proteins) in such a way as to align areas sharing common properties.

CLUSTAL W (1.83) multiple sequence alignment

The degree of relatedness, similarity between the sequences is predicted computationally or statistically

Software tools used for general sequences alignment tasks are  
**ClustalW**

Vector NTI

NCBI BLAST

Latest news: 12 Dec 2006 : New search options

Protein BLAST: search protein databases using a protein query - Windows Internet Explorer

File Modifica Visualizza Preferiti Strumenti ?

Preferiti Siti suggeriti Scarica altri add-on

Isoelectric focusing and two... Protein BLAST: search pr... X

Pagina Sicurezza Strumenti

## Basic Local Alignment Search Tool

BLAST Basic Local Alignment Search Tool My NCBI [Sign In] [Register]

Home Recent Results Saved Strategies Help

► NCBI/ BLAST/ blastp suite

blastn blastp blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#)

Reset page Bookmark

Enter Query Sequence

Enter accession number, gi, or FASTA sequence [?](#) Clear

Query subrange [?](#)

From  To

Or, upload file  [Sfoglia...](#) [?](#)

Job Title   
Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Database Non-redundant protein sequences (nr) [?](#)

Organism Optional Enter organism name or id--completions will be suggested   Exclude +  
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude Optional  Models (XM/XP)  Environmental sample sequences

Entrez Query

[Disclaimer](#)

**Standard protein BLAST** is designed for protein searches

Standard protein-protein BLAST (blastp) is used for both identifying a query amino acid sequence and for finding similar sequences in protein databases. Like other BLAST programs, blastp is designed to find local regions of similarity. When sequence similarity spans the whole sequence, blastp will also report a global alignment, which is the preferred result for protein identification purposes.

**PSI-BLAST** is designed for more sensitive protein-protein similarity searches.

**Position-Specific Iterated (PSI)-BLAST** is the most sensitive BLAST program, making it useful for finding very distantly related proteins or new members of a protein family. Use PSI-BLAST when your standard protein-protein BLAST search either failed to find significant hits, or returned hits with descriptions such as "hypothetical protein" or "similar to...".

**PHI-BLAST** can do a restricted protein pattern search.

Pattern-Hit Initiated (PHI)-BLAST is designed to search for proteins that contain a pattern specified by the user AND are similar to the query sequence in the vicinity of the pattern. This dual requirement is intended to reduce the number of database hits that contain the pattern, but are likely to have no true homology to the query.

# FASTA

- A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line (defline) is distinguished from the sequence data by a greater-than (">") symbol at the beginning. It is recommended that all lines of text be shorter than 80 characters in length. An example sequence in FASTA format is:

```
>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLLTTLVLVNAIYFKGMWKTAFAEDTREMPFHVTKQESKPVQMMCMNNSFNVATLPAE
KKKILELPLFASGDLSMLVLLPDEVSDLERIEKTINFELTEWTNPNTMEKRRVKVYLPQMKEEKYNLTS
VLMALGMTDLFIPSANLTGIISSAESLKISQAVHGAFMELSEDGIEMAGSTGVIEDIKHSPSEQFRADHP
FLFLIKHNPTNTIVYFGRYWSP
```

- Blank lines are not allowed in the middle of FASTA input.
- Sequences are expected to be represented in the standard IUB/IUPAC amino acid and nucleic acid codes, with these exceptions: lower-case letters are accepted and are mapped into upper-case; a single hyphen or dash can be used to represent a gap of indeterminate length; and in amino acid sequences, U and \* are acceptable letters (see below). Before submitting a request, any numerical digits in the query sequence should either be removed or replaced by appropriate letter codes (e.g., **N for unknown nucleic acid residue or X for unknown amino acid residue**).

# Bare Sequence

This may be just lines of sequence data, without the FASTA definition line, e.g.:

```
QIKDLLVSSSTDLDTTLVLVNNAIYFKGMWKTA  
FNAEDTREMPFHVTQESKPVQMMCMNNNSFN  
VATLPAEKMKILELPFASGDLSMLVLLP  
DEVSDLERIEKTINFEKLTEWTNPNTME  
KRRVKVYLPQMKIEEKYNLTS  
VLMALGMDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGVIEDIKHSPESEQFRADHP  
FLFLIKHNPTNTIVYFGRYWSP
```

It can also be sequence interspersed with numbers and/or spaces, such as the sequence portion of a GenBank/GenPept flatfile report:

```
1 qikdllvsss tdldttlv lv naiyfkgmwk tafnaedtre mpfhvtqes kp vqmmcmnn  
61 sfnvatlpae kmkilelpfa sgdlsm l vll p devsdler i ektinfeklt ewtnpntmek  
121 rrrvkvylpqm kieekynlts vlmalgmtl fipsanltgi ssaeslkisq avhgafmels  
181 edgiemagst gviedikhsp eseqfradhp flflikhnpt ntivyfgryw sp
```

# Identifiers

Normally these are simply accession, accession.version or gi's (e.g., p01013, AAA68881.1, 129295), but a bar-separated NCBI sequence identifier (e.g., gi|129295) will also be accepted.

These NCBI sequence identifiers have a very specific syntax as described in <ftp.ncbi.nlm.nih.gov/blast/documents/blastdb.html>

14

15

16

17

18

19

20

21

22



HOME SEARCH SITE MAP

Entrez

Structure

Protein

CDD

PubMed

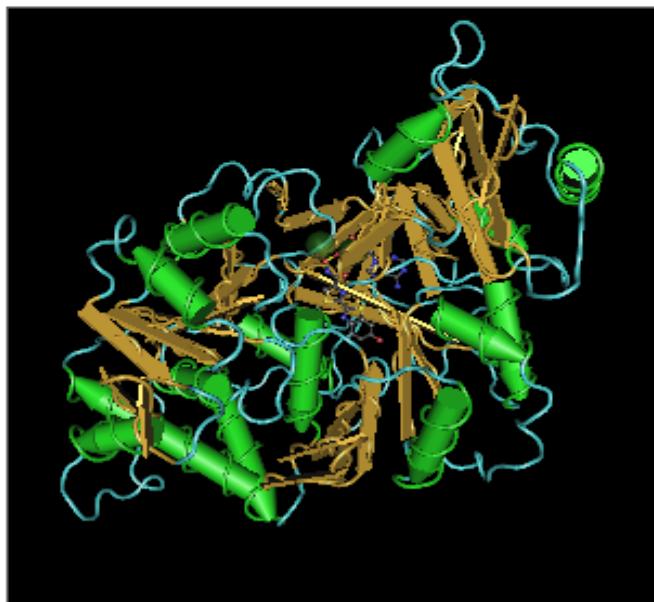
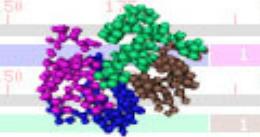
Taxonomy

PubChem

Help

Cn3D

# Structure Summary MMDB



MMDB ID: 38141

PDB ID: 2D1S

Search

PDB or MMDB ID

**Reference:** Nakatsu T, Ichiyama S, Hiratake J, Saldanha A, Kobashi N, Sakata K, Kato H *Structural basis for the spectral difference in luciferase bioluminescence* Nature v440, p.372-376

Fireflies communicate with each other by emitting yellow-green to yellow-orange brilliant light. The bioluminescence reaction, which uses luciferin, Mg-ATP and molecular oxygen to yield an electronically excited oxyluciferin species, is carried out by the enzyme luciferase. Visible light is emitted during relaxation of excited oxyluciferin to its ground state....

» View full abstract

**Description:** Crystal Structure Of The Thermostable Japanese Firefly Luciferase Complexed With High-Energy Intermediate Analogue.

**Deposition:** 2005/8/31 +

**Taxonomy:** *Luciola cruciata*

**Related Structure:** VAST

Structure View in Cn3D

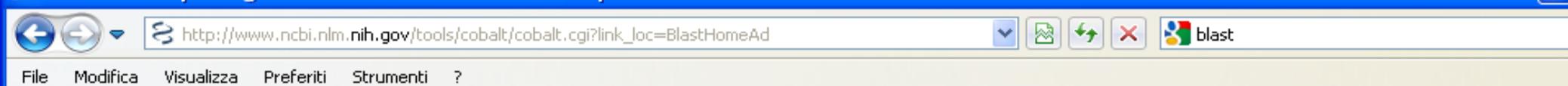
Structure View in RasMol

Tasks: Display Drawing: All Atoms

REFERENCE 2 (residues 1 to 548)

Recent activity

# COBALT:Multiple Alignment Tool - Windows Internet Explorer



File Modifica Visualizza Preferiti Strumenti ?

Preferiti Siti suggeriti Scarica altri add-on



## Constraint-based Multiple Alignment Tool

My NCBI  
[Sign In] [Re]

[Home](#) [Recent Results](#)

COBALT computes a multiple protein sequence alignment using conserved domain and local sequence similarity information. [more...](#)

[Reset page](#)

### Enter Query Sequences

Enter at least 2 protein accessions, gis, or FASTA sequences [?](#) [Clear](#)

Or, upload FASTA file

 [Sfoglia...](#)

Job Title

**Align**

Show results in a new window

► [Advanced parameters](#)

# Molecular Structures / Functional Families

**Tertiary structure** the overall shape of the protein (fold)

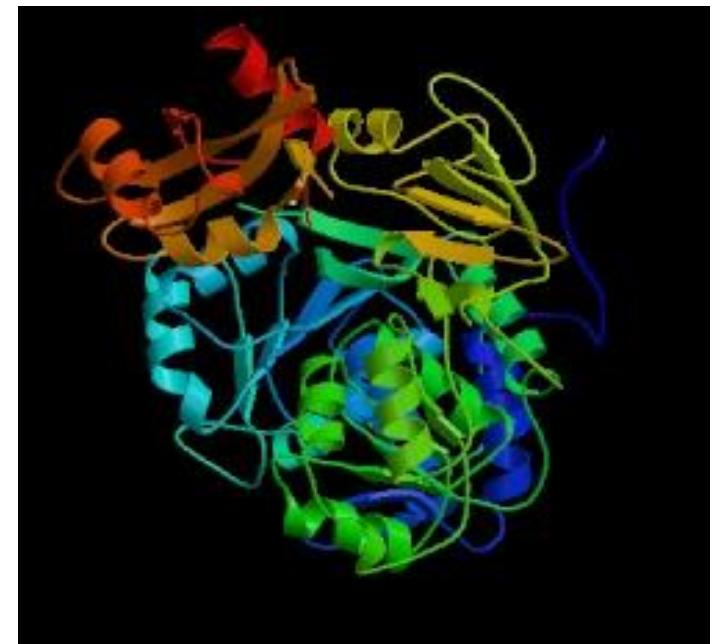
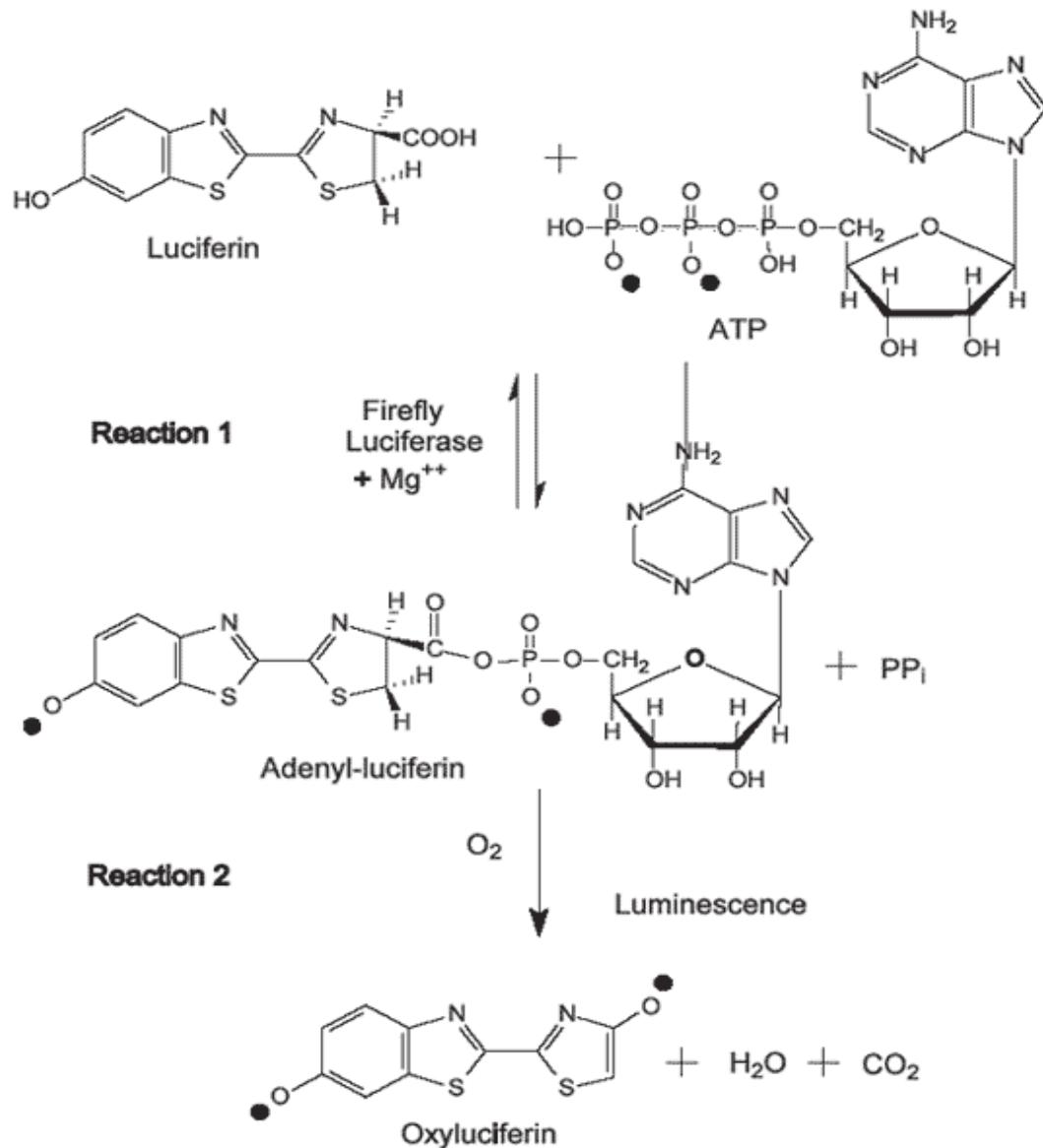
The **three-dimensional shape of the proteins** might be critical to their function. For example, specific binding sites for substrates on enzymes

Specific sequences that also confer unique properties and functions, **motifs** or **domains**

**Quaternary structure** -formation usually involves the "assembly" or "coassembly" of subunits that have already folded

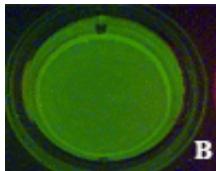
Incorrectly folded proteins are responsible for illnesses such as *Creutzfeldt-Jakob disease* and *Bovine spongiform encephalopathy (mad cow disease)*, and amyloid related illnesses such as *Alzheimer's*.

# E.g., luciferase



*3D-structure of firefly luciferase*

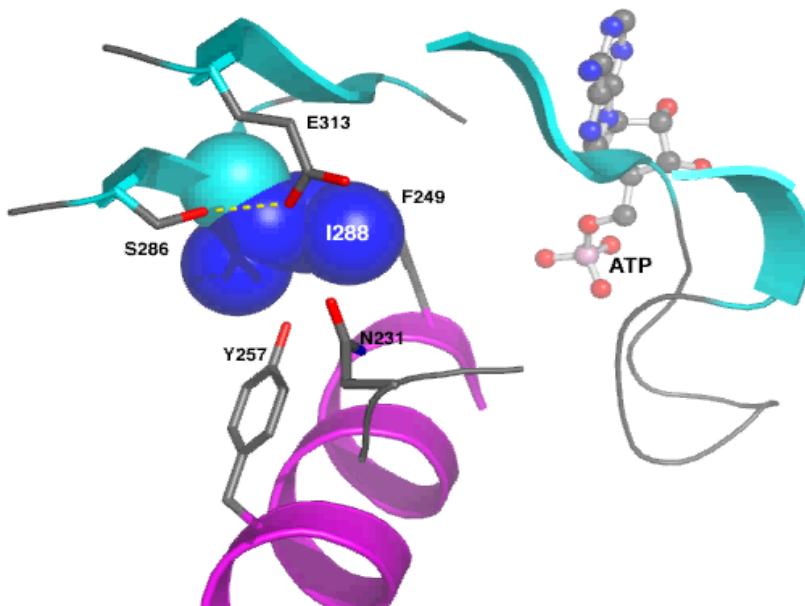
# Site-directed mutagenesis



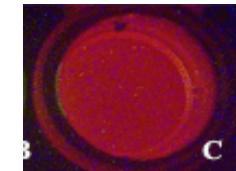
*L. cruciata* luciferase  
wt

Wild

(1) Before Reaction (ATP complex)

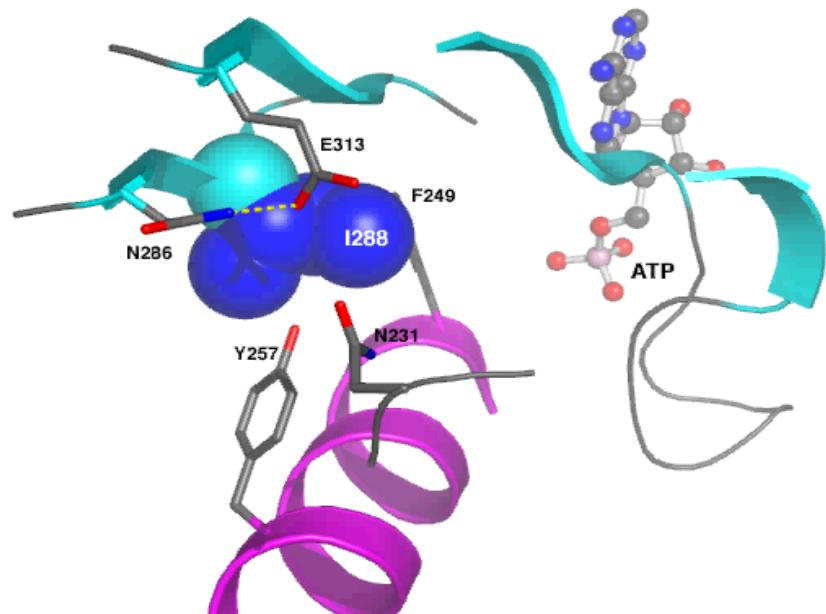


**RED MUTANT**  
**(single mutation)**



S286N

(1) Before Reaction (Calculated with Wild ATP complex)



A conformational change in Ile 288 occurs in WT luc (in complex with AMP and LH2 ) but not in the red mutant, shifting the emission wavelength.

The degree of molecular rigidity of the excited state



Colour of BL emission

(Nakatsu, Nature 2006)

# Functional families

Proteins can be grouped into functional families;  
proteins that carry out related functions



Structural



Signaling pathways



Metabolic



Transportation

Domains are clustered into families in which significant sequence similarity is detected as well as conservation of biochemical activity.

SCOP-a structural classification of proteins

---

By associating a novel protein with a protein family, one can predict the function of the novel protein

## Protein family classification databases:

**PROSITE**. Database of protein families and domain, defined by patterns and profiles, at ExPASY.

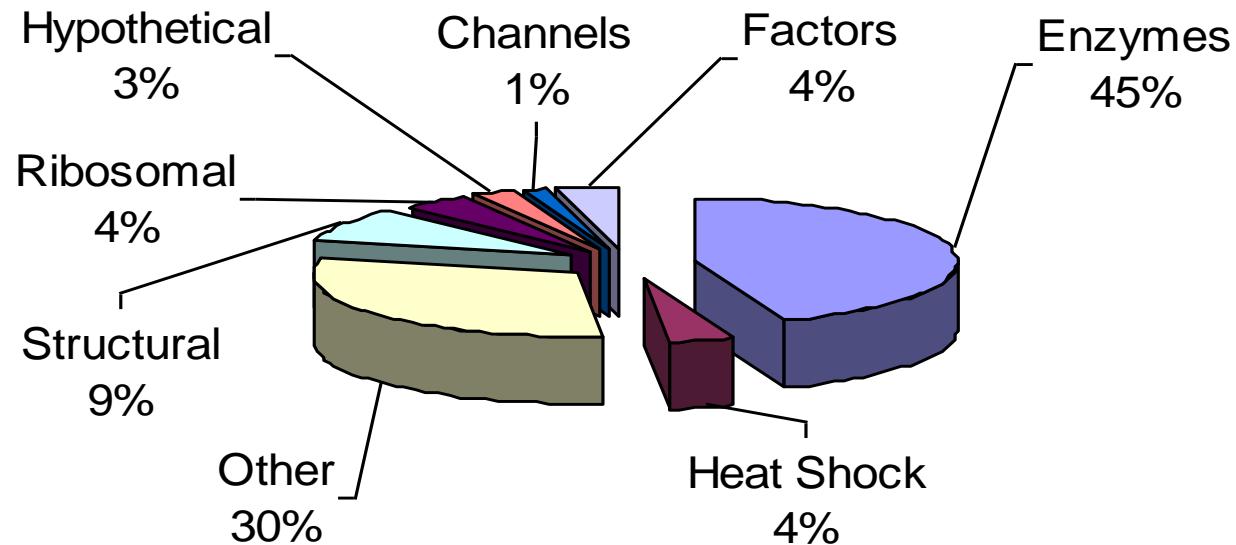
<http://au.expasy.org/prosite/>

**Pfam**. Pfam 27.0 (March 2013, 14831 families) is a large collection of protein , at Sanger Institute.

<http://pfam.xfam.org/>

**SMART** Simple Modular Architecture Research Tool, at EMBL. <http://smart.embl-heidelberg.de/>

# Protein function





# Database of protein domains, families and functional sites

ExPASy Home page Site Map Search ExPASy Contact us Swiss-Prot ENZYME

Search PROSITE for Go Clear

proSite Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More details / References / Disclaimer / Commercial users]. PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More details].

Release 20.3, of 09-Jan-2007 (1461 documentation entries, 1327 patterns, 696 profiles and 722 ProRule)

PROSITE access

e.g: PDOC00022, PS50089, SH3, zinc

add wildcard \*\*

Browse:

- by documentation entry
- by ProRule description \*\*\*
- by taxonomic scope \*\*\*
- by number of positive hit \*\*\*

SRS - Sequence Retrieval System

PROSITE tools

Scan a sequence against PROSITE - quick scan

Enter your sequence or a UniProtKB (Swiss-Prot or TrEMBL) ID or AC [ help ]:

Output includes graphical view and feature detection.

Scan Clear

exclude patterns with a high probability of occurrence

- Enzyme catalytic sites.
- Prosthetic group attachment sites (heme, pyridoxal-phosphate, biotin, etc).
- Amino acids involved in binding a metal ion.
- Cysteines involved in disulfide bonds.
- Regions involved in binding a molecule (ADP/ATP, GDP/GTP, calcium, DNA, etc.)



Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95, 5857-5864

Letunic et al. (2006) Nucleic Acids Res 34, D257-D260

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

#### SMART MODE:

NORMAL  
GENOMIC

Simple  
Modular  
Architecture  
Research  
Tool



### Sequence analysis

You may use either a [Uniprot](#)/[Ensembl](#) sequence identifier (ID) / accession number (ACC) or the protein sequence itself to request the SMART service.

#### Sequence ID or ACC

#### Sequence

HMMER searches of the SMART database occur by default. You may also find:

- [Outlier homologues](#) and homologues of known structure
- [PFAM domains](#)
- [signal peptides](#)
- [internal repeats](#)
- [intrinsic protein disorder](#)

If you have multiple sequences to analyze, try [batch access](#) to SMART.

### Architecture analysis

You can search for proteins with combinations of **specific domains** in different species or taxonomic ranges. You can input the domains directly into "Domain selection" box, or use "GO terms query" to get a list of domains. See [What's New](#) for more info.

#### Domain selection

Example: **TyrKc AND SH3 AND NOT SH2**

#### GO terms query

Example: **membrane AND signal**

[transduction](#)

#### Taxonomic selection

Select a taxonomic range via the selection box or type it into the text box below:

Examples: **Dictyostelium**

[discoideum](#), [Porifera](#)

You can try an [Advanced Query](#) if you're familiar with SQL.

### Alert SMART

If you want to be automatically informed each time a new protein with a defined domain composition is deposited in the database, please use '[Alert SMART](#)' (this facility is also available following an architecture

In this manner, threading tries to **predict the three-dimensional structure** starting from a given protein sequence. It is sometimes successful when comparisons based on sequences or sequence profiles alone fail to a too low similarity.

Human Genome Project

46 chromosomes

3 billion bp

~ 21,000-25,000 *protein-encoding genes*

Transcriptome: 40-100 000 mRNAs

Proteome: 100-400 000 proteins

Protein interactions:  $> 10^6$  interactions

Protein modifications:  $> 10^8$  post-translational modifications

One cell can contain between 1 and more than 100,000 copies of a single protein

The advances made in molecular biology over the past few decades greatly improved the study of proteins

- Recombinant molecular biological techniques allow:

- ✓ the **expression of proteins** or their domains in large quantities, necessary for biochemical and structural studies
- ✓ to **change primary protein sequences** by site-directed and random mutagenesis and to analyse the function of generated mutants *in vivo*
- ✓ to **investigate sub-cellular localisation, post-translational modifications and binding partners**
- ✓ to **study the function** of a protein of interest in cells, organs or organisms (knock in or knock out studies)

# Tools of Proteomics

- 📍 Protein separation technology

- Simplify complex protein mixtures

- Target specific proteins for analysis

- 📍 Mass spectrometry (MS)

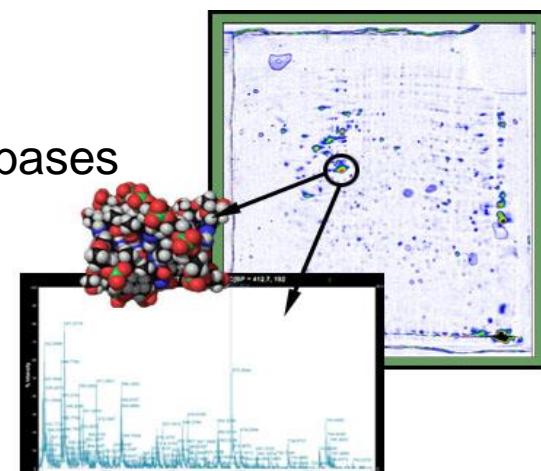
- Provide accurate molecular mass measurements of intact proteins and peptides

- 📍 Database

- Protein, EST, and complete genome sequence databases

- 📍 Software collection

- Match the MS data with specific protein sequences in databases

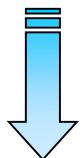


# Functional vs. structural proteomics

- **Functional Proteomics** : Study of change in protein expression within the proteome;
- **Structural Proteomics** : Study of the primary, secondary, and tertiary structure of the proteins in a proteome, functional predictions from primary structure; hightthroughput crystallography

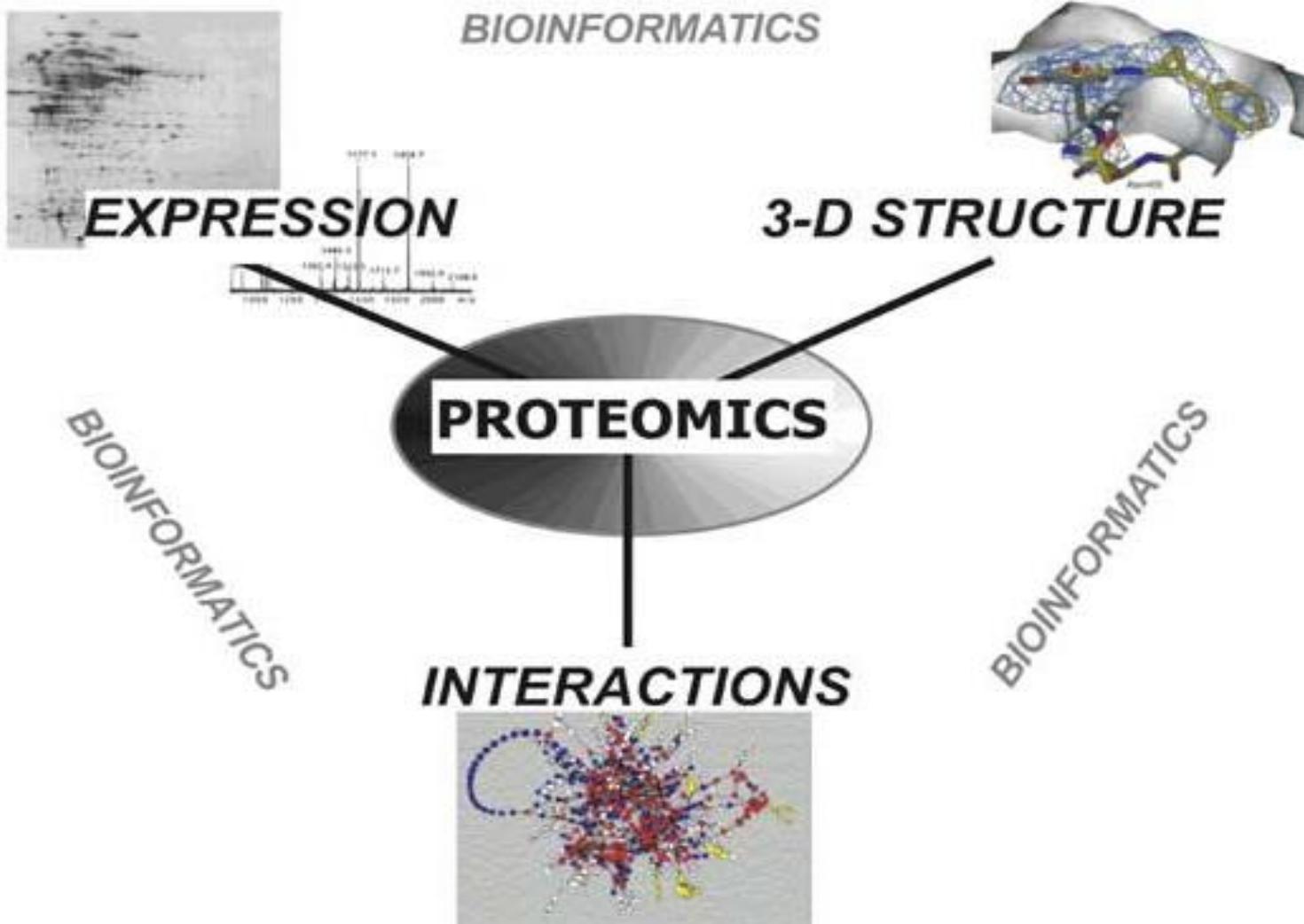
# Proteomics aims and applications

- Protein sample identification/confirmation
- Protein sample purity determination
- Detection of post-translational modifications
- Monitoring protein-ligand complexes/structure
- De novo peptide sequencing
- Biomarker discovery by detection of differences in protein expression among different classes of sample
- Identification of new target for drugs

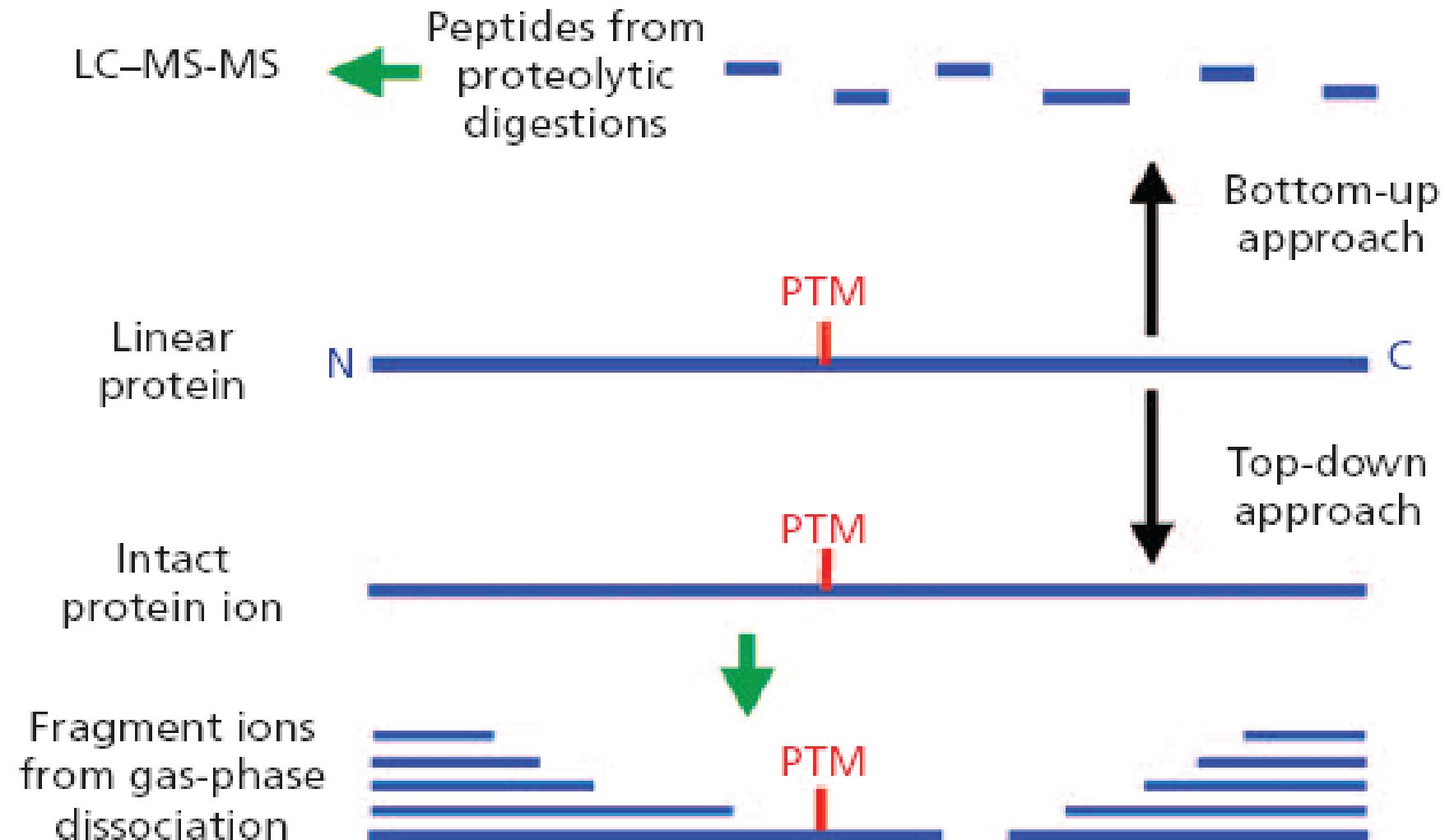


**Need for biochemical, bioanalytical, biomolecular and bioinformatic knowledges**

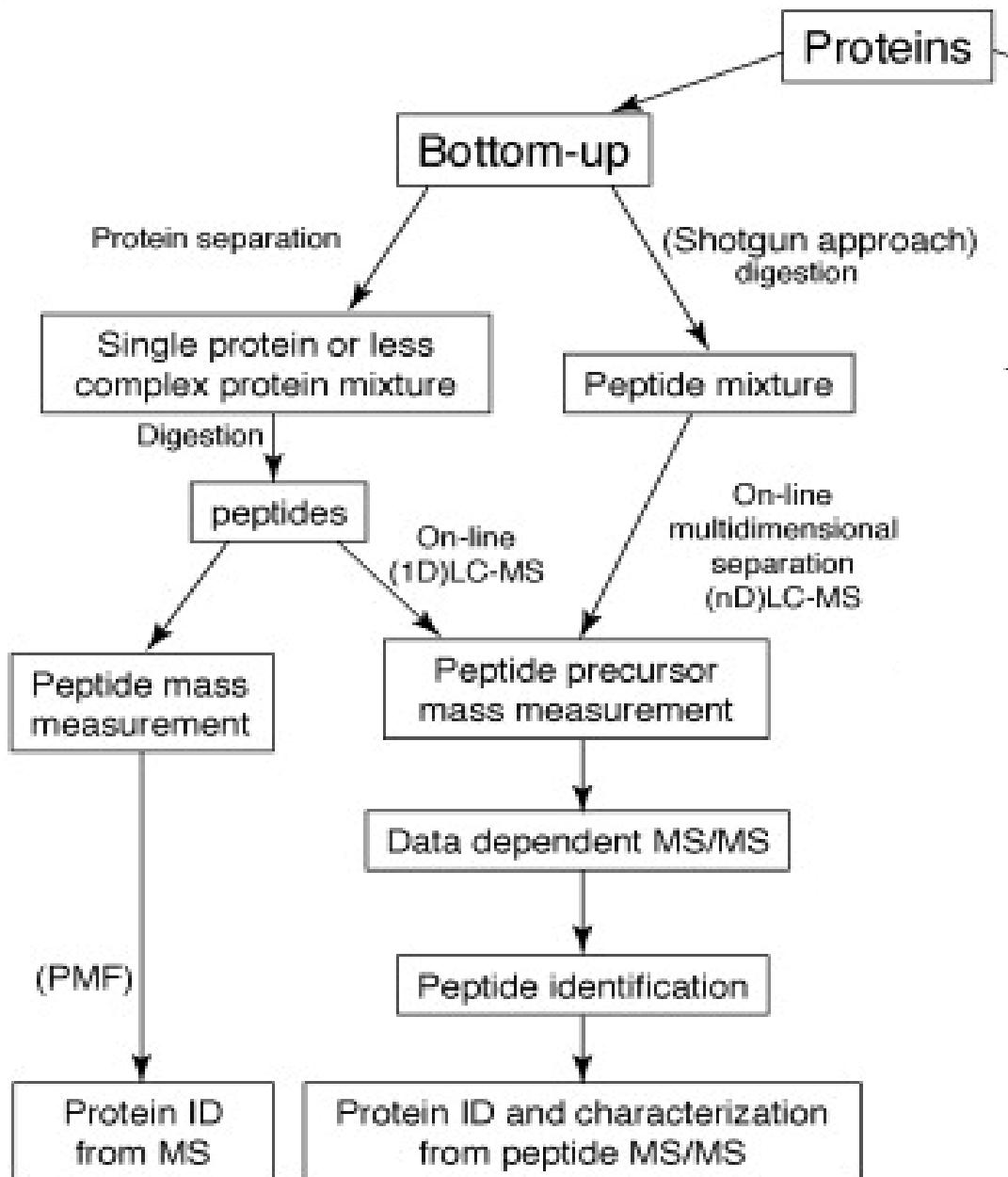
# Proteomics



# Proteins extracted from biological samples can be analyzed by bottom-up or top-down methods

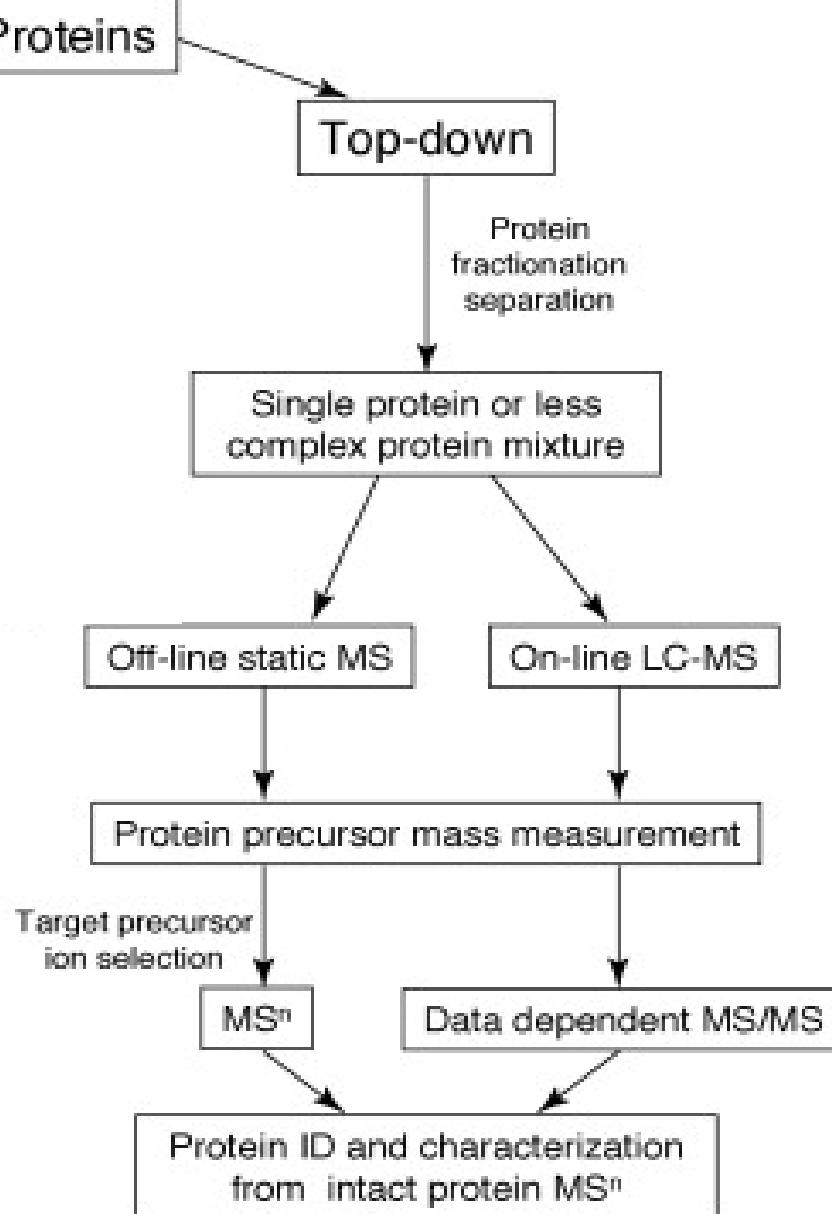


# Bottom-up approach



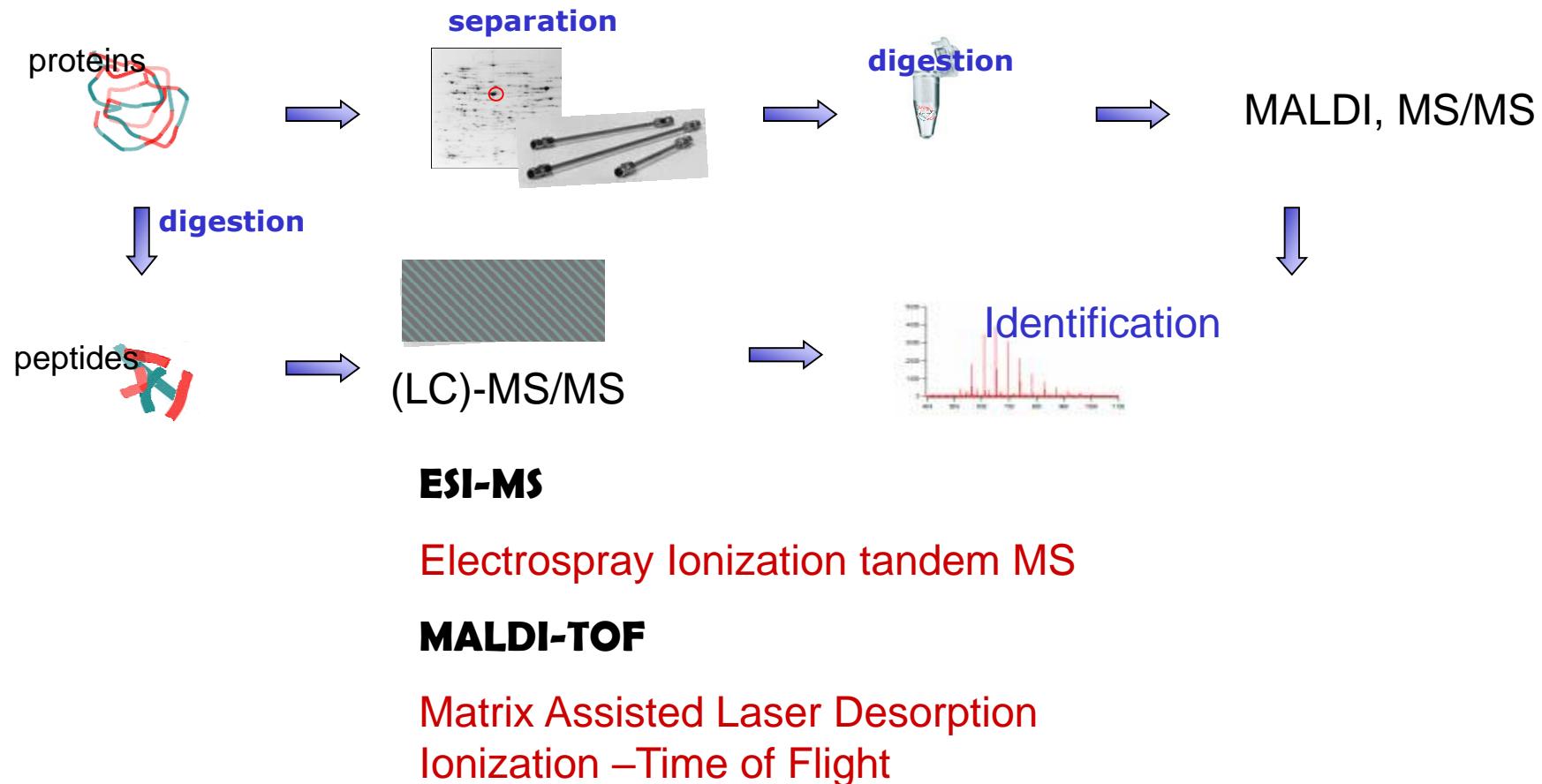
In the **bottom-up approach**, proteins in complex mixtures can be **separated before enzymatic (or chemical) digestion** followed by direct peptide mass fingerprinting-based acquisition or further peptide separation on-line coupled to tandem mass spectrometry. Alternatively, the protein mixture can be directly digested into a collection of **peptides** ('shotgun' approach), which are then separated by multidimensional chromatography on-line coupled to tandem mass spectrometric analysis.

# Top-down approach



In the **top-down approach**, proteins in complex mixtures are fractionated and separated into pure single proteins or less complex protein mixtures, followed by off-line static infusion of sample into the mass spectrometer for **intact protein mass measurement** and **intact protein fragmentation**. An on-line LC-MS strategy can also be used for large-scale protein interrogation.

# General workflow of proteomics analysis



# Separation techniques

## Separation techniques used with intact proteins

1D- and 2D-SDS PAGE

Preparative IEF isoelectric focusing

HPLC

Separating intact proteins to take advantage of their diversity in physical properties

## Separation techniques for peptides

MS-MS

HPLC (MudPIT)

SELDI

## Differential display proteomics

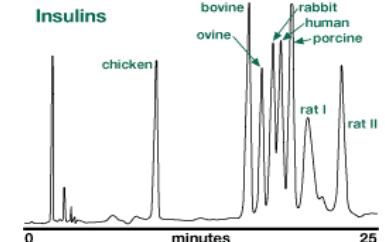
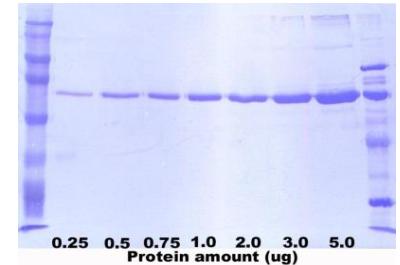
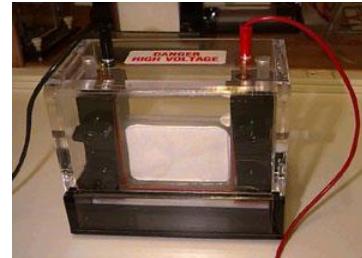
Difference gel electrophoresis (DIGE)

Isotope-coded affinity tagging (ICAT)

**The less complex a mixture of proteins is, the better chance we have to identify more proteins.**

# Protein separation techniques

- By size: gel electrophoresis  
gel filtration chromatography  
ultracentrifugation  
dialysis
- By charge: isoelectric focusing  
ion exchange chromatography
- By polarity: paper and reverse-phase chromatography  
hydrophobic interaction chromatography
- By specificity: affinity chromatography



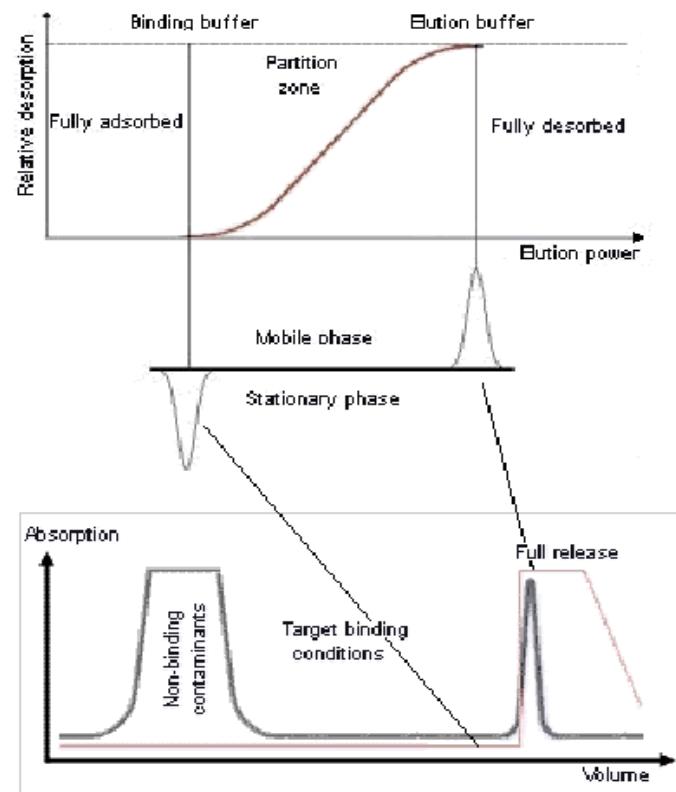
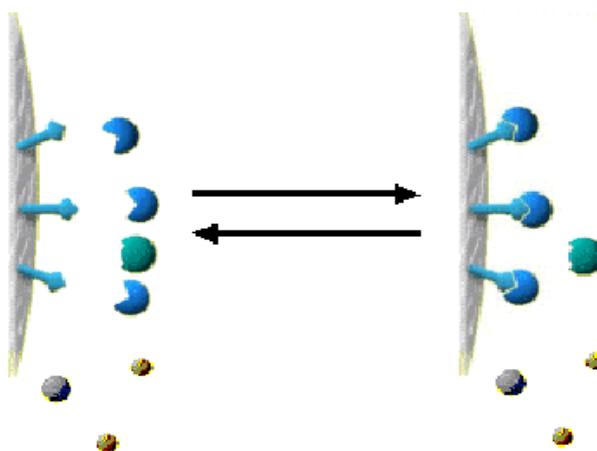
# Protein separation: Chromatography

HPLC

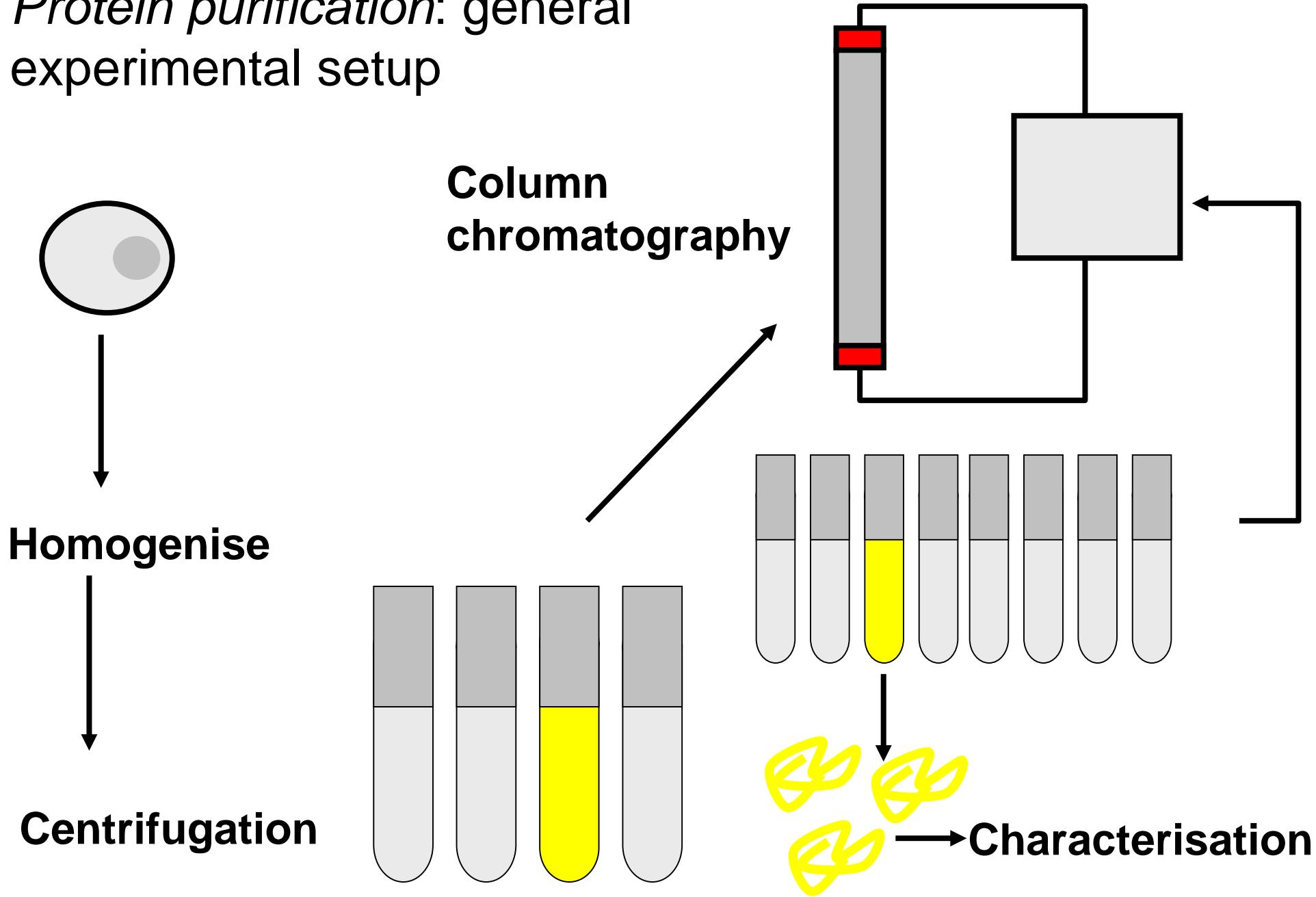
Affinity chromatography

Gel filtration

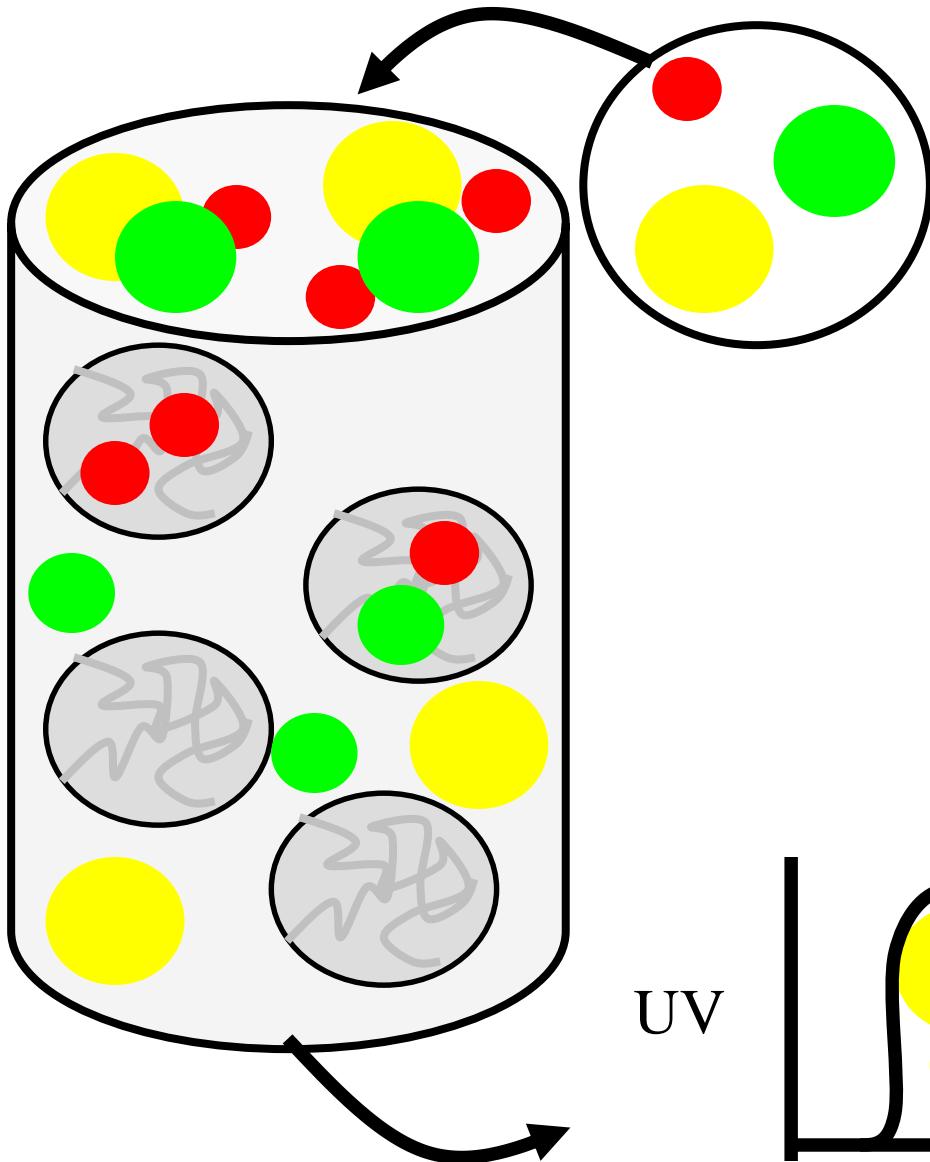
Ion exchange



# *Protein purification: general experimental setup*

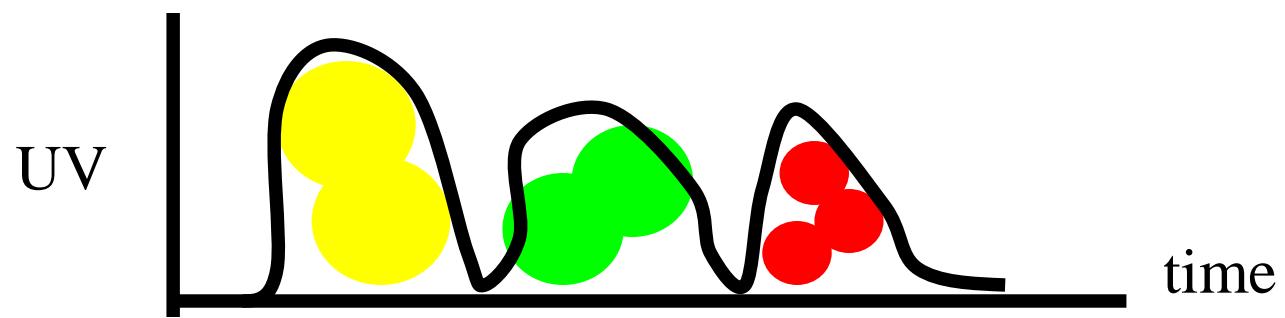


# *Gel permeation chromatography: separating on basis of size*

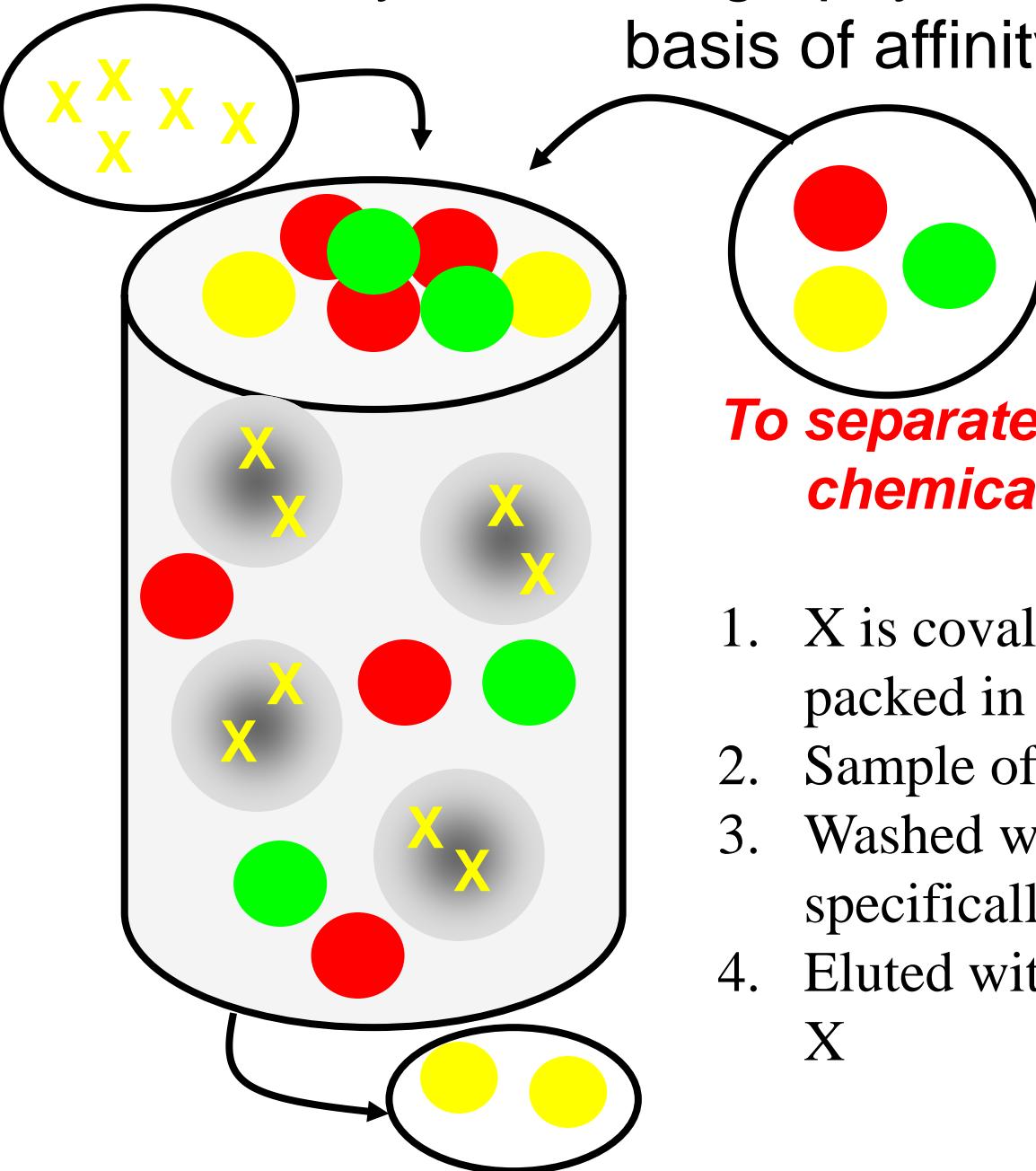


## **Mixture of proteins**

1. A mixture of proteins in a small volume is applied to a column filled with porous beads
2. Because large proteins cannot enter the beads, they emerge sooner than do small ones
3. A detector (e.g. UV) is used to detect protein fragments
4. Fragments are collected separately



# Affinity Chromatography: separating on the basis of affinity



***To separate proteins that recognize a chemical group X***

1. X is covalently attached to beads that are packed in a column
2. Sample of proteins is added
3. Washed with buffer to remove non specifically bound protein
4. Eluted with high concentration of soluble X

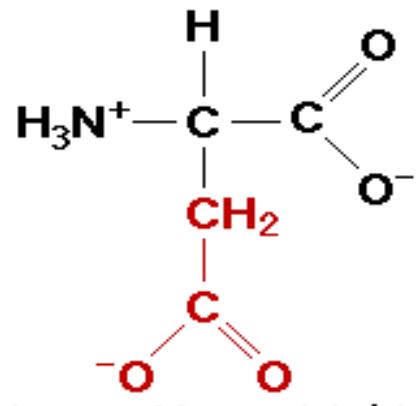
# Separation on the basis of charge

All proteins are charged. Their charges depend on the relative number of **acid** and **basic** amino acids in their primary structure.

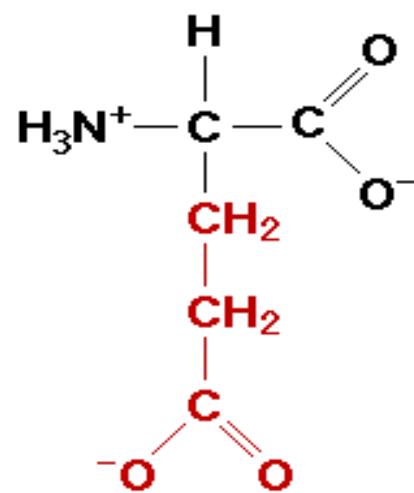
All proteins have a pH value where they are uncharged: the isolelectric point (pl)

H<sub>2</sub>N- **Met Ala Asn Cys His Glu Ser Thr Glu Arg**-COOH

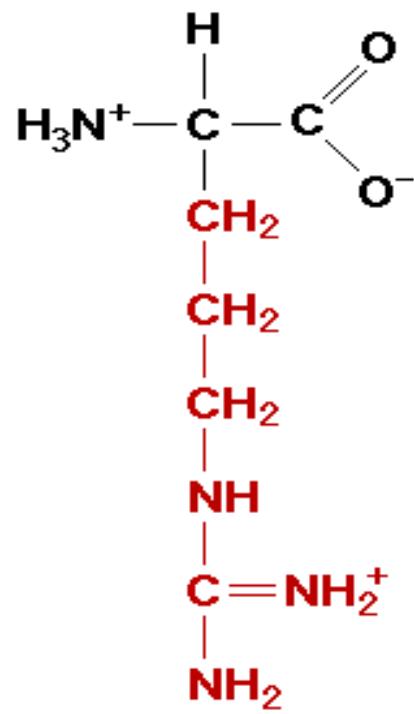
# Ionic amino acids



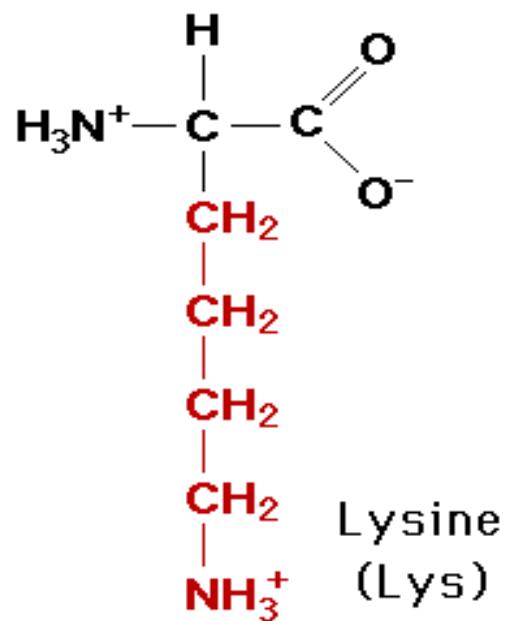
Aspartic acid (Asp)



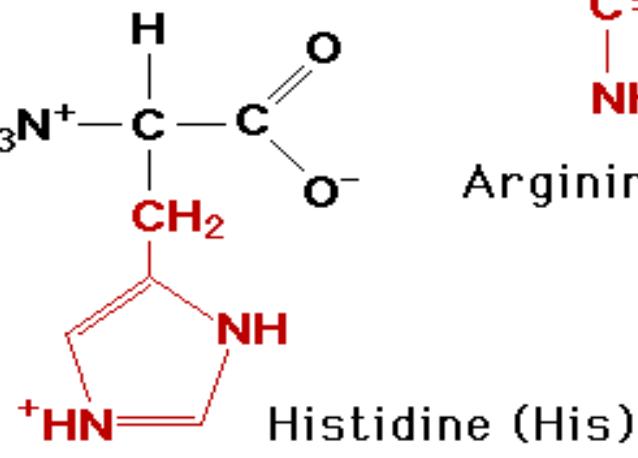
Glutamic acid (Glu)



Arginine (Arg)

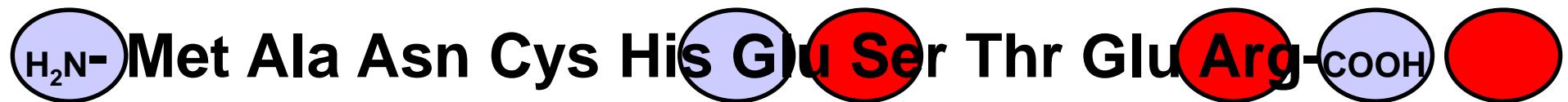


Lysine (Lys)



Histidine (His)

# Separation on the basis of charge



His: 6.0

Glu: 4.1

Arg: 12.5

N-terminal amine: 8.0

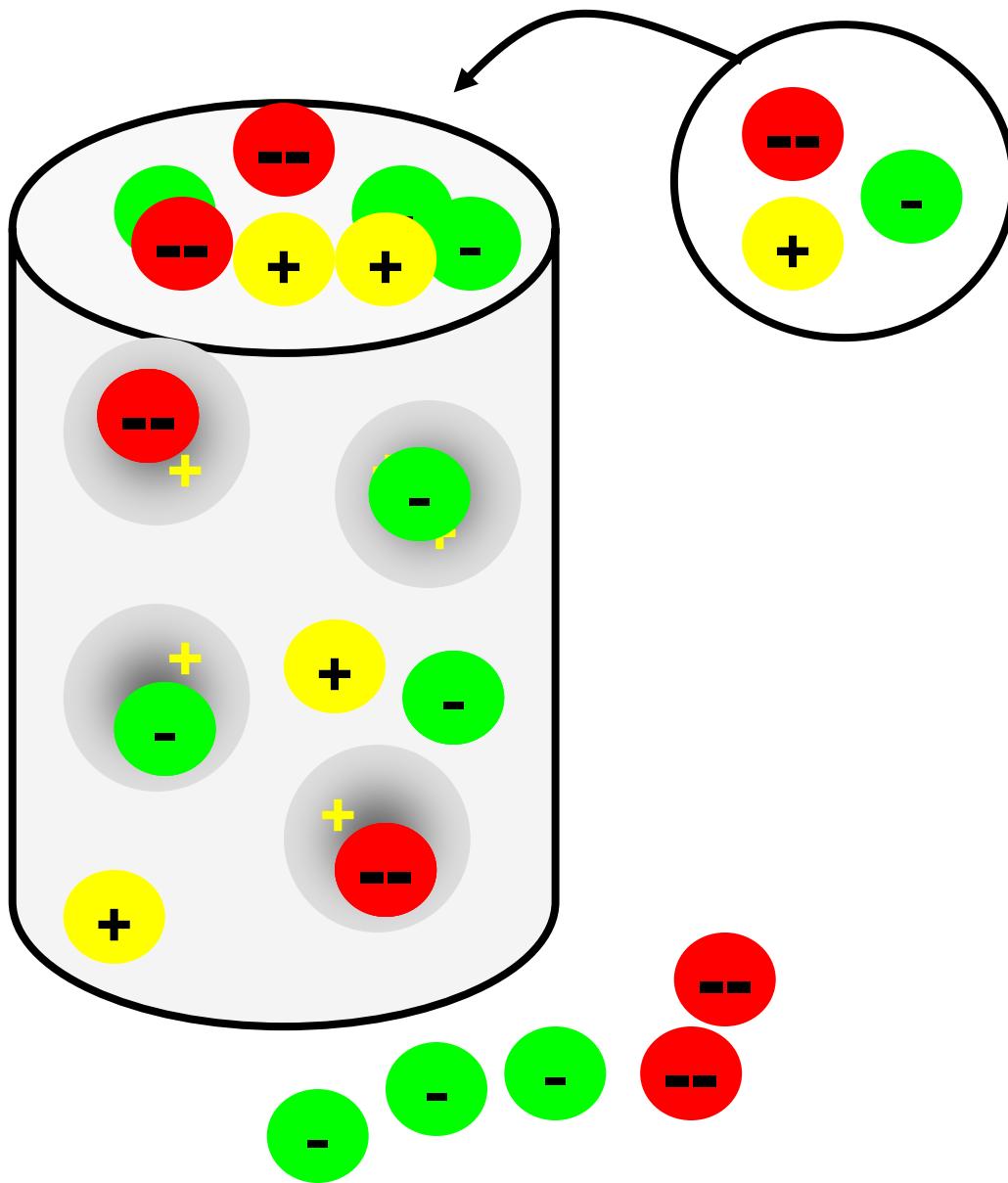
C-terminal acid: 3.1

For this peptide:  
 $pI = \sum pK_a / N = 6.3$

**Positively charged at  $pH < 6.3$**

**Negatively charged at  $pH > 6.3$**

# *Ion Exchange Chromatography:* separation on basis of net charge



1. Positive or negatively charged resin can be used for separation of positive or negatively charged proteins
2. Sample of proteins is added
3. Washed with buffer to remove non specifically bound protein
4. Elute with increasing concentration of salt
5. Proteins with highest net charge come of last

# Enrichment /Fractionation

For the detection of low-abundance proteins, a separation of complex mixtures into fractions with fewer components is necessary



- Enrichment from larger volumes

Selective precipitation

Selective centrifugation

Preparative approaches



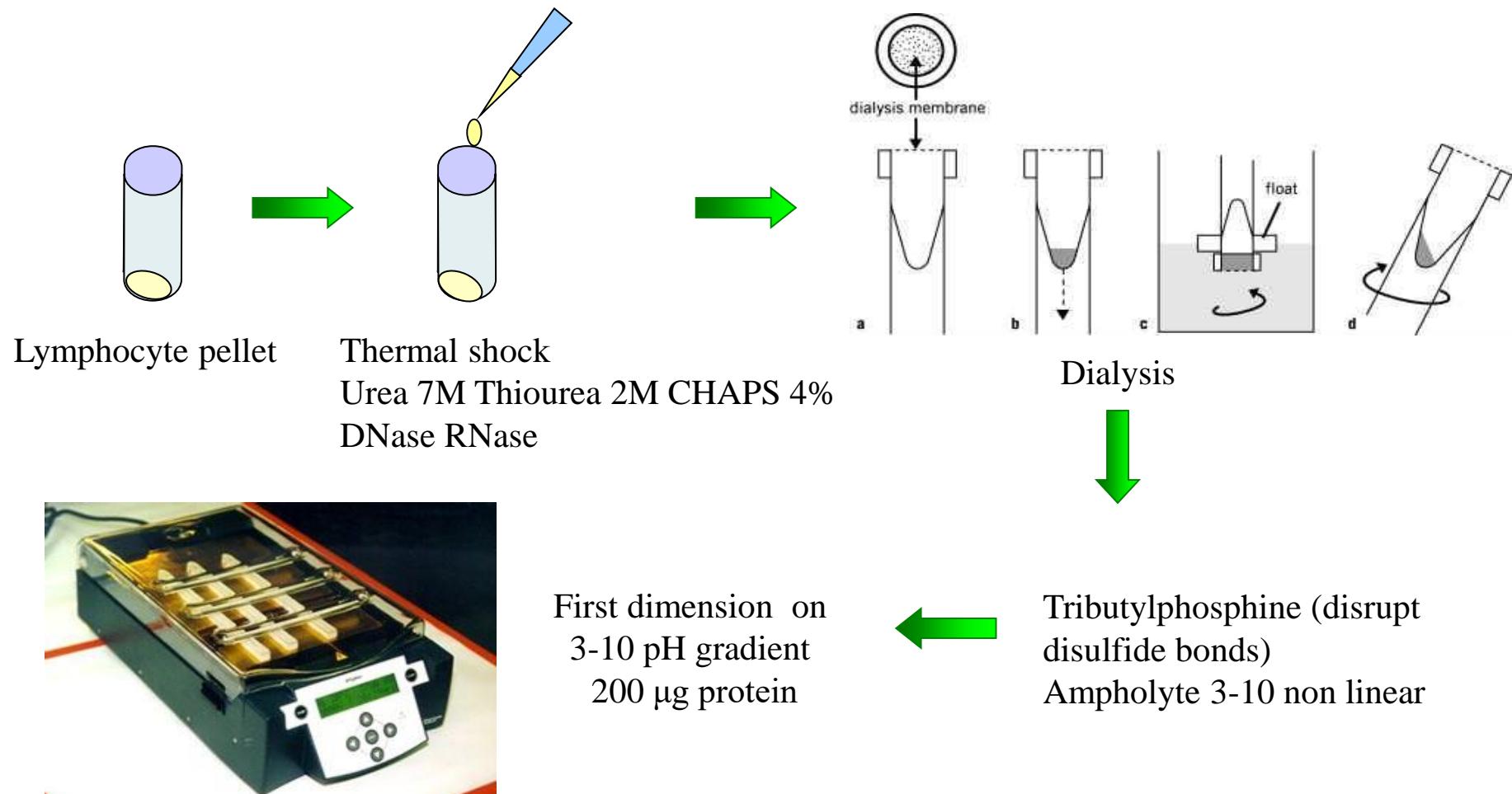
- Combination of 2DE with LC



- Multi-dimensional LC

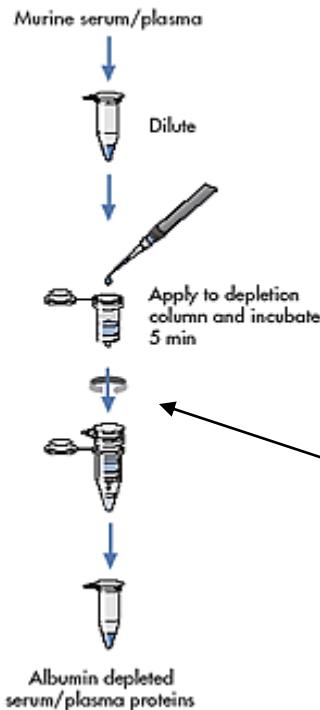
# Microdialysis

**Dialysis:** Diffusion through a semi-permeable cellulose membrane. Different pore sizes allow removal of molecules smaller than specific MW



# Albumin depletion

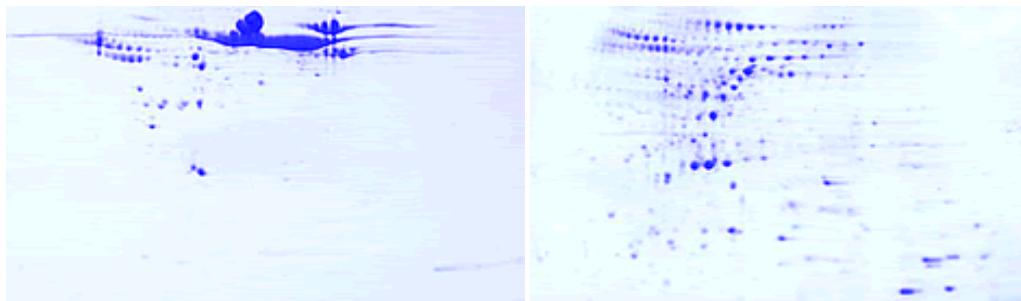
## Albumin Depletion Procedure



Diluted samples are applied to the resin in the spin columns, which are then sealed and incubated on an end-over-end shaker. Plasma and serum proteins, depleted of albumin, are recovered by centrifugation

Albumin in the sample is bound by antibodies immobilized on a solid support in the spin columns

## Albumin Depletion Facilitates Analysis of Low-Abundance Proteins



Coomassie stained 2D-PAGE gels showing non-depleted (left) and depleted (right) rat plasma samples.

# Types of electrophoresis

## **Filter paper electrophoresis:**

Proteins are easily denatured due to the high absorbance of filter paper

Works for small peptides or amino acids.

## **Thin layer electrophoresis (TLC):**

Chemically-modified cellulose

## **Gel electrophoresis**

1. Starch gel electrophoresis
2. Agarose gel electrophoresis
3. Polyacrylamide gel electrophoresis (PAGE)

# Main types of PAGE

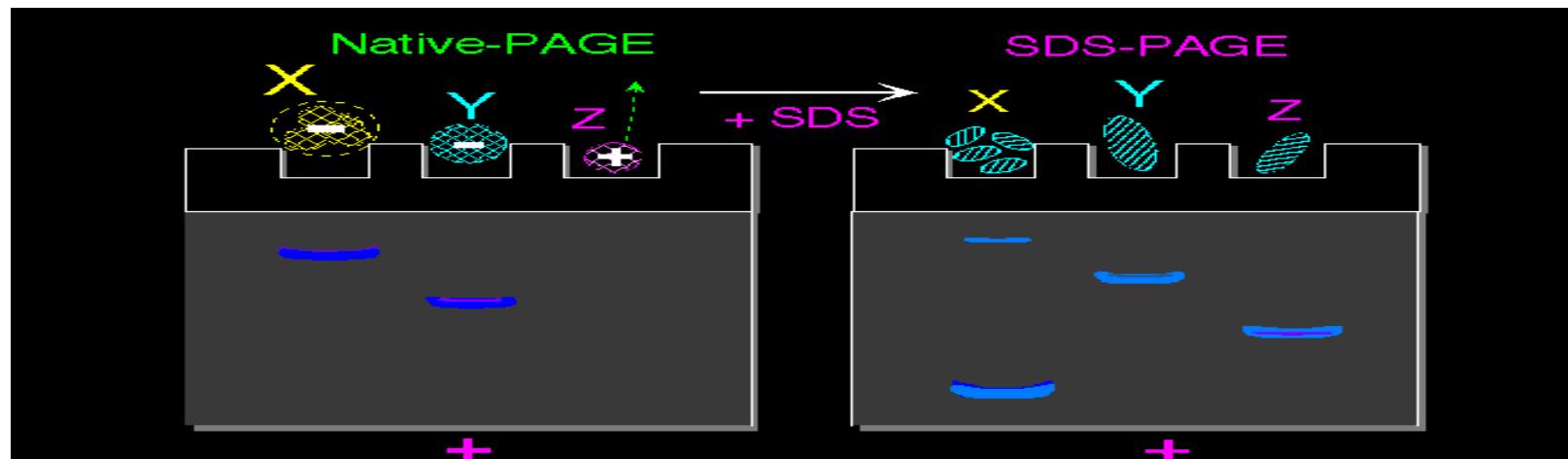
## Native-PAGE

- \* Enzyme activities are retained after electrophoresis, so enzymatic assays can be performed on separated proteins
- \* Factors affecting mobility: **charge; molecular weight and the shape of proteins**

## SDS-PAGE

- \* SDS (sodium dodecyl sulfate) coats the surface of proteins
- \* Proteins are denatured, so enzyme activities are lost after SDS-PAGE.
- \* Factors affecting mobility : **molecular weight**

## Separation of proteins in native and SDS-PAGE gels

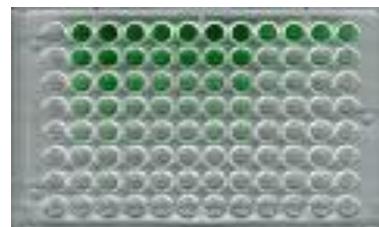
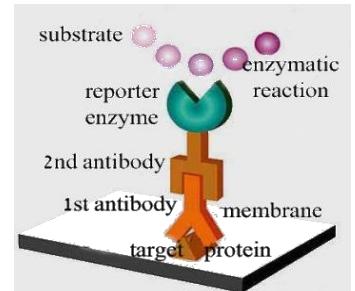
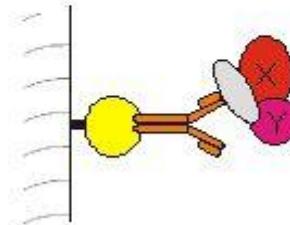


# Immunological techniques

Proteins are very strong antigens. Polyclonal and monoclonal antibodies have been developed to almost all known human proteins. They have been used extensively to study cellular functions of proteins. Immunological techniques are based on the specificity of antibody-antigen interaction

- Immunoprecipitation assay
- Immunoblotting (Western blotting)
- ELISA (Enzyme-Linked ImmunoSorbent Assays)

Co-immunoprecipitation



# Immunoprecipitation assay

## **What is IP assay?**

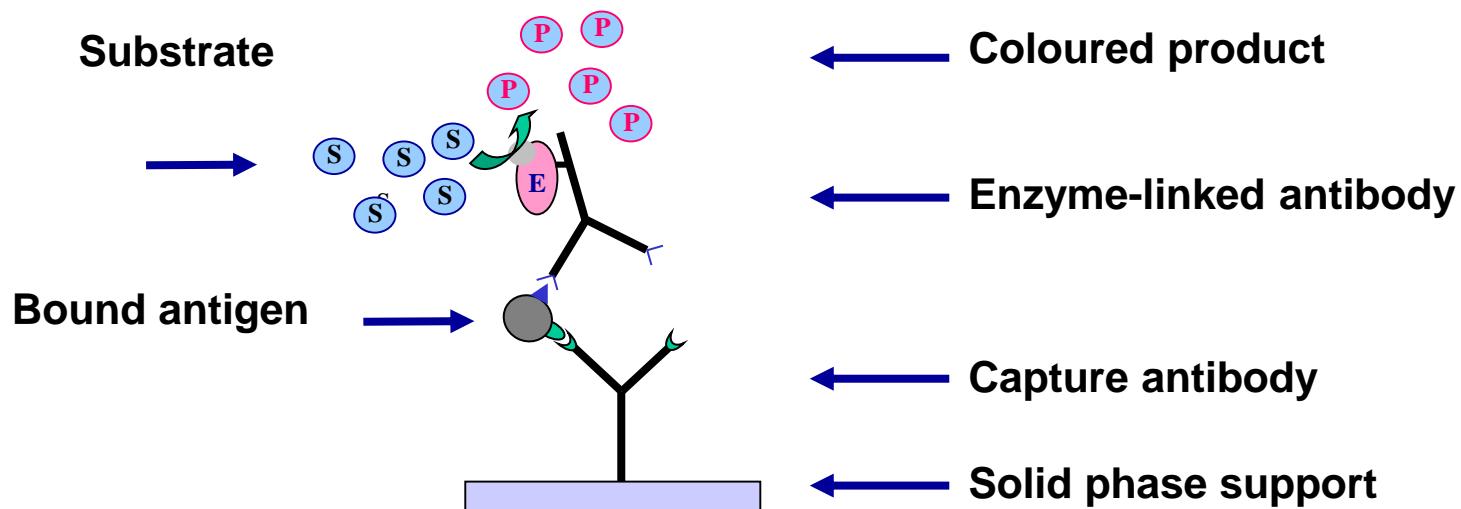
Immunoprecipitation is a procedure which permits the purification of the protein of interest with the use of specific antibody (poly- or monoclonal)

## **Where to apply?**

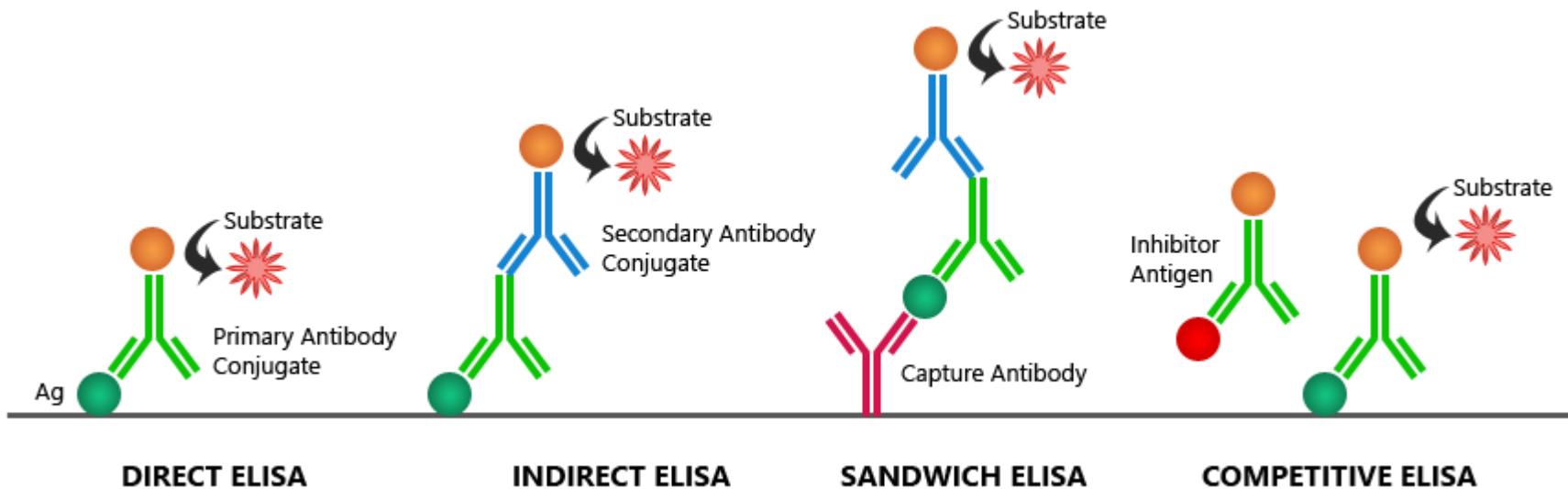
- \* To analyse the activity of immunoprecipitated proteins
- \* To prove the interactions between two proteins
- \* To isolate multienzyme complexes and to identify their components
- \* To analyse protein-DNA interactions (chromatin immunoprecipitation assay)

# ELISA

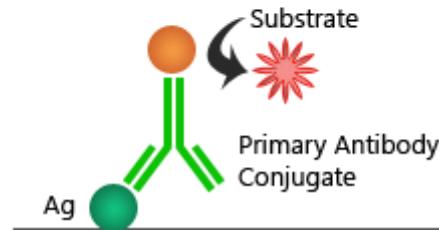
This technique combines the specificity of the antibody-antigen interaction with the sensitivity of **enzyme assays** using either an antibody or an antigen conjugated to an enzyme. The activity of the enzyme is measured by adding an appropriate chromogenic substrate, which is converted to a coloured product. **Automation** of the assay is straightforward, so it is often used in drug discovery programs and, thanks to its high sensitivity, is often used in diagnostic kits



# ELISA: general procedure



# Direct ELISA



For direct detection, an antigen coated to a multi-well plate is detected by an antibody that has been directly conjugated to an enzyme. This detection method is a good option if there is no commercially available ELISA kits for your target protein.

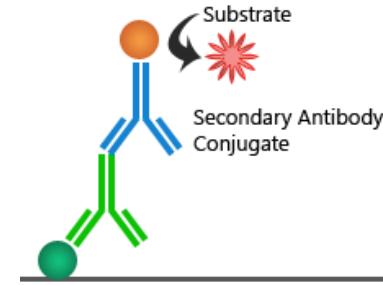
## 1. Advantages

1. Quick because only one antibody and fewer steps are used.
2. Cross-reactivity of secondary antibody is eliminated.

## 2. Disadvantages

1. Labeling primary antibodies for each specific ELISA system is time-consuming and expensive.
2. No flexibility in choice of primary antibody label from one experiment to another.
3. Minimal signal amplification.

# Indirect ELISA



For indirect detection, the antigen coated to a multi-well plate is detected in two stages or layers. First an unlabeled primary antibody, which is specific for the antigen, is applied. Next, an enzyme-labeled secondary antibody is bound to the first antibody. The secondary antibody is usually an anti-species antibody and is often polyclonal. The indirect assay, the most popular format for ELISA, has the advantages and disadvantages:

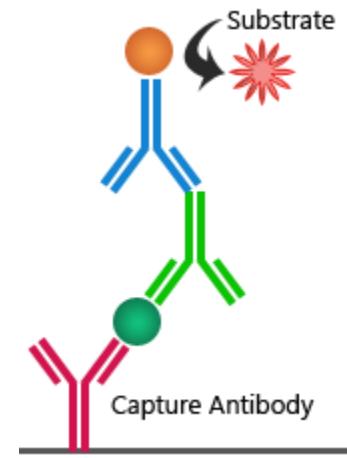
## Advantages

1. A wide variety of labeled secondary antibodies are available commercially.
2. Versatile because many primary antibodies can be made in one species and the same labeled secondary antibody can be used for detection.
3. Maximum immunoreactivity of the primary antibody is retained because it is not labeled.
4. Sensitivity is increased because each primary antibody contains several epitopes that can be bound by the labeled secondary antibody, allowing for signal amplification.

## Disadvantages

1. Cross-reactivity might occur with the secondary antibody, resulting in nonspecific signal.
2. An extra incubation step is required in the procedure.

## Sandwich ELISA

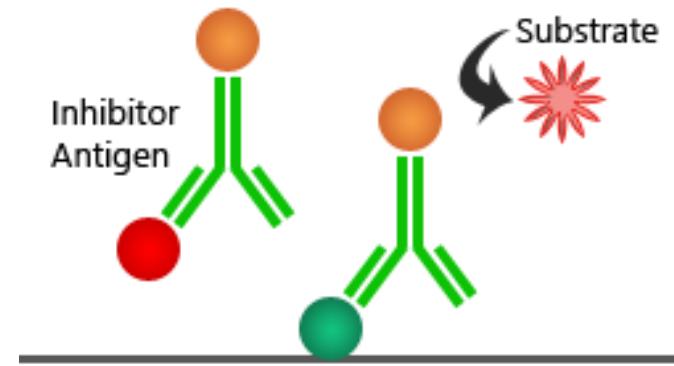


Sandwich ELISAs typically require the use of matched antibody pairs, where each antibody is specific for a different, non-overlapping part (epitope) of the antigen molecule. A first antibody (known as capture antibody) is coated to the wells. The sample solution is then added to the well. A second antibody (known as detection antibody) follows this step in order to measure the concentration of the sample.

This type of ELISA has the following advantages:

- High specificity: the antigen/analyte is specifically captured and detected
- Suitable for complex (or crude/impure) samples: the antigen does not require purification prior to measurement
- Flexibility and sensitivity: both direct or indirect detection methods can be used

# Competitive ELISA



The key event of competitive ELISA is the process of competitive reaction between the sample antigen and antigen bound to the wells of a microtiter plate with the primary antibody.

First, the primary antibody is incubated with the sample antigen and the resulting antibody–antigen complexes are added to wells that have been coated with the same antigen. After an incubation period, any unbound antibody is washed off. The more antigen in the sample, the more primary antibody will be bound to the sample antigen. Therefore, there will be a smaller amount of primary antibody available to bind to the antigen coated on the well, resulting in a signal reduction.

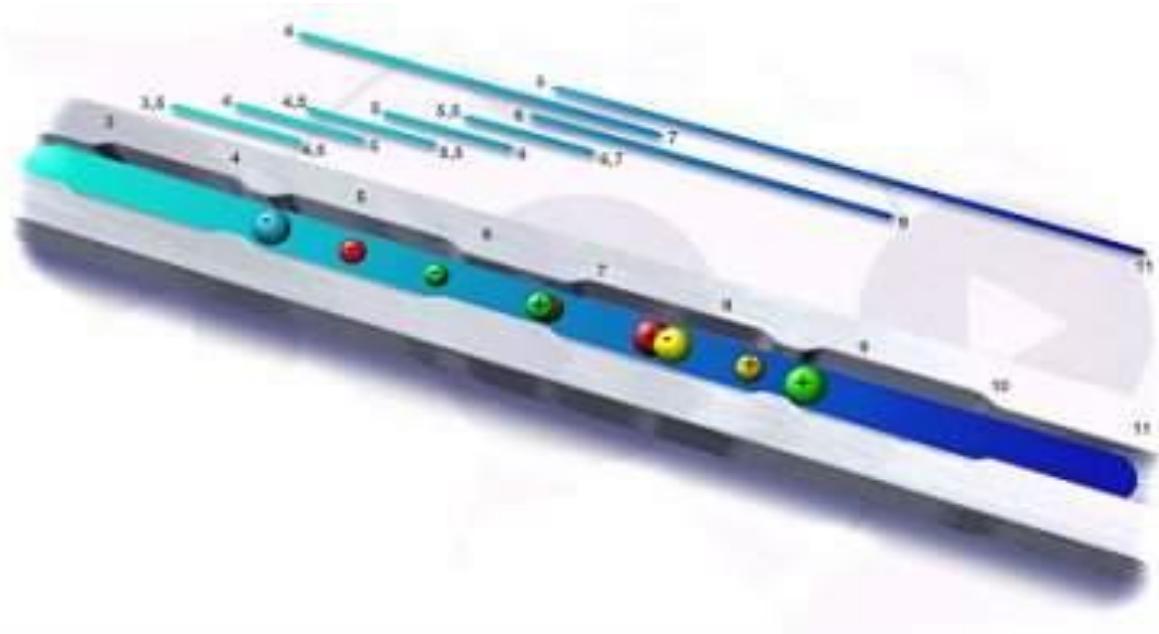
The main advantage of this type of ELISA arises from its **high sensitivity** to compositional differences in complex antigen mixtures, even when the specific detecting antibody is present in relatively small amounts.

# CLASSIC MS PROTEOMICS APPROACH

- SAMPLE SOLUBILIZATION
- IEF
- SDS - UREA 7M, TIOUREA 2M, CHAPS
- RESULT 4%
- GEL ANALYSIS
- MASS SPECTROMETRY

# CLASSIC APPROACH

- SAMPLE SOLUBILIZATION
- IEF
- SDS - PAGE
- VISUALIZZAZIONE
- ANALISI DEI GEI
- SPETTROMETRIA DI MASSA



# CLASSIC APPROACH

- SOLUBILIZZAZIONE
- IEF
- SDS - PAGE
- RESULTS VISUALISATION
- GEL ANALYSIS
- MASS SPECTROMETRY

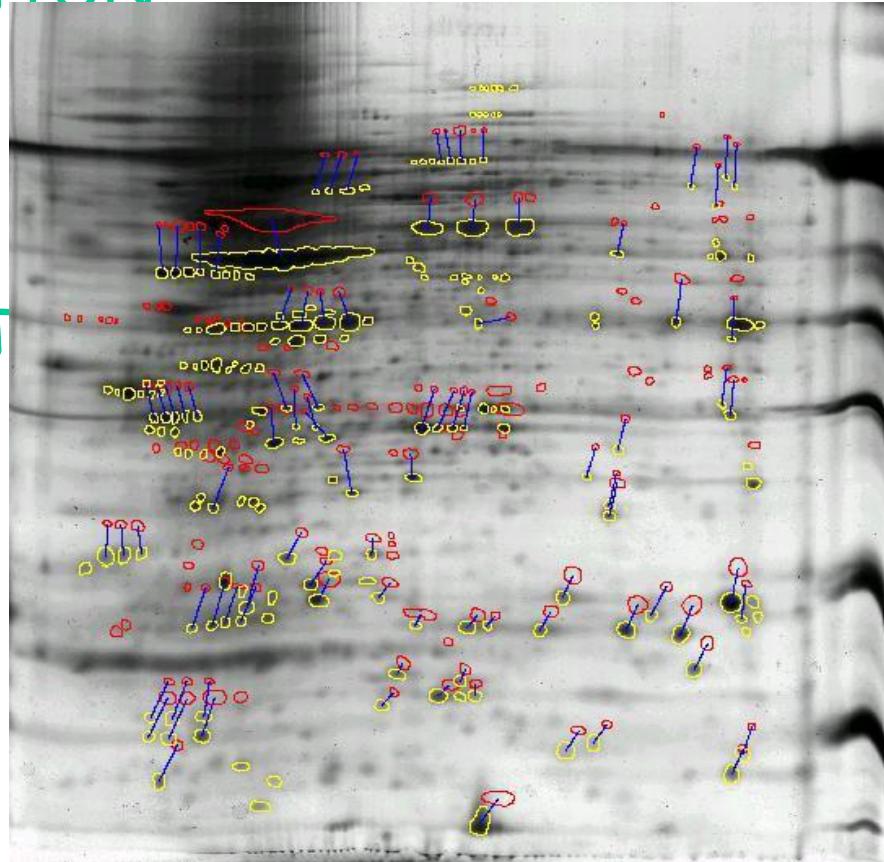


# CLASSIC APPROACH

- SOLUBILIZZAZIONE DEL CAMPIONE
- IEF
- SDS - PAGE
- RESULTS VISUALIZATION
  - *•immunoblotting*
  - *•Coomassie Blue or silver staining*

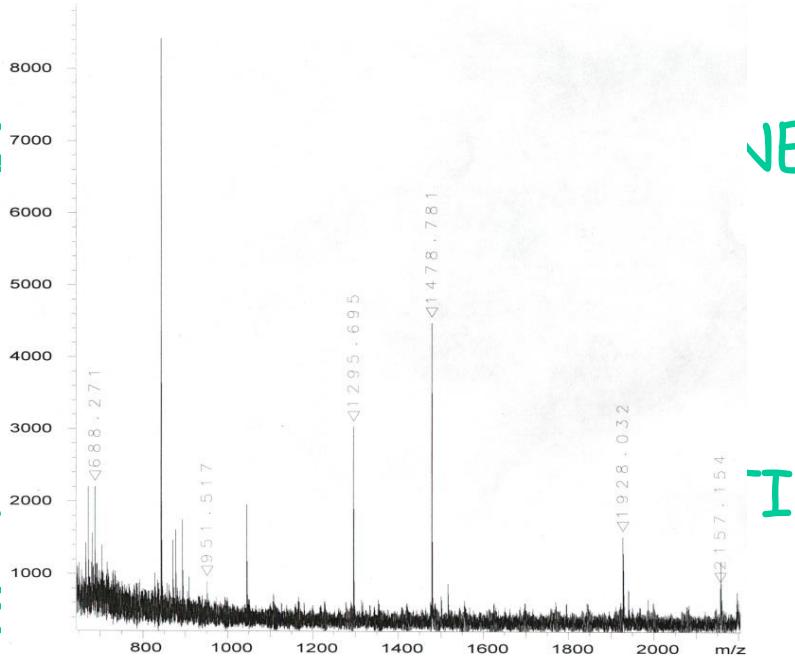
# L'APPROCCIO CLASSICO

- SAMPLE SOLUBILIZATION
- IEF
- SDS - PAGE
- RESULTS VISUALIZATION
- GEL ANALYSIS
- MASS SPECTOMETRY

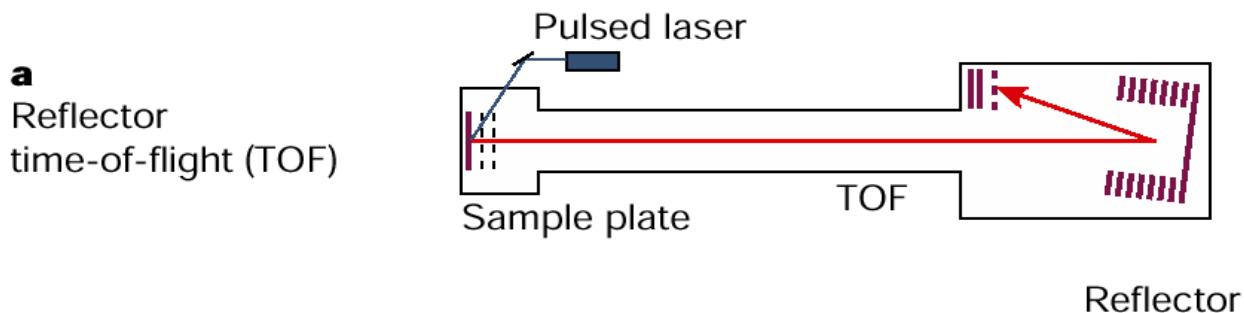


# CLASSIC APPROACH

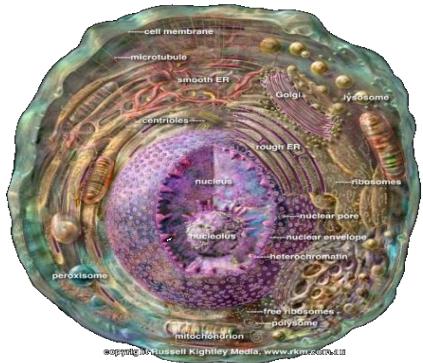
- SOLUBILIZZARE
- IEF
- SDS - PAGE
- VISUALIZZARE
- GEL ANALYSIS



- MASS SPECTROMETRY (MALDI)



# Sample preparation



EXTRACTION  
Different components  
or  
and  
SOLUBILISATION



An efficient sample preparation should:

1. **Solubilize in a reproducible way all classes of proteins, including the hydrophobic ones**
2. **Prevent protein aggregation and keep solubility during IEF**
3. **Prevent chemical and enzymatic modifications during extraction process**
4. **Digest or remove nucleic acids and other molecules that can interfere with the analysis**
5. **Enriches target proteins (e.g., by eliminating most abundant proteins like albumin)**



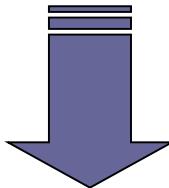
# Cell Disruption Methods

## Techniques used for the physical disruption of cells

Lysis Method	Description	Apparatus
Mechanical	Waring Blender Polytron	Rotating blades grind and disperse cells and tissues
Liquid Homogenization	Dounce Homogenizer Potter-Elvehjem Homogenizer French Press	Cell or tissue suspensions are sheared by forcing them through a narrow space
Sonication	Sonicator	High frequency sound waves shear cells
Freeze/Thaw	Freezer or dry ice/ethanol	Repeated cycles of freezing and thawing disrupt cells through ice crystal formation
Manual grinding	Mortar and pestle	Grinding plant tissue, frozen in liquid nitrogen

# Sample preparation

Proteins must be denatured and soluble

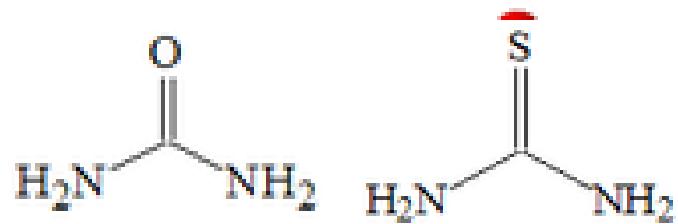


This can be obtained by specific treatments with

- Urea (denaturing agent)
- Thiol (reducing agent)
- Detergents
- protease inhibitors

Low ionic strength conditions to allow high voltages during isoelectrofocusing without producing high current

# Protein solubilization



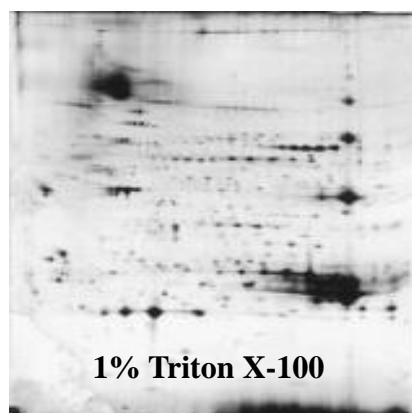
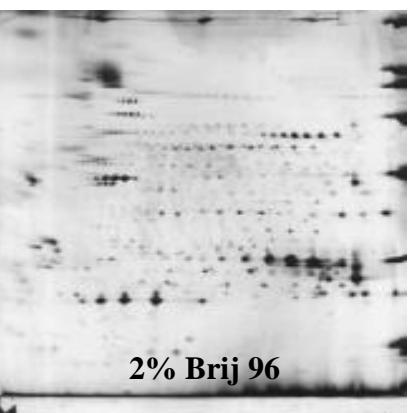
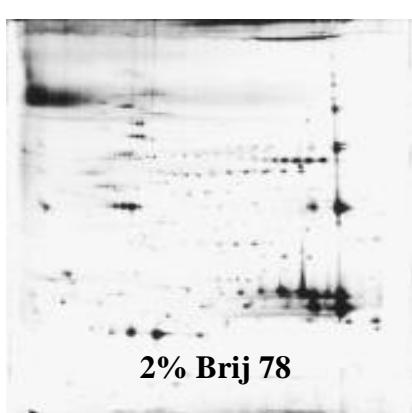
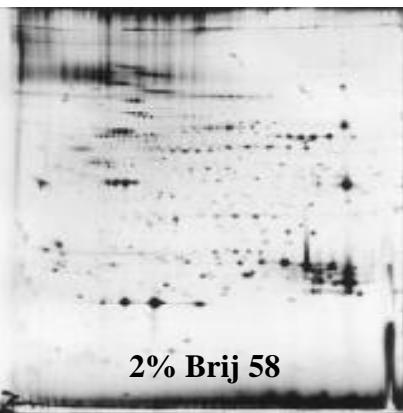
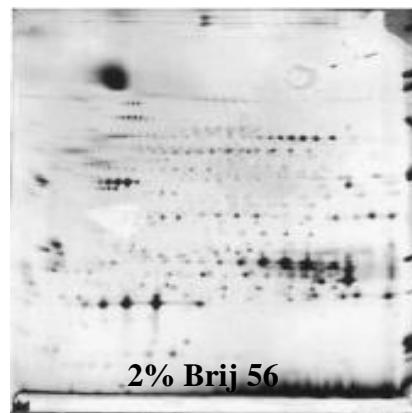
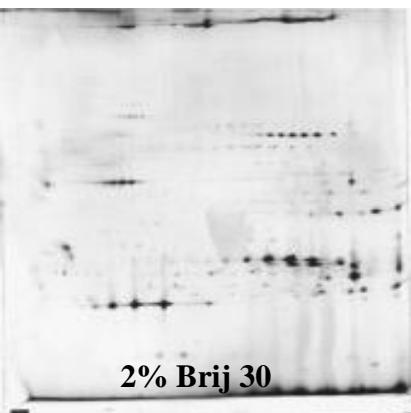
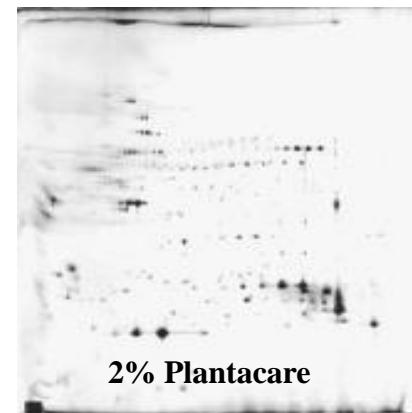
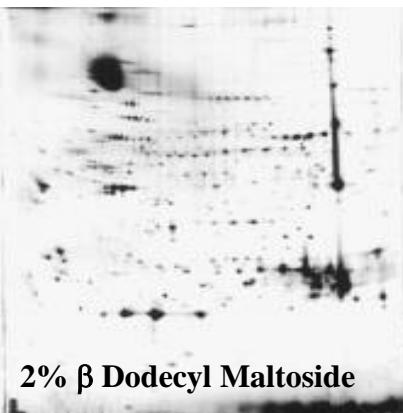
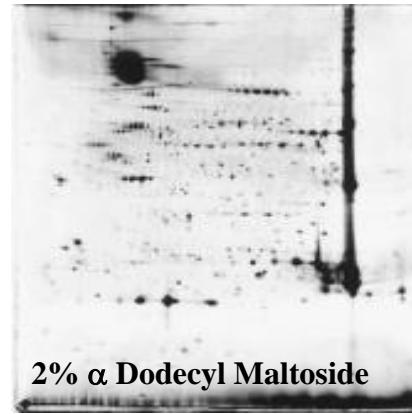
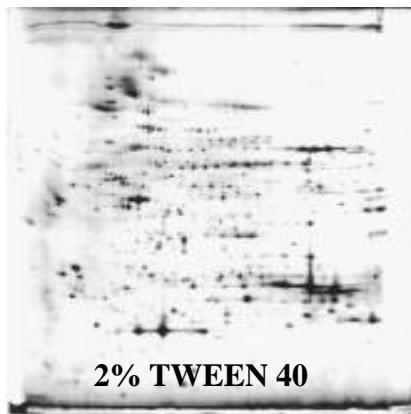
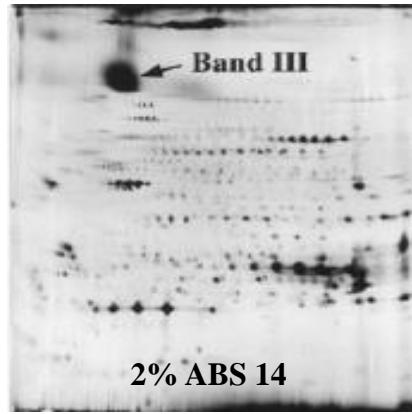
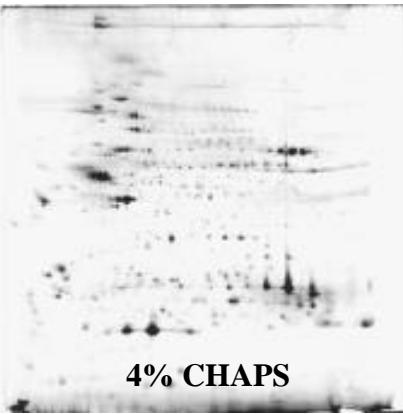
## UREA and THIOUREA

Chaotropic agents of choice for disrupting hydrogen bonds

**Urea 8M** or mixtures of thiourea 2M and urea 5-8M.

## DETERGENTS

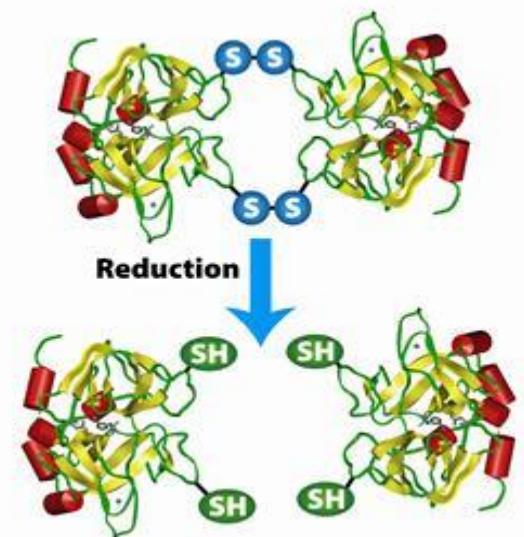
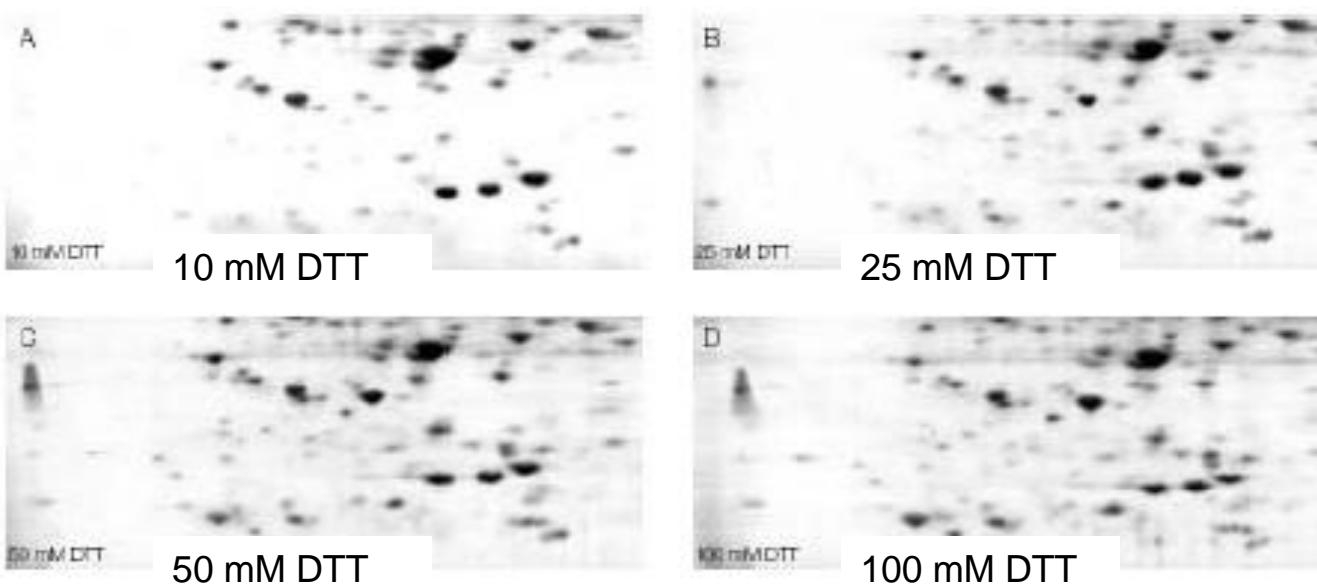
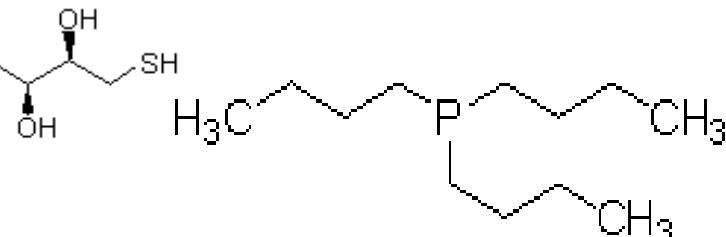
Detergents are a class of molecules whose unique properties enable manipulation (disruption or formation) of hydrophobic-hydrophilic interactions among molecules in biological samples. They are used to lyse cells (release soluble proteins), solubilize membrane proteins and lipids, control protein crystallization, prevent nonspecific binding in affinity purification and immunoassay procedures, and as additives in electrophoresis. Nonionic (uncharged) or zwitterionic (having both positively and negatively charged groups but with a net charge of zero) detergents must be used to allow protein migration according to its net charge (SDS cannot be used) .



# Dysulfide reducing agents

To break dysulfure bonds (and analyze separate protein subunits): dithiotreitol (DTT) o tributyl-phosphine (TBP)

A solution of DTT 50mM is enough for the majority of proteins, for difficult cases TBP must be used



At increasing DTT concentrations the number of spots increases

## **Proteolysis Protection**

- PMSF (phenylmethylsulphonyl fluoride)
- Pefabloc
- EDTA
- EGTA
- leupeptin

## **Contaminant Removal**

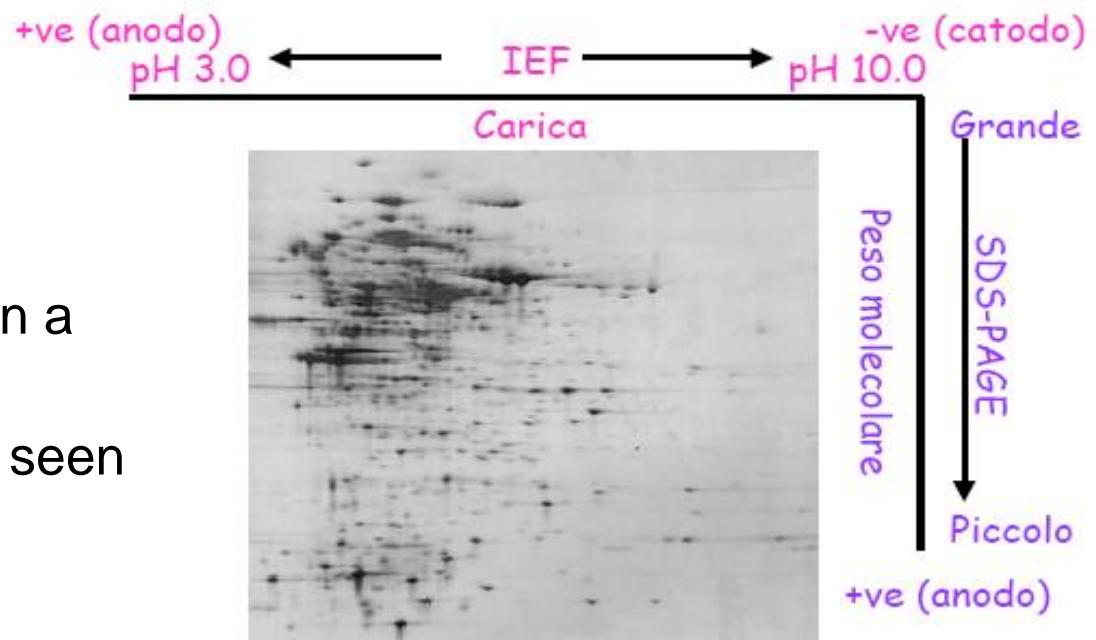
- Endonuclease treatment
- Dialysis
- Filtration
- Centrifugation
- Chromatography
- Solvent Extraction

***Prevent keratin contamination!!***

# 2D Gel Electrophoresis

Proteins have two important features that are exploited for 2D separation

1. Isoelectric point (pI)
2. Molecular weight (MW)



- Simultaneous separation and detection of ~2000 proteins on a 20x25 cm gel
- Up to 10,000 proteins can be seen using optimized protocols

# 1<sup>st</sup> dimension

## IsoElectric Focusing, IEF

### Immobilized pH gradients (IPGs)

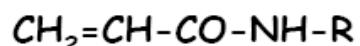


A pH gradient is generated by a limited number of well defined chemicals (**immobilines**) which are co-polymerized with the acrylamide matrix.

Migration of proteins in a pH gradient: protein stop at pH=pl

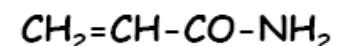
Individual strips:  
Loading quantities (18 cm strip)  
Use ~~14, 18, 11, 7 cm long~~ strips to  
Analytical run: 50-100 µg  
focus on particular pl range  
3 mm wide  
Micropreparative runs: 0,5 – 10 mg  
0,5 mm thickness

#### Immobiline:



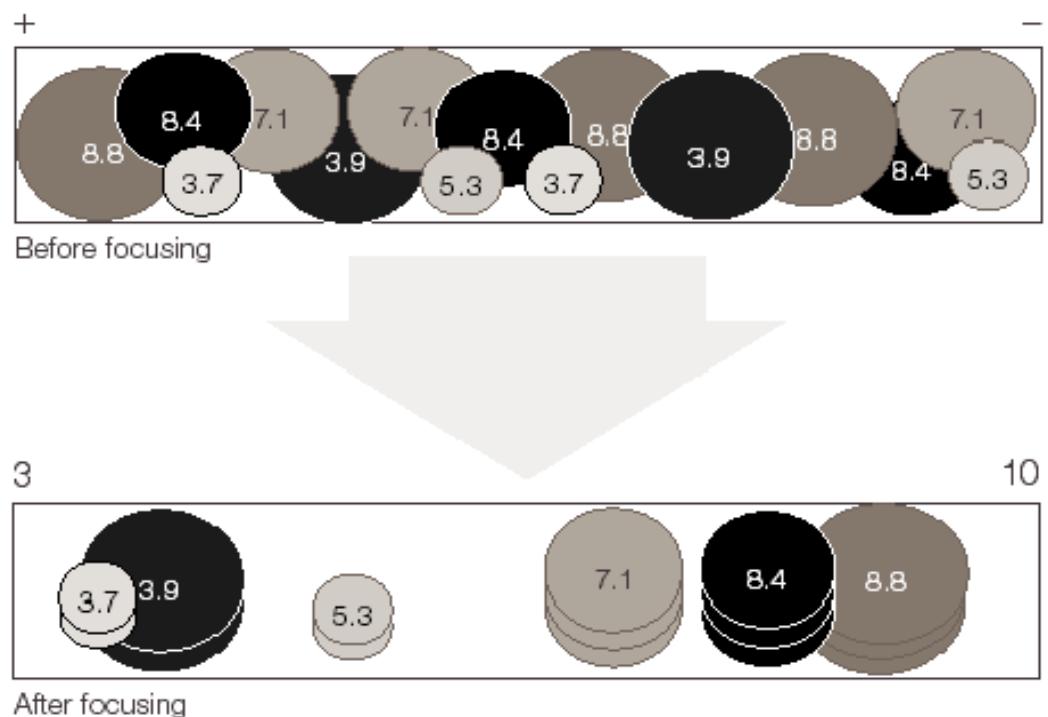
R = un **gruppo carbossilico** (-COOH)  
o una **ammina terziaria** (-NH<sub>2</sub>)

#### Acrilamide:

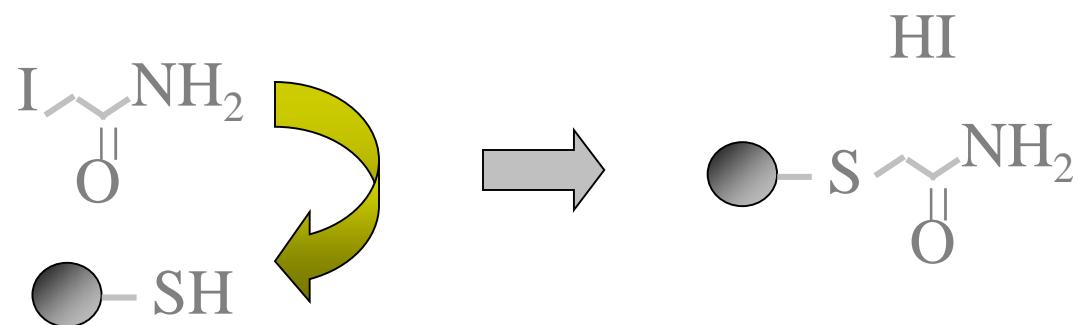
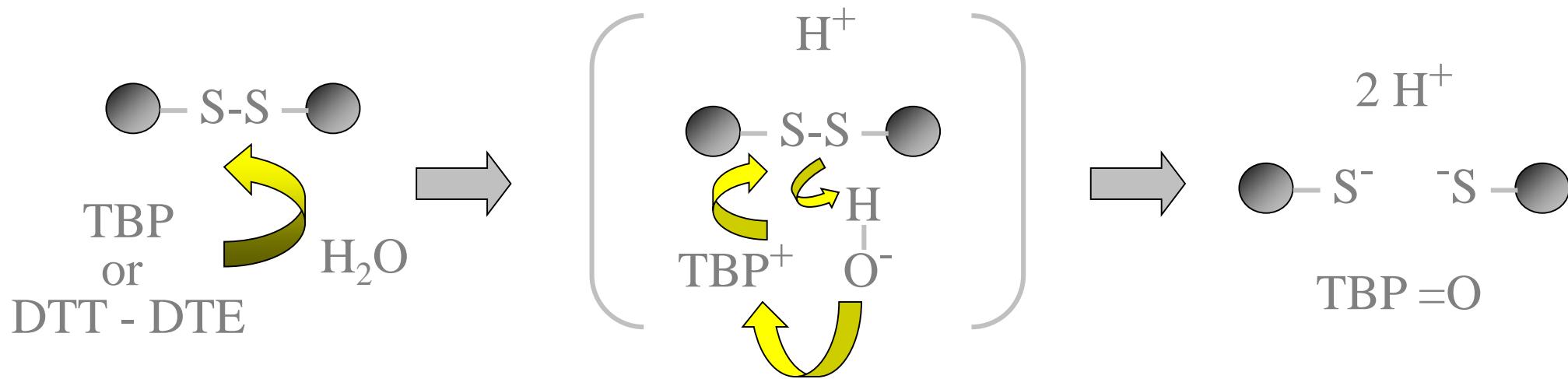


# IEF Principles

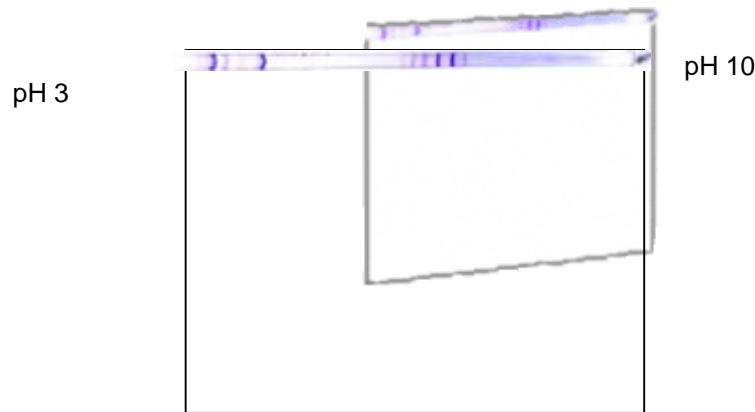
- Separation of basis of  $pI$ , not  $M_w$
- Requires very high voltages (5000V)
- Requires a long period of time (10h)
- Presence of a pH gradient is critical
- Degree of resolution determined by slope of pH gradient and electric field strength
- Uses ampholytes to establish pH gradient



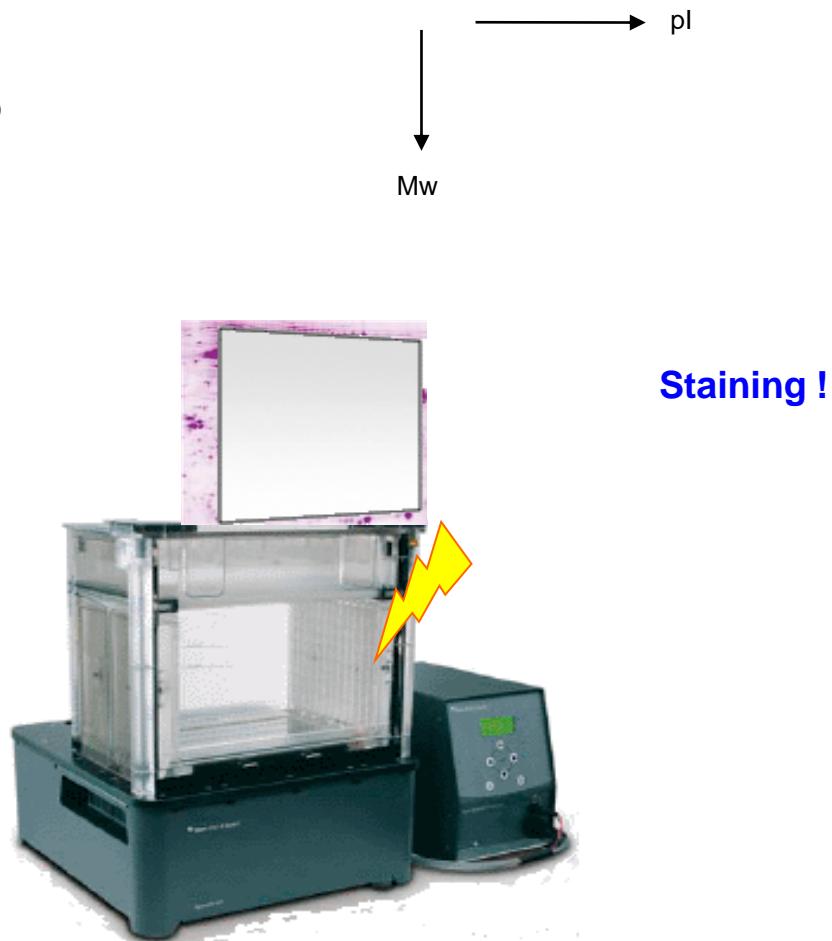
# Equilibration of the resolved IPG strips: reduction and alkylation reaction, SDS coating



# 2<sup>nd</sup> dimension



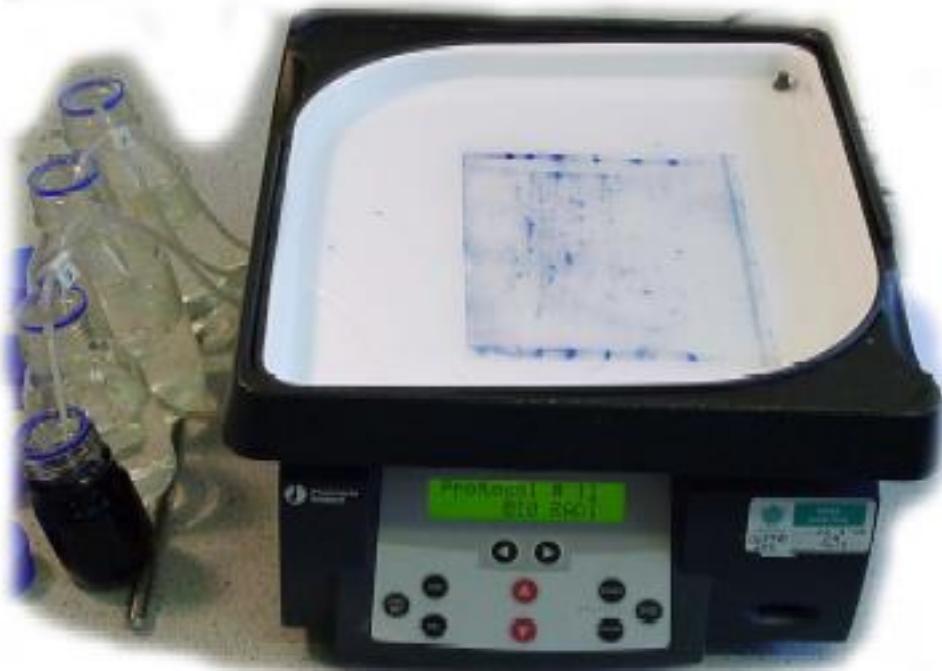
The strip is loaded onto a SDS gel



Proteins that were separated on IEF gel are next separated in the second dimension based on their molecular weights.

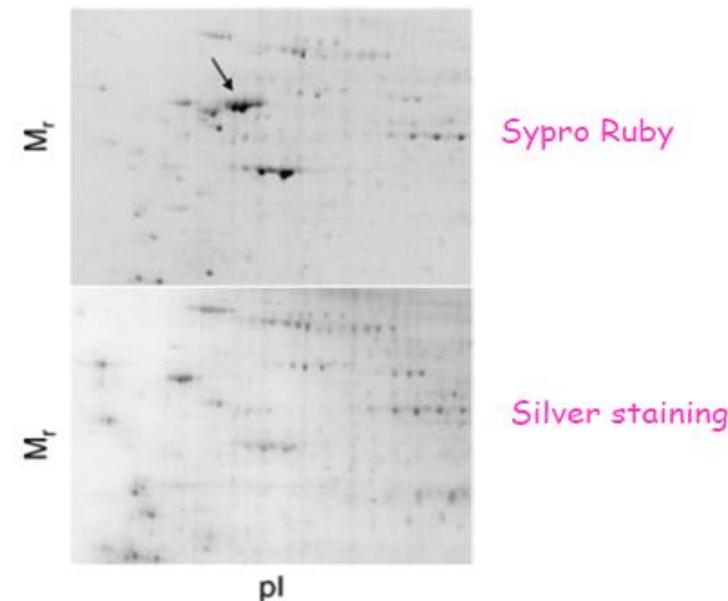
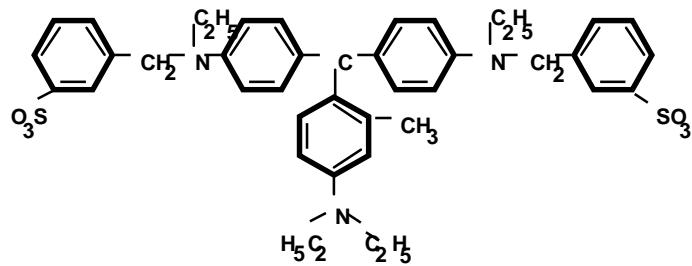
# Gel Staining

When the gel electrophoresis has completed its run the gel is stained



# Staining

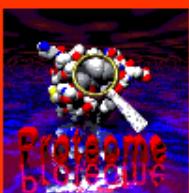
- Coomassie Blue R-250 (40 ng protein)
- Silver Stain Plus (1 ng protein)
- Fluorescent molecules (Sypro RubyStain (1 ng protein)



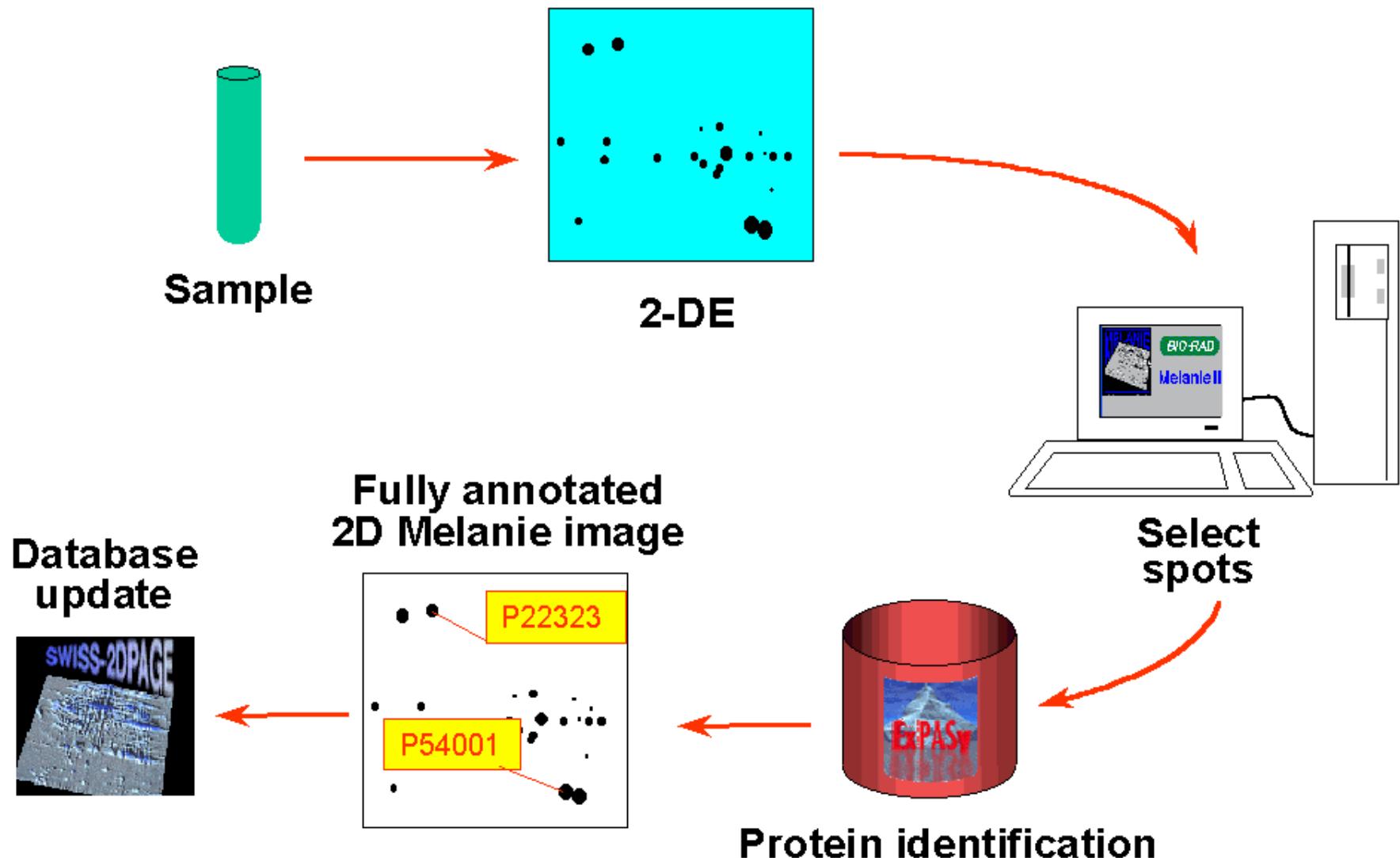
# Gel Stains - Summary

<b>Stain</b>	<b>Sensitivity (ng/spot)</b>	<b>Advantages</b>
Coomassie R-250	50-100	Simple, fast, consistent, cheap
Colloidal Coomassie	5-10	Simple, fast *
Silver stain	1-4	Very sensitive, awkward *
Copper stain	5-15	Reversible, 1 reagent negative stain
Zinc stain	5-15	Reversible, simple, fast high contrast neg. stain
SYPRO ruby	1-10	Very sensitive, fluorescent *

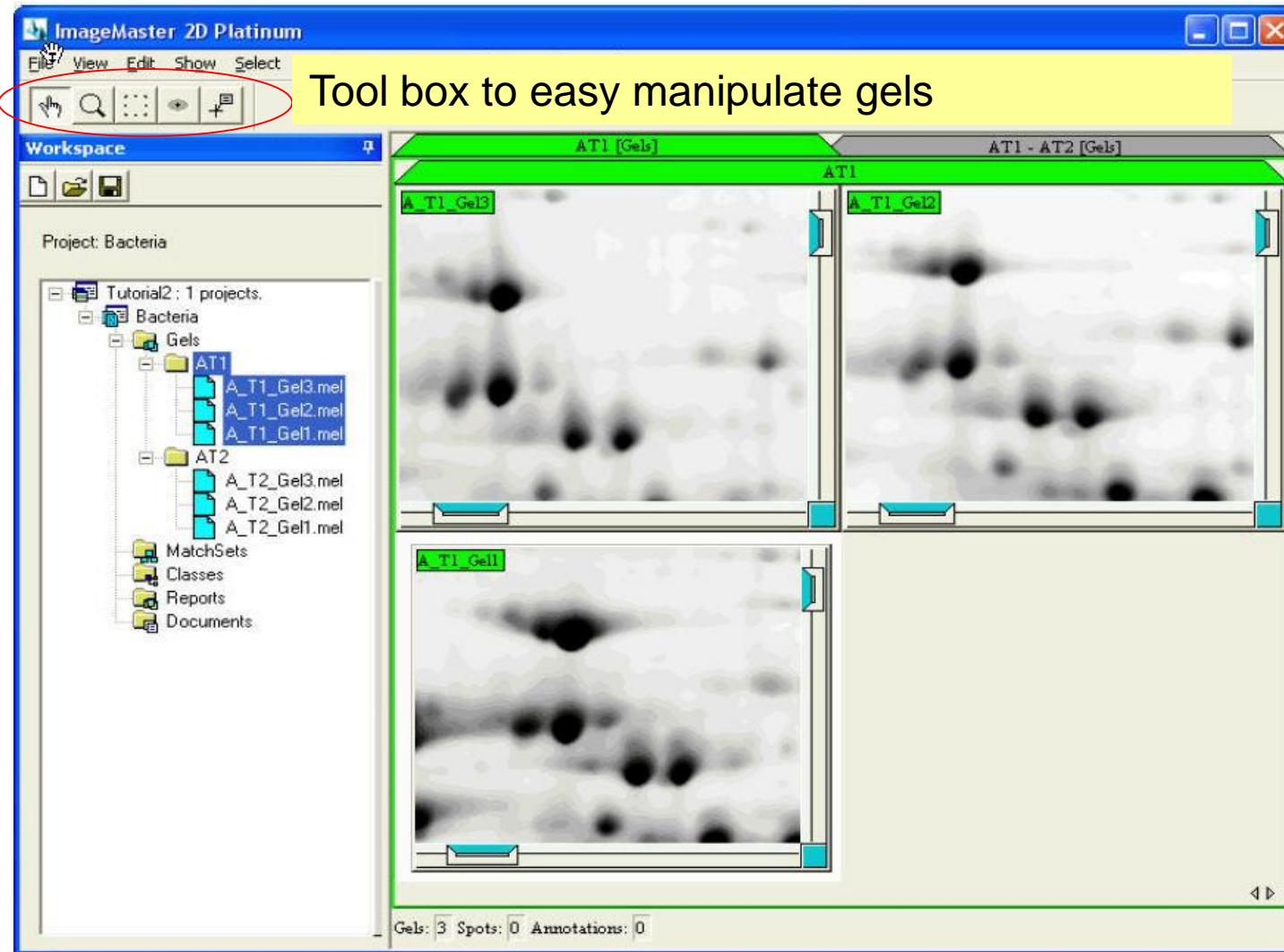
\* mass spectrometry compatible



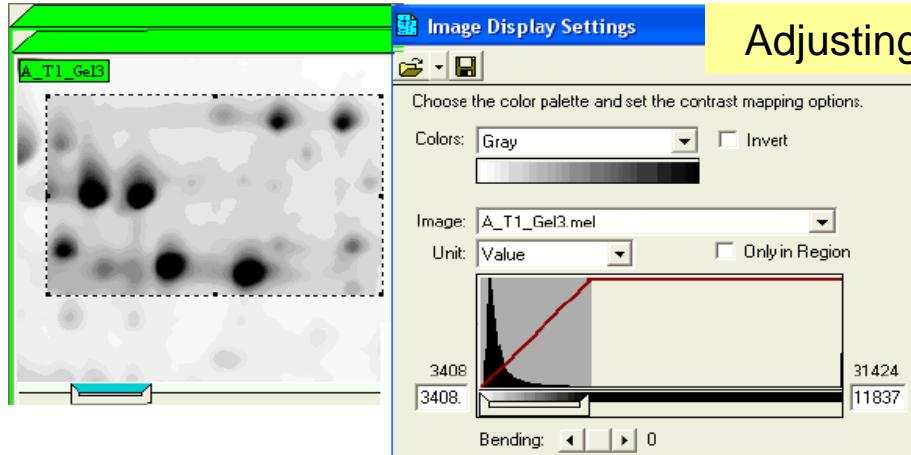
# Proteome data analysis



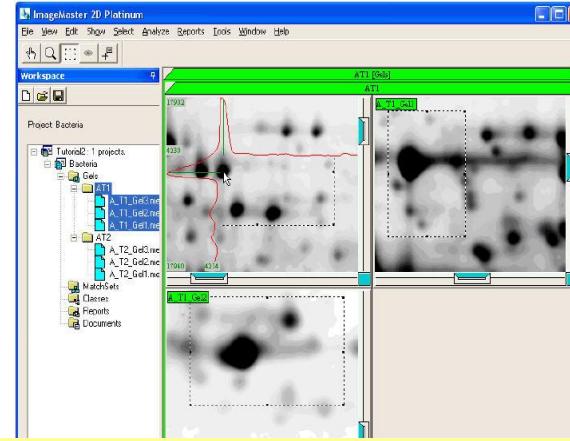
# Import gels



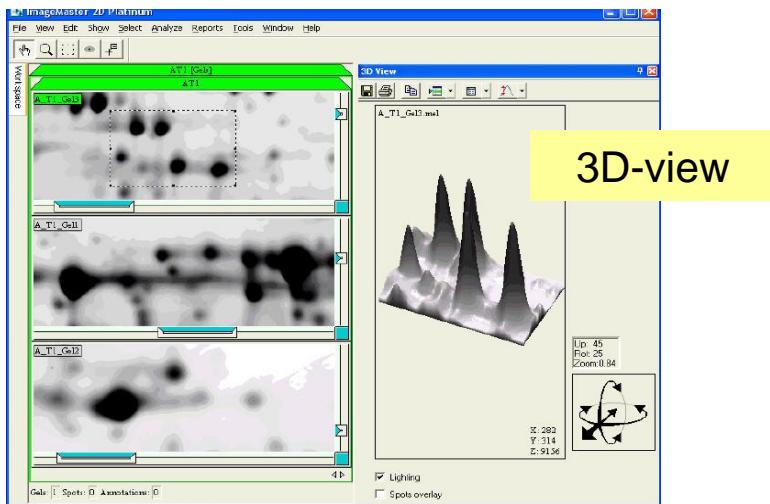
# Viewing and manipulating images



Adjusting contrast



Intensity variations in x- and y-direction



3D-view



Automatically subtracted background

# Spot detection

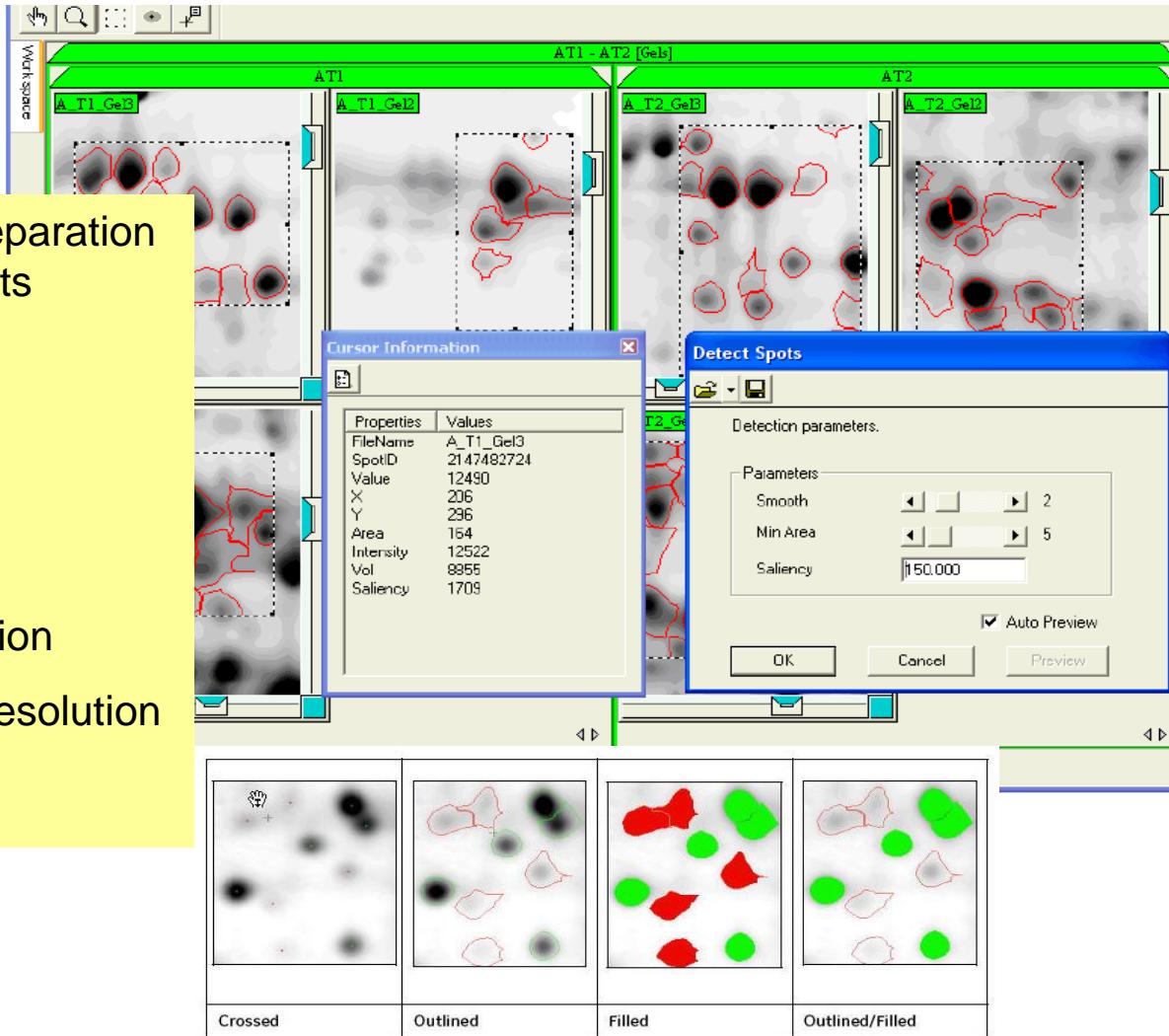
Adjust the separation between spots

Split overlap

Eliminate art affects/noise

Stain saturation

Incomplete resolution



# Spots report

Spot Report

The screenshot shows a software window titled "Spot Report". The window has a toolbar at the top with various icons. Below the toolbar is a status bar with the text "Information on selected" and a button labeled "Select on Gels". The main area is a table with the following data:

	FileName	SpotID	X	Y	Intensity ▲	Area	Vol	%Intensity	%Vol	Saliency
3	A_T2_Gel2	1477	1034	834	20942.2	8.06000	82534.8	1.04859	1.51769	2727.16
4	A_T2_Gel2	1363	1091	578	20823.0	22.0000	194809	1.04262	3.58226	6138.28
5	A_T2_Gel3	1380	1013	801	20793.0	43.4700	444740	1.22887	9.32264	4275.67
6	A_T2_Gel1	1776	676	567	20784.0	26.1000	242860	0.767693	3.43156	4634.07
7	A_T2_Gel1	1626	630	392	20775.0	18.6100	142491	0.767360	2.01337	5003.70
8	A_T2_Gel1	1795	1090	582	20665.5	19.2400	178681	0.763316	2.52471	4537.42

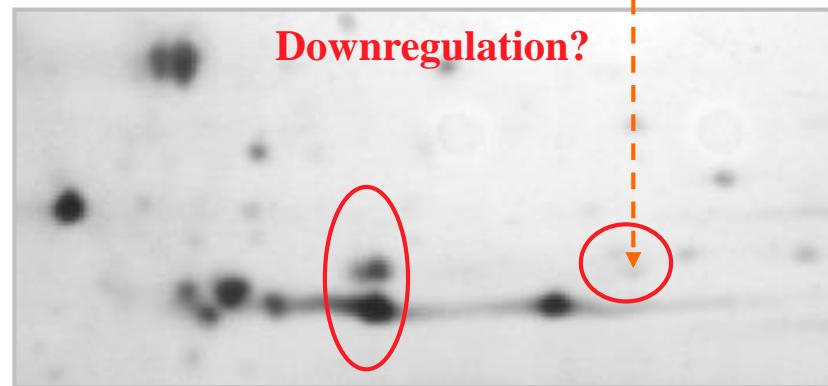
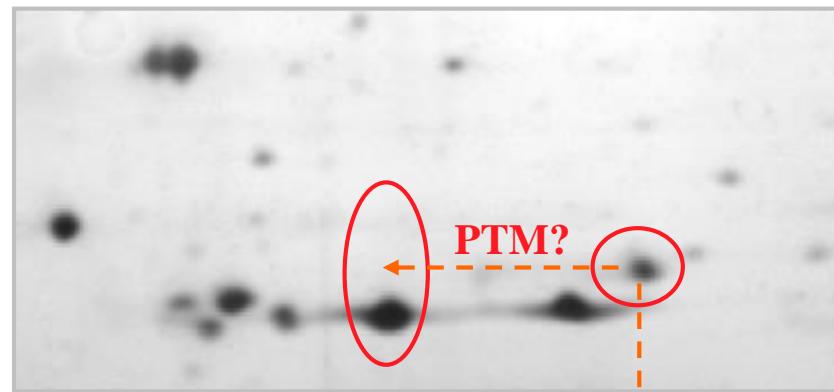
Spot Report A spot report summarizes the information about the selected spots

# Detection/matching

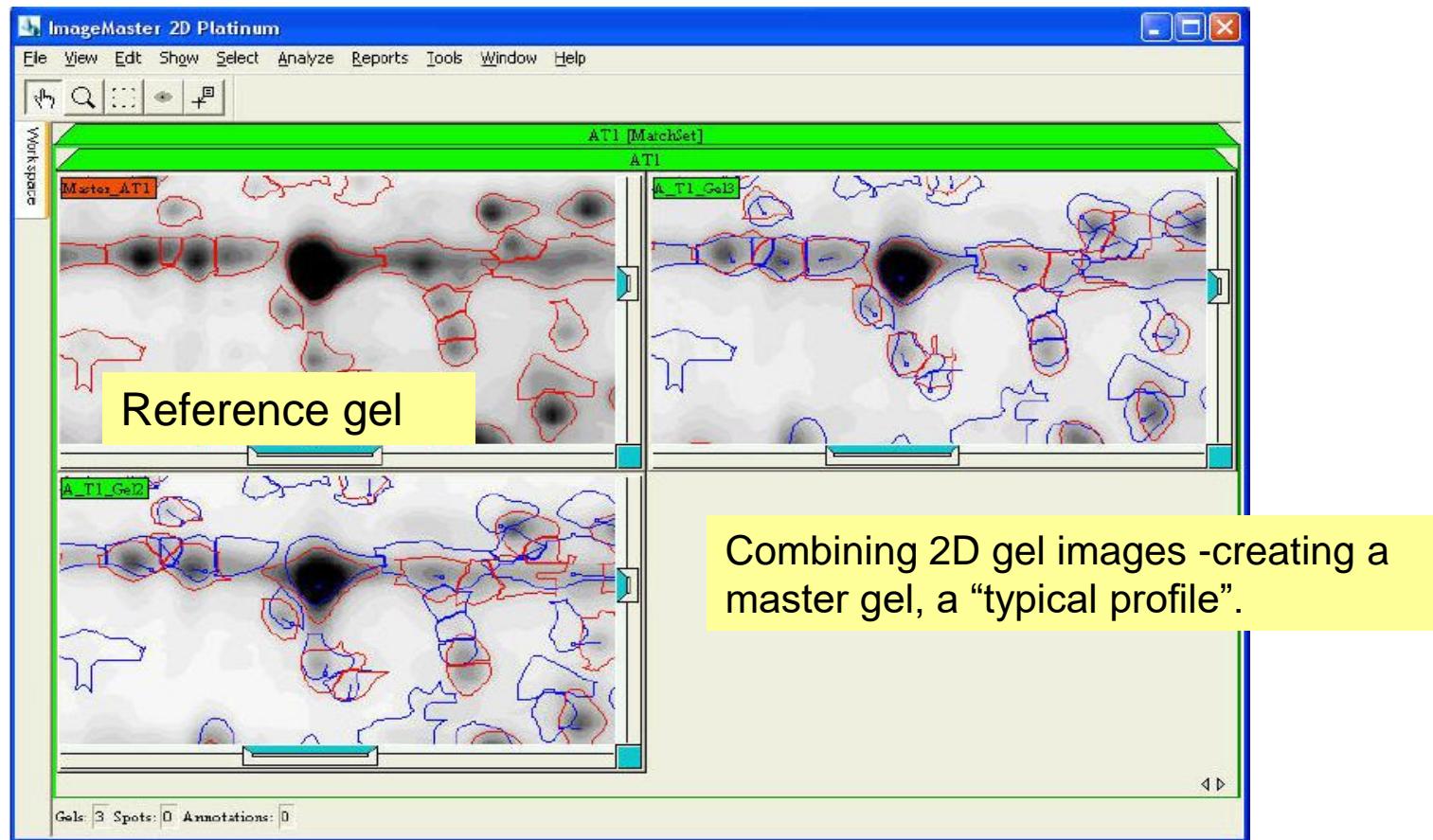
Spot detection

Spot matching

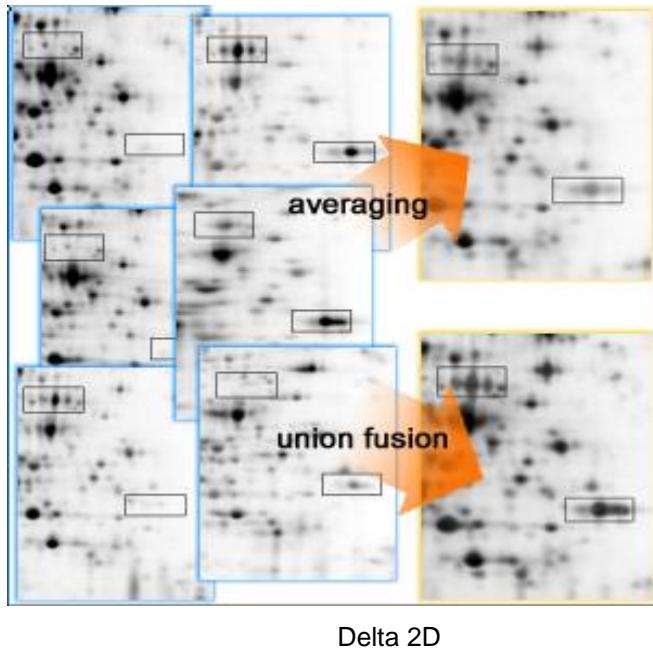
Normalization of spot intensities



# Matching

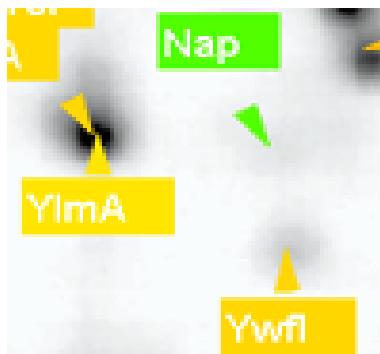


# Master gels



Combine several images, creating the master image

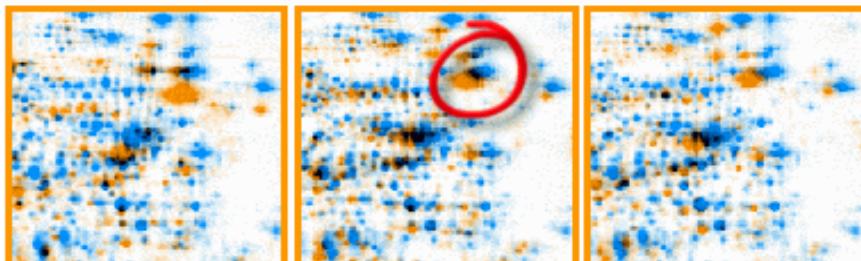
- all the spots on a single image
  - even those that will never be expressed at the same time,
- a summary of groups of replicate gels (average gel)



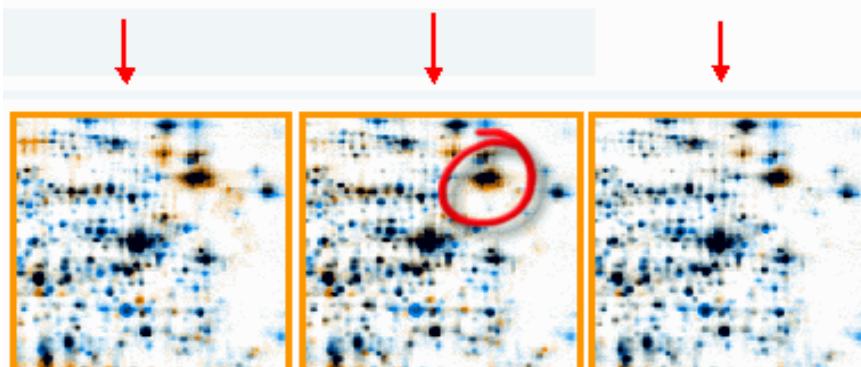
Any point on a gel can be labeled, and automatically transferred from one gel to another.

# Gel image warping

Variations in migration, protein separation, stain artifacts and stain saturation complicate gel matching and quantitation.

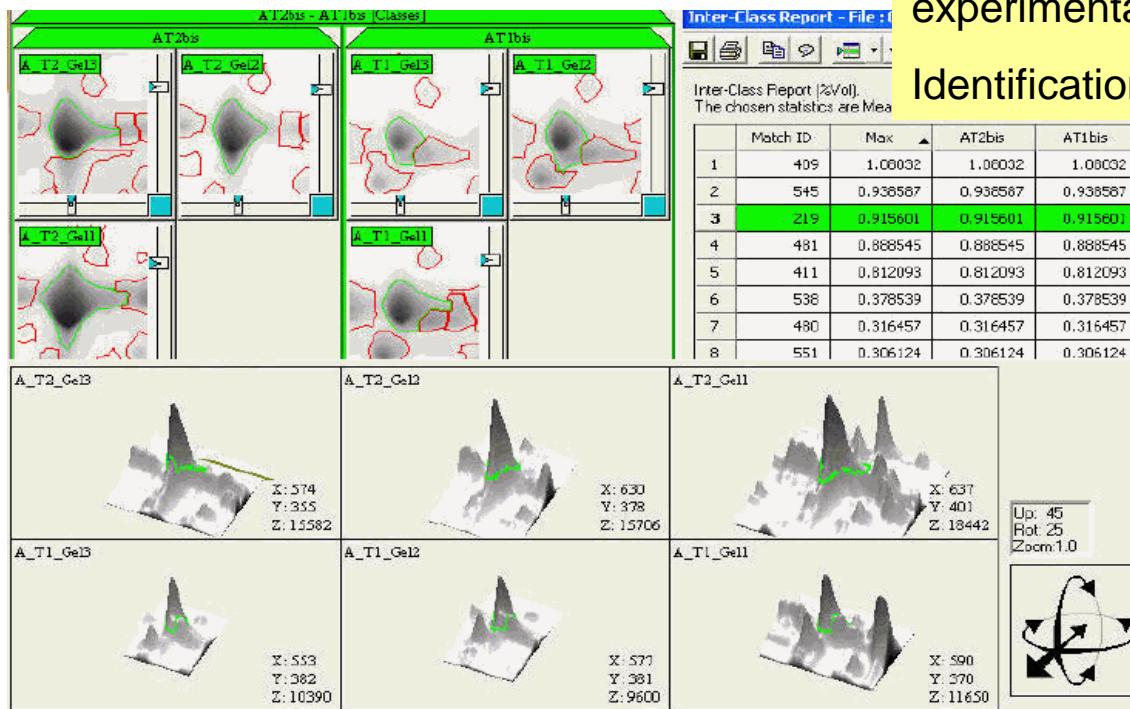


Compensates for running differences between gels

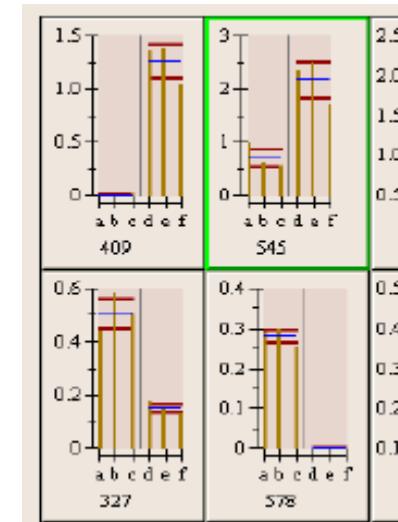


After warping, corresponding spots will have the same position on every image.

# Expression



Comparison of individual experimental gels to master gels.  
Identification of variant spots



# 2D Gel Databases

Swiss-2DPAGE [www.expasy.ch](http://www.expasy.ch)

GelBank <http://www.gelscape.ualberta.ca:8080/htm/gdbIndex.html>

Cornea 2D-PAGE <http://www.cornea-proteomics.com/>

World 2DPAGE, Index of 2D gel databases

<http://ca.expasy.org/ch2d/2d-index.html>

# DATABASE SEARCHING

A screenshot of a web browser window. The address bar shows "Indirizzo" and "http://www.expasy.org/ch2d/". Below the address bar is a navigation menu with links for "ExPASy Home page", "Site Map", "Search ExPASy", and "Contact us". A search bar contains the text "Search SWISS-2DPAGE" followed by a dropdown menu set to "for APOA1\_HUMAN" and buttons for "Go" and "Clear".



## SWISS-2DPAGE

### Two-dimensional polyacrylamide gel electrophoresis database

**SWISS-2DPAGE** contains data on proteins identified on various 2-D PAGE and SDS-PAGE reference maps. You can locate these proteins on the 2-D PAGE maps or display the region of a 2-D PAGE map where one might expect to find a protein from UniProtKB/Swiss-Prot [More details / References / Linking to SWISS-2DPAGE / Commercial users / Disclaimer].

Release 18.0, September 2006 and updates up to 02-Oct-2006 (contains 1265 entries in 36 reference maps from human, mouse, *Arabidopsis thaliana*, *Dictyostelium discoideum*, *Escherichia coli*, *Saccharomyces cerevisiae*, and *Staphylococcus aureus* (N315)).

[[Search](#)] [[Documents](#)] [[Services](#)] [[Software](#)] [[Related servers](#)] [[Other databases](#)] [[Job openings](#)]

#### Access to SWISS-2DPAGE

- by description (any word in the DE, OS, GN and ID lines)
- by accession number (AC lines)

#### SWISS-2DPAGE documents

- User manual

Release notes (September 26, 2006) 

# DATABASE SEARCHING

**SWISS-2DPAGE:** APOA1\_HUMAN

## APOA1\_HUMAN

### General information about the entry

[View entry in simple text format](#)

Entry name	<b>APOA1_HUMAN</b>
Primary accession number	<b>P02647</b>
integrated into SWISS-2DPAGE on	August 1, 1993 (release 0)
2D Annotations were last modified on	March 31, 2004 (version 2)
General Annotations were last modified on	September 26, 2006 (version 14)

### Name and origin of the protein

Description	<b>Apolipoprotein A-I (Apo-AI) (ApoA-I) [Contains: Apolipoprotein A-I(1-242)].</b>
Gene name	<b>Name=APOA1</b>
Annotated species	Homo sapiens (Human) [TaxID: <a href="#">9606</a> ]
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

### References

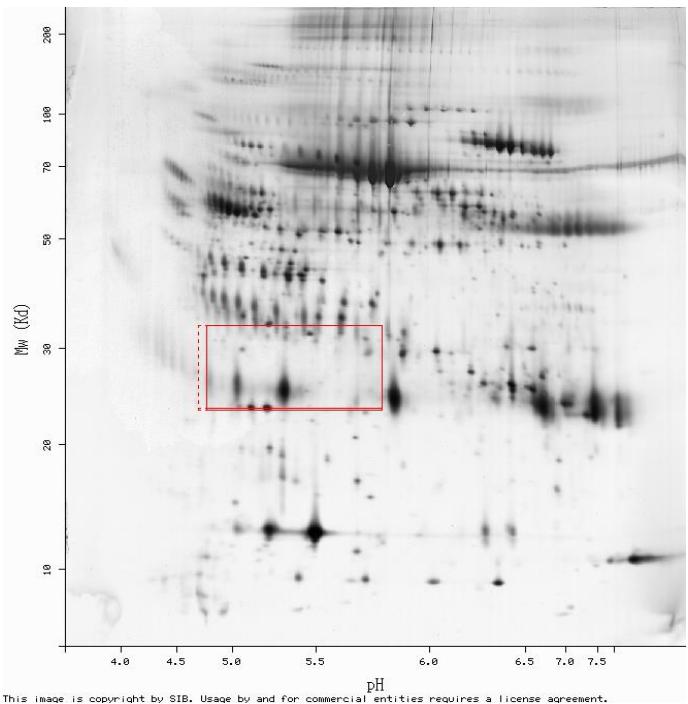
- [1] MAPPING ON GEL  
MEDLINE=93162045; PubMed=1286669; [[NCBI](#), [ExPASy](#), [EBI](#), [Israel](#), [Japan](#)]  
Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F., Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R., Appel R.D., Hughes G.J.  
"Human liver protein map: a reference database established by microsequencing and gel comparison"  
*Electrophoresis* 13:992-1001(1992)
- [2] MAPPING ON GEL  
MEDLINE=96007934; PubMed=7498157; [[NCBI](#), [ExPASy](#), [EBI](#), [Israel](#), [Japan](#)]  
Sanchez J.-C., Appel R.D., Golaz O.G., Pasquali C., Ravier F., Bairoch A., Hochstrasser D.F.  
"Inside SWISS-2DPAGE database"  
*Electrophoresis* 16:1131-1151(1995)

# DATABASE SEARCHING

## APOA1\_HUMAN

### Gel experimental information

Information Type	Data
Map name	CSF_HUMAN (Cerebrospinal Fluid)
Dimension	2-D
Related documents	<a href="#">Gel Preparation</a> / <a href="#">Gel Informatics</a>
species	<i>Homo sapiens (Human)</i> / TaxID:9606
Tissue (Swiss-Prot definition)	Cerebrospinal fluid
pI	start: 3.50 - end: 10.00
Mw	start: 10000 - end: 200000
Number of detected spots	1664
Number of identified spots	309
Number of identified proteins	30
Mapping Methods	{Gm} Gel matching
Reference	<a href="#">Bibliographic reference</a>
Protein list	<a href="#">Display list</a>
Graphical interface	



# SWISS 2D PAGE

Search on  
accession  
numbers,  
separation  
techniques,  
author....

## Access to SWISS-2DPAGE

- by description (any word in the DE, OS, GN and ID lines)
- by accession number (AC lines)
- by clicking on a spot: select one of our 2-D PAGE or SDS-PAGE reference maps, click on a spot and then get the corresponding information from the SWISS-2DPAGE database.
- by author (RA lines)
- by spot serial number (2D and 1D lines)
- by experimental pl/Mw range
- by experimental identification methods
- by full text search
- retrieve all the protein entries identified on a given reference map
- complex queries (SRS like)
- compute estimated location on reference maps for a user-entered sequence

# Swiss 2D PAGE viewer

The screenshot shows the Swiss-2DPAGE Viewer interface. At the top, there is a navigation bar with links to "ExPASy Home page", "Site Map", "Search ExPASy", "Contact us", and "SWISS-2DPAGE". Below the navigation bar is a search bar with the placeholder "Search SWISS-2DPAGE for" and "Go" and "Clear" buttons. To the right of the search bar is a link "Back to the search engine". On the left side of the main content area, there are two dropdown menus. The first dropdown under "Switch to Gel:" has options like LYMPHOCYTE\_HUMAN, LYMPHOMA\_HUMAN, NUCLEI\_LIVER\_HUMAN, NUCLEOLI\_HELA\_1D\_HUMAN, NUCLEOLI\_HELA\_2D\_HUMAN, PLASMA\_HUMAN, PLATELET\_HUMAN, RBC\_HUMAN, and U937\_HUMAN. The second dropdown under "View:" has options like all identities, protein, peptide, and sequence. A mouse cursor is hovering over the "LIVER\_MOUSE" option in the "Switch to Gel:" dropdown. To the right of the interface, a text box contains the question "Which gel you want to look at".

swiss 2Dpage SWISS-2DPAGE Viewer

Back to the  
search engine

Switch to Gel:

View: all identities

---

- LYMPHOCYTE\_HUMAN { Lymphocytes }
- LYMPHOMA\_HUMAN { Lymphoma }
- NUCLEI\_LIVER\_HUMAN { Soluble nuclear proteins and matrix from liver tissue }
- NUCLEOLI\_HELA\_1D\_HUMAN { SDS-PAGE of nucleolar proteins from Human HeLa cells }
- NUCLEOLI\_HELA\_2D\_HUMAN { 2D-PAGE of nucleolar proteins from Human HeLa cells }
- PLASMA\_HUMAN { Plasma }
- PLATELET\_HUMAN { Platelet }
- RBC\_HUMAN { Red blood cells }
- U937\_HUMAN { Macrophage Like Cell Line }

**Mus musculus (Mouse)**

- BAT\_MOUSE { Brown adipose tissue }
- ISLETS\_MOUSE { Pancreatic islet cells }
- LIVER\_MOUSE { Liver } (Mouse)
- MUSCLE\_MOUSE { Gastrocnemius muscle }
- NUCLEI\_LIVER\_MOUSE { Soluble nuclear proteins and matrix from liver tissue }
- WAT\_MOUSE { White adipose tissue }

**Saccharomyces cerevisiae (Baker's yeast)**

- YEAST { Saccharomyces cerevisiae }

**Staphylococcus aureus (strain N315)**

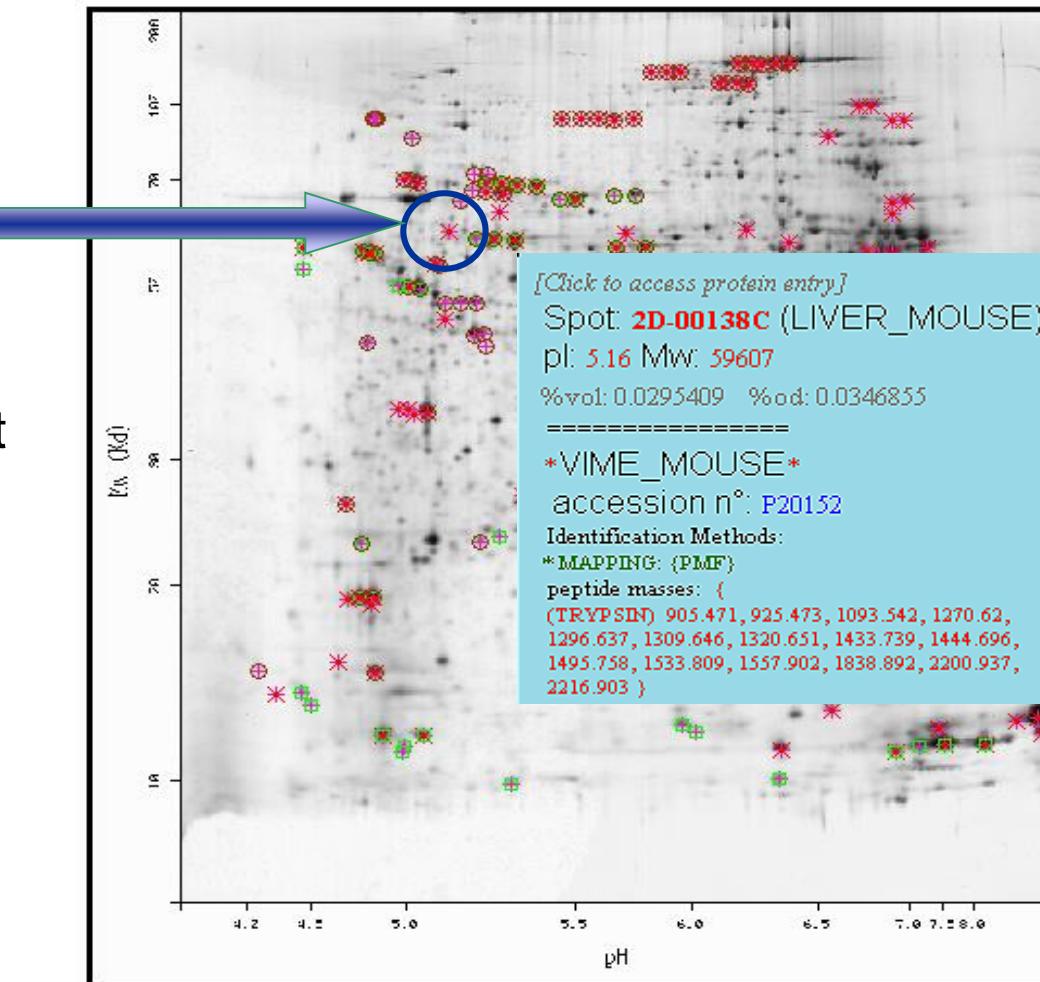
- STAPHYLOCOCCUS { Staphylococcus aureus }

Which gel you want to look at

# Swiss 2D PAGE



Point at a spot,  
and you'll  
automatically get  
accession  
number  
(P20152), id-  
method..



# Swiss-2D PAGE

This is the protein you've chosen  
Scroll down, and you get gel images of where this protein has been identified (bat-mouse, mouse muscle..)

## Search by

- [\[accession number\]](#)
- [\[description, ID or gene\]](#)
- [\[author's name\]](#)
- [\[spot ID / serial number\] ►](#)
- [\[identification methods\]](#)
- [\[pI / Mw range\]](#)
- [\[combined fields\]](#)

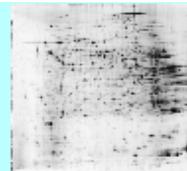
## Maps

- [\[experimental info\]](#)
- [\[protein list\]](#)
- [\[graphical interface\]](#)

### Query Remote Interfaces:

[World-2DPAGE](#)

Exclude local DBs  
*has only effect if a remote interface is selected*



map experimental info  
protein estimated location

Searching in 'SWISS-2DPAGE' for entry matching: **P20152**

**SWISS-2DPAGE:** P20152

# P20152

## General information about the entry

[View entry in simple text format](#)

Entry name	<b>VIME_MOUSE</b>
Primary accession number	<b>P20152</b>
integrated into SWISS-2DPAGE on	September 1, 1998 (release 7)
2D Annotations were last modified on	October 1, 2001 (version 4)
General Annotations were last modified on	September 26, 2006 (version 6)

## Name and origin of the protein

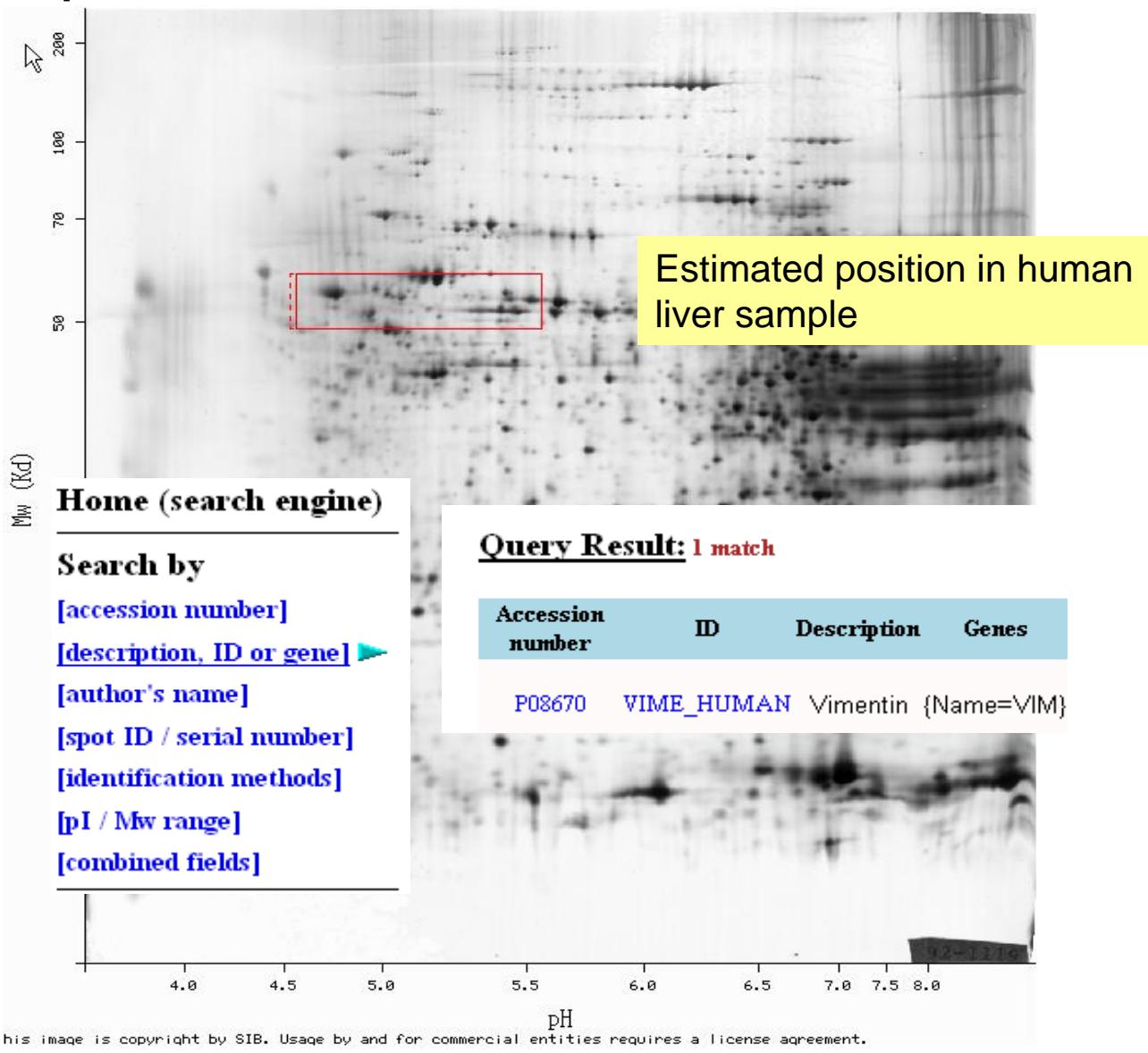
Description	<b>Vimentin.</b>
Gene name	<b>Name=Vim</b>
Annotated species	Mus musculus (Mouse) [TaxID: <a href="#">10090</a> ]
Taxonomy	Eukaryota; Metazoa; Chordata; Crania; Euteleostomi; Mammalia; Eutheria; Eu; Rodentia; Sciurognathi; Muroidea; Mu;

## References

- (the  
searched  
spot) [identification data]
- SPOT 2D-00138C: pl=5.16;  
Mw=59607
  - SPOT 2D-0013E4: pl=5.13; Mw=53059

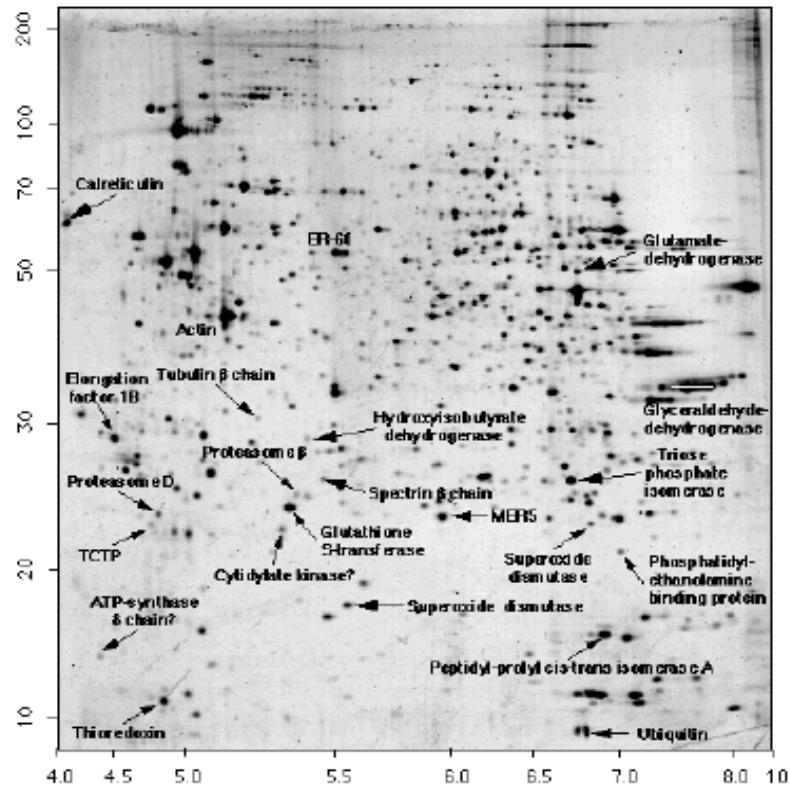
**MAPPING (identification):**

# Estimated position

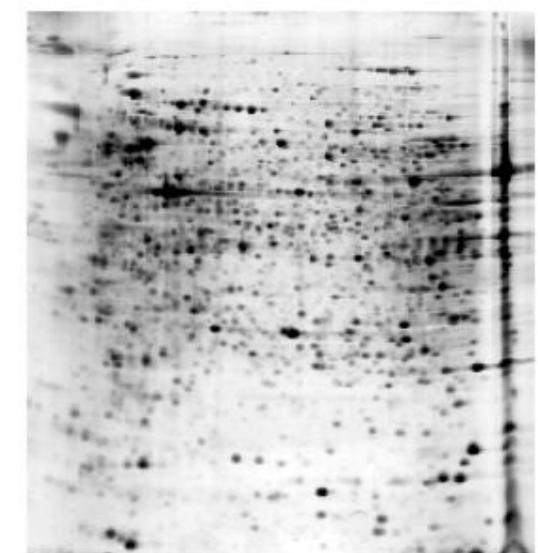


# Match with databases

# Differential analysis



Normal



Pathologic

# Limitations/difficulties with the 2D gel

## Reproducibility

Samples must be run at least in triplicate to rule out effects from gel-to-gel variation (statistics)

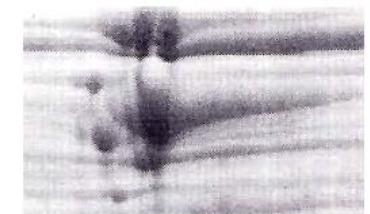
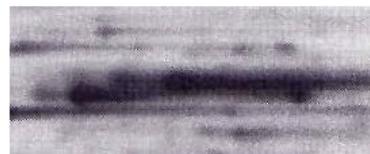
**Small dynamic range** of protein staining as a detection technique- visualization of abundant proteins while less abundant might be missed.

**Co-migrating** spots forming a complex region



**Streaking and smearing**

**Weak spots and background**

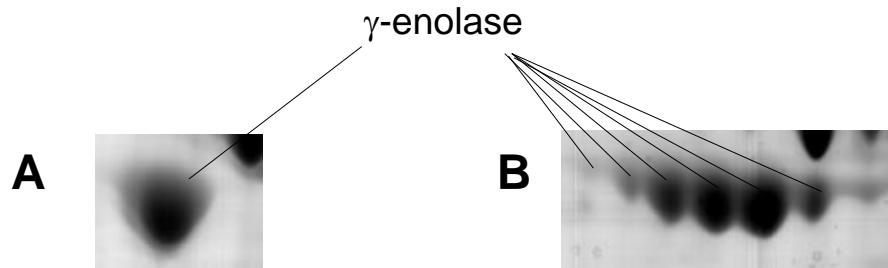


Posttranscriptional control mechanisms

Incompatibility of some proteins with the first dimension IEF step (**hydrophobic proteins**)

Marginal solubility leads to protein precipitation and degradation- smearing (Glycolysation, oxidation)

# Protein Heterogeneity



Partial 2D-gel images showing  $\gamma$ -enolase from human brain. The protein is represented by one spot when IEF was performed on pH 3-10 non-linear IPG strips (A), and by six spots when IEF was performed on pH 4-7 strips (B).

Increased Resolution and Detection of  
More Spots with the Use of Narrow pH  
Gradient Strips

# DIGE

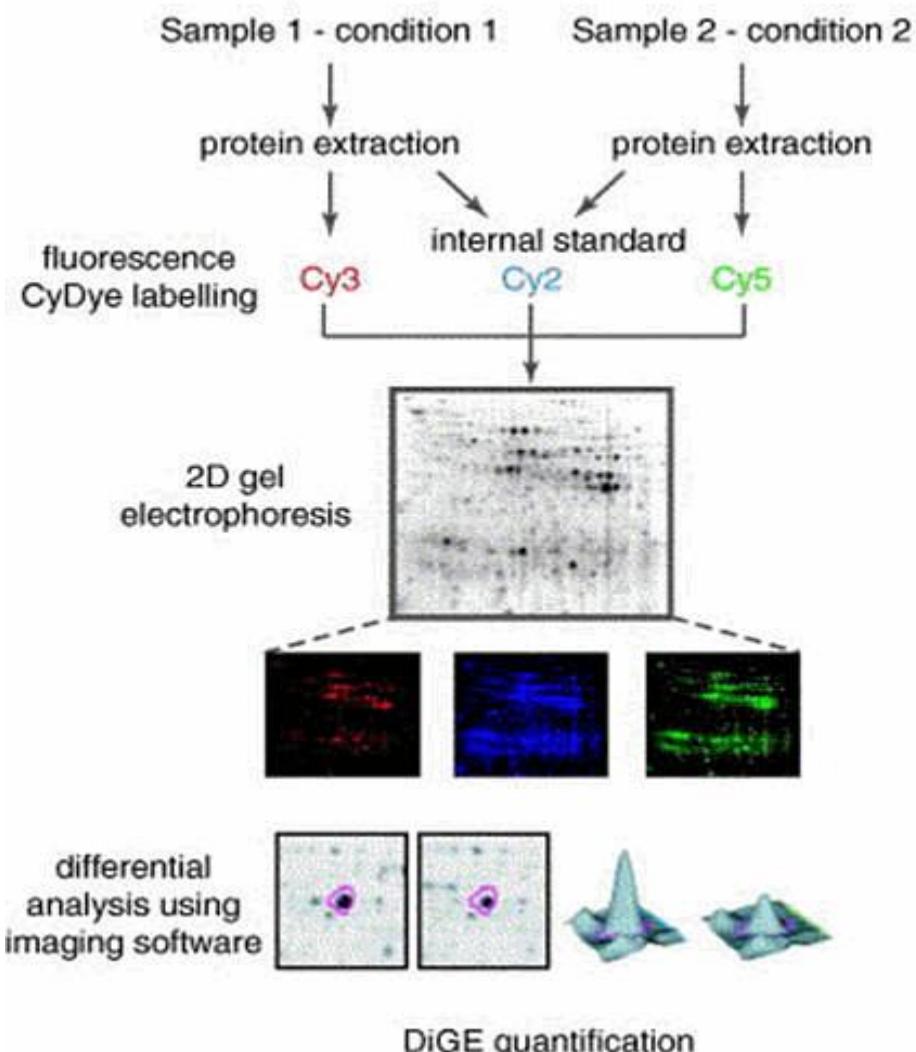
## 2D Fluorescence Difference Gel Electrophoresis

### Quantification of Spot Relative Levels

Proteins are labeled prior to running the first dimension with up to three different fluorescent cyanide dyes

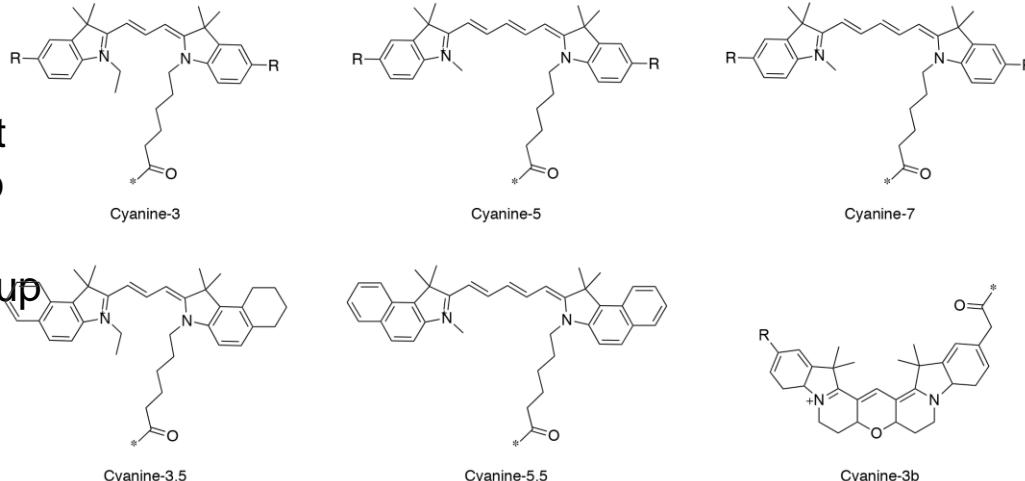
Allows use of an internal standard in each gel-to-gel variation, reduces the number of gels to be run

Adds 500 Da to the protein labeled



# DIGE

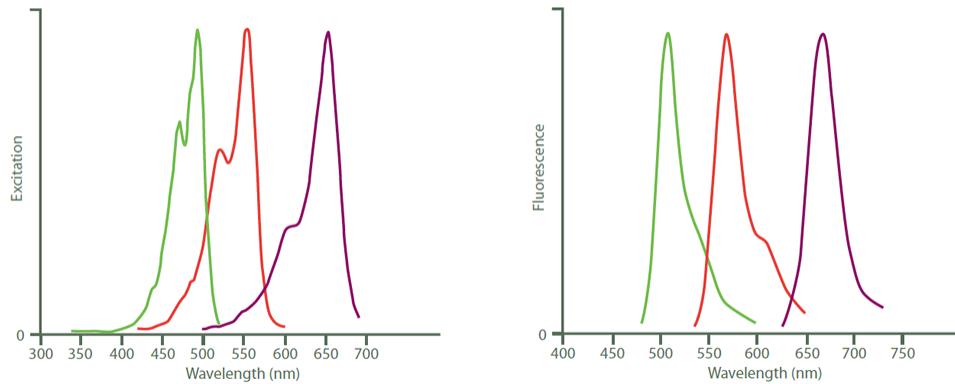
DIGE uses direct labeling of proteins with fluorescent dyes (known as CyDyes: Cy2, Cy3, and Cy5) prior to IEF. CyDyes are spectrally resolvable cyanine dyes carrying an N-hydroxysuccinimidyl ester reactive group that covalently binds the  $\epsilon$ -amino groups of lysine residues in proteins.



Dye concentrations are kept low, such that approximately one dye molecule is added per protein. The important aspect of the DIGE technology is its ability to label two or more samples with different dyes and separate them on the same gel, eliminating gel-to-gel variability.

This makes spot matching and quantitation much simpler and more accurate. Cy2 is used for a normalization pool created from a mixture of all samples in the experiment...

This Cy2-labeled pool is run on all gels, allowing spot matching and normalization of signals from different gels. The Cy2-labeled pool used in 2D-DIGE is useful because it provides a consistent spot map on all gels in an experiment, facilitating spot matching.



# Pubmed compound

<https://pubchem.ncbi.nlm.nih.gov/compound/4414>

NIH U.S. National Library of Medicine National Center for Biotechnology Information

**PubChem OPEN CHEMISTRY DATABASE**

Compound Summary for CID 44140555

STRUCTURE PATENTS

Cy3B dye N-hydroxysuccinimidyl ester

Cite this Record

**PubChem CID:** 44140555

**Chemical Names:** Cy3B dye NHS ester; SCHEMBL1312591; SCHEMBL18727749; CHEBI:52004; Cy3B dye N-hydroxysuccinimidyl ester; 14-{2-[(2,5-dioxopyrrolidin-1-yl)oxy]-2-oxoethyl}-16,16,18,18-tetramethyl-6,7,7a,8a,9,10,16,18-octahydrobenzo[2'',3'']indolizino[8'',7'',5',6']pyrano[3',2':3,4]pyrido[1,2-a]indol-5-ium-2-sulfonate More...

**Molecular Formula:** C<sub>35</sub>H<sub>35</sub>N<sub>3</sub>O<sub>8</sub>S

**Molecular Weight:** 657.738 g/mol

**InChI Key:** PLHHGVSUNRYQL-UHFFFAOYSA-N

Cy3B dye NHS ester is a fluorescent dye with a maximum emission wavelength of 572 nm, derived from a heteroheptacyclic ring system. It has a role as a fluorochrome. It is an iminium betaine, an



## Contents



- 1 2D Structure
- 2 3D Conformer
- 3 Names and Identifiers
  - 3.1 Computed Descriptors
    - 3.1.1 InChI
    - 3.1.2 InChI Key
    - 3.1.3 Canonical SMILES
  - 3.2 Molecular Formula
  - 3.3 Other Identifiers
    - 3.3.1 Wikipedia
- 3.4 Synonyms
  - 3.4.1 Depositor-Supplied Synonyms
    - 1. [Cy3B dye NHS ester](#)
    - 2. [SCHEMBL1312591](#)
    - 3. [SCHEMBL18727749](#)
    - 4. [CHEBI:52004](#)
    - 5. [Cy3B dye N-hydroxysuccinimidyl ester](#)
    - 6. [14-{2-\[\(2,5-dioxopyrrolidin-1-yl\)oxy\]-2-oxoethyl}-16,16,18,18-tetramethyl-6,7,7a,8a,9,10,16,18-octahydrobenzo\[2",3"\]indolizino\[8",7":5',6'\]pyrano\[3':2':3,4\]pyrido\[1,2-a\]indol-5-iium-2-sulfonate](#)
    - 7. [2-{2-\[\(2,5-dioxopyrrolidin-1-yl\)oxy\]-2-oxoethyl}-16,16,18,18-tetramethyl-6,7,7a,8a,9,10,16,18-octahydrobenzo\[2",3"\]indolizino\[8",7":5',6'\]pyrano\[3':2':3,4\]pyrido\[1,2-a\]indol-5-iium-14-sulfonate](#)
- 4 Chemical and Physical Properties
  - 4.1 Computed Properties
- 5 Related Records
  - 5.1 Related Compounds with Annotation

## 3.4 Synonyms



### 3.4.1 Depositor-Supplied Synonyms



1. [Cy3B dye NHS ester](#)
2. [SCHEMBL1312591](#)
3. [SCHEMBL18727749](#)
4. [CHEBI:52004](#)
5. [Cy3B dye N-hydroxysuccinimidyl ester](#)
6. [14-{2-\[\(2,5-dioxopyrrolidin-1-yl\)oxy\]-2-oxoethyl}-16,16,18,18-tetramethyl-6,7,7a,8a,9,10,16,18-octahydrobenzo\[2",3"\]indolizino\[8",7":5',6'\]pyrano\[3':2':3,4\]pyrido\[1,2-a\]indol-5-iium-2-sulfonate](#)
7. [2-{2-\[\(2,5-dioxopyrrolidin-1-yl\)oxy\]-2-oxoethyl}-16,16,18,18-tetramethyl-6,7,7a,8a,9,10,16,18-octahydrobenzo\[2",3"\]indolizino\[8",7":5',6'\]pyrano\[3':2':3,4\]pyrido\[1,2-a\]indol-5-iium-14-sulfonate](#)

▶ from PubChem

## 4 Chemical and Physical Properties

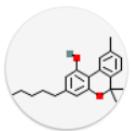


### 4.1 Computed Properties



# Cannabinol

 Cite this Record



STRUCTURE



VENDORS



PHARMACOLOGY



LITERATURE



PATENTS



BIOACTIVITIES

PubChem CID: 2543

Chemical Names: Cannabinol; 521-35-7; Cannabinolo [DCIT]; Cannabinolum [INN-Latin]; Cannabinol [INN:BAN]; UNII-7UYP6MC9GH More...

Molecular Formula: C21H26O2

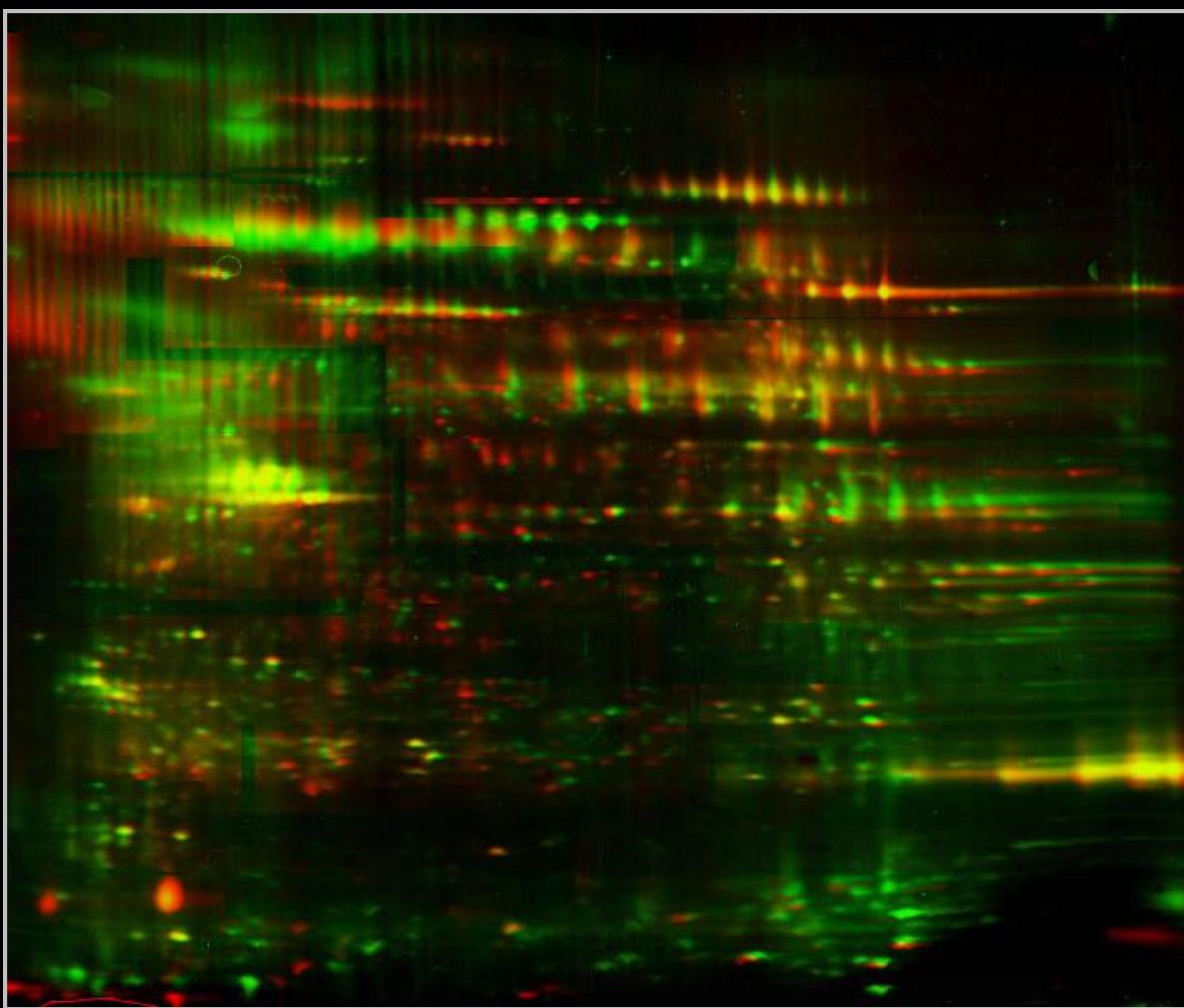
Molecular Weight: 310.437 g/mol

InChI Key: VBGLYOIFKLUMQG-UHFFFAOYSA-N

Substance Registry: FDA UNII

Pubmed compound,  
Choose a molecule and take a tour

# Comparison of “Secretoma” from Normal and Pharmacological Treatment of bronchial epithelial cells



Control [Cy5]

Pharmacological Treatment [Cy3]

# Proteomics: 2D analytical limits

- Solubility of target proteins (es. Membrane proteins)
- Presence of interfering substances (es. lipids, salts, nucleic acids)
- Concentration of proteins in the sample
- Biological value of densitometric variations and their statistic significance (relative quantification)

# Disadvantages of 2D electrophoresis

2D gels

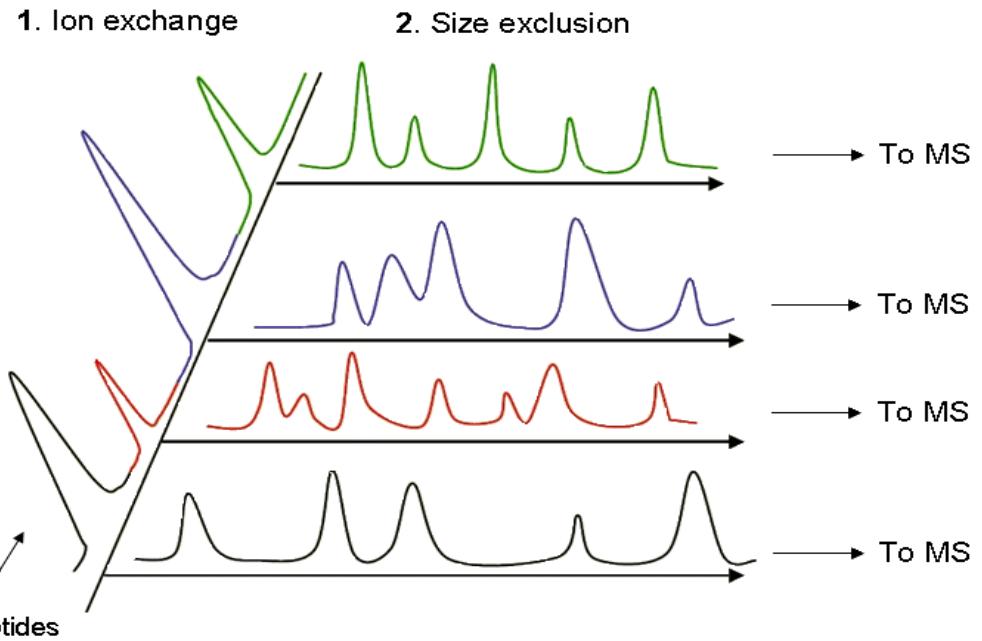
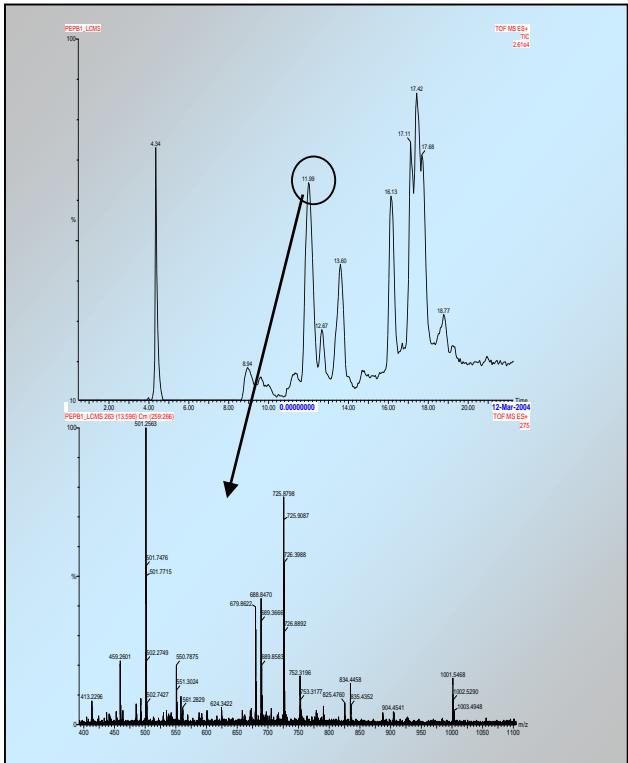


- ✓ Reproducibility Issues
- ✓ Difficult to automate
- ✓ Limited quantitation
- ✓ MW & Expression level
- ✓ Difficulties with low MW proteins (<15 kDa)
- ✓ Requires Mass Spectrometry for the identification of the digested proteins [Trypsin]

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

# Multidimensional LC:

First dimension: ionic exchange chromatography



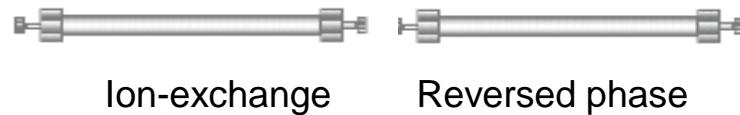
Each fraction containing peptides from ionic exchange column are then separated on a reverse phase column and then revealed with a detector ESI-Q/TOF

# Multidimensional HPLC

## Mud PIT

### Multidimensional Protein Identification Techniques or Tandem HPLC

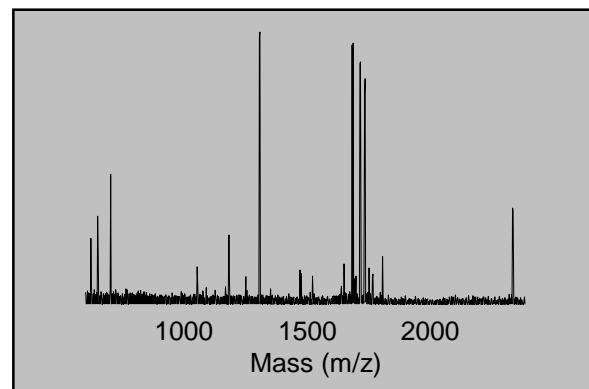
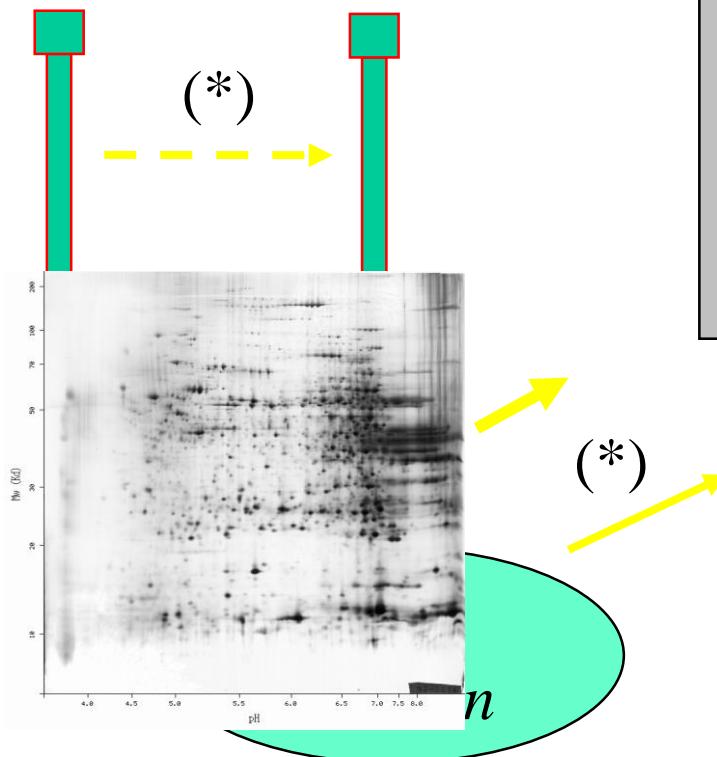
the combination of dissimilar separation modes will allow a greater resolution of peptides in mixture.



- **Reversed phase**, hydrophobicity
- **Ion exchange**, net positive/negative charge
- **Size exclusion**, peptide size, molecular weight
- **Affinity chromatography**, interaction with specific functional groups

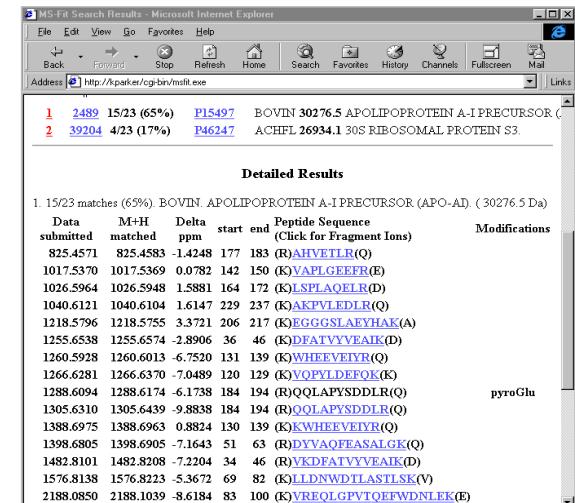
# Multidimensional chromatography: a valid alternative to 2D PAGE

Chromatographic separations



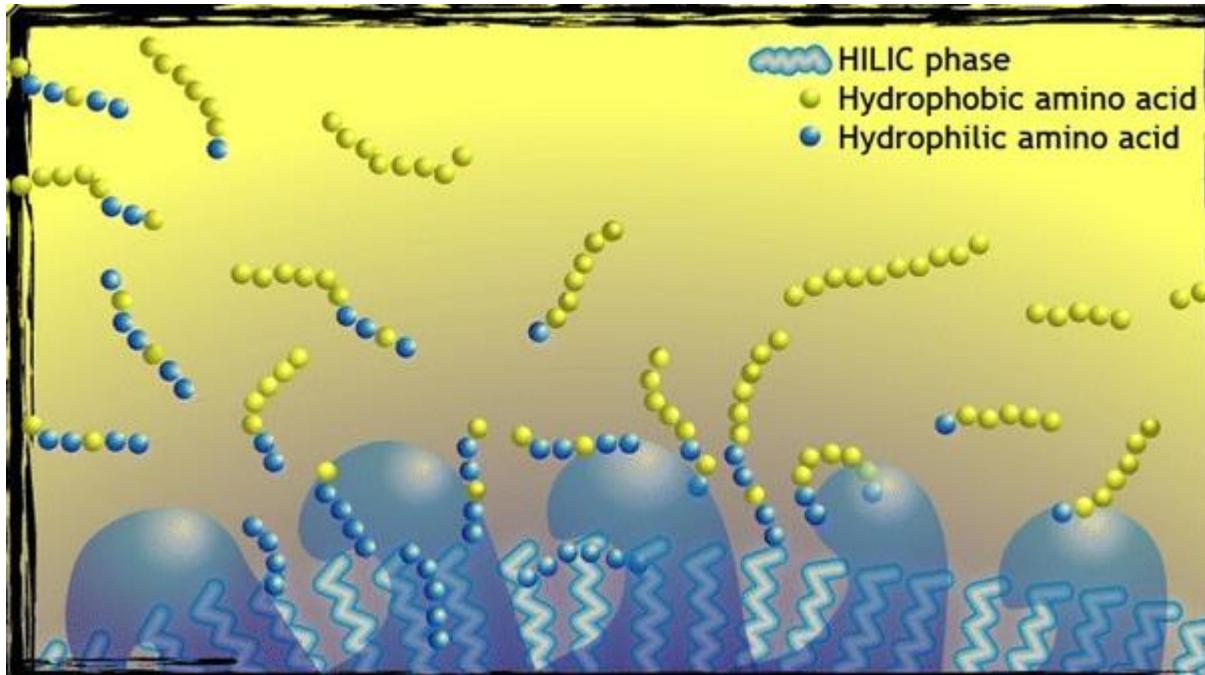
mass analysis

(\*)proteolytic digestion



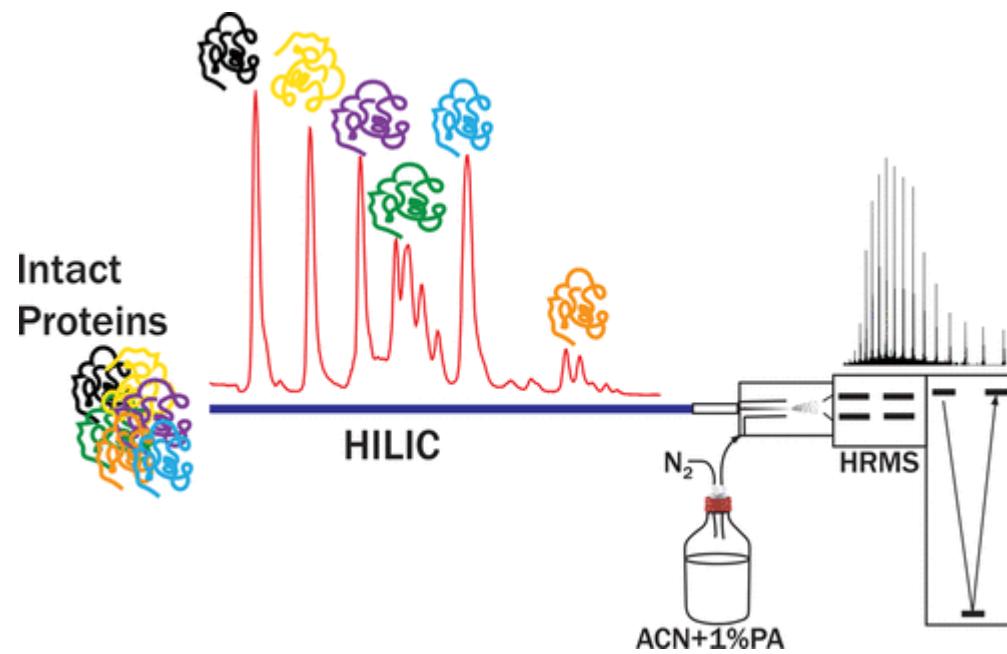
# Hydrophilic interaction liquid chromatography (HILIC)

HILIC has recently emerged as a popular chromatographic mode for the separation of hydrophilic analytes. HILIC operates on the basis of hydrophilic interactions between the analytes and the hydrophilic stationary phase, with either highly polar, or hydrophilic compounds interacting most strongly.



# HILIC

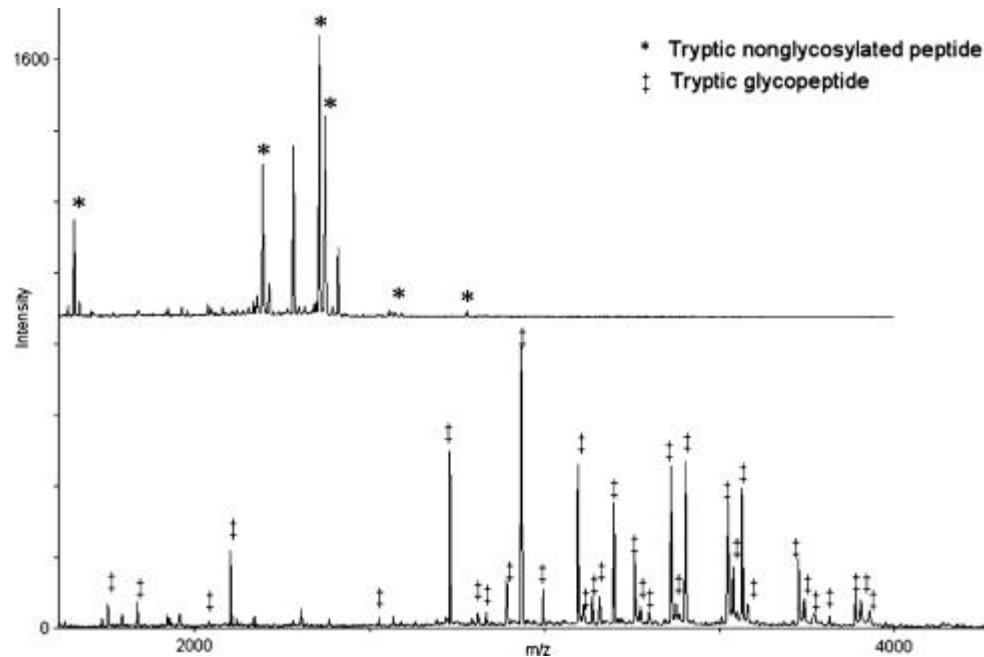
HILIC can simply be seen as a form of normal-phase (NP) chromatography. However, the acronym HILIC was suggested to distinguish it from NP, as NP is typically performed with nonaqueous, non-water-miscible solvent buffers, while HILIC is performed with water-miscible solvents and elution is achieved by a water gradient



*Capillary HILIC-MS: A New Tool for Sensitive Top-Down Proteomics.*

Gargano AFG, Roca LS, Fellers RT, Bocxe M, Domínguez-Vega E, Somsen GW. *Anal Chem.* 2018 Jun 5;90(11):6601-6609

# HILIC VS reverse Phase



MALDI-TOF spectra after SPE of a tryptic digest of TIMP-1 by RP (*upper spectrum*) and HILIC (*lower spectrum*) microcolumns. HILIC clearly enriches for glycopeptides.