

## Proteome

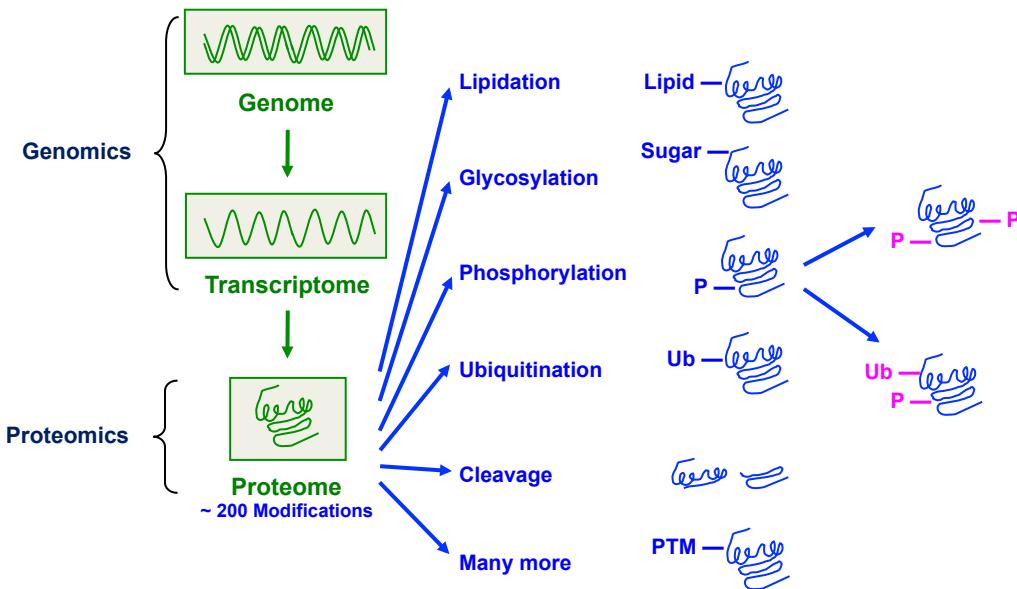
The **proteome** is the set of proteins expressed from a determined genome.

The **genome** is essentially constant through the organism.

The **proteome** is dynamic.

- ✓ In *E. coli* we find 1.3 proteins per gene.
- ✓ In yeast we find 3 proteins per gene.
- ✓ In Hepatocytes we find 5 to 10 proteins per gene.

## Proteomics

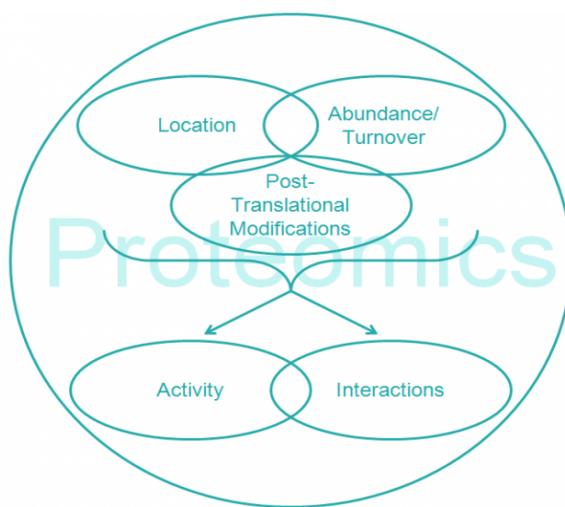


## Protein Chemistry vs Proteomics

### Differences Between Protein Chemistry and Proteomics

Protein chemistry	Proteomics
<ul style="list-style-type: none"> <li>• Individual proteins</li> <li>• Complete sequence analysis</li> <li>• Emphasis on structure and function</li> <li>• Structural biology</li> </ul>	<ul style="list-style-type: none"> <li>• Complex mixtures</li> <li>• Partial sequence analysis</li> <li>• Emphasis on identification by database matching</li> <li>• Systems biology</li> </ul>

## Areas of proteomics



Proteomic experiments generally collect data on three properties of proteins in a sample:  
**location, abundance/turnover and post-translational modifications.**

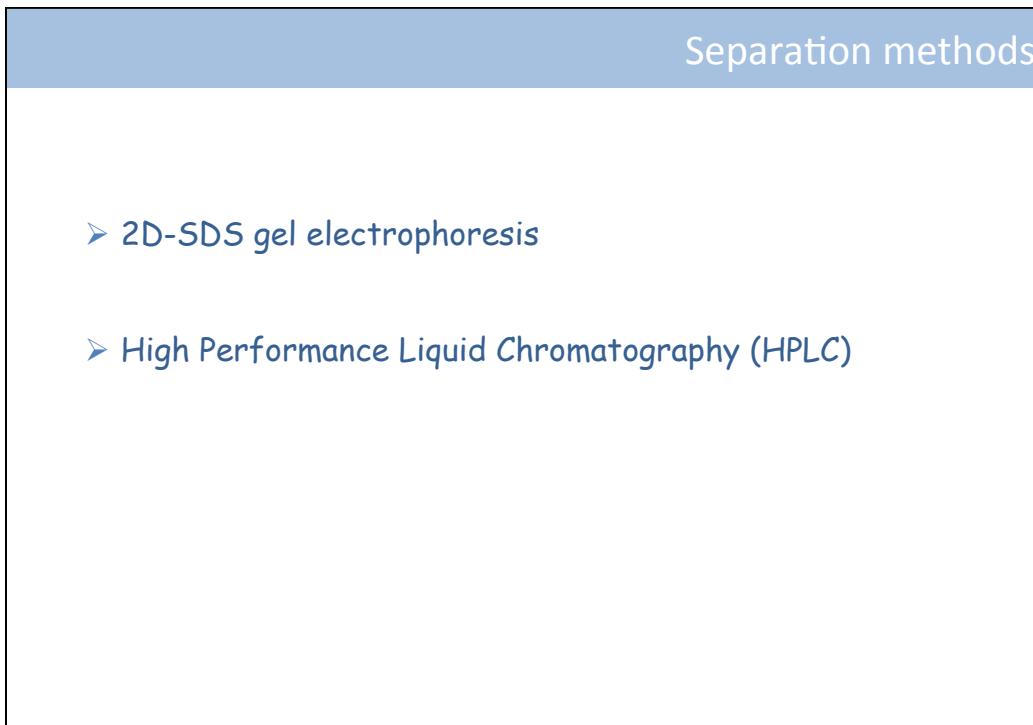
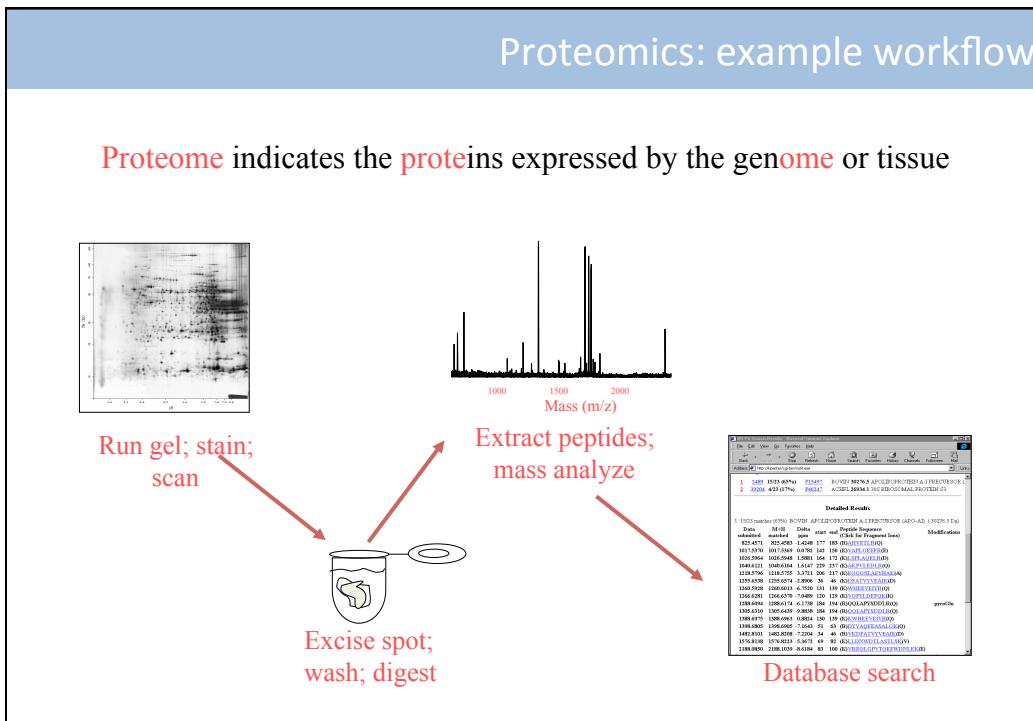
## Proteomics

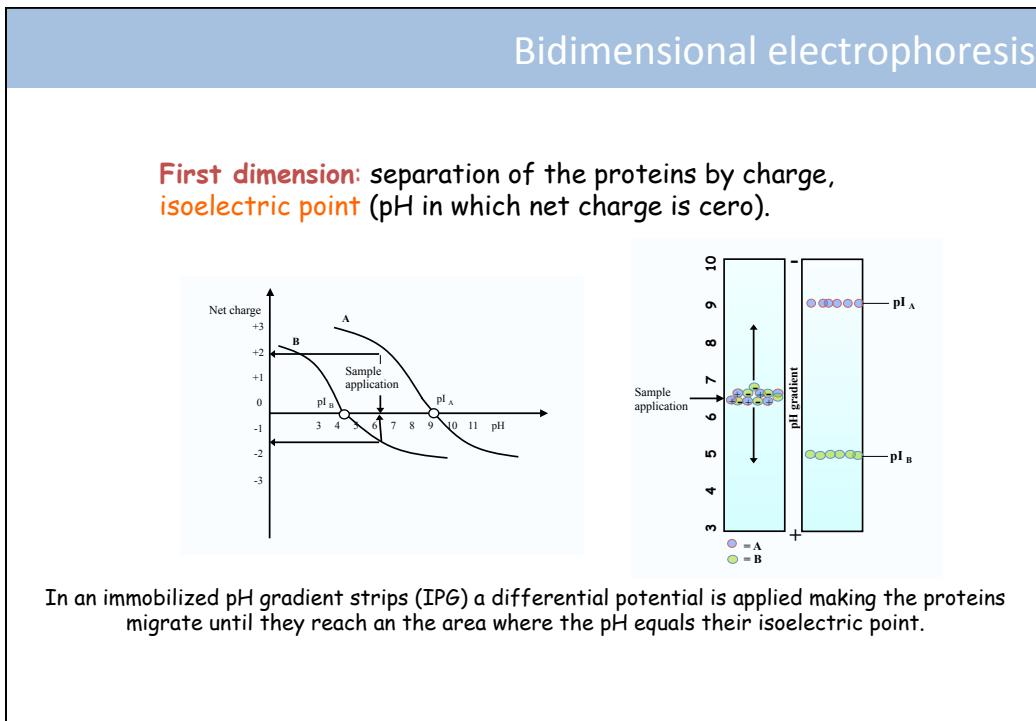
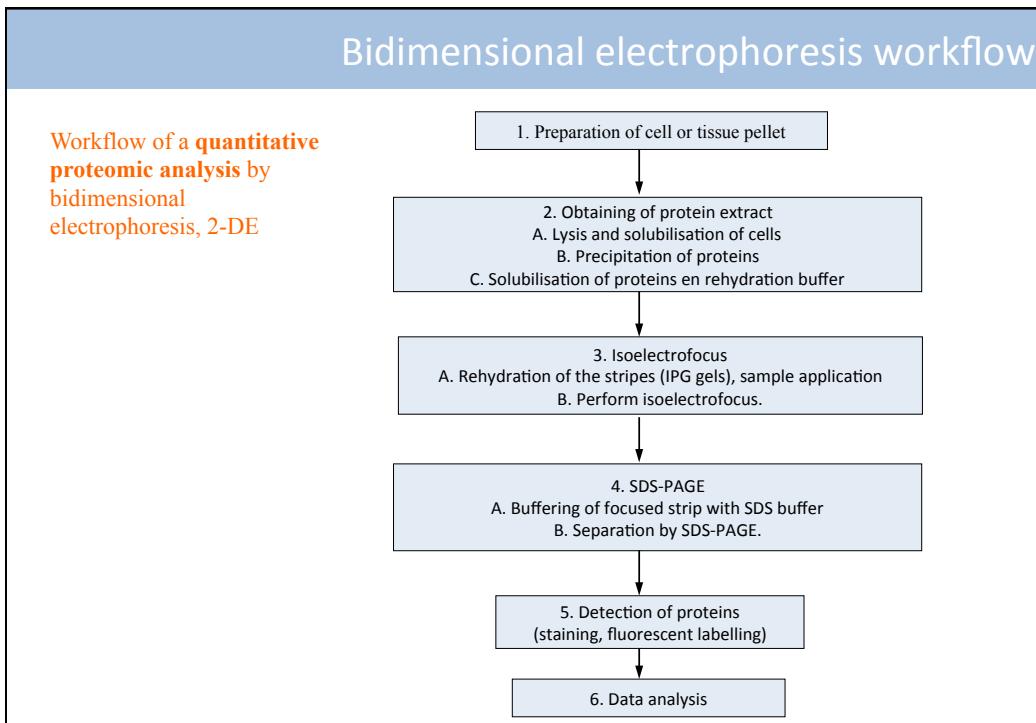
### What we can do?

- ✓ **Protein Mining:** Identification of many different proteins in complex samples.
- ✓ **Protein Expression Profiling:** Comparison of protein abundance level under determined conditions (i.e. search for protein candidate for biomarkers of diseases).
- ✓ **Protein Network Mapping:** Approach to look at the interaction between proteins from different systems.
- ✓ **Mapping of Protein Modifications:** Characterization of post-translational modifications and site mutation localization.

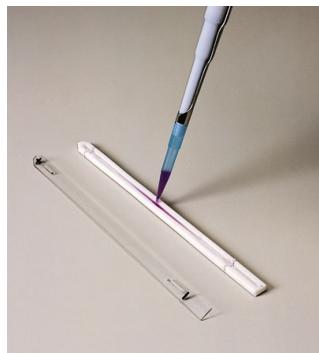
## Basic tools in Proteomics

- ✓ **Separation and selection of target proteins**
- ✓ **Measure of peptide masses**
- ✓ **Comparison with available databases**

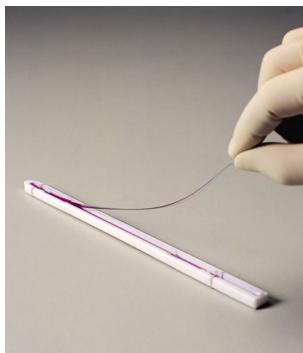




## Bidimensional electrophoresis



## Application of the rehydration buffer in the sarcophagus



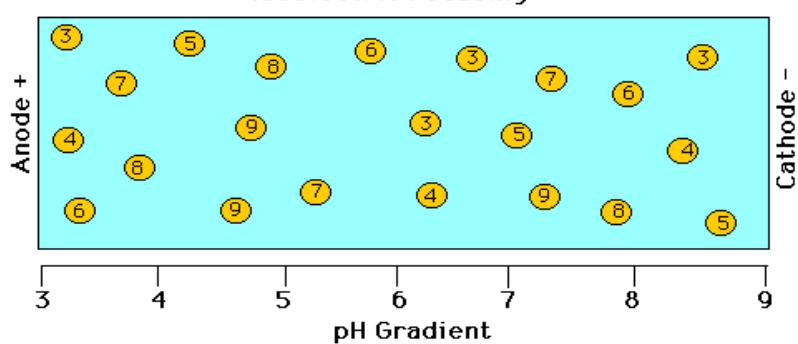
## Deposition of the IPG stripe in the rehydration buffer

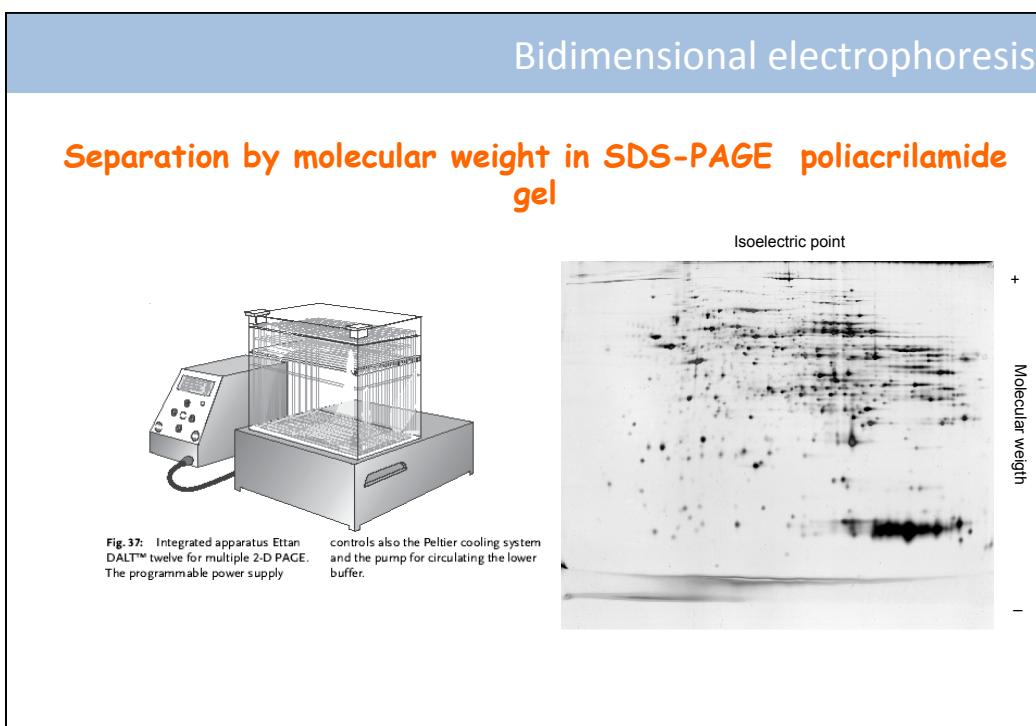
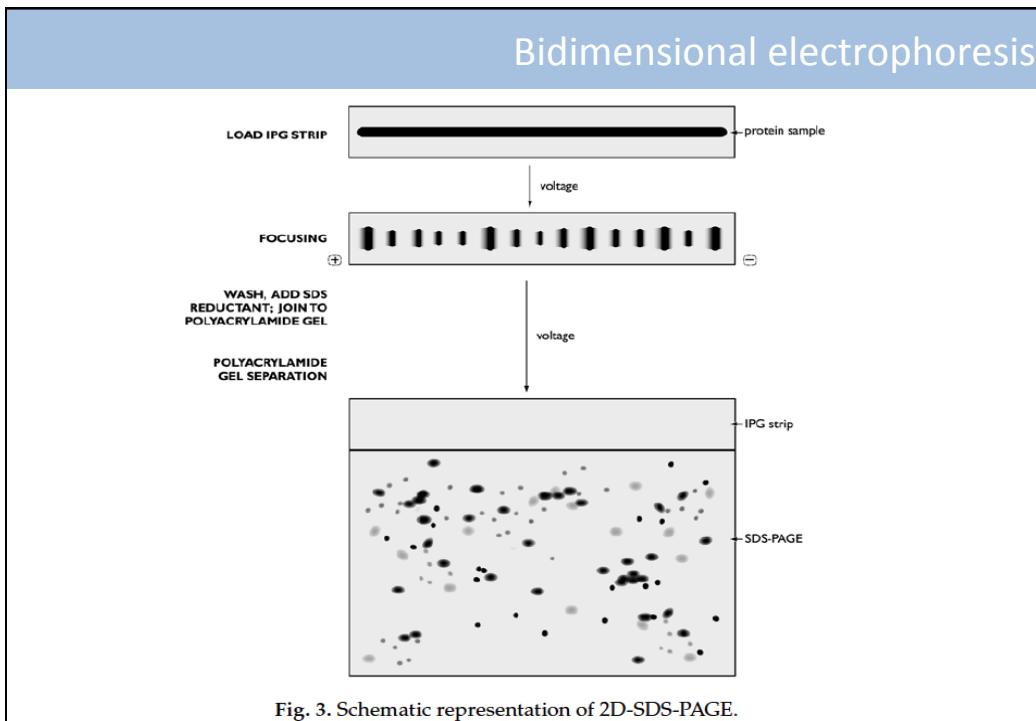


Example of an isoelectrofocusing apparatus

## Bidimensional electrophoresis

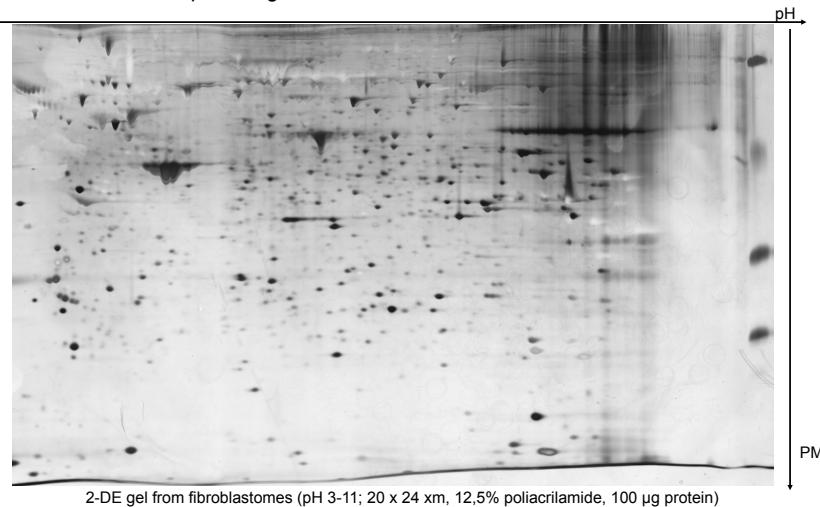
### Isoelectric Focusing



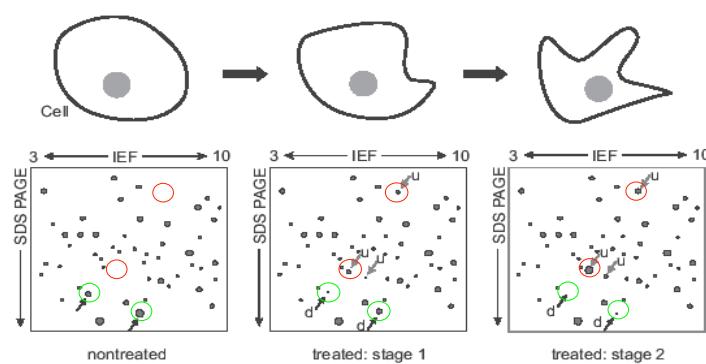


## Bidimensional electrophoresis

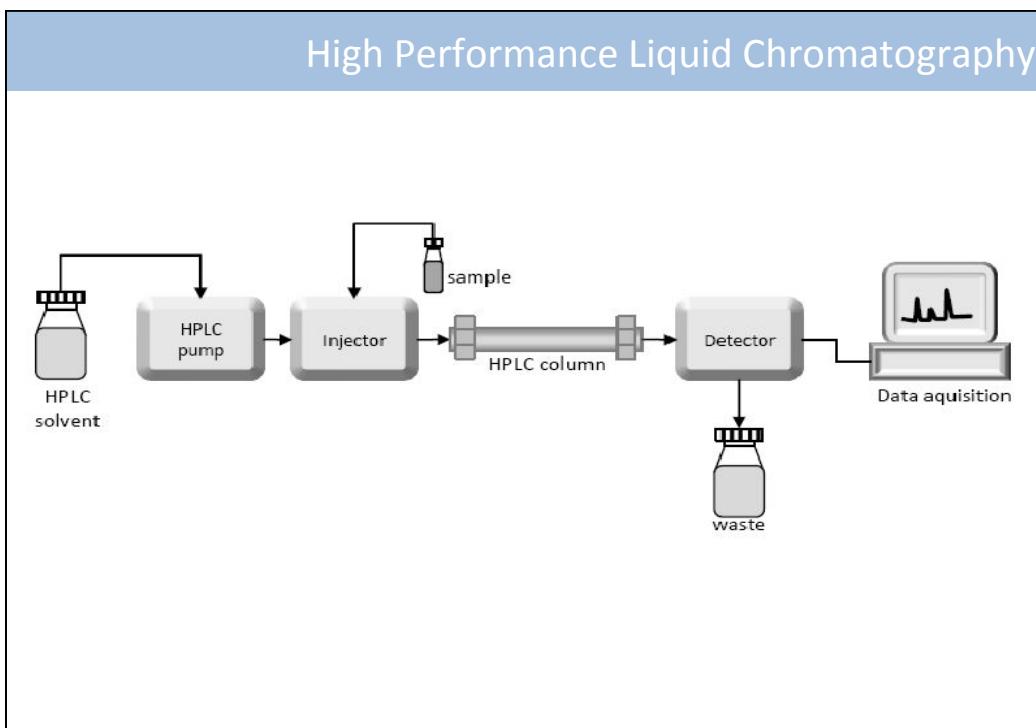
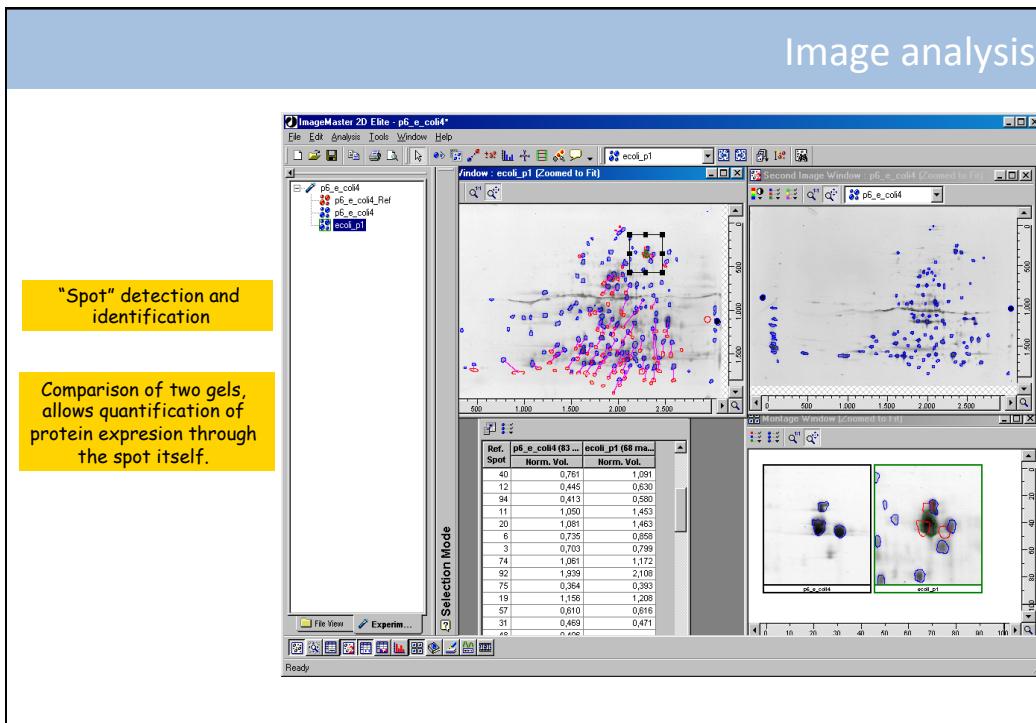
Example of a gel stained with silver nitrate



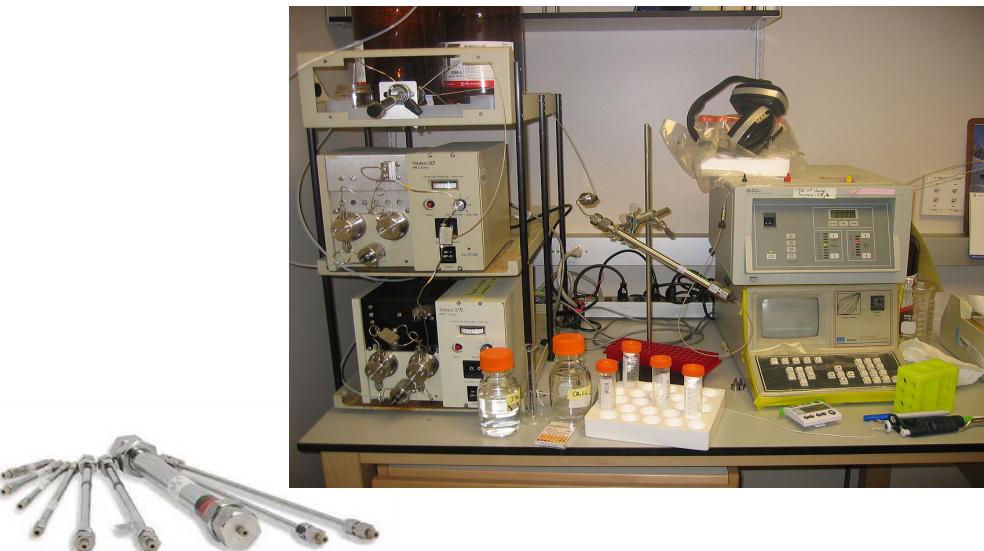
## Image analysis



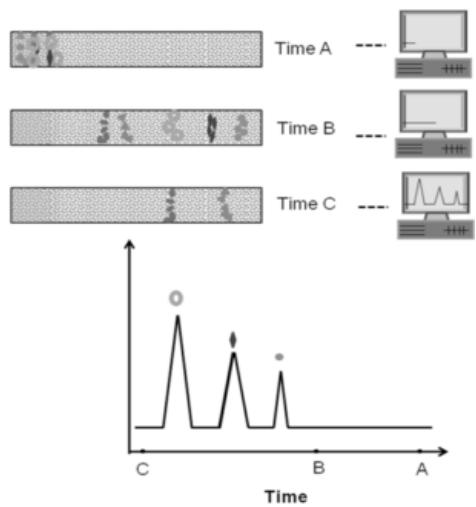
**Fig. 1:** Schematic representation of a series of 2-D gels showing different stages of a proteome. Up and down regulated gene products are marked with u and d respectively.



## High Performance Liquid Chromatography



## High Performance Liquid Chromatography



The different fractions separated are kept for further analysis.

## High Performance Liquid Chromatography

### Chromatography methods

- **RP:** hydrophobicity
- **Anion/cation exchange:** net negative or positive charge
- **Size exclusion:** peptide size/molecular weight
- **Affinity chromatography:** interaction with functional groups

**Tandem HPLC** separations can combine two different methods to improve resolution.

## Protein fragmentation

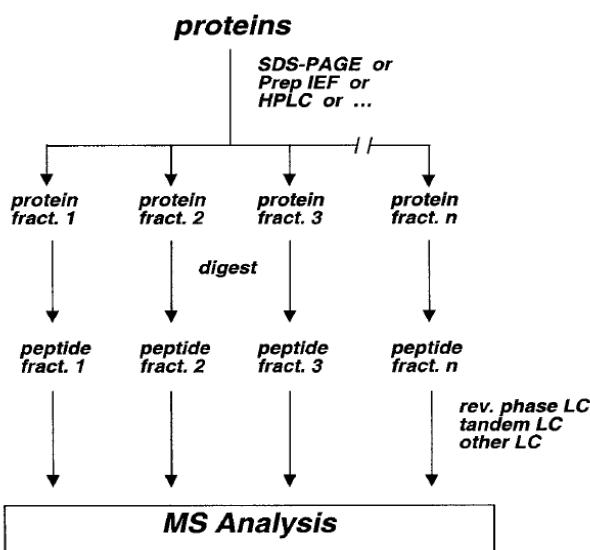


Fig. 8. Generic approach to protein/peptide fractionation.

## Protein fragmentation

Why digest proteins?

1. The greater the mass the greater the error.
2. Large hydrophobic proteins difficult to measure.
3. Sensitivity of measurements for intact protein masses not as good as sensitivity for peptide masses.

Optimal size 6 to 20 aminoacids.

## Group Activity

1. There are four nucleotides and 20 genetically encoded amino acids. Combinations of nucleotides code for each amino acid. How many amino acids could be coded for by 1 nucleotide? By a string of 2 nucleotides? By a string of 3 nucleotides? 4? What is the equation to describe the # of amino acids that can be uniquely coded as a function of the number of nucleotides in the string?

*As an example, if translation worked by mapping a single nt to the amino acid sequence (coded by one nucleotide), then you could have, at most, four amino acids in the proteome, since you only have four nucleotides.*

2. The goal of analytical proteomics is to quantitatively measure proteins. This is done by counting short peptides (short strings of amino acids). If human proteins were random sequences of amino acids, what is the minimum peptide length you would need to uniquely identify each protein?

*As an example, if the human genome had only ten genes, you could possibly uniquely identify each gene with only a single amino acid (if each gene was composed of only one aa).*

*However, if the human genome had 23 genes, it would be impossible to uniquely identify each gene using one aa. But you could with a string of two AAs.*

3. Write a regular expression (pattern matching code) to search for Open Reading Frames in a genome. You have to identify the start codon, the stop codon, and the stop codon has to be in-frame with the start codon.

4. You could use the above as your 'database' of predicted proteins, or you could use ORFs in known transcripts (eg: from RNA-seq). What is one advantage of the former? Of the latter?

## Protein fragmentation

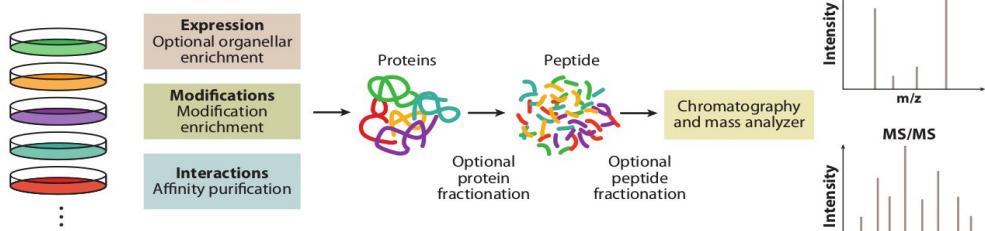
Cleavage at known sites gives additional information

### Proteases and Their Cleavage Specificities

Enzyme	Cleavage specificity
Trypsin	/K-, /R-, \P
Chymotrypsin	/W-, /Y-, /F-, \P
Glu C (V8 protease)	/E-, /D <sup>a</sup> -, \P
Lys C	/K-, \P
Asp N	/D-

<sup>a</sup>Cleavage after aspartate and glutamate in sodium phosphate buffer; otherwise cleavage only after glutamate.

## Mass Spectrometry Workflow



- Mass spec is a technique used to **identify and quantify** compounds
- It can be applied to study the **protein composition** of biological samples

goals:

determine expression level (abundance) of each protein  
determine post-translational modifications

## Mass Spectrometry output

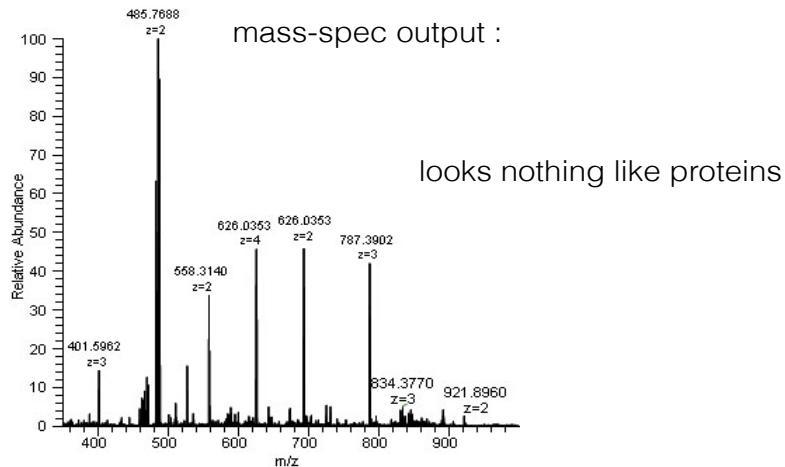
mass-spec (proteomics) is hard

sequencer (transcriptome) output:

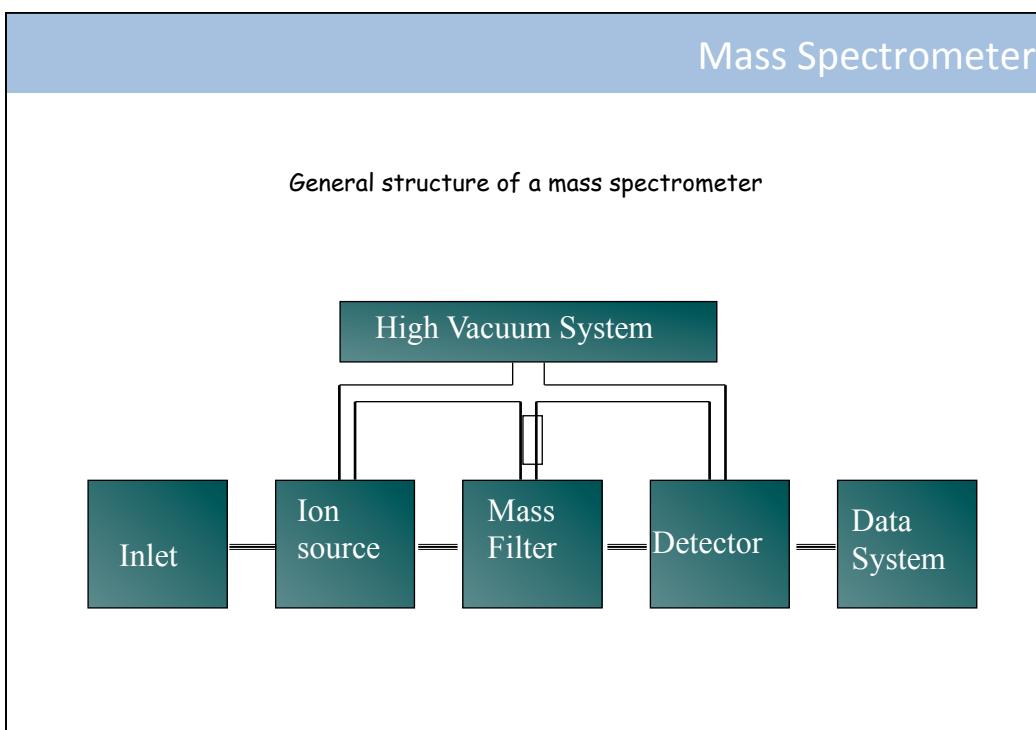
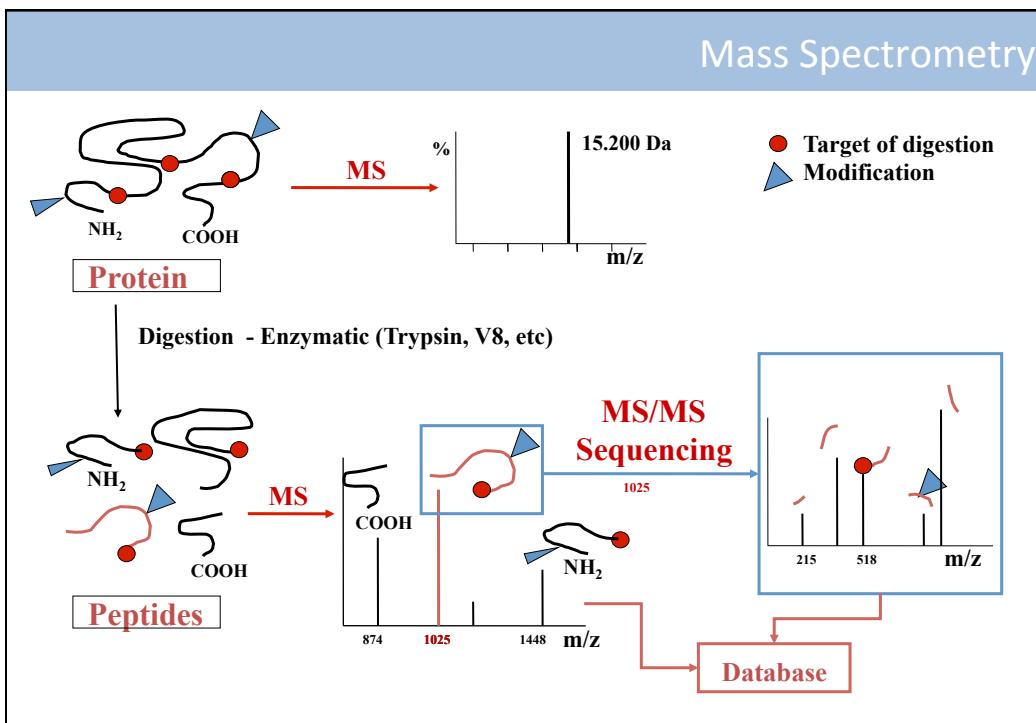
looks like genes/DNA/mRNA

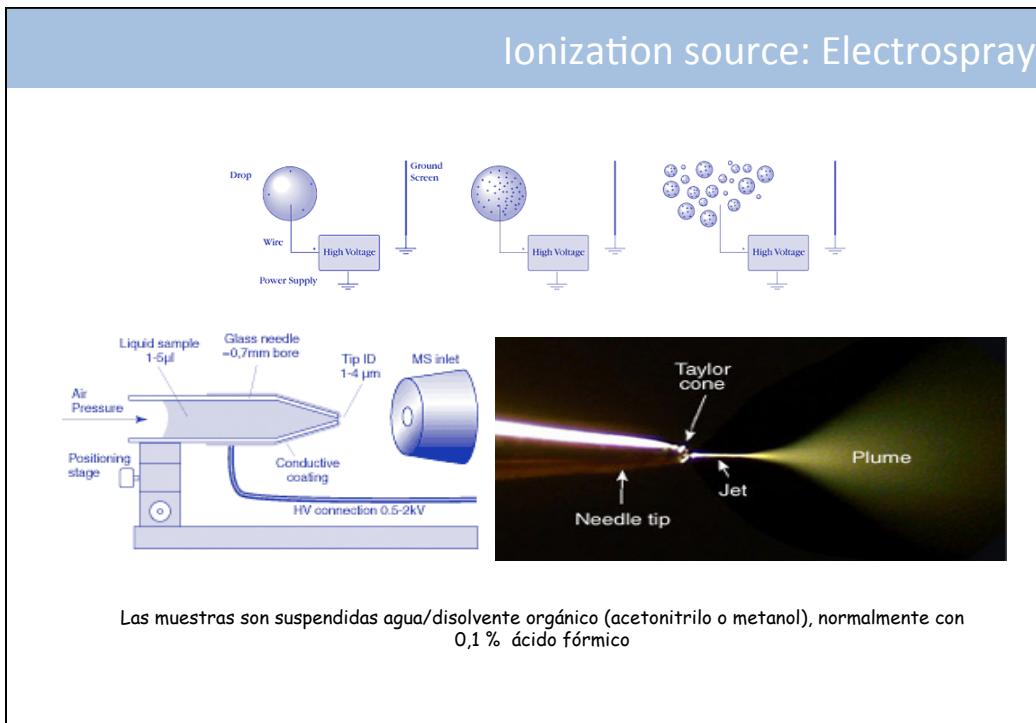
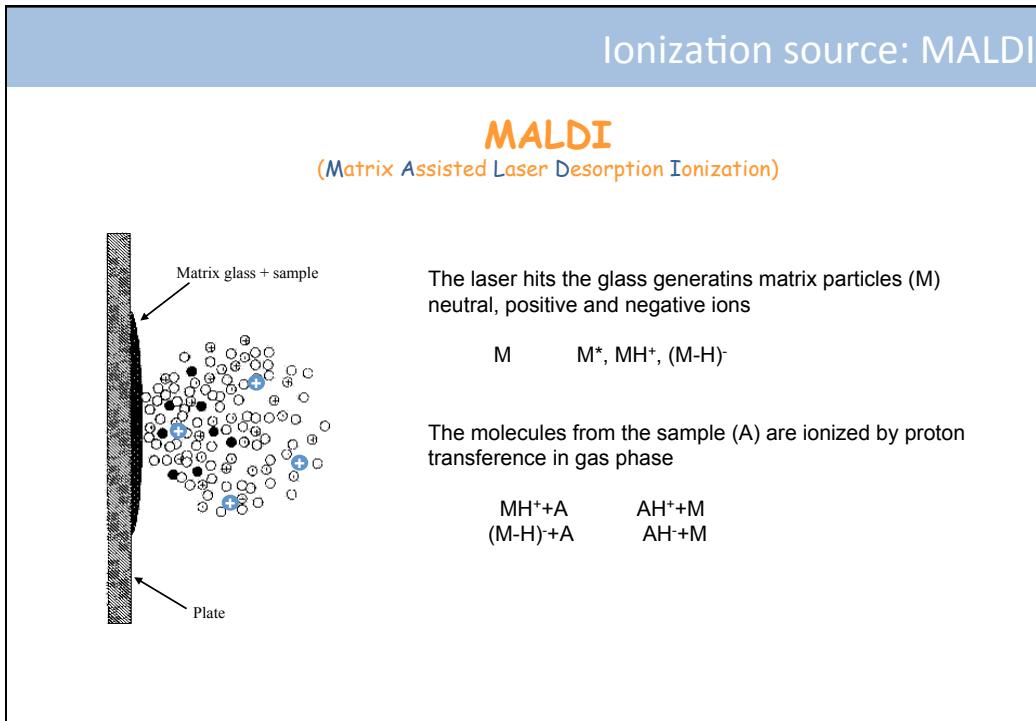
# Mass Spectrometry output

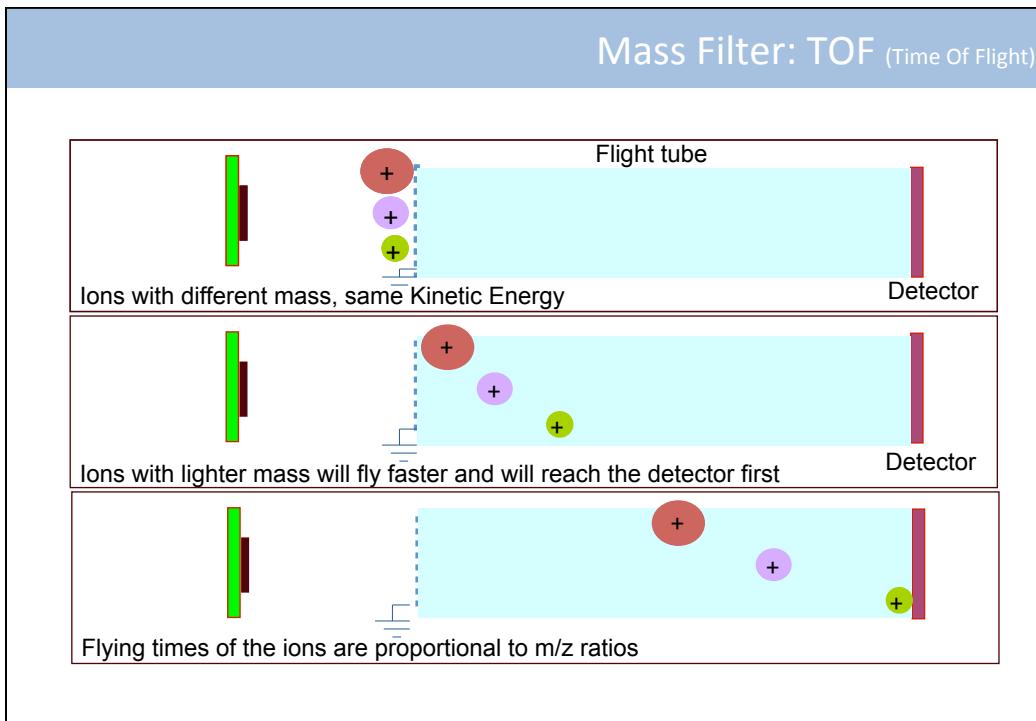
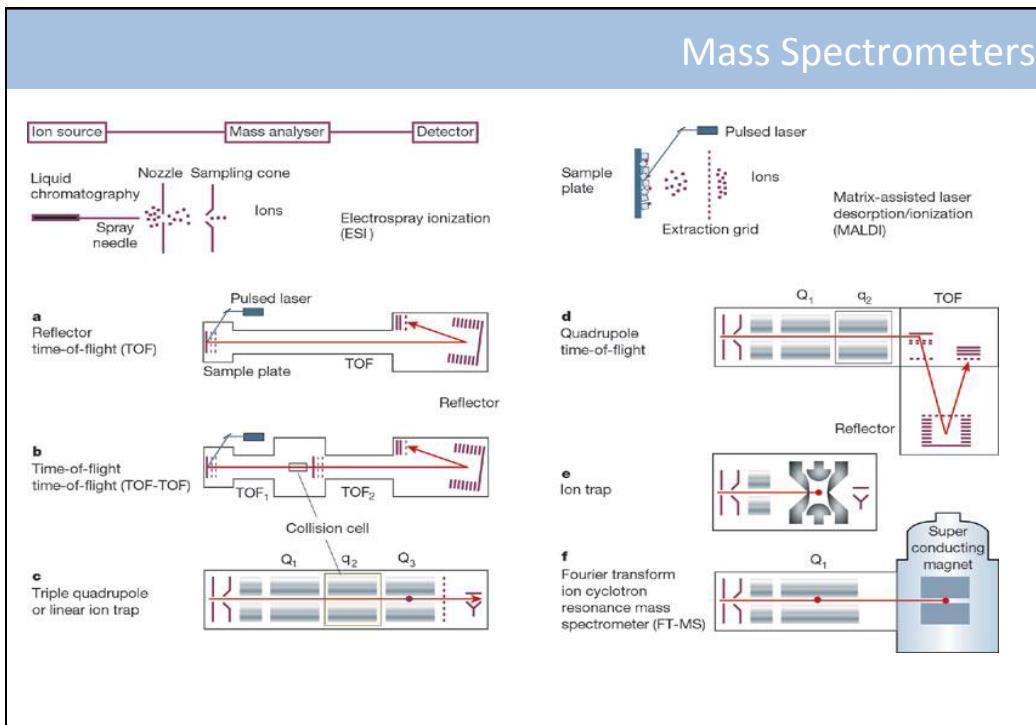
mass-spec (proteomics) is less popular, partially because it's hard

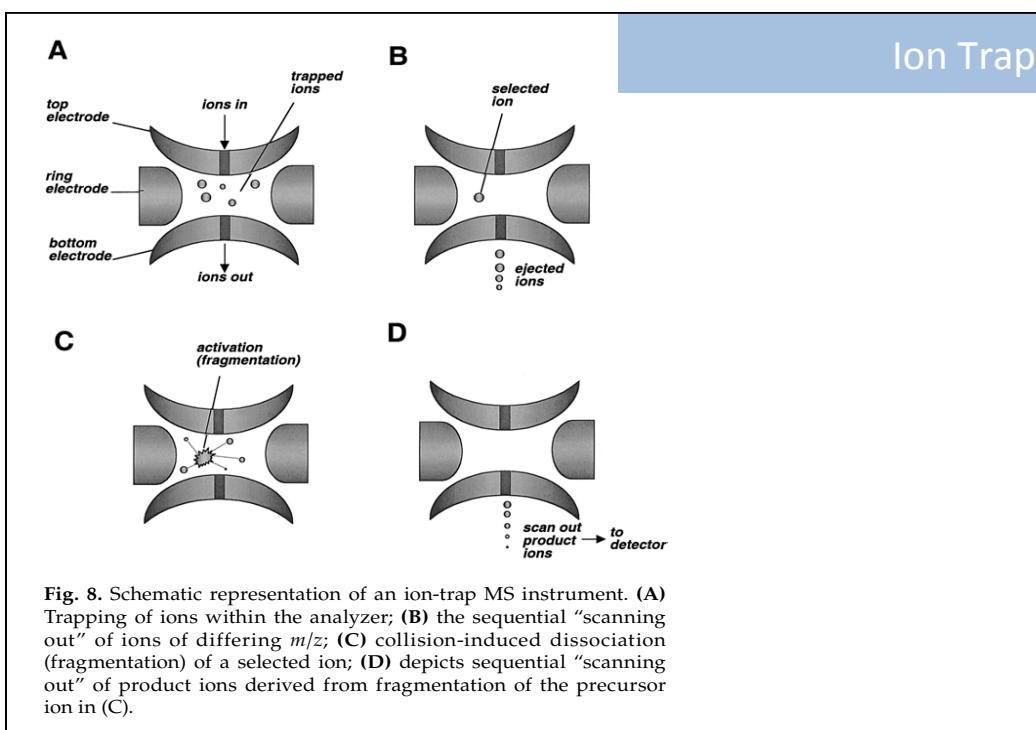
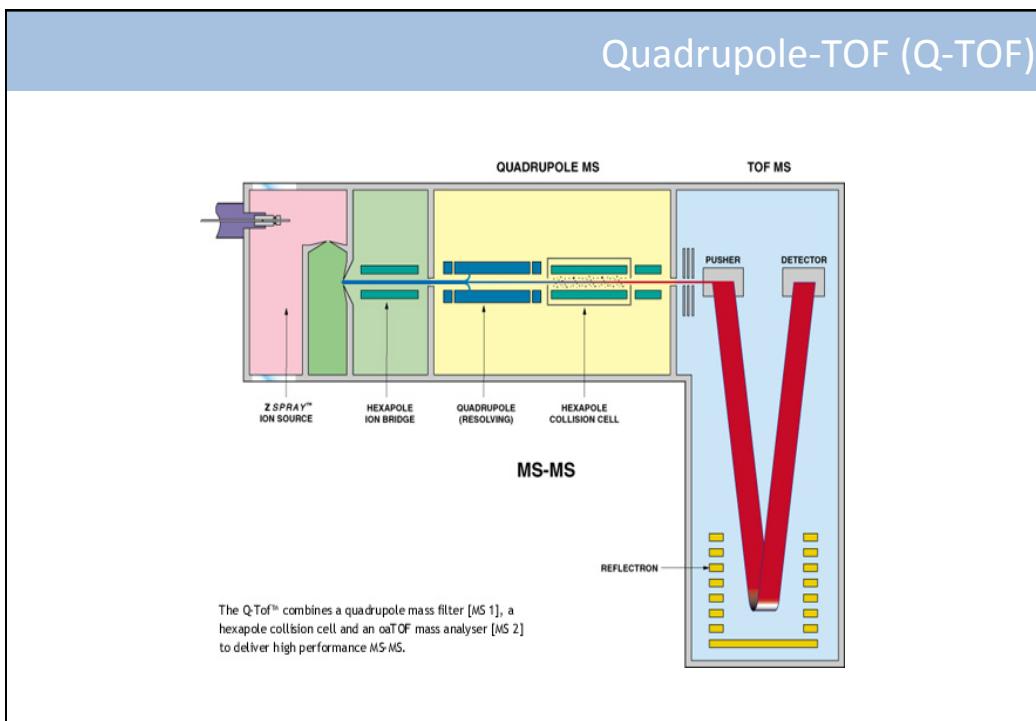


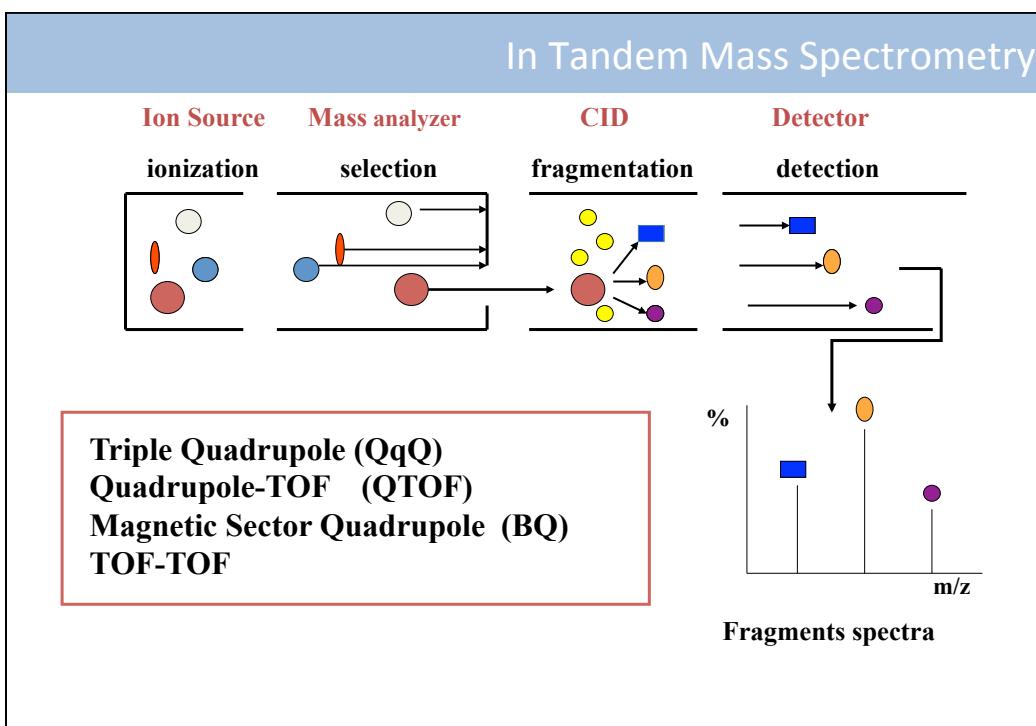
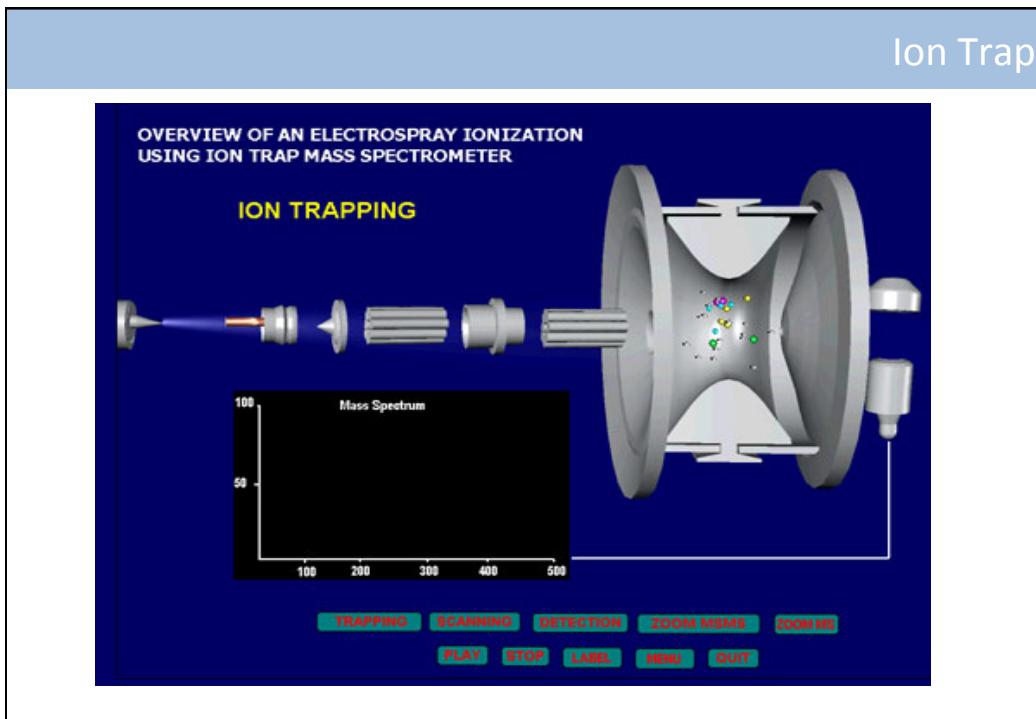
**How do we go from this to proteome quantification?**



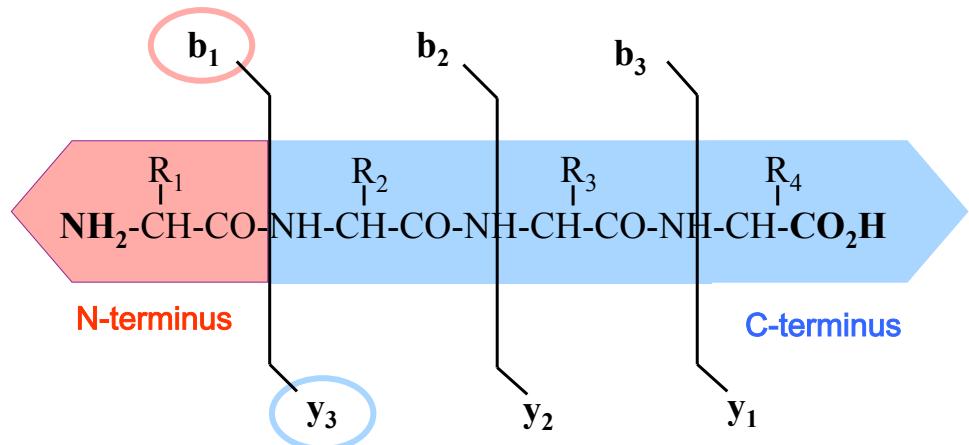




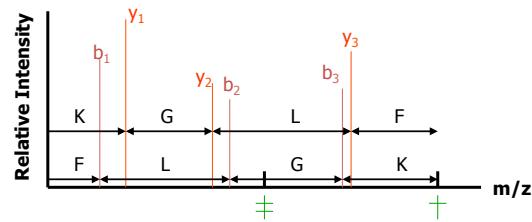
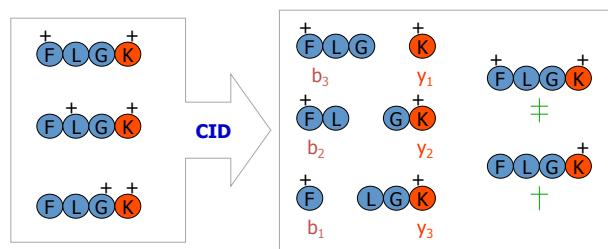




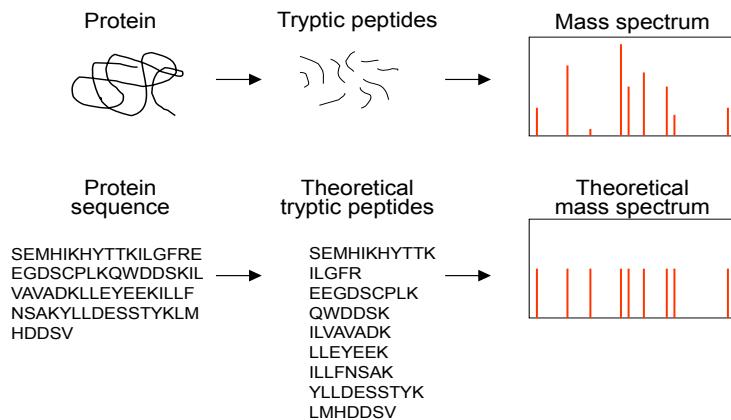
## Fragmentation pattern of peptides



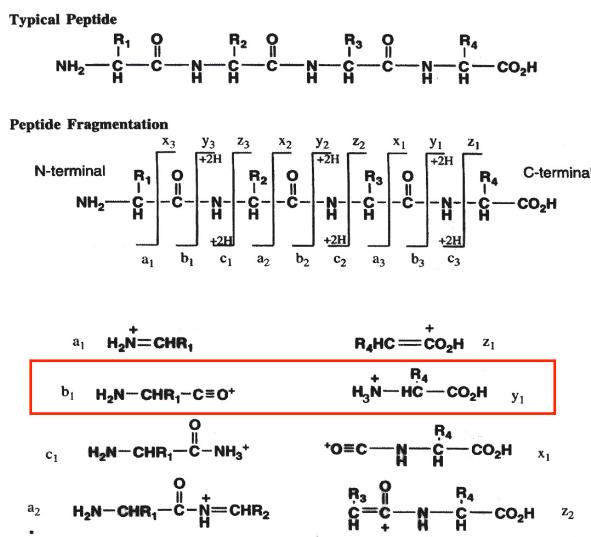
## Theoretical CID of a Tryptic Peptide



## Peptide Mass fingerprinting (PMF)



## Naming of peptide fragments



### Fragmentación del esqueleto

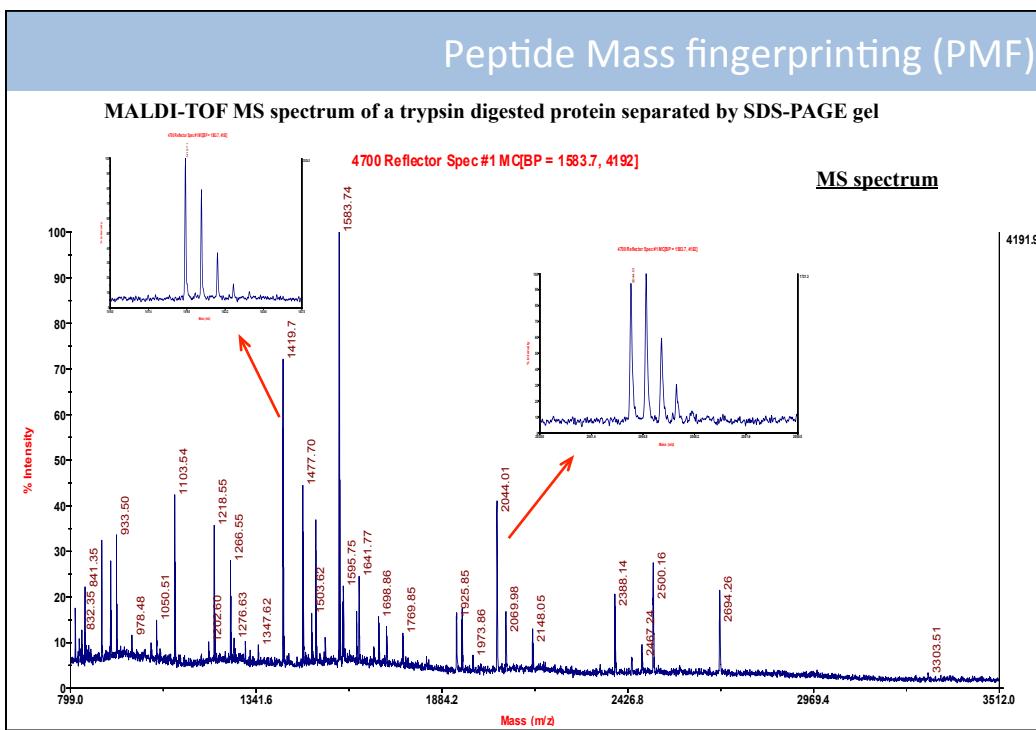
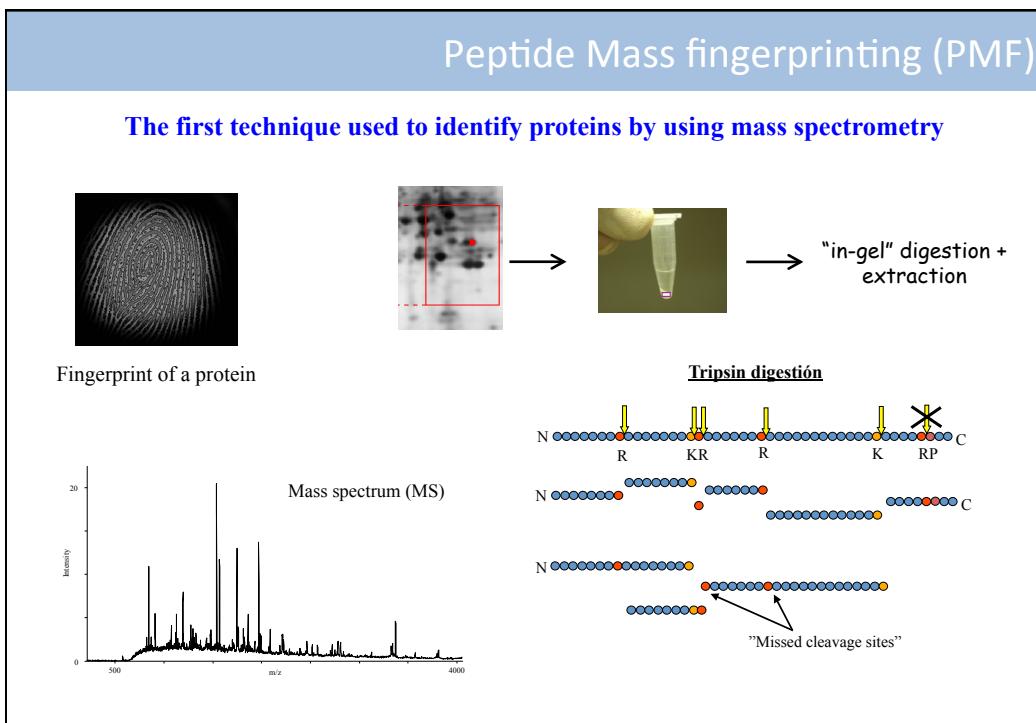
iones  $a, b, c$ : carga N-terminal

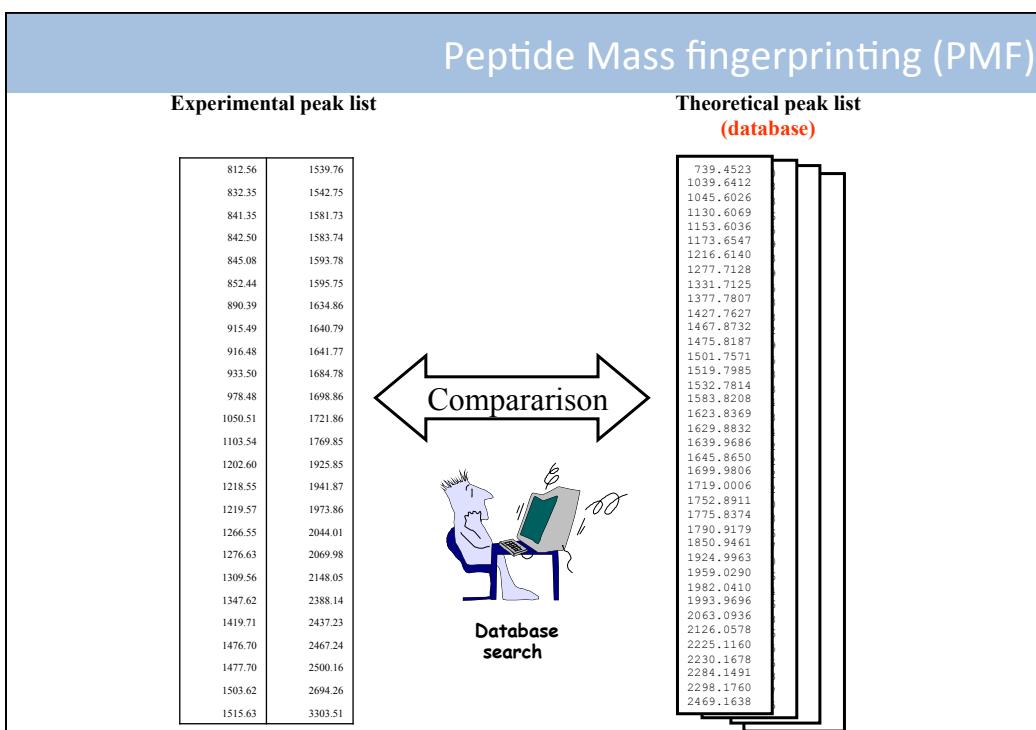
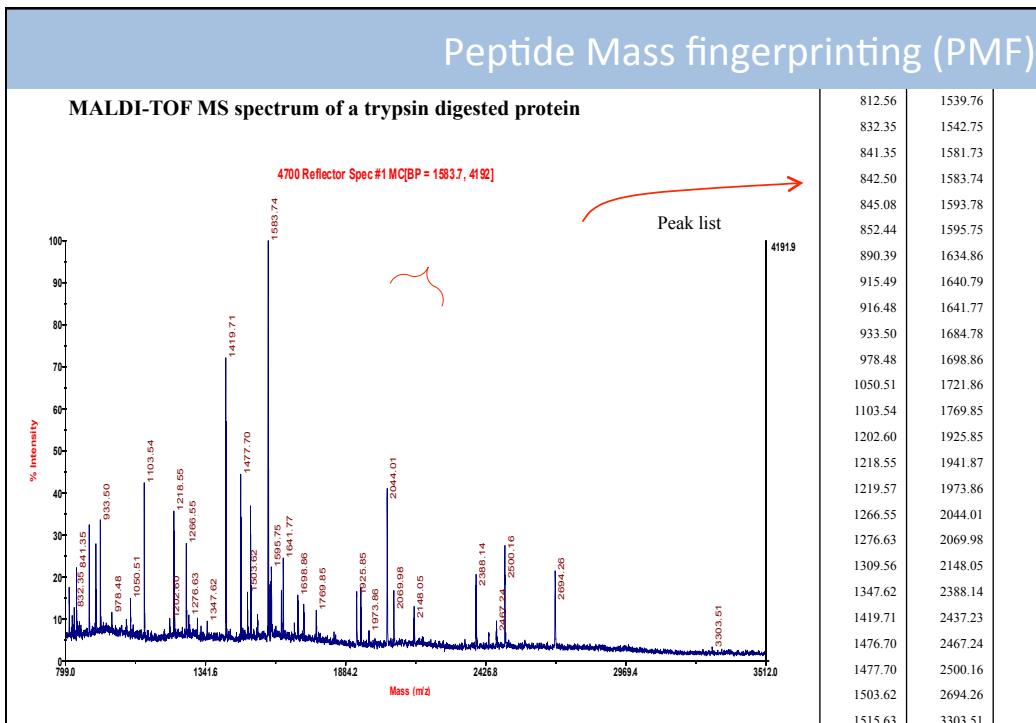
iones  $x, y, z$ : carga C-terminal

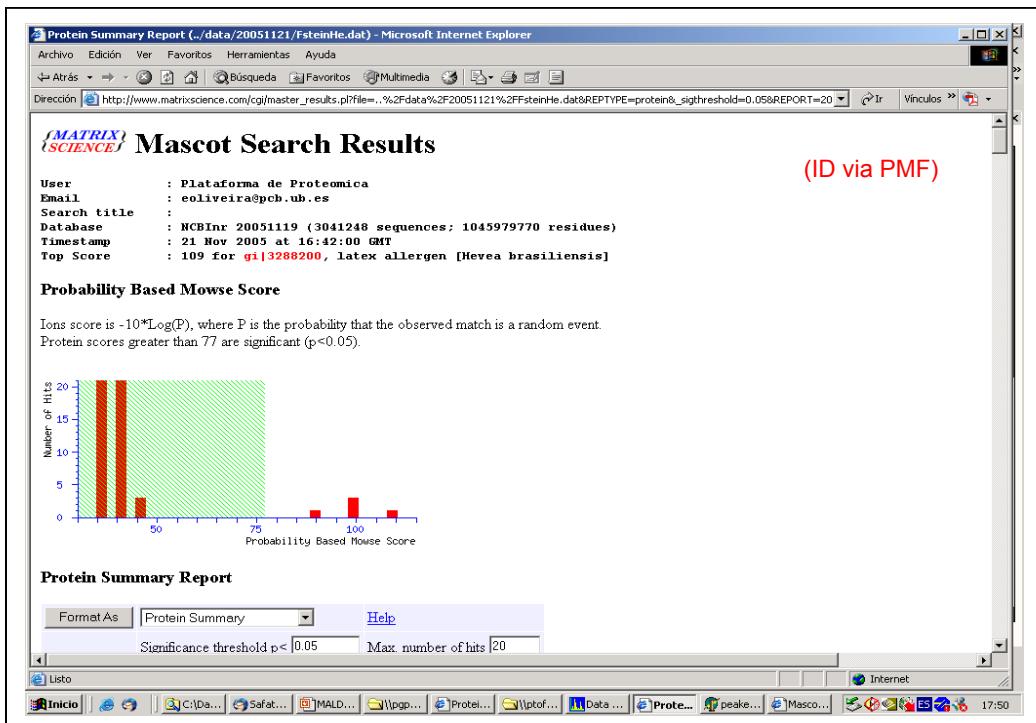
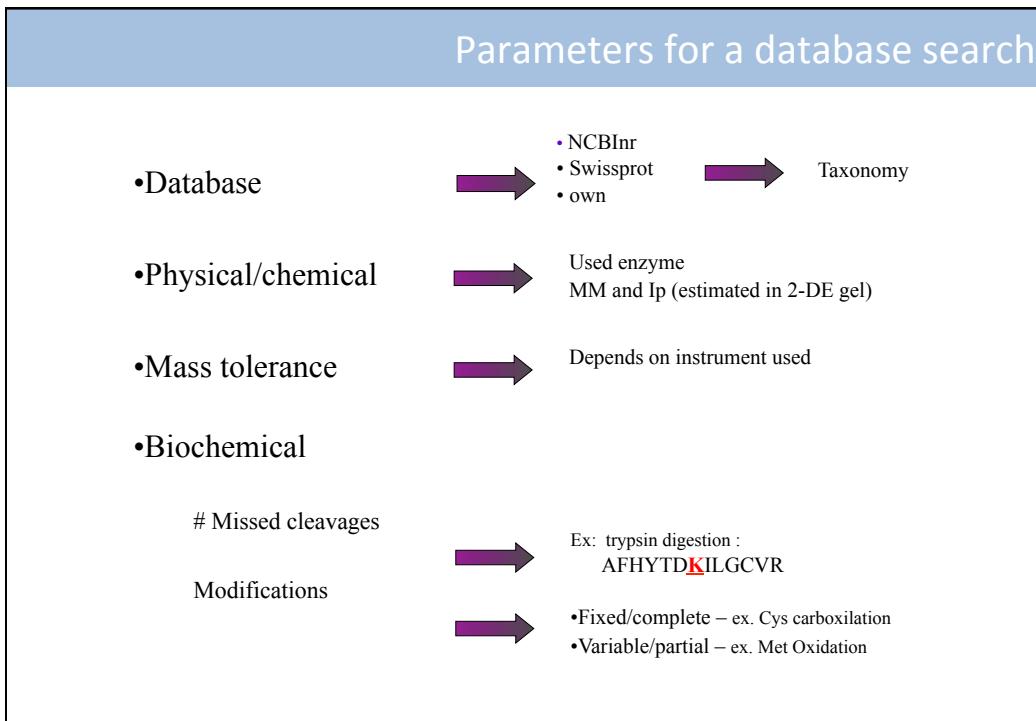
iones  $a, b, y$  son más comunes en las fragmentaciones de baja energía realizadas en el CID

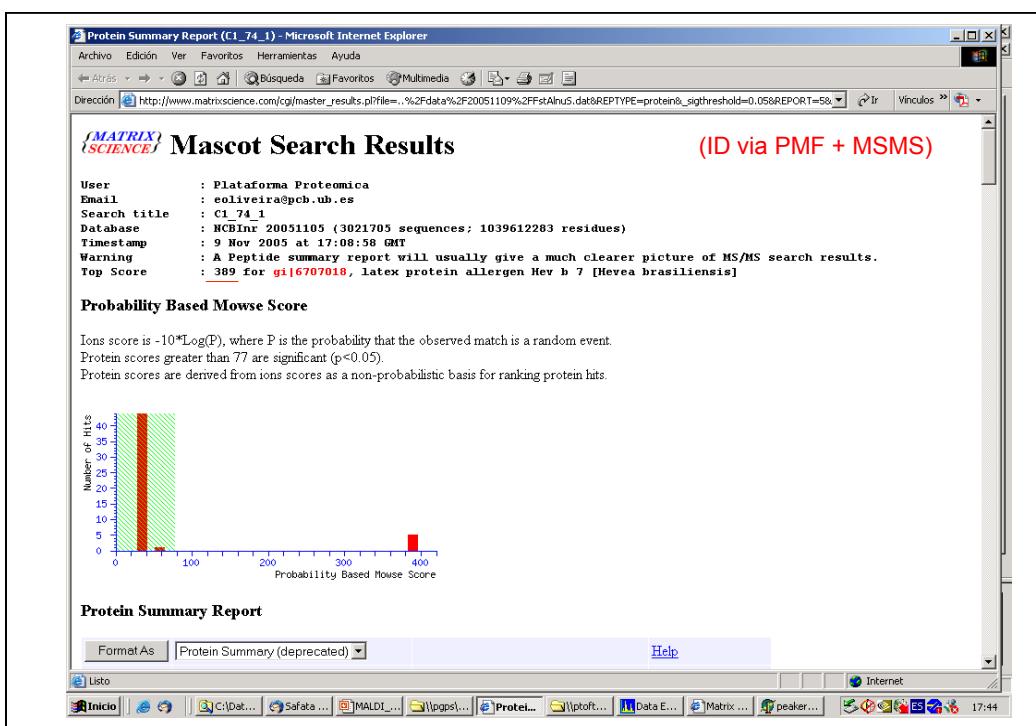
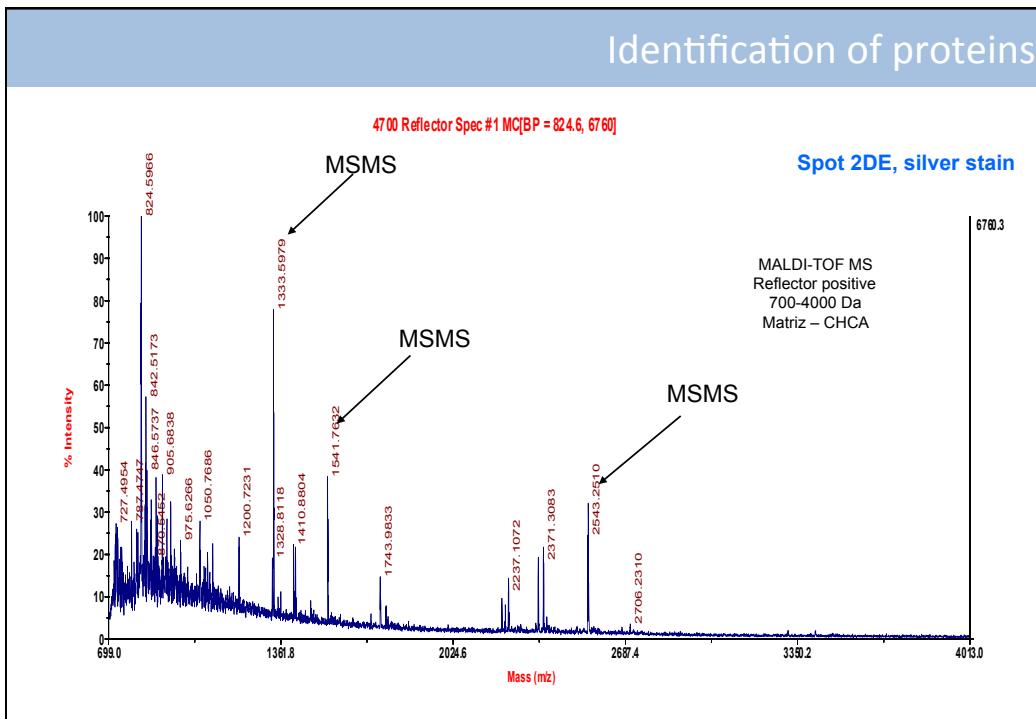
$x, z$  son poco comunes y  $c$  es raramente observado

Los iones "y" suelen ser los más intensos en los espectros de MSMS de péptidos trípicos









**Mascot Search Results: Protein View - Microsoft Internet Explorer**

Archivo Edición Ver Favoritos Herramientas Ayuda

← Atrás → 🔍 Búsqueda Favoritos Multimedia 📁 🗂️ 📎

Dirección: [http://www.matrixscience.com/cgi/protein\\_view.pl?file=../data/20051109/FstAlnu5.dat&hit=1](http://www.matrixscience.com/cgi/protein_view.pl?file=../data/20051109/FstAlnu5.dat&hit=1) 🔍 Ir Vínculos 🔍

**(MATRIX)  
(SCIENCE) Mascot Search Results**

### Protein View

Match to: **gi|6707018 Score: 389 Expect: 3.8e-33**  
Latex protein allergen Hev b 7 [*Hevea brasiliensis*]

Nominal mass ( $M_z$ ): **43107**; Calculated pI value: **5.00**  
NCBI BLAST search of [gi|6707018](#) against nr  
Unformatted [sequence string](#) for pasting into other applications

Taxonomy: [Hevea brasiliensis](#)

Fixed modifications: Carbamidomethyl (C)  
Variable modifications: Oxidation (M)  
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P  
Sequence Coverage: **49%**

Matched peptides shown in **Bold Red**

```

1 MMTGSTPLTO GKKTVLISID GGGIRGTYIPG TILASLESKI QDLDGPDART
51 ADYFDIILAGT STGGGLITML TAPNEDKRPW YQADKIFDY LENCPKIPPK
101 ESRDNYDPPIH SIGPTYDY IRELCHNLIK DLTVKDLTD WIPTFDIKL
151 LLPVIFESSDA AKCNALKHAR LDPCISTSA APVLLFAHST TTEDDKNIHT
201 FELIDOGGVAA TNPTILLATH IRNEIIRQNPF RFIGAMLTES KSRLVLSLGT
251 GKSEKVEKY ADNTSKWY RL NWALYGNNSP AVDIFSNASS DNVDFRLSAL
301 FKSLSDCEDY LRIQDDTLTG ESSHHIATE EHQLRVLEIG TELLERQESR
351 IHLDTGES IPGAPTHEAAKTAKFLARLESE ERKLRQLK

```

Show predicted peptides also

Sort Peptides By  Residue Number  Increasing Mass  Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
1-351	MMTGSTPLTO GK <b>KTVLISID GGGIRGTYIPG TILASLESK<b>I QDLDGPDART</b></b>					
51-60	ADYFDIILAGT STGGGLITML TAPNEDKRPW YQAD <b>KIFD<b>Y LENCPKIPPK</b></b>					
101-110	ESRDNYDPPIH SIGPTYD <b>Y IRELCHNL<b>IK DLTVKD<b>LTD WIPTFD<b>IKL</b></b></b></b>					
151-160	LLPVIFESSD <b>A AKCNALKHAR LDPCISTSA APVLLFAHS<b>T TTEDDKNI<b>HT</b></b></b>					
201-210	FELIDOGGVAA TNPTILLATH IRNEII <b>RQNPF RFI<b>GAMLTES KSRL<b>VLSL<b>GT</b></b></b></b>					
251-260	G <b>KSEKVEK<b>Y ADNTSKW<b>Y RL NWALYGN<b>NSP AVDIFSNASS DNVD<b>FRLS<b>AL</b></b></b></b></b></b>					
301-310	FKS <b>LSD<b>CED<b>Y LRI<b>QD<b>D<b>TLTG E<b>SSHH<b>IA<b>TE E<b>HQLR<b>VL<b>E<b>IG TELLER<b>Q<b>ESR</b></b></b></b></b></b></b></b></b></b></b></b></b></b></b>					
351-360	IHL <b>D<b>T<b>G<b>E<b>S IPGAP<b>T<b>HEA<b>AK<b>TA<b>K<b>FL<b>AR<b>LE<b>SE ERKL<b>R<b>Q<b>LK</b></b></b></b></b></b></b></b></b></b></b></b></b></b></b></b></b>					

Internet

Inicio 🔍 C:\da... 🔍 Safat... 🔍 MALD... 🔍 \pg... 🔍 Prote... 🔍 \ptof... 🔍 Data... 🔍 Condi... 🔍 Peaks... 🔍 Masc... 🔍 ESI... 🔍 ES... 🔍 17:46

## Why don't we have 100% of the sequence?

51 MKLKDTNLNG KTFEPPMRAGL PTKEPVMWQRE WEYAKLYQRR QELNQGKPHF  
51 TLHDGPPYAN GNIVHGAMHN KISKDIIVRS KSMMSGFYAPF IPGWDTHGLP  
101 IEQVLSKQGV KRKEMDLVEY LKLCLREYALS QVDKQREDFK RLGVSGDWEN  
151 PIVVTLTDPYE AAQIRVFGEML ANKGKYIYRGA KPVYWSNSSE SALAAEIEY  
201 HDLVSTSLYY ANKVKGDKGV LDTDYIYVWV TTTFFITIAS RGLTVGADID  
251 YLVLPQPAGEA RKFVVAEELL TSLSKEFGWA DVQVLEYTRG QELNHVITEH  
301 PWDATAVEEL ILGDHVTTDS QVHGTAVHP FGEDDDNVGI ANNLEAVVT  
351 DERGIMKNA PEGFEQGKHY KVUPVTEICKN HLLNQAEEQ SHSYSPFDWRT  
401 KKPIIWRAPP QWFASVSKFR QEIILDEIEKV KFHSEWKGVR LYMINRDGRD  
451 NVVISRQAWG VPLPIFYAED GTAIMVAETI EHVAQLFEEH GSSIWIWERDA  
501 KDLLPEGFTH PGSPNGEFKK ETIDIMDVWFD SGSSWNQVVV NRPELTYPAD  
551 LYLEGSDQYR GWFNSSLITS VANHGVPAYK QILSQGFEALD GKGEKMSKSL  
601 GNTIAPSDEV KQFGAEIILRL WTVTSVDSSND VRISMDSILSQ VSETYRKIRN  
651 TLRFILANTS DFNPAQDTVA YDELRSVDKY MTIRFNRNLVK TIRDAYADFE  
701 FLITYIKALVN FINVUDLSAFLY LDFAKDVMVYI EGAKSRLERQ MQTVFYDILV  
751 KITKLNLTPIL PHTAAEISWY LEFETEDFVQ LSELPEVQTF ANQEELILDWT  
801 EKDFRDRGQA KQALEEARNEA VKİWG SLEAH LTVPYNEVVK TLBEAVNSVN  
851 AQILIVSELT IAEGPAPEA LSFDVAFTV ERATGEVCR CRRIDPTTA

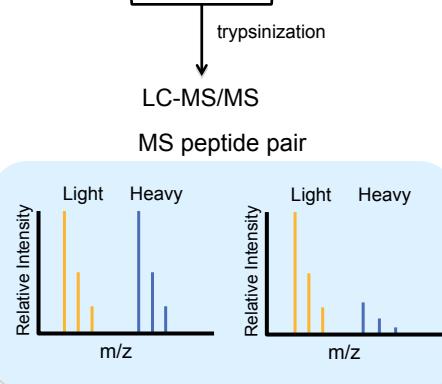
- mass range ( $m/z$ )
  - suppression effect (preferential ionization of some components over others)
  - posttranslational modifications (PTM)

## Activity 2: Individual

### Protein fingerprinting with MASCOT.

Instructions and data files in moodle.

## Protein quantification by MS: stable isotope labelling



### Comparison of peak intensity

- peptide pairs labelled with stable light/heavy isotopes
- same peptide in different MS analysis

### Isotope labelling can be used for:

- relative comparison of samples
- absolute quantification with marker peptides

### Advantages of isotope labelling ( $^{13}\text{C}$ ; $^{15}\text{N}$ ; $^2\text{H}$ )

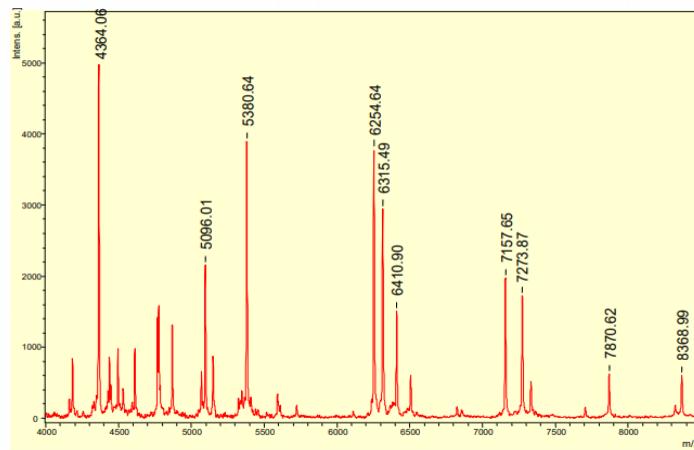
- the peptide pair is chemically very similar
- samples can be combined before or during preparation
- samples are analysed in the same experiment

## Protein quantification by MS: labelling methods

Labelling Method	Labelling sample	Mass spectrometry analysis		
<b>SILAC</b> (S <table border="1"><tr><td>table</td><td>Isotope Labeling by Amino acids in Cell culture)</td></tr></table>	table	Isotope Labeling by Amino acids in Cell culture)	Cell level	MS based quantification
table	Isotope Labeling by Amino acids in Cell culture)			
<b>iCAT</b> (Isotope Coded Affinity Tag)	Protein level			
<b>iTRAQ, TMT</b> (Isotope tagged relative quantitation)	Peptide level	MS/MS based quantification		

## MS Microorganism identification

### RESULT: Bacterial Protein “Fingerprint”



ghk@bdal.com

Bruker Daltonics

