VU medisch centrum

ESC Summer School Cardiovascular Sciences June 15, 2011

PROTEOMICS

Prof dr Anton J.G. Horrevoets

Molecular Cell Biology and Immunology

VU Medical Center Amsterdam

Institute for Cardiovascular Research

ICaR-VU

vrije Universiteit amsterdam

http://genomebiology.com/2001/2/11/comment/2008.1

Opinion

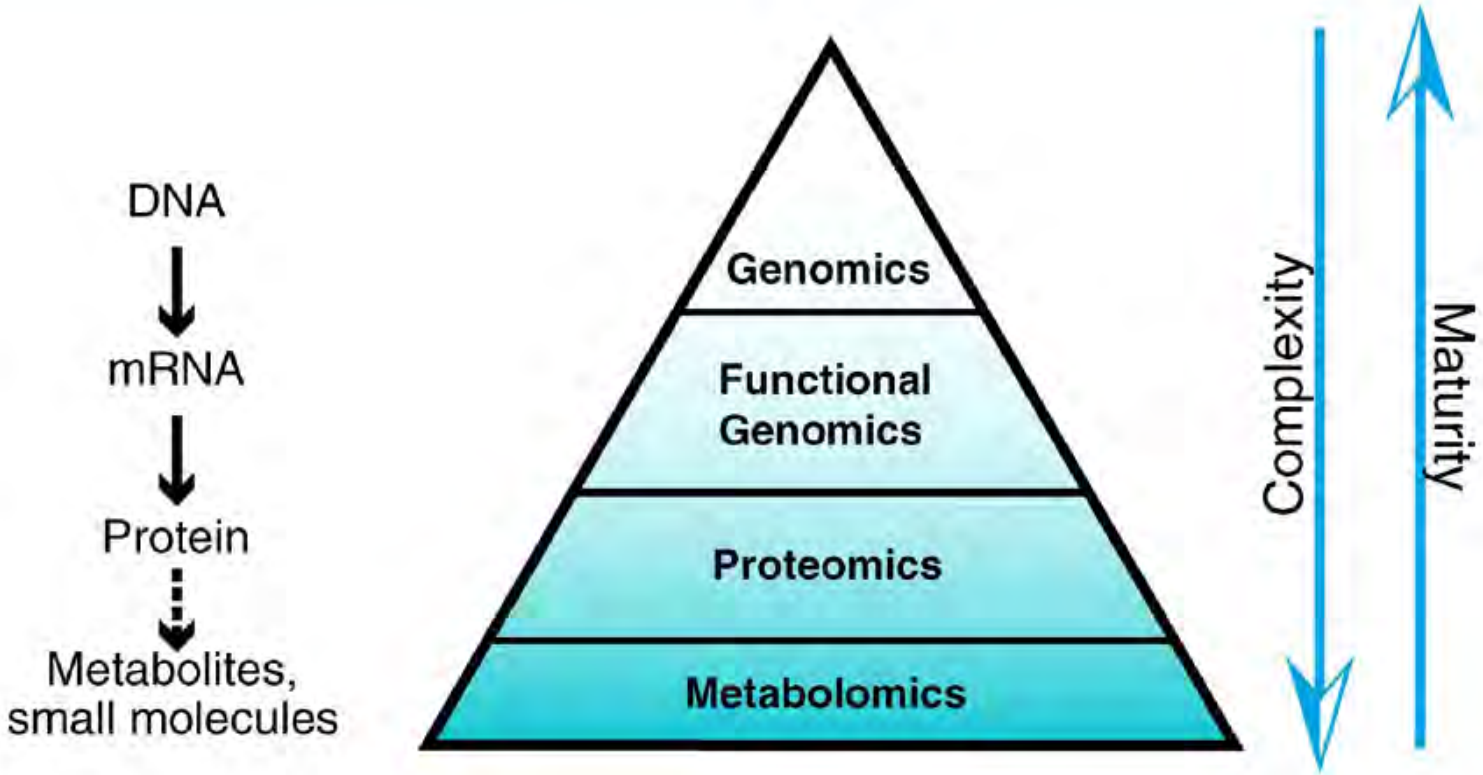
What does a worm want with 20,000 genes?

Jonathan Hodgkin

One Genome – Two Proteomes

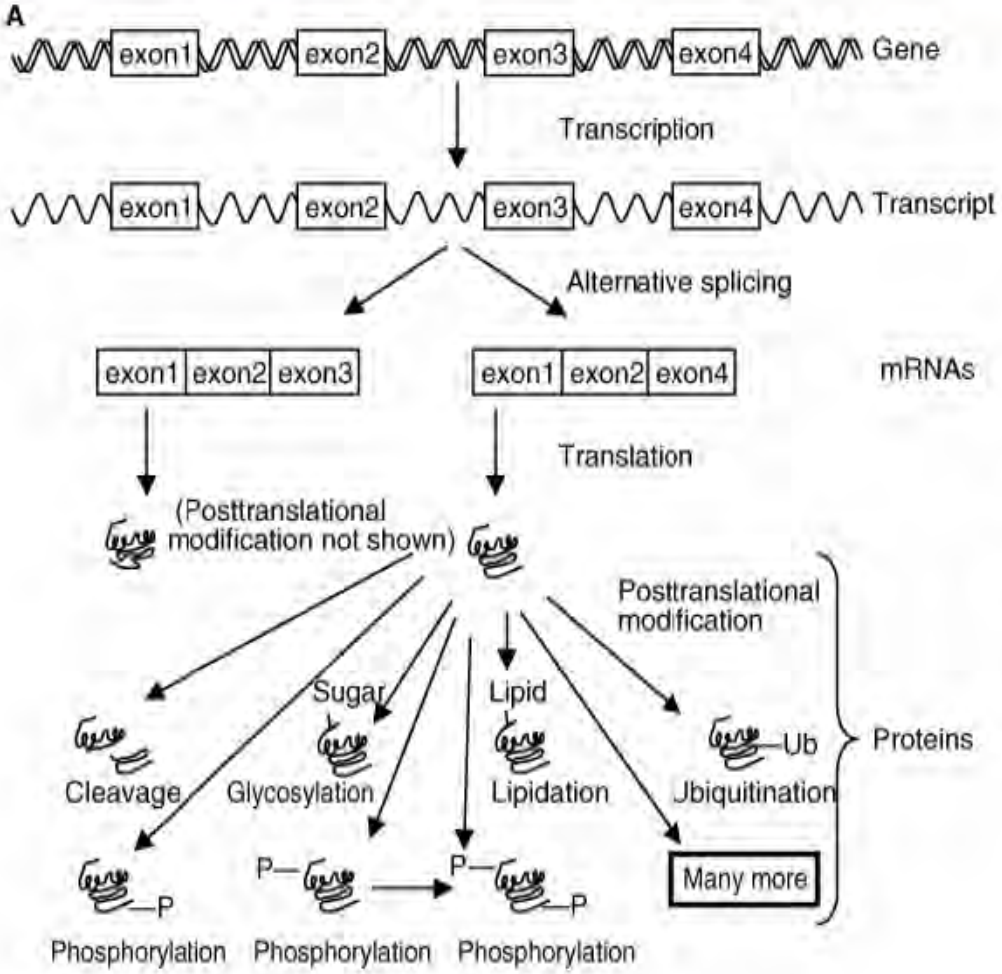
Inachis io

Moving from genomics to proteomics



de Hoog CL, Mann M Annu Rev Genomics Hum Genet. 2004;5:267-93

Complexity of the proteome



What can happen

What might happen

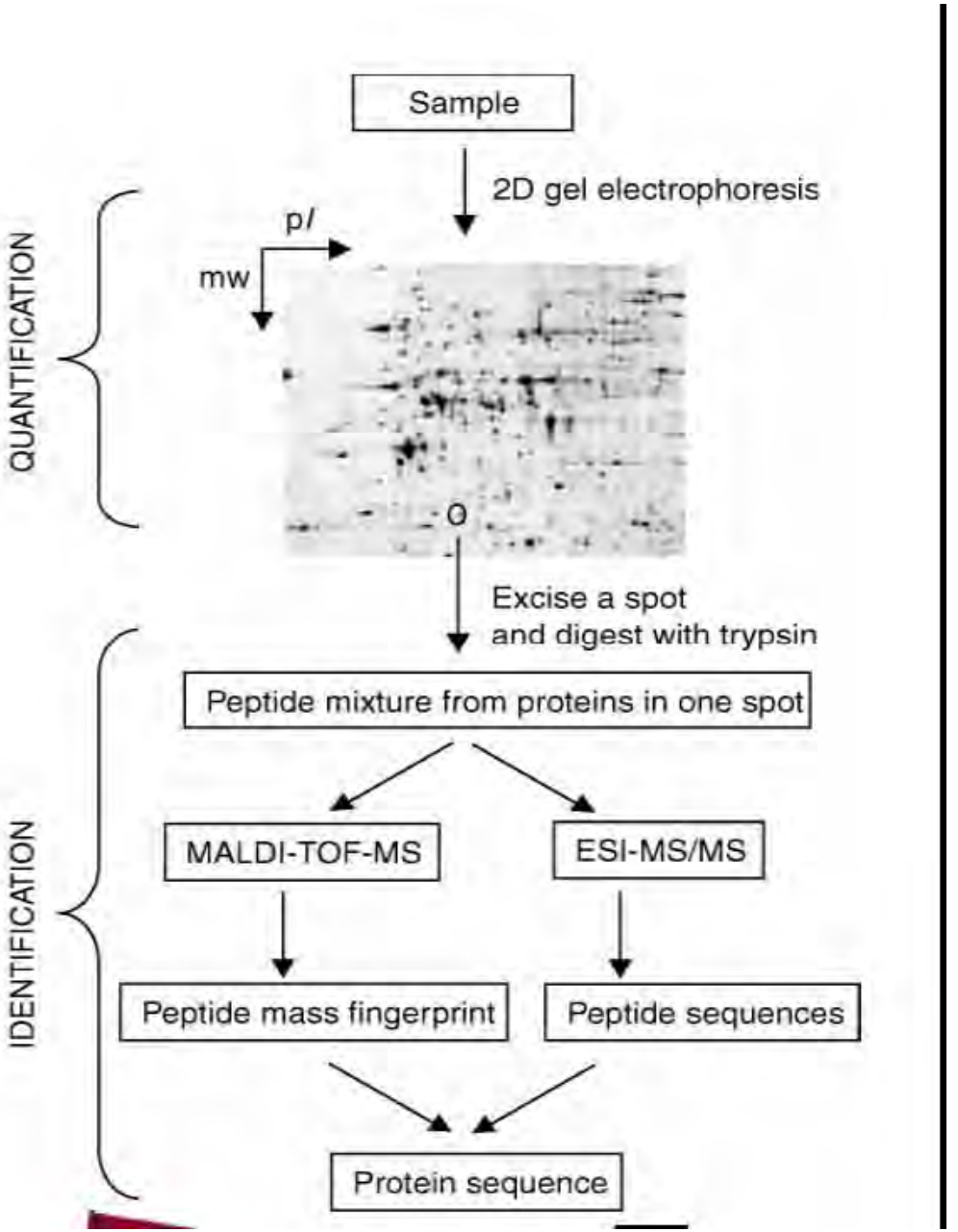
What is likely going to happen

What is happening

Proteome analysis using two-dimensional gel electrophoresis and mass spectrometry (2DE/MS)

"a major goal of proteomics is the global and quantitative measurement of the proteins expressed in cells or tissues"

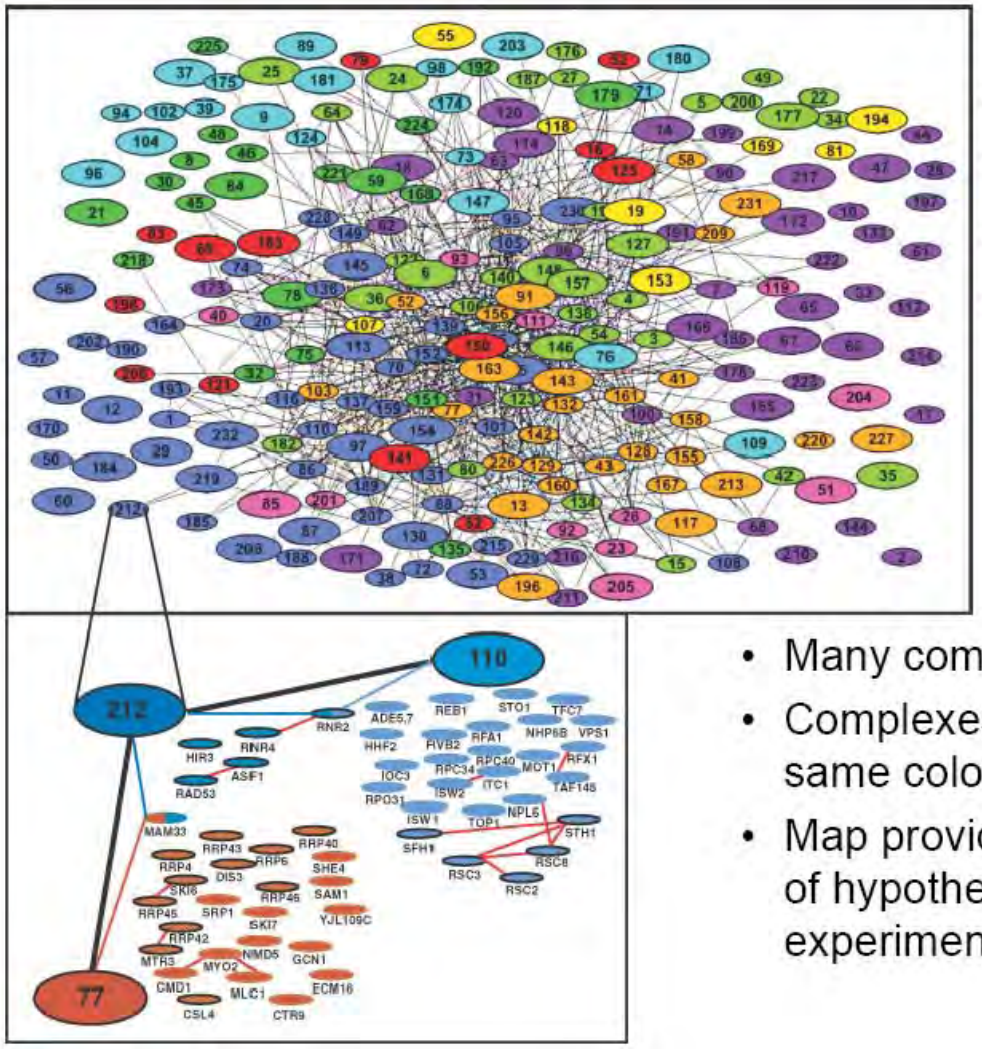
"the qualitative and quantitative comparison of proteomes under different conditions to further unravel biological processes"



Protein mass spectrometry develops in various stages

- What is the primary structure and the identity of the protein?
- What is the higher order structure of the protein, which structural domains can be discerned
- What is the subunit composition of functional protein complexes
- What is the topology of a protein and protein complexes
- Which functional domains are present and what is the function of proteins and assemblies of proteins
- How is the function of a protein regulated by post translational modifications
- How is the function of a protein regulated by interaction with other proteins

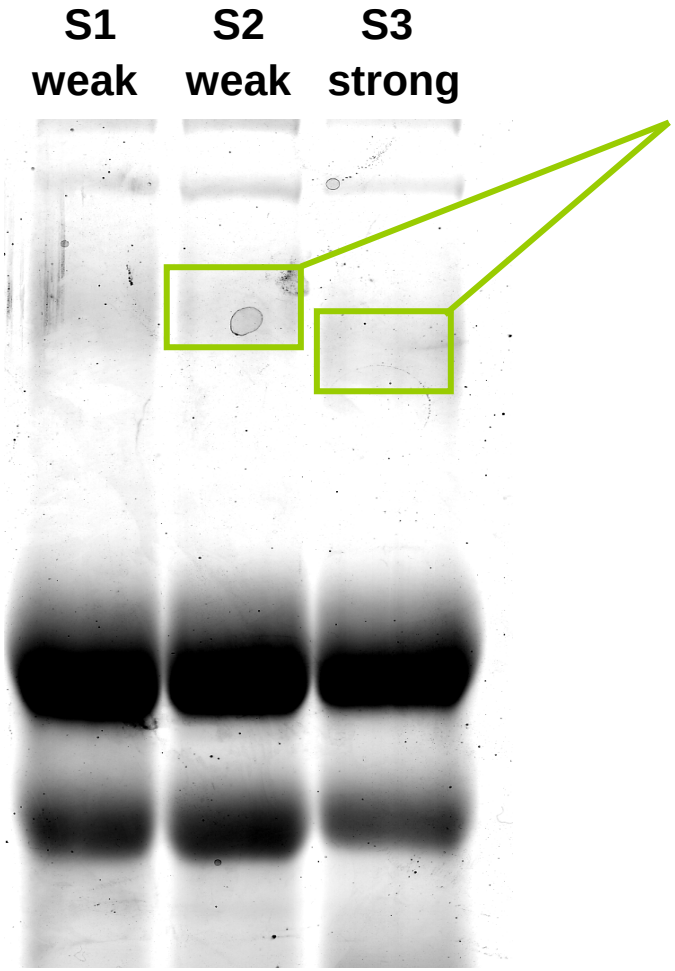
The protein complex network in *S. cerevisiae*



- 16830 proteins identified
- 1140 distinct gene product
- 25 % ORF's
- 232 complexes

- Many complexes share components
- Complexes in related processes have the same color
- Map provides information for generation of hypothesis leading to further experimental investigations

Identification



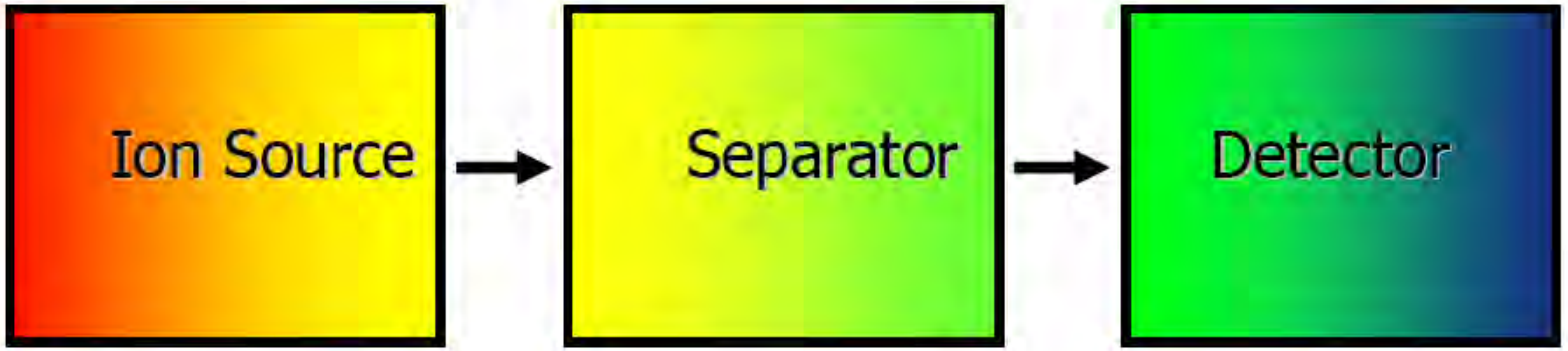
Coomassie SDS-PAGE

What is this protein ‘band’???

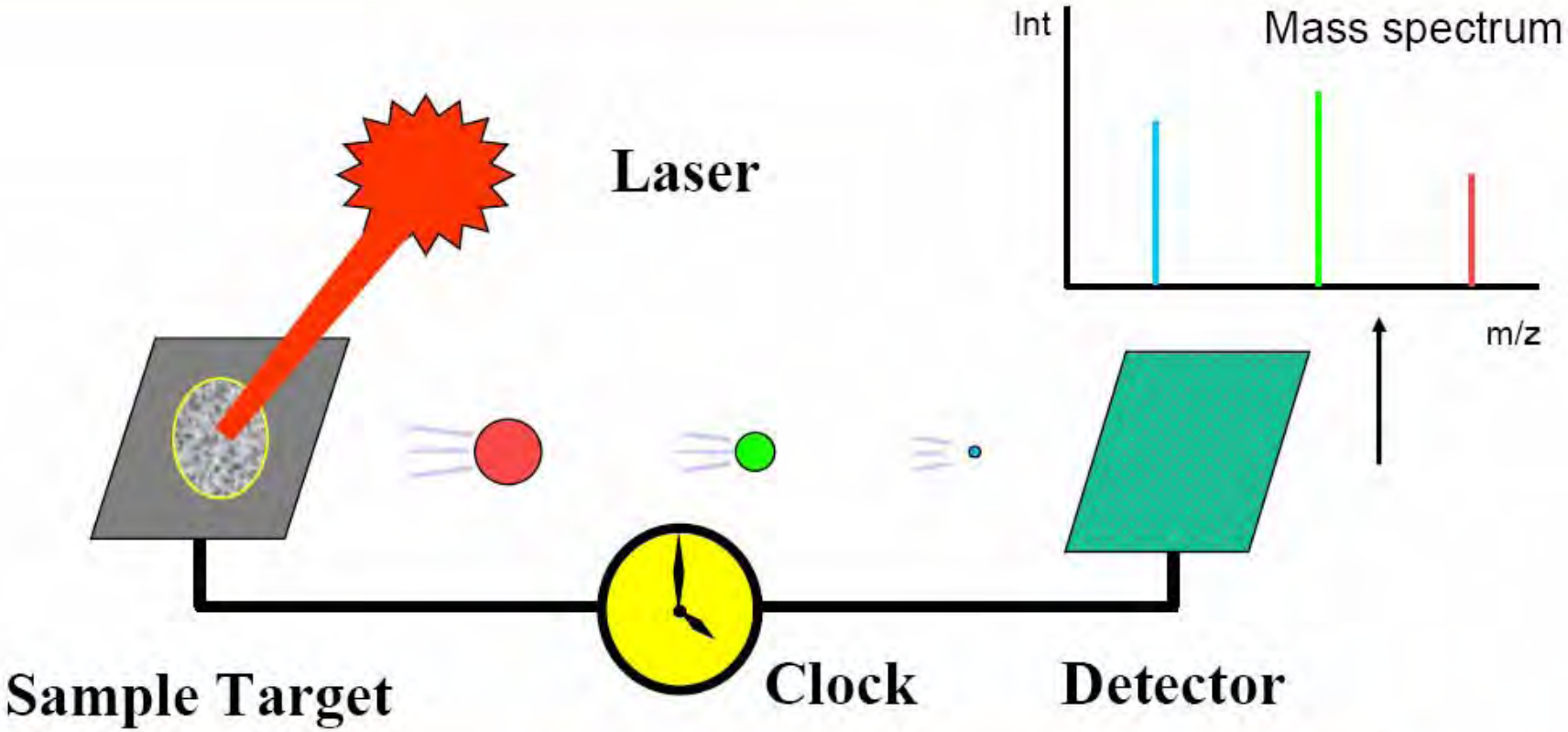
How would you solve this question?

Principle of Biomolecular Mass Spectrometry

- transfer biomolecule into the gas phase
- ionize biomolecule
- apply electromagnetic fields
- analyse ion movements
- determine m/z of the biomolecule and the number of ions



Principle MALDI-TOF-MS

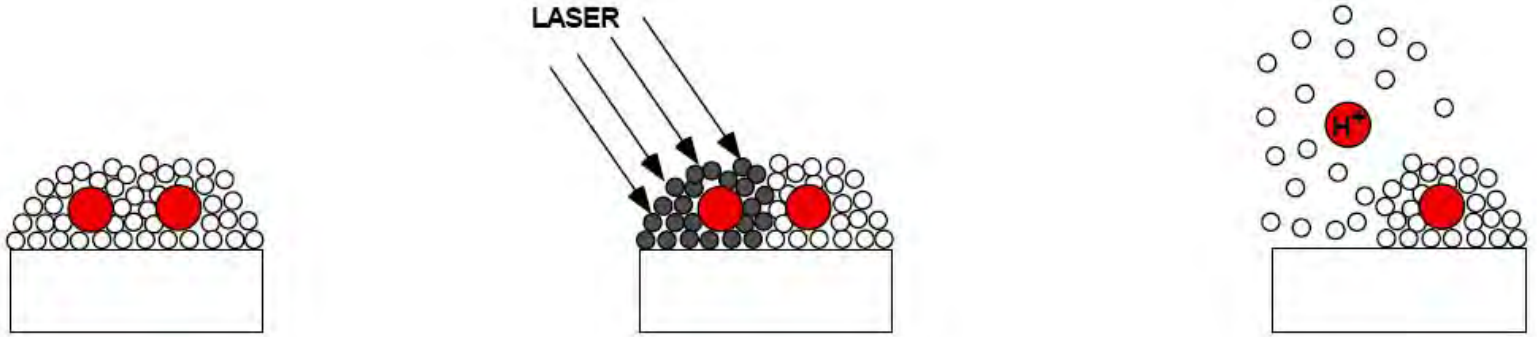


Matrix-Assisted Laser Desorption Ionization (MALDI)

Sample embedded in light-absorbing matrix

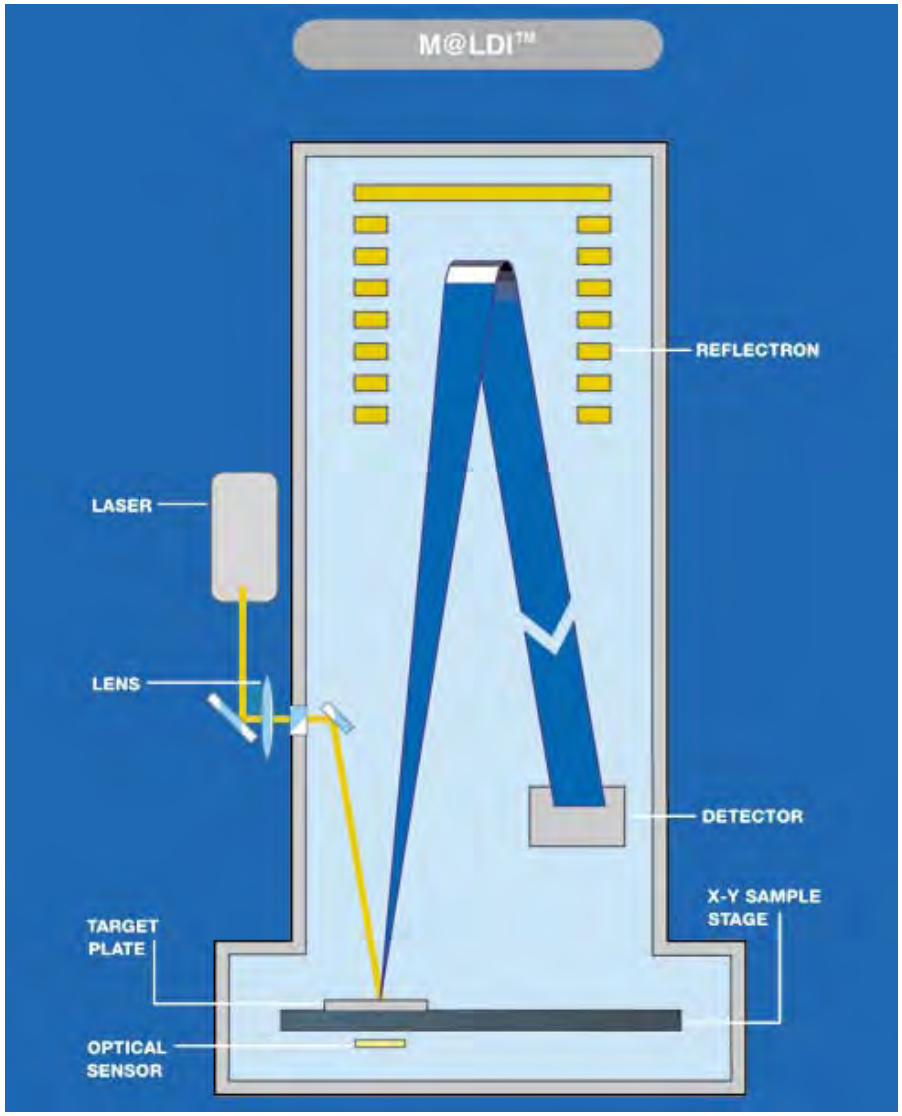
LASER-excitation of matrix molecules

Sample desorption and protonation



Matrix for Proteins: sinapinic acid, dihydroxybenzoic acid etc.
Matrix for Peptides: 4-hydroxy-α-cyanocinnamic acid, DHB

it co-crystallizes, absorbs laser energy, evaporates and acts as acid



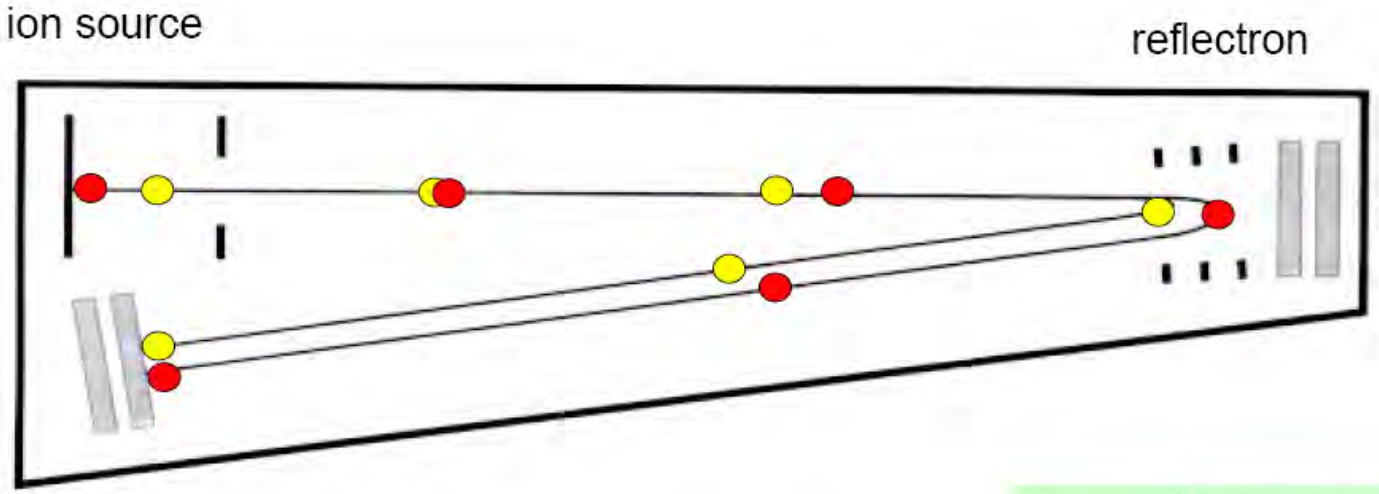
MALDI-TOF MS

- Matrix-Assisted Laser Desorption Ionization Time Of Flight ms.
- sample in UV absorbing matrix hit by UV laser pulse.
- peptideH^+ desorbed, accelerated, enters field-free flight tube
- flight time = $A.m^{1/2}$
- reflectron increases mass accuracy

Identical ions with slightly different kinetic energies are focused by the reflectron

- ions with a relatively high kinetic energy travel deeper into the reflectron than identical ions with a relatively low kinetic energy
- therefore, the faster "high energy ions" reside for longer times in the reflectron then the slower "low energy ions"
- as a result, identical ions with slightly different kinetic energy reach the detector at the same time, under properly choosen conditions

MALDI-TOF-MS Resolution: Ion trajectories in a reflectron time-of-flight mass spectrometer, where U is the kinetic energy and dU the difference in kinetic energy of two ions with mass m



$$t_r = 2\frac{\sqrt{2m(U+dU)}}{zeE_r}$$

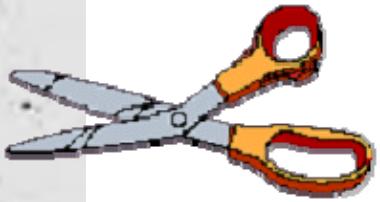
where t_r is the flight time in the reflectron field E_r

U U+dU

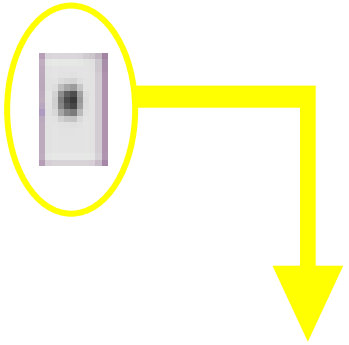
UNIVERSITEIT VAN AMSTERDAM

Sample

2D Page



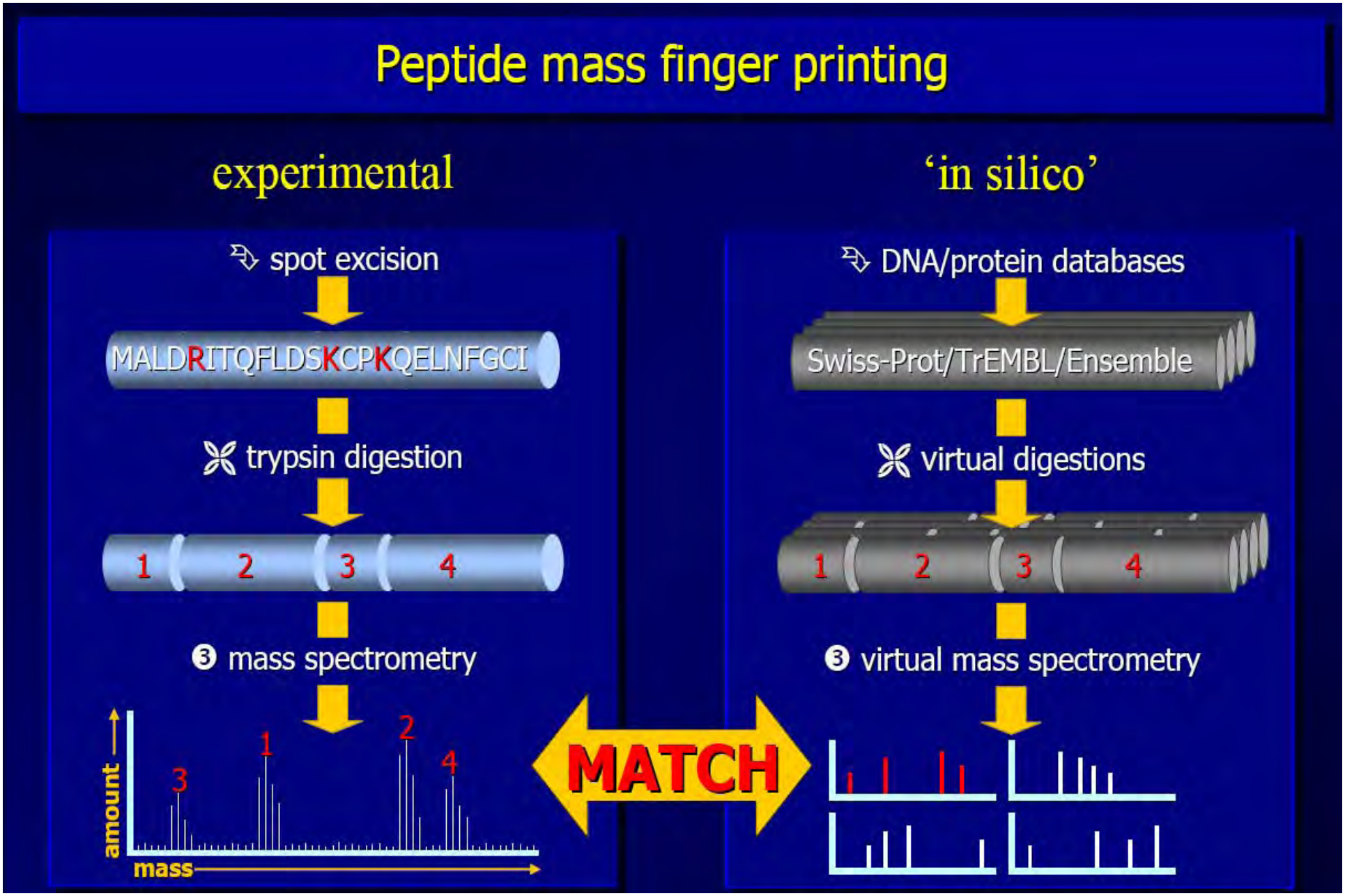
Cut



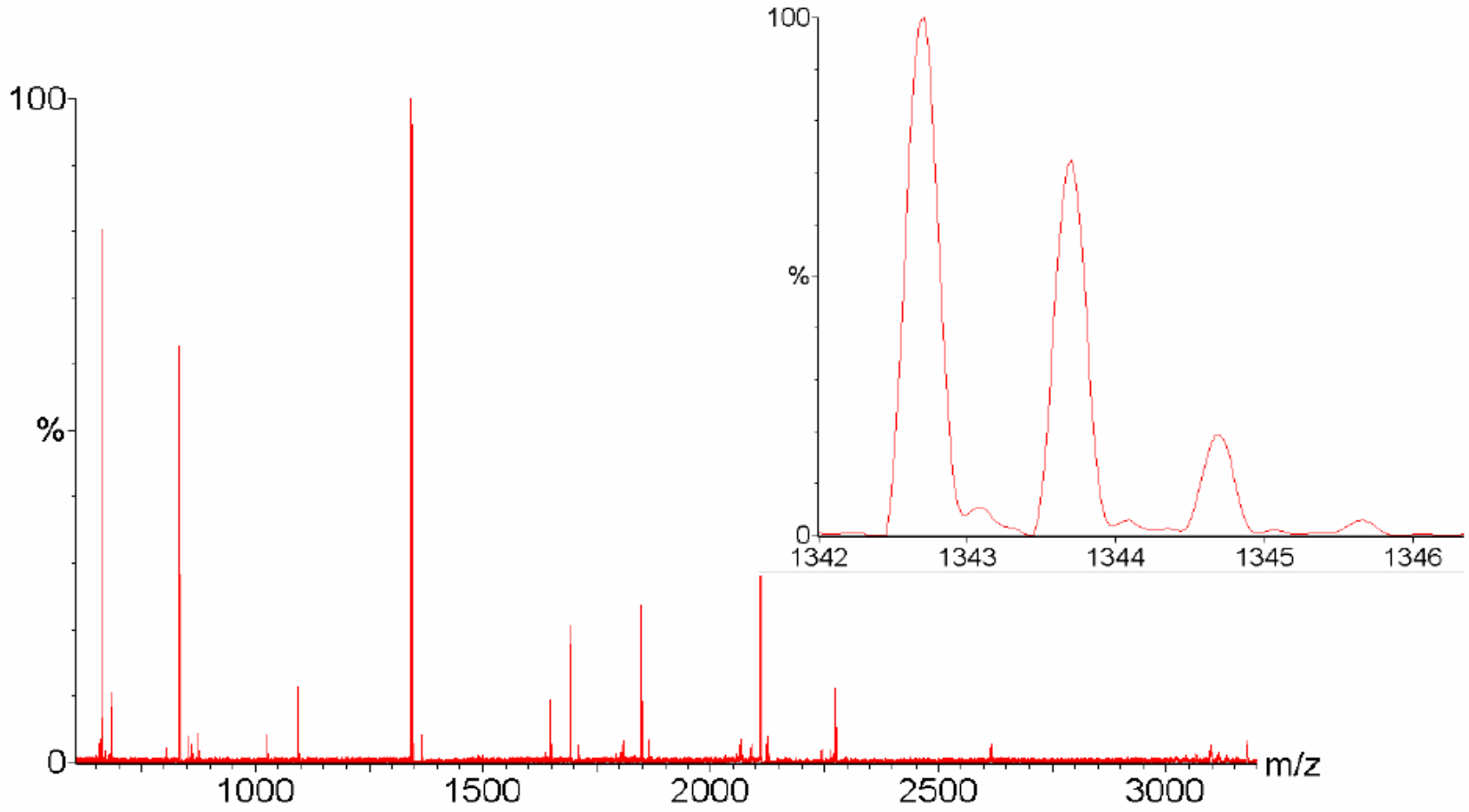
Isolate

Digest

Peptides

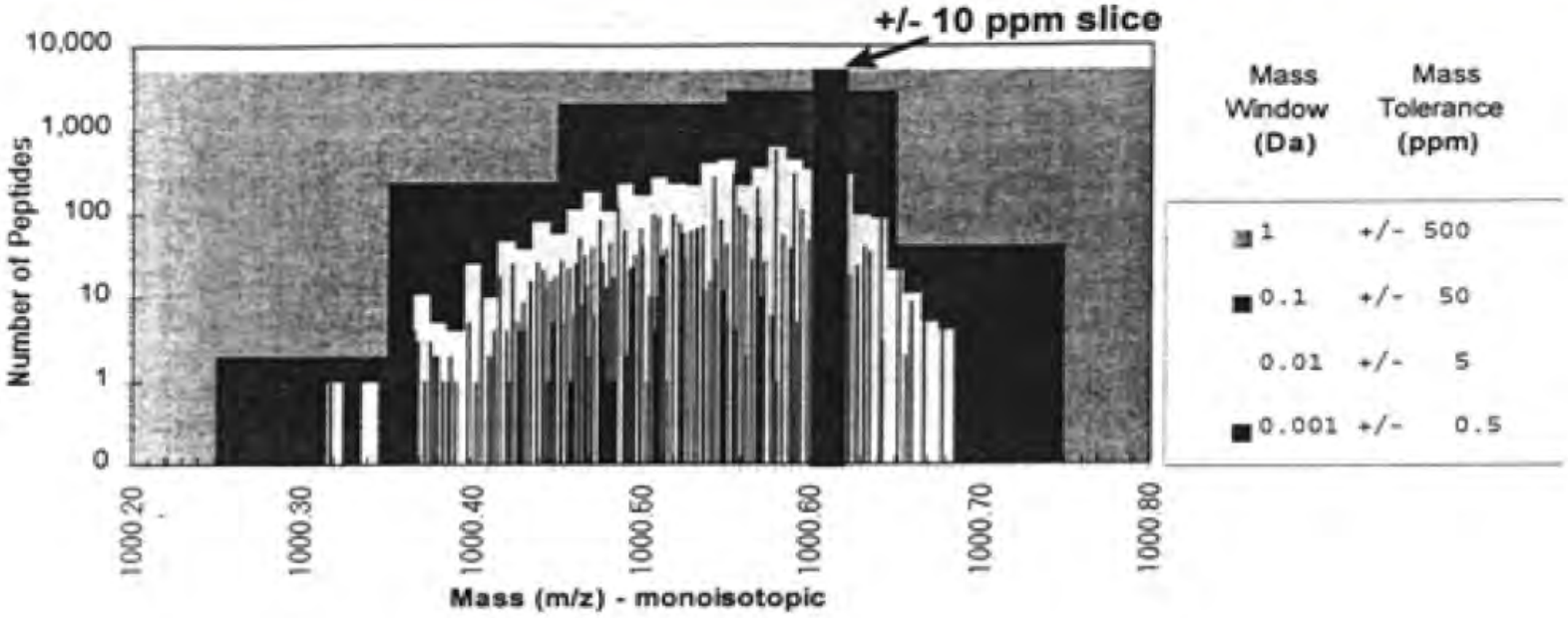

Peptide mass finger printing
experimental
'in silico'
spot excision
MALDRITQFLDSKCPKQELNFGCI
trypsin digestion
1
2
3
4
mass spectrometry
amount
mass
MATCH
DNA/protein databases
Swiss-Prot/TrEMBL/Ensemble
virtual digestions
virtual mass spectrometry

Mass spectrum of tryptic digest of aldolase obtained by MALDI-TOF MS



N.B. a mass spectrum like this contains enough information to identify the protein if present in a database

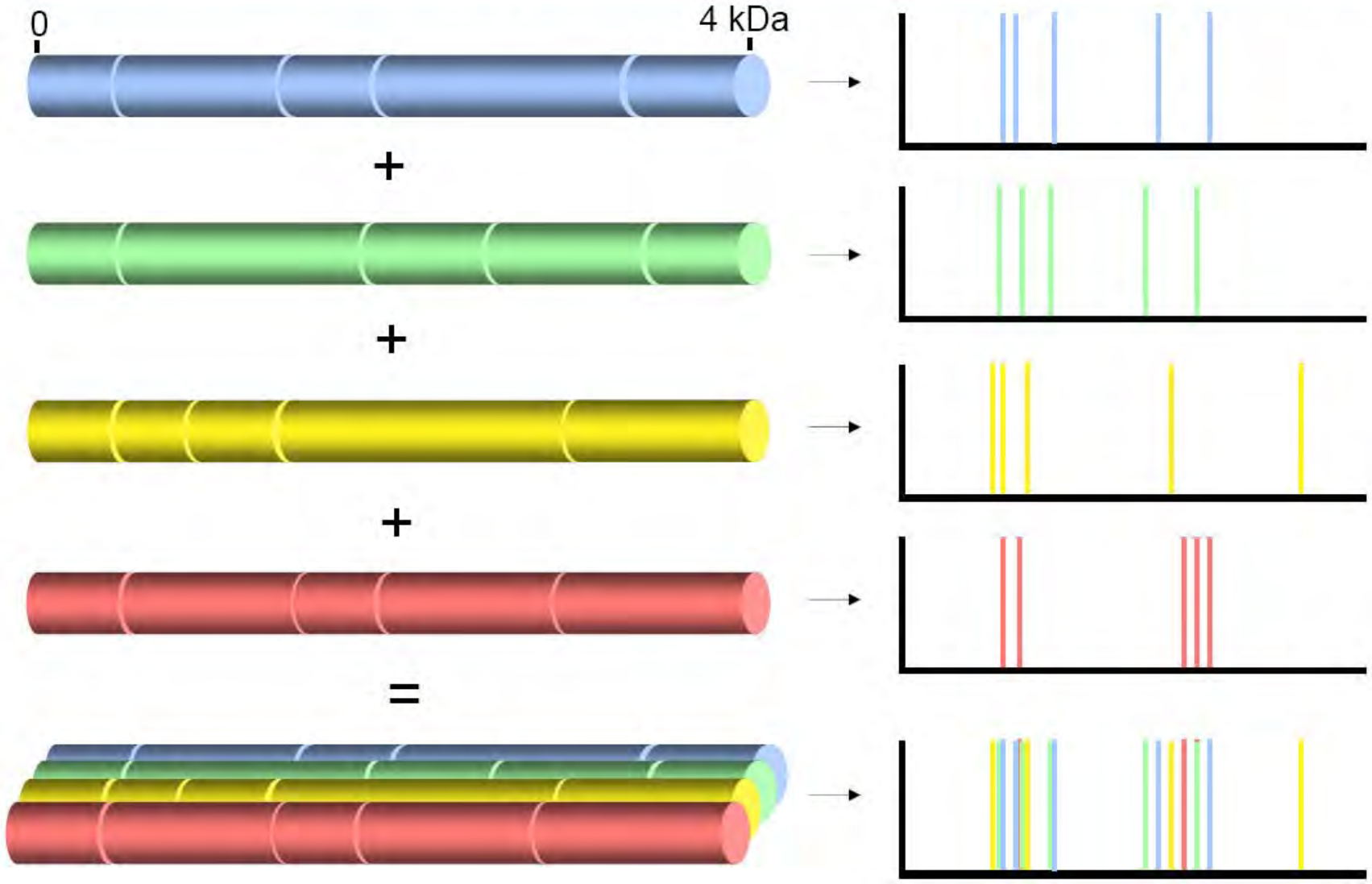
Number of tryptic peptides of nominal mass 1000 in a database with 210.000 entries



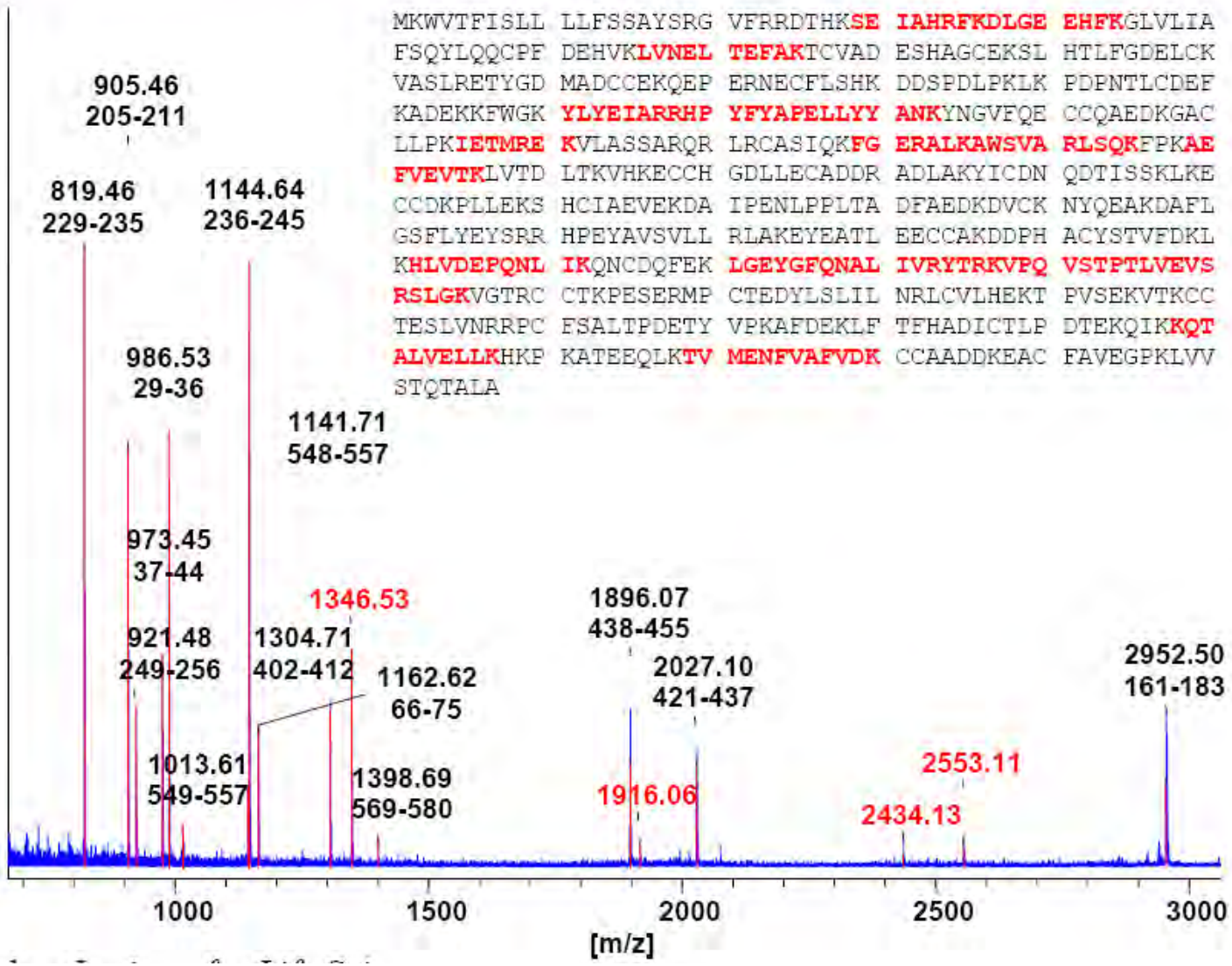
Suppression effect

- Different peptides have different propensities to ionize
- The ionization of some peptides is suppressed by the presence of other, better ionizable peptides
- So, the signal intensity of a peak in a mass spectrum of a mixture of peptides is not a measure of the concentration of the peptide concerned

Mass Accuracy and PMF: database searching



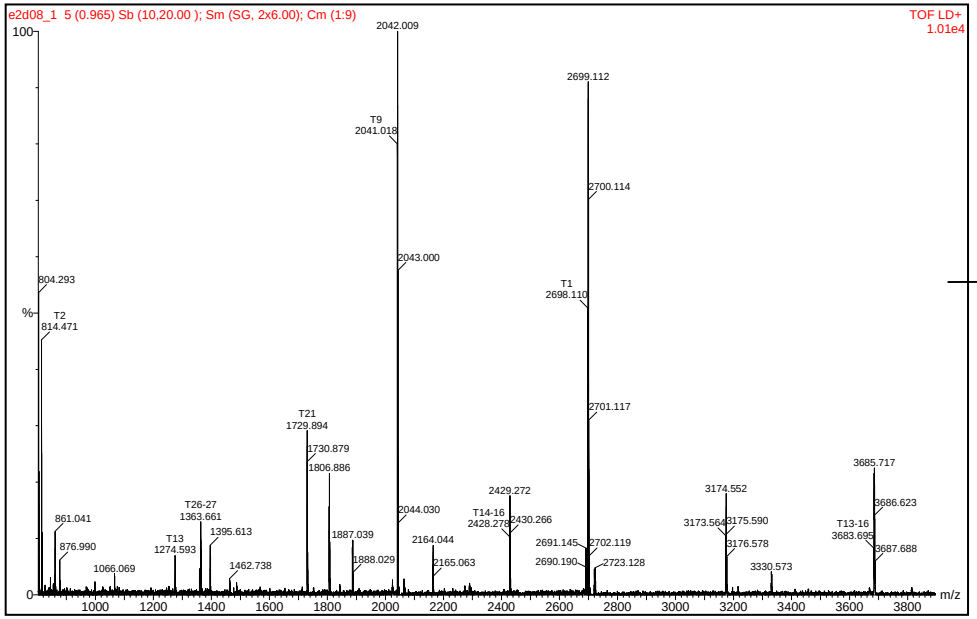
Lys C Digest of Bovine Serum Albumine



Peptide mass fingerprinting

ion

int.



m/z

Ionisation efficiency !

814.489
861.105
1274.612
1363.655
1395.664
1729.908
1806.013
1806.869
1886.063
2041.015
2042.155
2163.046
2428.265
2690.126
2697.335
2698.127
2699.038
2719.405
2721.204
3172.451
3173.639
3682.893
3683.684

Protein dBase

Peptide dBase

compare

match

Example: HSP26_yeast

Mass analysis by MALDI-TOF-MS

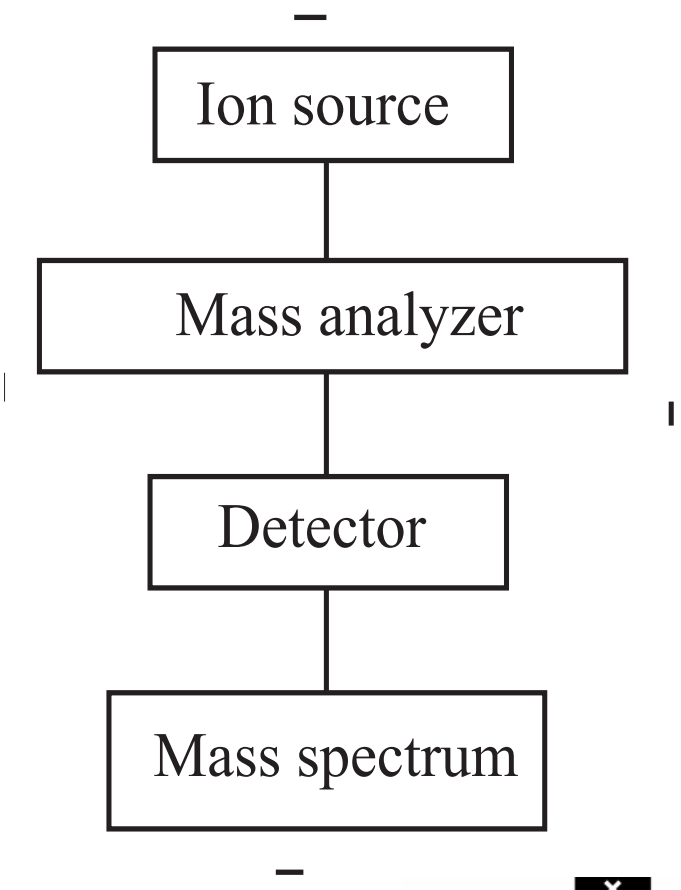
- Mostly singly charged analyte ions $(M + H)^+$, minor or no fragmentation
- Mass range of analytes: 500-300000 Da
- Resolution: $m/\Delta m$; Δm: "full width at half maximum" (fwhm)
- Resolution with MALDI-TOF-MS ~20.000: no "isotopic resolution" at m/z values > 10.000
- Accurate mass measurements requires internal standards
- Mass accuracy
  - 500-3000 Da: ~ 30 ppm (~ 0.1 Da at 3000)
  - 3-10 kDa: ~ 100 ppm (~ 1 Da at 10000)
  - 10-30 kDa: ~ 300 ppm (~ 10 Da at 30000)
- Peptide mass fingerprinting demands ±0.2 Da mass accuracy
- Sensitivity for peptides: < 10 fmol

Electrospray LC-MSMS and sequence tag analysis

What is a mass spectrometer?

A mass spectrometer is an instrument wherein

- Ions are generated
- Ions are separated as a function of mass over charge ration m/z
- Ions are detected
- The mass over charge ratio and the amount of ions are determined



Electrospray ionisation

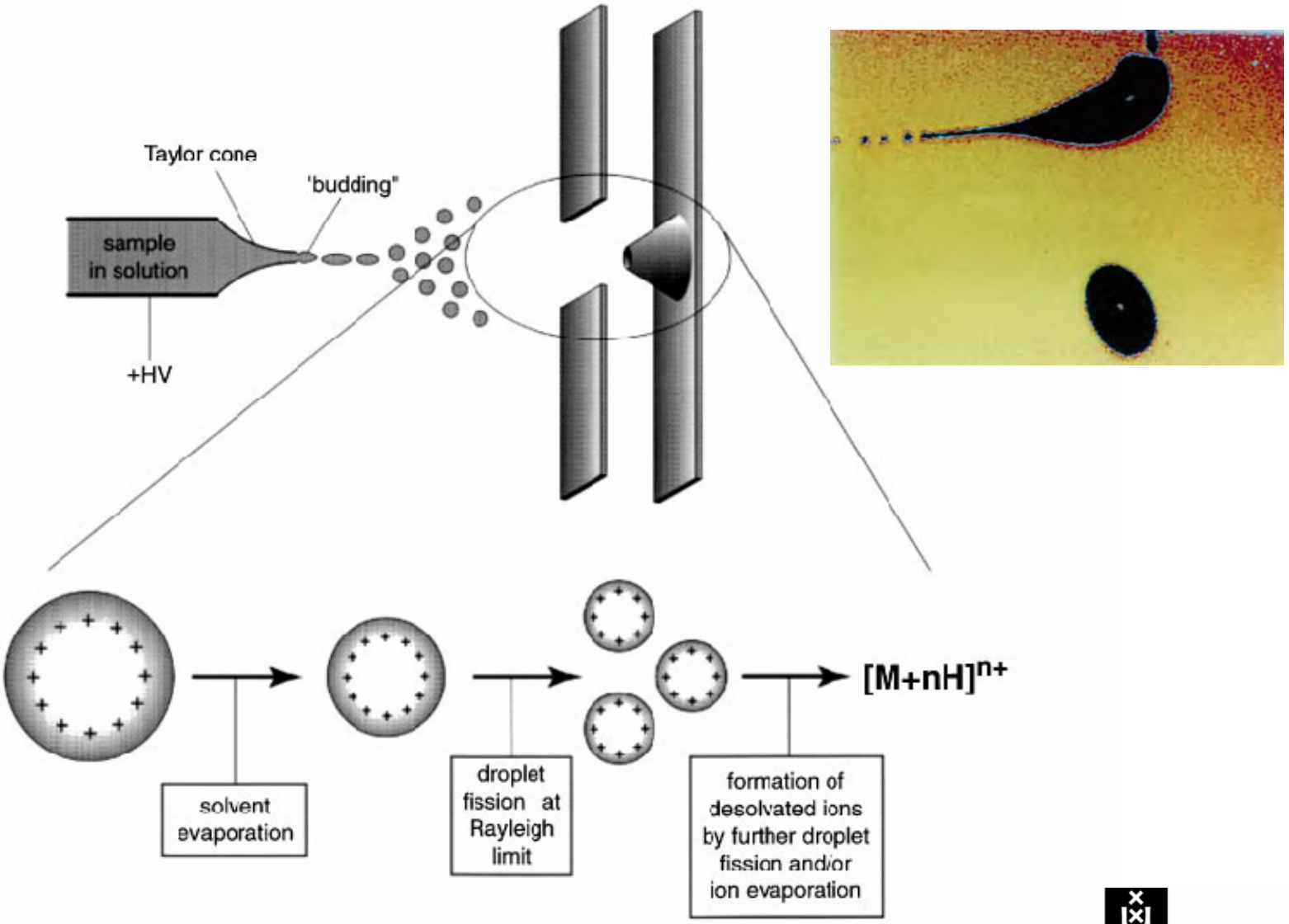
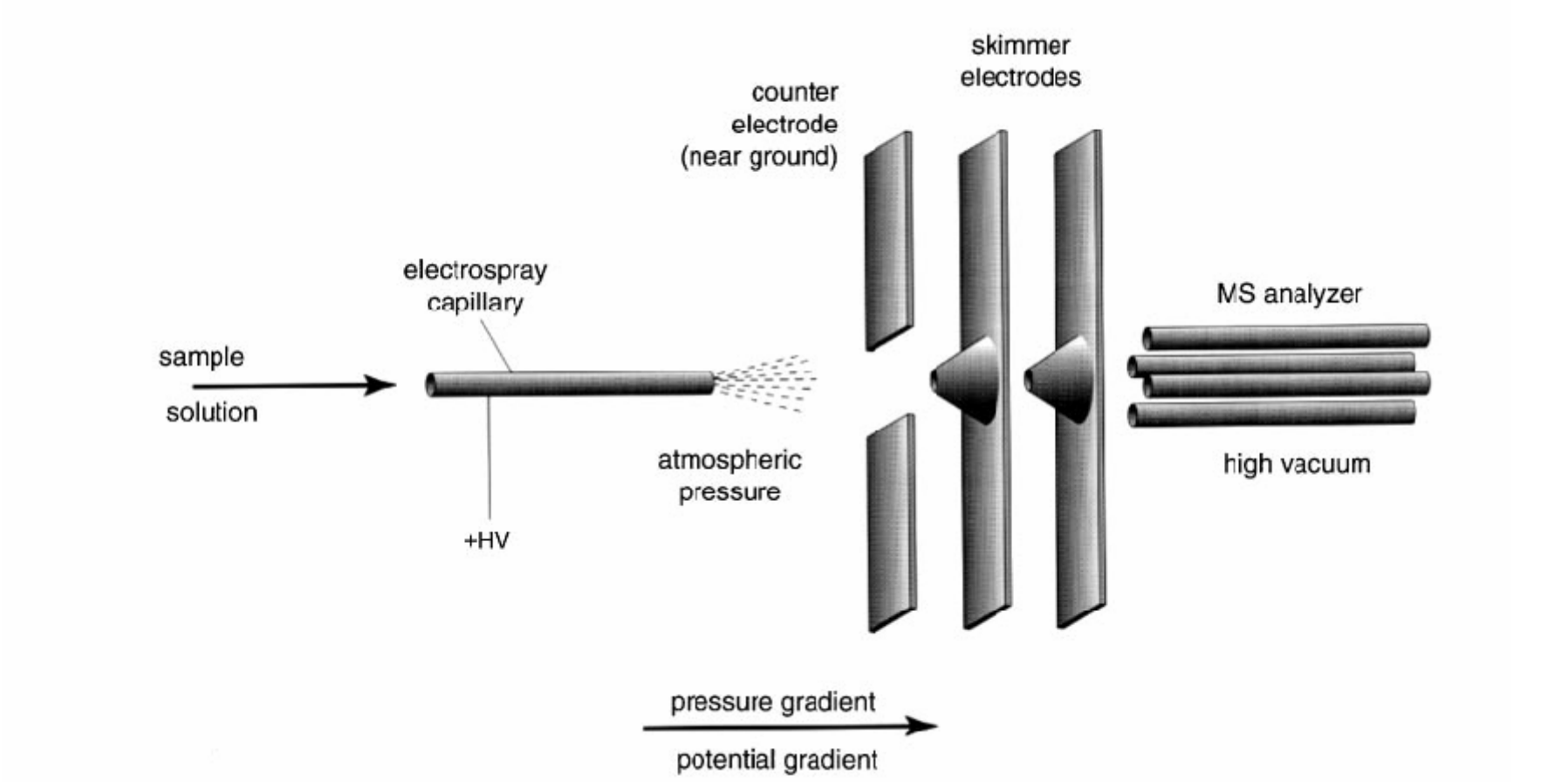
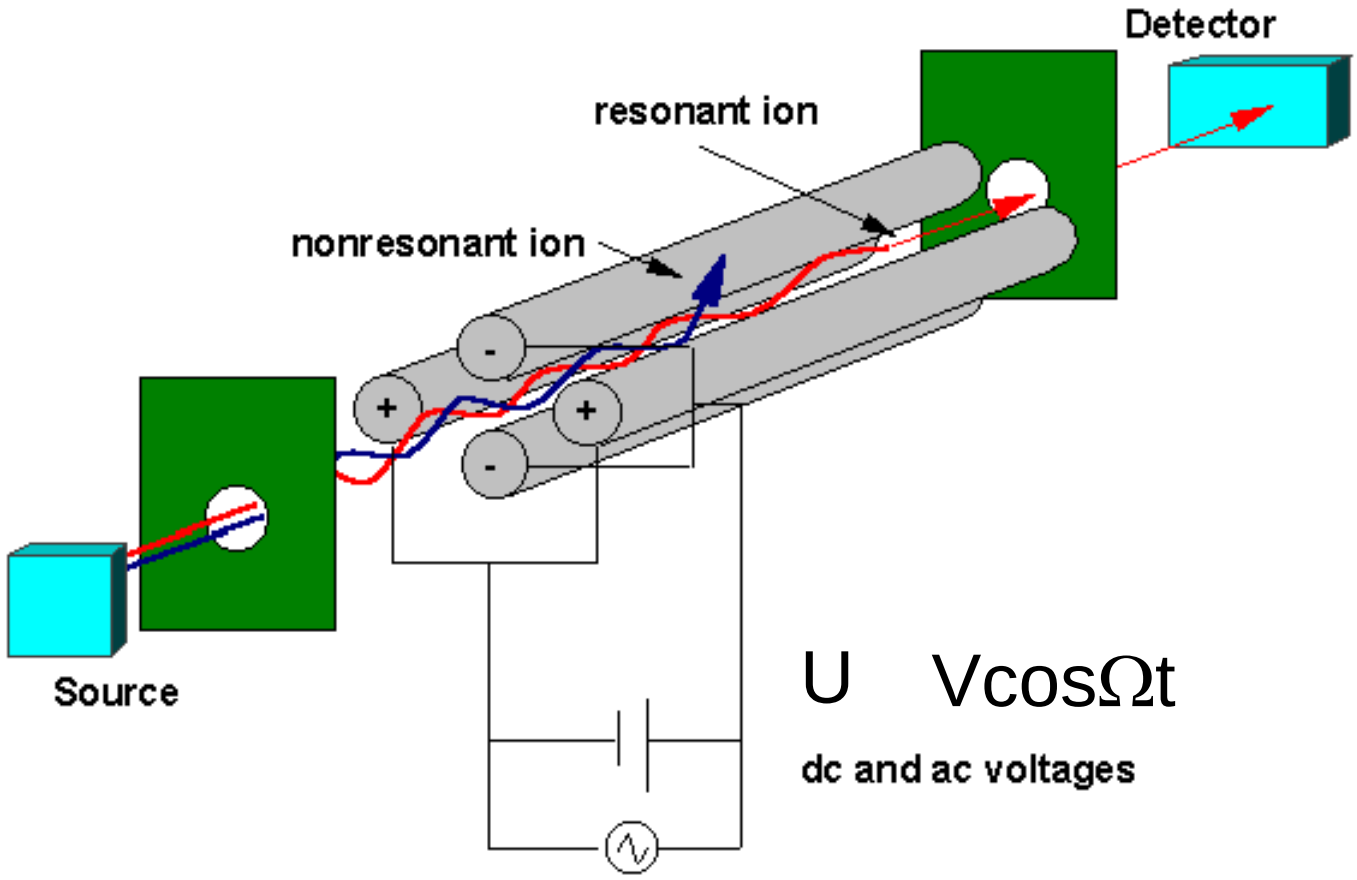


Figure 2. Droplet production in the electrospray interface.

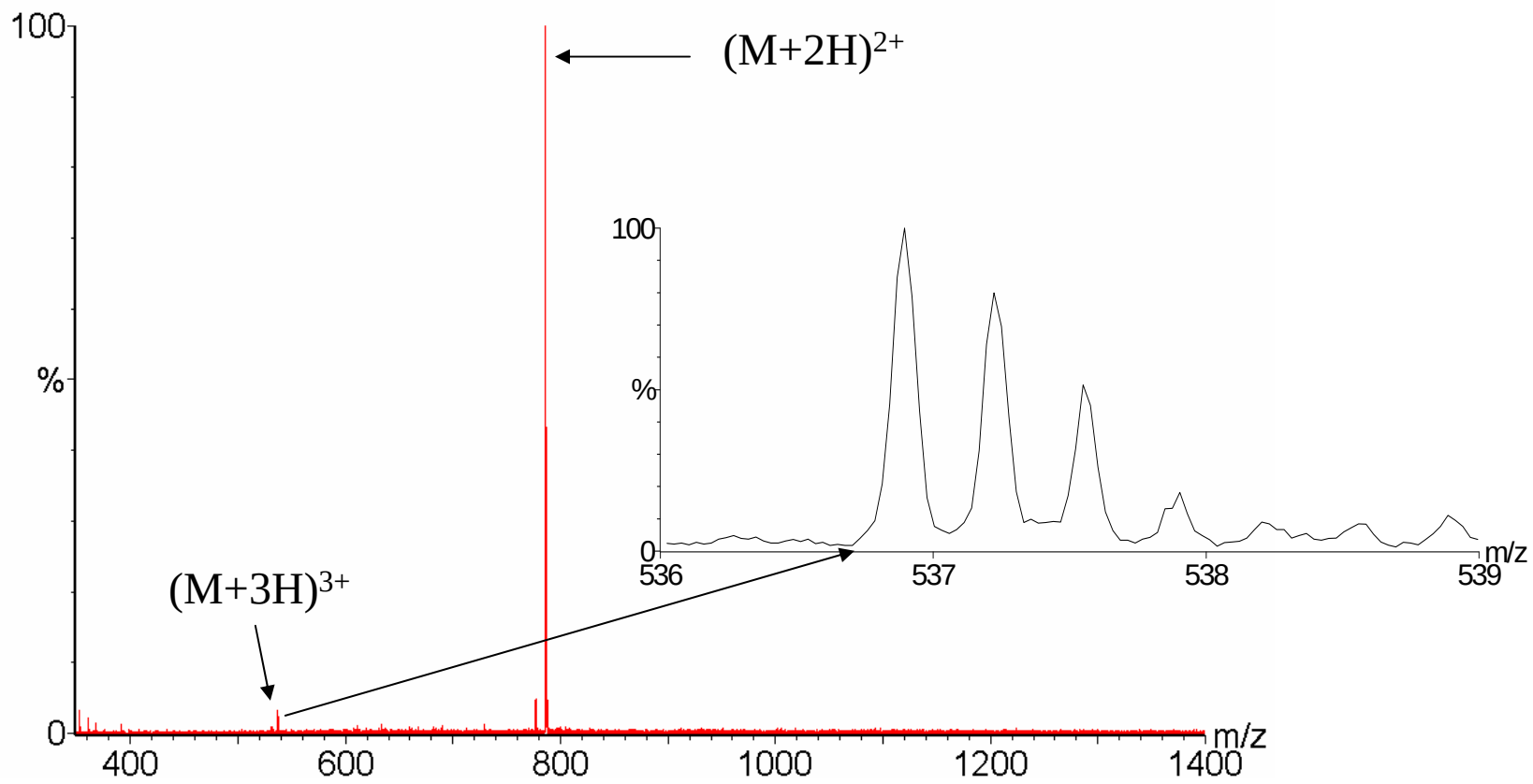
Electrospray ionisation



Quadrupole analyser



ESI-TOF mass spectrum of Glu-fibrinopeptide B



Swammerdam Institute for Life Sciences

UNIVERSITEIT VAN AMSTERDAM

Protein sequencing by mass spectrometry

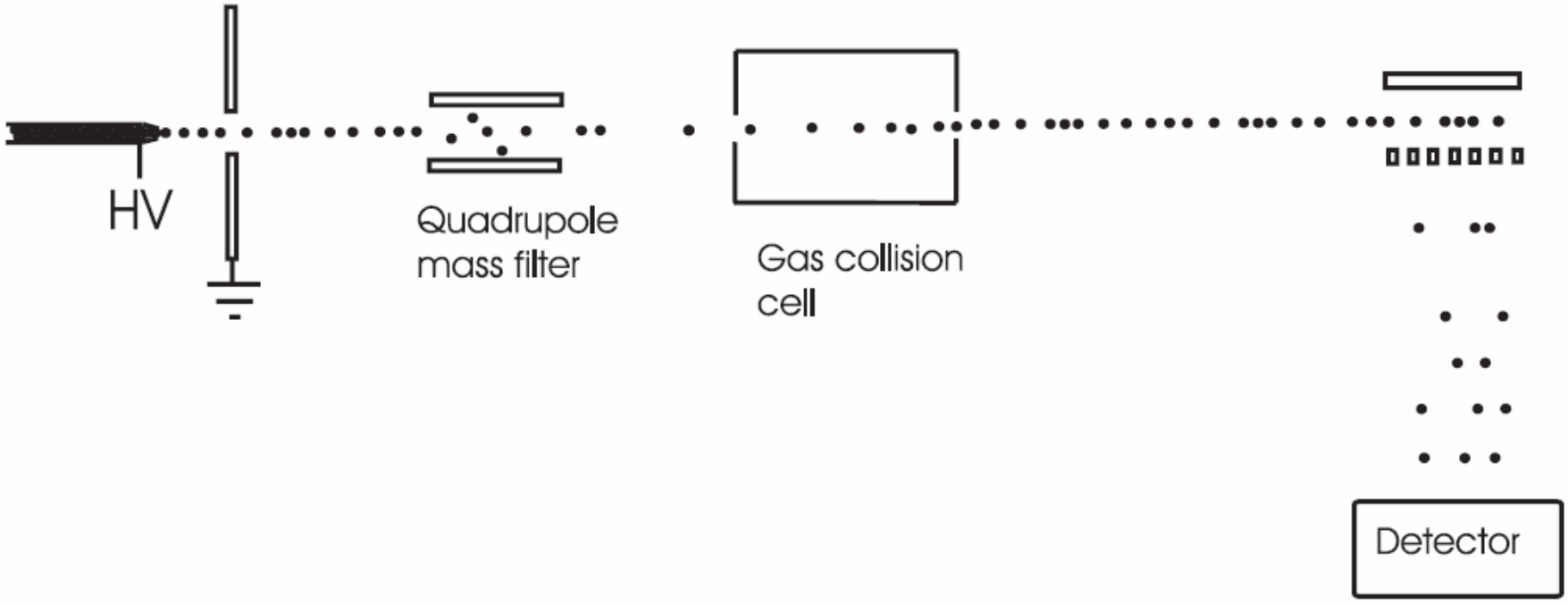
- electrospray ionisation of proteins and peptides
- MSMS principles and instrumentation to fragment gas phase ions
- amino acid sequence determination of peptides

Cleavage sites of trypsin in profilin

Swammerdam Institute for Life Sciences

UNIVERSITEIT VAN AMSTERDAM

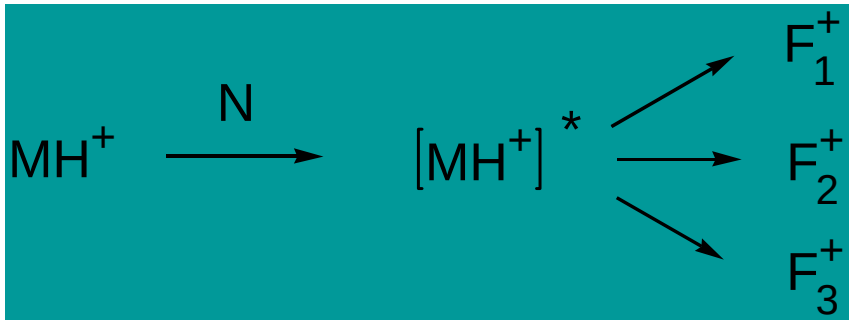
Simplified schematic drawing of an electrospray ionization quadrupole time of flight mass spectrometer (ESI Q-TOF MS) in the MS/MS or MS^2 mode



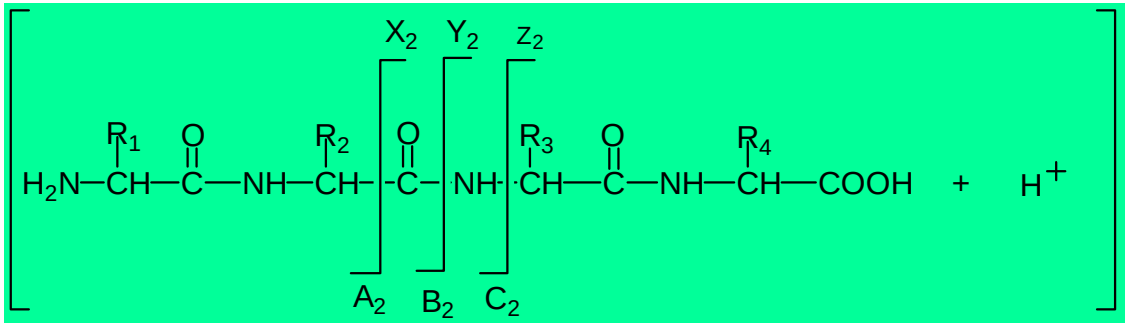
Collision-induced Dissociation

Collisions between ions with a certain kinetic energy and noble gas atoms or small molecules, such as N_2 and O_2.

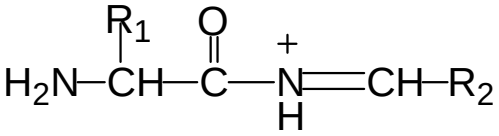
Conversion of kinetic energy into internal energy results in dissociation of the mass selected ion.



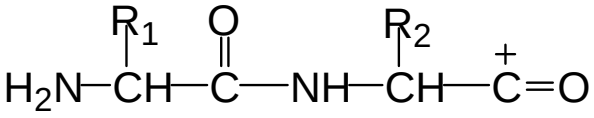
Assumed structures of ions formed by "backbone" fragmentation



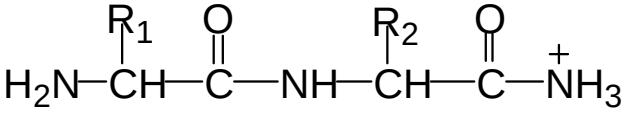
A_2 / a_2 ions:



B_2 / b_2 ions:



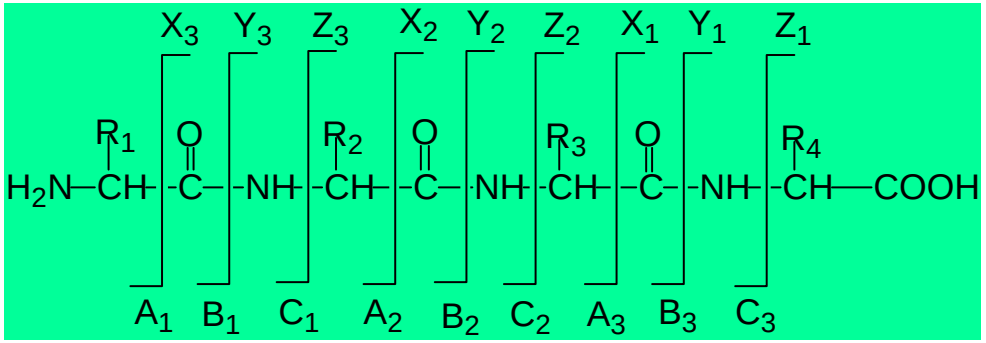
C_2" / c_{2+2} ions:



Peptide fragment ions

Nomenclature: P. Roepstorff and J. Fohlman, Biomed. Mass Spectrom., 11 (1984) 601); K. Biemann, Biomed. Environ. Mass Spectrom., 16 (1988) 99.

Roepstorff, Fohlman:

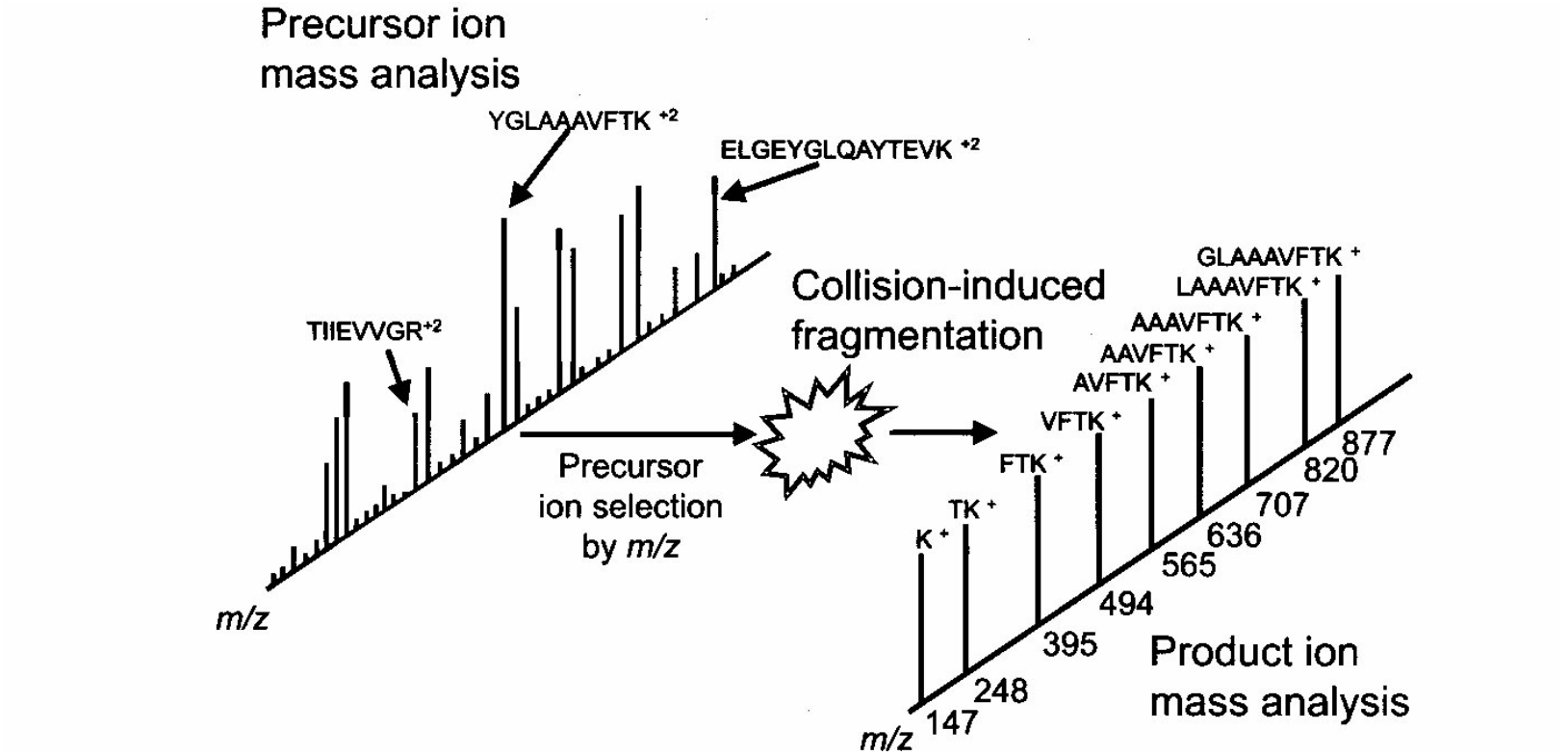


Biemann:

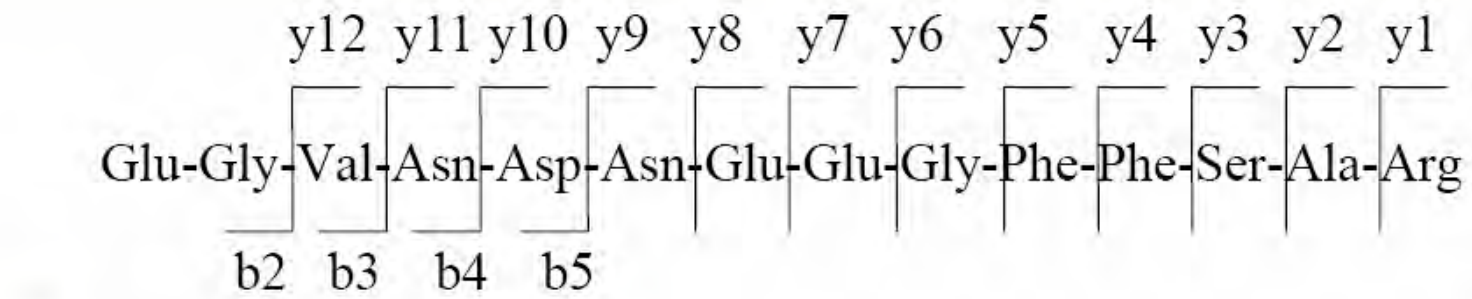
non-capital letters : a_n, b_n, c_n, x_n, y_n, z_n

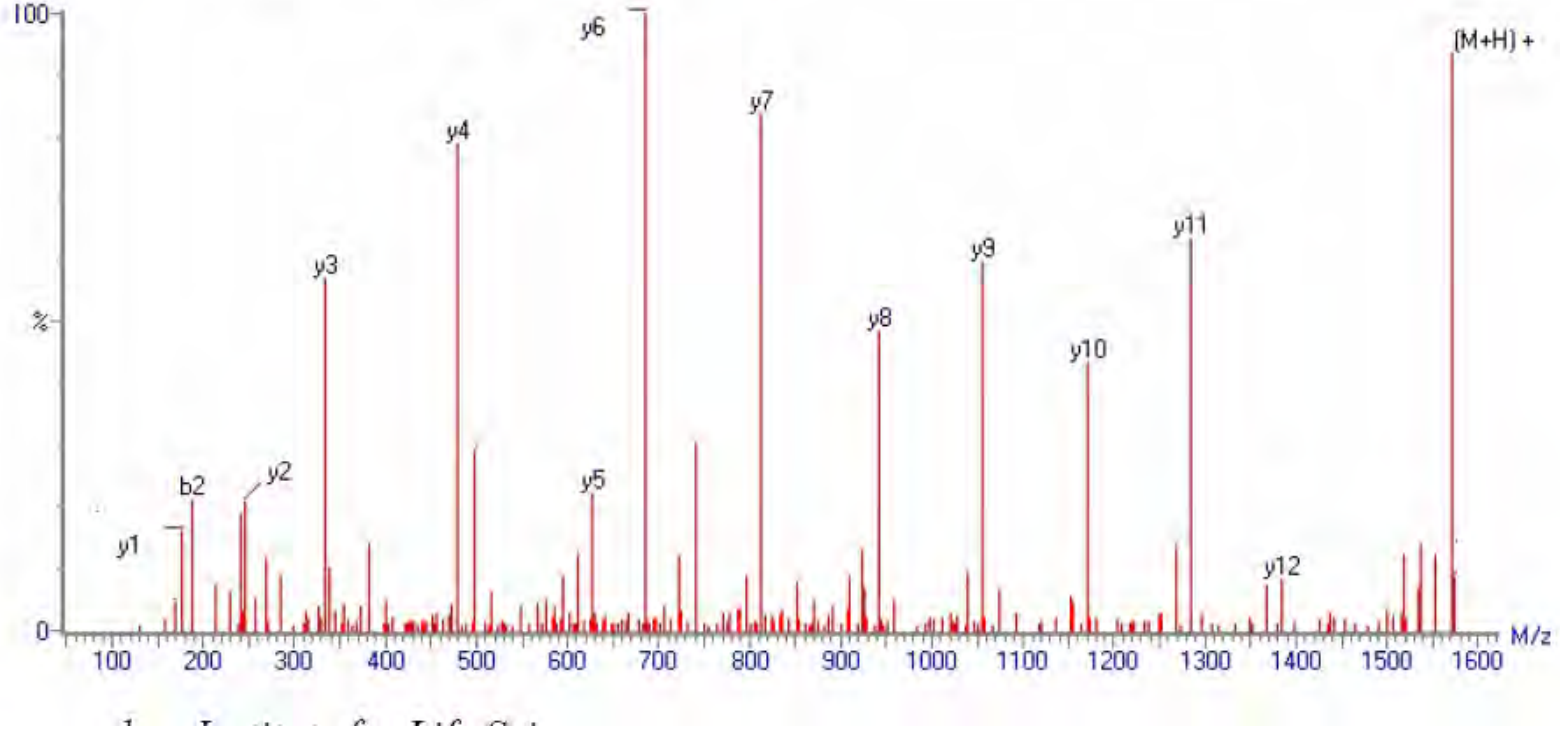
y_{n+2} instead of Y_2"

MS-MS spectra of protonated peptides



MS/MS spectrum of the $[M + 2H]^{2+}$ ion of GluFib

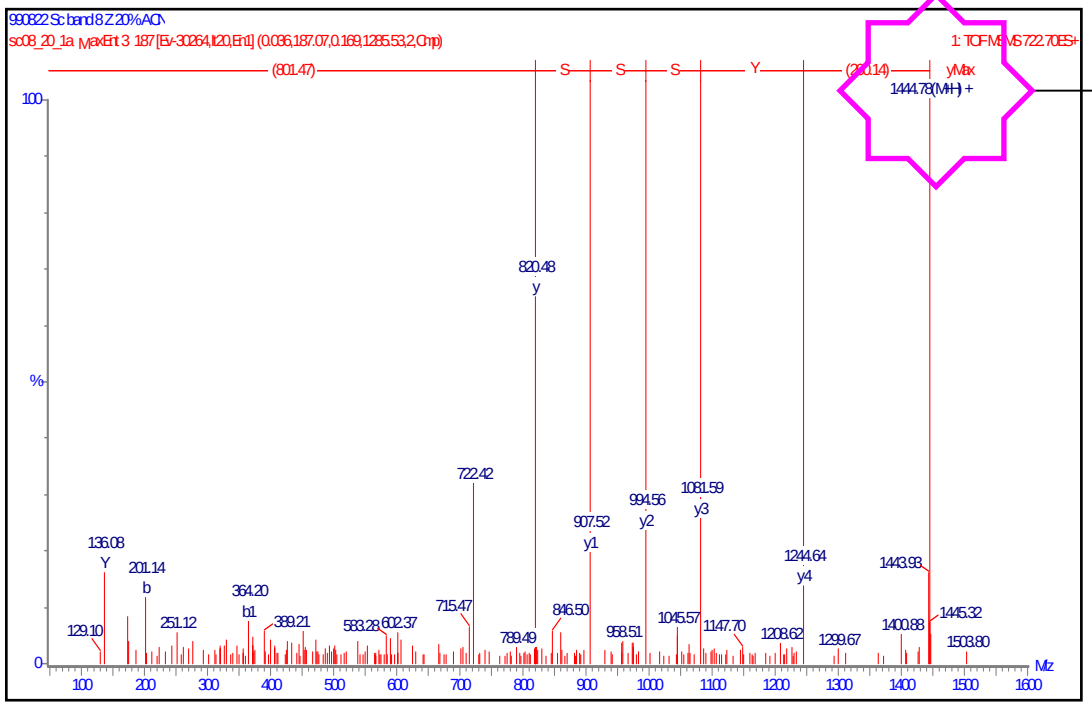


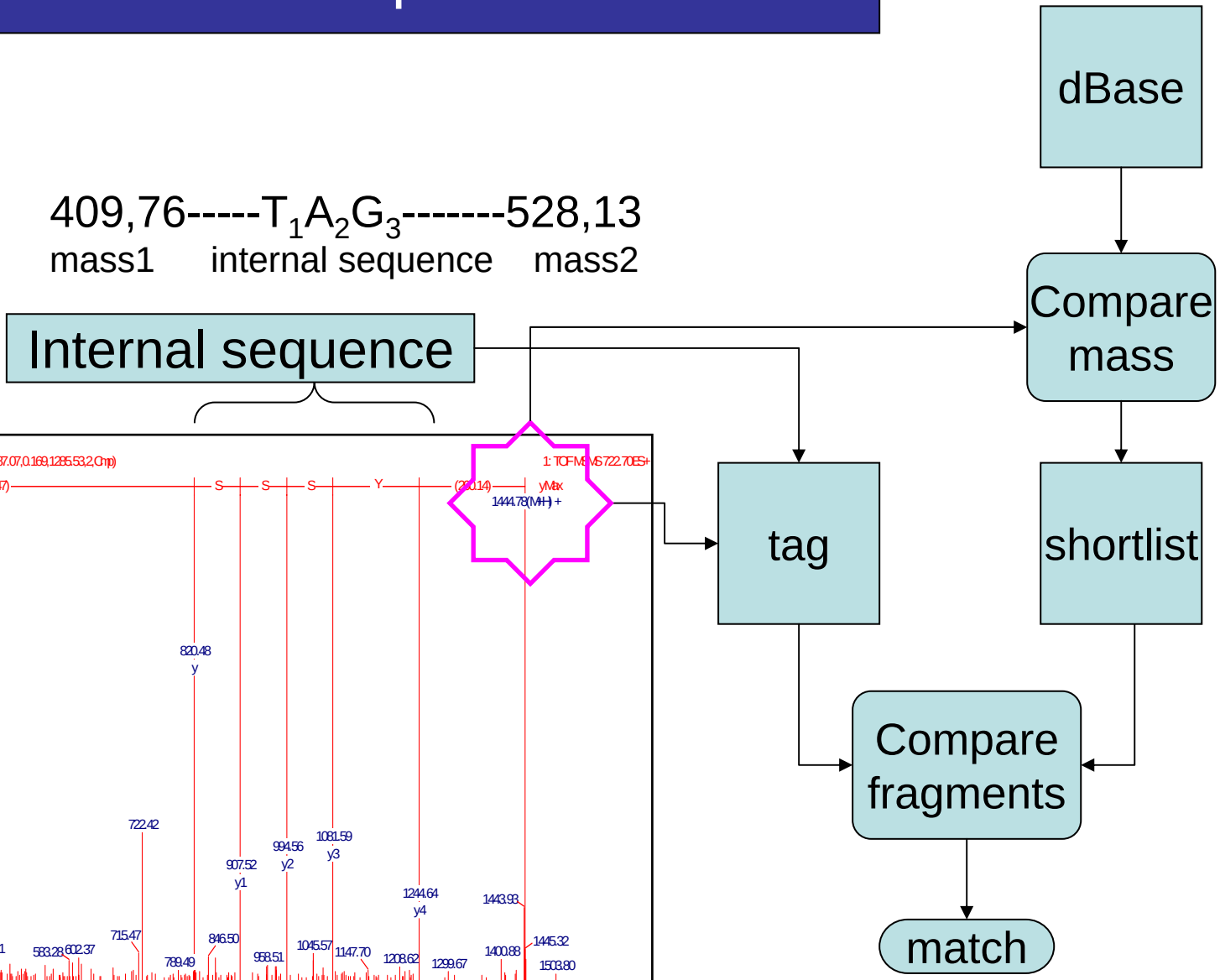


MSMS: classical sequence TAG

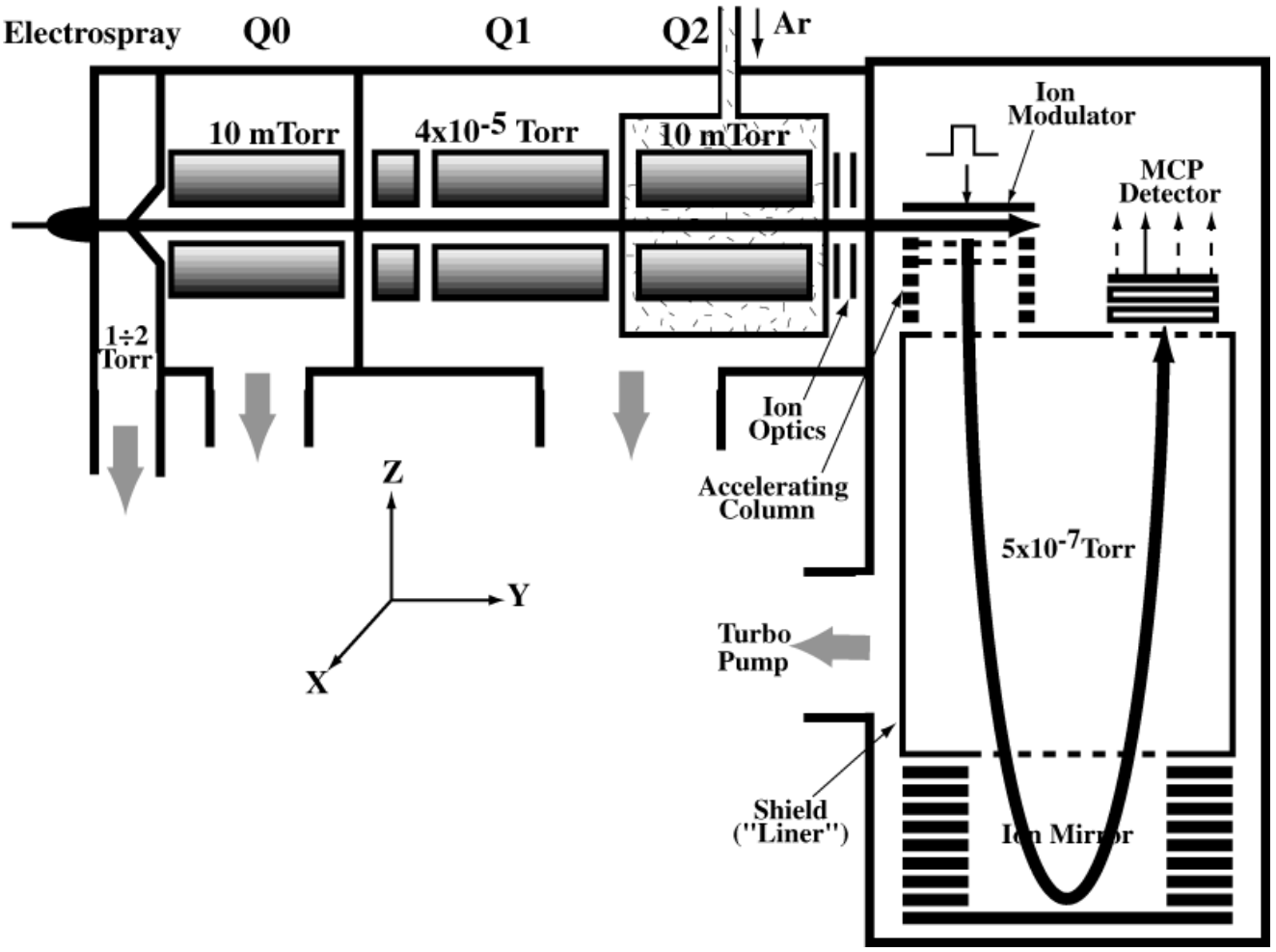
Format: 409,76-----$T_1A_2G_3$-------528,13

mass1 internal sequence mass2





Electrospray ionisation tandem mass spectrometry



Quantitative proteomics

How much do you see?

Suppression effect

- Different peptides have different propensities to ionize
- The ionization of some peptides is suppressed by the presence of other, better ionizable peptides
- So, the signal intensity of a peak in a mass spectrum of a mixture of peptides is not a measure of the concentration of the peptide concerned

A method for quantitative proteomics should …

- Be reproducible
- Be reliable
- Be sensitive
- Cover many proteins
- Have a large dynamic range

ABSOLUTE vs relative

- Absolute :
  - Concentration
  - Copy number / cell
- Relative:
  - up – down regulation
  - comparison of states

Methodology

- Gel based
  - Staining + Image analysis
  - Prederivatized proteins
- Without gels
  - Isotope free vs isotope labeling
  - Multi-dimensional LC
  - Affinity separations

Gel based methods

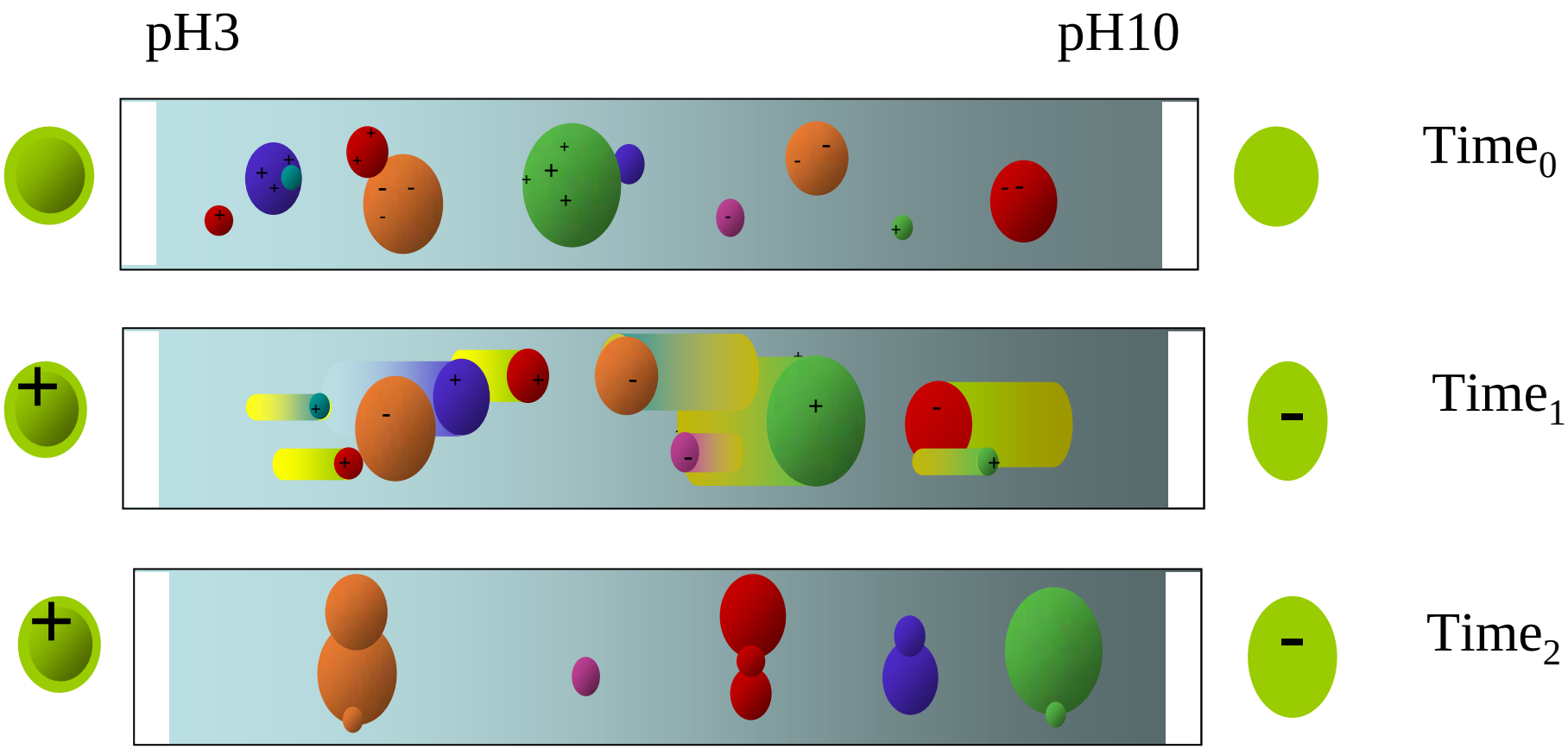
1. Stain (CBB/Ag/Sypro)
2. Image analysis
3. Repeat for reproducibility

Alternatively:

Fluorescent derivatization prior

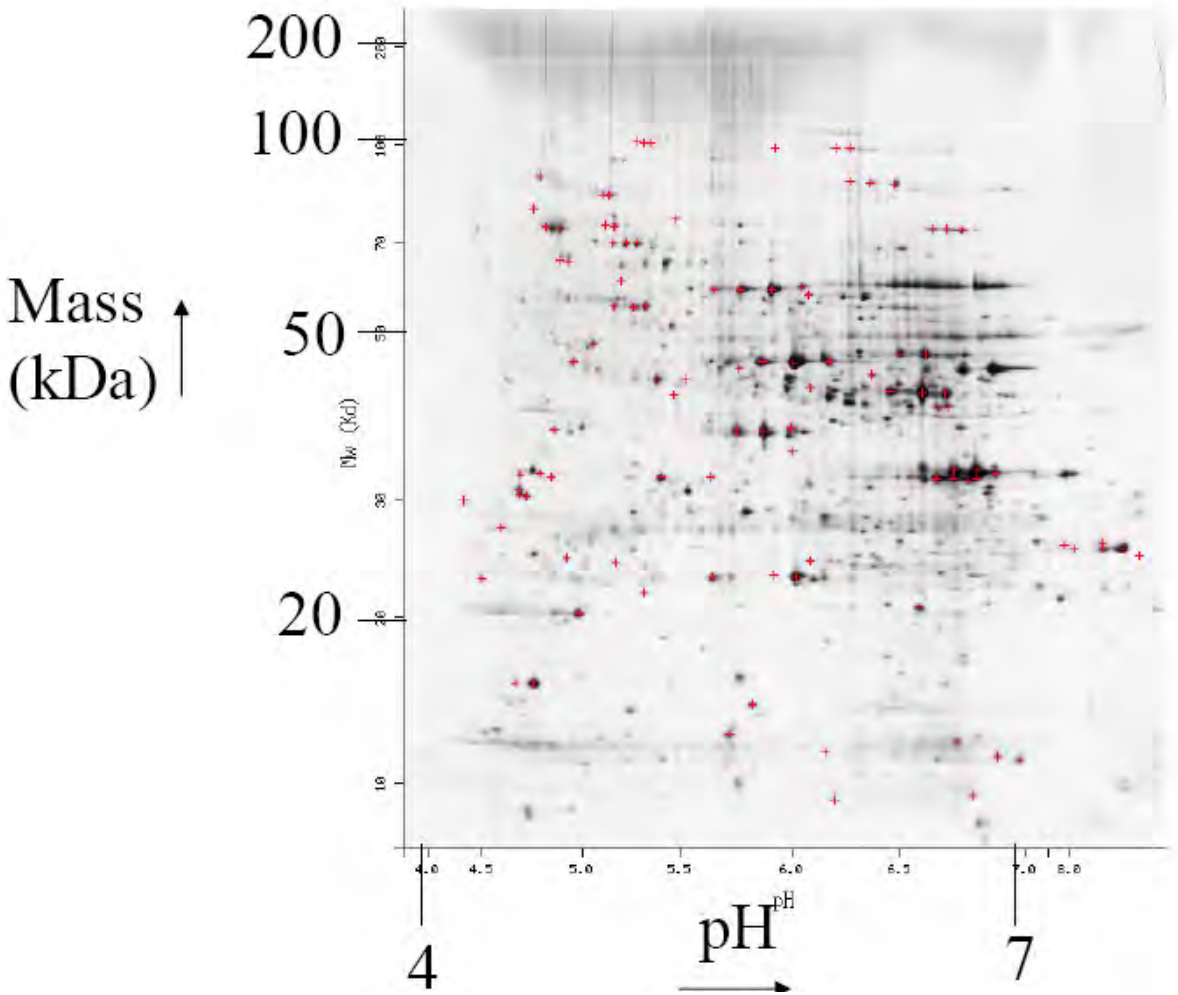
to 2D-GE

1st Dimension: Iso Electric Focussing

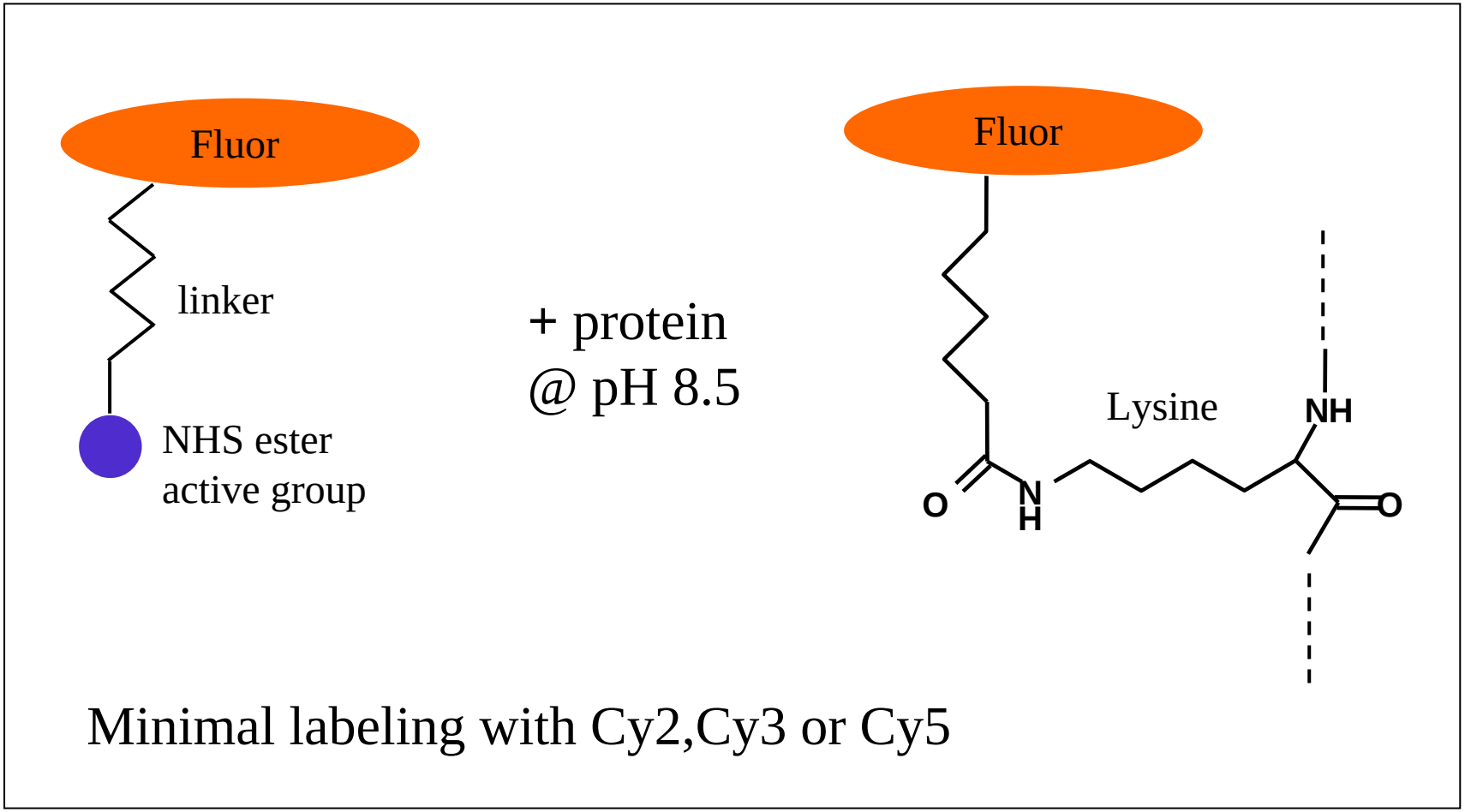


2nd Dimension: SDS Page

2-D gel of *S. cerevisiae* proteins



CyDye DIGE fluor Structure



Minimal labeling with Cy2,Cy3 or Cy5

Sensitivity: 125 pg/protein, linear response over 10^5 range.
Compare: Coomassie >100 ng and Silver >10 ng.

DIGE: Difference Gel Electrophoresis

Protein extract 1

Label with fluor 1

Mix labeled extracts

Protein extract 2

Label with fluor 2

Separate by 2-D PAGE

Excitation wavelength 1

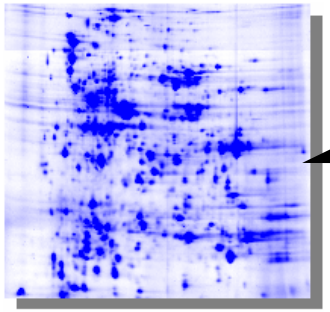
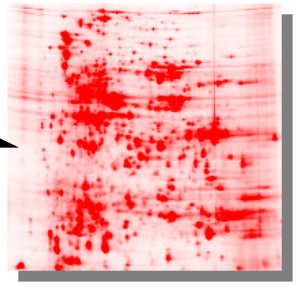
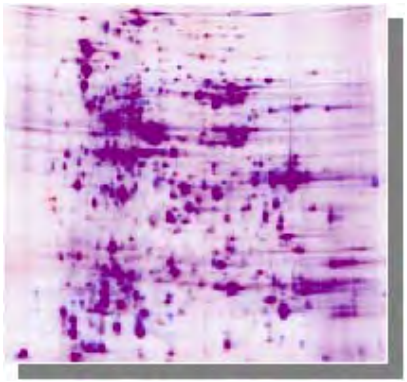


Image gel



Excitation wavelength 2

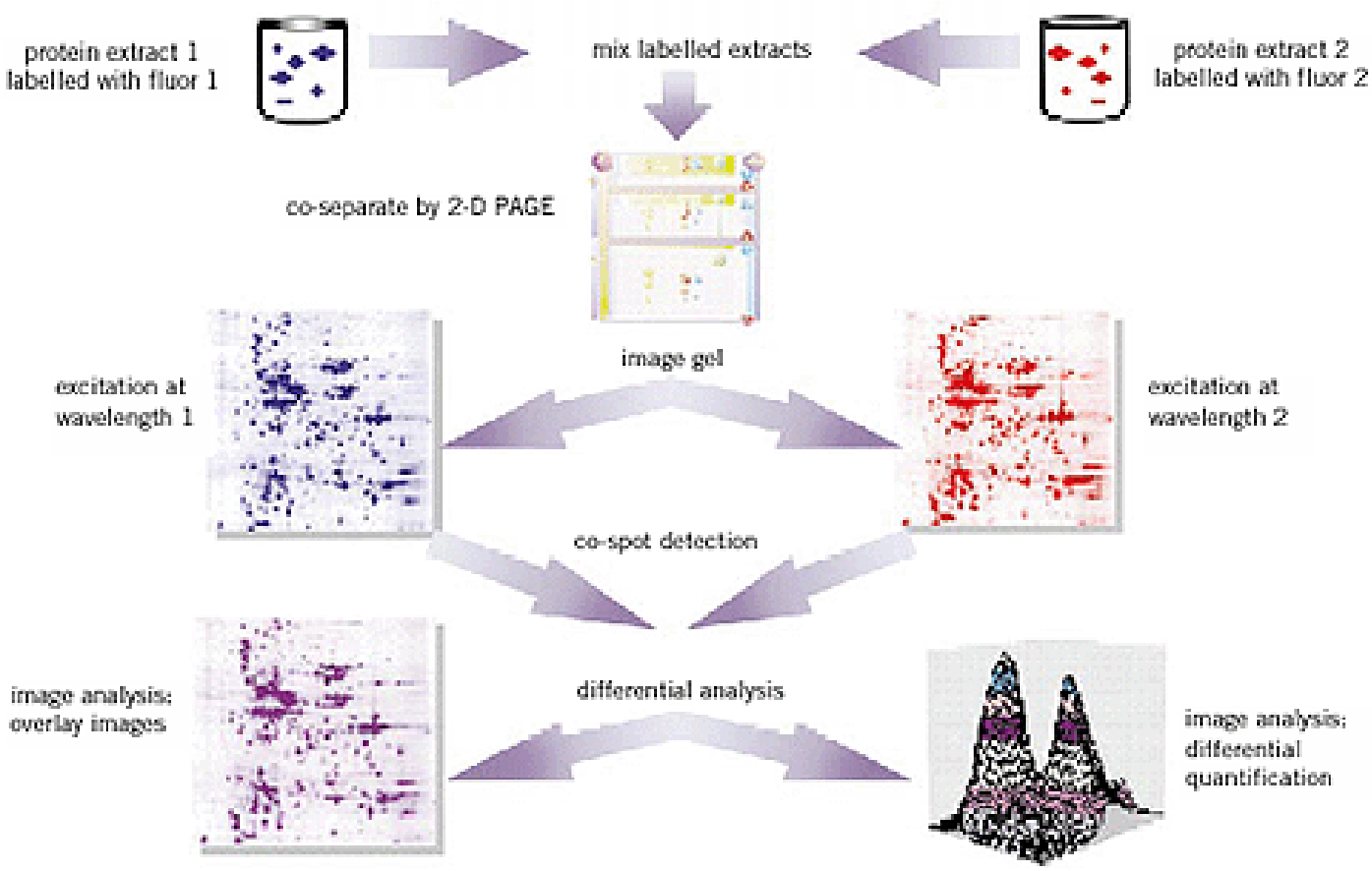
Image analysis: overlay images



Analysis of difference

Image analysis: data quantitation

DeCyder procedure



Gel based methods

Dynamic range	Stain: 10^3 (?) Fluor: ???
Reproducibility	Stains: worse than 2 Fluor: up to 1.3 (ideal)
Separating power	Thousands of SPOTS (not equal to proteins)
Bias	Soluble, abundant

Limitations of 2-D gel electrophoresis

- large proteins (> 100 kDa) are underrepresented
- hydrophobic membrane proteins (30% of entire proteome) tend to aggregate during the first dimension and do not enter the second dimension gel
- highly basic and acid proteins are problematic
- relatively low dynamic range $10^3 - 10^4$ of 2-D gels prevents detection of less abundant proteins

Limitations of 2-D gel electrophoresis

- relatively low dynamic range $10^3 – 10^4$ of 2-D gels prevents detection of less abundant proteins

Isotope labels or isotope free

Mass spectrometry is not quantitative

- Ionisation propensity
- Co-suppression
- Co-enhancing

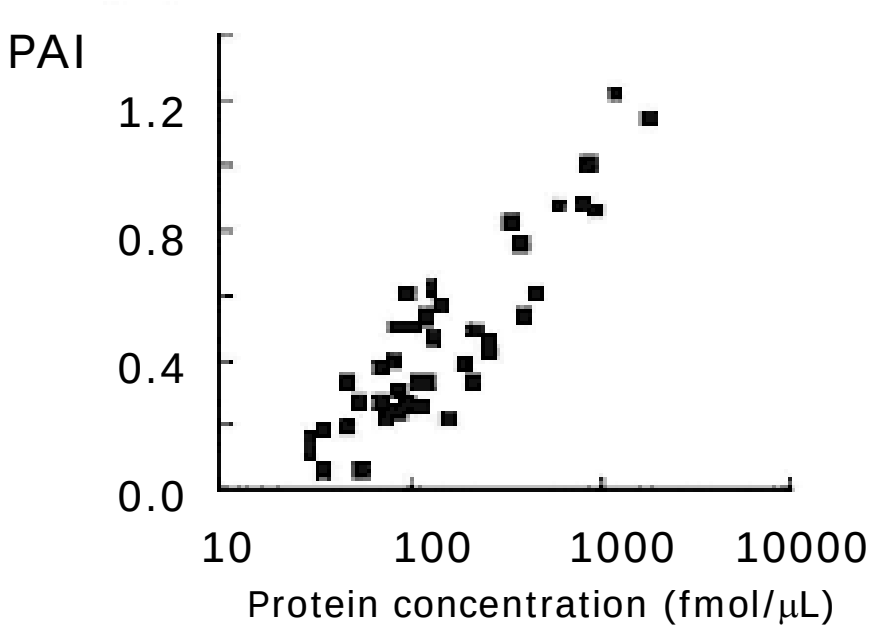
Use of (stable) isotopes serves as an internal standard (chemically identical)

Protein Abundance Indices (PAI)

$$PAI = \frac{N_{obsd}}{N_{obsbl}}$$

N_{obsd} the number of observed peptides per protein

N_{obsbl} the number of observable peptides per protein



Isotope labels

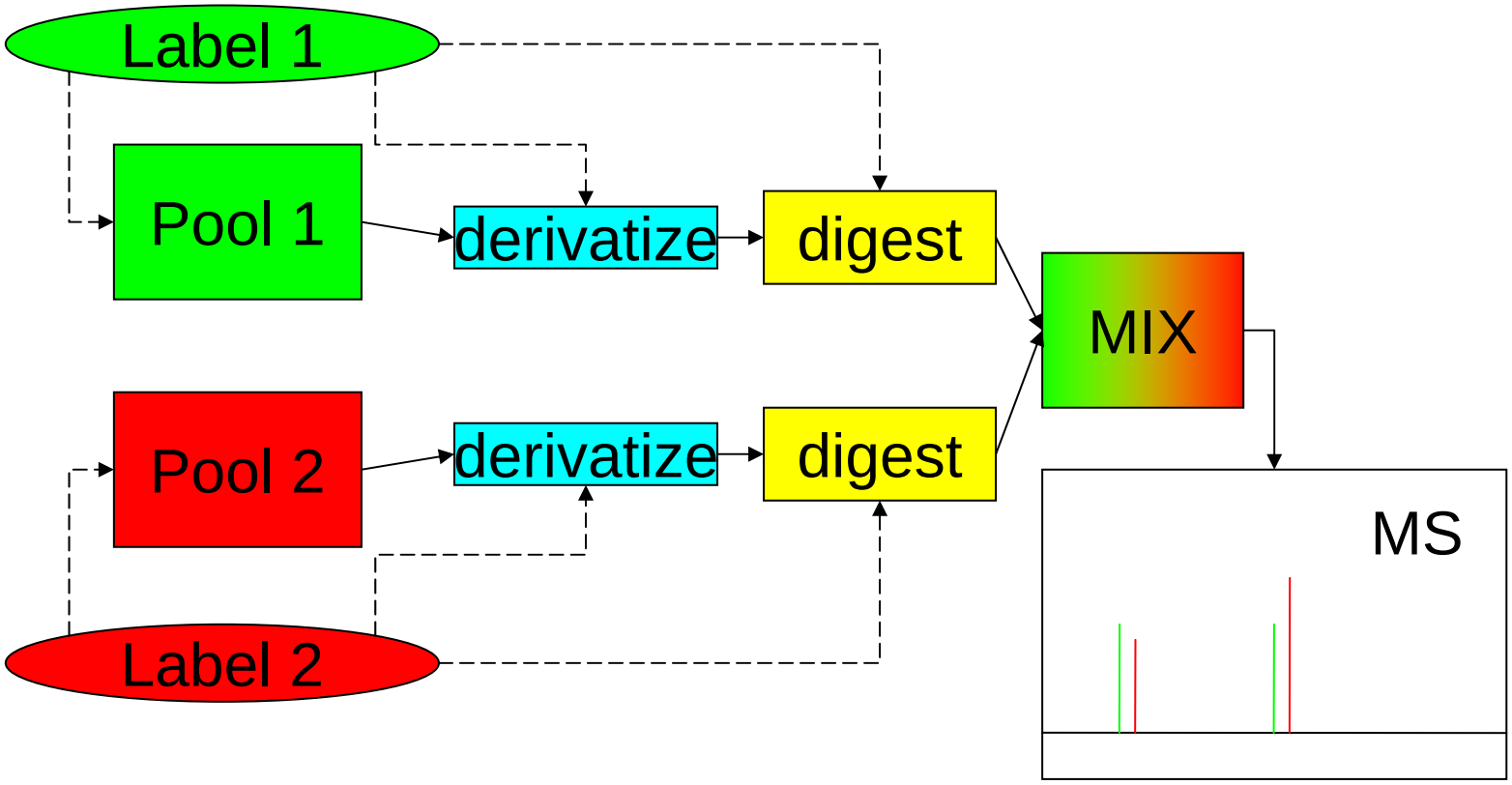
Labels can be incorporated via:

1 Culture media

2 Derivatization of reactive groups on the proteins

3 Proteolytic cleavage

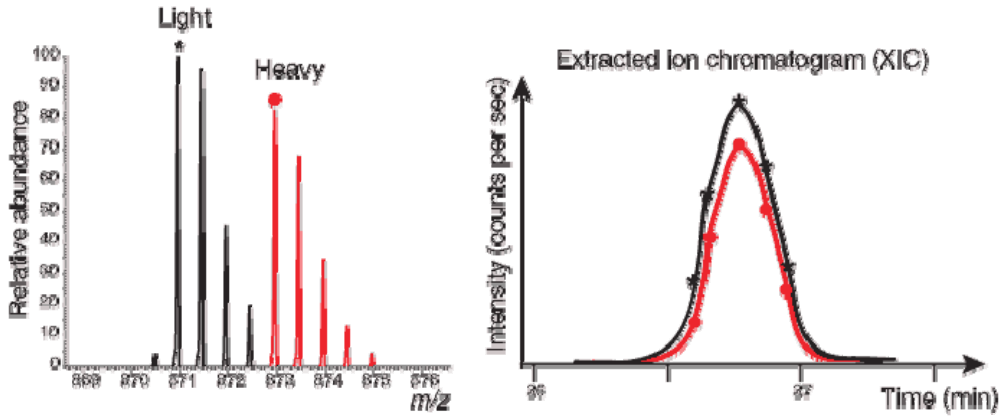
Label introduction: techniques



Isotope ratio for quantification

Isotope labels

- Extracted ion chromatograms (XIC)
  - Peak area of a selected *m/z* value



When to mix in the label?

Culture media:

-early in the procedure

-identical treatment throughout the preparation process

-keeping standards standard may be difficult

Chemical probes/digestions

-irreproducibility's: check by inverse labeling

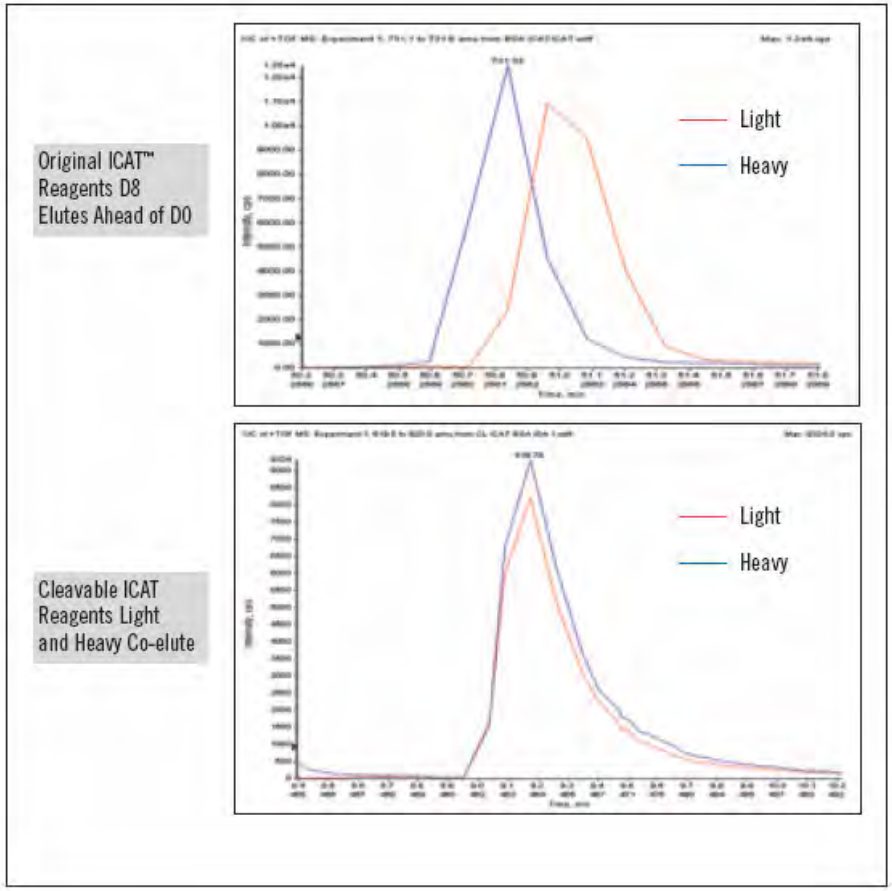
Natural isotopic abundances of common elements

hydrogen	H	1	99.985
	D	2	0.015
carbon	^{12}C	12	98.9
	^{13}C	13	1.1
nitrogen	^{14}N	14	99.64
	^{15}N	15	0.36
oxygen	^{16}O	16	99.76
	^{17}O	17	0.04
	^{18}O	18	0.2
fluorine	^{19}F	19	100
phosphorus	^{31}P	31	100
sulpher	^{32}S	32	95.02
	^{33}S	33	0.76
	^{34}S	34	4.22
chlorine	^{35}Cl	35	75.77
	^{37}Cl	37	24.33
bromine	^{79}Br	79	50.5
	^{81}Br	81	49.5
iodine	^{127}I	127	100

Examples of isotope post-hoc labels

H/D	ICAT; D-labeled amino-acids (SILAC) Caution: D changes RP-HPLC retention
$^{12}C/^{13}C$	ICAT the sequel
$^{14}N/^{15}N$	Incorporated via growth media Can be used to calculate # of nitrogens
$^{16}O/^{18}O$	Incorporated by proteases Can be combined with other procedures

Deuterium alters RP elution



Isotope labels

Labels can be incorporated via:

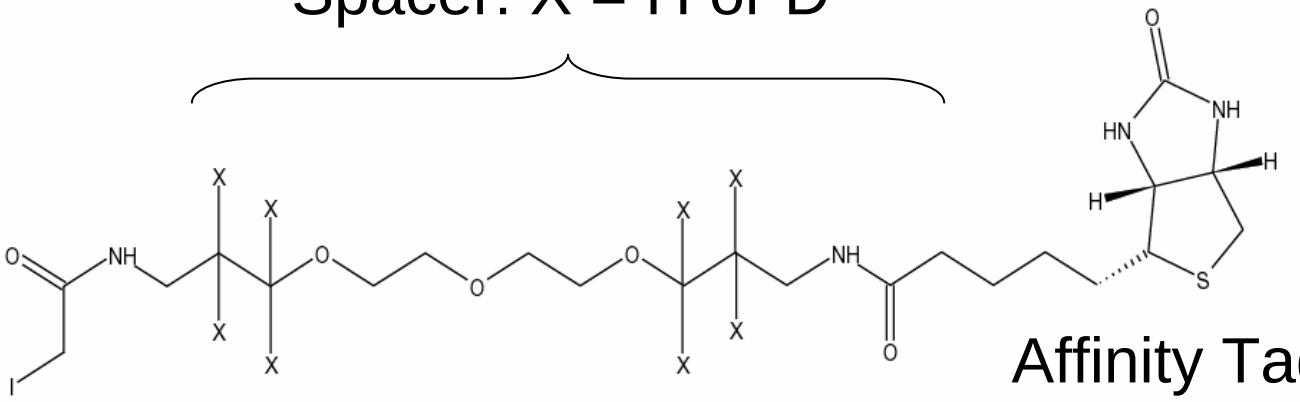
1 Culture media

2 Derivatization of reactive groups on the proteins

3 Proteolytic cleavage

ICAT structure

Spacer: X = H or D

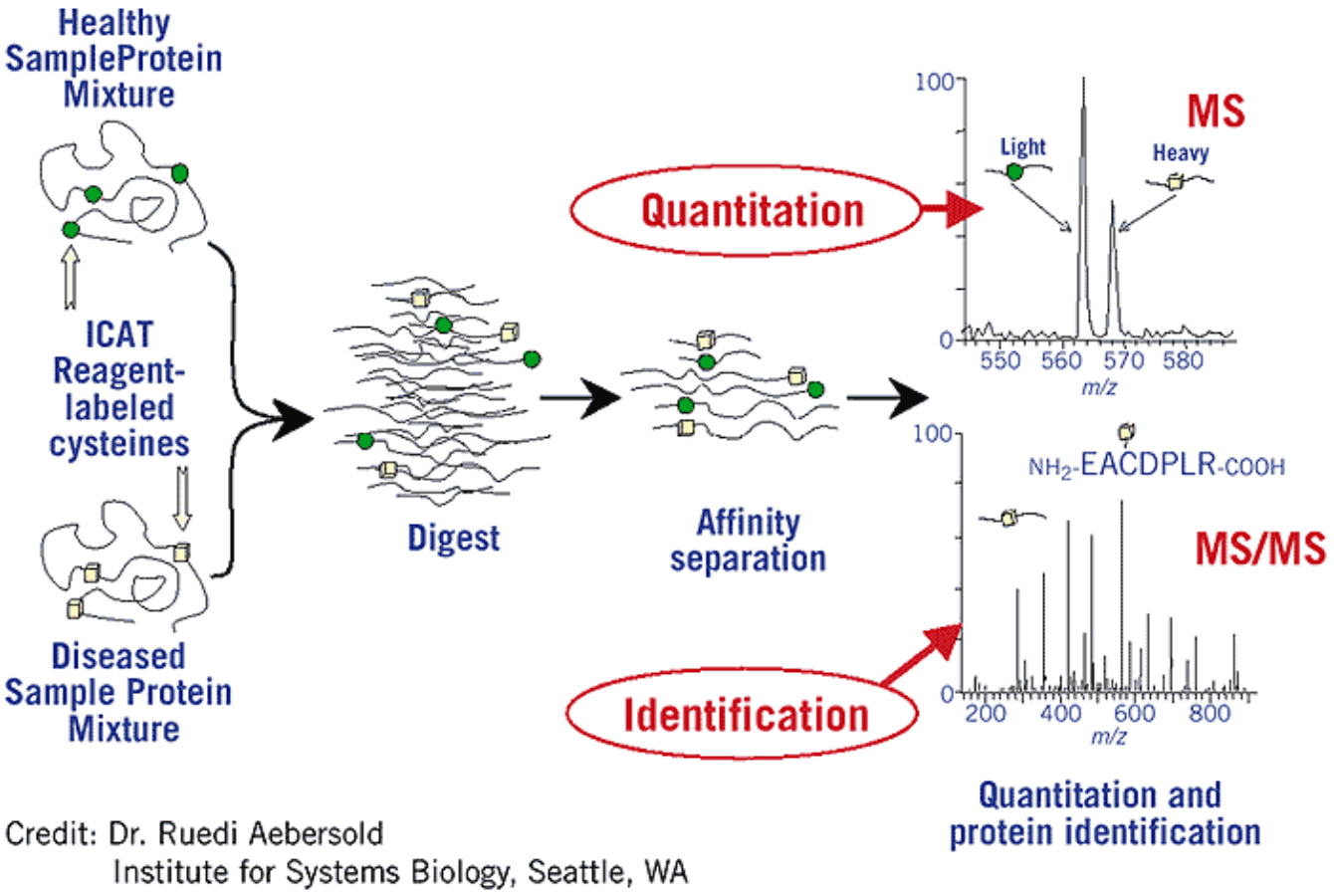


SH reactive

Affinity Tag
(Biotin)

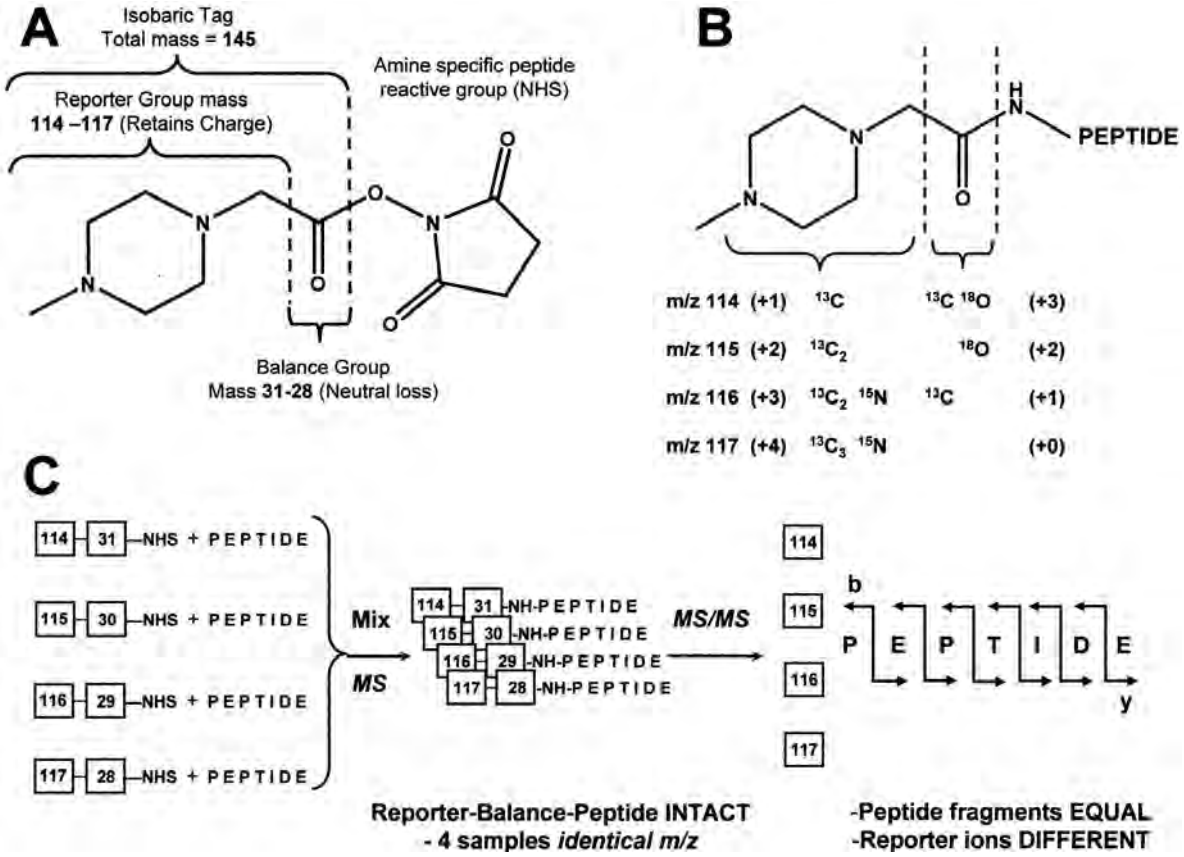
D0:+$C_{20}H_{35}N_4SO_5$ / -H
mass change = 442.225/cystein (mono) ; 442.5799 (av).
D8:+$C_{20}H_{27}D_8N_4SO_5$ / -H
mass change = 450.2752/cystein (mono) ; 450.63 (av).

ICAT procedure



Credit: Dr. Ruedi Aebersold
Institute for Systems Biology, Seattle, WA

iTRAQ



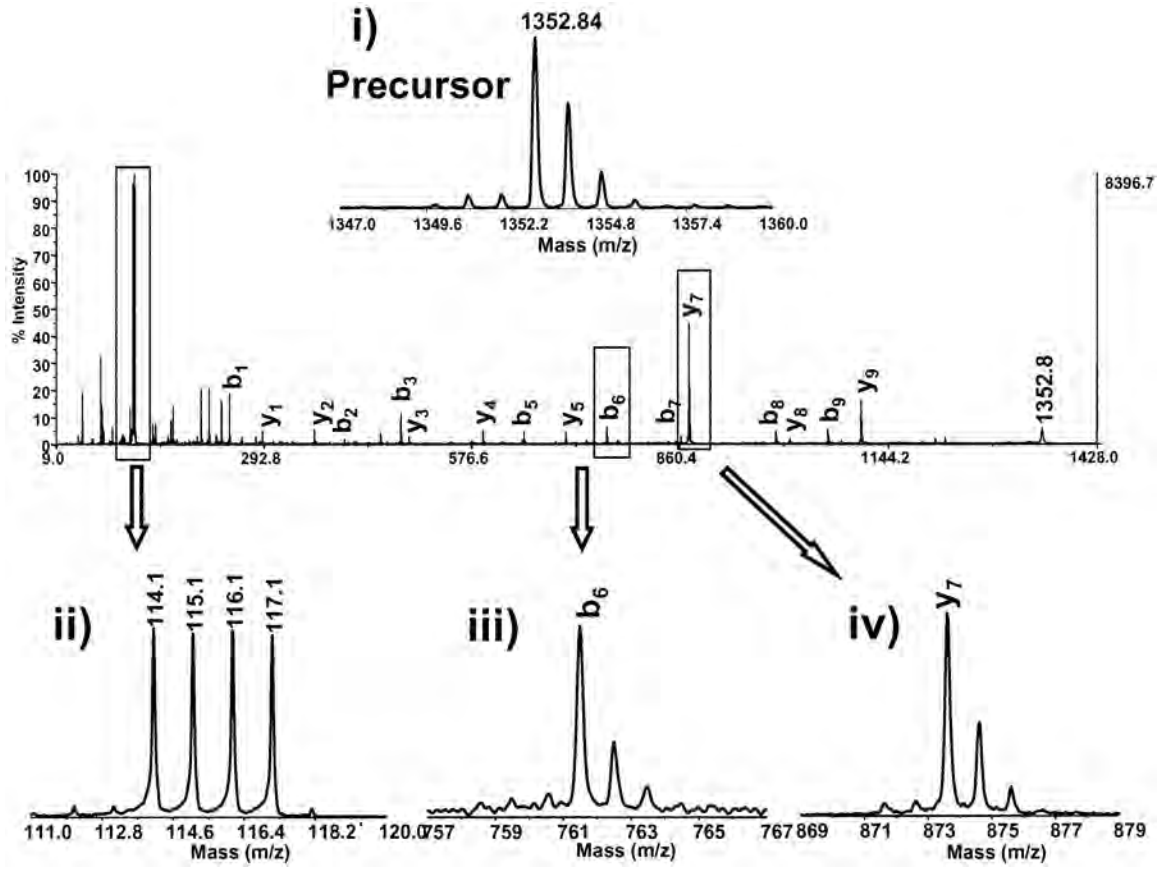
Signal is concentrated in MS overview:

High sensitivity

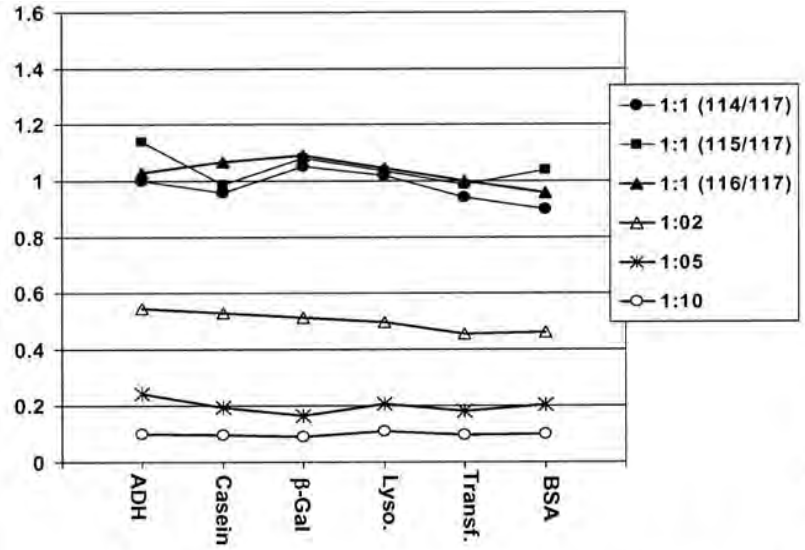
iTRAQ

Quantification on daughter ions in MSMS:

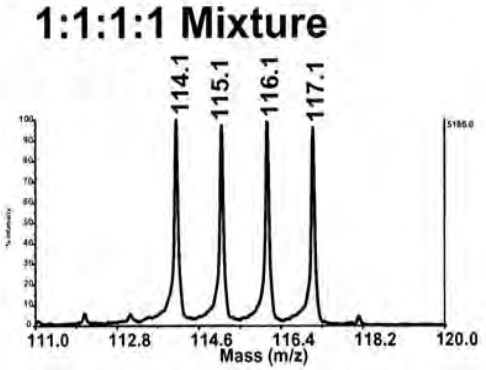
Improved S/N ratio

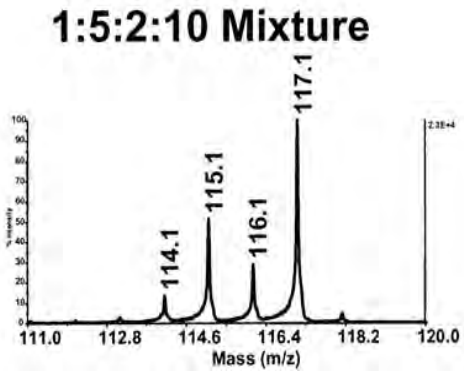


iTRAQ



Ratio	Mean	SD
1:1	1.03	0.16
1:2	0.514	0.12
1:5	0.204	.045
1:10	0.097	0.023

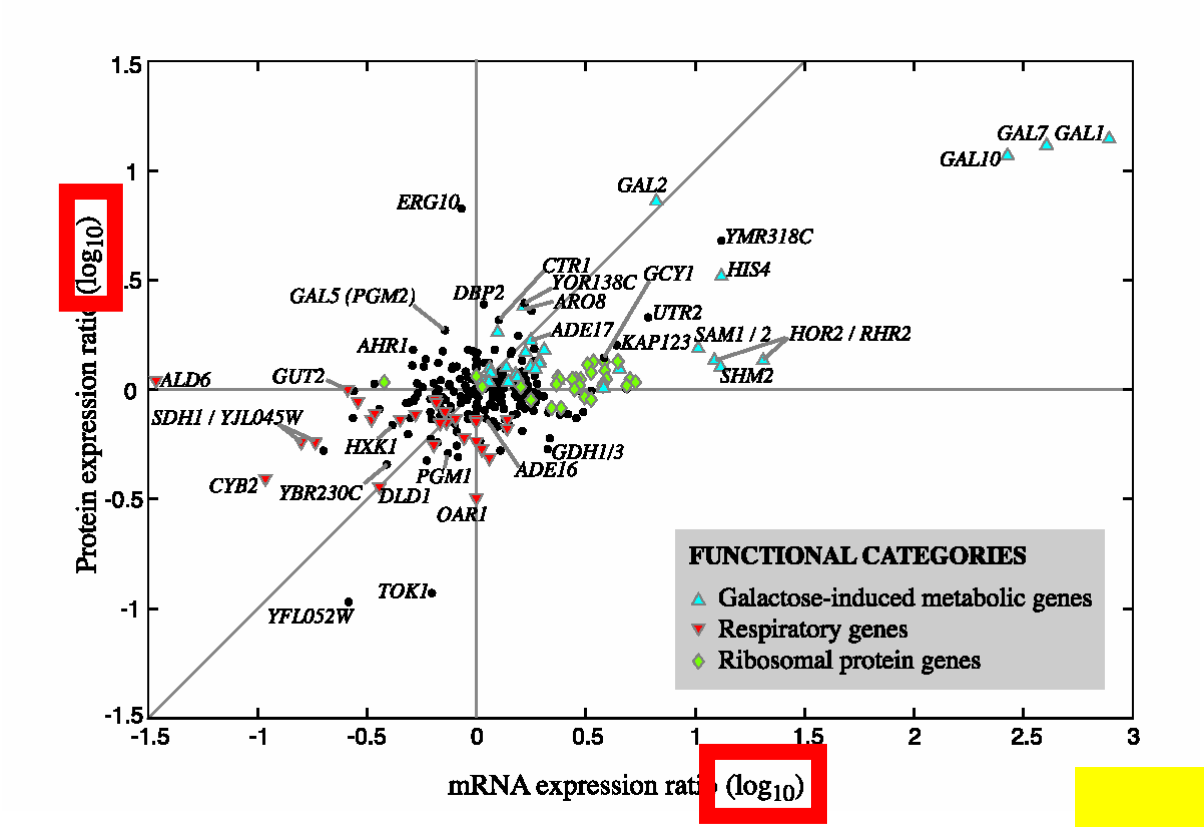




Isotope label methods

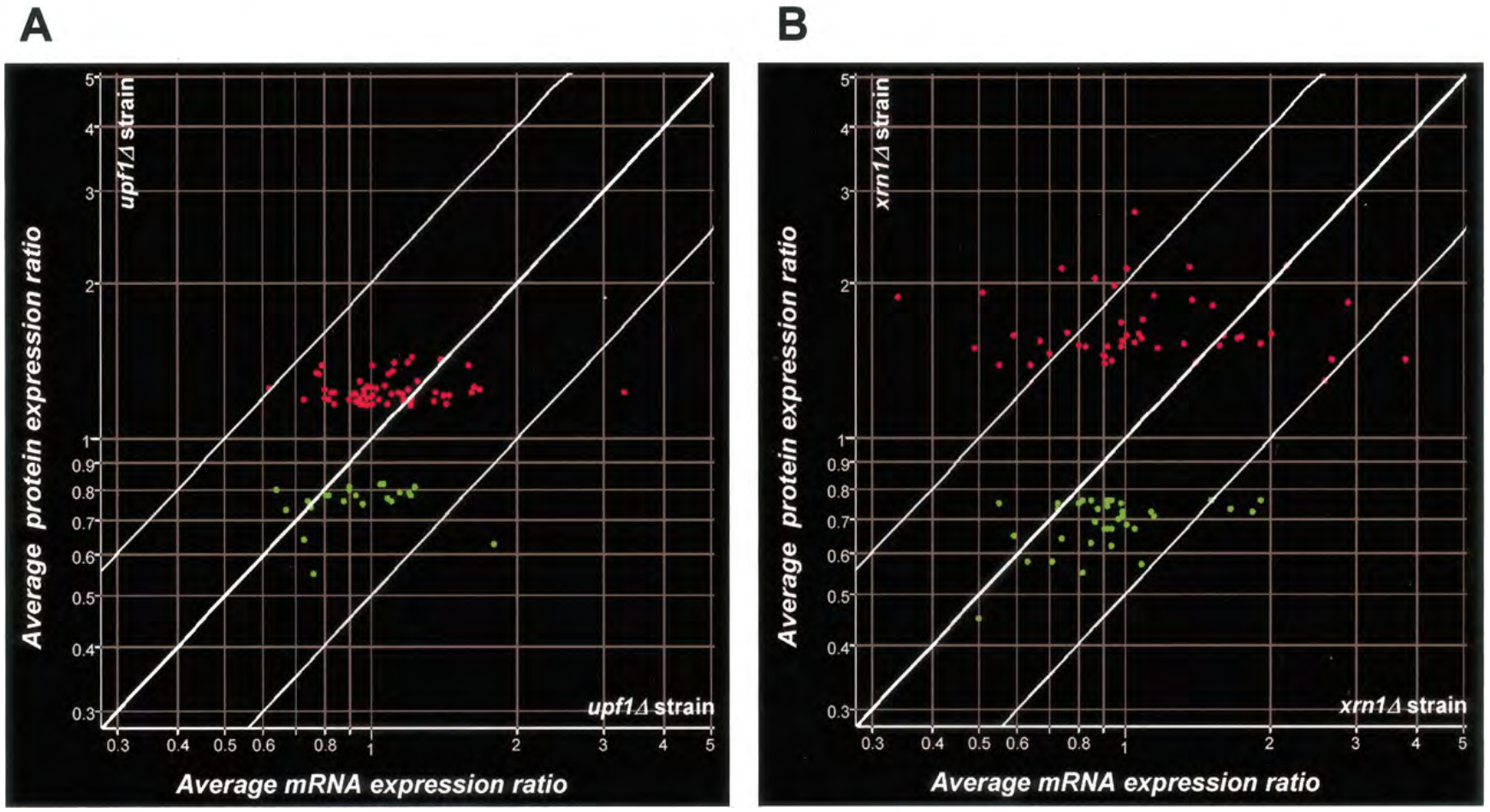
Dynamic range	10^3 : > 10 both up and down
Reproducibility	Approximately 20% is attainable
Separating power	Thousands of PEPTIDES (not equal to proteins)
Bias	Less biased than gels

Correlation RNA/protein



$r = 0.61$

Or from the iTRAQ experiment



Isotope labels

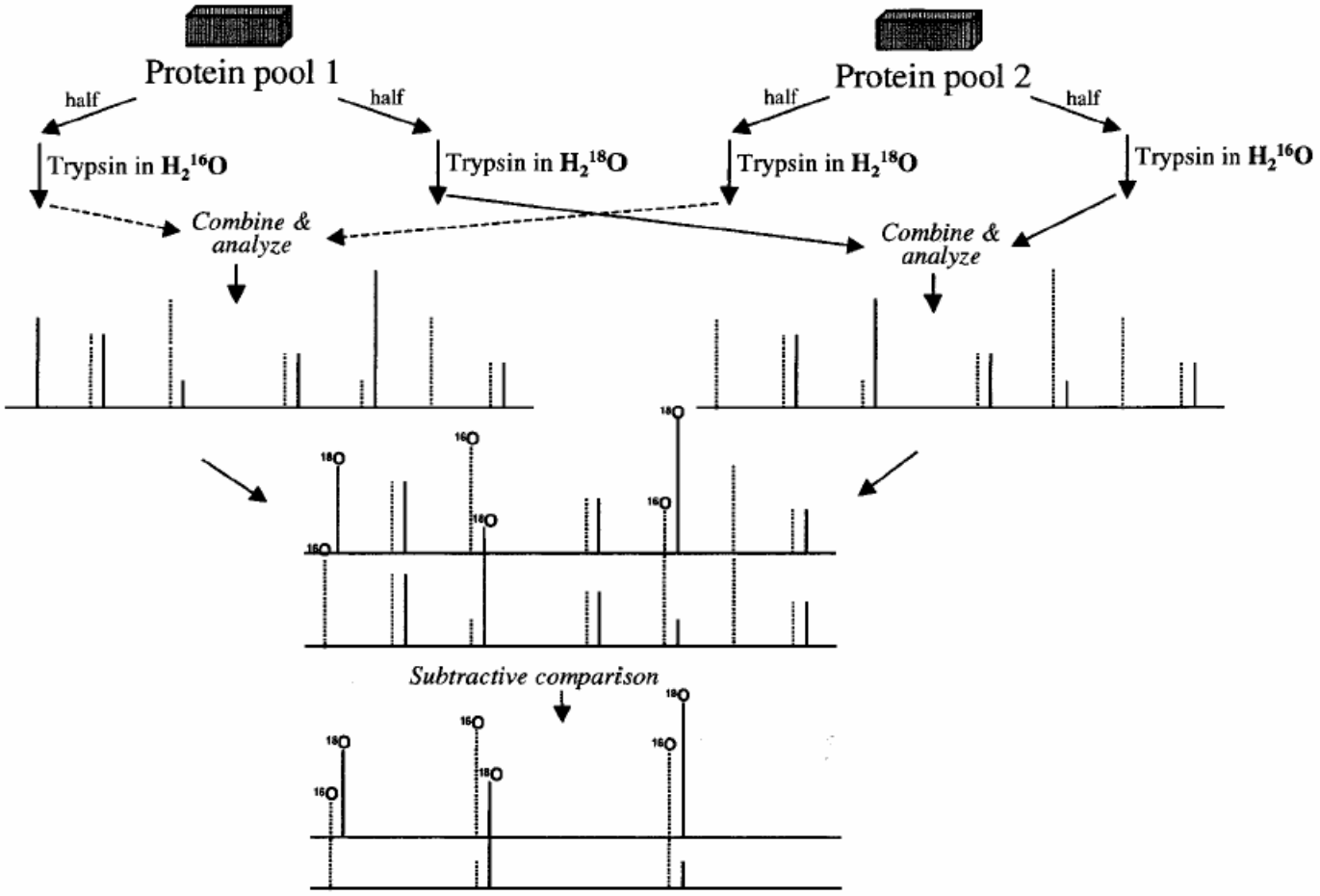
Labels can be incorporated via:

1 Culture media

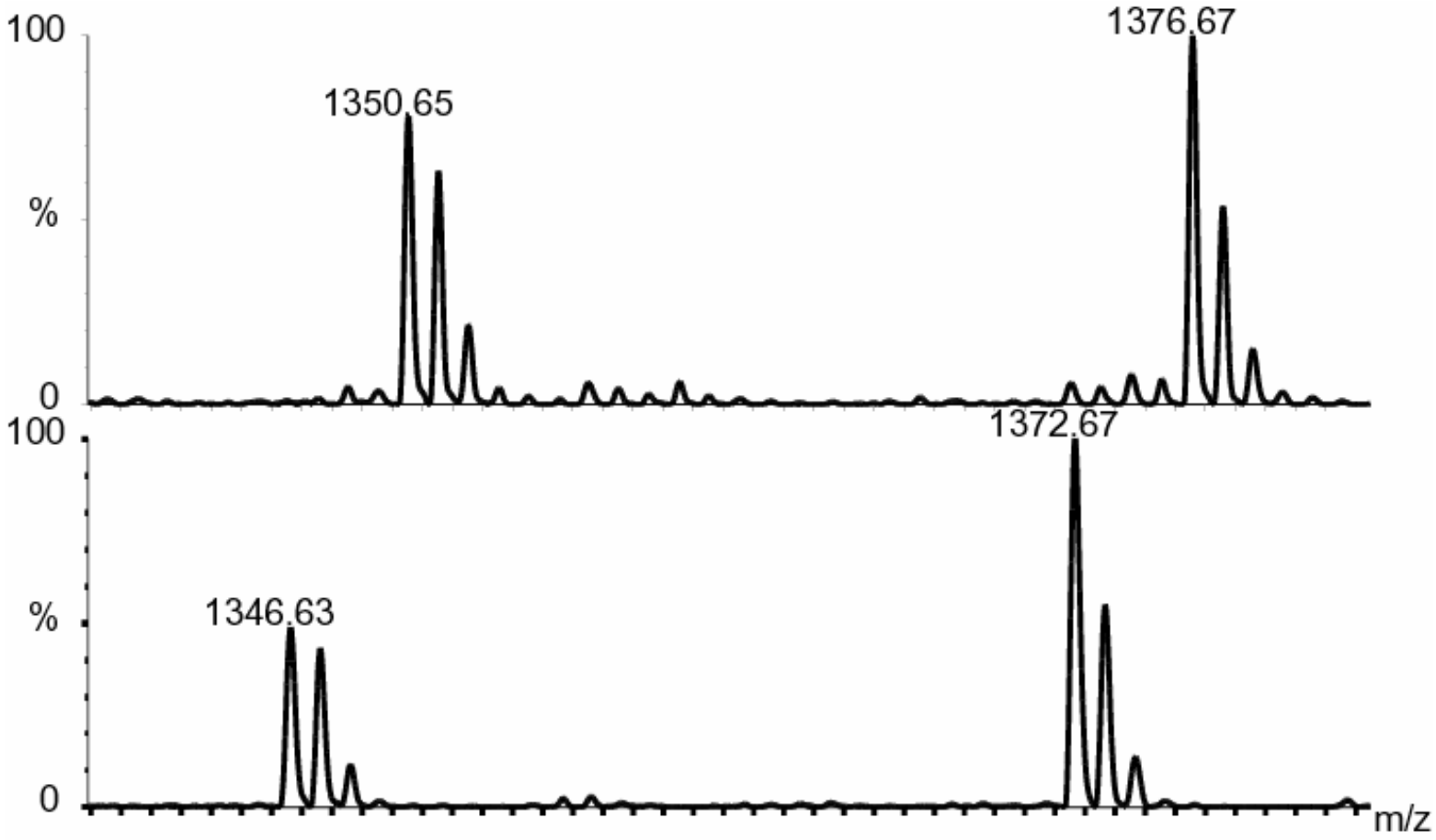
2 Derivatization of reactive groups on the proteins

3 Proteolytic cleavage

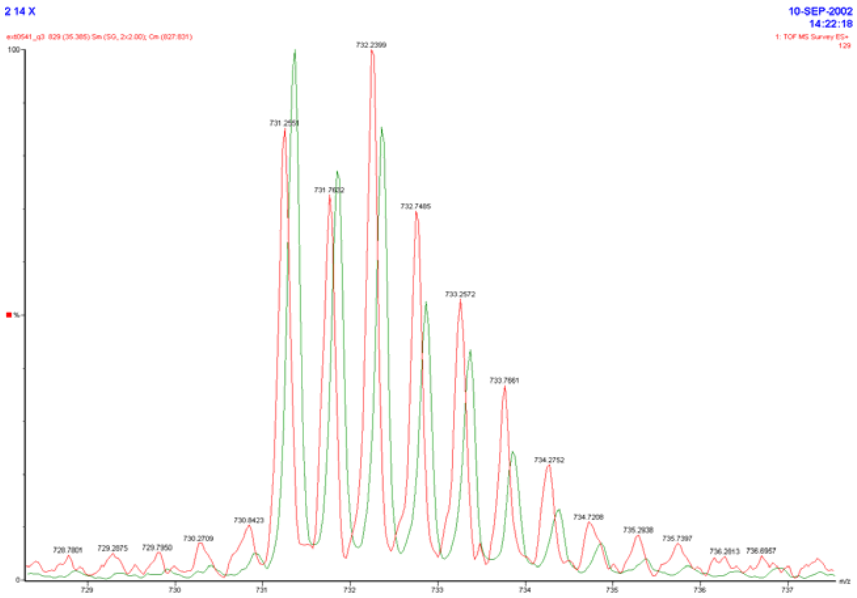
Inverse labeling with $H_2{}^{18}O$

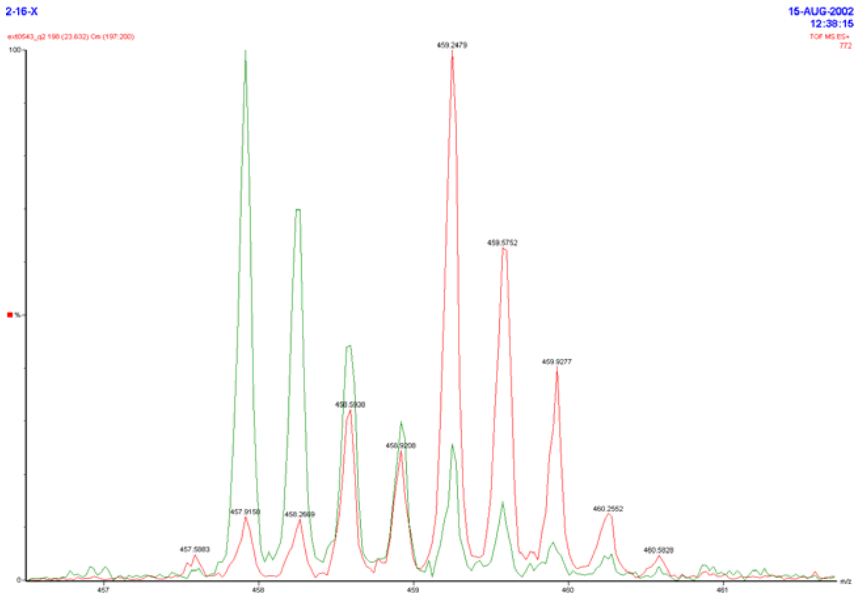


Trypsin exchanges two oxygens



Equal and different amounts





Examples of isotope labels

H/D	ICAT; D-labeled amino-acids (SILAC) Caution: D changes RP-HPLC retention
$^{12}C/^{13}C$	ICAT the sequel
$^{14}N/^{15}N$	Incorporated via growth media Can be used to calculate # of nitrogens
$^{16}O/^{18}O$	Incorporated by proteases Can be combined with other procedures

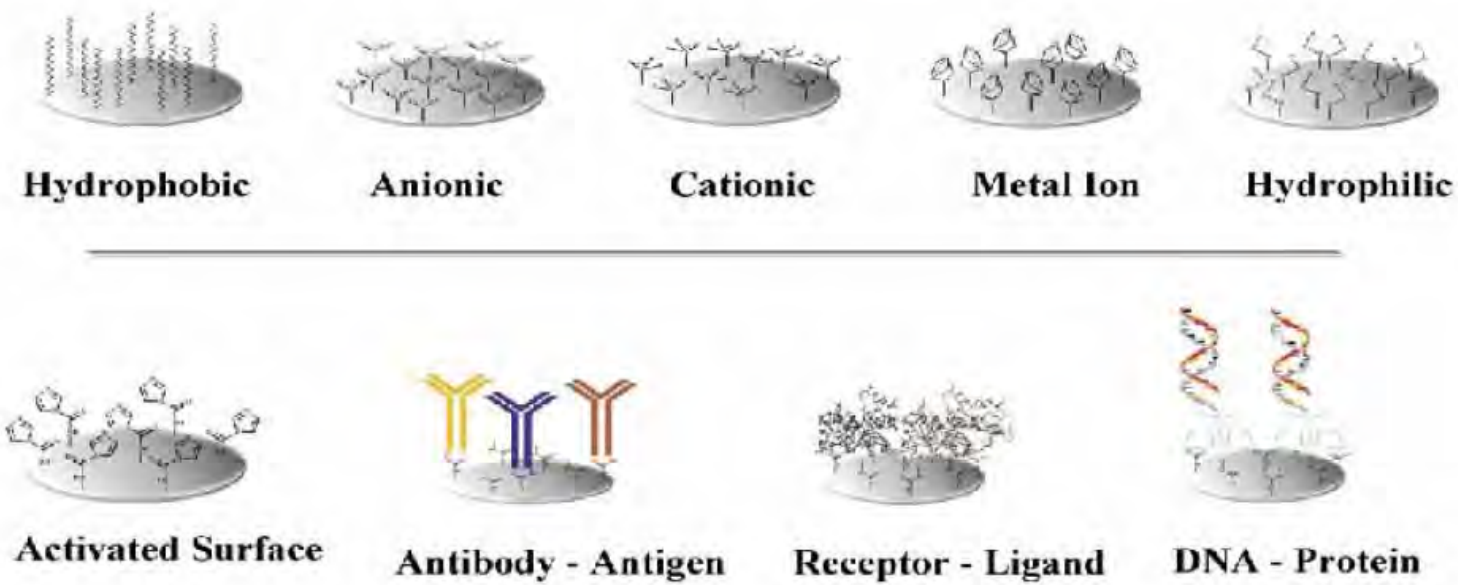
Literature

- Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol. 2005 Oct;1(5):252-62.
- Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics. 2005 Sep;4(9):1265-72.

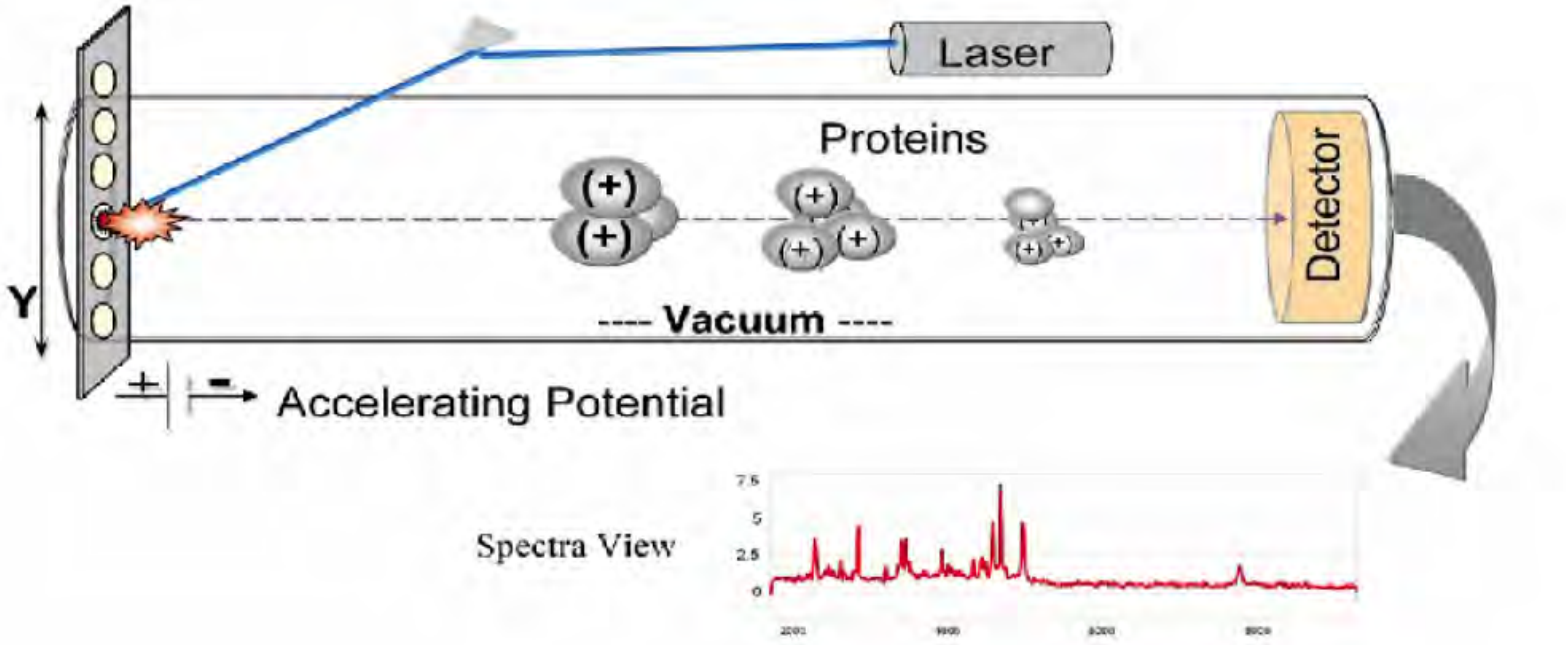
Some recent developments

- Surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS): protein profiling and biomarker identification

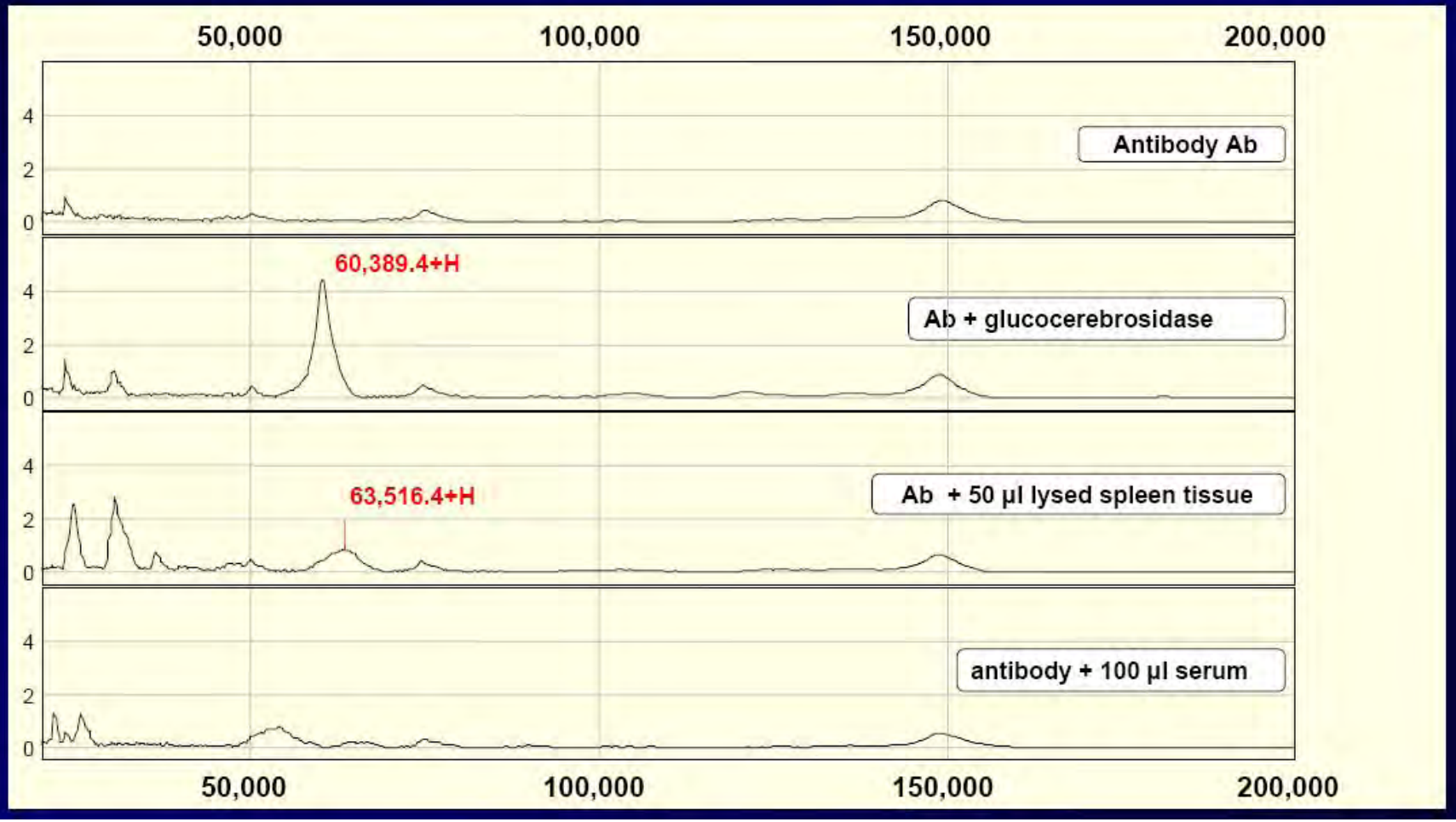
SELDI-TOF-MS : Chemically modified MALDI targets are used to retain a group of proteins



SELDI-TOF-MS : The SELDI mass spectrometer



Capture of glucocerebrosidase by antibody immobilized on PS20 chips

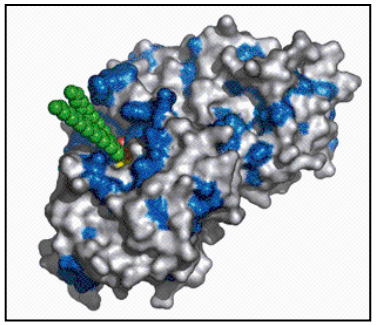


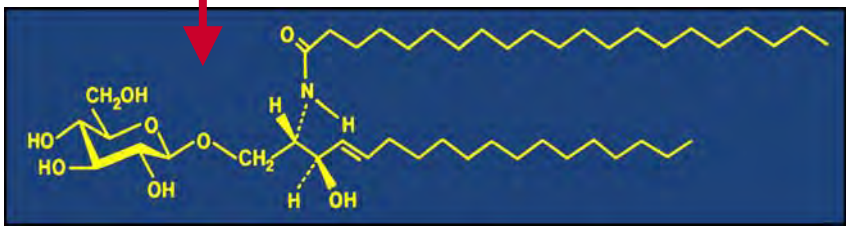
Gaucher Disease

Lysosomal storage disorder

Deficiency in lysosomal glucocerebrosidase

Substrate : glucosylceramide





Flux of glycosphingolipids is largest in monocyte-macrophage system

These cells become major storage sites (Gaucher cells >> mainly found in liver, spleen and bone marrow)

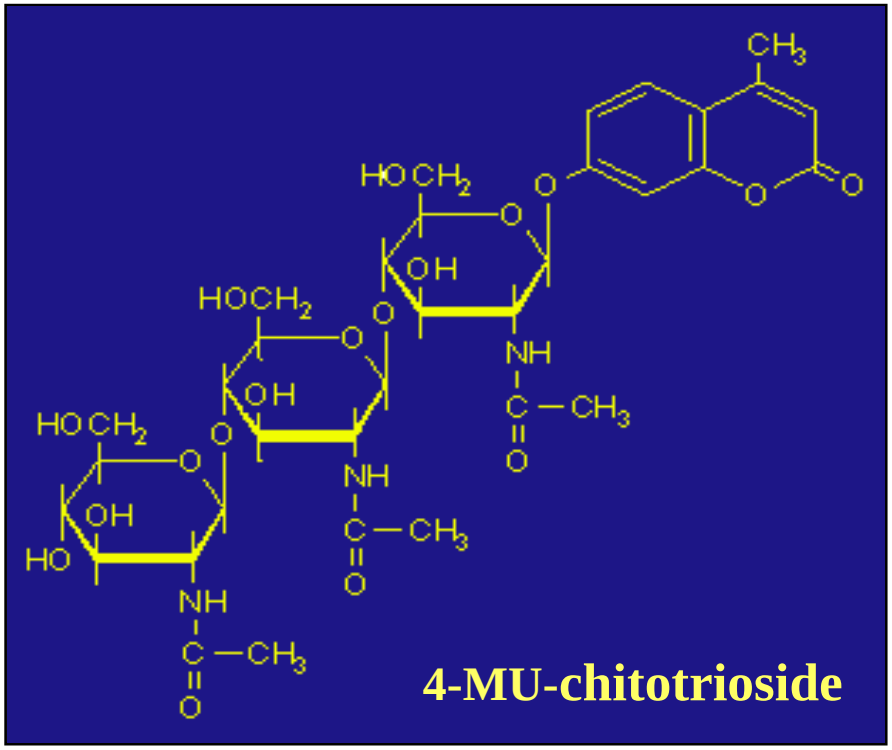
Treatment of lysosomal storage diseases

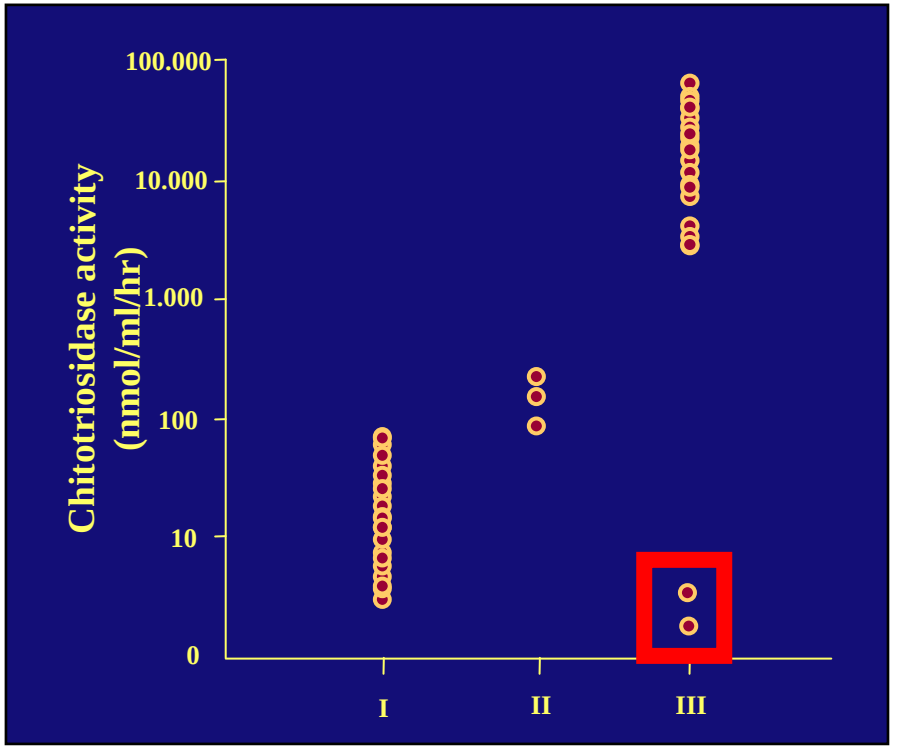
Discovery of chitotriosidase

Gaucher disease : search for markers

Patient plasma : 1000-fold increased 4-MU-chitotrioside hydrolysis

Unfortunately 10-20 % of people deficient for this enzyme





Search for novel surrogate markers

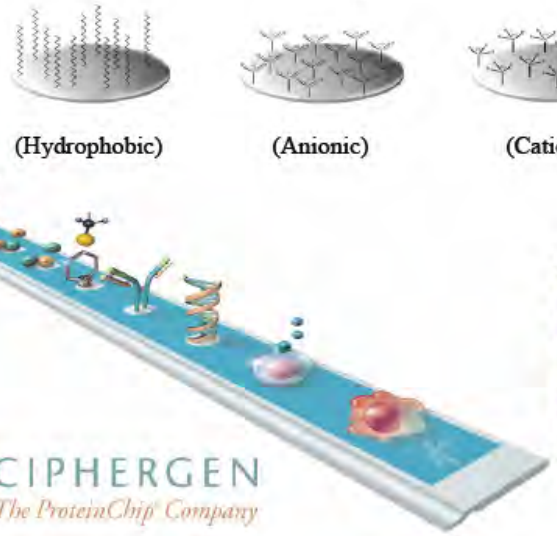
SELDI-TOF-MS

Surface enhanced laser disorption ionisation

Time of flight

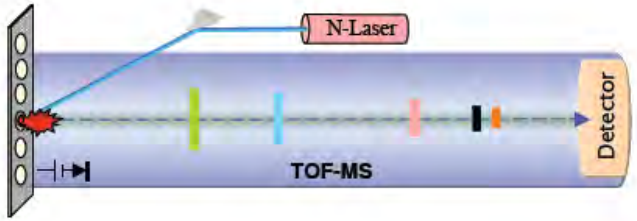
Mass spectrometry

Chemical Surfaces – Protein Expression Profiling:

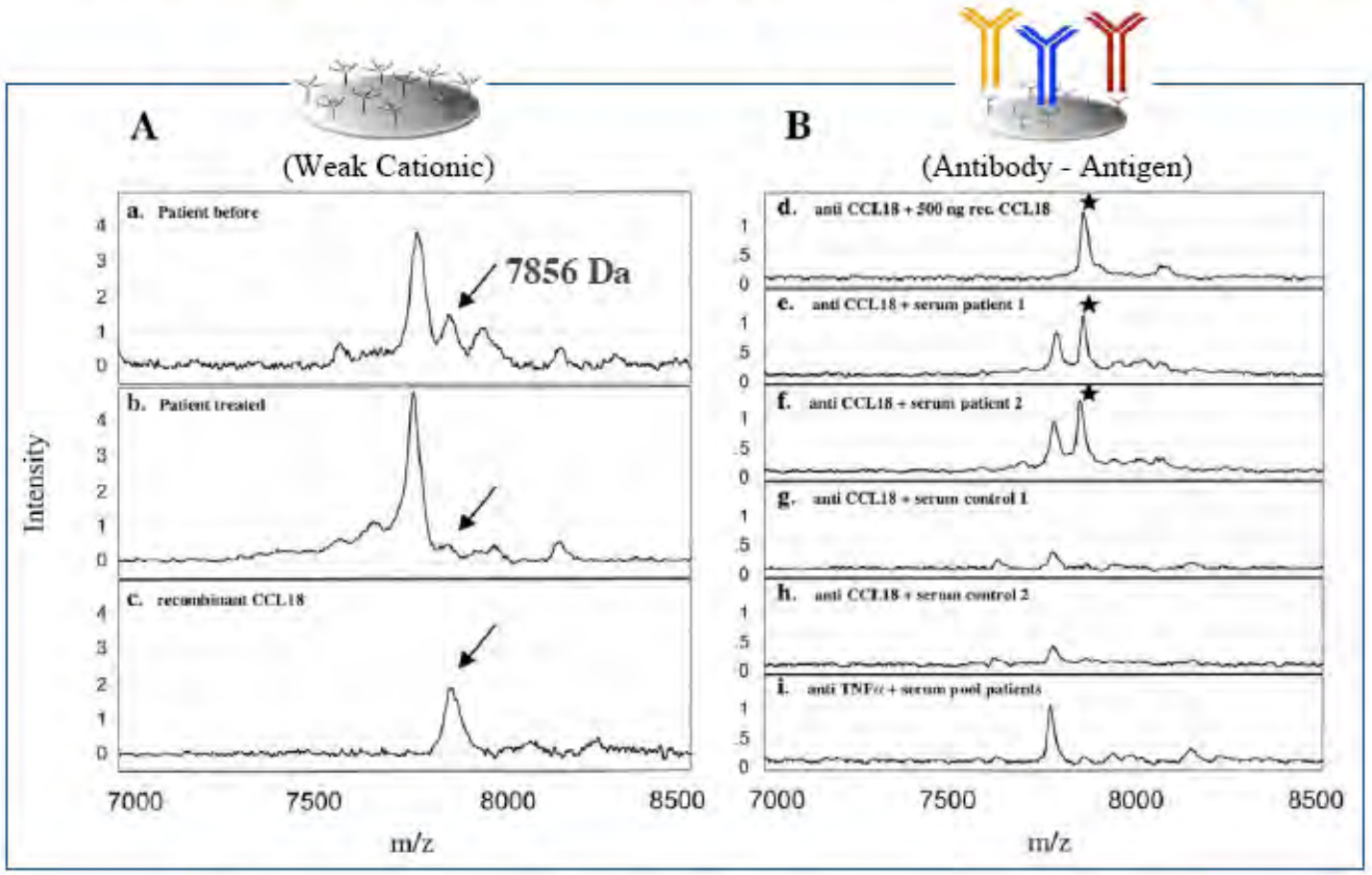


(Hydrophobic) (Anionic) (Cationic) (Metal Ion) (Normal Phase)

CIPHERGEN
The ProteinChip® Company



SELDI-TOF mass spectrometry profiling of Gaucher plasma

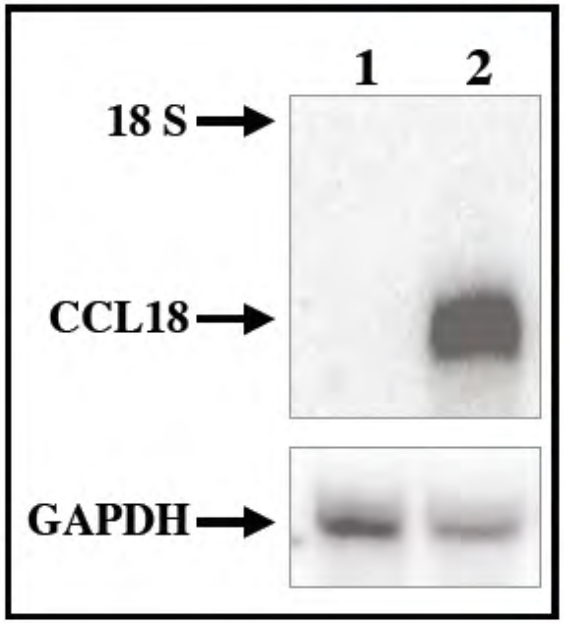


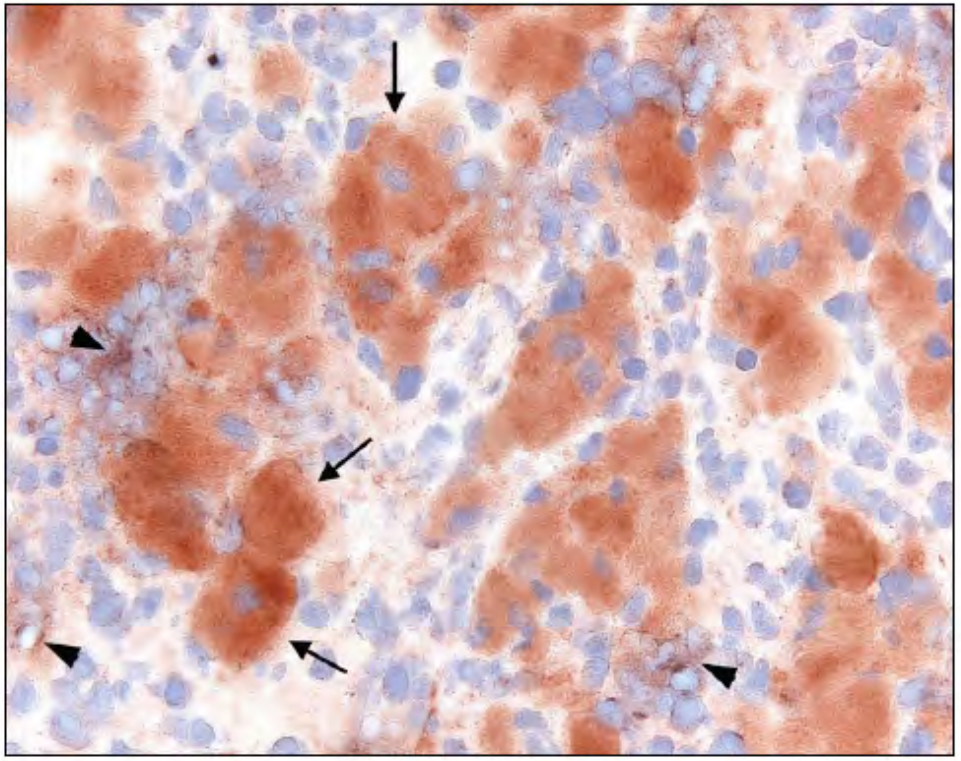
Molecular mass and basic pI similar to CCL18 (PARC)

Identified to be upregulated in Gaucher spleen

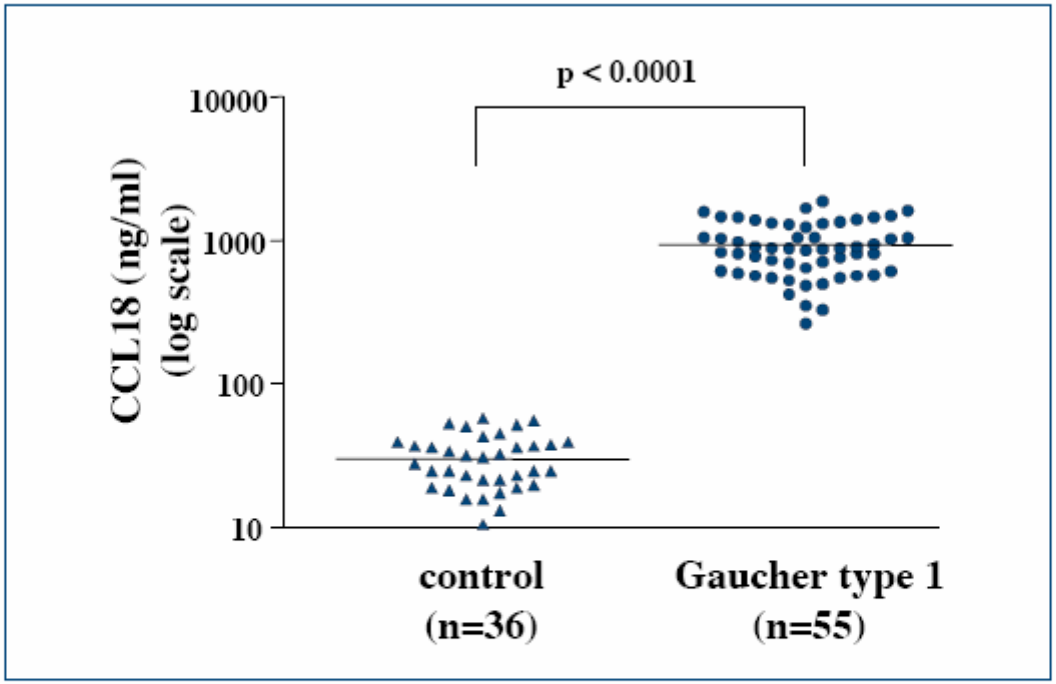
(Moran et al. (2000) Blood 96:1969-1978)

Expression of CCL18 in Gaucher spleen





Plasma CCL18 levels in controls and Gaucher patients



Sandwich ELISA: Median control plasma level 33 ng/ml (10-72 ng/ml)

Median Gaucher plasma level 948 ng/ml (237-2285 ng/ml)

CCL18 levels on average 30 fold elevated compared to control

No Overlap with control values

Protein mass spectrometry develops in various stages

- What is the primary structure and the identity of the protein?
- What is the higher order structure of the protein, which structural domains can be discerned
- What is the subunit composition of functional protein complexes
- What is the topology of a protein and protein complexes
- Which functional domains are present and what is the function of proteins and assemblies of proteins
- How is the function of a protein regulated by post translational modifications
- How is the function of a protein regulated by interaction with other proteins