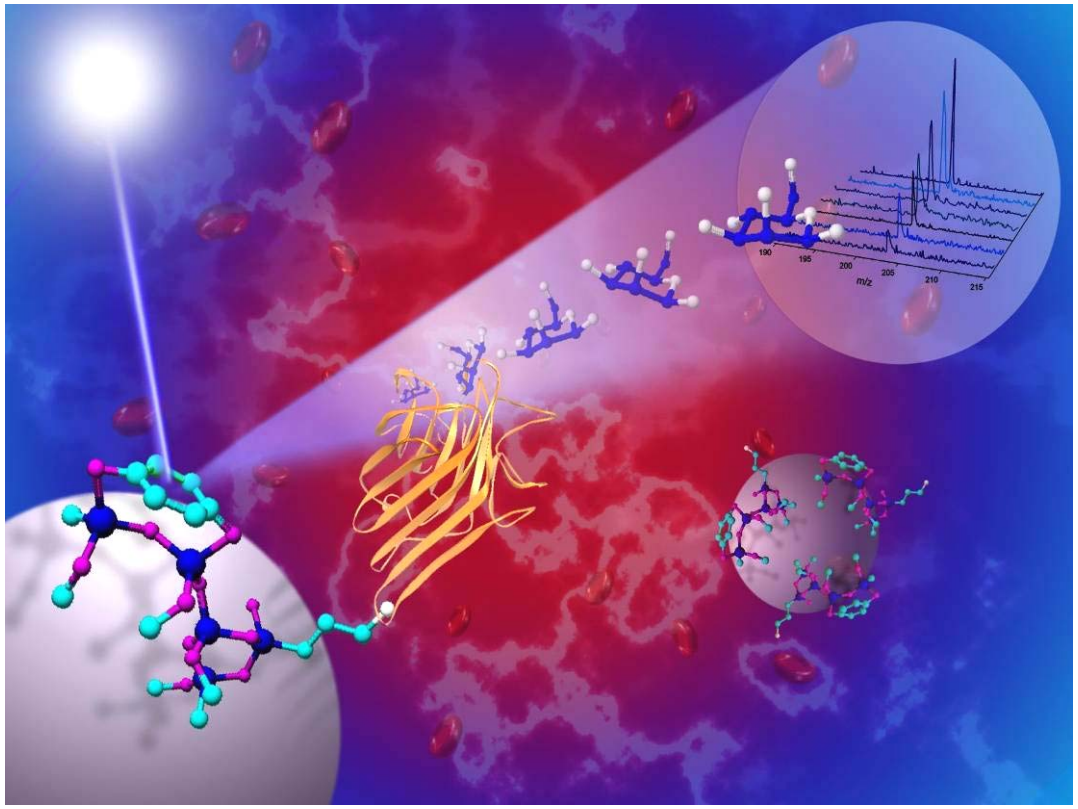
Mass Spectrometry-based Proteomics

- Introduction to Proteomics
- Introduction of Mass Spectrometry
- Protein Identification by Mass Spectrometry
- Quantitation Strategy

Yu-Ju Chen
Institute of Chemistry

Nanoprobe-based Affinity Mass Spectrometry for Protein Separation



Yu-Ju Chen

Institute of Chemistry

Interdisciplinary Nanotechnology in Biomedical Research

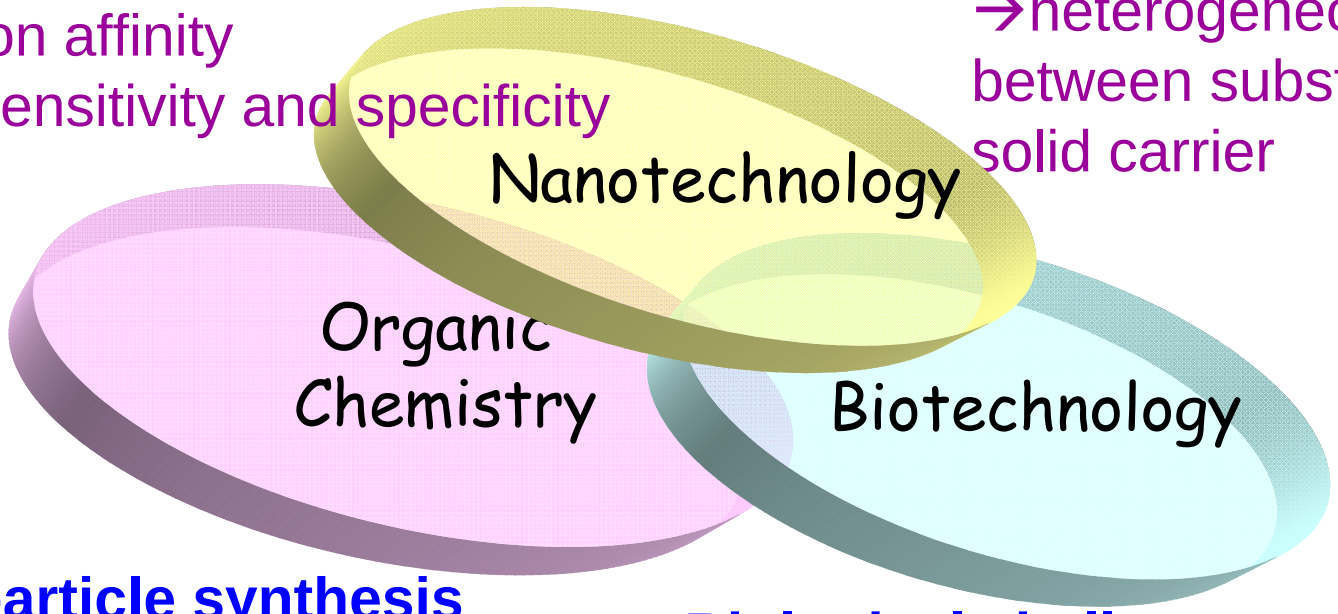
•High surface area to volume ratio

→High ligand density and interaction affinity

→High sensitivity and specificity

•Pseudo-homogeneity in water

→heterogeneous interface between substrate and solid carrier



•<u>Nanoparticle synthesis</u>

→uniform shape and composition

→narrow size distribution

•<u>Surface modification</u>

→high reaction efficiency

→high biocompatibility

<u>Biological challenges</u>

→low abundant protein enrichment

→site-directed protein modification

→low-affinity protein-ligand interaction

→low molecular-weight molecules identification

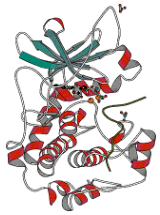
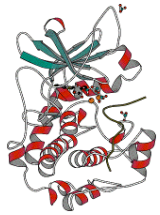
Example:
Immunoassay for Disease Diagnosis

In the post-genome era, the rapid evolution of proteomics research has opened new horizons because it promises to accelerate the discovery of new protein disease markers. The recognition that every disease induces a specific pattern of change in a proteomics microenvironment has important implications on the early detection and progression of diseases. Many clinical diagnostic assays, such as **ELISA**, correlate the concentration of specific protein markers with the onset of disease.

Disease Biomarker for Cancer Diagnostics

Serum Tumor Markers	Primary Clinical Applications	Other Related Cancer Type
Alpha-Fetoprotein (AFP)	Hepatocellular carcinoma (HCC) and germ-cell (nonseminoma) tumor monitoring and diagnosing	
CA 15-3	Breast cancer monitoring[1]	colorectal, liver, lung, ovarian, pancreatic cancer
CA 19-9	Colorectal and pancreatic cancer monitoring	breast, gastric, hepatobiliary, hepatocellular, and ovarian cancer
CA 125	Endometrial and ovarian cancer monitoring[2]	breast, cervical, colorectal, gastrointestinal, lung, pancreatic cancer
Prostate Specific Antigen (PSA)	Prostate cancer monitoring and diagnosing[3]	

Why Mass Spectrometry for Protein Detection and Identification?

PR  TE  MICS

蛋白質體學

Protein activity, modifications, localizations, amd interactions of proteins in complexes

Proteomics can be defined as *the qualitative and quantitative comparison of proteomes under different conditions to further unravel biological processes*

Estimated Number of Proteins per Genome

• Haemophilus	1742
• *E. coli*	4413
• Yeast	6600
• Caenorhabditis	18000
• Drosophila	13000
• Human	>1000000 (35000 genes)

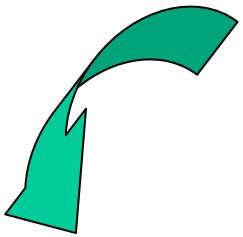
What Do We See?

Technology Platform V.S. ***Complex Proteome***

A Key Technical Challenge in Proteomics

A complex biological problem

Diverse solution
(10^5 – 10^6 for protein abundance)

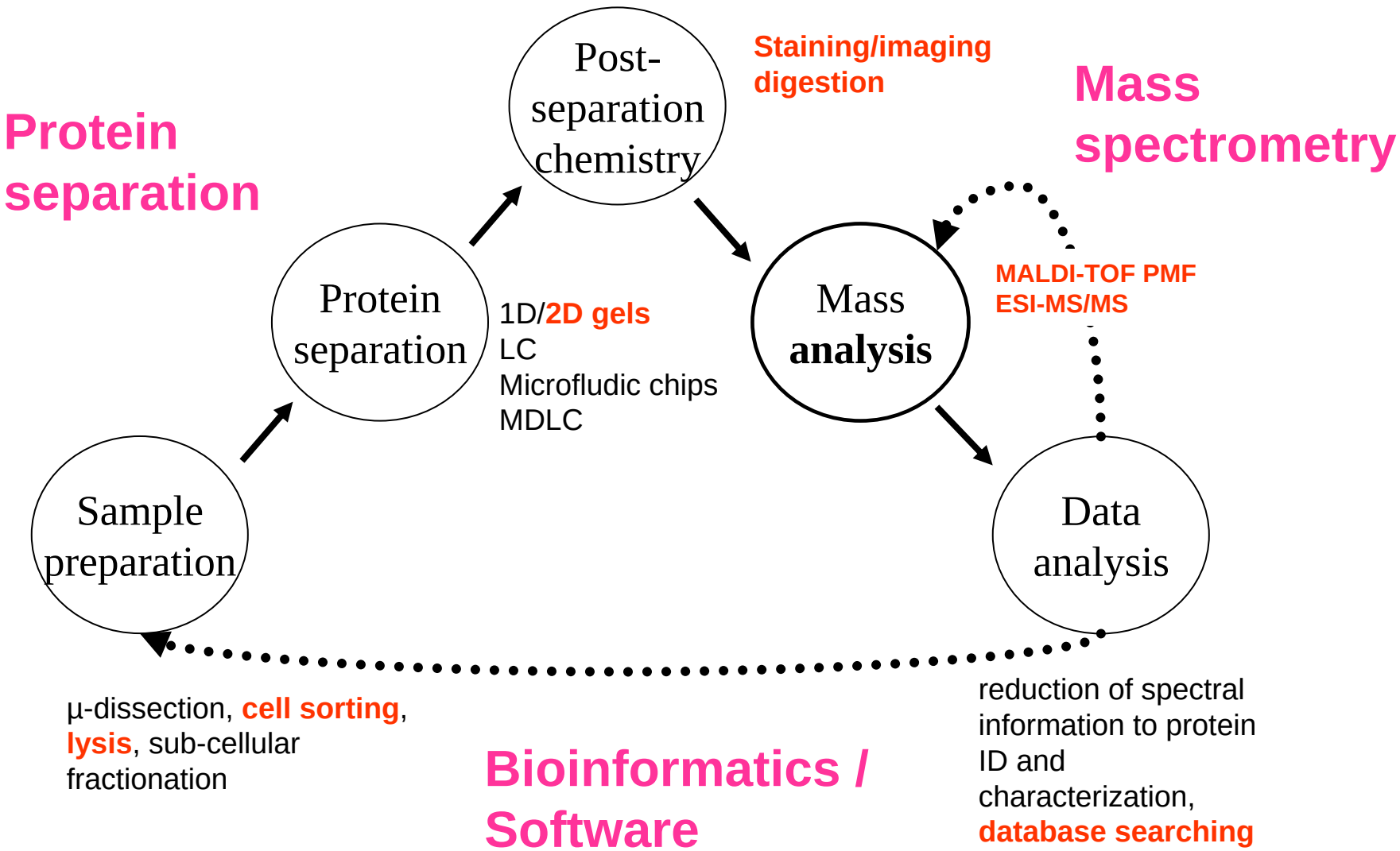


Mass-spectrometry strategy for more complex mixture ?

+

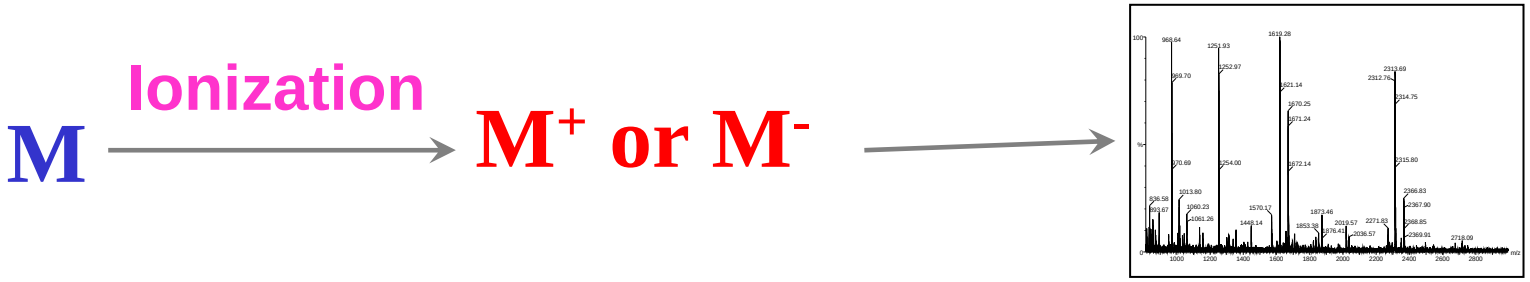
Sample prefractionation

Three Key Technologies in Proteomics



Principle of Mass Spectrometry

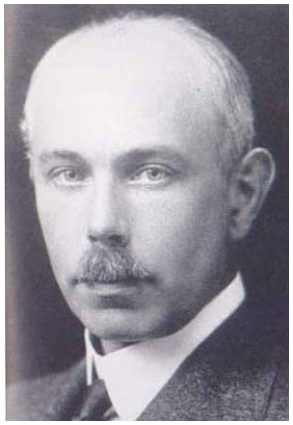
What is mass spectrometry?



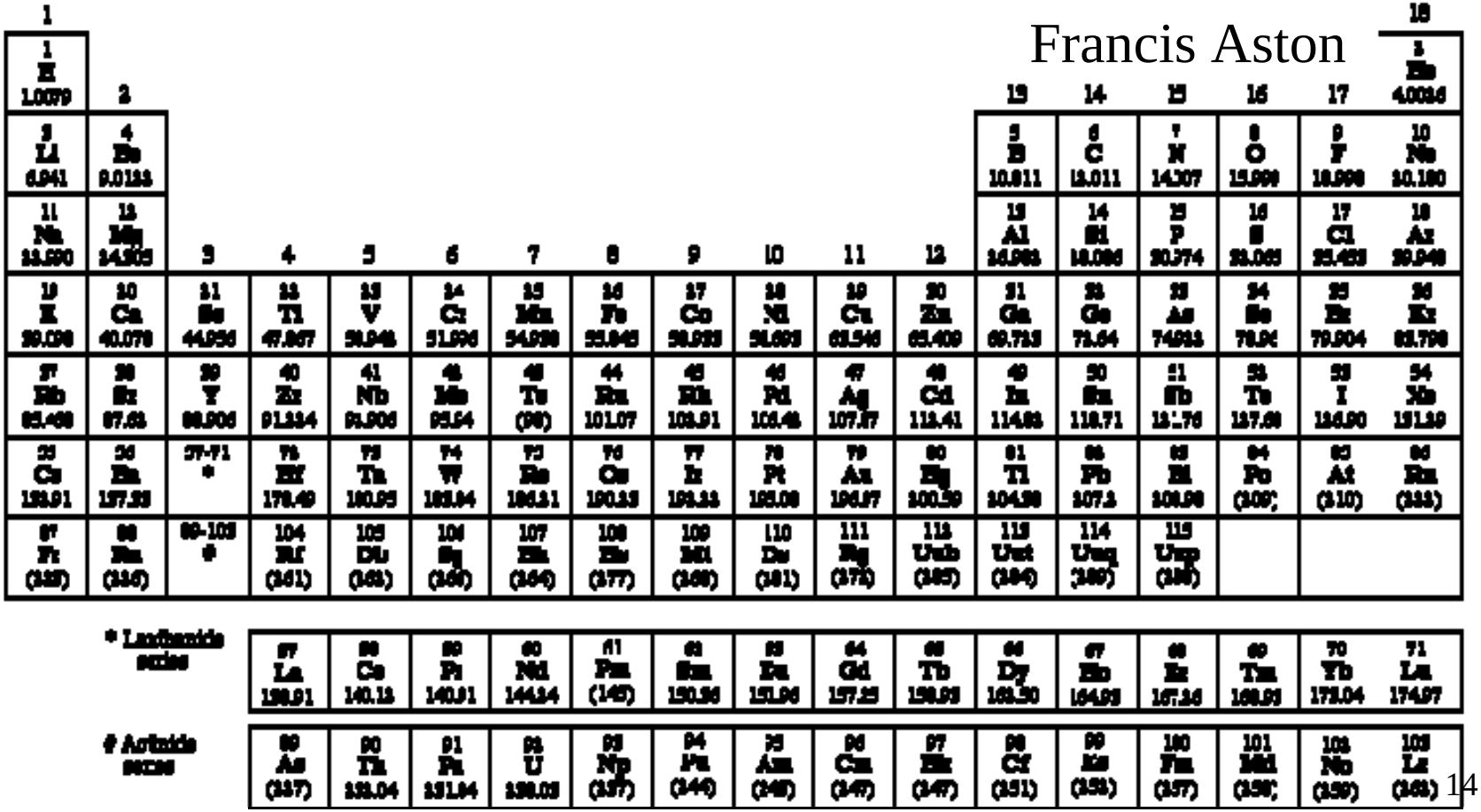
Mass-to-charge Ratio (***m/z***)

MS is an analytical tool that measure the molecular weight of molecules based on the motion of charged particle in an electrical or magnetic field.

- ***Francis Aston*** is awarded the Nobel Prize in chemistry for his discovery of isotopes of "inactive elements."



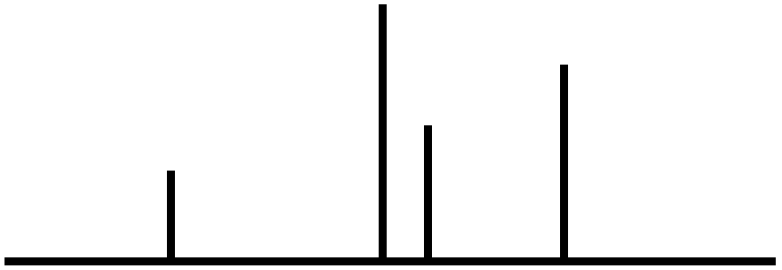
Francis Aston



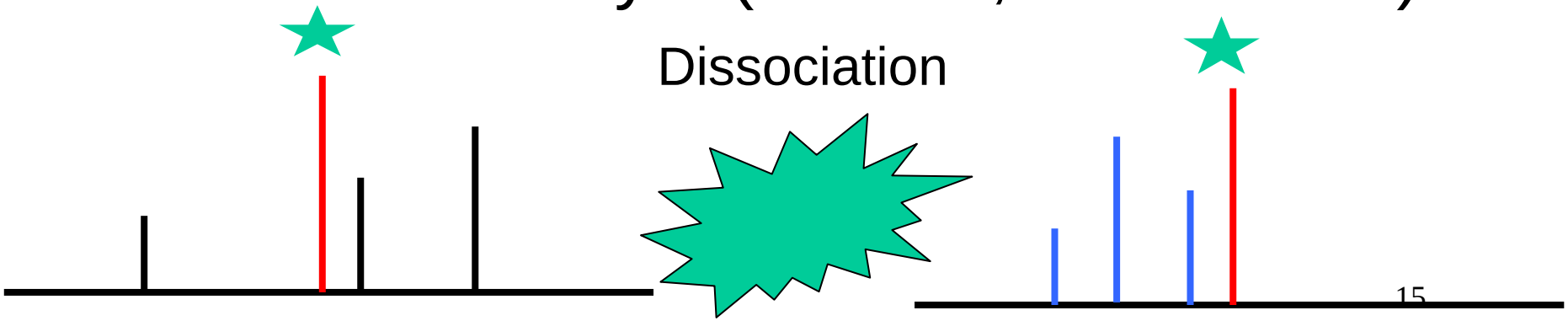
Two different ways of measurement

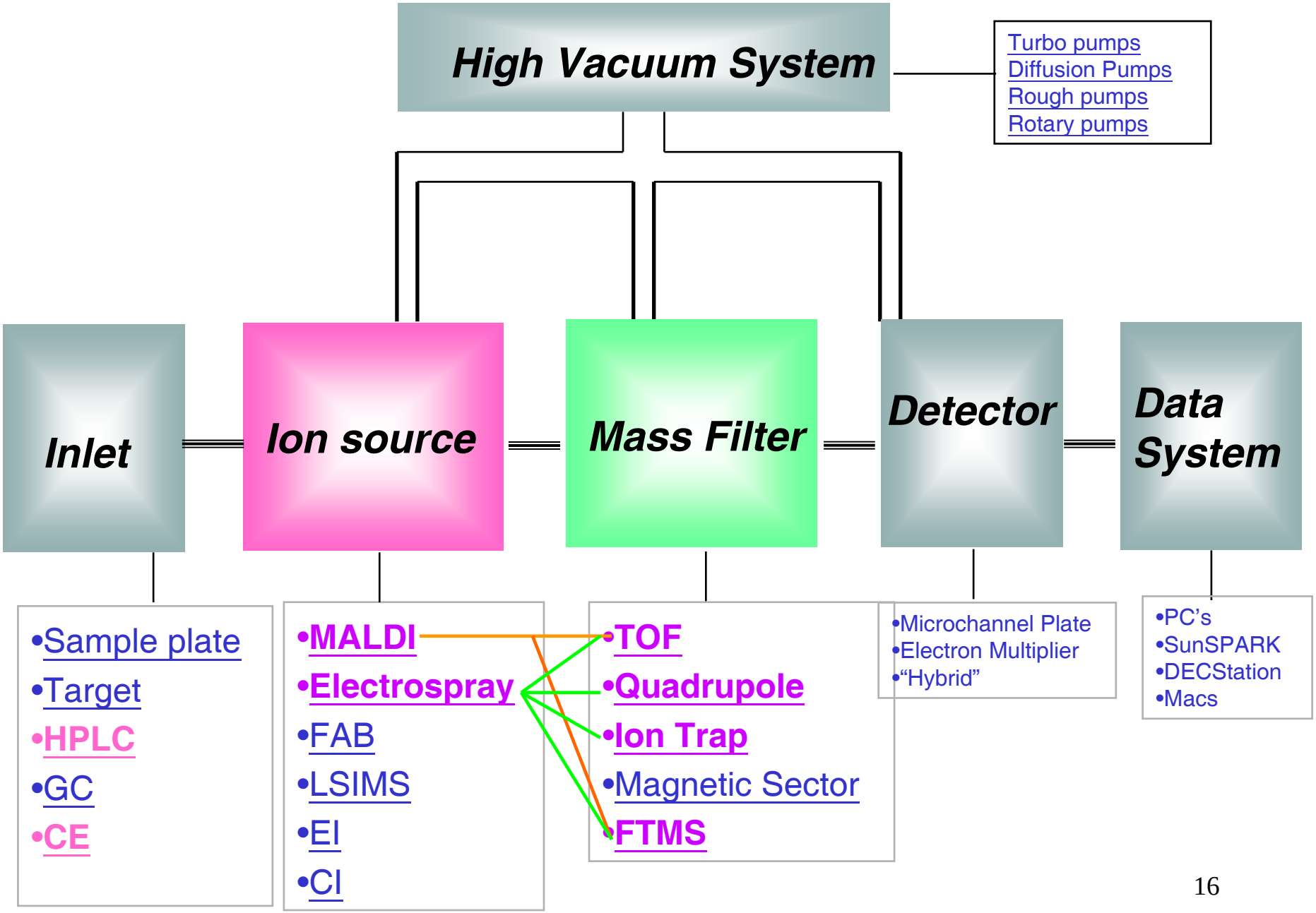
- Composition of analyte (MS)

  e.g. Peptide mass fingerprinting



- Structure of analyte (MS/MS, tandem MS)




High Vacuum System
Turbo pumps
Diffusion Pumps
Rough pumps
Rotary pumps
Inlet
Ion source
Mass Filter
Detector
Data System
•Sample plate
•Target
•HPLC
•GC
•CE
•MALDI
•Electrospray
•FAB
•LSIMS
•EI
•CI
•TOF
•Quadrupole
•Ion Trap
•Magnetic Sector
•FTMS
•Microchannel Plate
•Electron Multiplier
•"Hybrid"
•PC's
•SunSPARK
•DECStation
•Macs

Ionization Methods

- Electron Impact (EI)
- Fast Atom Bombardment (FAB)
- Electrospray Ionization (ESI)
- Matrix-Assisted Laser Desorption Ionization (MALDI)

Advances in Modern Mass Spectrometry

Limitations of traditional MS on biological applications

- High molecular weight >50,000
- Amount of Sample $< 10^{-12} - 10^{-15}$ mole

Intact Molecule

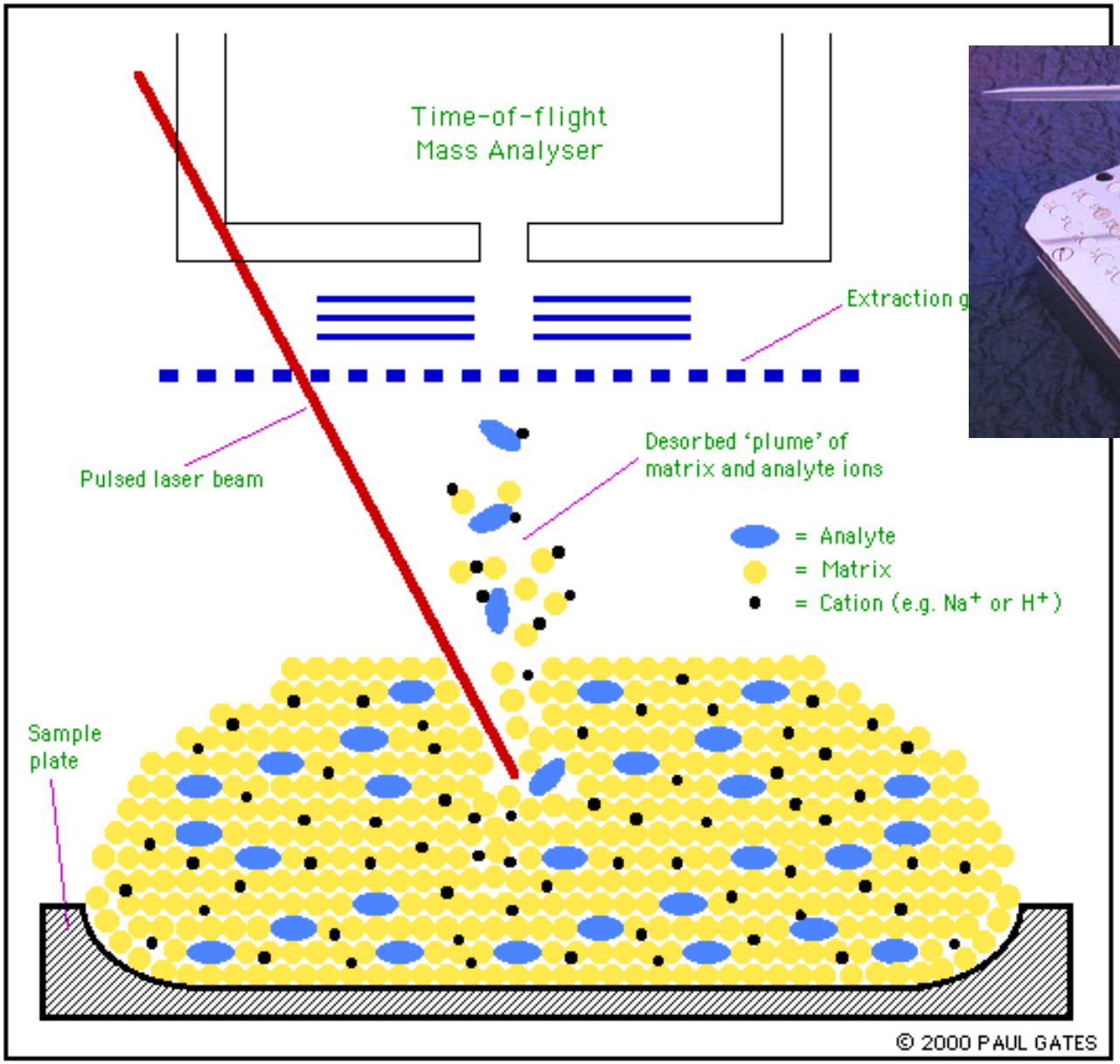
Non-covalent Complex

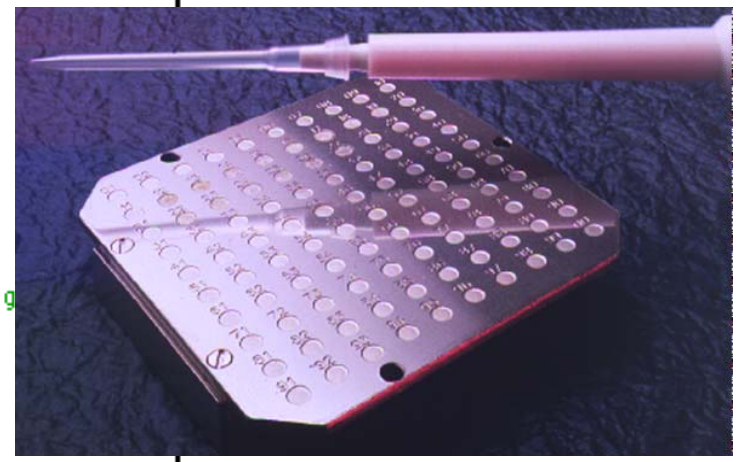
Electro**S**pray **I**onization MS

Matrix **A**ssisted **L**aser **D**esorption **I**onization MS

2002 Nobel Prize in Chemistry

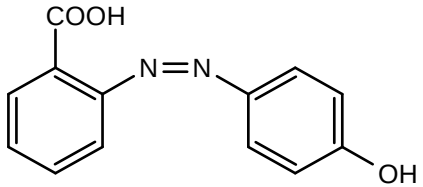
*M*atrix-*A*ssisted *L*aser *D*esorption *I*onization (*MALDI*)



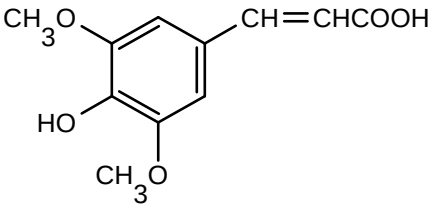


Matrix selection

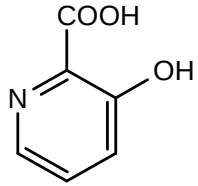
α-Cyano-4-hydroxy-cinnamic acid (CHCA)	Peptides<10kDa
Sinapinic Acid	Proteins >10kDa
2,5-Dihydroxybenzoic acid (DHB)	Neutral Carbohydrates, Synthetic Polymers
"Super DHB"	Proteins, Glycosylated proteins
3-Hydroxypicolinic acid	Oligonucleotides
HABA	Proteins, Oligosaccharides



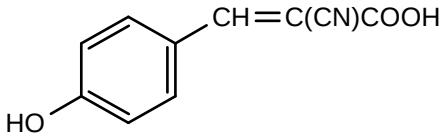
2-(4-hydroxyphenylazo)-benzoic acid (HABA)



Sinapinic acid (3,5-Dimethoxy-4-hydroxy cinnamic acid)

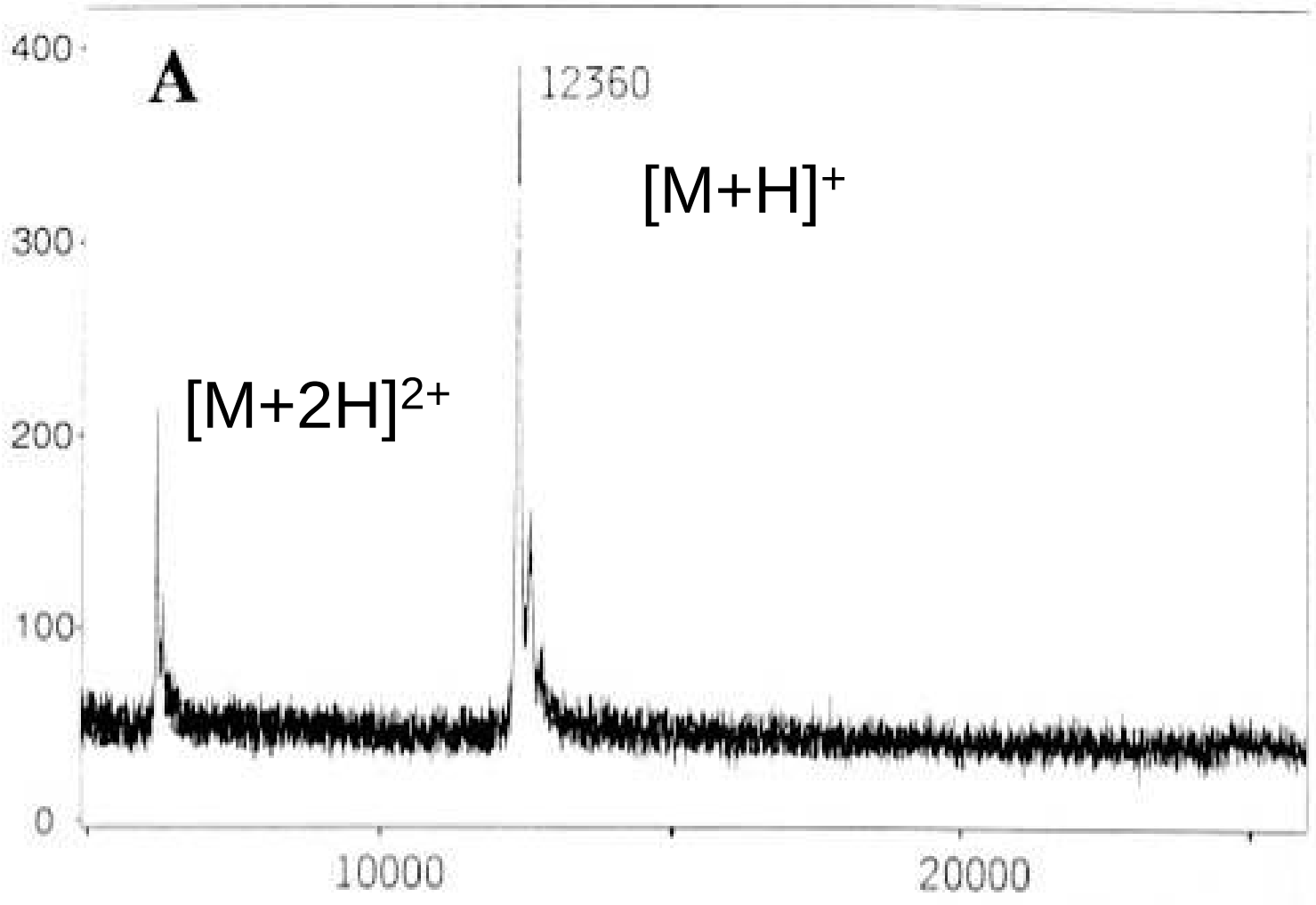


3-hydroxypicolinic acid (3-HPA)

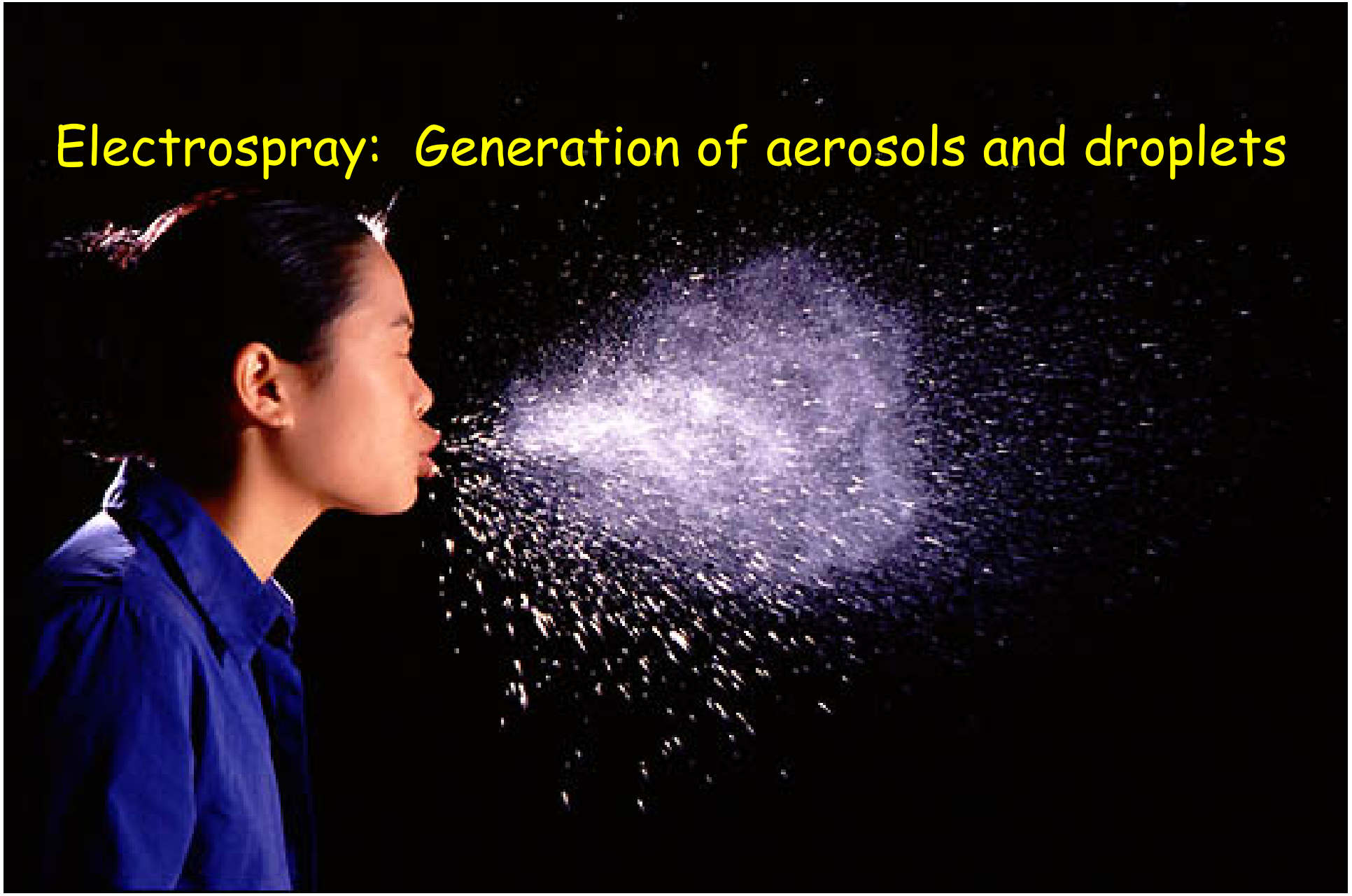


α-cyano-4-hydroxycinnamic acid

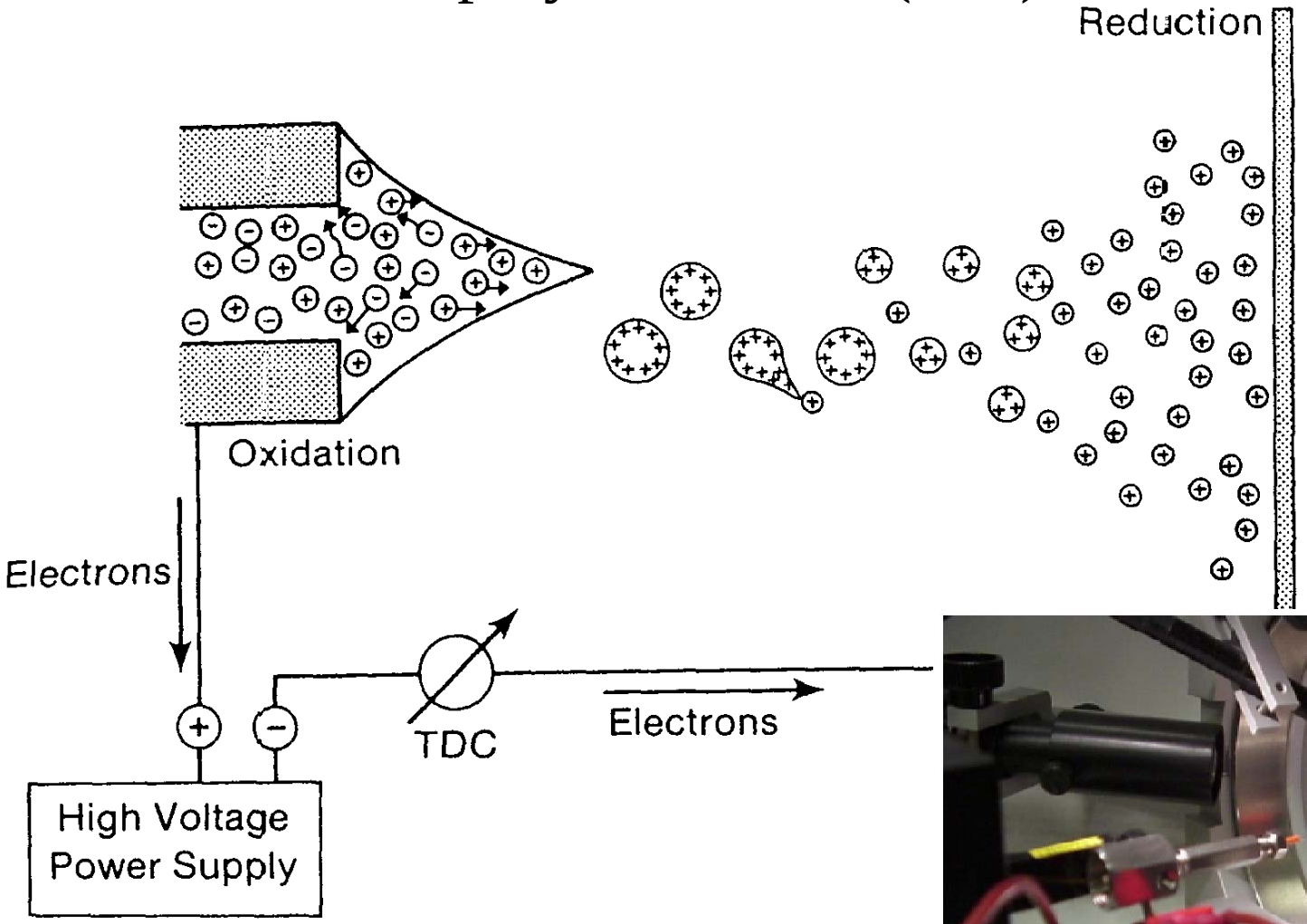
A typical mass spectra

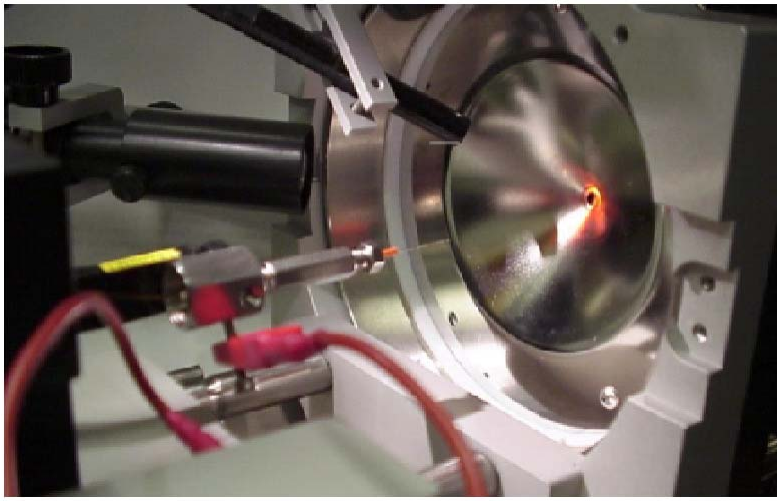


Electrospray: Generation of aerosols and droplets

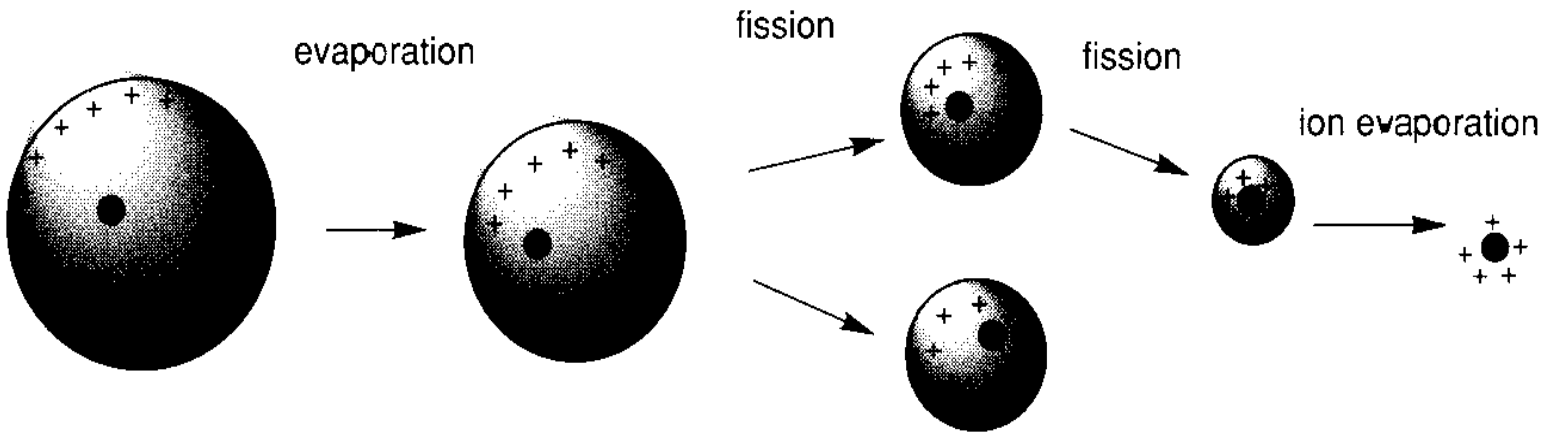


*E*lectro*S*pray *I*onization (**ESI**)



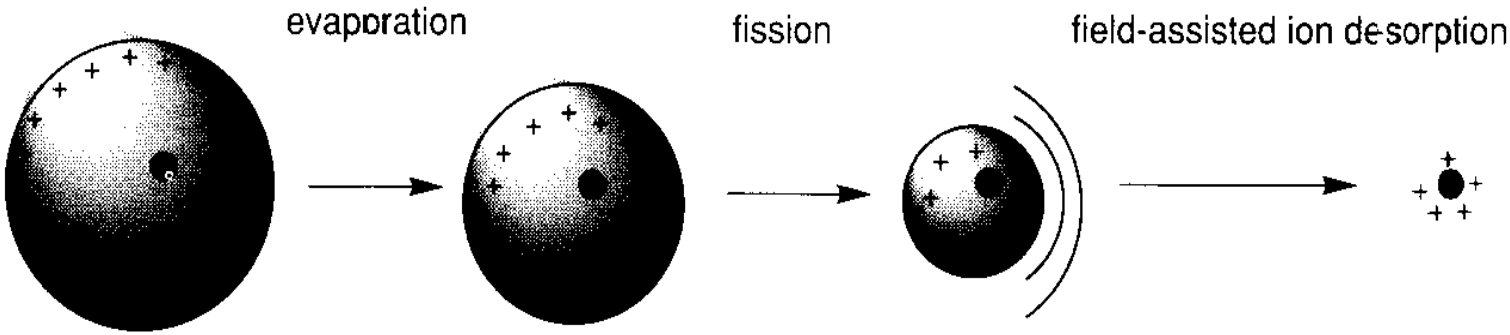


Charge Residue Model



(a)

Coulomb repulsion : Charge repulsion > surface tension

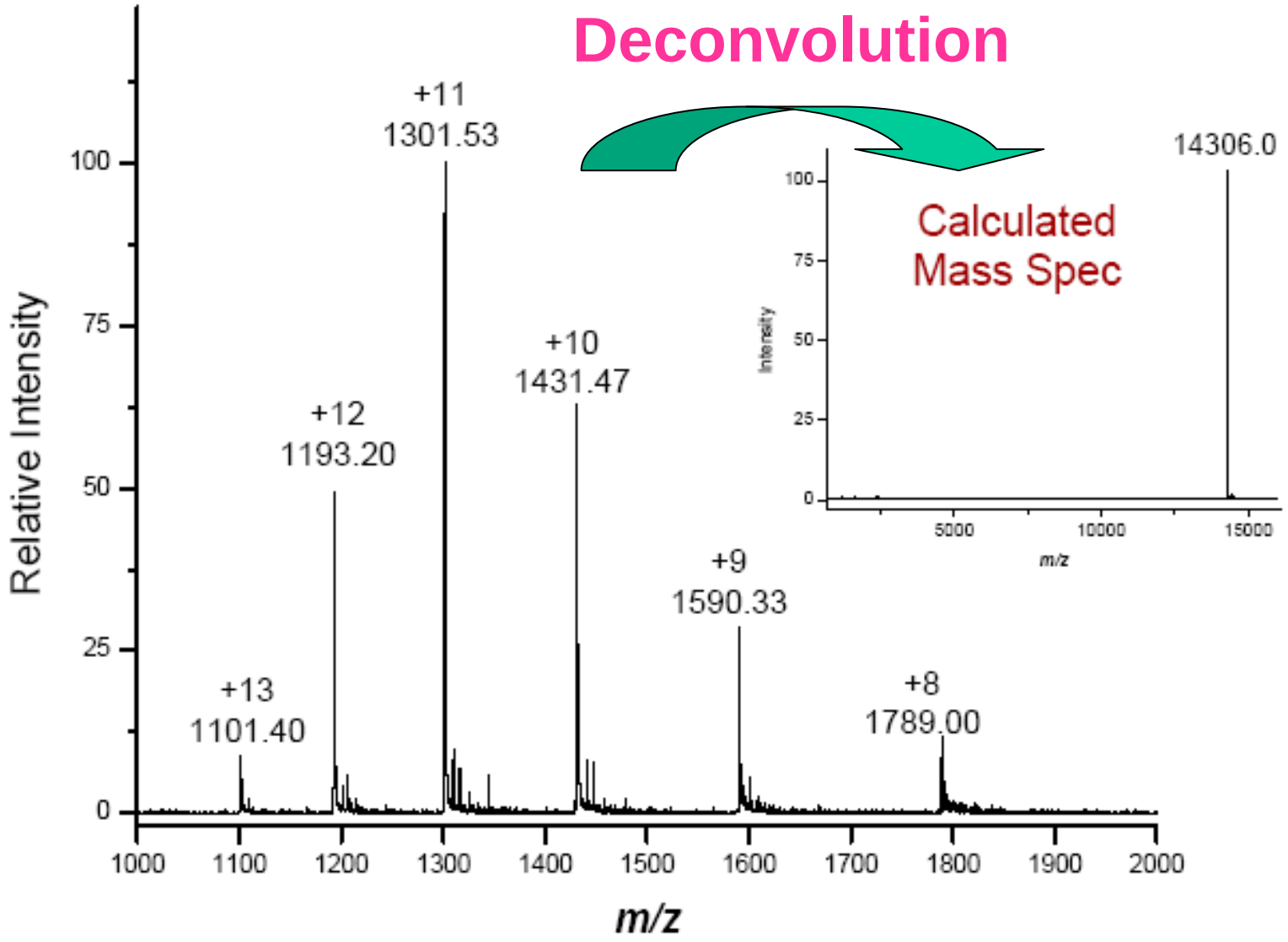


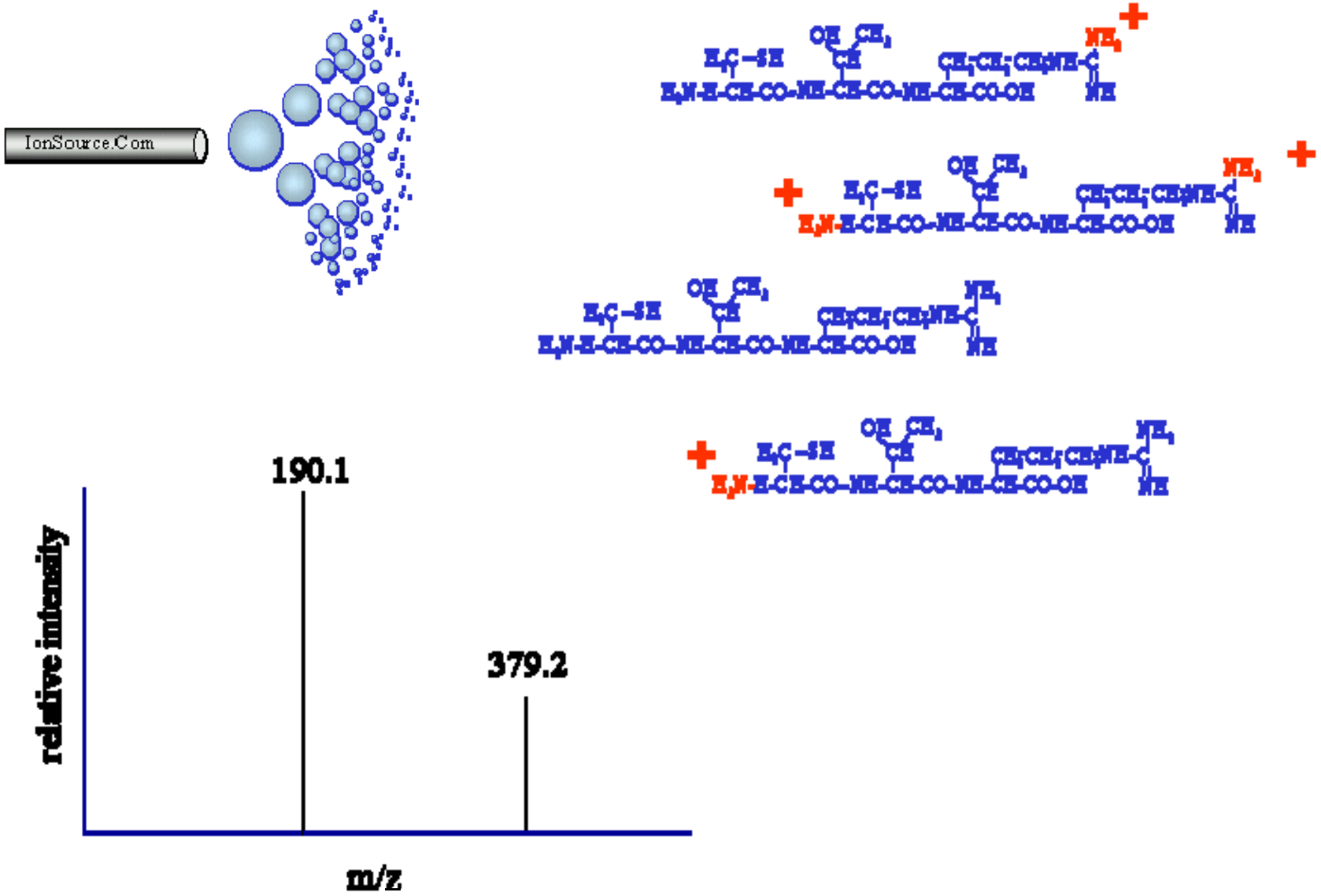
(b)

Ion Desorption Model

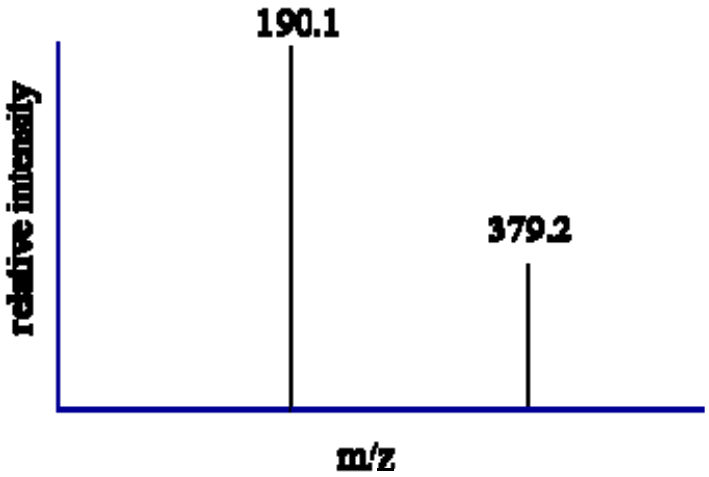
Ion desorption from the droplet surface

Multiple-charged ESI spectra




IonSource.Com
190.1
379.2
relative intensity
m/z

Deconvolution



Step 1: assume 190.1 = mass/charge

379.2 = mass/charge

Step 2: assume the two peaks are related

190.1 = [m+(z+1)]/(z+1)

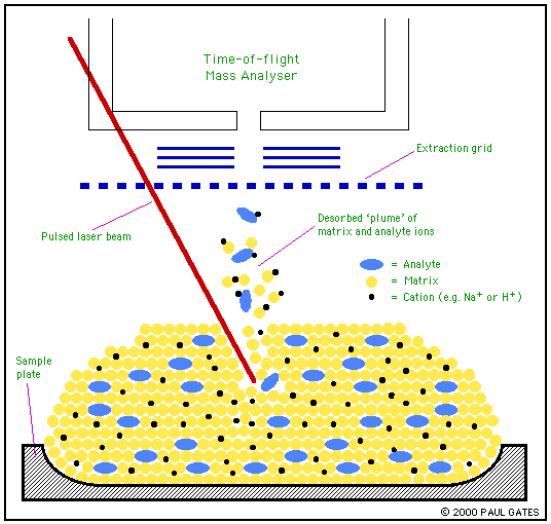
379.2=m+z/z

Step 3: solve m and z

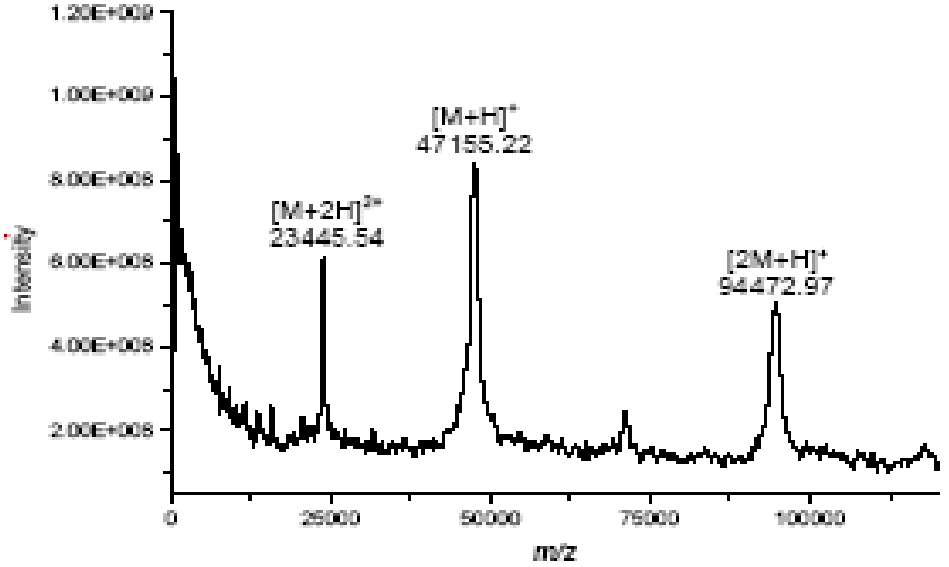
m=378.2, z=1

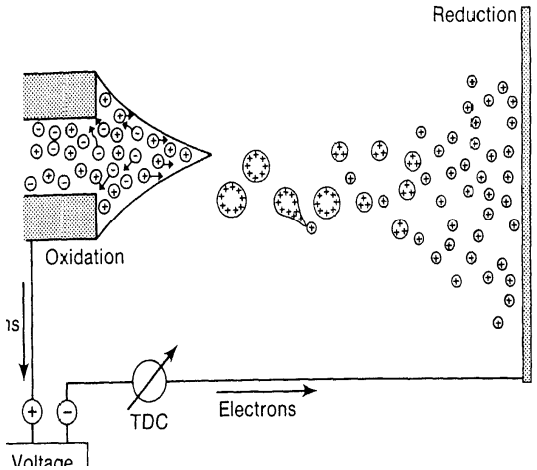
Charge state	Calculation	Unprotonated mass
+1	(379.2-1)*1	378.2
+2	(190.1-1)*2	378.2
	average	378.2

Molecular Weight of Proteins

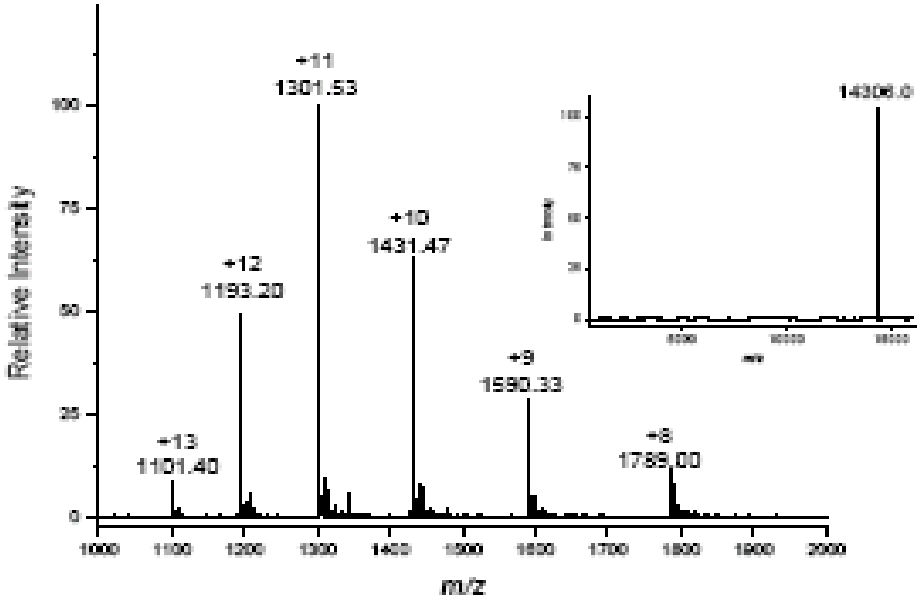


MALDI



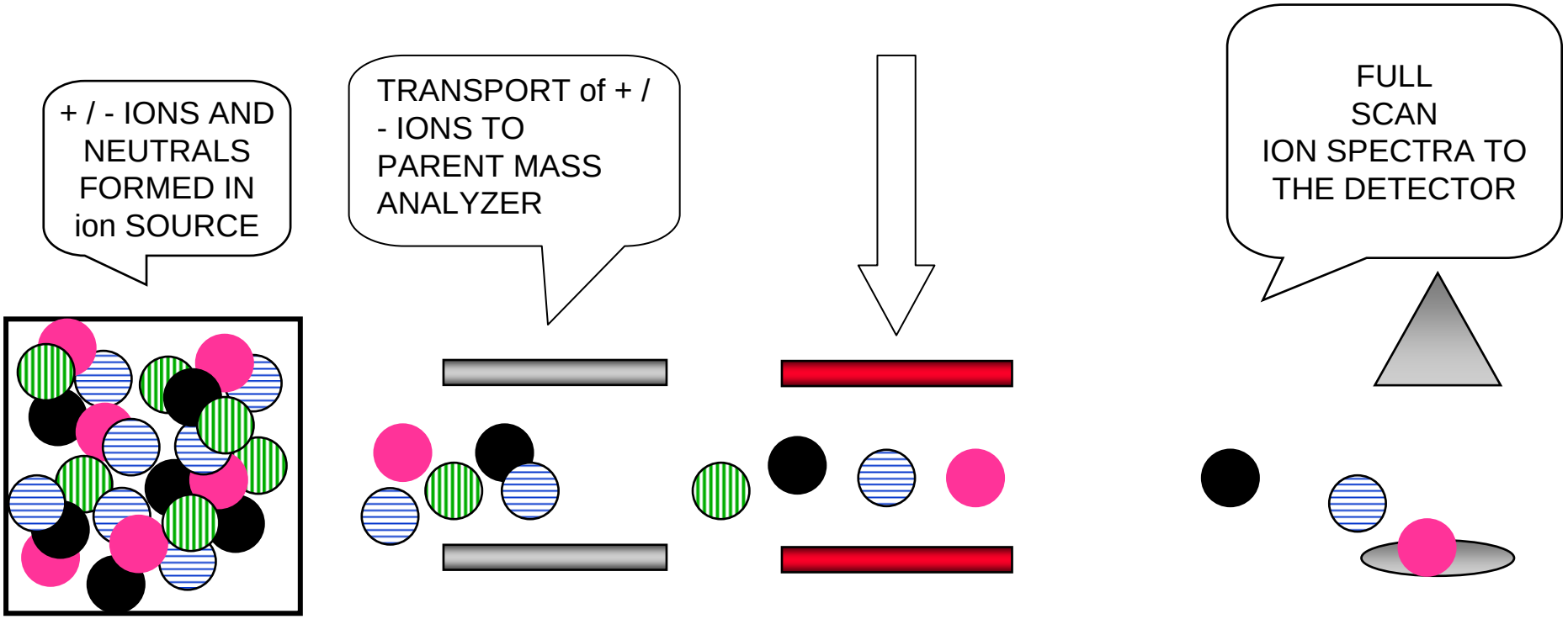


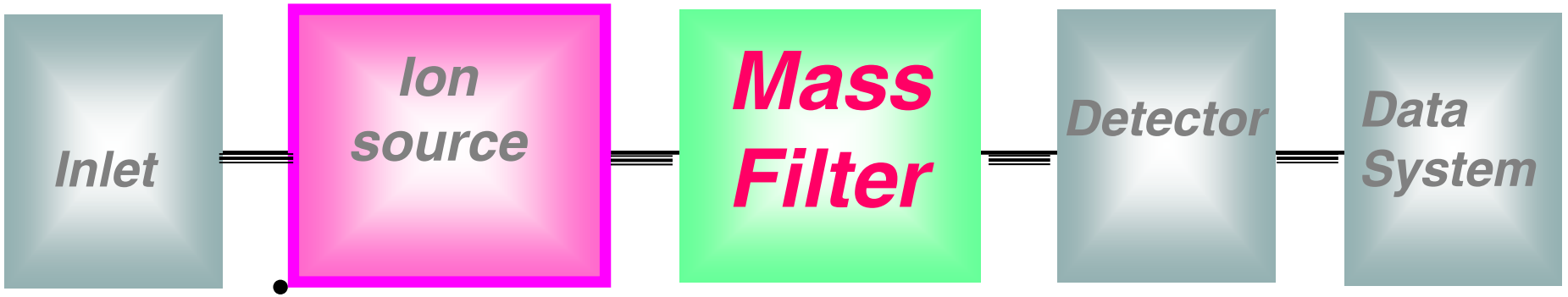
ESI



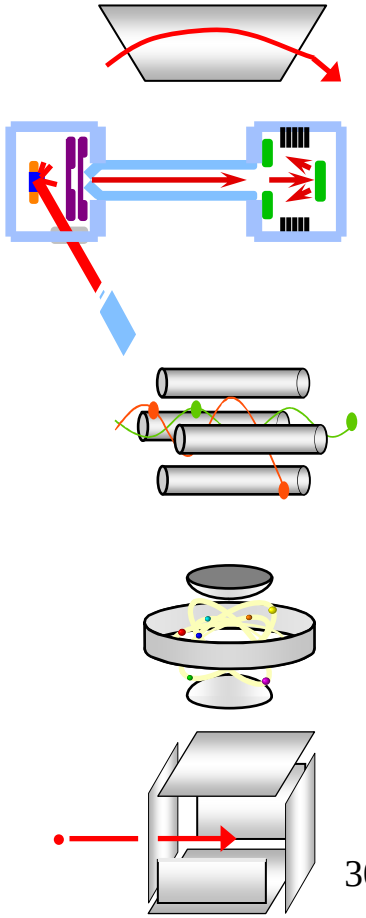
Mass Analysis

Ions are separated according to their mass-to-charge (*m/z*)





- Sector Instruments
- Time-of-flight Analyzer
- Quadrupole Mass Filter
- Ion-Trap Instrument
- FT-ICR



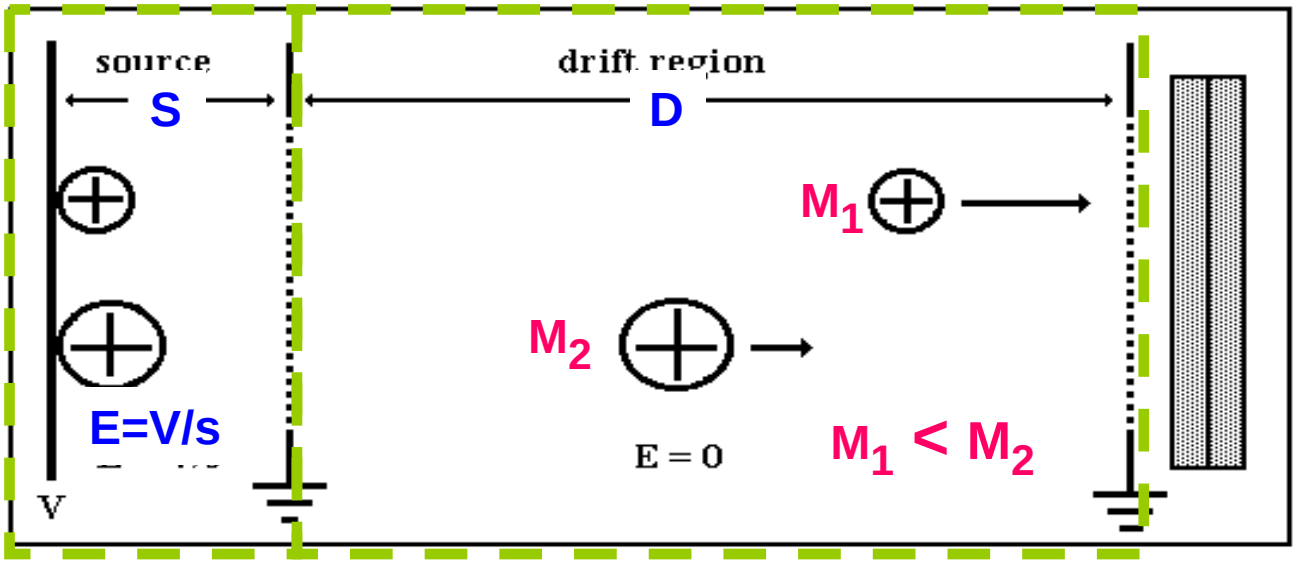
Time-of-flight Analyzer

- Ions are generated in the source zone of the instrument.
- A potential (V) is applied across the source to extract and accelerate the ions from the source into the field-free 'drift' zone of the instrument.
- Ions travel with velocity **v = d/t** ; d:tube distance, t:time
- All ions produced will leave the source at the same time with the same kinetic energy (**KE = ½ *m*v^2 = *z*V**), due to their having been accelerated through the same potential difference (ideally).
- The time-of-flight of the ions produced will only be dependent on the mass and the charge of the produced ion.

$$m/z = [2\ t^2\ V] / d^2$$

- The larger the ion, the slower its velocity and thus the longer it takes to traverse the field-free drift zone.

Mass-to-Charge (m/z) is a Function of Flight Time



In a electric field [Potential Energy] ⇨ **In drift region [Kinetic Energy]**

$$E = \frac{V}{s}$$ *where V = accelerating voltage*

$$\frac{1}{2}mv^2 = qV = zeEs \longrightarrow v = \left(\frac{2zeEs}{m}\right)^{\frac{1}{2}}$$

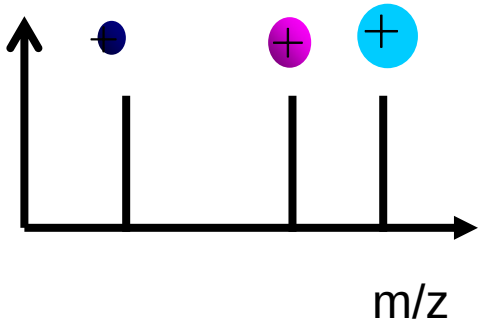
→ **Flight time** $$t = \left(\frac{m}{2zeEs}\right)^{\frac{1}{2}} D$$

Relation between m/z and t

$$\left(\frac{m}{z}\right) = \left(\frac{2eEs}{D^2}\right)t^2$$

Mass spectrum

$$t = \left(\frac{m}{2zeEs} \right)^{\frac{1}{2}} D$$



t= 100 μ second for mass 50 (small moleculae)

1000 μ second for mass 5,000 (peptide, polymer…)

3000 μ second for mass 50,000 (protein, DNA….)

Protein Identification and Quantification

Identification (ID)

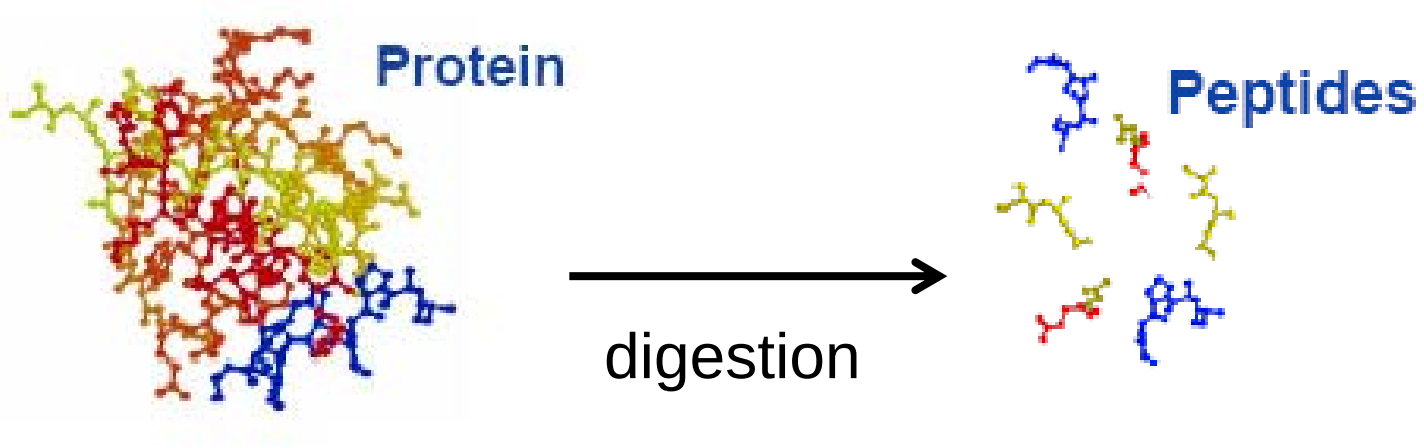
✓Identify the sequence of an unknown protein

✓Determine the site of post-translational modification

Quantification

✓Analyze the relative concentration of a protein under different conditions

Protein Analysis by Mass Spectrometry



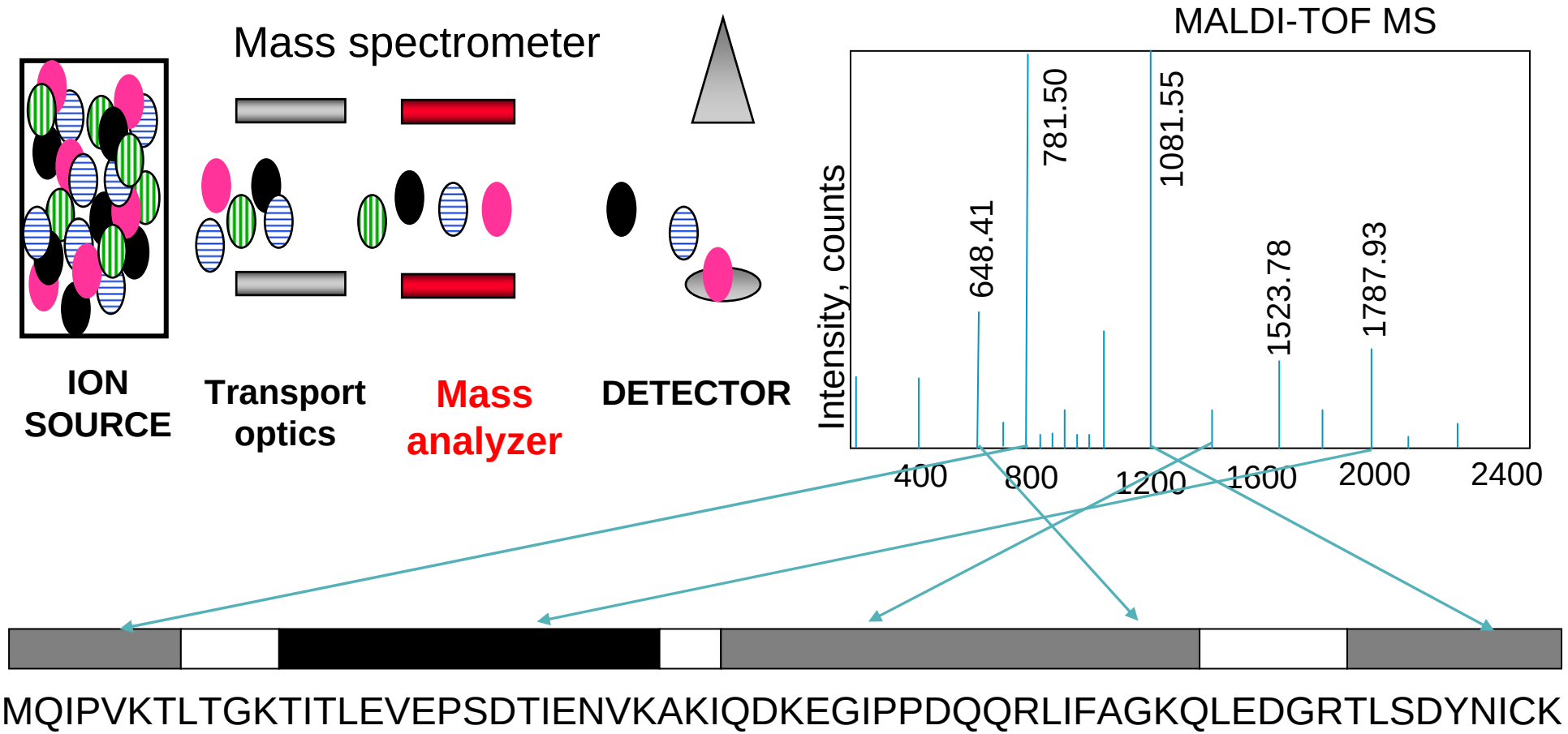
Molecular weight

Identification
Sequence information

- ✓ MALDI TOF MS
- ✓ ESI MS

- ✓Peptide mass fingerprinting (MALDI TOF MS)
- ✓Peptide Sequencing (Tandem MS)

Two Ways of Measurement—
1. Peptide Mass Fingerprint (by M.W. measurement)



Peptide Mass Fingerprint

Left to right:

William J. Henzel; ***Protein Chemistry;*** Genentech

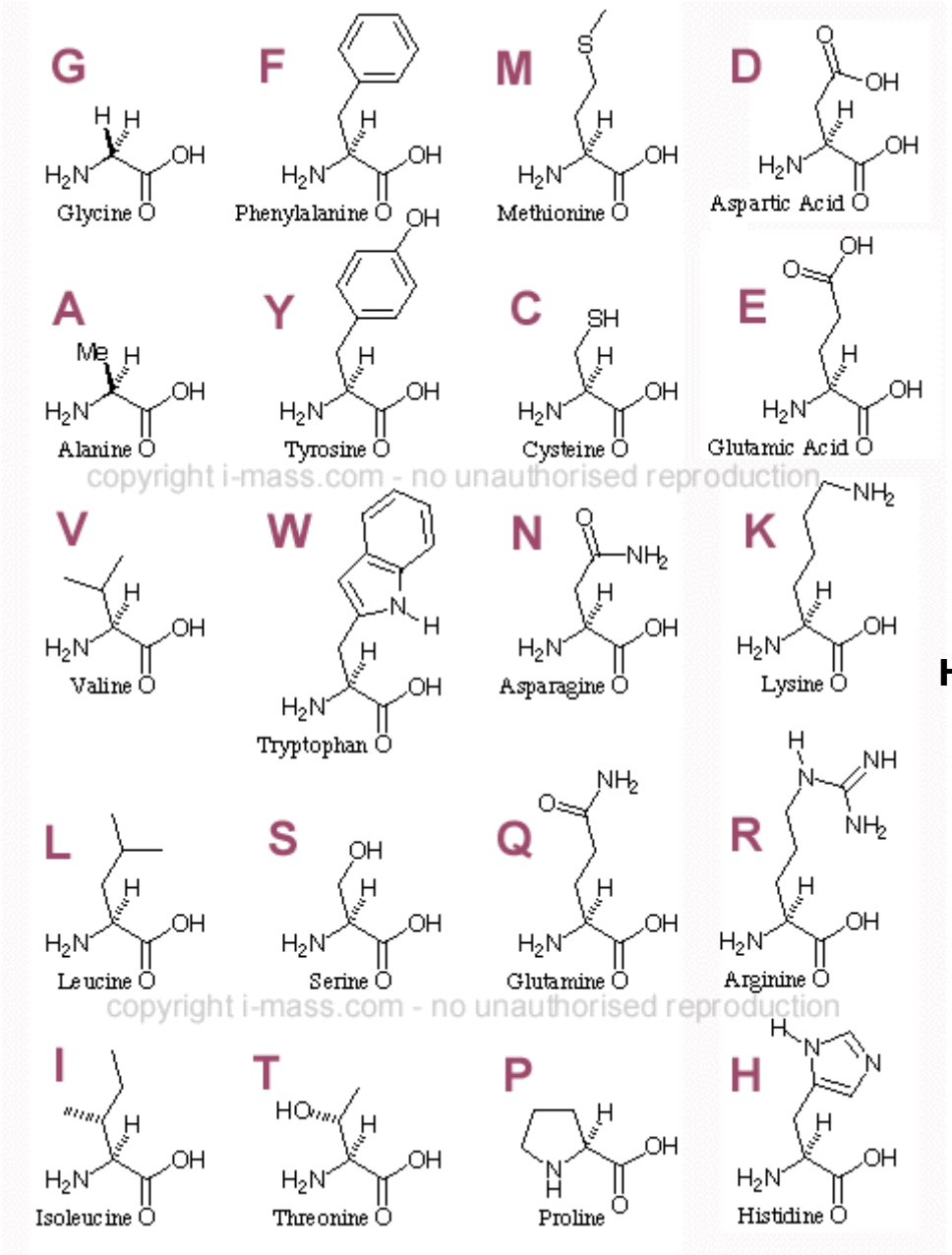
John T. Stults; ***Analytical Chemistry***; Genentech

Colin Watanabe; ***Software Engineer;*** Genentech

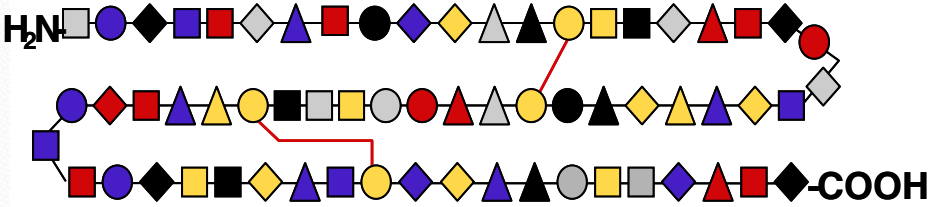
- This methodology was the first to allow protein identification without the need for time-consuming Edman sequencing or immunoaffinity probes..

- Landmark Study(1993)

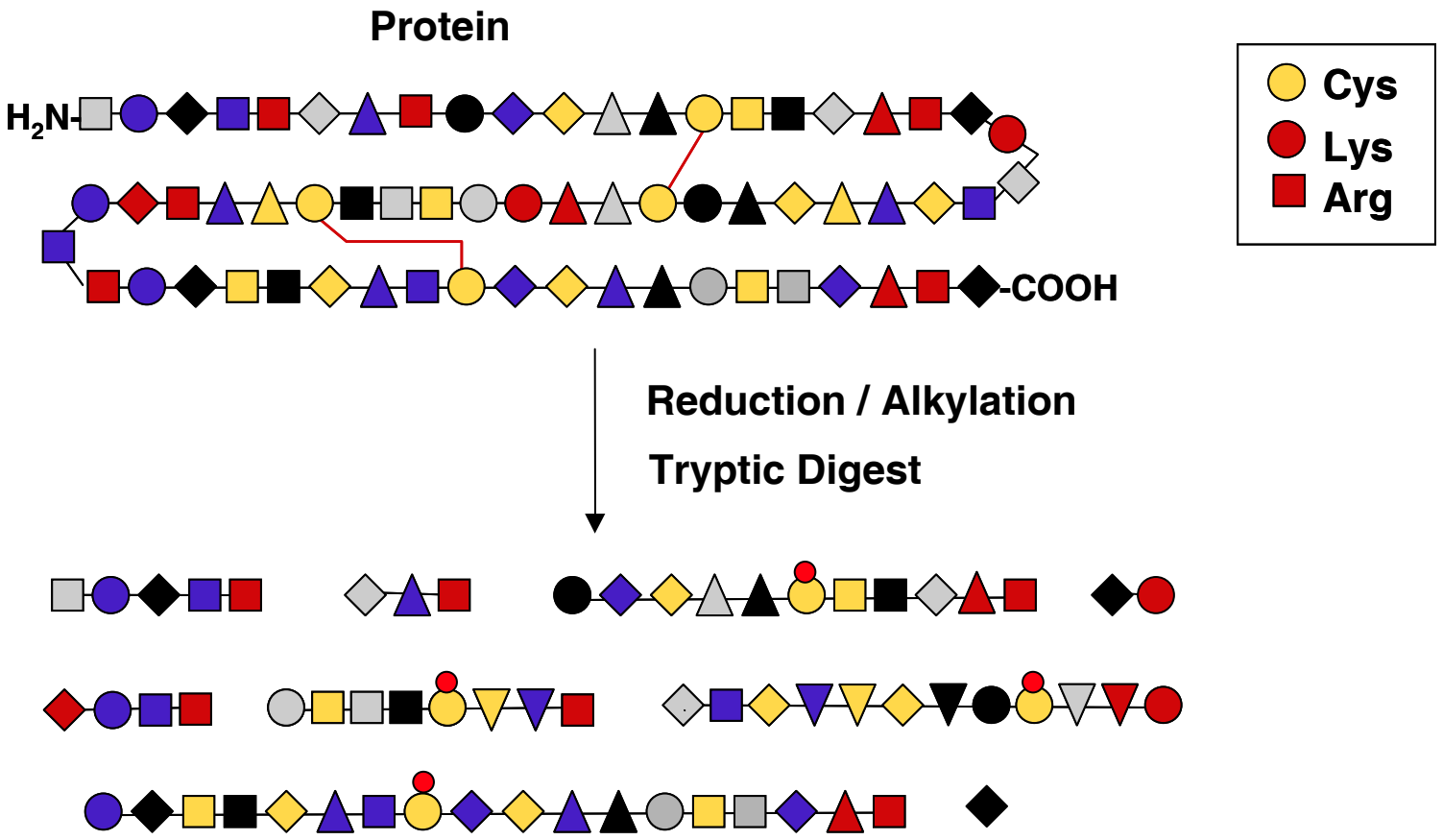
MS approaches alone could be used to analyze proteins from 2DE



Peptide Mass Fingerprinting



Step 1: Enzyme digestion or chemical fragmentation

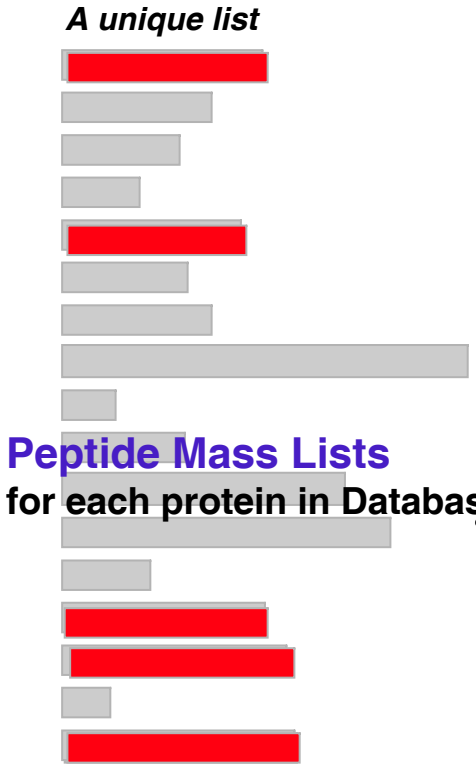


Every protein generate a set of unique peptides (every peptide has different mass)

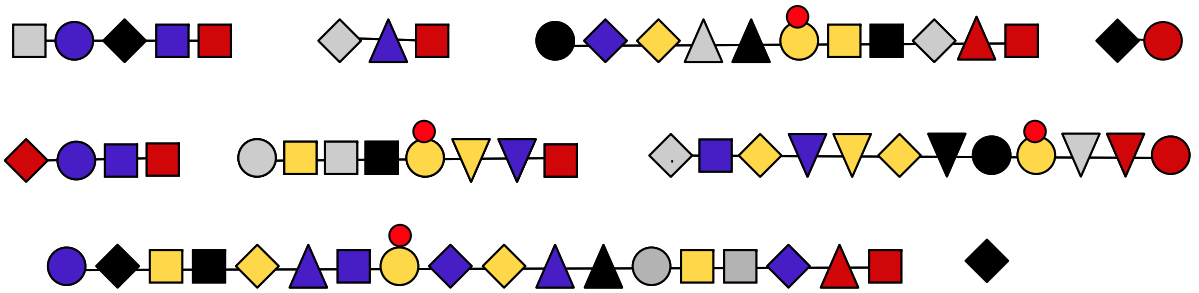
Residue Mass of Amino Acids

Table9.2	Symbols and residue masses of the protein amino acids		
Name	**Symbol**	**Residue mass**	**Side-chain**
Alanine	A,Ala	71.079	CH_3-
Arginine	R,Arg	156.188	$HN=C(NH_2)-N-(CH_2)_3$
Asparagine	N,Asn	114.104	$H_2N-CO-CH_2$-
Aspartic acid	D,Asp	115.089	$HOOC-CH_2$-
Cysteine	C,Cys	103.145	$HS-CH_2$-
Glutamine	Q,Gln	128.131	$H_2N-CO-(CH_2)_2$-
Glutamic acid	E,Glu	129.116	$HOOC-(CH_2)_2$-
Glycine	G,Gly	57.052	H-
Histidine	H,His	137.141	Imidazole-CH_2-
Isoleucine	I,Ile	113.16	$CH_3-CH_2-CH(CH_3)$-
Leucine	L,Leu	113.16	$(CH_3)_2-CH-CH_2$-
Lysine	K,Lys	128.17	$H_2N-(CH_2)_4$-
Methionine	M,Met	131.199	$CH_3-S-(CH_2)_2$-
Metsulphoxide	Met.SO	147.199	$CH_3-S(O)-(CD_2)_2$-
Phenylalanine	F,Phe	147.177	Phenyl-CH_2-
Proline	P,Pro	97.117	Phrrolidone-CH-
Serine	S,Ser	87.078	$HO-CH_2$-
Threonine	T,Thr	101.105	$CH_3-CH(OH)$-
Tryptophan	W,Trp	186.213	ndole-NH-CH=C-CH
Tyrosine	Y,Tyr	163.176	4-OH-Phenyl-CH_2-
Valine	V,Val	99.133	$CH_3-CH(CH_2)$-

Step 2: Mass spectrometry analysis



Peptide Mass Lists
for each protein in Database

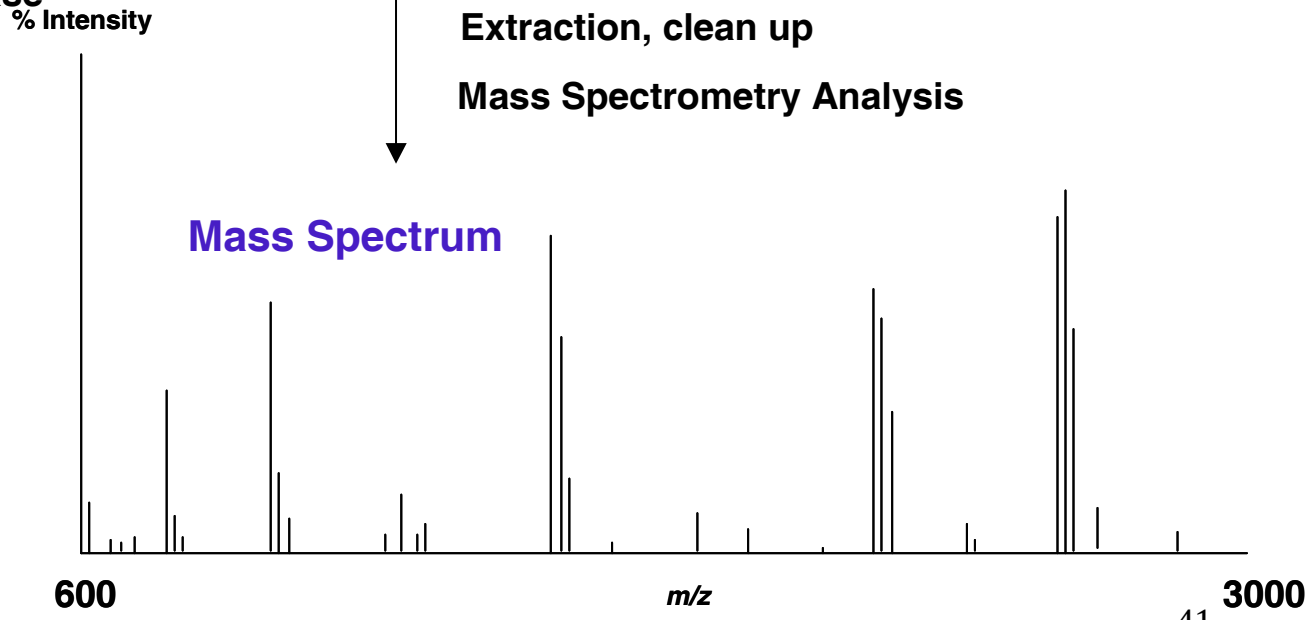


Peptide Mixtures

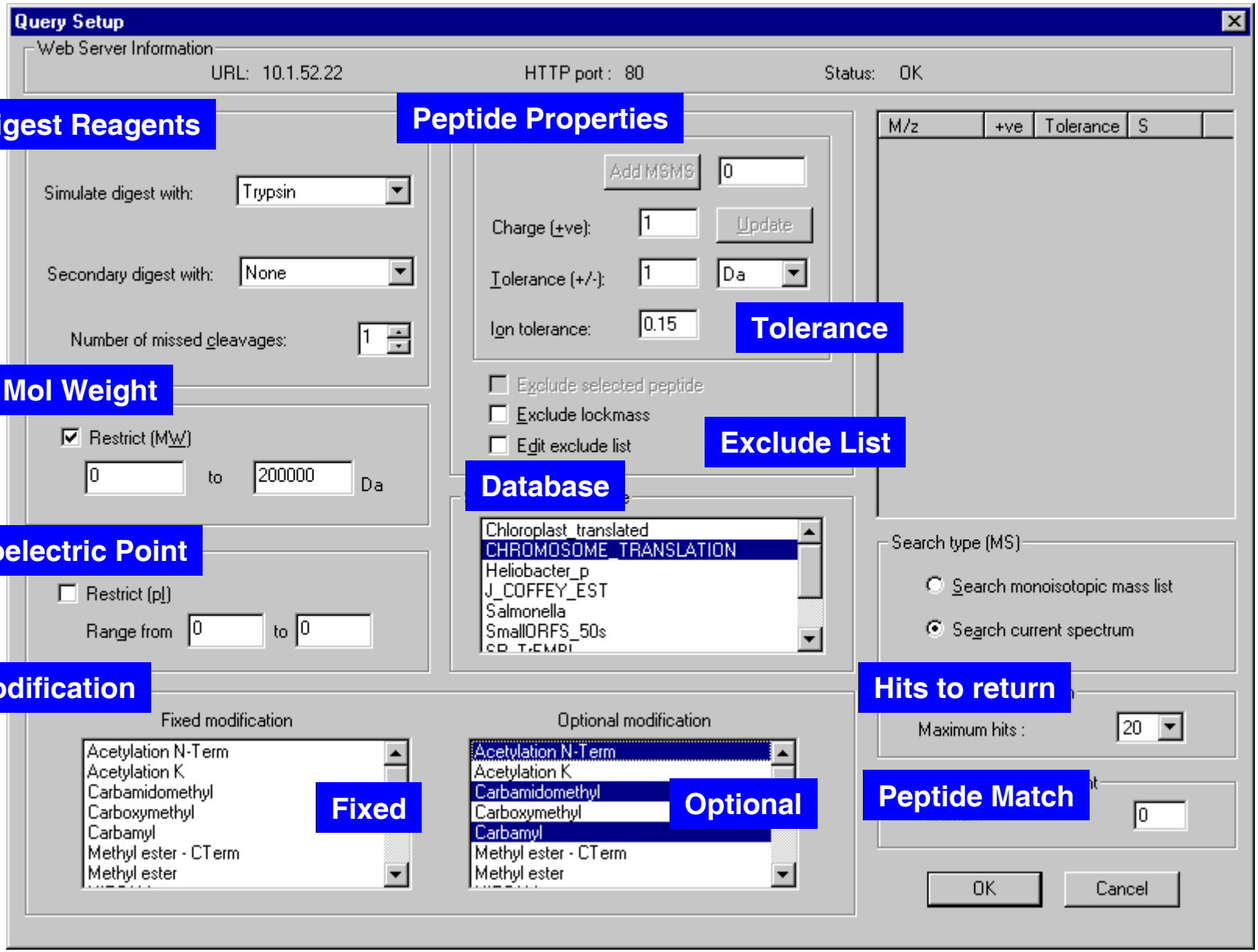
Extraction, clean up

Mass Spectrometry Analysis

Mass Spectrum



Step 3: Database match

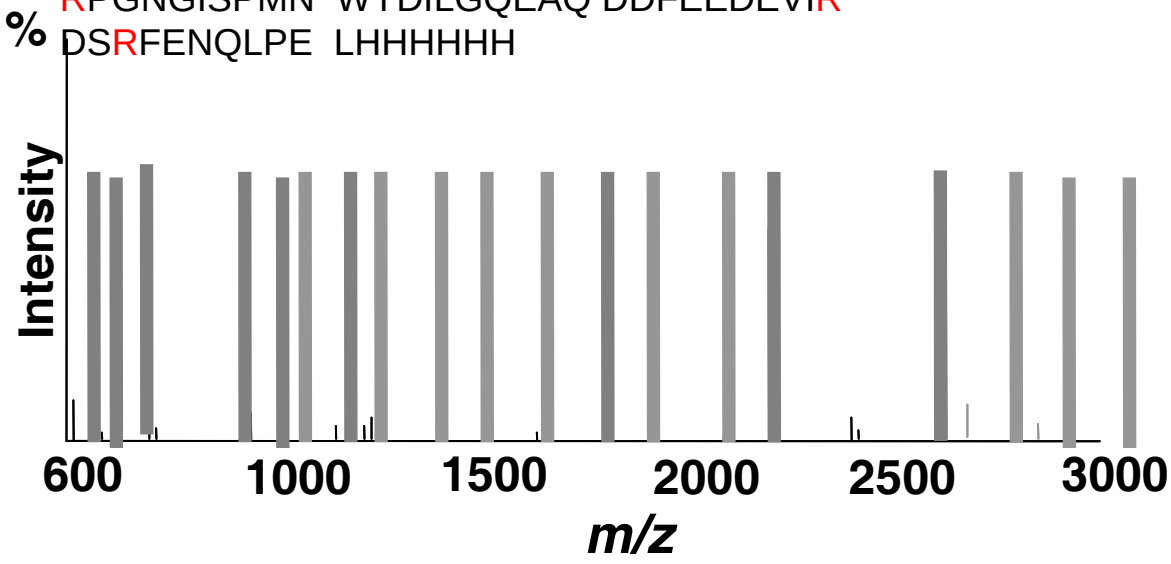


Step 3: In-Silica digestion

Protein Sequence

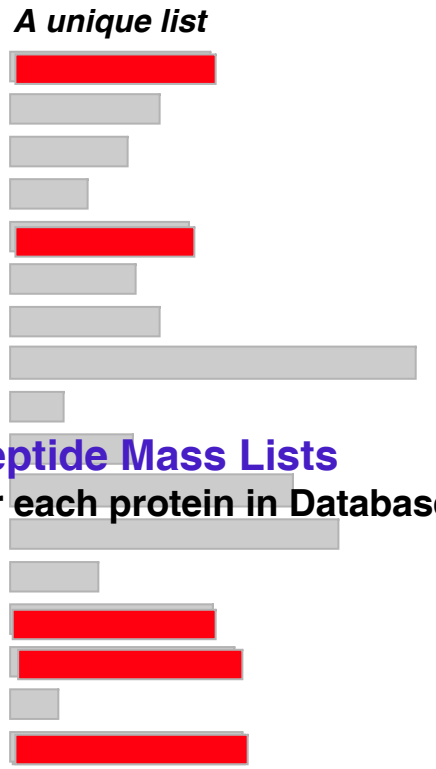
MVYIIAEIGC NHNGDINLAK KMVDVAVSCG
VDAVKFQTFK AEKLISKFAP KAEYQKATTG
TADSQLEMTK RLELSFEEYL EMRDYAISKG
VETFSTPFDE ESLEFLISTD MPIYKIPSGE
ITNLPYLEKI GKQQKKVILS TGMAVMEEIH
QAVNILRQNG TTDISILHCT TEYPTPYPSL
NLNVIHTLKD EFKDLTIGYS DHSIGSEVPI
AAAAMGAEVI EKHFTLDTNM
EVPDHKASAT
PDILAALVKG FALLNQALGR FEKIPDPVEE
KNKIVARKSV VALKPIKKGD IYSIENITVK
RPGNGISPMN WYDILGQEAQ DDFEEDEVIR
DSRFENQLPE LHHHHHH

Digestion

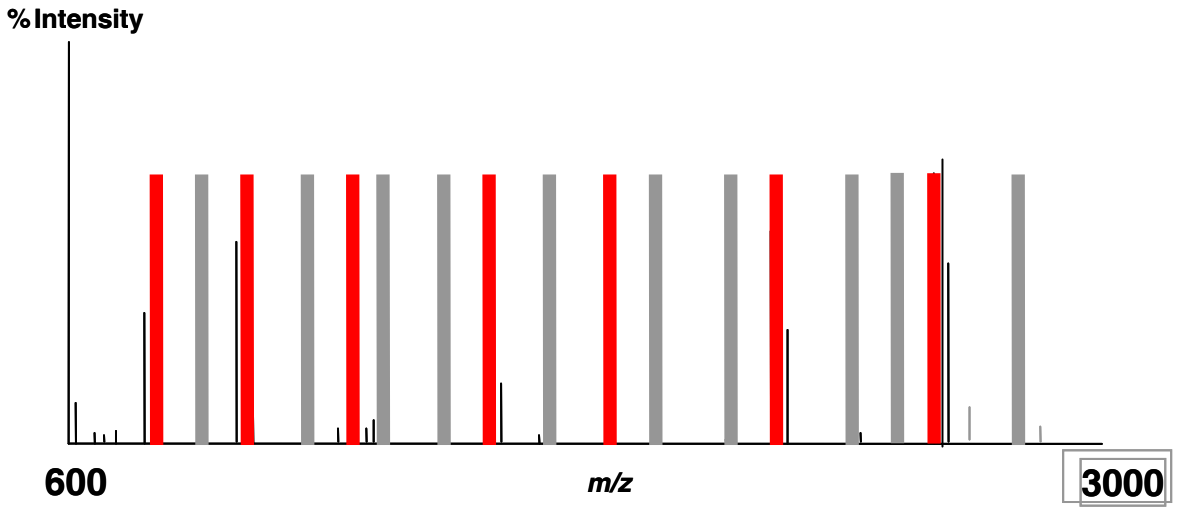


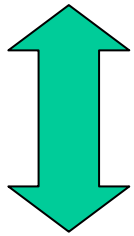
Mass	Peptide Sequence
3583.8	QNGTTDISILHCTTEYPTPY PSLNLNVIHTLK
3493.6	RPGNGISPMNWYDILGQEAQ DDFEEDEVIR
2995.4	GVETFSTPFDEESLEFLIST DMPIYK
2944.5	DLTIGYSDHSIGSEVPIAAA AMGAEVIEK
2324.2	VILSTGMAVMEEIHQAVNIL R
2188.1	MVYIIAEIGCNHNGDINLAK
1811.8	FENQLPELHHHHHH
1683.8	HFTLDTNMEVPDHK
1573.8	IPSGEITNLPYLEK
1558.7	LELSFEEYLEMR
1453.7	ATTGTADSQLEMTK
1392.7	MVDVAVSCGVDAVK
1351.7	GDIYSIENITVK
1269.7	ASATPDILAALVK
1159.7	GFALLNQALGR
954.63	SVVALKPIK
926.48	IPDPVEEK
696.36	DYAISK
670.36	FQTFK
638.31	AEYQK
538.25	DEFK

Step 4: MS match to Database



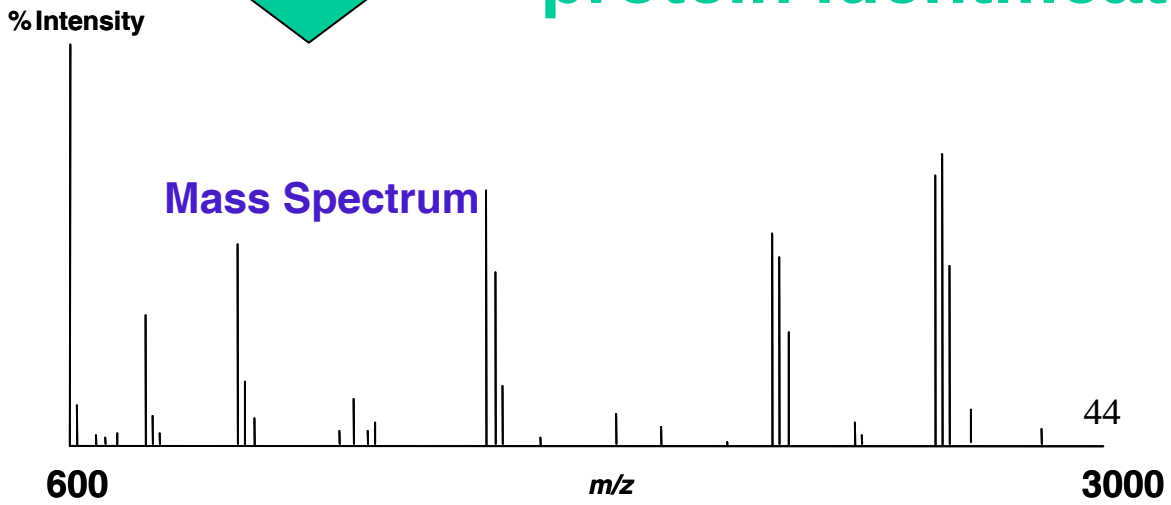
Peptide Mass Lists
for each protein in Database



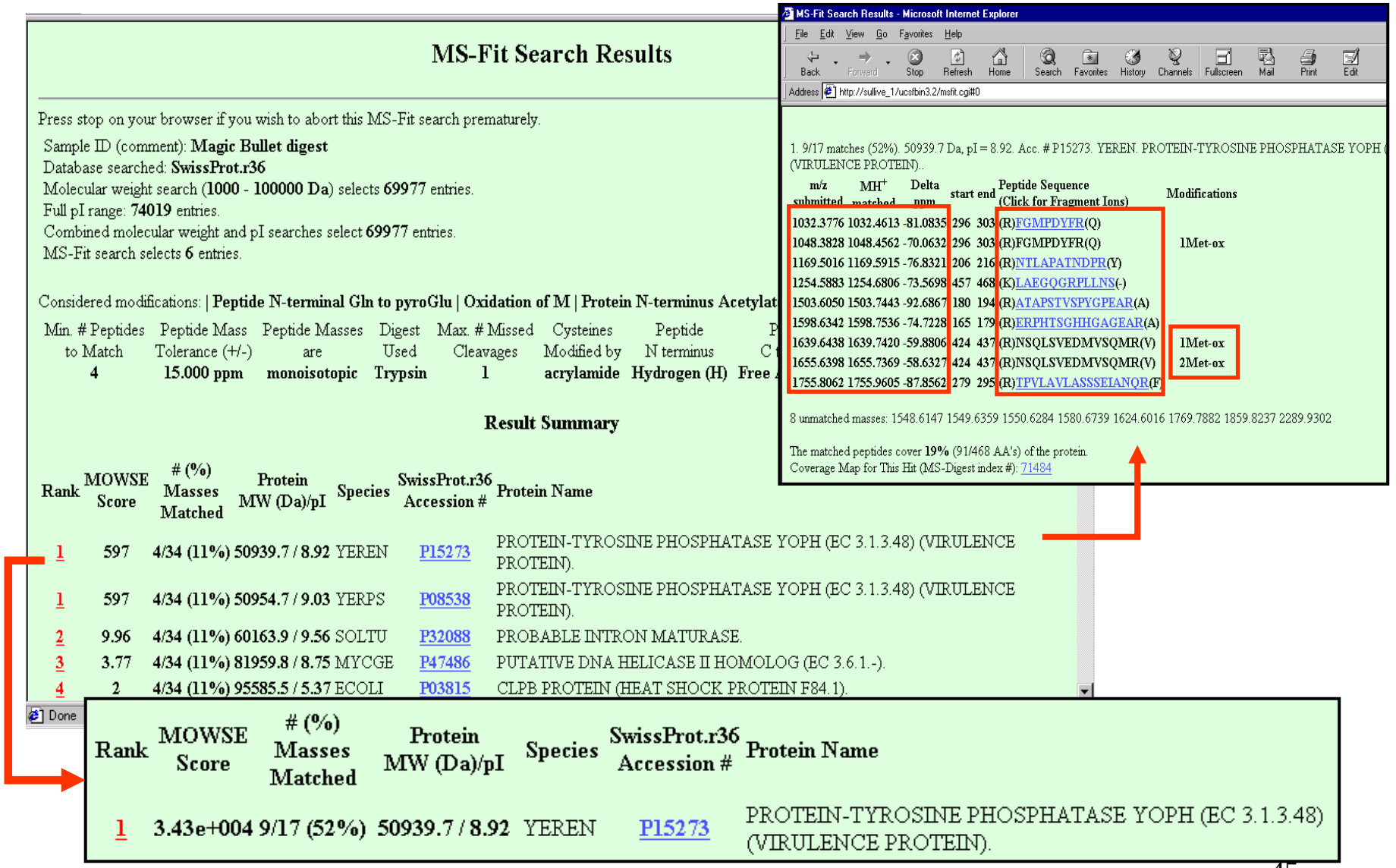


Match ----

protein identification



Search Result



MS-Fit Search Results

Press stop on your browser if you wish to abort this MS-Fit search prematurely.

Sample ID (comment): **Magic Bullet digest**
Database searched: **SwissProt.r36**
Molecular weight search (**1000 - 100000 Da**) selects **69977** entries.
Full pI range: **74019** entries.
Combined molecular weight and pI searches select **69977** entries.
MS-Fit search selects **6** entries.

Considered modifications: | **Peptide N-terminal Gln to pyroGlu** | **Oxidation of M** | **Protein N-terminus Acetylat**

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest Used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	P C
4	**15.000 ppm**	**monoisotopic**	**Trypsin**	**1**	**acrylamide**	**Hydrogen (H)**	**Free**

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	SwissProt.r36 Accession #	Protein Name
1	597	4/34 (11%)	50939.7 / 8.92	YEREN	P15273	PROTEIN-TYROSINE PHOSPHATASE YOPH (EC 3.1.3.48) (VIRULENCE PROTEIN).
1	597	4/34 (11%)	50954.7 / 9.03	YERPS	P08538	PROTEIN-TYROSINE PHOSPHATASE YOPH (EC 3.1.3.48) (VIRULENCE PROTEIN).
2	9.96	4/34 (11%)	60163.9 / 9.56	SOLTU	P32088	PROBABLE INTRON MATURASE.
3	3.77	4/34 (11%)	81959.8 / 8.75	MYCGE	P47486	PUTATIVE DNA HELICASE II HOMOLOG (EC 3.6.1.-).
4	2	4/34 (11%)	95585.5 / 5.37	ECOLI	P03815	CLPB PROTEIN (HEAT SHOCK PROTEIN F84.1).

Done

MS-Fit Search Results - Microsoft Internet Explorer

File Edit View Go Favorites Help

Back Forward Stop Refresh Home Search Favorites History Channels Fullscreen Mail Print Edit

Address http://sullive_1/ucsfbin3.2/msfit.cgi#0

1. 9/17 matches (52%). 50939.7 Da, pI = 8.92. Acc. # P15273. YEREN. PROTEIN-TYROSINE PHOSPHATASE YOPH (VIRULENCE PROTEIN)..

m/z submitted	MH+ matched	Delta ppm	start	end	Peptide Sequence (Click for Fragment Ions)	Modifications
1032.3776	1032.4613	-81.0835	296	303	(R)FGMPDYFR(Q)	
1048.3828	1048.4562	-70.0632	296	303	(R)FGMPDYFR(Q)	1Met-ox
1169.5016	1169.5915	-76.8321	206	216	(R)NTLAPATNDPR(Y)	
1254.5883	1254.6806	-73.5698	457	468	(K)LAEGQGRPLLNS(-)	
1503.6050	1503.7443	-92.6867	180	194	(R)ATAPSTVSPYGPEAR(A)	
1598.6342	1598.7536	-74.7228	165	179	(R)ERPHTSGHHGAGEAR(A)	
1639.6438	1639.7420	-59.8806	424	437	(R)NSQLSVEDMVSQMR(V)	1Met-ox
1655.6398	1655.7369	-58.6327	424	437	(R)NSQLSVEDMVSQMR(V)	2Met-ox
1755.8062	1755.9605	-87.8562	279	295	(R)TPVLAVLASSSEIANQR(F)	

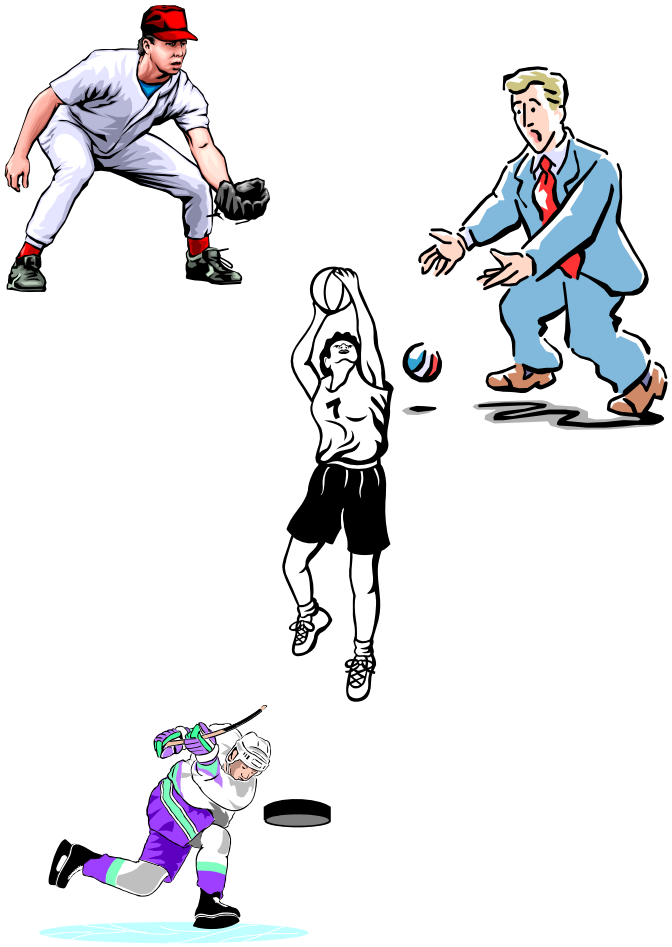
8 unmatched masses: 1548.6147 1549.6359 1550.6284 1580.6739 1624.6016 1769.7882 1859.8237 2289.9302

The matched peptides cover **19%** (91/468 AA's) of the protein.
Coverage Map for This Hit (MS-Digest index #): 71484

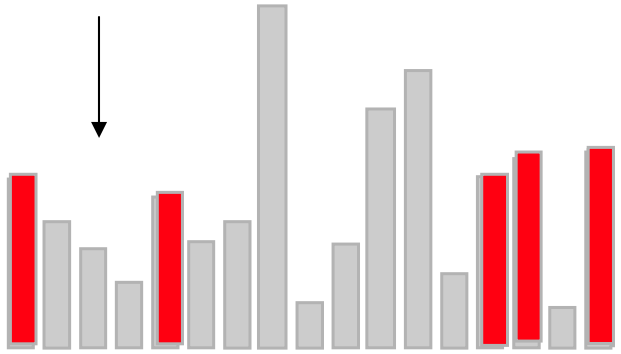
Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	SwissProt.r36 Accession #	Protein Name
1	3.43e+004	9/17 (52%)	50939.7 / 8.92	YEREN	P15273	PROTEIN-TYROSINE PHOSPHATASE YOPH (EC 3.1.3.48) (VIRULENCE PROTEIN).

Peptide Mass Fingerprinting (PMF)

– *Identify the sequence (Identity) of an unknown protein*



In-Gel Digestion

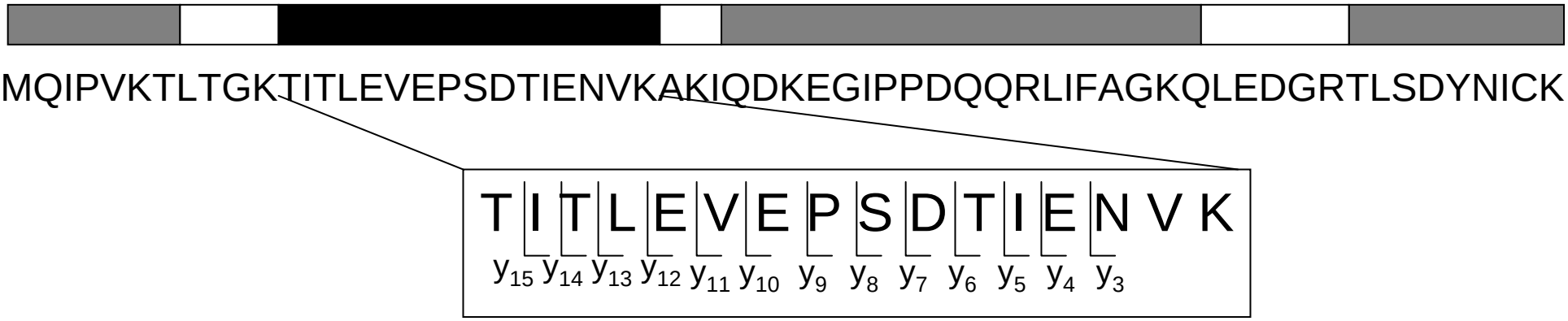


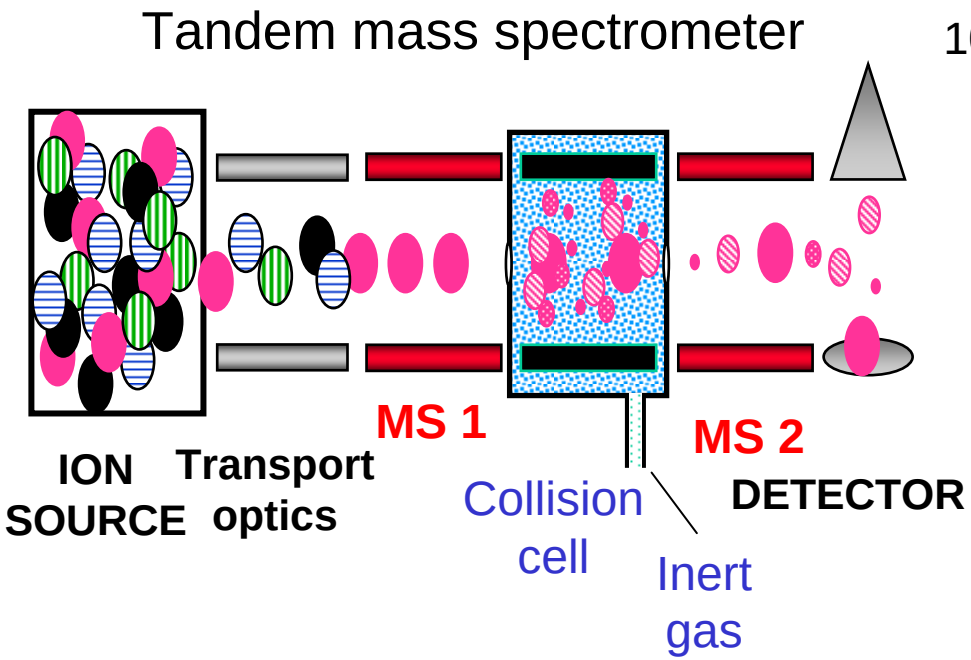
Extract peptides; mass analyze

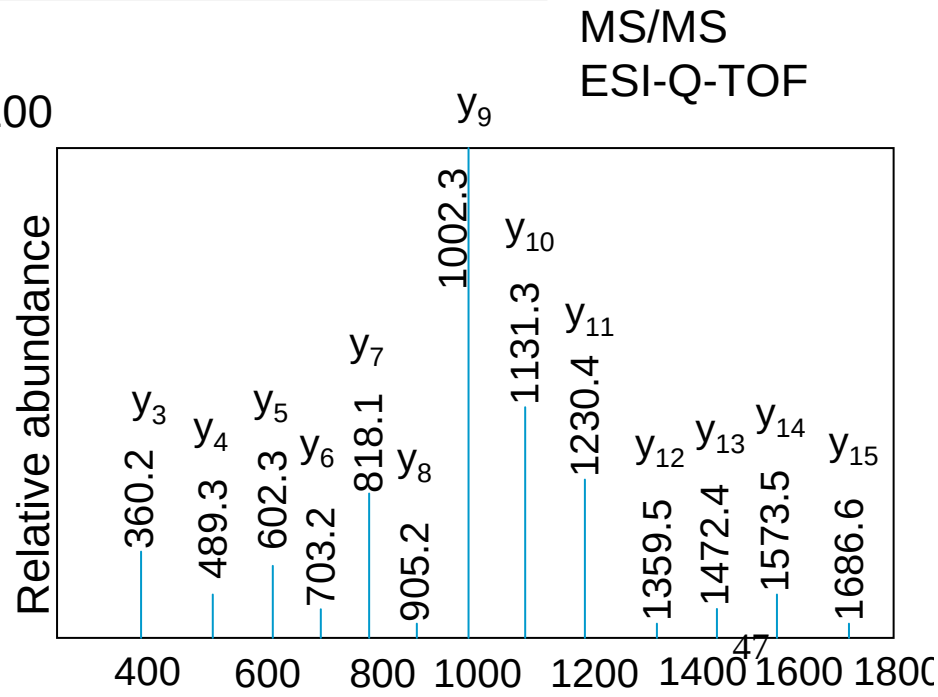
Each protein has a unique peptide mass list

Database search

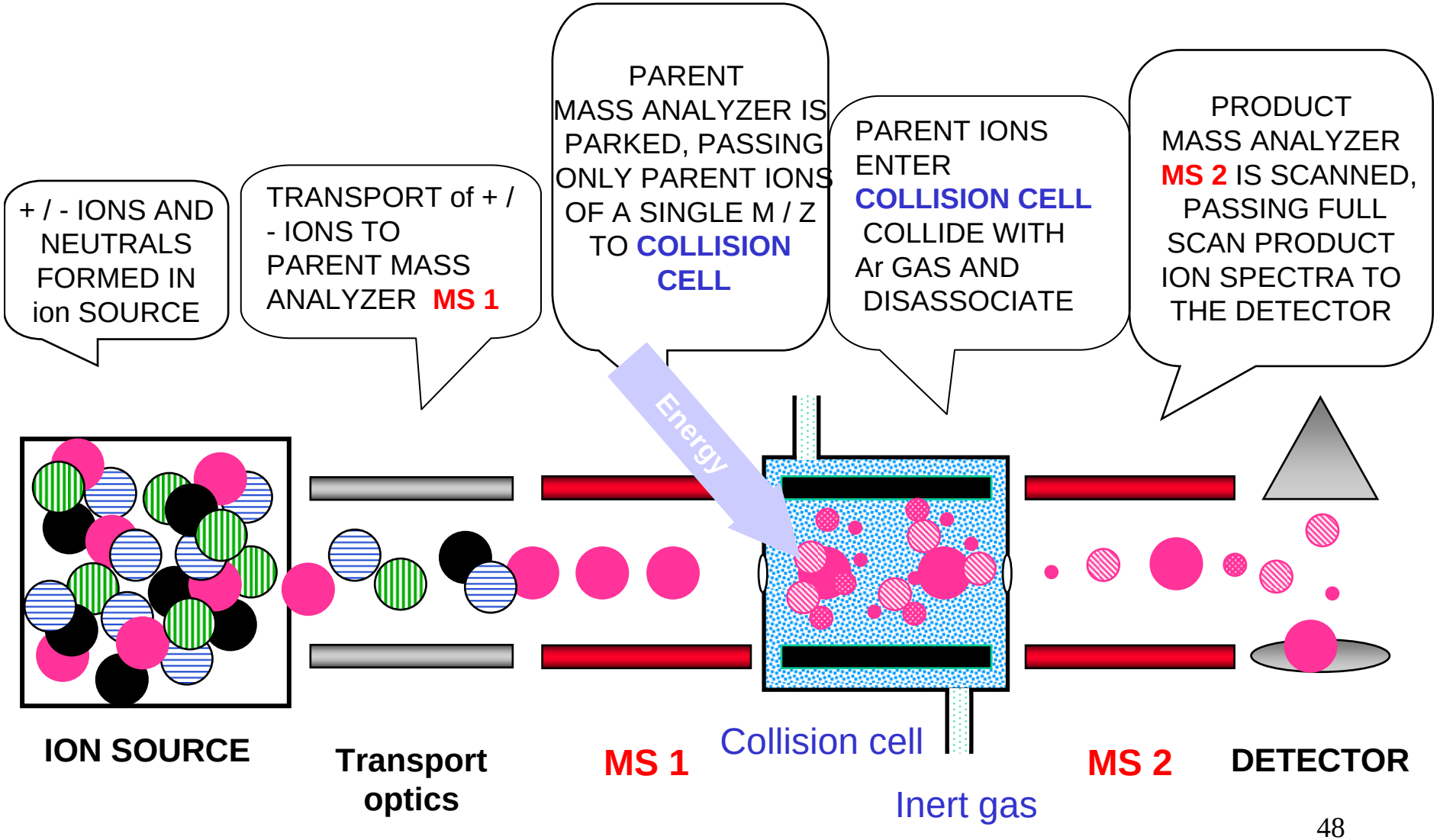
Two Ways of Measurement—
2. Peptide Sequencing (by MS/MS, or Tandem MS)





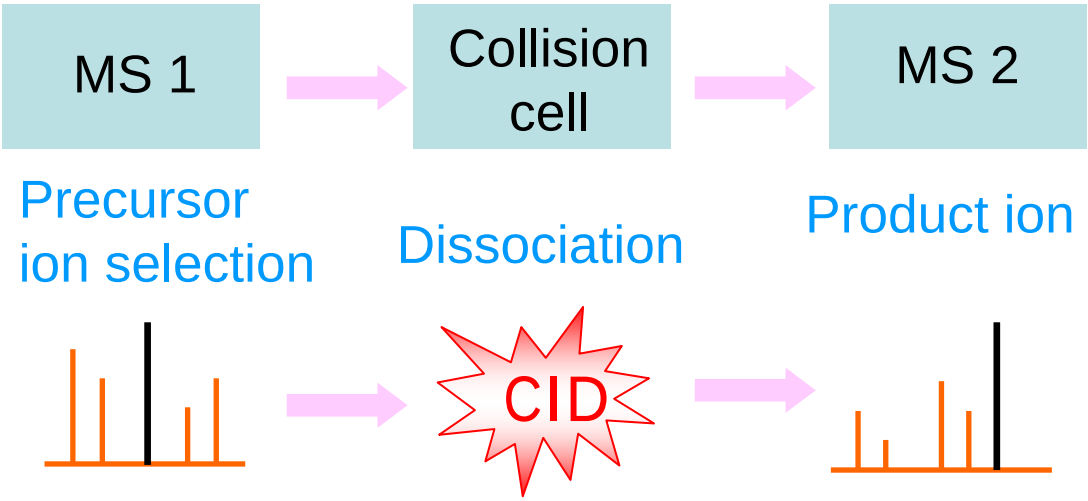


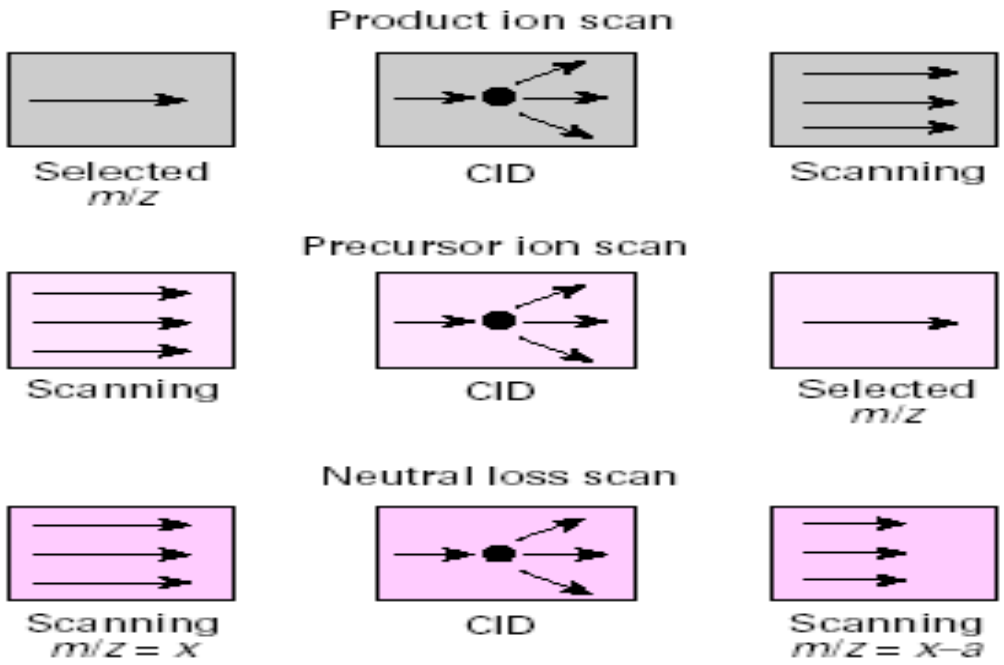
Tandem Mass Spectrometry



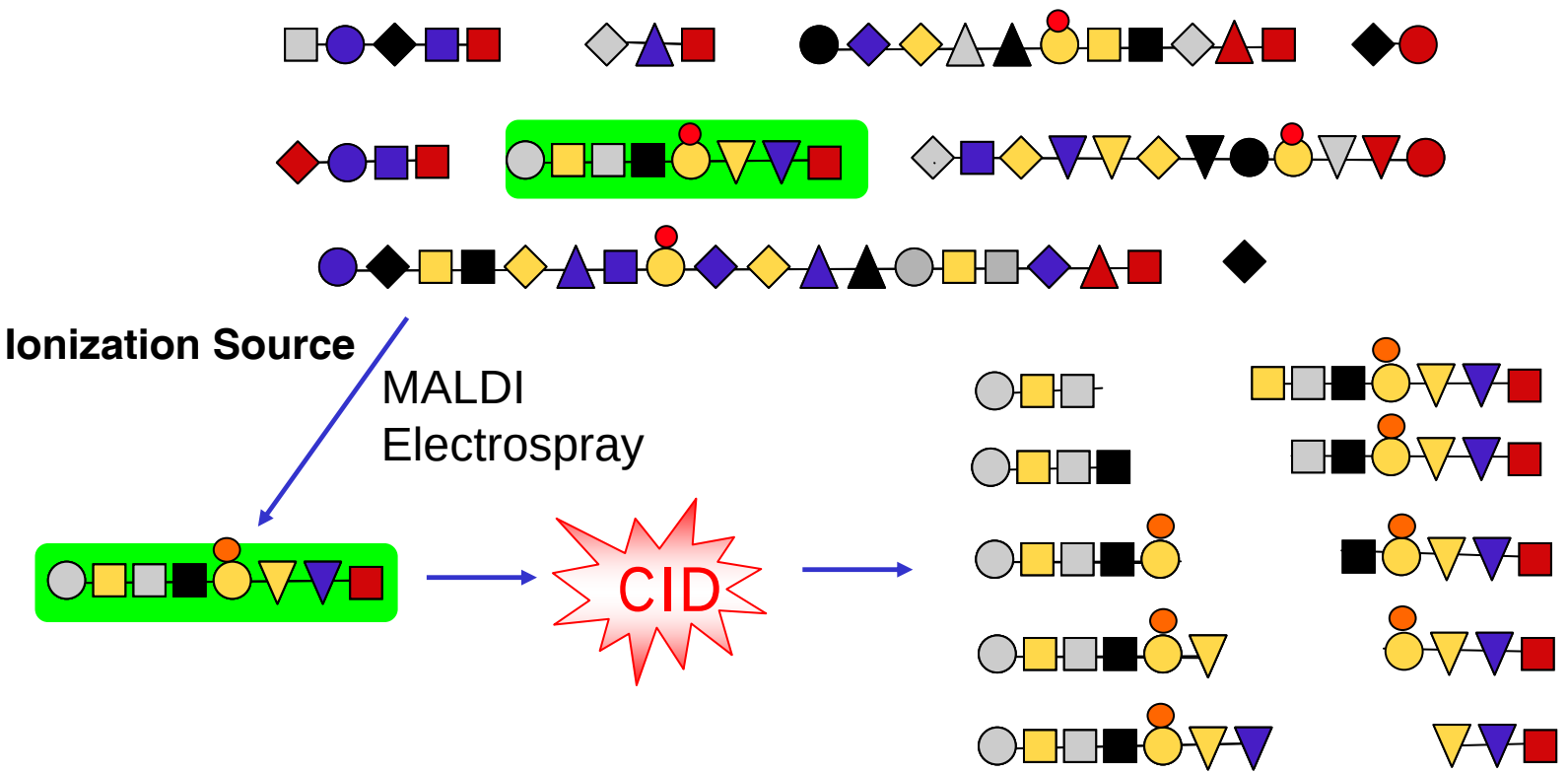
Methods of MS-MS

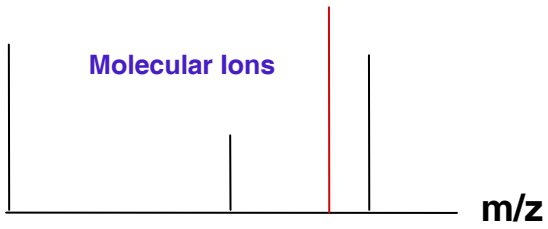
In space

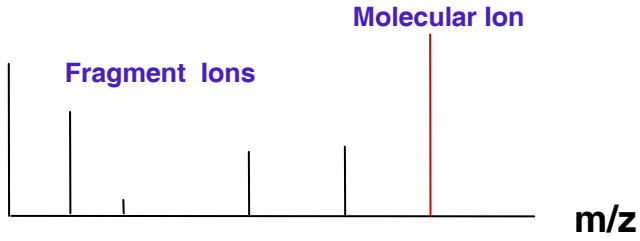




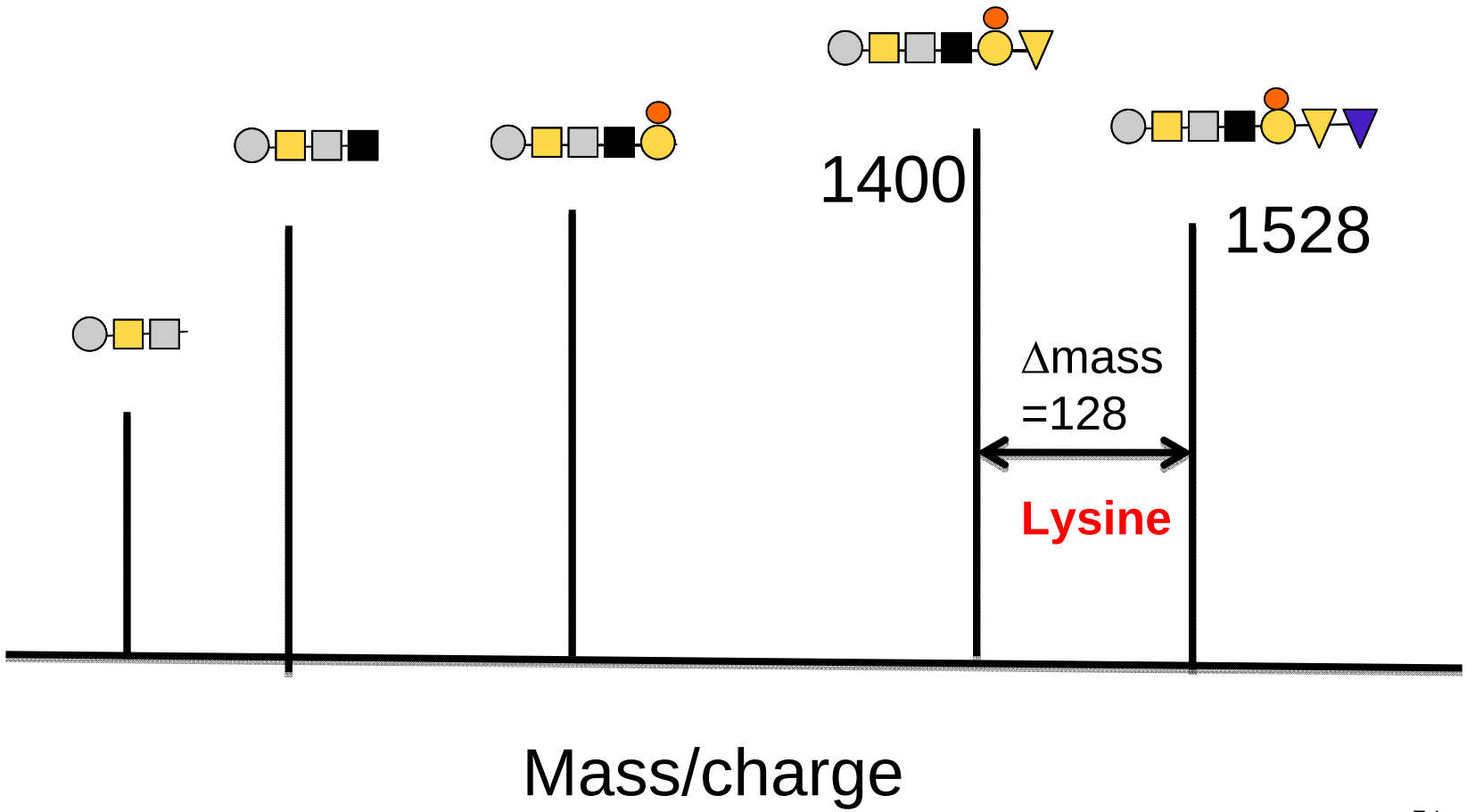
Tandem Mass Spectrometry (MS/MS)

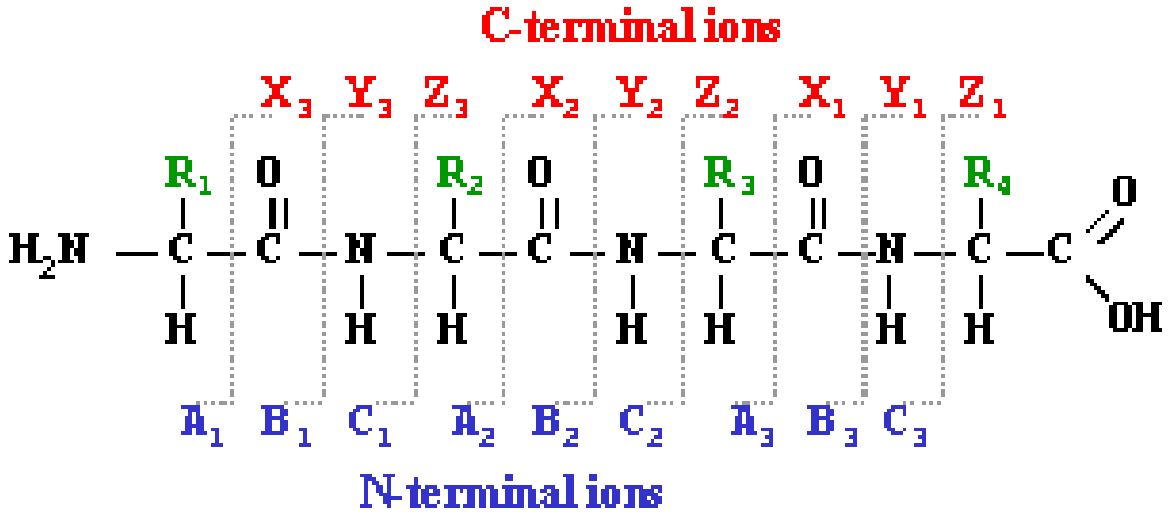






Peptide Fragment Mass Spectrum





Peptide Fragmentation

a,b,c – N-terminal side

x,y,z – C-terminal side

S-P-A-F-D-S-I-M-A-E-T-L-K

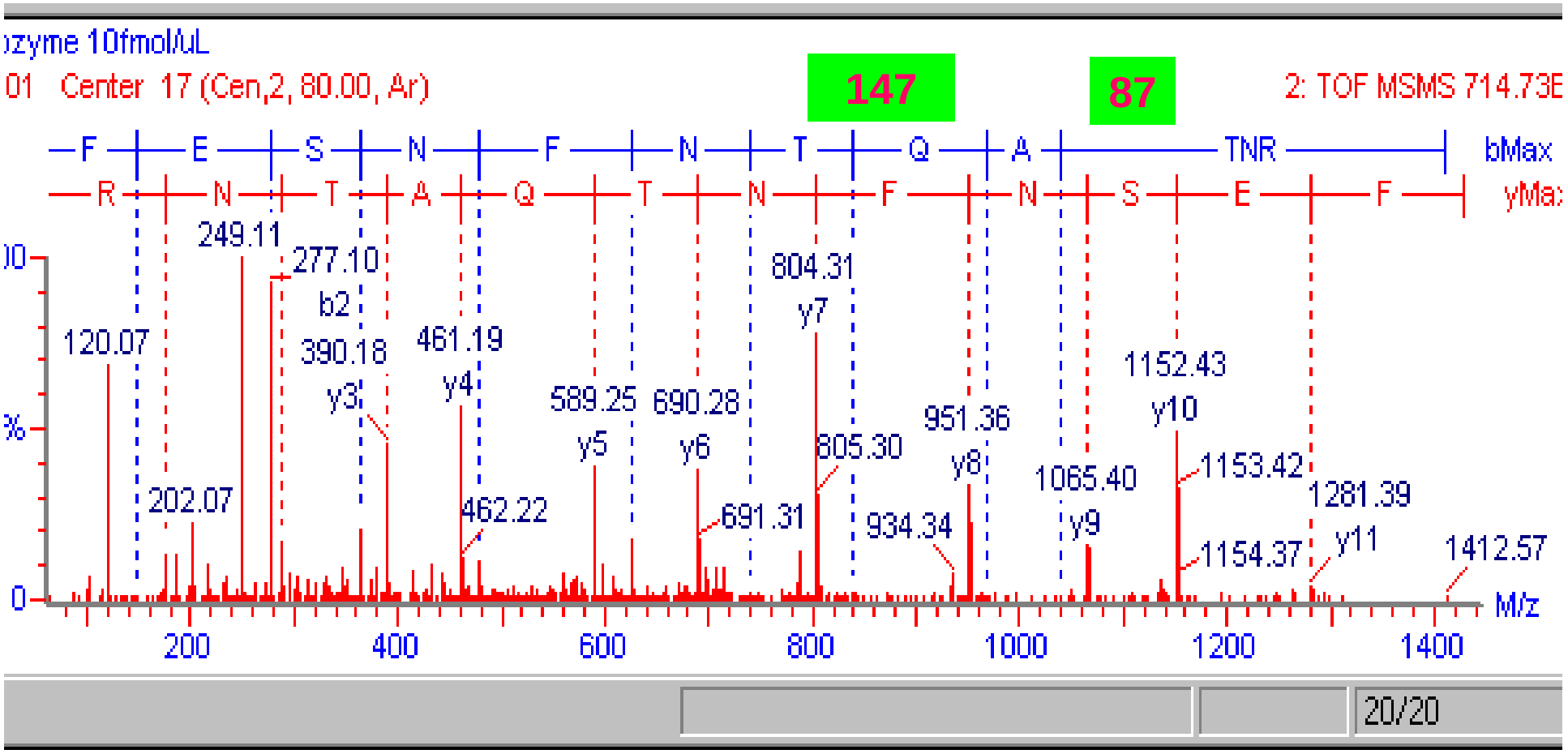
(protonated mass 1410.6)

mass+	b-ions	y-ions	mass+
88.1	S	PAFDSIMAETLK	1323.6
185.2	SP	AFDSIMAETLK	1226.4
256.3	SPA	FDSIMAETLK	1155.4
403.5	SPAF	DSIMAETLK	1008.2
518.5	SPAFD	SIMAETLK	893.1
605.6	SPAFDS	IMAETLK	806.0
718.8	SPAFDSI	MAETLK	692.3
850.0	SPAFDSIM	AETLK	561.7
921.1	SPAFDSIMA	ETLK	490.6
1050.2	SPAFDSIMAE	TLK	361.5
1151.3	SPAFDSIMAET	LK	260.4
1264.4	SPAFDSIMAETL	K	147.2

Residue Mass of Amino Acids

Table9.2	Symbols and residue masses of the protein amino acids		
Name	**Symbol**	**Residue mass**	**Side-chain**
Alanine	A,Ala	71.079	CH_3-
Arginine	R,Arg	156.188	$HN=C(NH_2)-N-(CH_2)_3$
Asparagine	N,Asn	114.104	$H_2N-CO-CH_2$-
Aspartic acid	D,Asp	115.089	$HOOC-CH_2$-
Cysteine	C,Cys	103.145	$HS-CH_2$-
Glutamine	Q,Gln	128.131	$H_2N-CO-(CH_2)_2$-
Glutamic acid	E,Glu	129.116	$HOOC-(CH_2)_2$-
Glycine	G,Gly	57.052	H-
Histidine	H,His	137.141	Imidazole-CH_2-
Isoleucine	I,Ile	113.16	$CH_3-CH_2-CH(CH_3)$-
Leucine	L,Leu	113.16	$(CH_3)_2-CH-CH_2$-
Lysine	K,Lys	128.17	$H_2N-(CH_2)_4$-
Methionine	M,Met	131.199	$CH_3-S-(CH_2)_2$-
Metsulphoxide	Met.SO	147.199	$CH_3-S(O)-(CD_2)_2$-
Phenylalanine	F,Phe	147.177	Phenyl-CH_2-
Proline	P,Pro	97.117	Phrrolidone-CH-
Serine	S,Ser	87.078	$HO-CH_2$-
Threonine	T,Thr	101.105	$CH_3-CH(OH)$-
Tryptophan	W,Trp	186.213	ndole-NH-CH=C-CH
Tyrosine	Y,Tyr	163.176	4-OH-Phenyl-CH_2-
Valine	V,Val	99.133	$CH_3-CH(CH_2)$-

Peptide Sequencing



F (phenylalanine): 147.177

S (Serine) : 87.078