Analyzing 2D-PAGE gels and Mass Spectrometry Data

Christian Karlsen

Department of molecular biotechnology

Outline

- 2D-PAGE analysis ***Two-dimensional polyacrylamide gel electrophoresis***
  - General introduction
  - How we generate the data
  - Bioinformatics

- Mass spectrometry analysis
  - General introduction
  - How we generate the data
  - Bioinformatics

What is Proteomics?

Proteome: *PROTEin complement expressed by a given genOME*

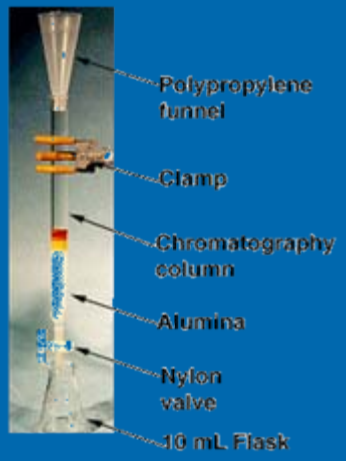
The concept of the proteome is fundamentally different from that of the genome: while the genome is virtually static and can be well defined for an organism, the proteome continually *changes in response to external and internal events*

Alternatively spliced varients

Post translational modifications

Expression in response to cellular state

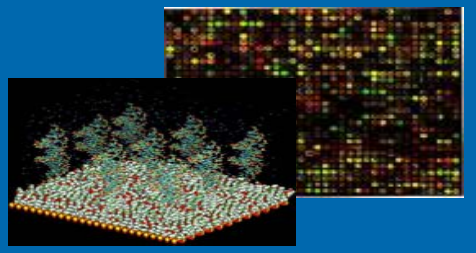
Proteomics



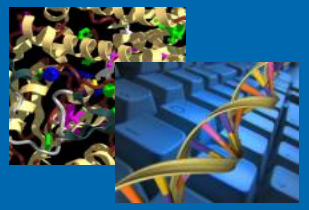
Chromatography



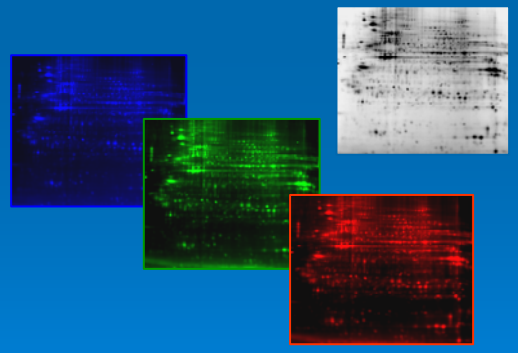
Microfluidics



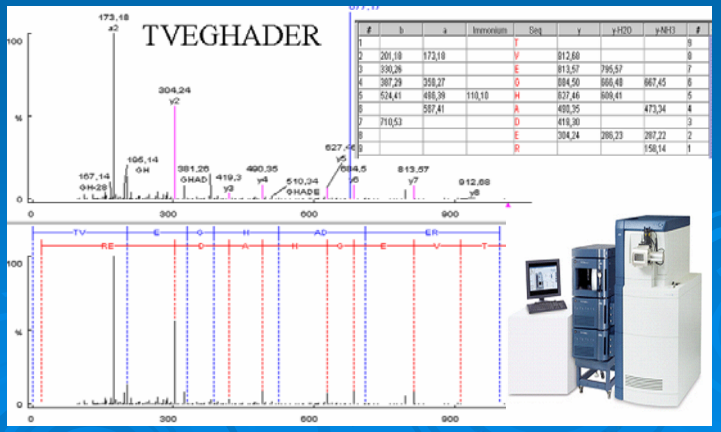
Microarrays



Bioinformatics



Electrophoresis

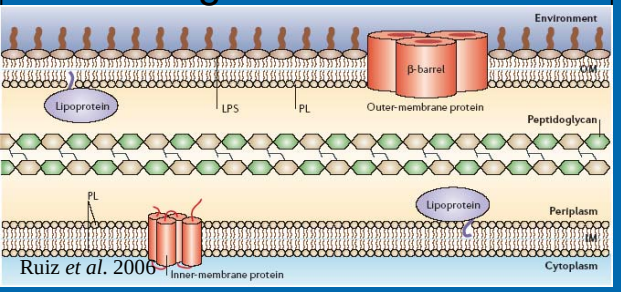


Proteomic - Steps

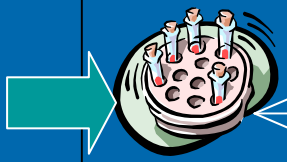
Various culture conditions:

Temp, nutrients, pH, etc.

G-neg. bacteria



Ruiz *et al.* 2006



OM proteins

Intracellular proteins

Extracellular proteins

2-DE problems

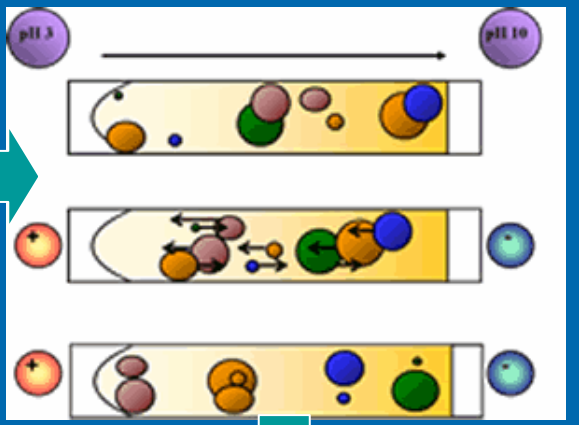
Extracting

Solubilise

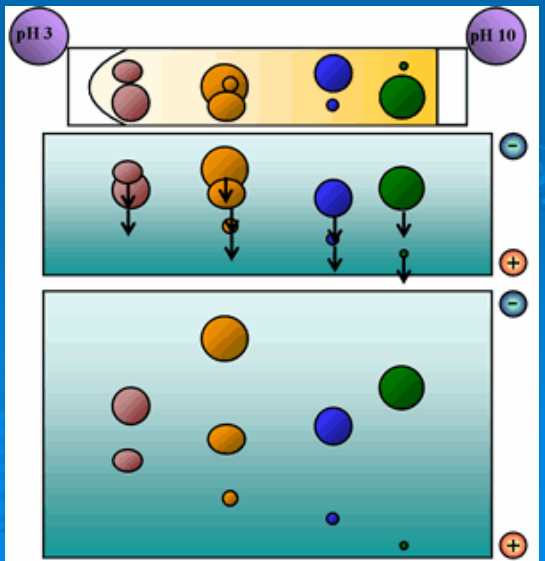
Precipitation

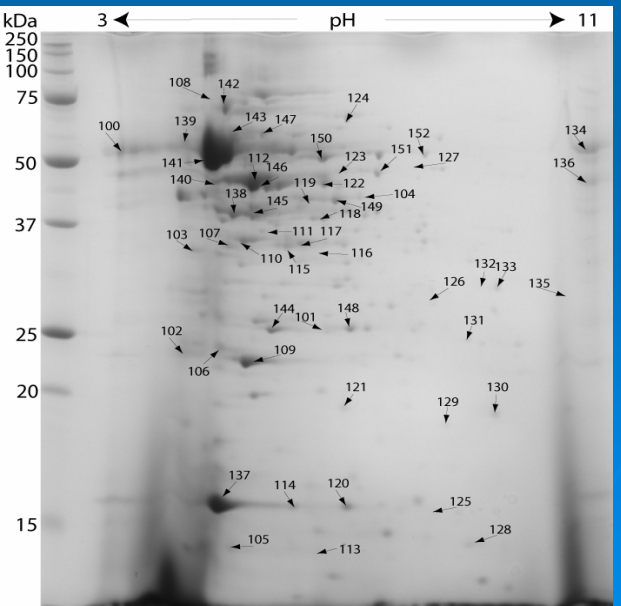
Two-dimensional gelelectrophoresis

First Dimension: IEF



Second Dimension: SDS-PAGE





Spot detection:

Coomassie blue

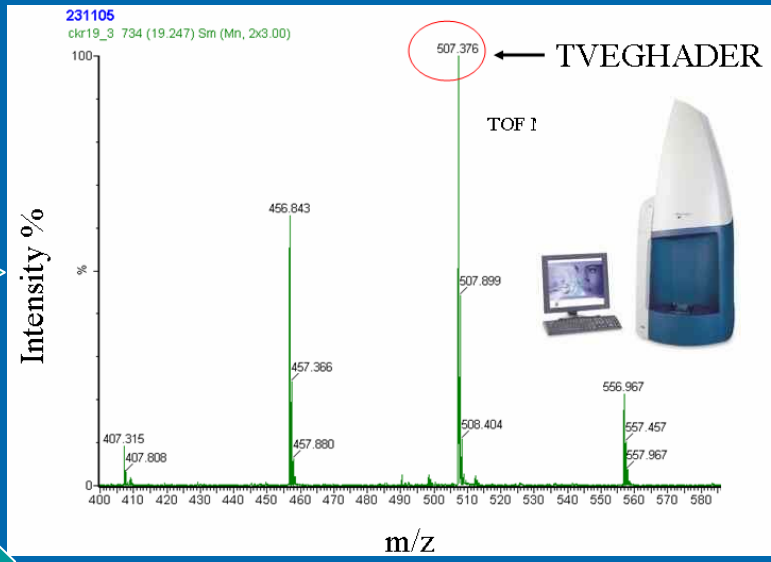
Gel scanning and image analysis: PDQuest



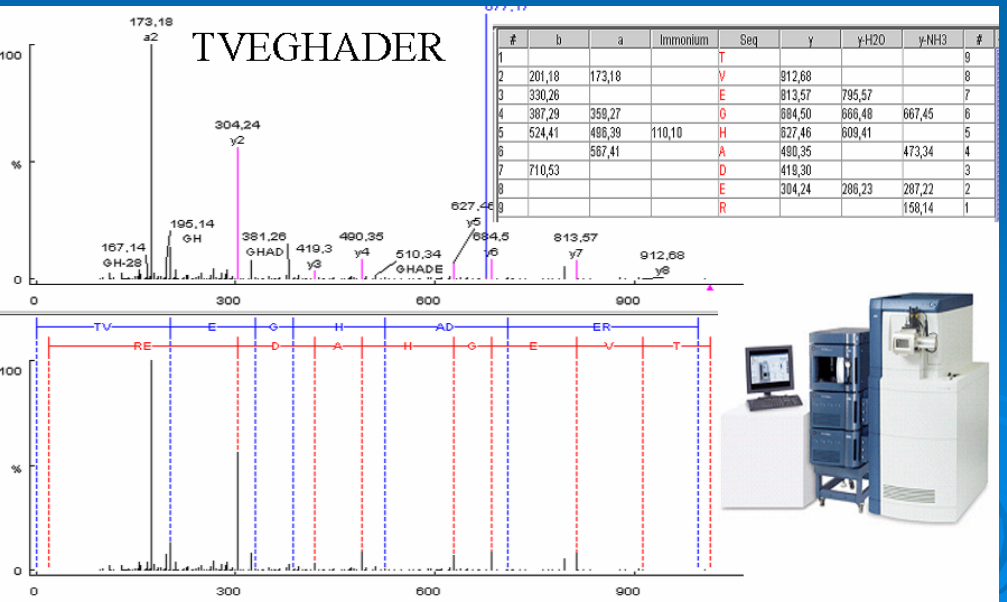
Spot picking

Tryptic digestion

Protein identification: MS



MS/MS



#	b	a	Immonium	Seq	y	y-H2O	y-NH3	#
1				T				9
2	201,18	173,18		V	912,68			8
3	330,26			E	813,57	795,57		7
4	387,29	359,27		G	684,50	666,48	667,45	6
5	524,41	496,39	110,10	H	627,46	609,41		5
6		567,41		A	490,35		473,34	4
7	710,53			D	419,30			3
8				E	304,24	286,23	287,22	2
9				R			158,14	1

MALDI-TOF Samples on slides	Q-TOF MS/MS Samples in solution
Measure mass of intact peptide ion PMF	Measure mass of peptide fragments MS/MS ions – aa sequence data
Protein ID by peptide mass fingerprint (PMF) Expensive	Protein ID by cross correlation algorithms Very expensive
Good for sequenced genomes	Good for unsequenced genomes

The proteomic challenge

- The proteome is dynamic, expression of proteins reflects its cellular state
- A given cell may express thousands to a few ten-thousands of proteins
- Large dynamic range of protein expression, at least 5 orders of magnitude
- MS-based proteomics needs to deal with a up to a million peptides in a given cell-line

Why do 2DE in proteomics

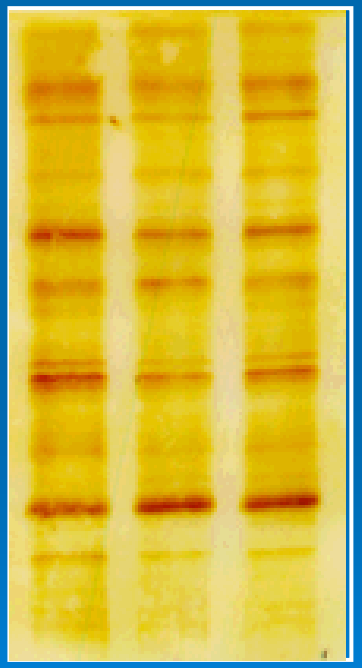
- Proteomes are highly complex, need a separation method that:
  - Is reproducible
  - Is of high resolution
  - Allows quantitative comparison between samples
  - Allows detection of PTMs

The protein sample

- Either in cells or in solution
- Removal of unwanted biomolecules
- Fractionation
  - Cytosolic proteins
  - Membrane proteins
  - Cytosceletal proteins
  - Nuclear proteins
- Protein concentration / removal of salts/detergents
  - precipitation

We do 2D PAGE to get better protein separation

I-DE



2-DE

3 ⟵ pH ⟶ 11



What is a 2D gel?

- 1st and 2nd dimension separate proteins on the basis of different ”molecular properties”

- 1st dimension = isolectriic point (pI)
  - Isoelectric focusing
- 2nd dimension = apparent molecular weight
  - SDS-PAGE

Isoelectric focusing

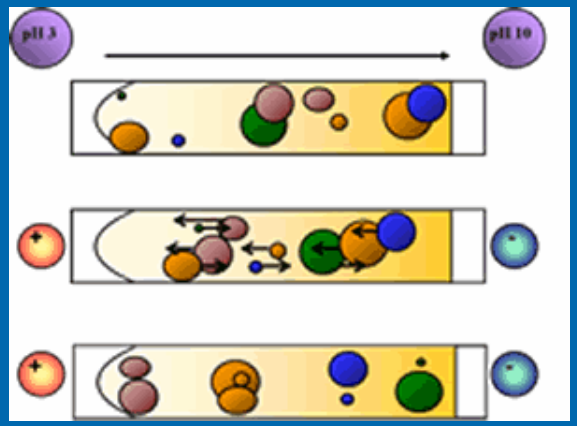
- Proteins are amphoteric, meaning that their charge varies according to the pH of their surrendings
- Now immobilised pH gradient strips (IPG strips) are available and have solved many problems

SDS-PAGE

- Proteins are coated with SDS (anionic detergent), which mask their instrinsic charge
- All proteins therfore migrate to the anode when under influence of an electric field
- Relative rates of protein migration are determined by their relative mass and the pore size of the acrylamide/cross-linker matrix

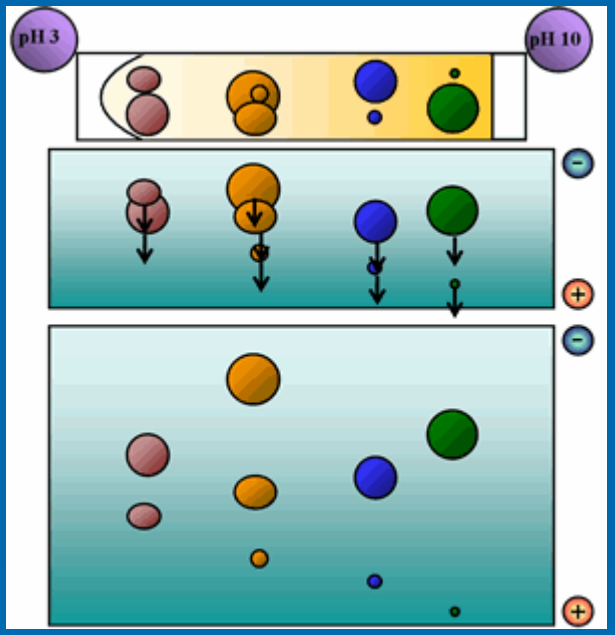
Isoelectric focusing

1st dimension

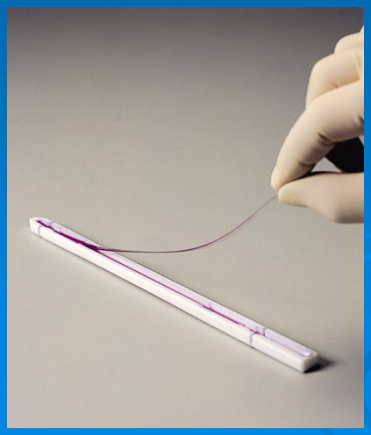


SDS-PAGE

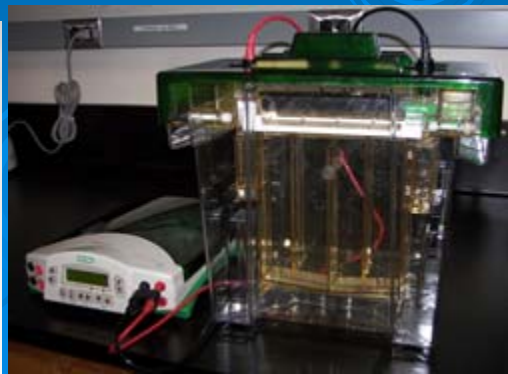
2nd dimension











Sample preparation

- General Principals:
  - Prevent protein modification
  - Remove particulate matter
  - Disrupt protein aggregates and denature completely
  - Keep salt concentration low

Samples are diluted into rehydration solution (RS)

- Most RS for IEF contain the following:
  - Chaotropic denaturant
  - Non-ionic or zwitterionic detergent
  - Reducing agent
  - Carrier ampholytes
  - Bromphenol blue

A typical solution might contain:

7M urea, 2M thiourea, 4% CHAPS, 65mM DTT, 0.8% carrier ampholytes, trace bromphenol blue

Equilibration

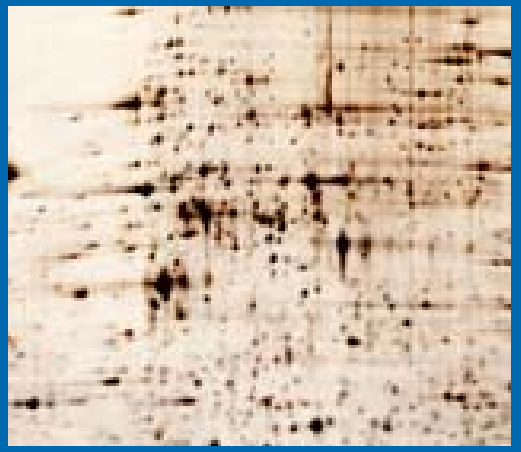
- Following IEF and before SDS-PAGE, samples are equilibrated, this serves two purposes:

1. Coating of each protein in SDS
2. Reduction and modification of cycteine residues to prevent anomalous recombination of disulphide bonds during the second dimension

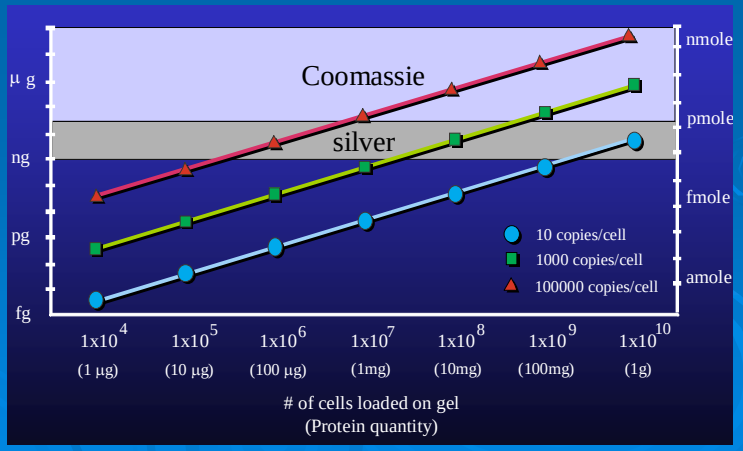
Staining

- Sensitivity
- Simlicity
- Specificity
- Speed

- Coomassie
  - Sypro Ruby
  - Silver staining
  - radioactivity



Relationship between protein copy number and cell level



Problems with 2DE

- 2DE not very compatible with:
  - Hydrophobic proteins
  - Proteins with extreme pIs
  - Low abundance proteins
  - Very large or small proteins

- Proteomics – the study of the proteome
  - The collection of proteins found in a particular cell type under a particular type of stimulation

- Protein expression profiling
  - Identify differentially expressed proteins between proteomes of the same cell to changes in the environment
  - Reveal biochemical pathways involved and affected by the parameter change
  - Assign new proteins to known pathways
  - Indicate function to a protein if previously function is unknown

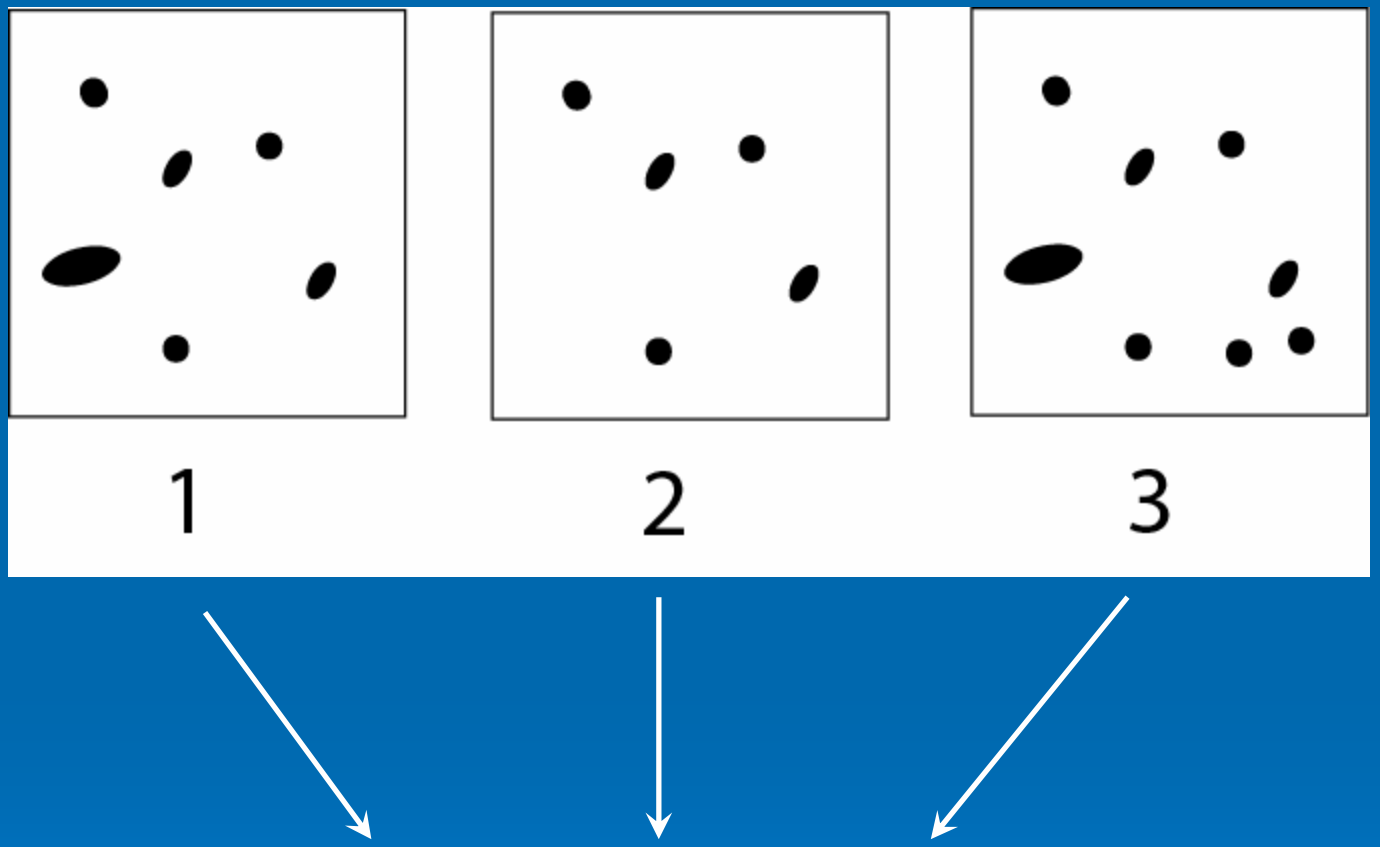
Assume any cell contains 30'000 proteins in average:

When analyzing those 30'000 proteins on a 2D-gel we see less than 1'500, *i. e.* only 5 %.

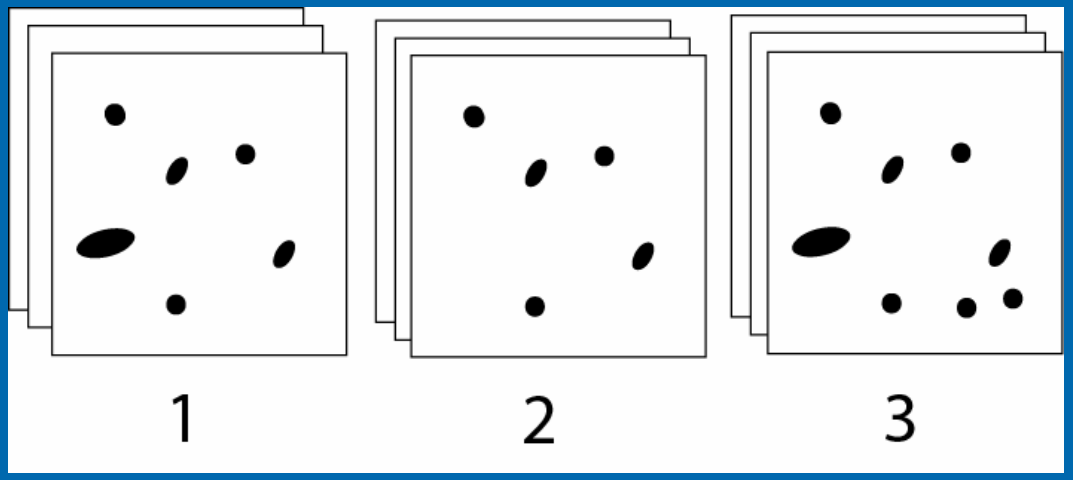
Which of the proteins do we see??

Obviously, to isolate individual specific proteins from a cell lysate, an efficient pre-enrichment step must be applied.

Problem to solve?



Compare expression profiles
Up- and down regulated



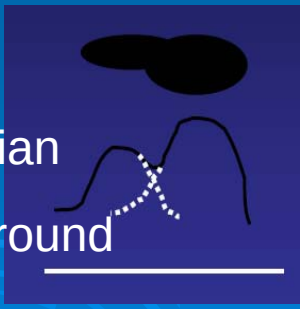
Raw gels

↓

Filtered gels (- faint spots etc)

↓

Gaussian gels (Separation of overlapping spots and accurate calculation of relative amounts of protein)



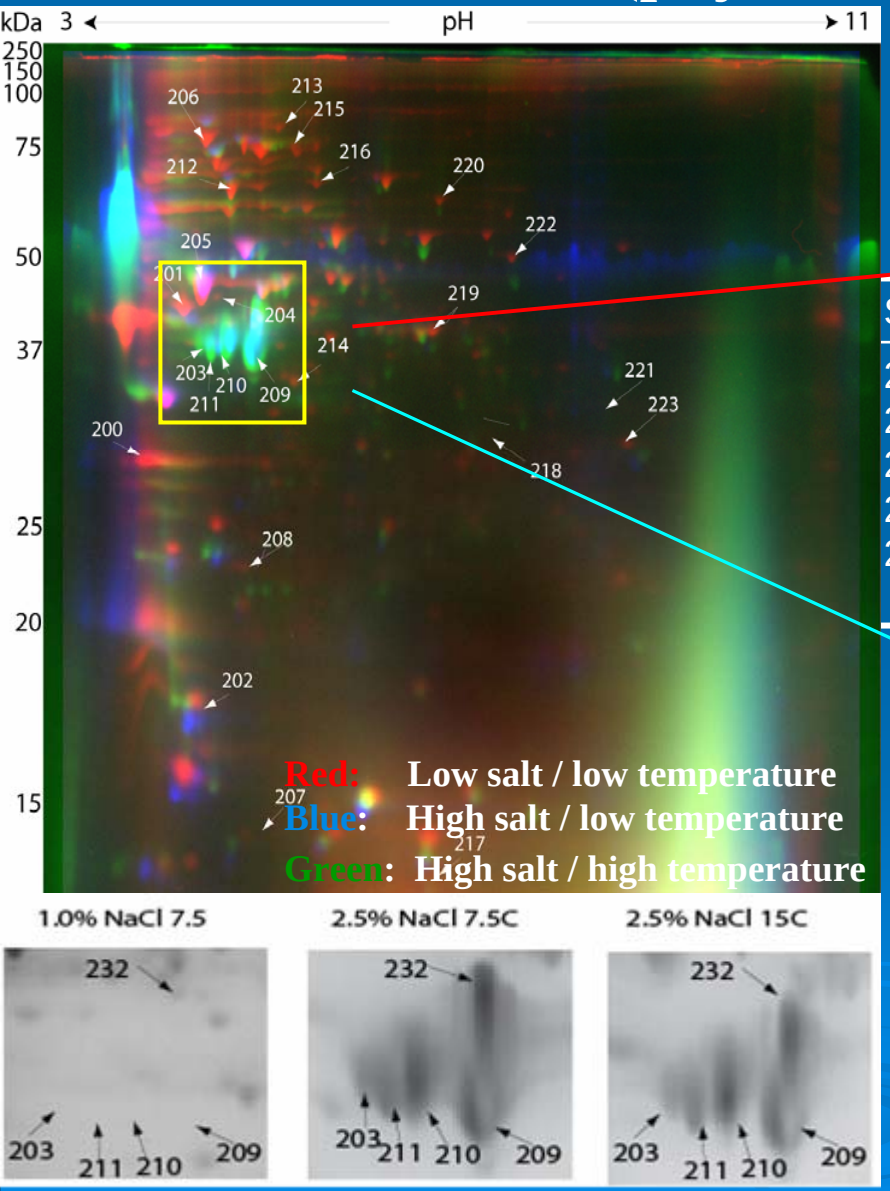
↓

Master gel (one digital gel contains information from all gels)

Functions and principles

- Image fitting and Gaussian modeling.
- Spot detection
- Normalization and removing lines and speckles
- Matching of gels (e.g. normal, diseased, and
- treated tissues)
- Analysis of data (changes in expression)
- Annotation of spots (linking of data to spots intensity and mass spectrometry data etc.)

Flagellin proteins are present at 2.5% NaCl (seawater-like), but not at 1% NaCl (physiological like) condition



Spot/ Protein
203 FlaF
209 FlaC
210 FlaE
211 FlaD
232 FlaA

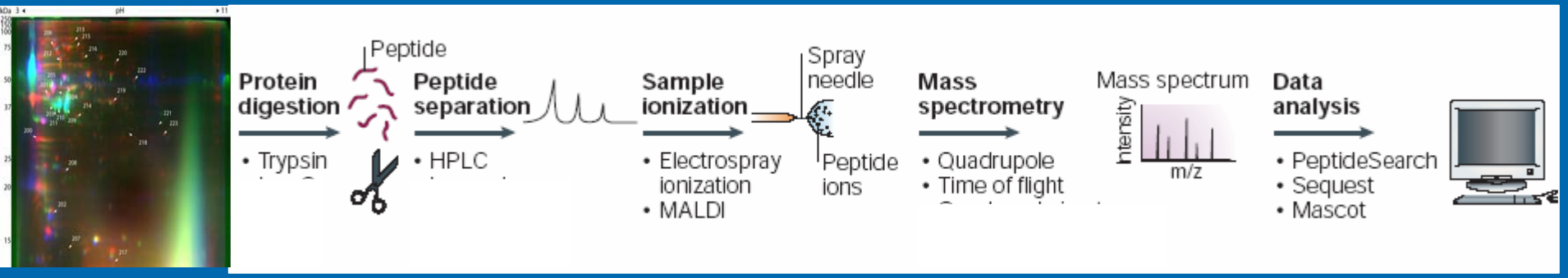


Mass Spectrometry

A mass spectrometer is an analytical device that separates molecular ions according to their mass-to-charge ratio (m/z) in the gas phase

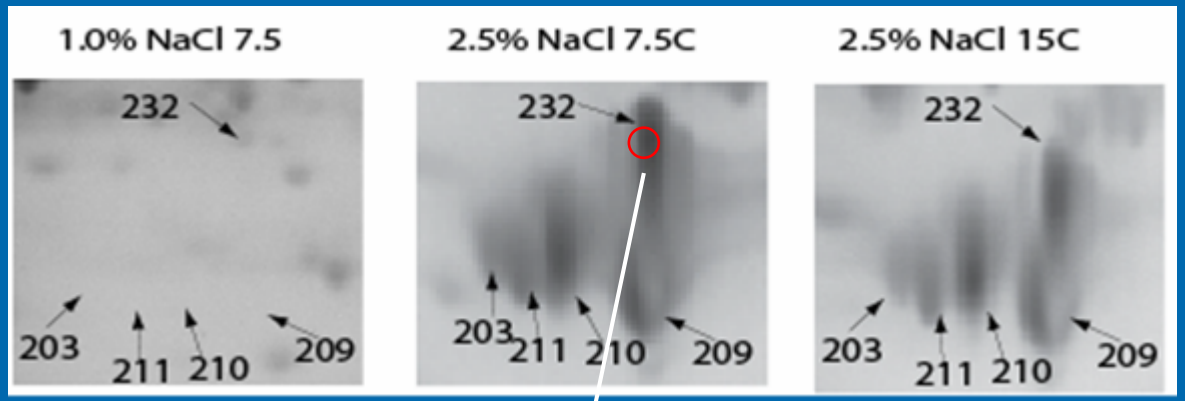
In order to detect/record an molecule using a mass spectrometer
- Ionize the molecule (Introduce a charge)
- Have the ion in the gas phase

Ions can be manipulated/directed by voltage gates



The in gel digestion protocol

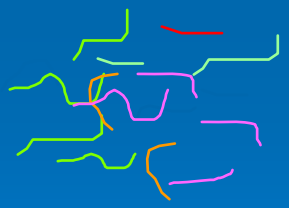
- Excise spots
  - Protect your sample against human keratin, crosscontamination, etc.
- Wash and dehydrate
  - This will remove SDS, salts, detergents, and coloring agent from the gel
  - Dehydration facilitates the absorption of active agents into the gel spots during rehydration
  - Reduce and alkylate
    - Dithiotreitol (DTT) is used to break Cys-Cys disulphid bridges
    - Iodoacetamide alkylate free cysteins
- Wash and dehydrate (again)
  - Add trypsin (night digestion)
  - Extract peptides





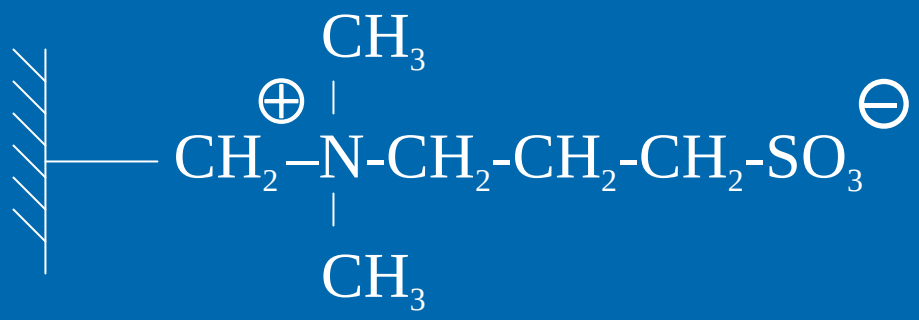
Protein

Trypsin



Peptides

Purification/desalting



- Peptides are bound in a high content of organic solvent – Elution at low organic content



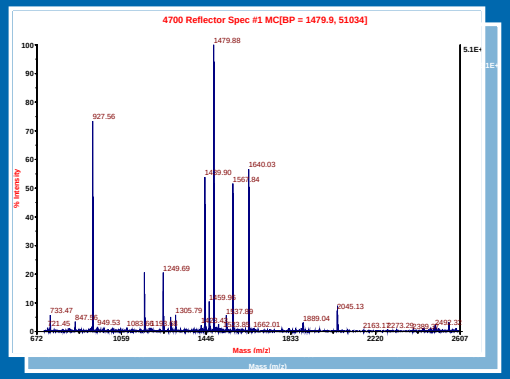
Figure 1. The ZipTip pipette tip.

Measuring peptide masses

Ionisation Techniques

Solid Samples

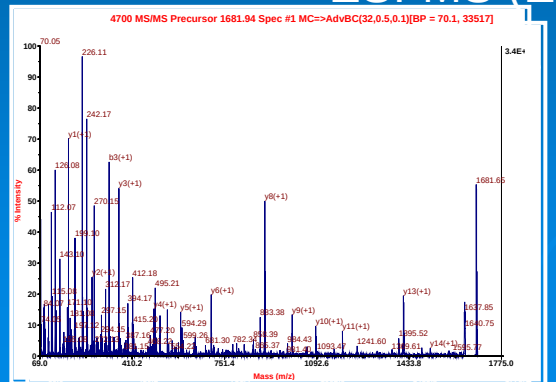
MALDI – ToF MS (Matrix Assisted Laser Desorption/Ionization – Time of Flight Mass Spectroscopy)



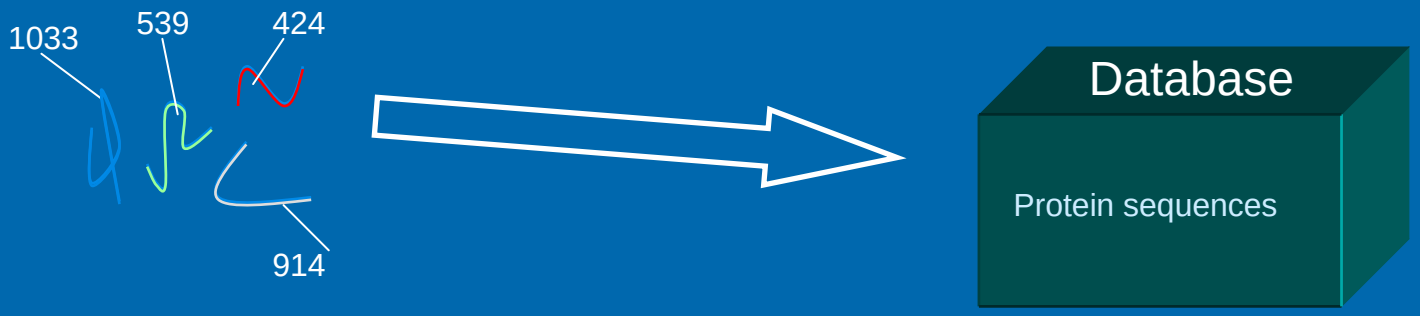
Peptide mass fingerprint (pmf)

Samples in Solution

ESI-MS (ElectroSpray Ionization – Mass Spectrometry)



Fragmentation spectra



Protein sequence

SSPEQPRPSTSKAVSPPHLDGPPSPRSPVIGSEVFLPNSNHVASGAGK

1 Predicted tryptic peptides

539 1033 424 914

SSPEQPRPSTSK AVSPPHLDGPPSPR SPVIGSR EVFLPNSNHVASGAGK

2

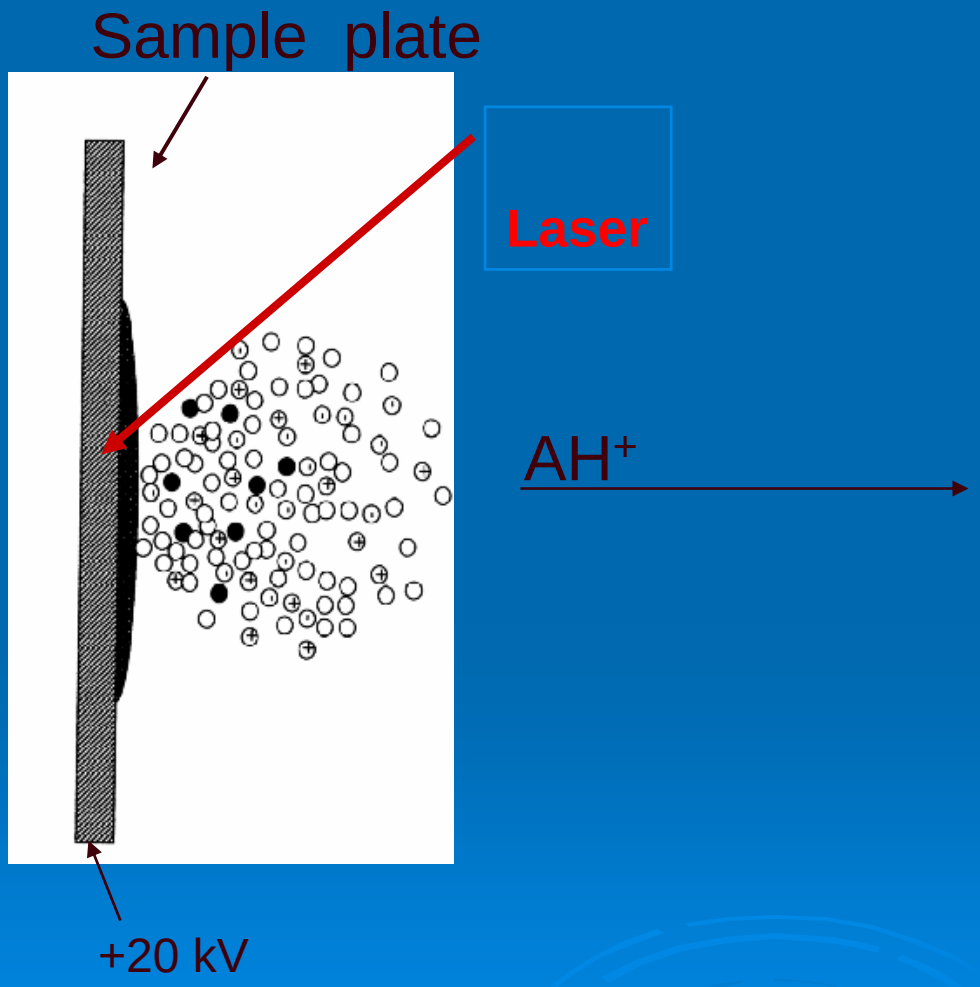


SSPEQPRPSTSK AVSPPHLDGPPSAR SPVIGSR EVFLPNSNHVASGAGK

A correlation between theoretical and the actual masses obtained in a MS experiment would provide very convincing evidence for a database match

MALDI: Matrix Assisted Laser Desorption Ionization

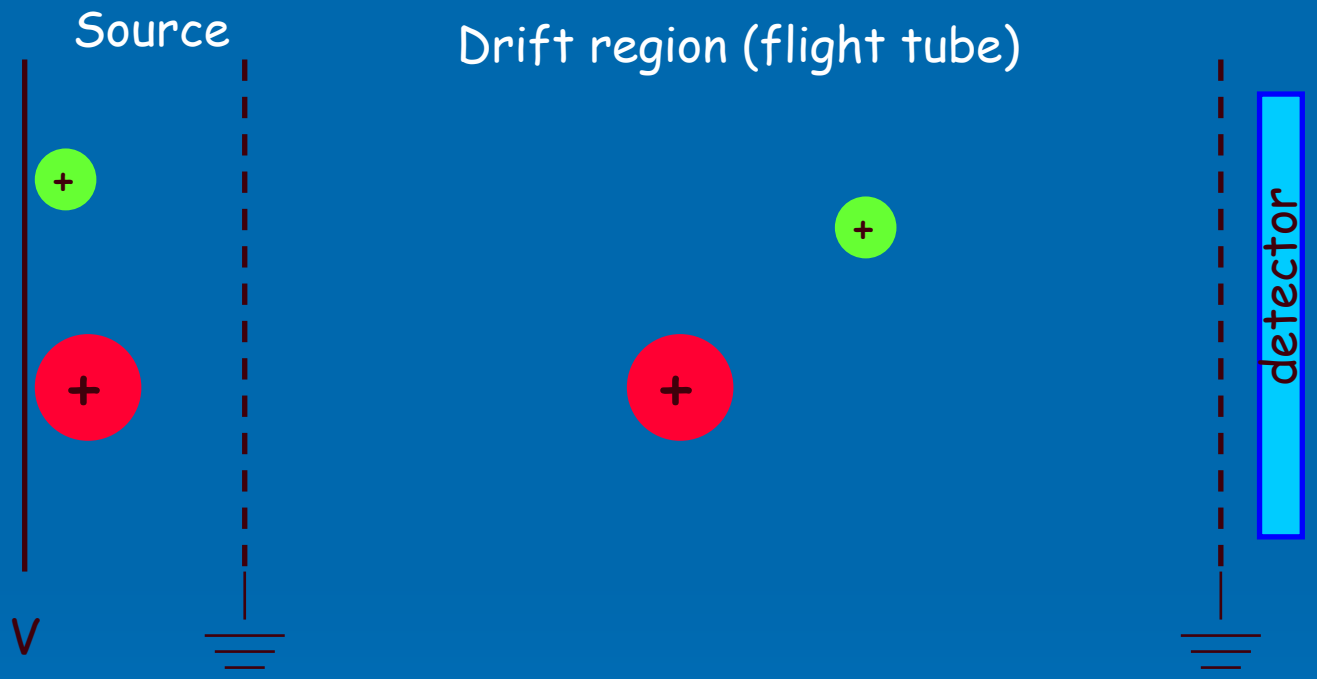
Molecules are brought from a surface into the gas phase ***(desorbed)*** and ***ionised*** at the same time



1. Sample is mixed with excess matrix and dried on a MALDI plate.
2. The matrix absorbs energy and therefore protects the analyte
3. Sample molecules are ionized by proton transfer from matrix:

   $MH^+ + A \rightarrow M + AH^+$.

Instrument Overview

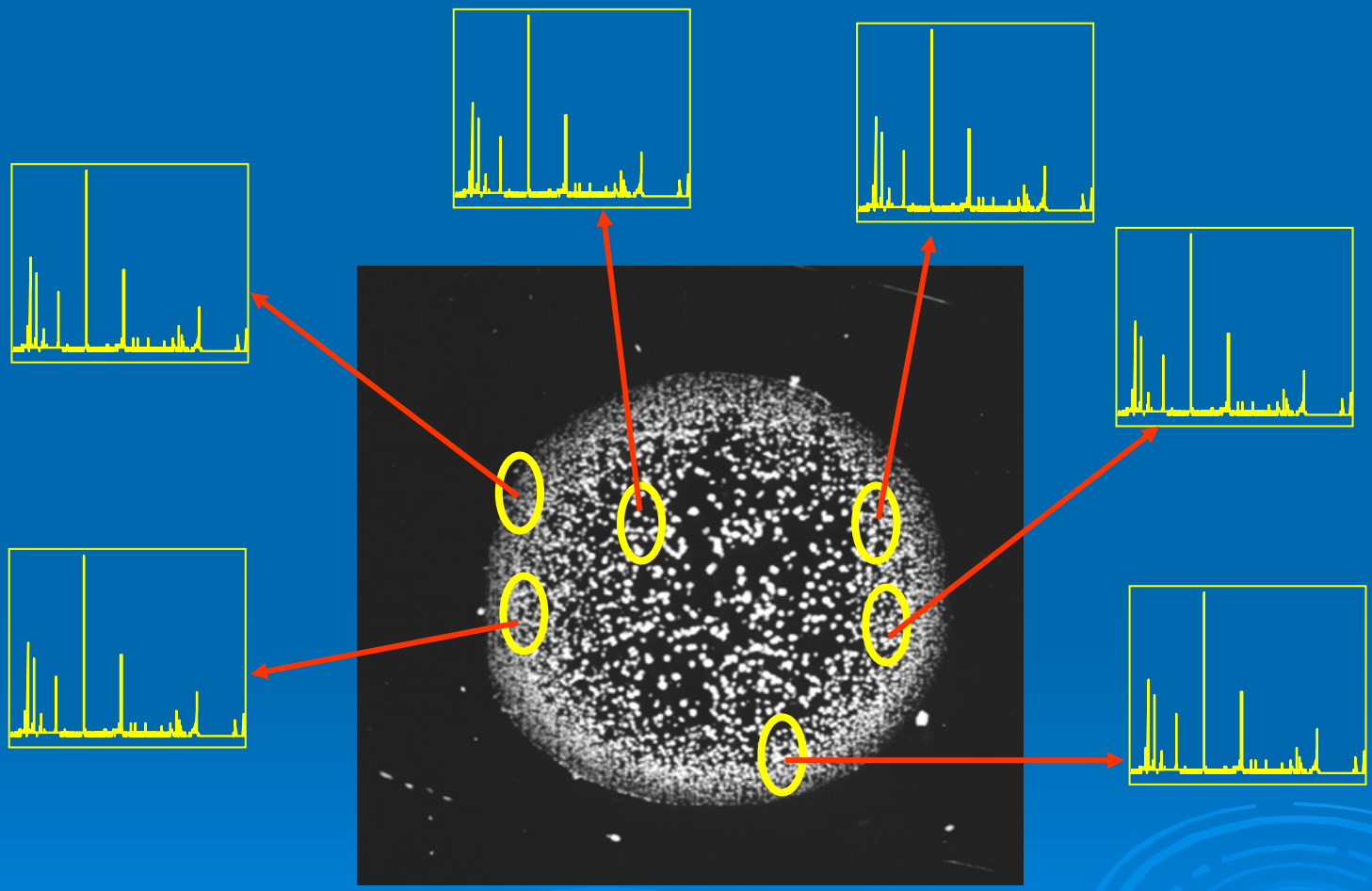


Ions are accelerated in pulses

Measures the time for ions to reach the detector

Small ions reach the detector before large ones

Reproducibility

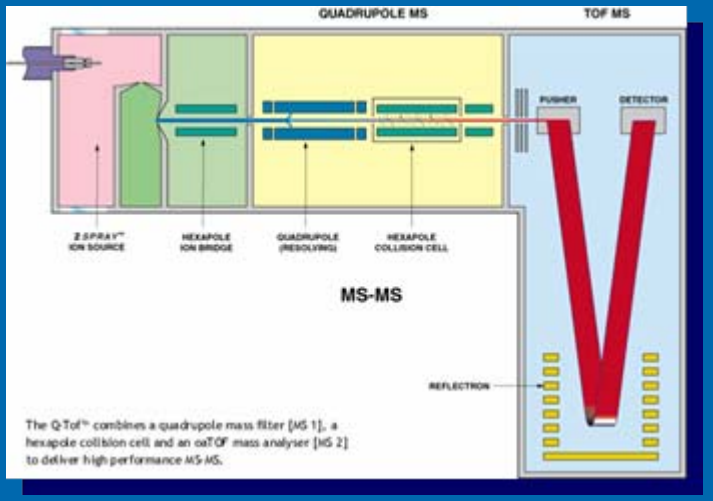


CHCA

PMF – Peptide Mass Fingerprinting

- One can now search with the obtained masses in protein databases via the Mascot search engine. A comparison is then done between the tryptic profile of proteins in the databases and peptide profile obtained from the sample.

Electrospray ionization (ESI)



Electrosprayed
'aerosol'

Liquid sample

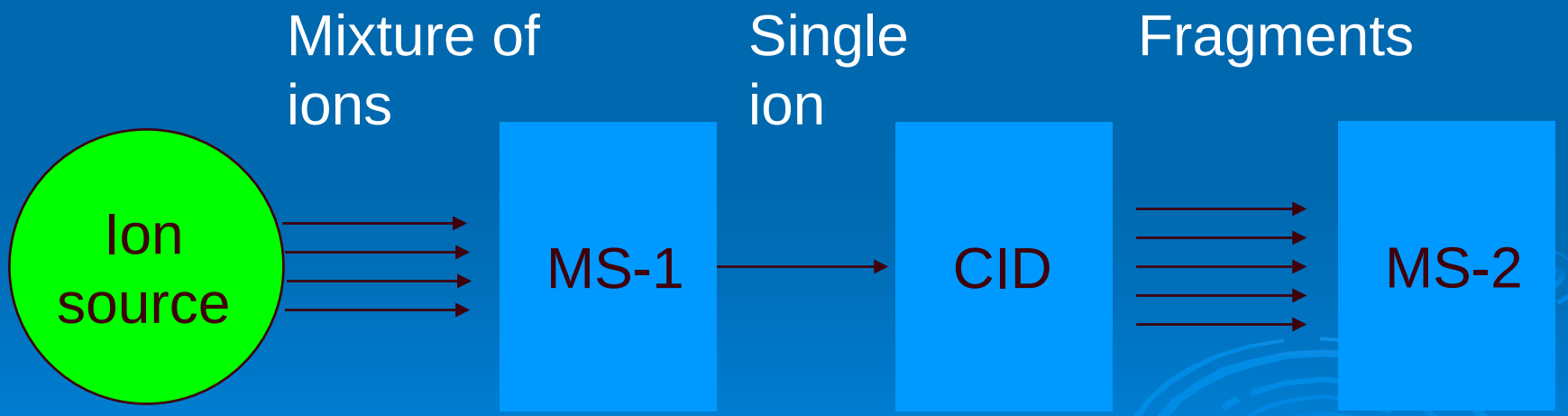
Mass spectrometer

1-3 kV
needle potential

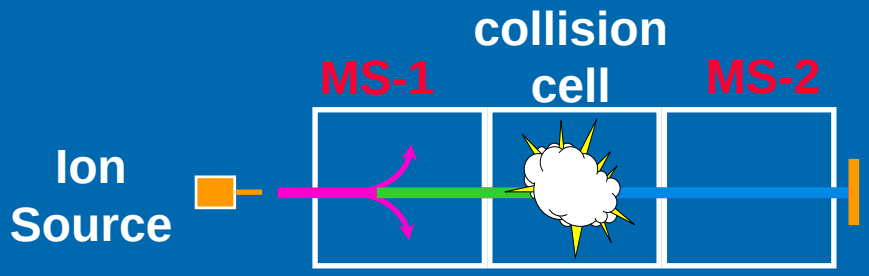
Gas-phase ions

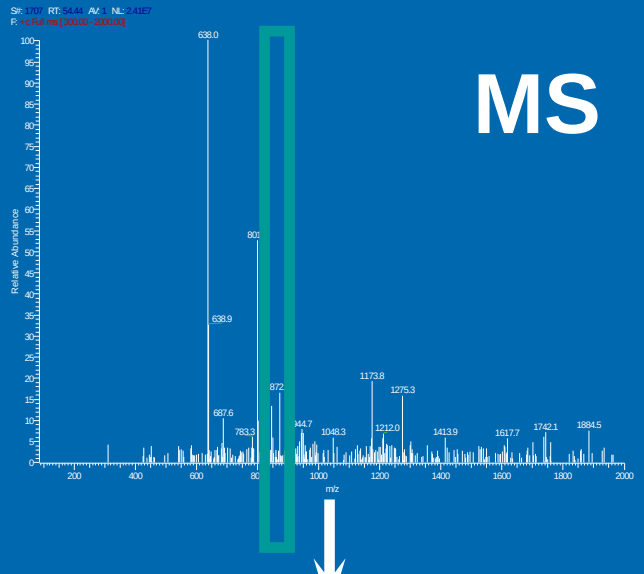
What is MSMS?

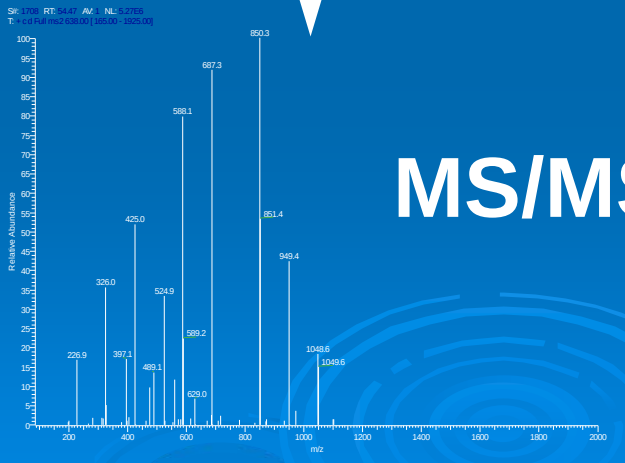
MS/MS means using two mass analyzers (combined in one instrument) to select an analyte (ion) from a mixture, then generate fragments from it to give structural information



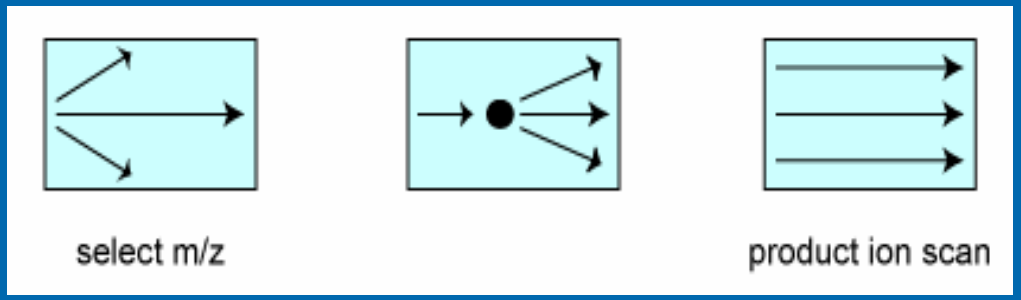
Tandem Mass Spectrometry

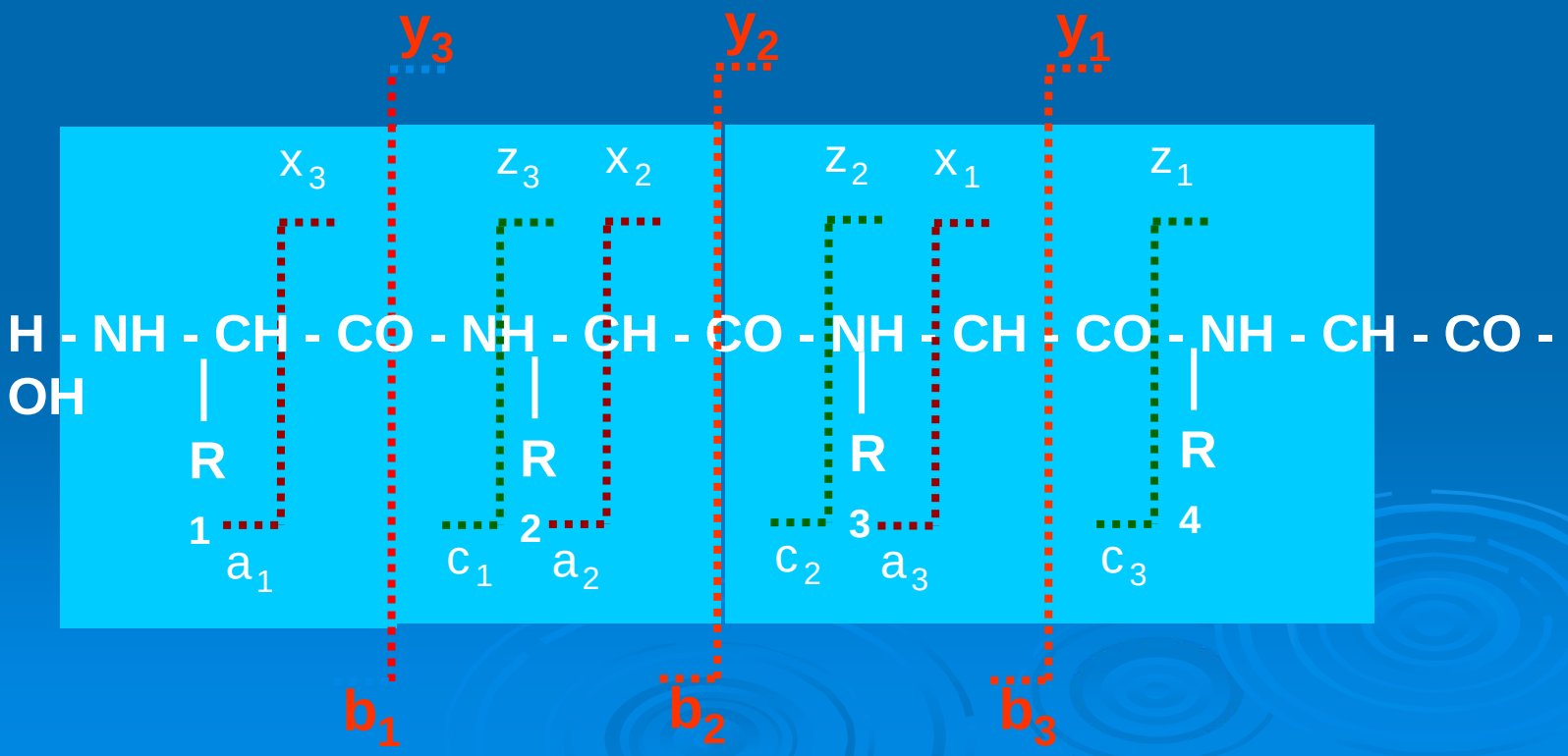


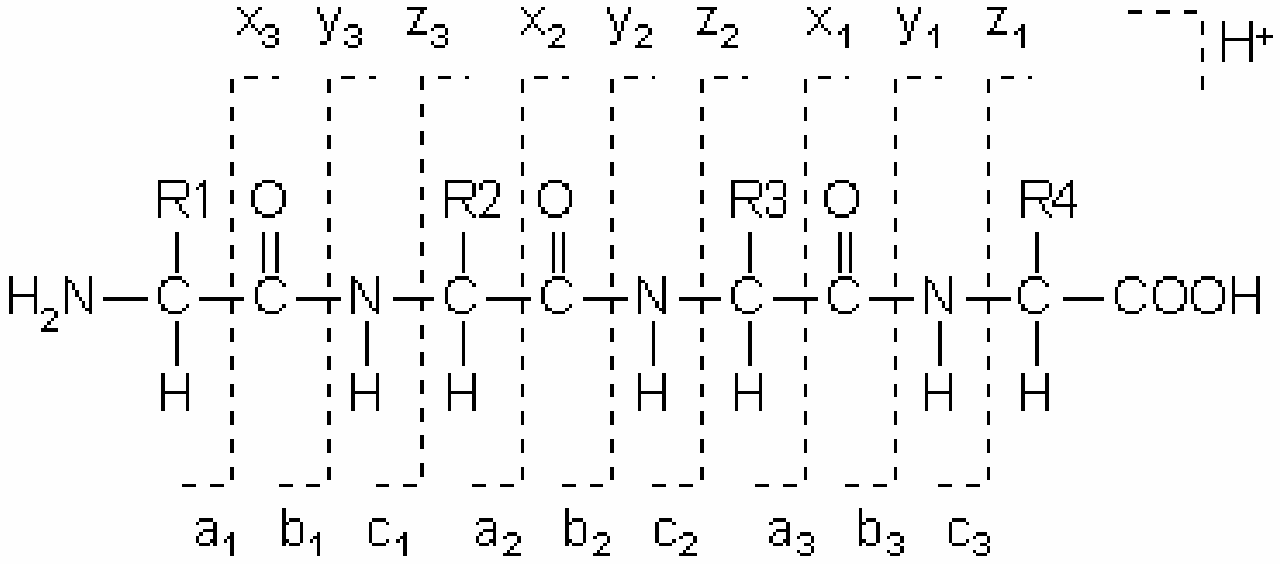




MS/MS


select m/z
product ion scan


y3
y2
y1
x3
z3
x2
z2
x1
z1
H - NH - CH - CO - NH - CH - CO - NH - CH - CO - NH - CH - CO - OH
R
1
R
2
R
3
R
4
a1
c1
a2
c2
a3
c3
b1
b2
b3



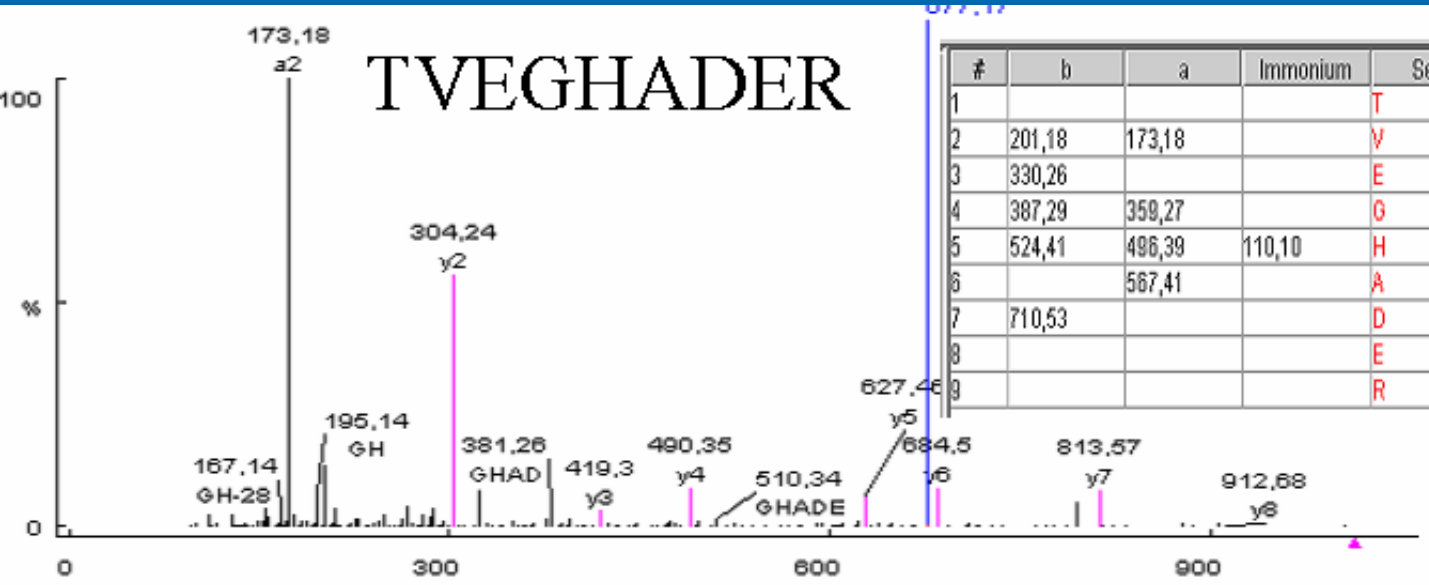
a, b and c fragments carry the charge at the N-terminus

x, y and z fragments carry the charge at the C-terminus

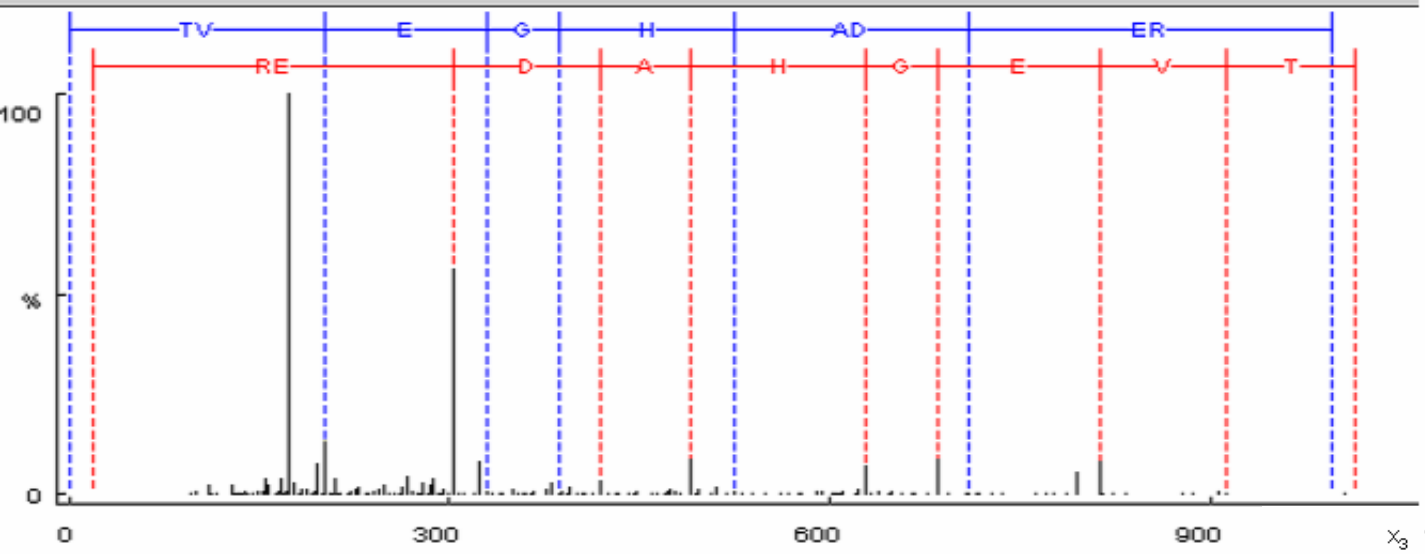
Softwares are used to interprete spectrums!

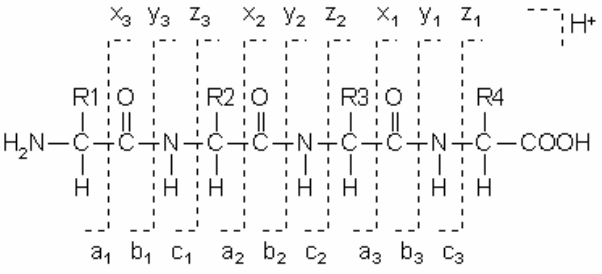
Ion Type	Ion Mass
a	[N]+[M]-CO
a*	a-NH_3
a°	a-H_2O
a^{++}	(a+H)/2
b	[N]+[M]
b*	b-NH_3
b°	b-H_2O
b^{++}	(b+H)/2
c	[N]+[M]+NH_3
d	a-partial side chain
v	y-complete side chain
w	z-partial side chain
x	[C]+[M]+CO
y	[C]+[M]+H_2
y*	y-NH_3
y°	y-H_2O
y^{++}	(y+H)/2
z	[C]+[M]-NH

TVEGHADER



#	b	a	Immonium	Seq	y	y-H2O	y-NH3	#
1				T				9
2	201,18	173,18		V	912,68			8
3	330,26			E	813,57	795,57		7
4	387,29	359,27		G	684,50	666,48	667,45	6
5	524,41	496,39	110,10	H	627,46	609,41		5
6		567,41		A	490,35		473,34	4
7	710,53			D	419,30			3
8				E	304,24	286,23	287,22	2
9				R			158,14	1

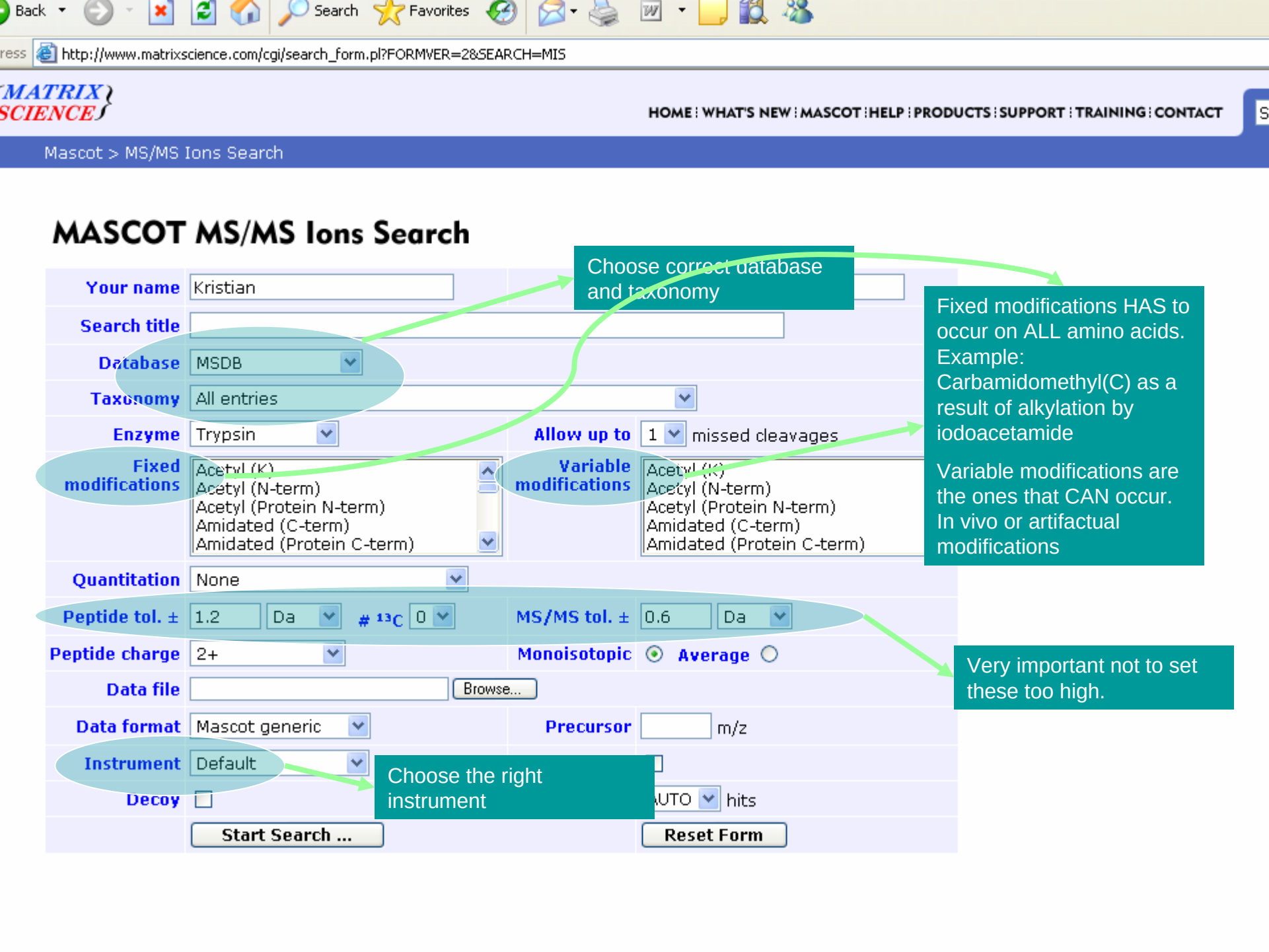




Using Mascot to search your data

Freely available online

- A version of the search-software Mascot is available online, at www.matrixscience.com
- Proteomics labs typically have their own license, and own servers
- Mascot is the most widely used program for searching MS/MS data. Can also be used for PMF


Back
Search
Favorites
http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS
MATRIX SCIENCE
HOME | WHAT'S NEW | MASCOT | HELP | PRODUCTS | SUPPORT | TRAINING | CONTACT
Mascot > MS/MS Ions Search
MASCOT MS/MS Ions Search
Your name
Kristian
Search title
Database
MSDB
Taxonomy
All entries
Enzyme
Trypsin
Allow up to
1
missed cleavages
Fixed modifications
Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Variable modifications
Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Quantitation
None
Peptide tol. ±
1.2
Da
13C
0
MS/MS tol. ±
0.6
Da
Peptide charge
2+
Monoisotopic
Average
Data file
Browse...
Data format
Mascot generic
Precursor
m/z
Instrument
Default
Decoy
AUTO
hits
Start Search ...
Reset Form
Choose correct database and taxonomy
Fixed modifications HAS to occur on ALL amino acids. Example: Carbamidomethyl(C) as a result of alkylation by iodoacetamide
Variable modifications are the ones that CAN occur. In vivo or artifactual modifications
Very important not to set these too high.
Choose the right instrument

How does Mascot work?

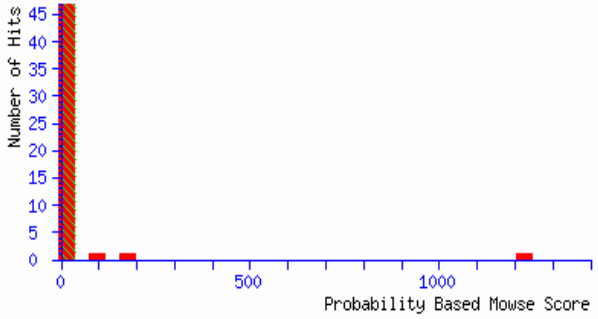
- The user enters a set of input spectra (OR only one) and search parameters.
- Mascot then presents the user with a list of hits. For MS/MS searching, Mascot identifies the *peptides* but organize them according to proteins.
- Each peptide has a score (numeric value). For example 45. Mascot supplies an *identity threshold.* IF the score is above the threshold, the score of this peptide is statistically significant.

Mascot Search Results

```
User            :
Email           :
Search title    : MS/MS Example
MS data file    : C:\Auto MSMS output\Sample 1.pkl
Database        : SwissProt 51.6 (257964 sequences; 93947433 residues)
Timestamp       : 19 Feb 2007 at 14:09:47 GMT
Protein hits    : CH60_HUMAN 60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat sho
                  CH60_DROME 60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat sho
                  CH60_CAEEL Chaperonin homolog Hsp-60, mitochondrial precursor (Heat shock protein 60) (HSP-60) - Caenorhabd
                  CH60_XANAC 60 kDa chaperonin (Protein Cpn60) (groEL protein) - Xanthomonas axonopodis pv. citri
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 38 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As: Peptide Summary			Help
	Significance threshold p< 0.05	Max. number of hits AUTO	
	Standard scoring ◉ MudPIT scoring ○	Ions score or expect cut-off 0	Show sub-sets 0
	Show pop-ups ◉ Suppress pop-ups ○	Sort unassigned Decreasing Score	Require bold red ☐

Select All | Select None | Search Selected

Peptide Summary Report (MS/MS Example) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.matrixscience.com/cgi/master_results.pl?file=../data/F981123.dat
Go Links

Select All | Select None | Search Selected | Error tolerant

1. CH60_HUMAN **Mass:** 61016 **Score:** 1225 **Queries matched:** 31

60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock

Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
11	417.1822	832.3498	832.3828	-0.0329	0	45	0.016	1	K.APGFGDNR.K
12	422.7433	843.4720	843.5066	-0.0346	0	46	0.017	1	K.VGEVIVTK.D
13	430.7328	859.4510	859.4837	-0.0327	0	36	0.15	1	K.IPAMTIAK.N + Oxidation (M)
15	451.2499	900.4853	900.5280	-0.0428	0	52	0.0039	1	K.LSDGVAVLK.V
16	456.7806	911.5467	911.5804	-0.0337	0	59	0.00056	1	K.VGLQVVAVK.A
21	480.7447	959.4748	959.5036	-0.0288	0	45	0.017	1	R.VTDALNATR.A
24	595.7855	1189.5565	1189.6012	-0.0447	0	(57)	0.0011	1	K.EIGNIISDAMK.K
25	603.7720	1205.5294	1205.5962	-0.0668	0	60	0.00048	1	K.EIGNIISDAMK.K + Oxidation (M)
26	608.3099	1214.6052	1214.6507	-0.0455	0	73	2.2e-05	1	K.NAGVEGSLIVEK.I
27	617.2857	1232.5569	1232.5885	-0.0316	0	81	4e-06	1	K.VGGTSDVEVNEK.K
31	672.8375	1343.6605	1343.7085	-0.0480	0	64	0.00016	1	R.TVIIEQSWGSPK.V
34	714.8884	1427.7623	1427.8058	-0.0435	0	(65)	0.00014	1	R.GVMLAVDAVIAELK.K
35	714.8938	1427.7730	1427.8058	-0.0327	0	(73)	2.1e-05	1	R.GVMLAVDAVIAELK.K
36	722.8849	1443.7552	1443.8007	-0.0455	0	75	1.2e-05	1	R.GVMLAVDAVIAELK.K + Oxidation (M)
37	722.8934	1443.7722	1443.8007	-0.0285	0	(73)	2.2e-05	1	R.GVMLAVDAVIAELK.K + Oxidation (M)
39	752.8643	1503.7141	1503.7490	-0.0349	0	90	4.3e-07	1	K.TLNDELEIIEGMK.F
40	760.8461	1519.6777	1519.7439	-0.0662	0	(89)	4.7e-07	1	K.TLNDELEIIEGMK.F + Oxidation (M)
45	640.3281	1917.9625	1918.0636	-0.1010	0	102	2.1e-08	1	K.ISSIQSIVPALEIANAHR.K
46	960.0327	1918.0509	1918.0636	-0.0127	0	(87)	5.1e-07	1	K.ISSIQSIVPALEIANAHR.K
48	1019.5106	2037.0067	2037.0153	-0.0087	0	52	0.0015	1	R.IQEIIEQLDVTTSEYEK.E
51	1057.0537	2112.0929	2112.1323	-0.0394	0	116	6.8e-10	1	R.ALMLQGVDLLADAVAVTMGPK.G
52	1065.0399	2128.0653	2128.1272	-0.0619	0	(72)	1.7e-05	1	R.ALMLQGVDLLADAVAVTMGPK.G + Oxidation (M)
53	1065.0623	2128.1100	2128.1272	-0.0172	0	(26)	0.63	1	R.ALMLQGVDLLADAVAVTMGPK.G + Oxidation (M)
54	1073.0477	2144.0809	2144.1221	-0.0412	0	(93)	1.3e-07	1	R.ALMLQGVDLLADAVAVTMGPK.G + 2 Oxidation (M)
58	789.1062	2364.2968	2364.3264	-0.0296	0	(56)	0.00052	1	R.KPLVIIAEDVDGEALSTLVLNR.L
59	1183.1570	2364.2994	2364.3264	-0.0270	0	(65)	5.3e-05	1	R.KPLVIIAEDVDGEALSTLVLNR.L
60	789.1094	2364.3063	2364.3264	-0.0201	0	95	6.3e-08	1	R.KPLVIIAEDVDGEALSTLVLNR.L
61	828.1238	2481.3495	2481.3942	-0.0447	0	(26)	0.42	1	R.TALLDAAGVASLLTAEVVVTEIPK.E
62	828.1322	2481.3748	2481.3942	-0.0194	0	48	0.003	1	R.TALLDAAGVASLLTAEVVVTEIPK.E
64	854.0588	2559.1545	2559.2413	-0.0868	0	75	5.1e-06	1	K.LVQDVANNTNEEAGDGTTTATVLAR.S
65	1038.5031	3112.4873	3112.5023	-0.0150	0	13	6.7	1	K.DMAIATGGAVFGEEGLTLNLEDVQPHDLGK.V + Oxidation (M)

2. CH60_DROME **Mass:** 60771 **Score:** 174 **Queries matched:** 4

60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock

Check to include this hit in error tolerant search

Done Internet
start Microsoft PowerPoint ... https://proxy-gw.pri... Peptide Summary Re... http://www.bioinfo.n... [CoCoozo] Help - Moz... EN 10:57

What identifies a protein

- For MS/MS data, at least one peptide must be identified with a significant threshold.
- Often, people (journals etc.) require two peptides per protein. Depends on the application.

MATRIX SCIENCE Mascot Search Results

Protein View

Match to: **CH60_HUMAN** Score: **1225**
60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock
Found in search of C:\Auto MSMS output\Sample 1.pkl

Nominal mass (M_r): **61016**; Calculated pI value: **5.70**
NCBI BLAST search of CH60_HUMAN against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Homo sapiens

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 MLRLPTVFRQ MRPVSRVLAP HLTRAYAKDV KFGADARALM LQGVDLLADA
    51 VAVTMGPKGR TVIIEQSWGS PKVTKDGVTV AKSIDLKDKY KNIGAKLVQD
   101 VANNTNEEAG DGTTTATVLA RSIAKEGFEK ISKGANPVEI RRGVMLAVDA
   151 VIAELKKQSK PVTTPEEIAQ VATISANGDK EIGNIISDAM KKVGRKGVIT
   201 VKDGKTLNDE LEIIEGMKFD RGYISPYFIN TSKGQKCEFQ DAYVLLSEKK
   251 ISSIQSIVPA LEIANAHRKP LVIIAEDVDG EALSTLVLNR LKVGLQVVAV
   301 KAPGFGDNRK NQLKDMAIAT GGAVFGEEGL TLNLEDVQPH DLGKVGEVIV
   351 TKDDAMLLKG KGDKAQIEKR IQEIIEQLDV TTSEYEKEKL NERLAKLSDG
   401 VAVLKVGGTS DVEVNEKKDR VTDALNATRA AVEEGIVLGG GCALLRCIPA
   451 LDSLTPANED QKIGIEIIKR TLKIPAMTIA KNAGVEGSLI VEKIMQSSSE
   501 VGYDAMAGDF VNMVEKGIID PTKVVRTALL DAAGVASLLT TAEVVVTEIP
   551 KEEKDPGMGA MGGMGGGMGG GMF
```

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
38 - 58	1057.0537	2112.0929	2112.1323	-0.0394	0	R.ALMLQGVDLLADAVAVTMGPK.G (Ions score 116)
38 - 58	1065.0399	2128.0653	2128.1272	-0.0619	0	R.ALMLQGVDLLADAVAVTMGPK.G Oxidation (M) (Ions score 72)
38 - 58	1065.0623	2128.1100	2128.1272	-0.0172	0	R.ALMLQGVDLLADAVAVTMGPK.G Oxidation (M) (Ions score 26)
38 - 58	1073.0477	2144.0809	2144.1221	-0.0412	0	R.ALMLQGVDLLADAVAVTMGPK.G 2 Oxidation (M) (Ions score 93)
61 - 72	672.8375	1343.6605	1343.7085	-0.0480	0	R.TVIIEQSWGSPK.V (Ions score 64)
97 - 121	854.0588	2559.1545	2559.2413	-0.0868	0	K.LVQDVANNTNEEAGDGTTTATVLAR.S (Ions score 75)
143 - 156	714.8884	1427.7623	1427.8058	-0.0435	0	R.GVMLAVDAVIAELK.K (Ions score 65)
143 - 156	714.8938	1427.7730	1427.8058	-0.0327	0	R.GVMLAVDAVIAELK.K (Ions score 73)
143 - 156	722.8849	1443.7552	1443.8007	-0.0455	0	R.GVMLAVDAVIAELK.K Oxidation (M) (Ions score 75)
143 - 156	722.8934	1443.7722	1443.8007	-0.0285	0	R.GVMLAVDAVIAELK.K Oxidation (M) (Ions score 73)
181 - 191	595.7855	1189.5565	1189.6012	-0.0447	0	K.EIGNIISDAMK.K (Ions score 57)
181 - 191	603.7720	1205.5294	1205.5962	-0.0668	0	K.EIGNIISDAMK.K Oxidation (M) (Ions score 60)
206 - 218	752.8643	1503.7141	1503.7490	-0.0349	0	K.TLNDELEIIEGMK.F (Ions score 90)
206 - 218	760.8461	1519.6777	1519.7439	-0.0662	0	K.TLNDELEIIEGMK.F Oxidation (M) (Ions score 89)
251 - 268	640.3281	1917.9625	1918.0636	-0.1010	0	K.ISSIQSIVPALEIANAHR.K (Ions score 102)
251 - 268	960.0327	1918.0509	1918.0636	-0.0127	0	K.ISSIQSIVPALEIANAHR.K (Ions score 87)
269 - 290	789.1062	2364.2968	2364.3264	-0.0296	0	R.KPLVIIAEDVDGEALSTLVLNR.L (Ions score 56)
269 - 290	1183.1570	2364.2994	2364.3264	-0.0270	0	R.KPLVIIAEDVDGEALSTLVLNR.L (Ions score 65)
269 - 290	789.1094	2364.3063	2364.3264	-0.0201	0	R.KPLVIIAEDVDGEALSTLVLNR.L (Ions score 95)
293 - 301	456.7806	911.5467	911.5804	-0.0337	0	K.VGLQVVAVK.A (Ions score 59)
302 - 309	417.1822	832.3498	832.3828	-0.0329	0	K.APGFGDNR.K (Ions score 45)

A single peptide identification

Mascot Search Results: Peptide View - Microsoft Internet Explorer

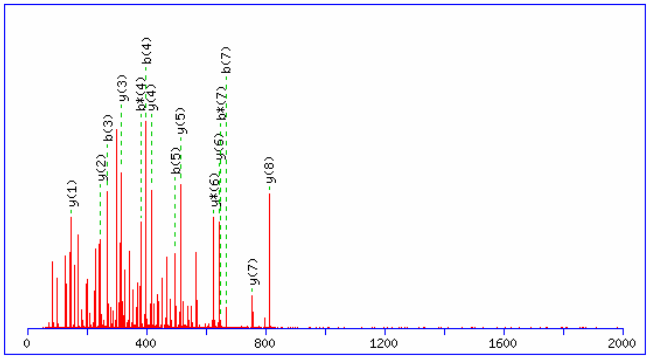
File Edit View Favorites Tools Help

Back Search Favorites

Address http://www.matrixscience.com/cgi/peptide_view.pl?file=../data/F981123.dat&query=16&hit=1&index=CH60_HUMAN&px=1&section=5

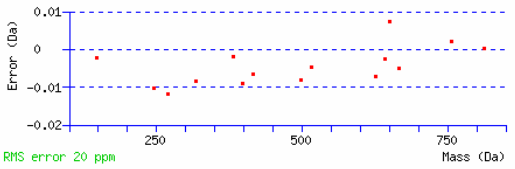
Go Links

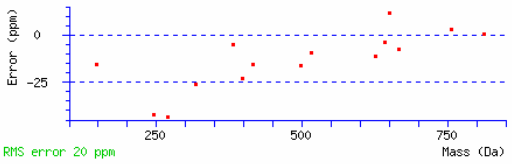
Or, Plot from 0 to 2000 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 911.5804
Ions Score: 59 **Expect:** 0.00056
Matches (Bold Red): 15/58 fragment ions using 109 most intense peaks

#	b	b^{++}	b*	b*$^{++}$	Seq.	y	y^{++}	y*	y*$^{++}$	#
1	100.0757	50.5415			V					9
2	157.0972	79.0522			G	**813.5193**	407.2633	796.4927	398.7500	8
3	**270.1812**	135.5942			L	**756.4978**	378.7525	739.4713	370.2393	7
4	**398.2398**	199.6235	**381.2132**	191.1103	Q	**643.4137**	322.2105	**626.3872**	313.6972	6
5	**497.3082**	249.1577	480.2817	240.6445	V	**515.3552**	258.1812	498.3286	249.6679	5
6	596.3766	298.6920	579.3501	290.1787	V	**416.2867**	208.6470	399.2602	200.1337	4
7	**667.4137**	334.2105	**650.3872**	325.6972	A	**317.2183**	159.1128	300.1918	150.5995	3
8	766.4822	383.7447	749.4556	375.2314	V	**246.1812**	123.5942	229.1547	115.0810	2
9					K	**147.1128**	74.0600	130.0863	65.5468	1





Done Internet

- peptide mass fingerprinting is limited to the identification of proteins for which sequences are already know
- If you are working with an organism which genome is not sequenced you would probably get more significant results from MS/MS
  - *de novo* sequencing

You have high quality MS data, but cannot get a high-confidence hit

- Protein may be absent from the database
- Unknown or unanticipated post-translational modifications
- Nonspecific proteolysis
- Contaminating proteins

False identifications

- If the user submit 100 MS/MS spectra, and 30 of these are identified, are any of these wrong?
  - Yes. Most likely. The standard error rate for Mascot is estimated to be 5% (at the 95% confidence setting). This can be changed.
  - So, from the 30 identified spectra, 30x0.05=1.5 false hits can be expected.
- In practice the real number of false hits is often estimated using *decoy* databases. The same spectra are searched against a “normal” and a “randomized” database.

Some guidelines

- Choice of database:
  - Prefer small databases when possible.
- Modifications:
  - Add the relevant ones, take care not to include too many (un-necessary) variable modifications.
- Reason:
  - Mascot`s identity threshold *increase* with larger databases and more variable modifications