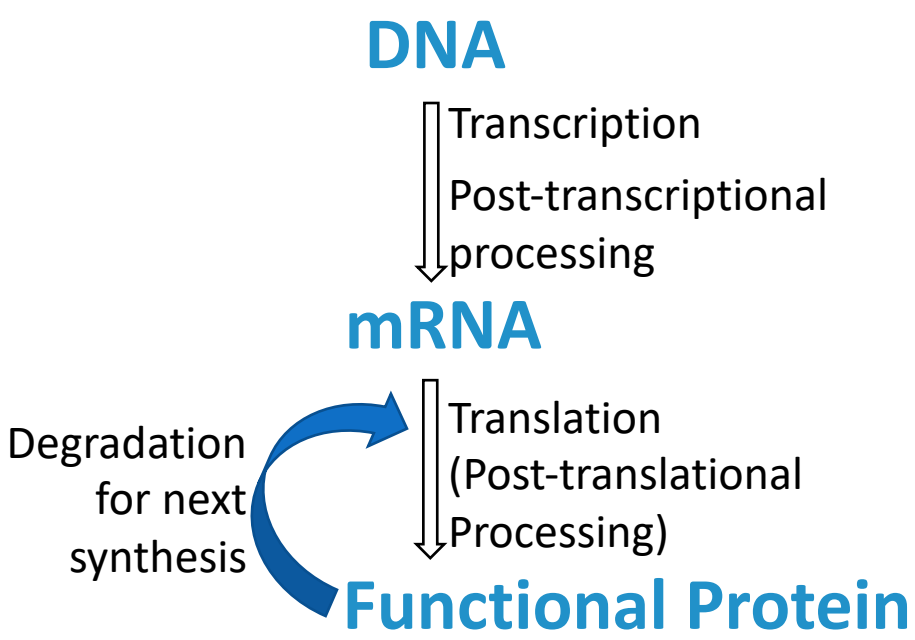
LECTURE 2-1: IDENTIFICATION OF PROTEINS IN COMPLEX MIXTURES - A MASS SPECTROMETRY APPROACH

Bio312

Instructor: Dr. Lanlan Han
E-mail: lanlan.han@xjtlu.edu.cn

Introduction of Proteomics

- Determine the functions of genes and their products, allowing them to be linked into pathways and networks, and ultimately providing a detailed understanding of how biological systems work.



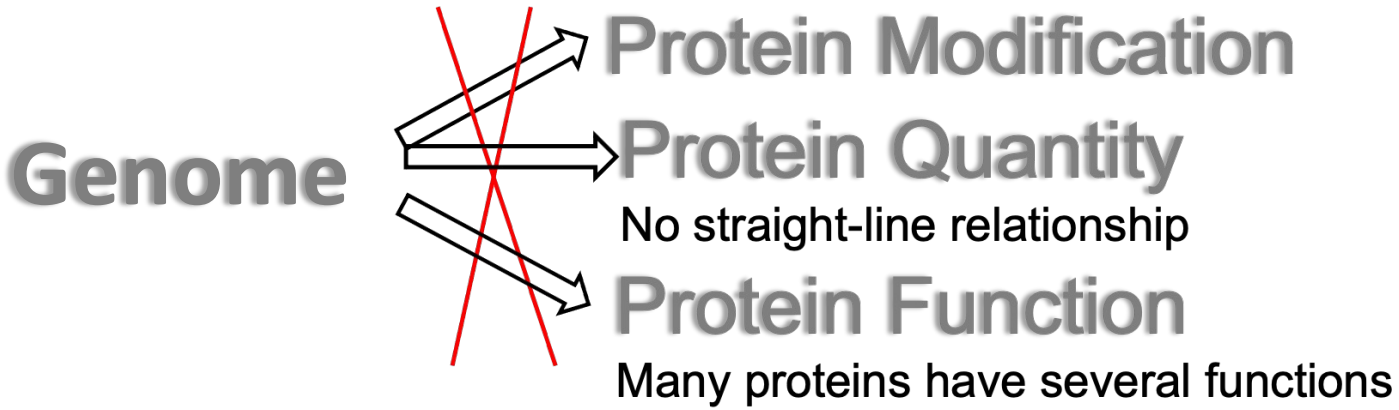
Genome **“Genomics”**

Transcriptome **“Transcriptomics”**

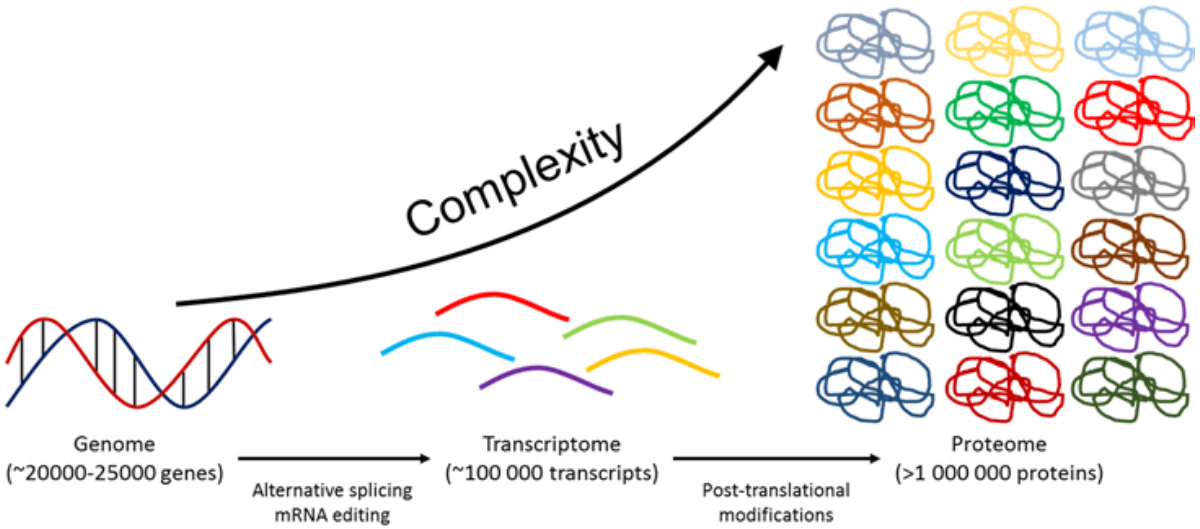
Proteome **“Proteomics”**

From Genomics to Proteomics

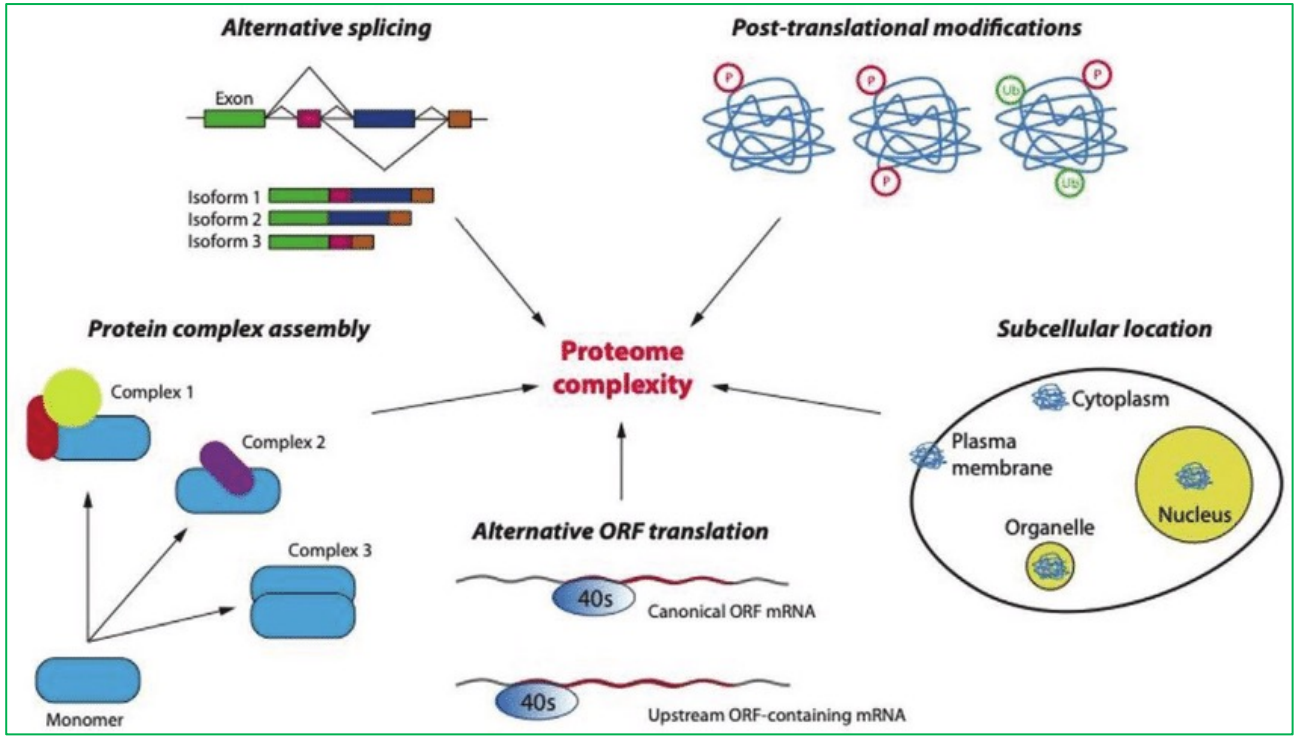
- Proteomics is a rapid growing area of molecular biology that is aimed to characterize the entire proteins of a cell line, tissue, or organism.



Complexity of Proteomics



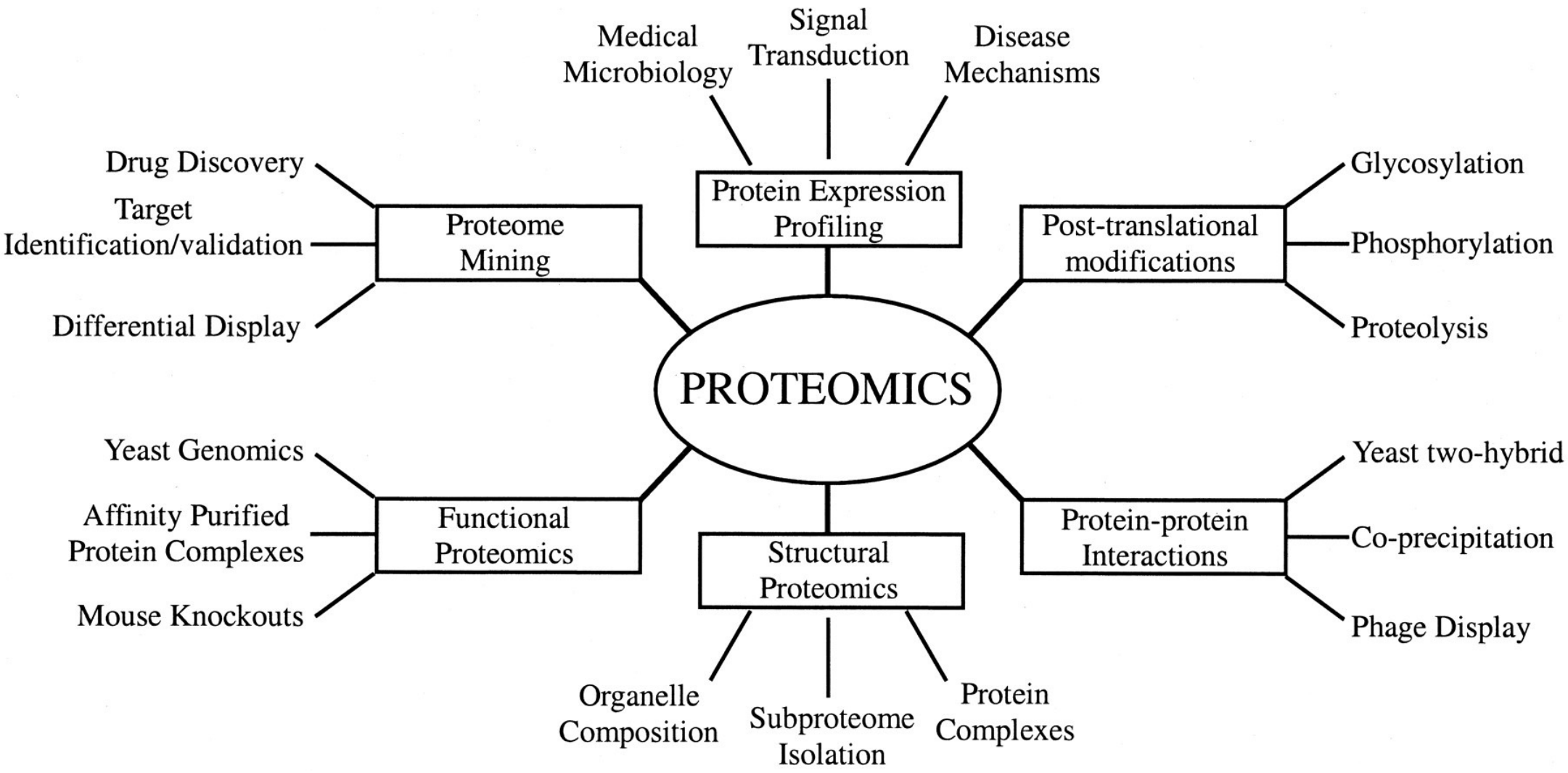
1. Numbers of proteins
2. Diversity of cells or tissues
3. Dynamic changes in protein levels



https://www.researchgate.net/profile/Juan-Pelta/publication/342275076/figure/fig1/AS:904991556530176@1592778209035/Overview-of-proteome-complexity-Numerous-factors-contribute-to-the-generation-of-complex.ppm

https://www.researchgate.net/profile/Floris-Van-Den-Brink/publication/313535513/figure/fig4/AS:460194112118787@1486730234132/The-human-proteome-contains-many-more-species-compared-to-the-human-genome-making-it.png

Proteomics Types and Application



Graves PR, Haystead TA. Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev*. 2002;66(1):39-63. doi:10.1128/MMBR.66.1.39-63.2002

Proteomics – How do we want to achieve?

1. Identify and quantify proteins in complex mixtures/complexes
   MS and MS/MS

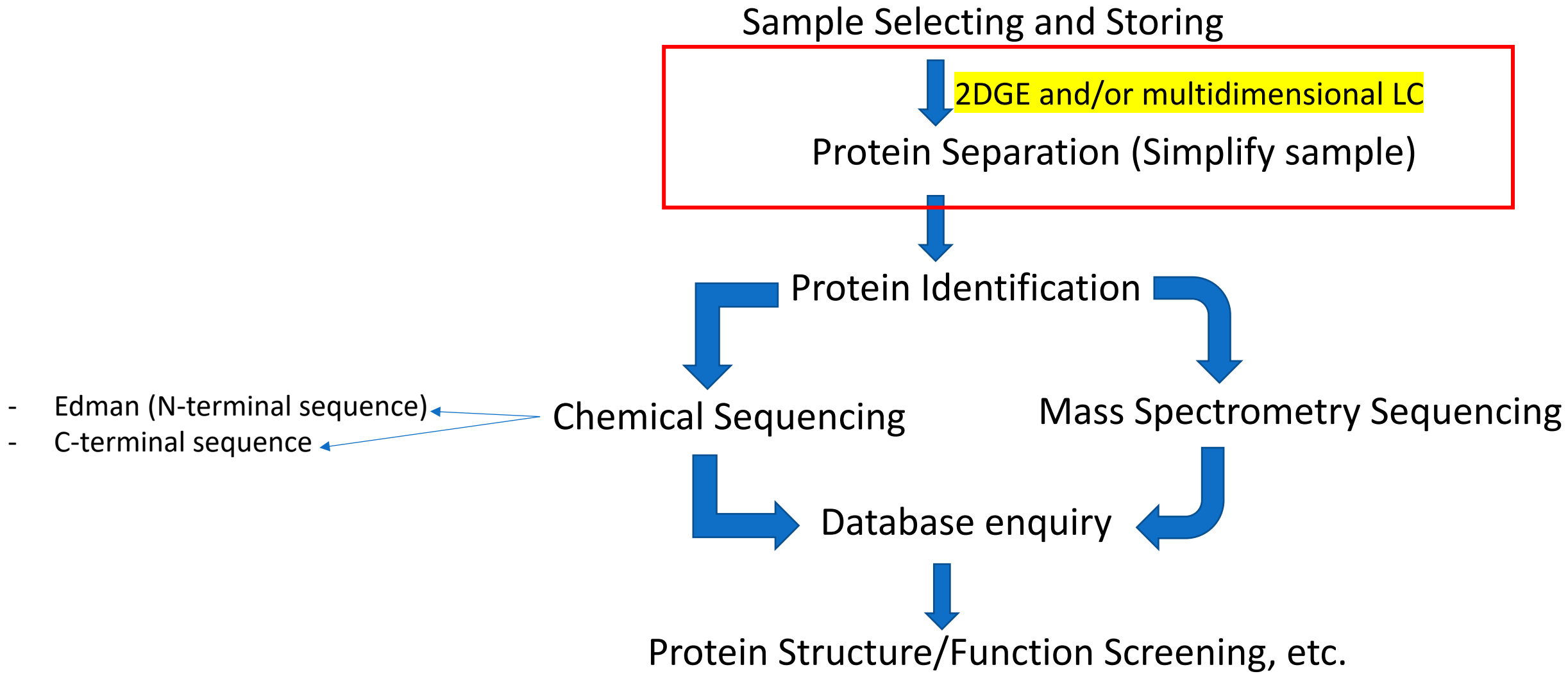
2. Identify global protein-protein interactions
   MS and MS/MS, Y2H

3. Define protein localizations within cells
   High-throughput microscopy, organelle pull-down

4. Measure and characterize post-translational modifications
   MS techniques

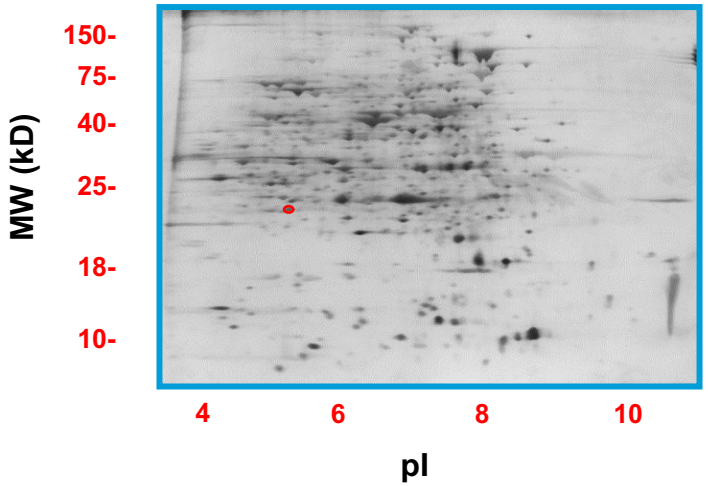
5. Measure and characterize activity (e.g., substrate specificity, etc.)
   Protein arrays

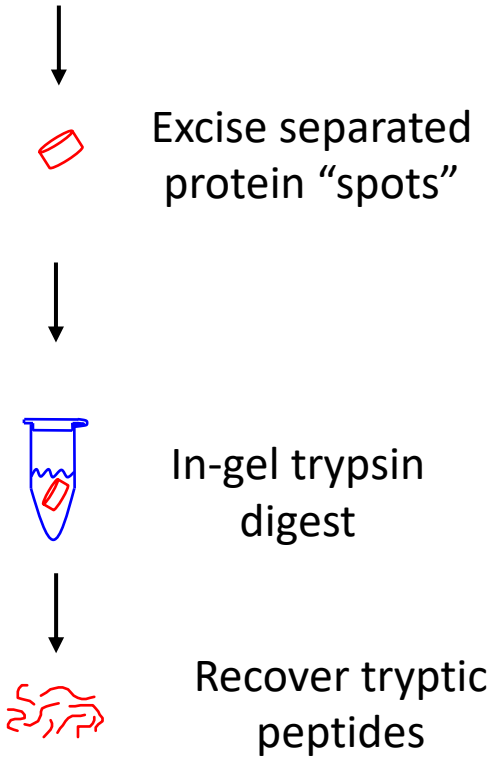
General Workflow in Proteomics Analysis



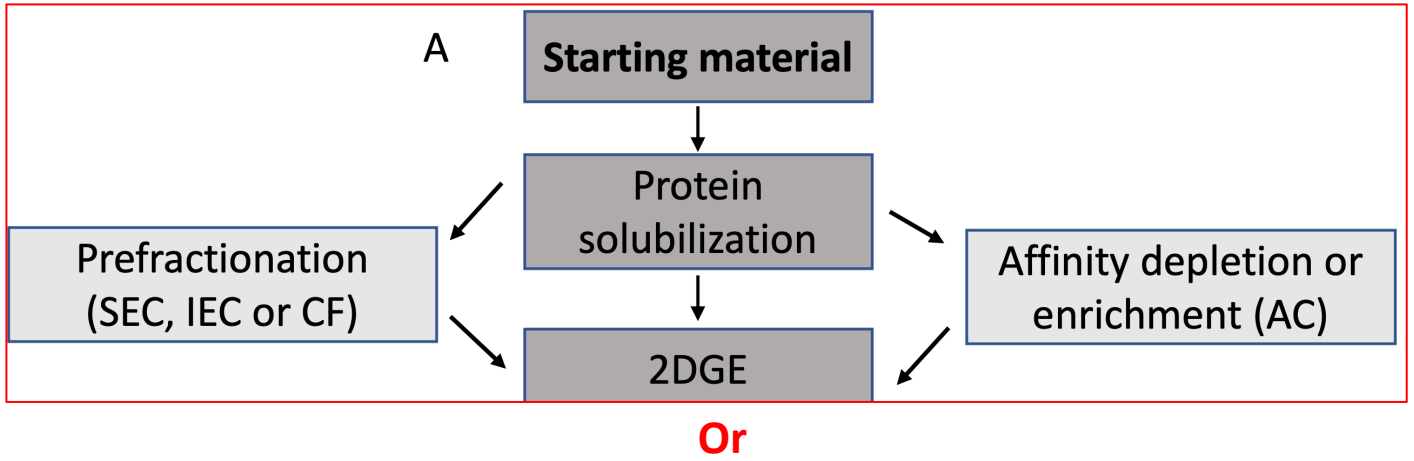
Strategies for Protein Separation[1]

2-D Gel Electrophoresis





Strategies for Protein Separation$_2$



Multidimensional Liquid chromatography

- The **charged groups**, **hydrophobic region**, **size** and **specific ligand binding affinity** largely determine the purification behavior of proteins.

- **ExPASy - ProtParam tool** (https://web.expasy.org/protparam/):
  A computation tool to calculate various physical and chemical parameters for a given protein

ExPASy - ProtParam tool

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID)

Or you can paste your own amino acid sequence (in one-letter code) in the box below:

```
MGKRILLLEKERNLAHFLSLELQKEQYRVDLVEEGQKALSMALQTDYDLILLNVNLGDMM
AQDFAEKLSRTKPASVIMILDHWEDLQEELEVVQRFAVSYIYKPVLIENLVARISAIFRGRDFI
DQHCSLMKVPRTYRNLRIDVEHHTVYRGEEMIALTRREYDLLATLMGSKKVLTREQLLES
VWKYESATETNIVDVYIRYLRSKLDVKGQKSYIKTVRGVGYTMQE
```

RESET Compute parameters

Number of amino acids: 229

Molecular weight: 26880.19

Theoretical pI: 6.55

Amino acid composition: CSV format

Ala (A)	12	5.2%
Arg (R)	18	7.9%
Asn (N)	6	2.6%
Asp (D)	13	5.7%
Cys (C)	1	0.4%
Gln (Q)	11	4.8%
Glu (E)	21	9.2%
Gly (G)	9	3.9%
His (H)	5	2.2%
Ile (I)	14	6.1%
Leu (L)	31	13.5%
Lys (K)	15	6.6%
Met (M)	9	3.9%
Phe (F)	5	2.2%
Pro (P)	3	1.3%
Ser (S)	12	5.2%
Thr (T)	11	4.8%
Trp (W)	2	0.9%
Tyr (Y)	12	5.2%
Val (V)	19	8.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 34
Total number of positively charged residues (Arg + Lys): 33

Atomic composition:

Carbon	C	1201
Hydrogen	H	1939
Nitrogen	N	327
Oxygen	O	350
Sulfur	S	10

Formula: $C_{1201}H_{1939}N_{327}O_{350}S_{10}$
Total number of atoms: 3827

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

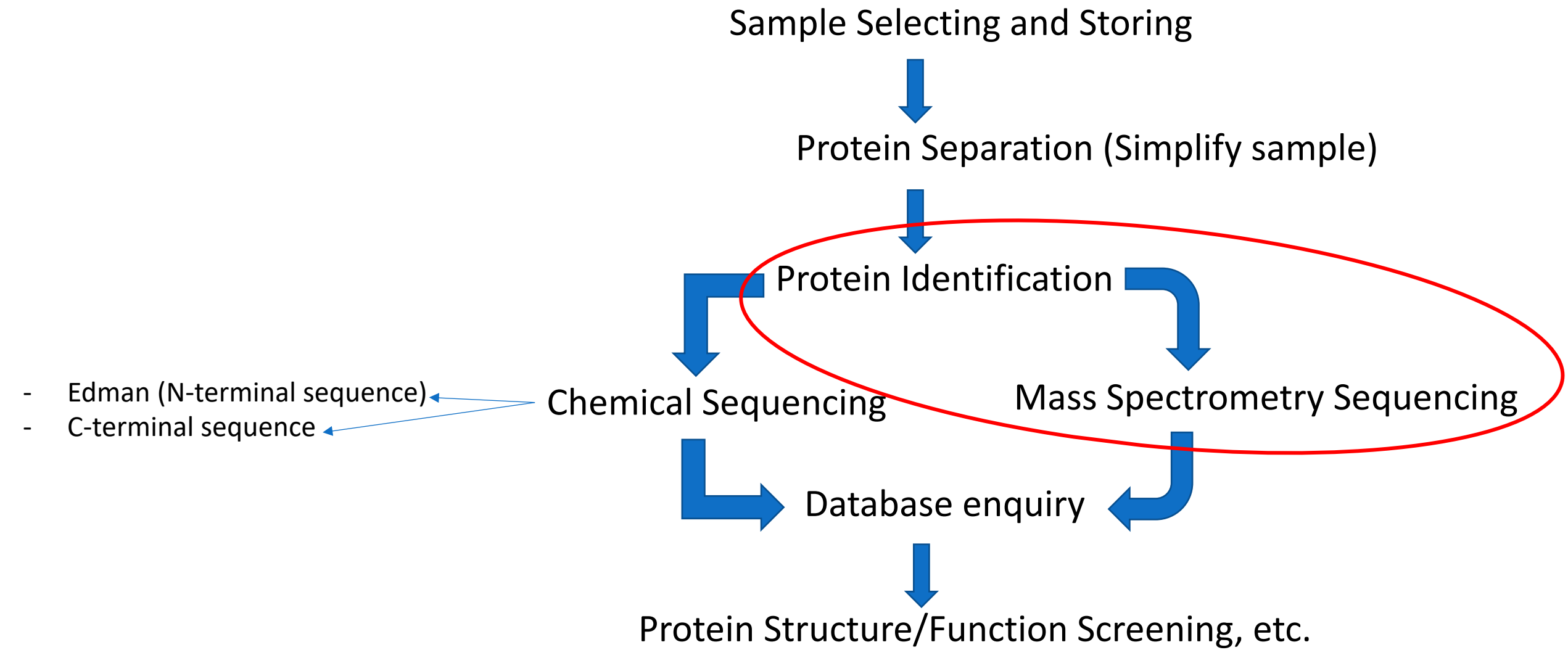
Ext. coefficient	28880
Abs 0.1% (=1 g/l)	1.074, assuming all pairs of Cys residues form cystines
Ext. coefficient	28880
Abs 0.1% (=1 g/l)	1.074, assuming all Cys residues are reduced

Beer's Law
A = ebc

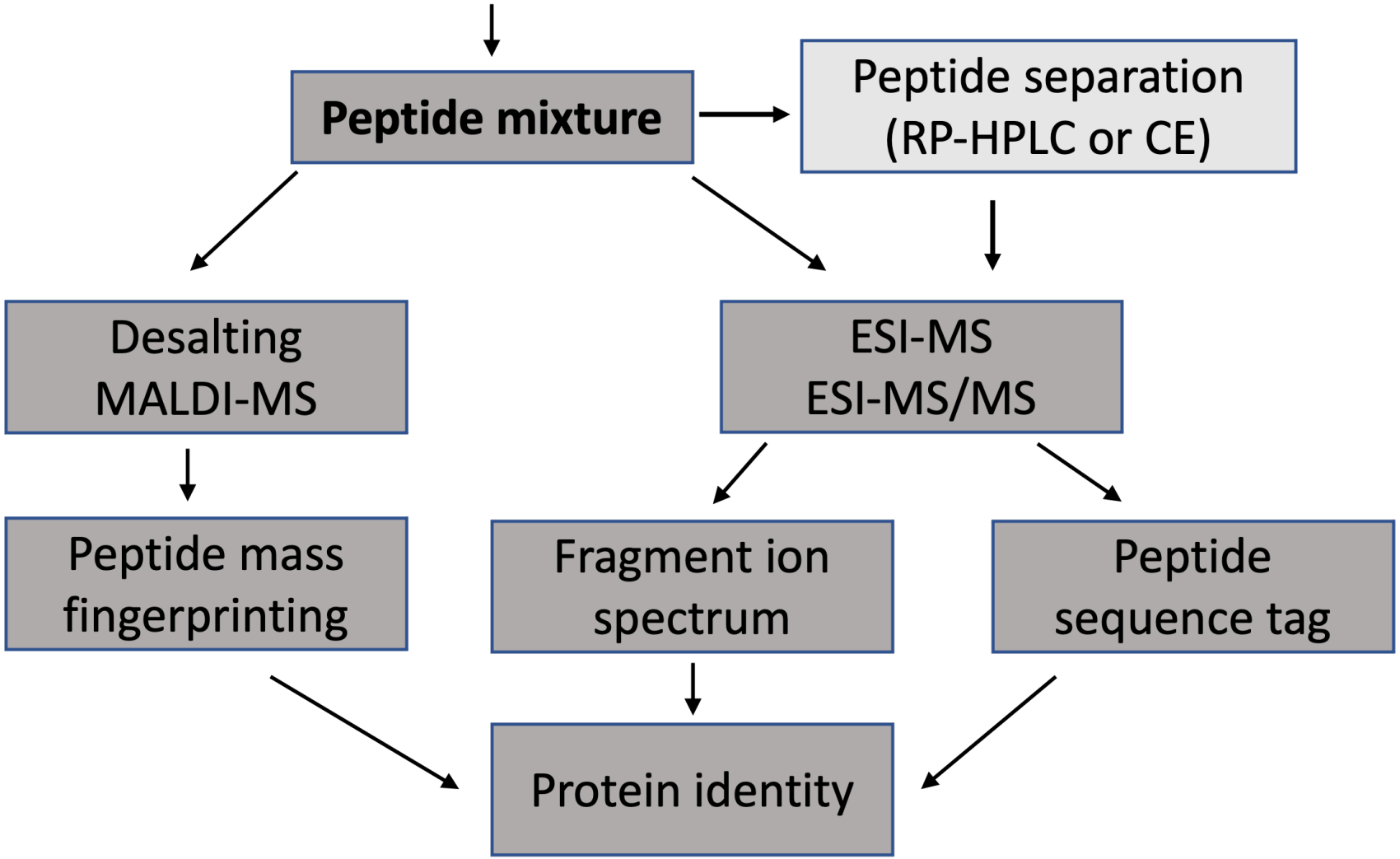
Multi-Dimensional Liquid Chromatography

- Pros:
  - Large sample volume
  - Detect low abundancy protein
  - Good for separating membrane proteins or very basic proteins
  - Separate both proteins and peptides
  - Connect directly to MS
- Cons:
  - Can not know the pI and MW of proteins
  - Need to consider the compatibility of buffers or solvents in different steps

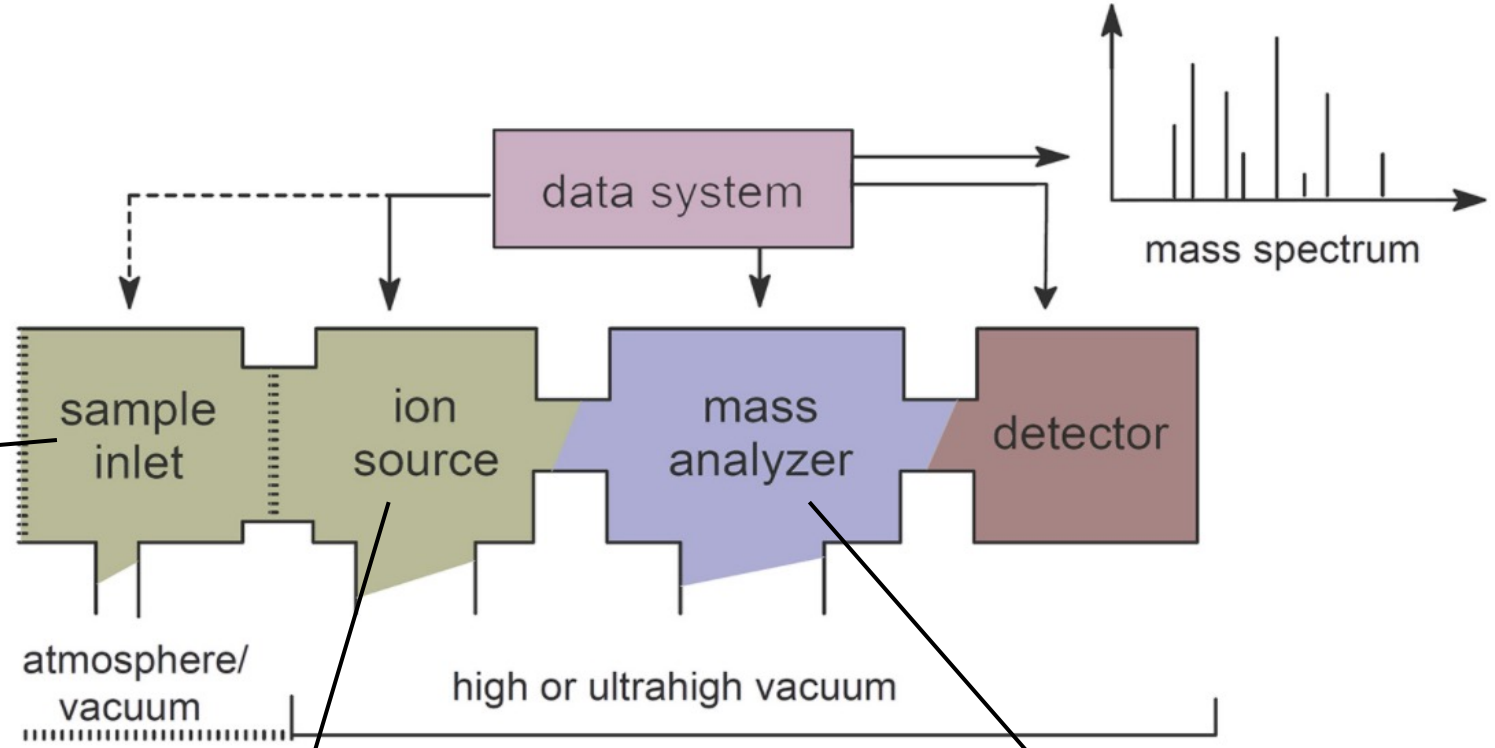
General Workflow in Proteomics Analysis



Mass Spectrometry-based Protein Identification



General Scheme of a Mass Spectrometer



RP-HPLC, Sample Plate, CE, etc.

ElectroSpray Ionization (ESI)

Matrix-Assisted Laser Desorption/Ionization (MALDI) etc.

Triple Quadrupole, Time of Flight (TOF), Ion trap, Fourier transform ion cyclotron resonance (FT-ICS), etc.

https://dgms.eu/en/about-dgms/this-is-mass-spectrometry/mass-spectrometers/

MS Basics

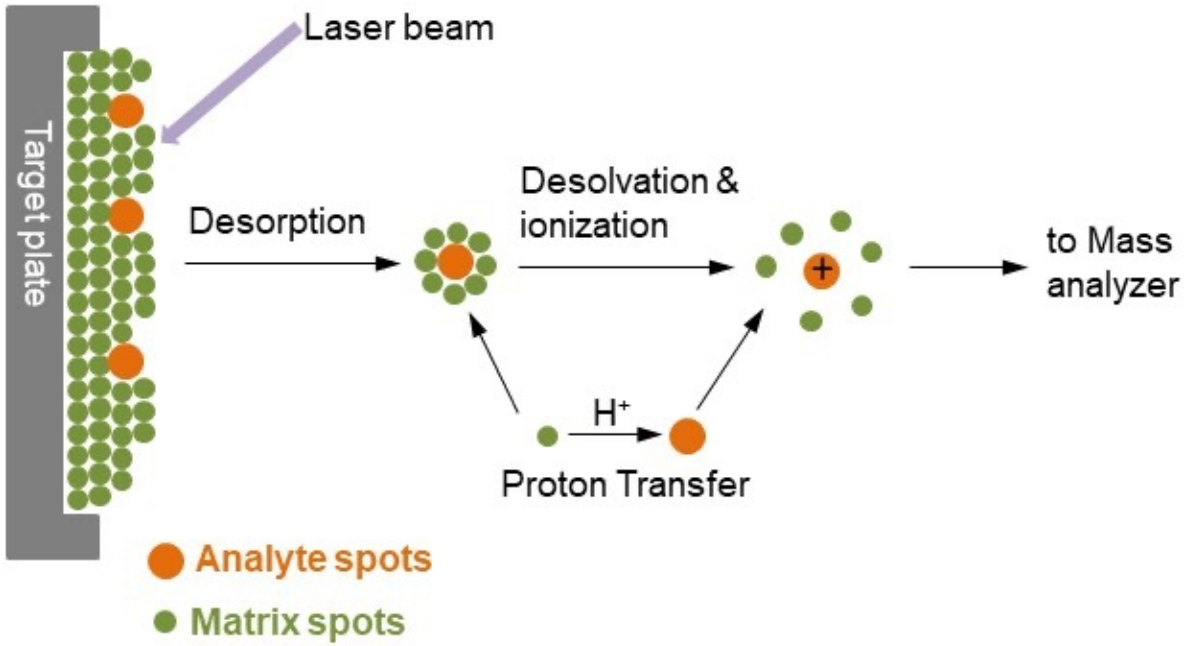
1. Soft Ionization

Goal: ionize (i.e., charge) peptide fragments without destroying molecule

- **Positive ionization** (protonate amine groups)
- Negative ionization (deprotonate carboxylics and alcohols)

- Two common ionization methods used in proteomics:
  - MALDI (Matrix-Assisted Laser Desorption/Ionization)
  - ESI (ElectroSpray Ionization)

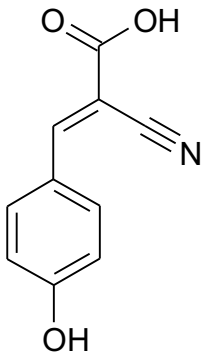
MALDI



- Peptide/protein analytes of interest are mixed with an excess of an aromatic matrix molecule and are co-crystallized on the MALDI target plate
- The crystals are targeted by a shorter laser pulse (UV pulse)
- Matrix molecules absorb energy and the heat released results in the desorption (sublimation)
- Analyte is ionized by gas-phase proton transfer (perhaps from ionized matrix molecules)
  - Protonation with one charge (H^+)

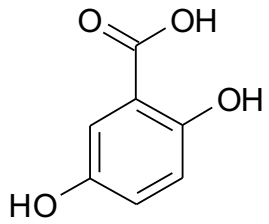
https://www.creative-proteomics.com/images/MALDI-TOF-mass-spectrometry-1-MALDI.png

MALDI Matrices



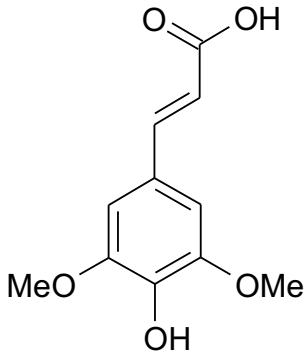
4-hydroxy-α-cyanocinnamic acid ("alpha-cyano" or 4-HCCA)

peptides



2,5-dihydroxybenzoic acid (DHB)

peptides and **proteins**



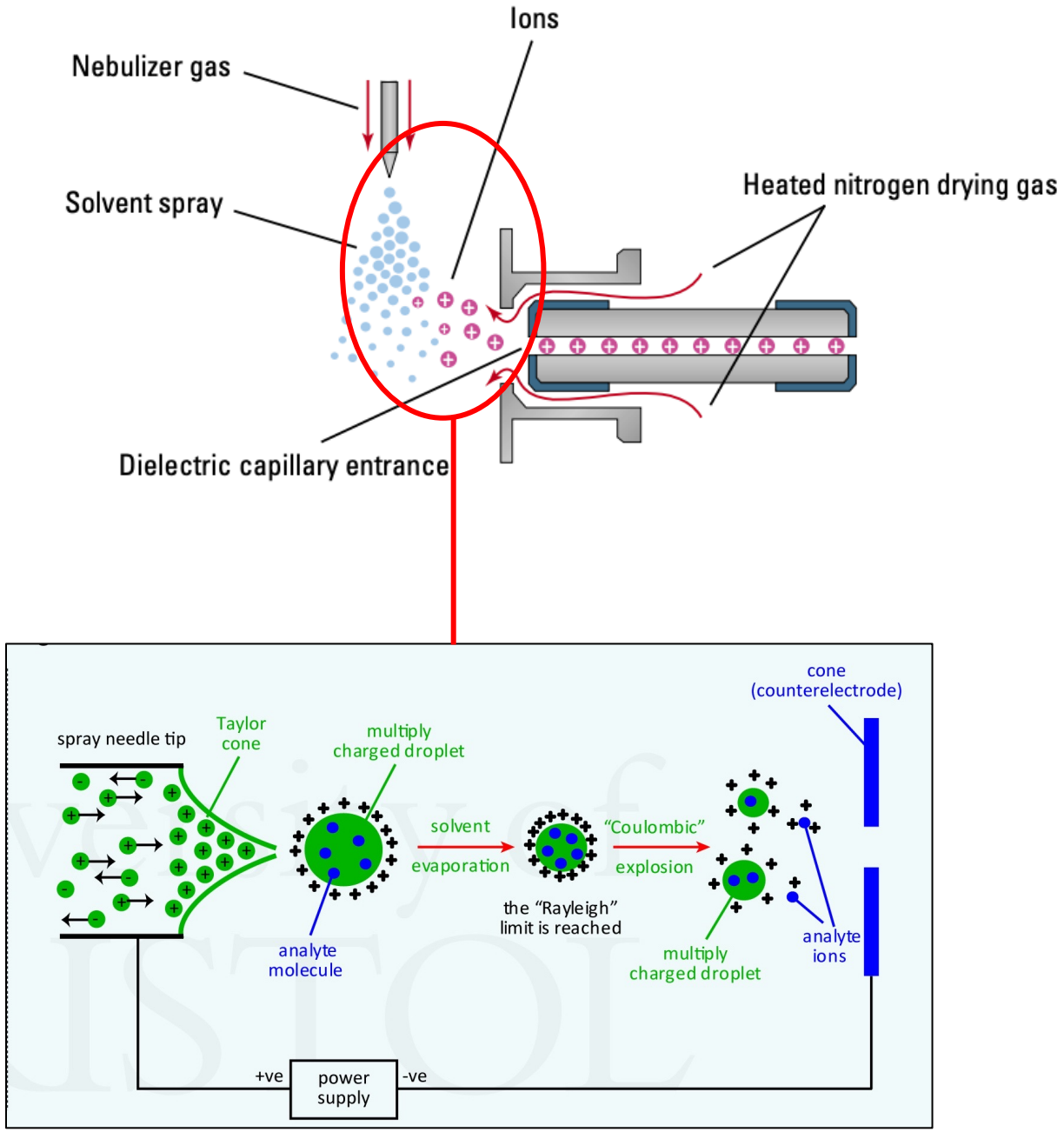
3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid)

proteins

matrices for 337 nm irradiation

ESI

- Analyte dissolved in a suitable solvent flows through a small diameter capillary tube
  - ESI is compatible with liquid chromatography (LC)
- Liquid in the presence of a high electric field generates a fine “mist” or aerosol spray of highly charged droplets
  - Multiple charged ions are possible (1^+, 2^+, 3^+, 4^+, etc.)



MS Basics

2. MS analyzers

Goal: Separate ions based on m/z (mass/charge) ratio

- Time-of-flight (TOF)
- Triple quadrupole
- Ion trap
- Fourier Transform Ion Cyclotron Resonance (FT-ICR)

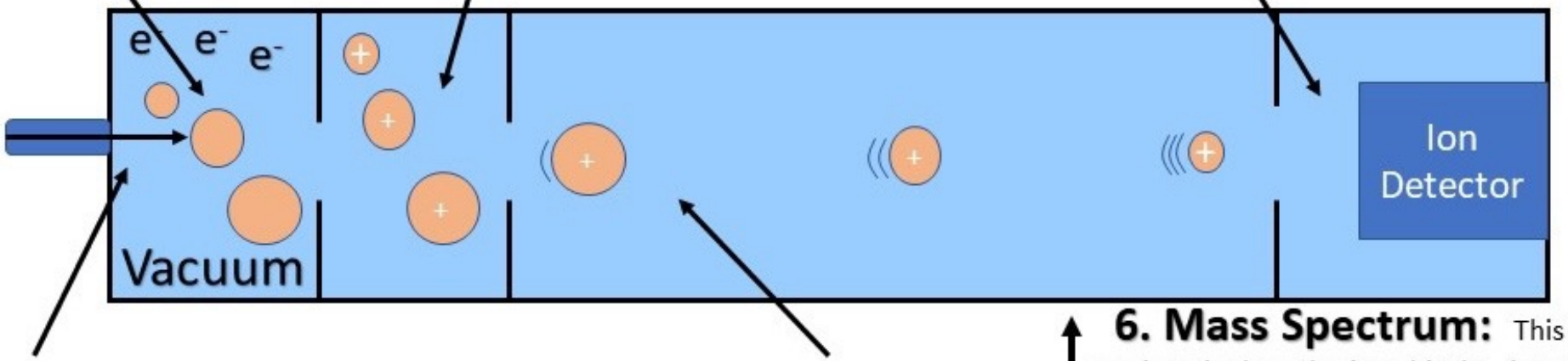
TOF

How does a Time of Flight Mass Spectrometer work?

2. Ionisation: They are then ionised into positive ions by firing electrons at them (knocking electrons from the outer energy level) This is called ***Electron Impact Ionisation***.

3. Electric Field: The positive ions are ***accelerated*** by an electric field. Lighter ions will accelerate greater than heavier ones.

5. Detector: Finally they hit the detector. Lighter ions will hit first, heavier ions later. The ions gain an electron when they hit the detector creating an electric current that can be detected.

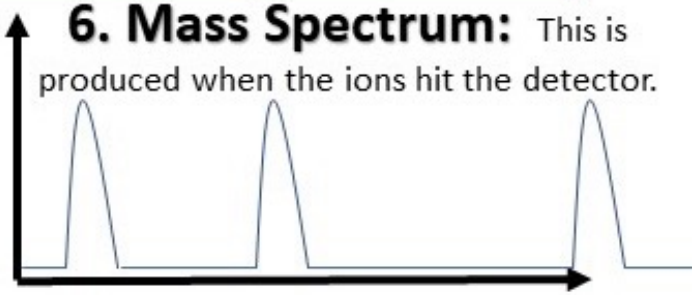


1. Vapourisation:

Samples of unknown substances (elements or compounds) are injected into a vacuum.

4. Ion Drift: They then drift with ***no*** electric field at the same speed they were travelling.

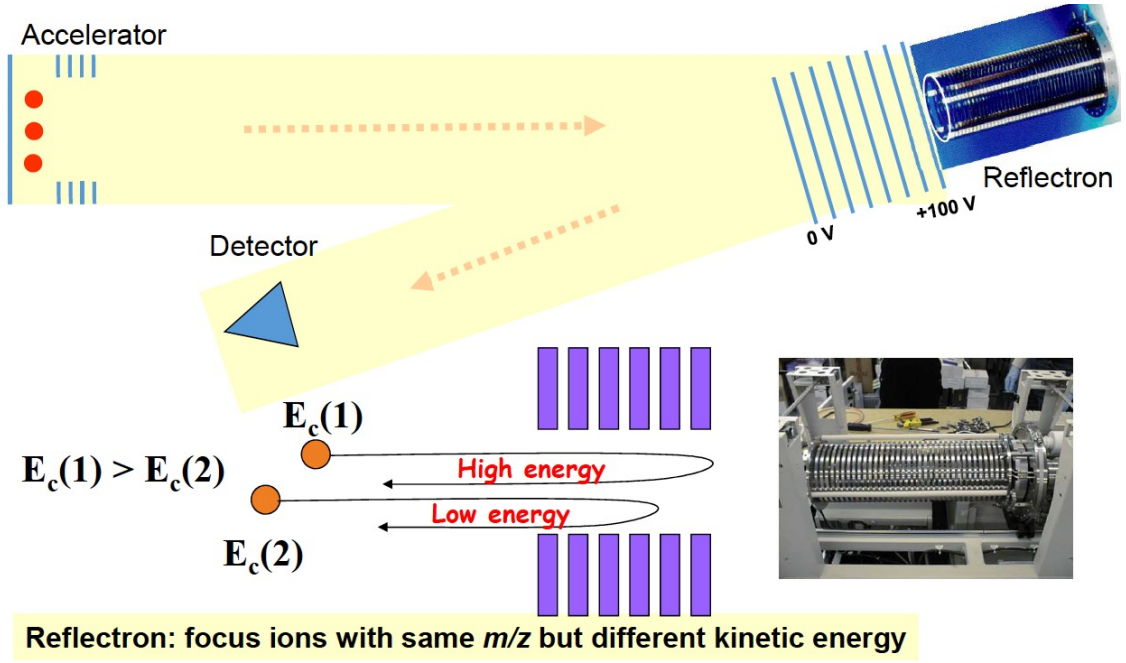
6. Mass Spectrum: This is produced when the ions hit the detector.



https://www.tes.com/teaching-resource/mass-spectrometer-time-of-flight-12386261

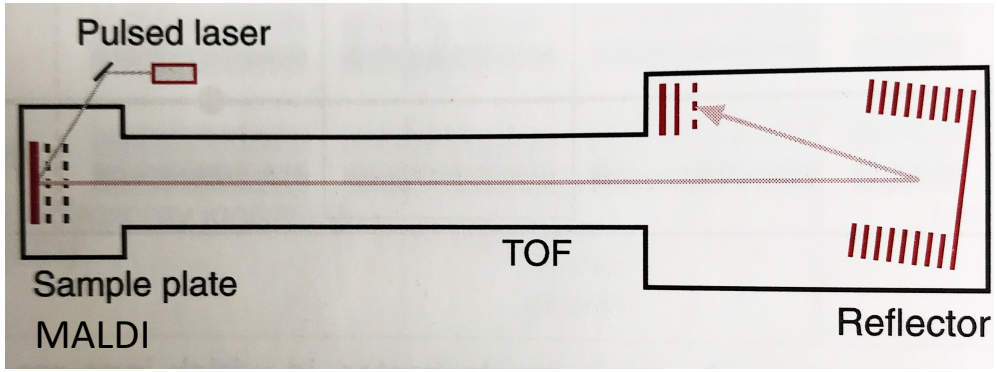
TOF

- Increase of resolution



Reflectron: focus ions with same *m/z* but different kinetic energy

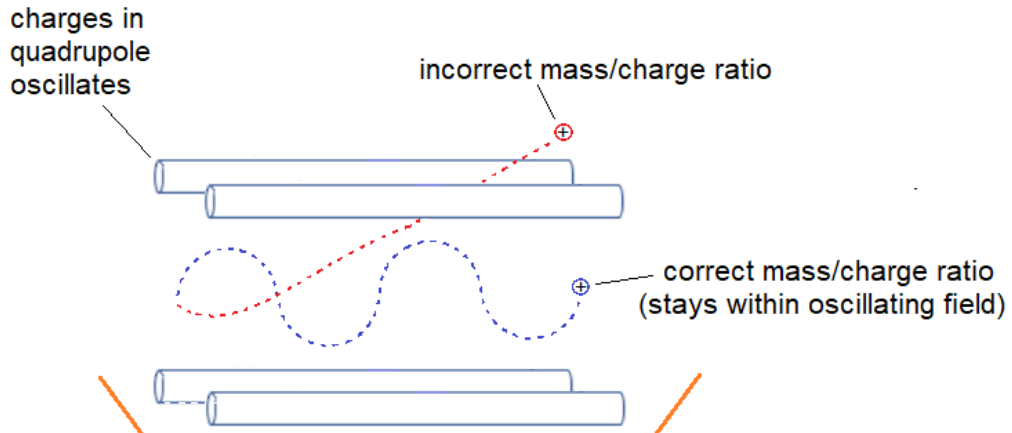
- **MALDI-TOF**
  - MALDI produces singly charged ions
  - The time of flight of any ions in TOF is inversely proportional to the square foot of the molecular mass.



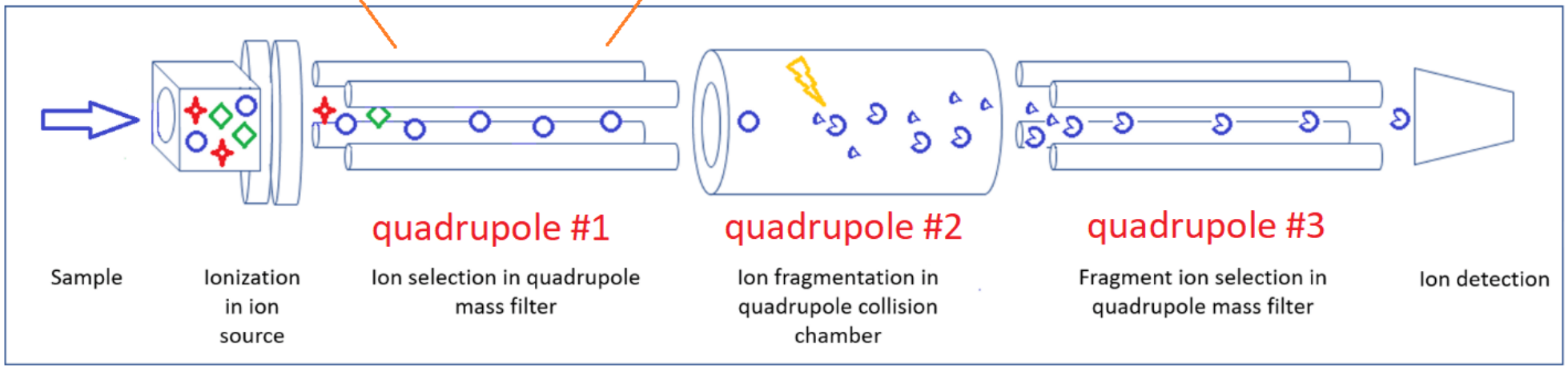
  - MALDI-tandem TOF or MALDI-hybrid Quadrupole-TOF analyzers are more sensitive for high-throughput analysis.

Quadrupole

Video on how triple quadrupole works:
https://www.youtube.com/watch?v=og2DUvF29zI

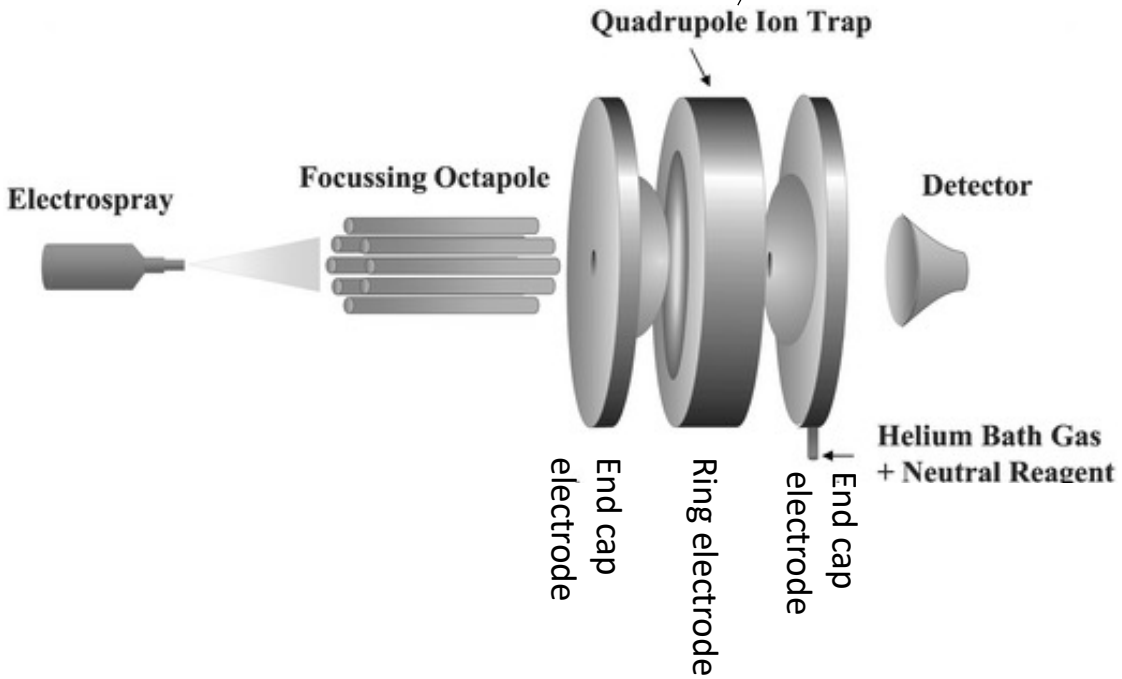


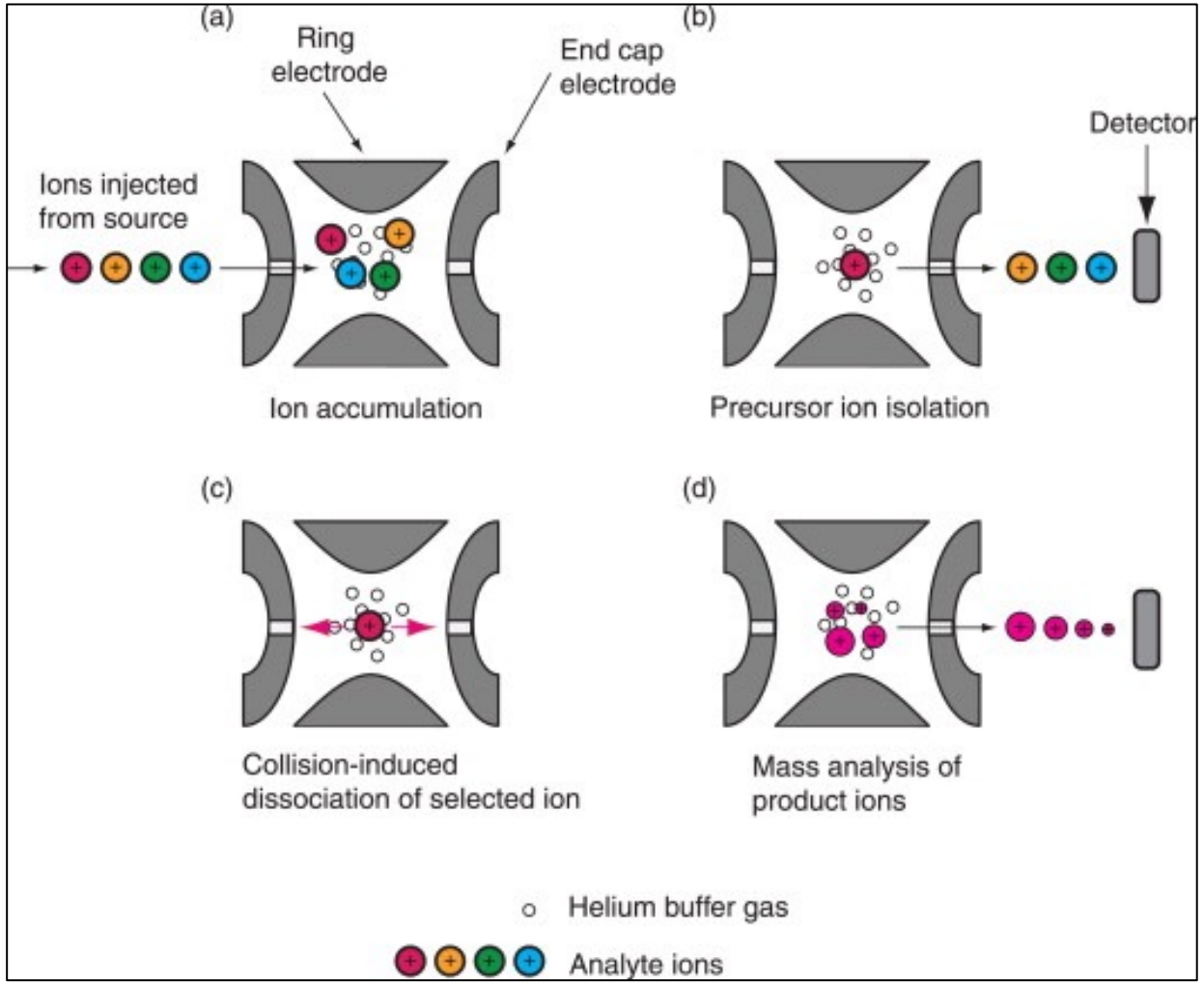
Triple quadrupole



Ion Trap

Collects and stores ions in order to perform MS/MS analyses on them.





- To monitor the ions coming from the source, the trap continuously repeats a cycle of filling the trap with ions and scanning the ions according to their m/z values.

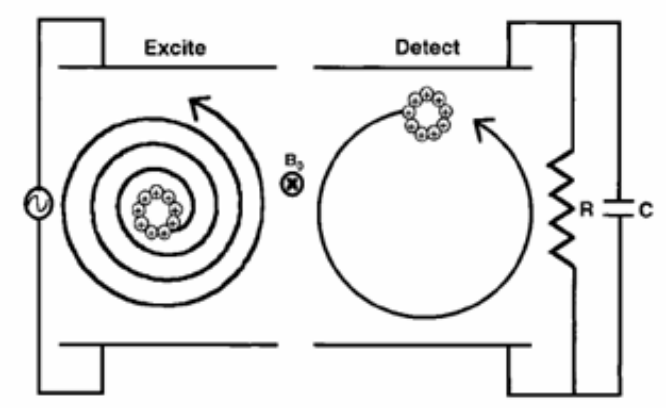
https://pubs.rsc.org/image/article/2006/cc/b516348j/b516348j-f1.gif

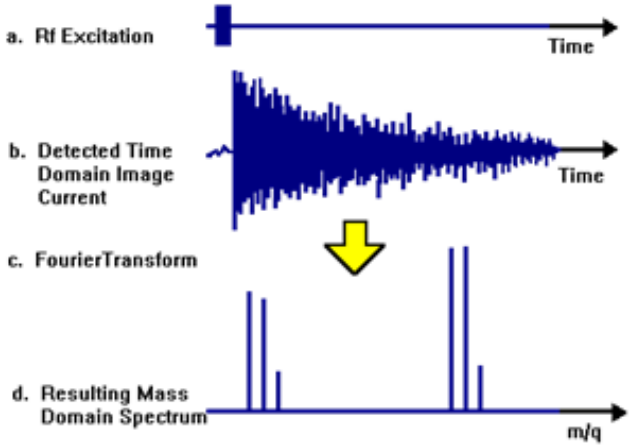
Fourier Transform Ion Cyclotron Resonance

- A mass analyzer for determining the mass-to-charge ratio (m/z) of ions based on the cyclotron frequency of the ions in a fixed magnetic field.

Ions are injected into a magnetic field , that causes them to travel in circular paths.

Excitation with oscillating electrical field increases the radius and enables a frequency measurement



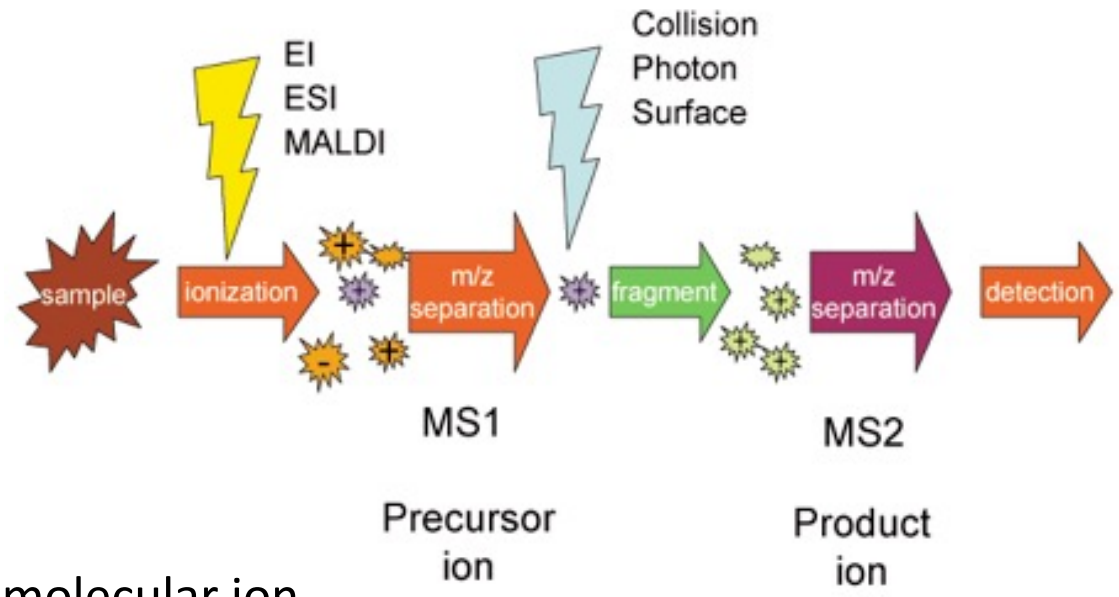


Video: https://jlab.chem.yale.edu/research/techniques/fourier-transform-ion-cyclotron-resonance-mass-spectrometry

- High resolution
- High accuracy
- Very sensitive (the minimal quantity for detection is in the order of several hundred ions)
- Non destructive – the ions don't hit the detection plate so they can be selected for further fragmentation

MS/MS (Tandem MS) Terminology

- Tandem mass spectrometry, also known as MS/MS or MS^2, is a technique in instrumental analysis where two or more mass analyzers are coupled together using an additional reaction step to increase their abilities to analyze samples.



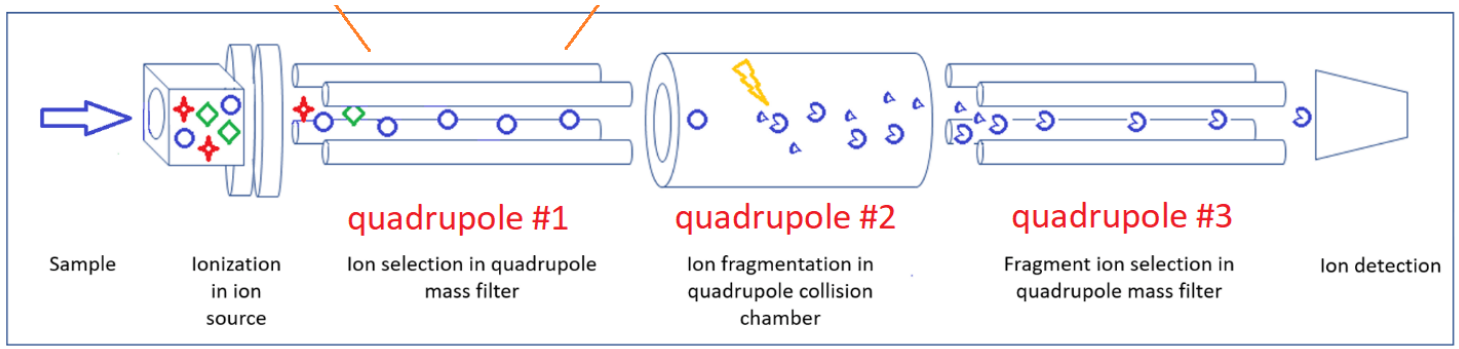
- Molecular ion / precursor ion (parent ion)

  Ion formed by ionization of the analyte species

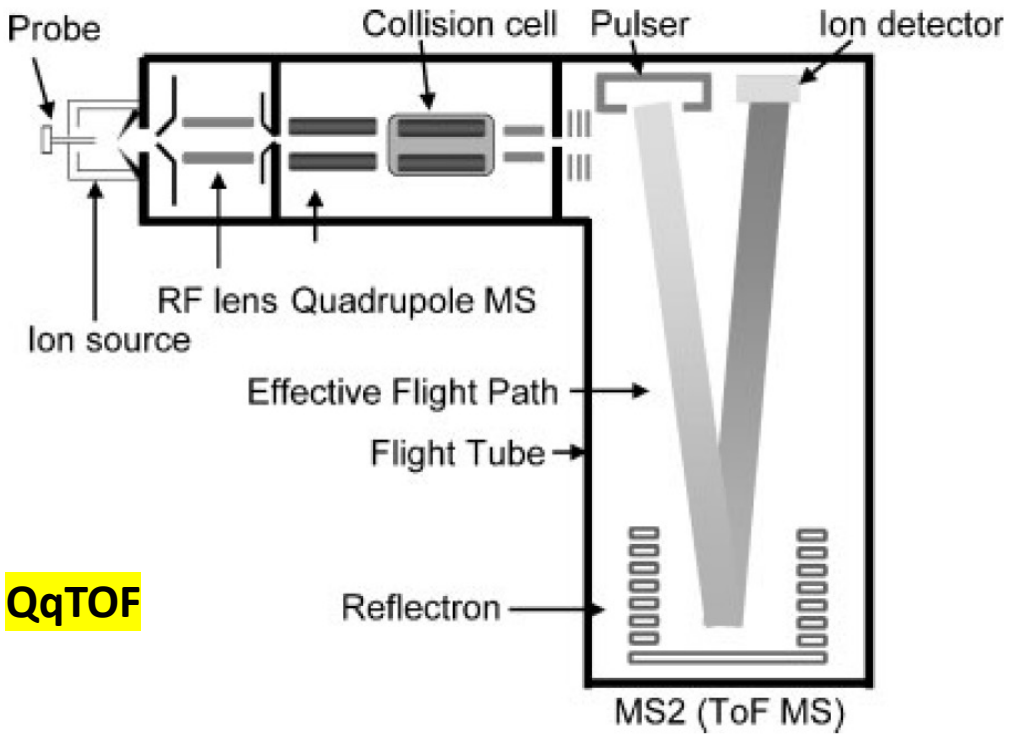
- Fragment ions / product ions (daughter ions)

  Ions formed by the gas-phase fragmentation of the molecular ion

https://nationalmaglab.org/images/users/icr/techniques/tandem_scheme.png



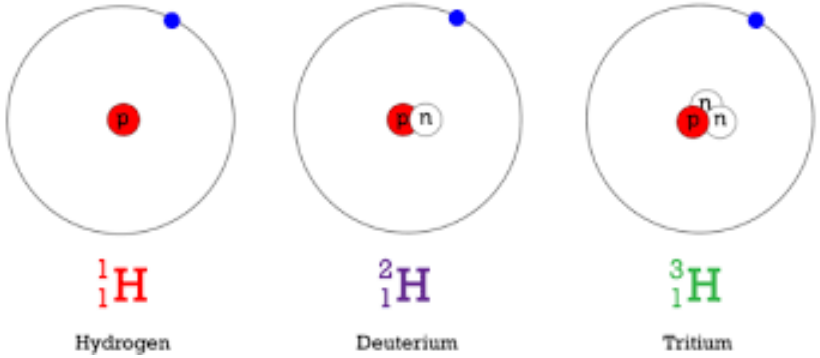
Triple Quadrupole (QqQ)

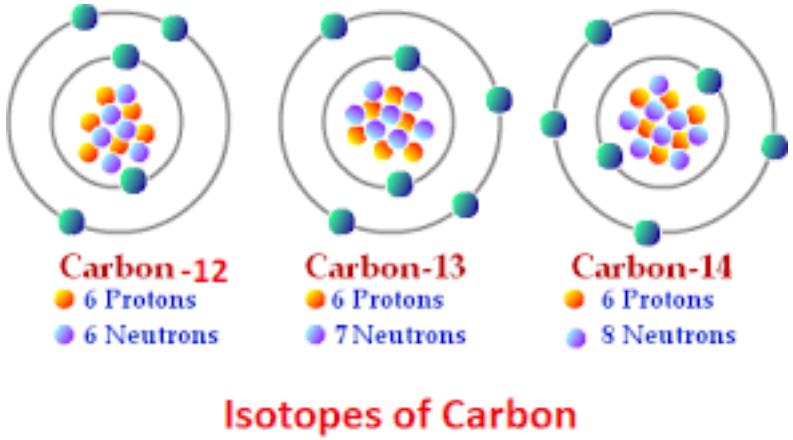


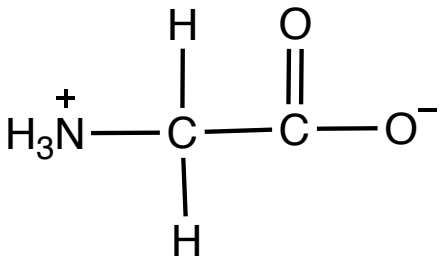
QqTOF

How to Calculate Mass?

- Elements and their isotopes have unique masses.







Glycine

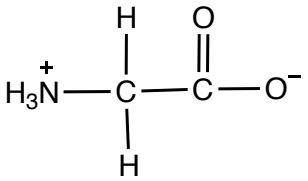
For Glycine ($C_2H_5NO_2$), the nominal mass is 12x2+1x5+14x1=75

If one of the carbons is ^{13}C, the nominal mass of Gly is 12x1+13x1+1x5+14x1=76

How to Calculate Mass?

- Elements and their isotopes have unique masses.
- **Monoisotopic Mass**: *Exact mass* of an ion or molecule *calculated* using the mass of the most abundant isotope of each element.
- **Average Mass**: Mass of an ion or molecule weighted for its isotopic composition.
- **Accurate Mass**: *Experimentally* determined mass of an ion of known charge

For Glycine, ($C_2H_5NO_2$),



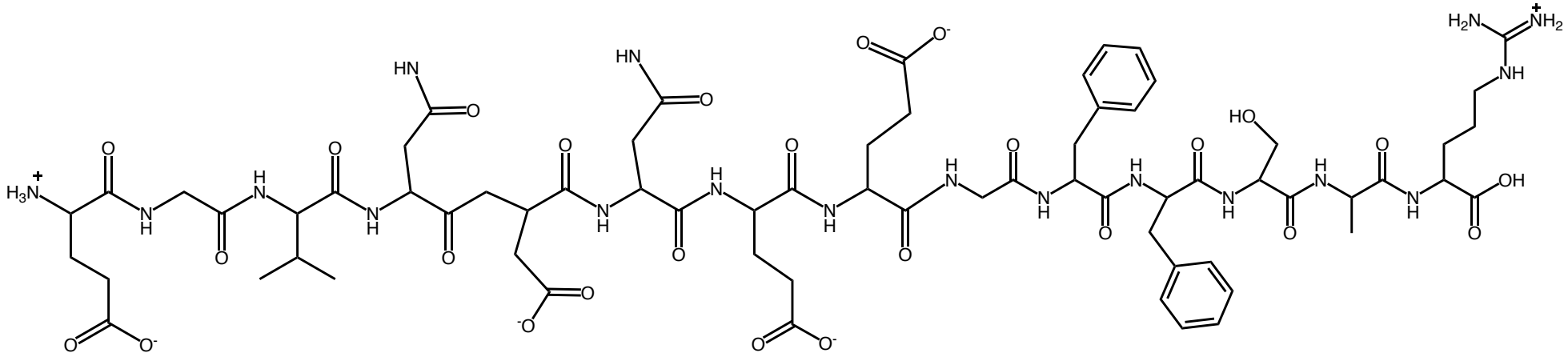
ATOM	Nominal Mass	Average Mass	Monoisotopic Mass
C	12 x 2 = 24	12.0011 x 2 = 24.022	12.0000 x 2 = 24.0000
H	1 x 5 = 5	1.008 x 5 = 5.04	1.0078 x 5 = 5.039
O	16 x 2 = 32	15.999 x 2 = 31.998	15.9949 x 2 = 31.9898
N	14 x 1 = 14	14.0067 x 1 = 14.0067	14.0031 x 1 = 14.0031
S	32 x 0 = 0	32.065 x 0 = 0	31.9720 x 0 = 0
	75	75.0667	75.0319
	Low Resolution	Chemistry Calculations	High Resolution

Recognizing Multiple Charged Ions

- Mass spectrometers operate on the basis of mass-to-charge ratio (*m/z*, unit: Thomson or Th). In other words, the analytes (peptides) need to be charged (ions).

$$m/z = \frac{m+z}{z}$$

- Single charge (MALDI) $m/z = (M+H^+)$
- Double charge (ESI) $m/z = (M+2H^+)/2$
- n charge (ESI) $m/z = (M+nH^+)/n$

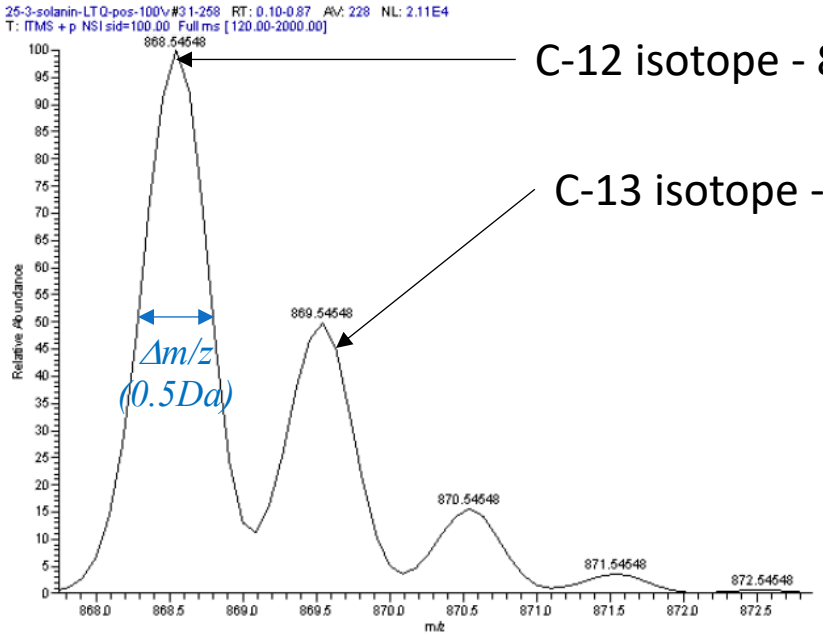


Mass Resolution

$$\textbf{Resolution} = \frac{m/z}{\Delta m/z}$$

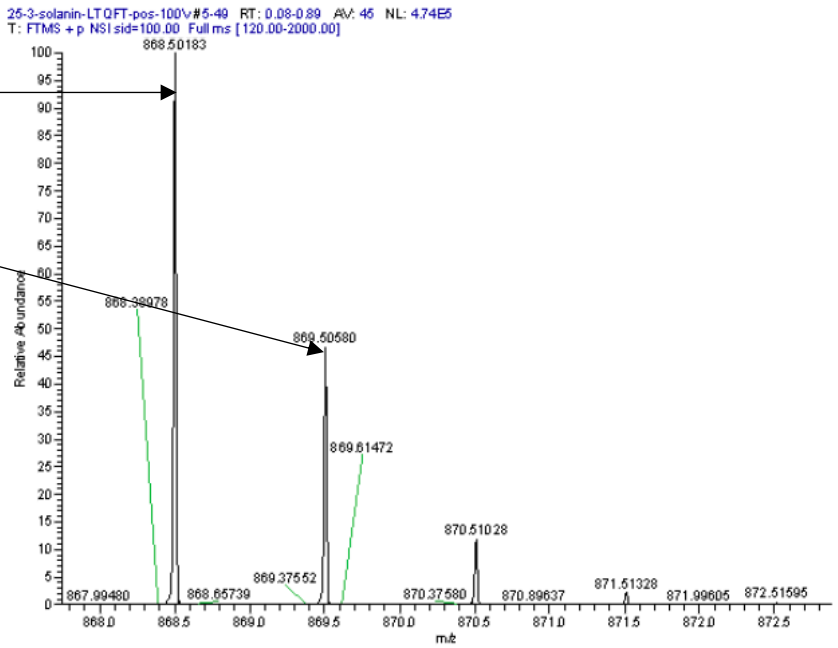
full width at half maximum

- The ability of the instrument to resolve two closely placed peaks.



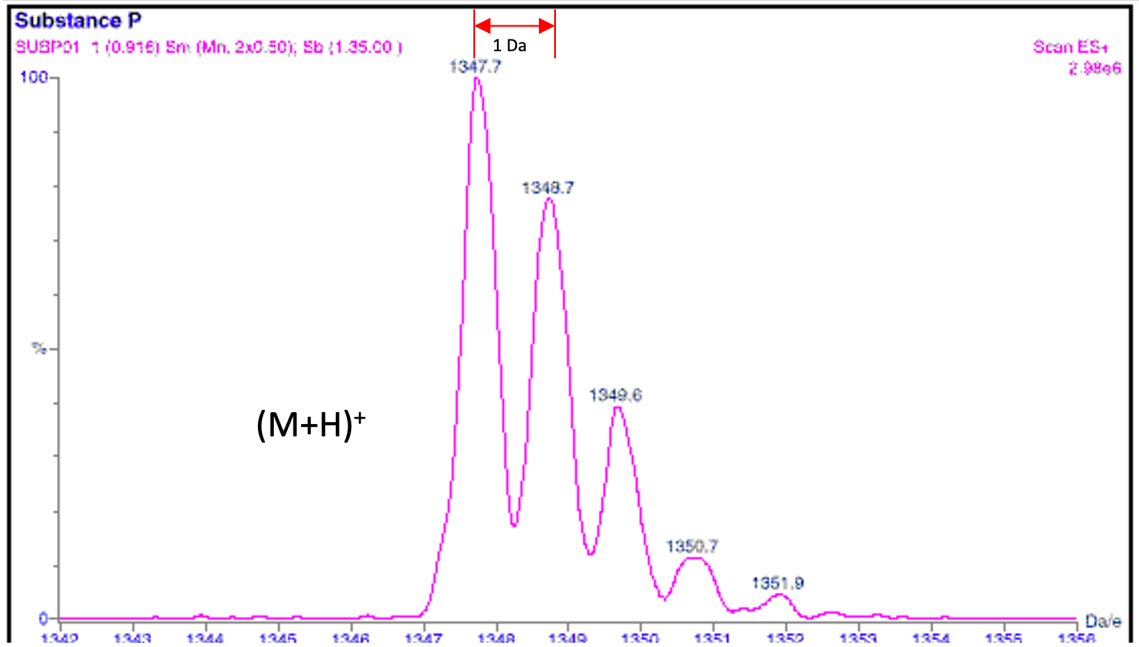
C-12 isotope - 868.5 Da

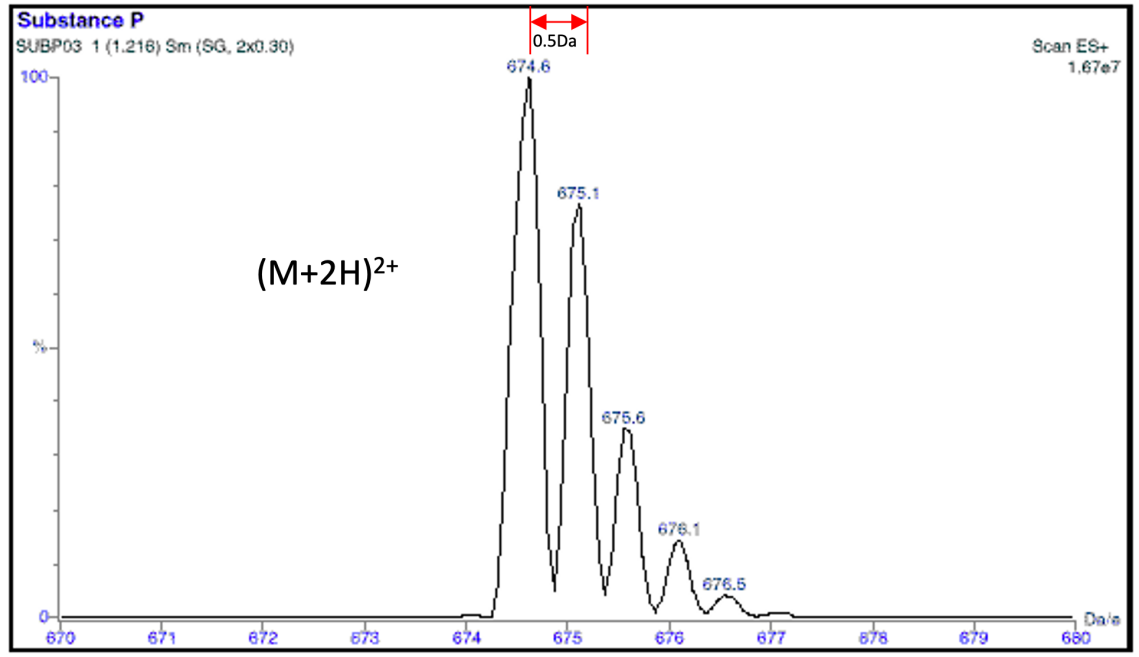
C-13 isotope - 869.5 Da



low resolution, Resolution =1737 (868.5/0.5)

high resolution, resolution = 48,250 (868.5/0.018)
m/z 868.5, peak width ~0.018



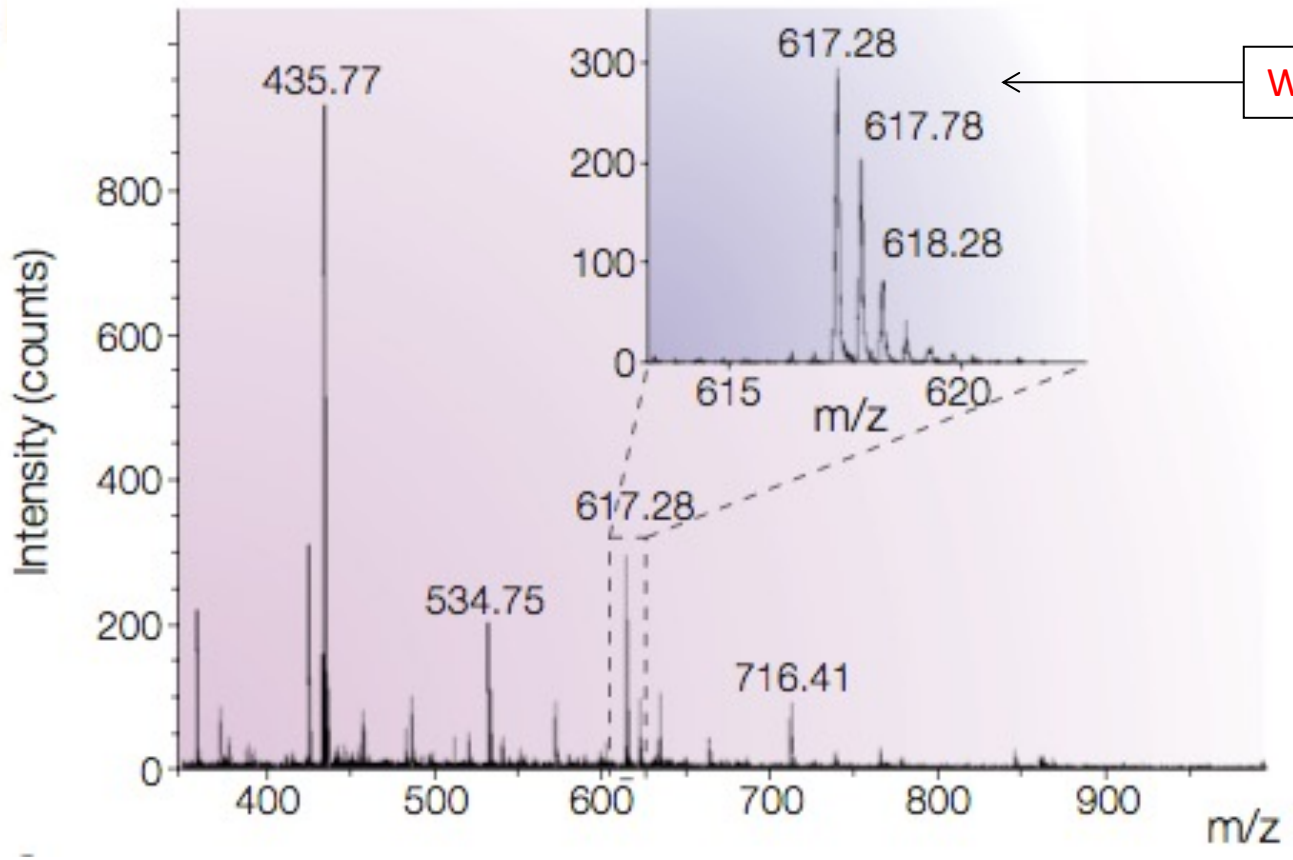


$$m/z=\frac{m+z}{z}$$

Isotopes of singly charged ions are separated by 1 Da

Isotopes of doubly charged ions are separated by 0.5 Da

MS Spectrum (i.e., peptide ions)



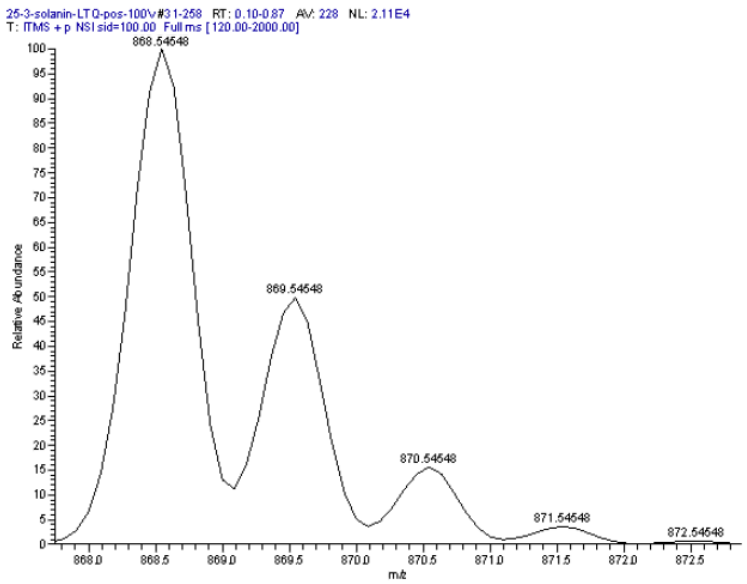
What is the charge state of this peptide?

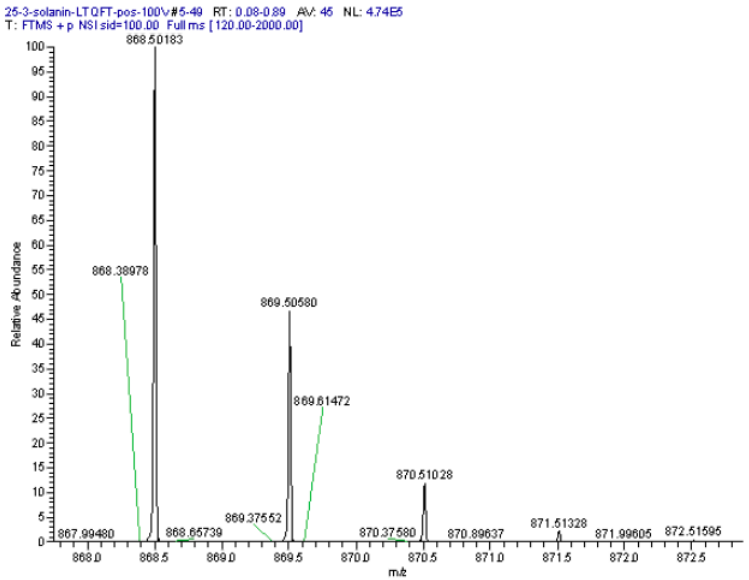
Each peak is a different peptide, separated based on m/z
A single peptide is selected by the instrument for the second MS

Mass Accuracy

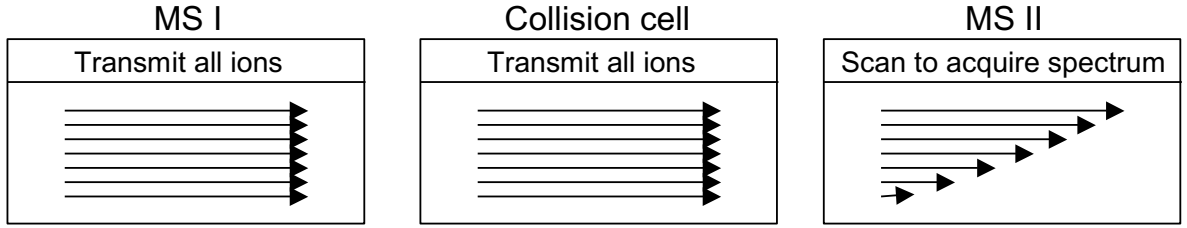
$$\frac{M_{\exp} - M_{true}}{M_{true}} \times 10^6$$

- The relative percent difference between the measured mass and the true mass (usually represented in ppm).
- The lower the number the better the mass accuracy



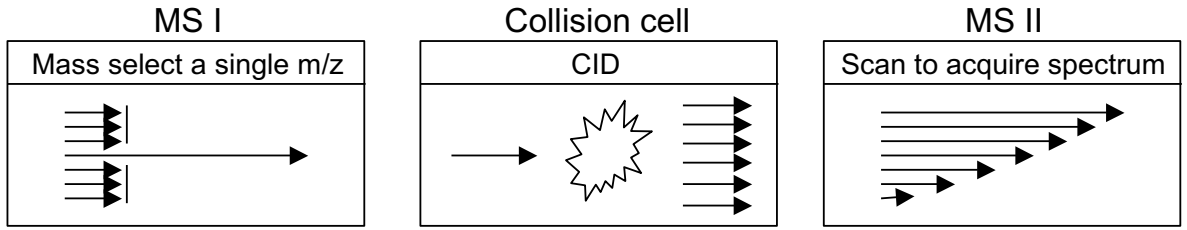


A. Mass spectrum scan



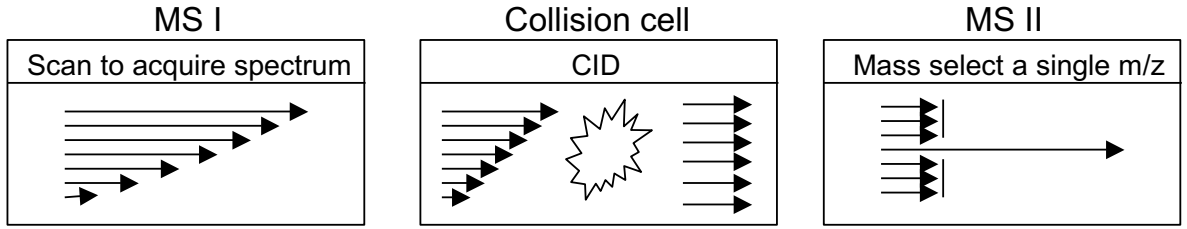
Protein ID, DDA-Data Dependent Analysis.

B. Product ion scan



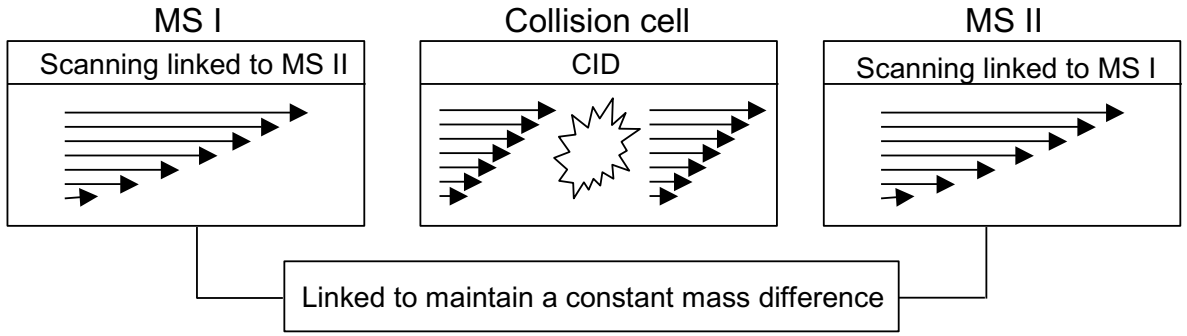
A scan which determines, in a single experiment, all the product ion m/z that are produced by the reaction of a selected precursor ion.

C. Precursor ion scan



A scan which determines, in a single experiment, all the precursor ion m/z that react to produce a selected product ion m/z (sometimes called a ‘precursor ion scan’).

D. Neutral loss scan



Protein post-translational modification analysis.

Protein Identification Using Data from MS

1. Peptide mass fingerprinting (aka. Peptide mapping)

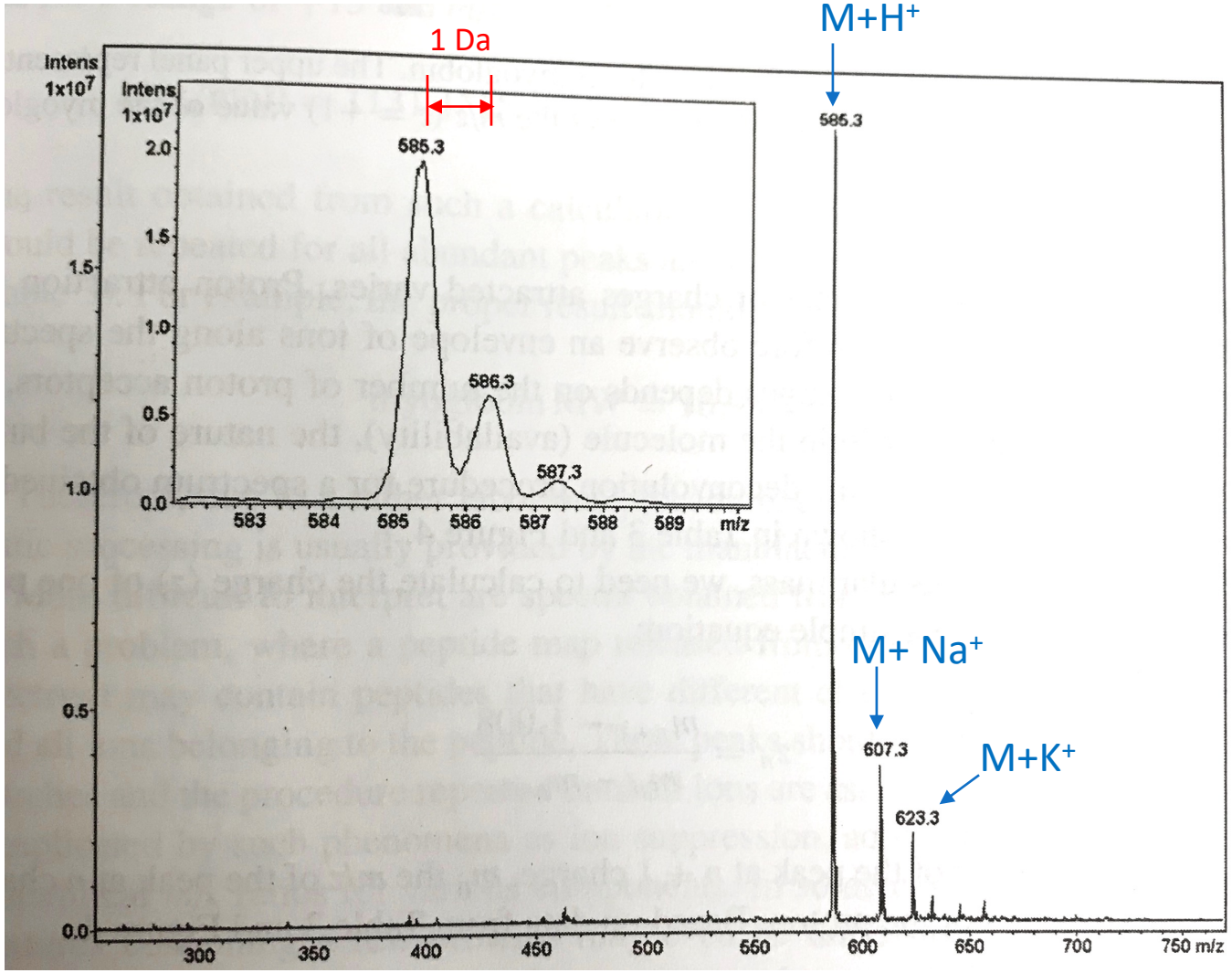
2. Fragment ions analysis

   Determine protein sequence *de novo* by MS

1. Peptide Mass Fingerprinting (PMF)

- Principle: *each protein can be uniquely identified by the masses of its constituent peptides.*

- A single protein or a simple mixture
  e.g., a spot on 2D gel or a single LC fraction
- The sample is digested by 1 or more specific cleavage reagent (e.g., trypsin)
- The masses of the peptides are determined, usually by **MALDI-TOF**.
- Search databases for correlative searching
- The algorithm carries out a virtual digest of each protein in the databases with the same specific cleavage and calculate the theoretical peptide masses
- The algorithm attempts to correlate the theoretical masses with the experimentally determined ones. Then Rank proteins from the database in order of best correlation (number of peptides matched).

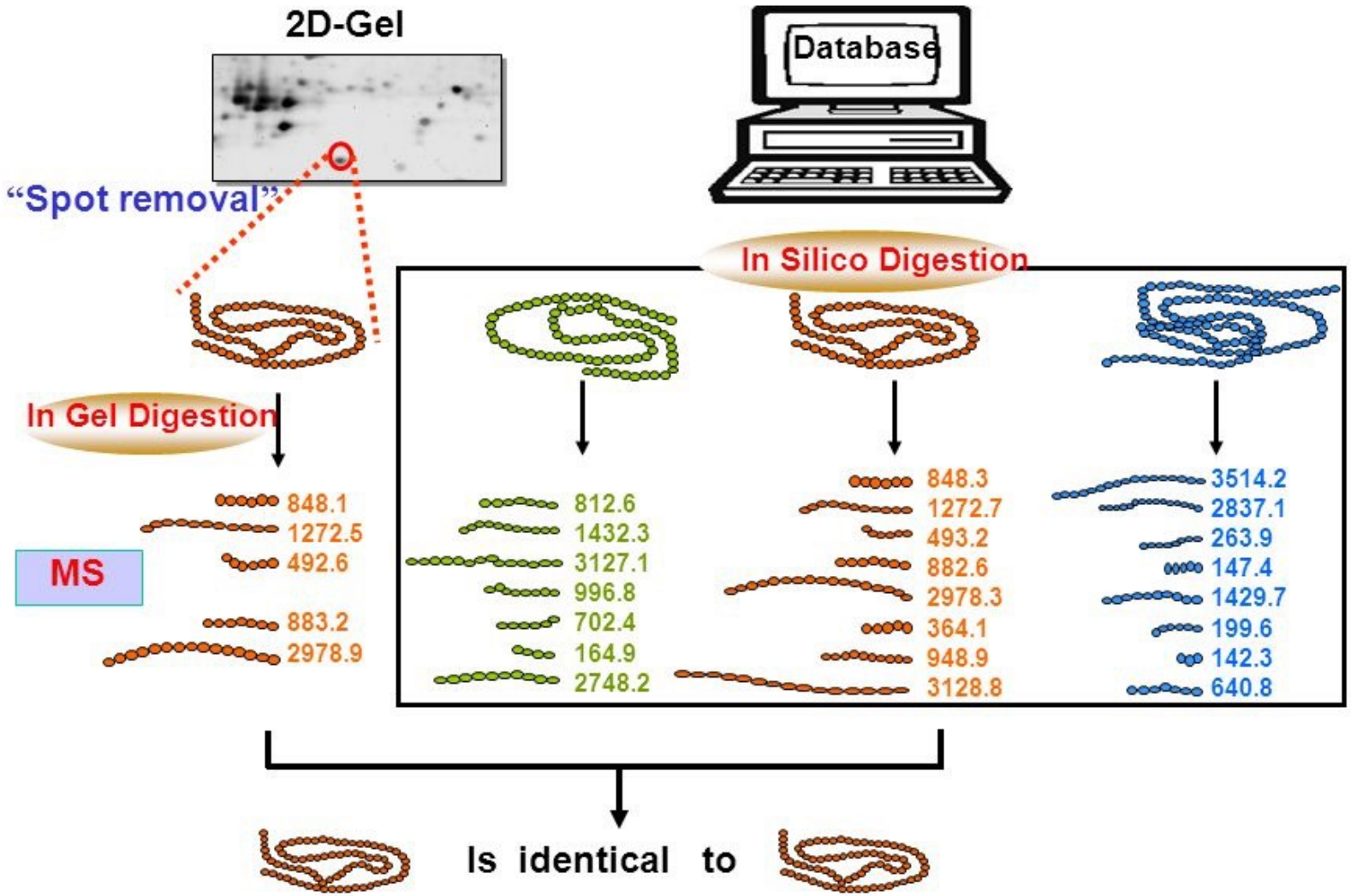


Mass spectrum of the peptide FGGFTG.
The MW of the peptide is 584.26.

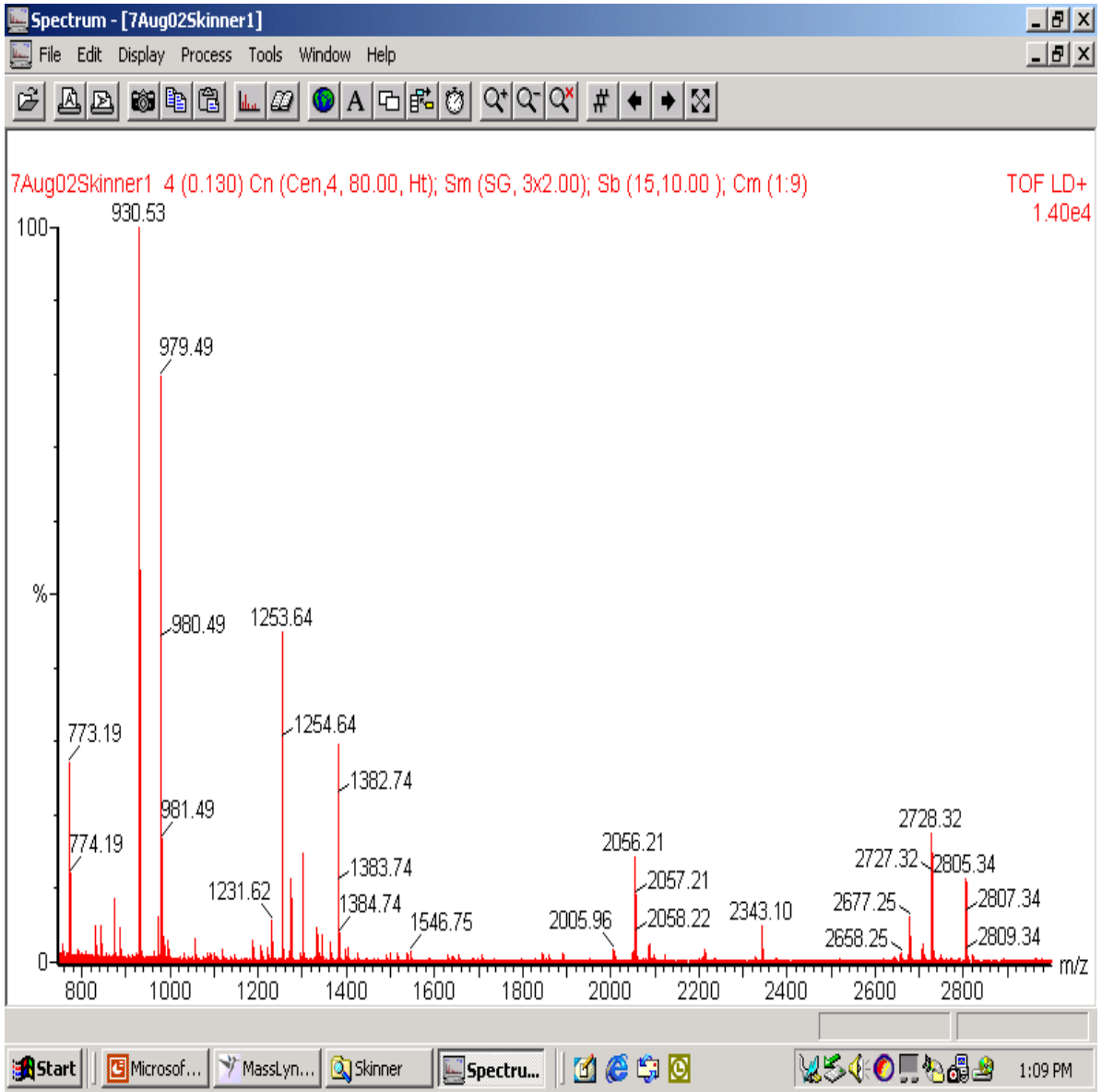
- Proteins and peptides are generally analyzed in the positively mode, which favors production of protonated ($M+H^+$) ions
- In the presence of some buffer ions, sodium ($M+ Na^+$), potassium ($M+K^+$), and ammonium ($M+ NH_4^+$) adducts may be formed.

Website for peptide mass calculation:
https://web.expasy.org/peptide_mass/

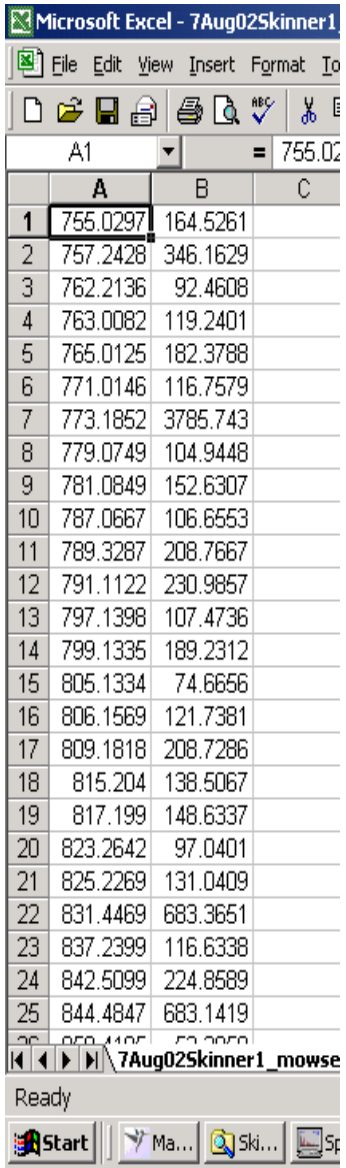
Peptide Mass Fingerprinting



MALDI-MS Database Searching

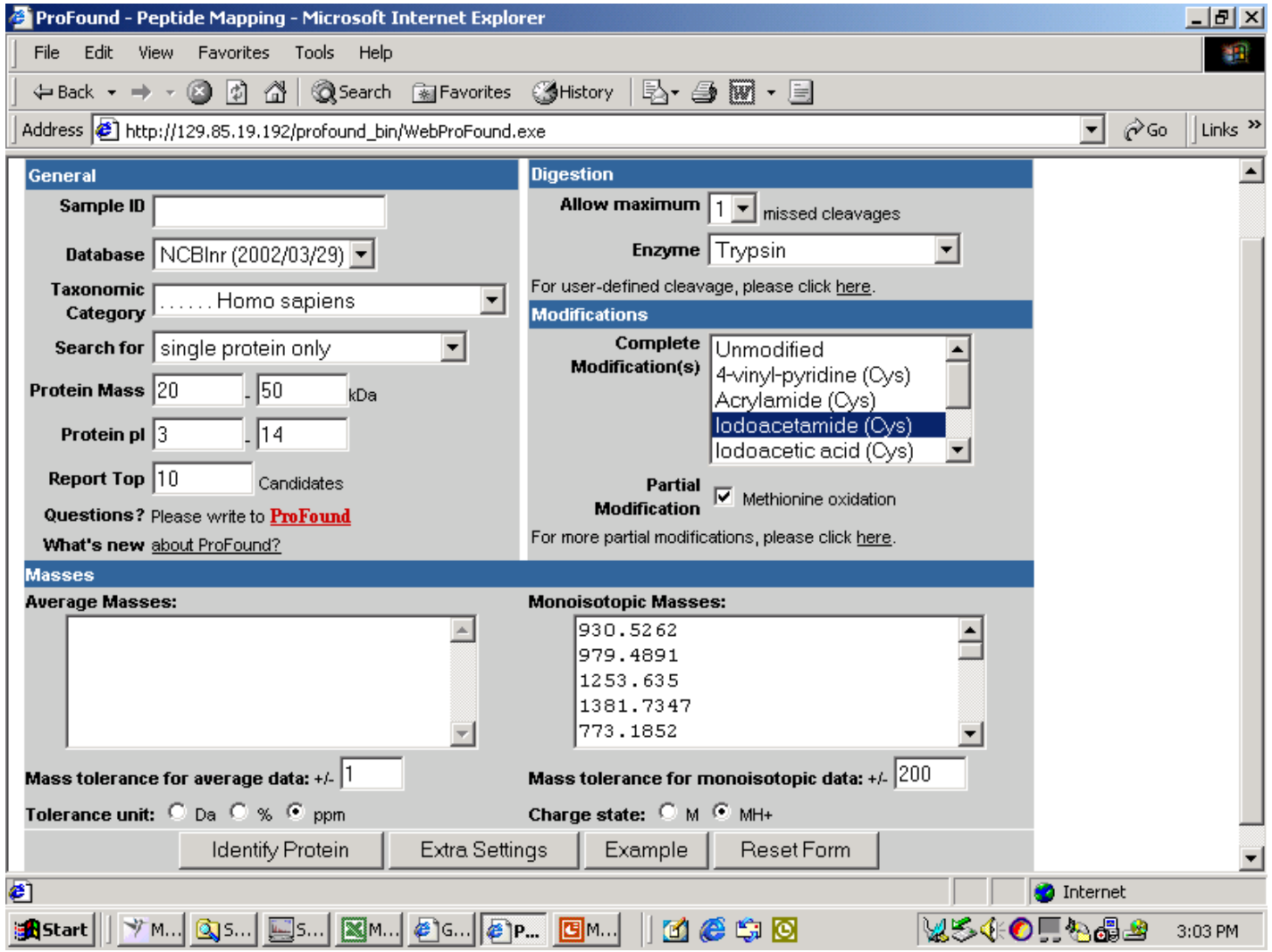


Peaklist



	A	B	C
1	755.0297	164.5261	
2	757.2428	346.1629	
3	762.2136	92.4608	
4	763.0082	119.2401	
5	765.0125	182.3788	
6	771.0146	116.7579	
7	773.1852	3785.743	
8	779.0749	104.9448	
9	781.0849	152.6307	
10	787.0667	106.6553	
11	789.3287	208.7667	
12	791.1122	230.9857	
13	797.1398	107.4736	
14	799.1335	189.2312	
15	805.1334	74.6656	
16	806.1569	121.7381	
17	809.1818	208.7286	
18	815.204	138.5067	
19	817.199	148.6337	
20	823.2642	97.0401	
21	825.2269	131.0409	
22	831.4469	683.3651	
23	837.2399	116.6338	
24	842.5099	224.8589	
25	844.4847	683.1419	

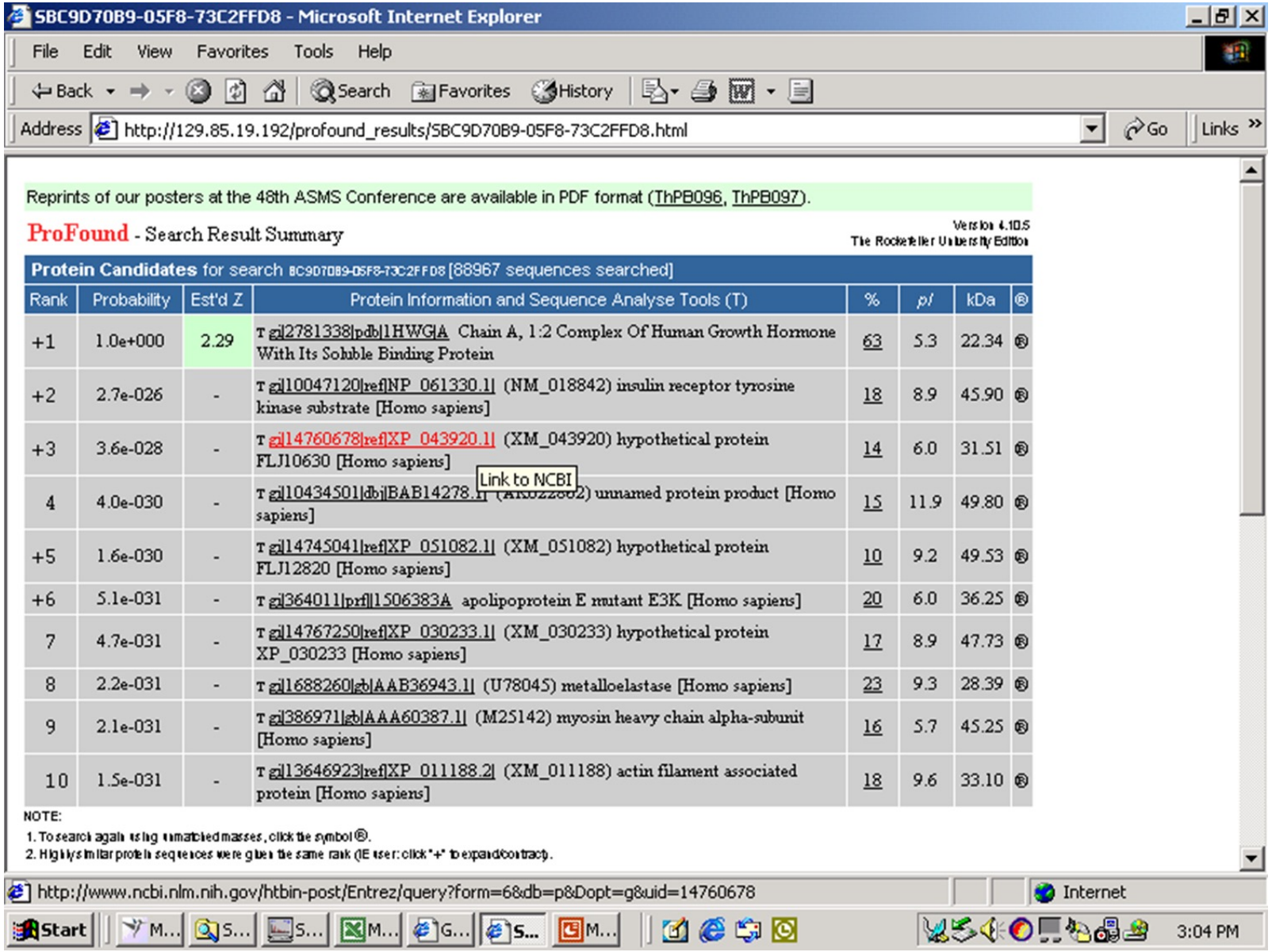
Web-based ProFound Search Engine



The majority of the available search engines allow one to define certain experimental parameters to optimize a particular search.

- Minimum number of peptides to be matched
- Allowable mass error
- Monoisotopic versus average mass data
- Mass range of starting protein
- Type of protease used for digestion
- Information about potential protein modification, such as N- and C-terminal modification, carboxymethylation, oxidized methionine, etc.

ProFound Search Result



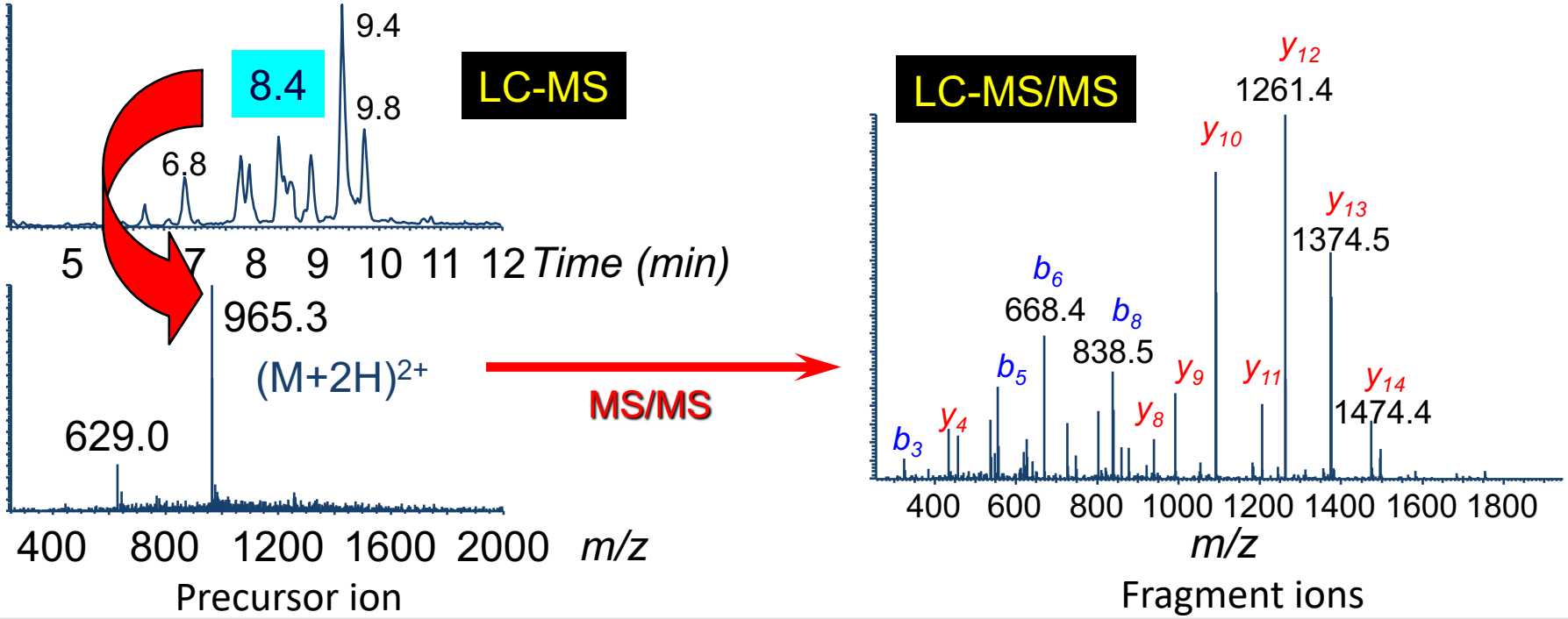
Searching with Peptide Mass Fingerprints (PMF): Limitations

- Most protein databases contain primary sequence information only
  - Any shift in mass incorporated into the primary sequence as a result of post-translational modification, amino acids substitution will result in an experimental mass that is in *disagreement with the theoretical mass,* even a protein with a great deal of homology in the database can not be identified.

- Non-specific cleavage, isobaric peptide (same mass but different order), etc.

LC-MS/MS for Protein Identification

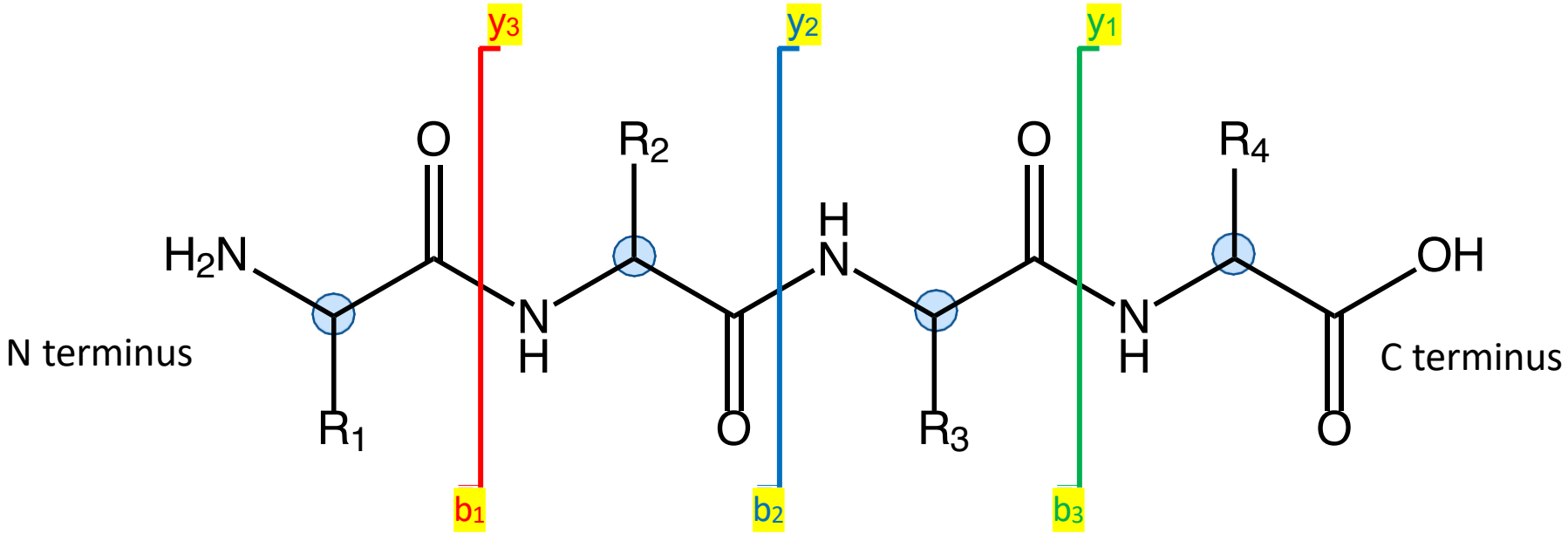
- An improvement in throughput of the overall method can be obtained by performing LC-MS/MS in the data-dependent mode.
  – As full scan mass spectra are acquired continuously in LC-MS mode, any ion detected with a signal intensity above a pre-defined threshold will trigger the mass spectrometer to switch over to MS/MS mode. Thus, the mass spectrometer switches back and forth between MS (molecular mass information) and MS/MS mode (sequence information) in a single LC run.



The data-dependent scanning capability can dramatically increase the capacity and throughput for protein identification.

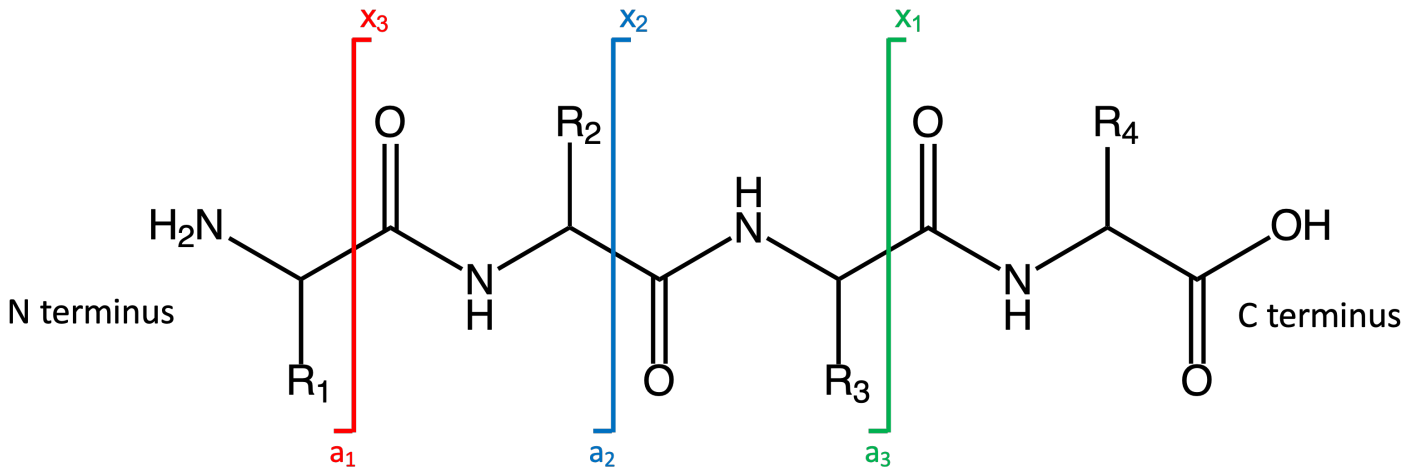
2. Fragment Ion Analysis$_1$

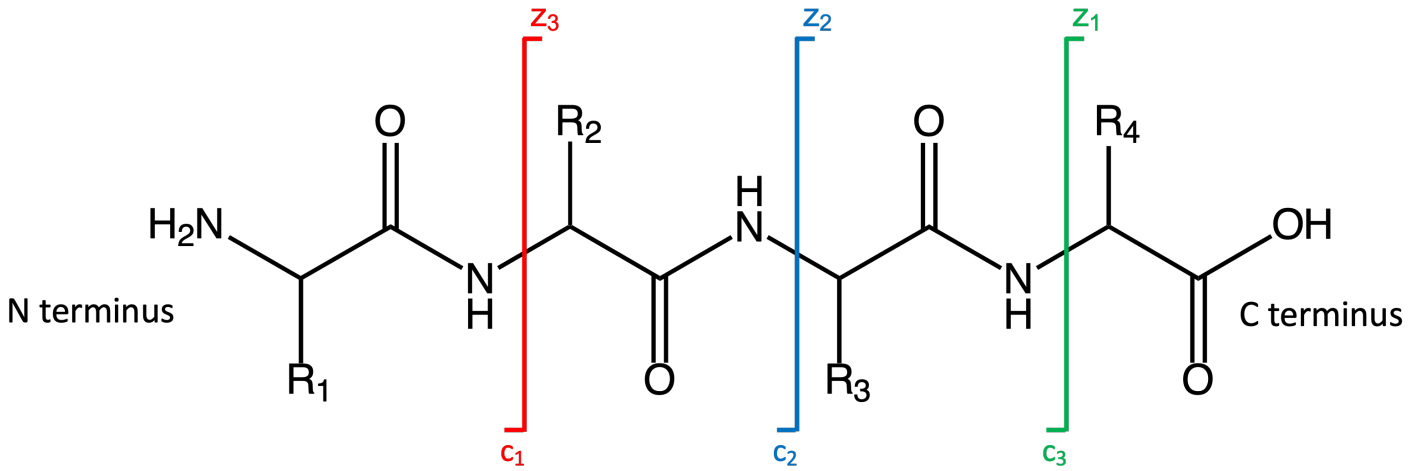
- Peptide can be fragmented by collision-induced dissociation (CID) (and other methods)
  - Collisions with neutral inert gas molecules (nitrogen, argon, etc.)
  - Charge stays on *either* the 'left' (a, b, or c) or 'right' (x, y, or z) side of cleavage
  - *Cleavage along the **CO-NH bond** is most common, generating 'b' and 'y' ions*



- Letter: Indicates the bond broken and the terminus contained in the fragment
- Number: Indicates the number of Cα in the fragment

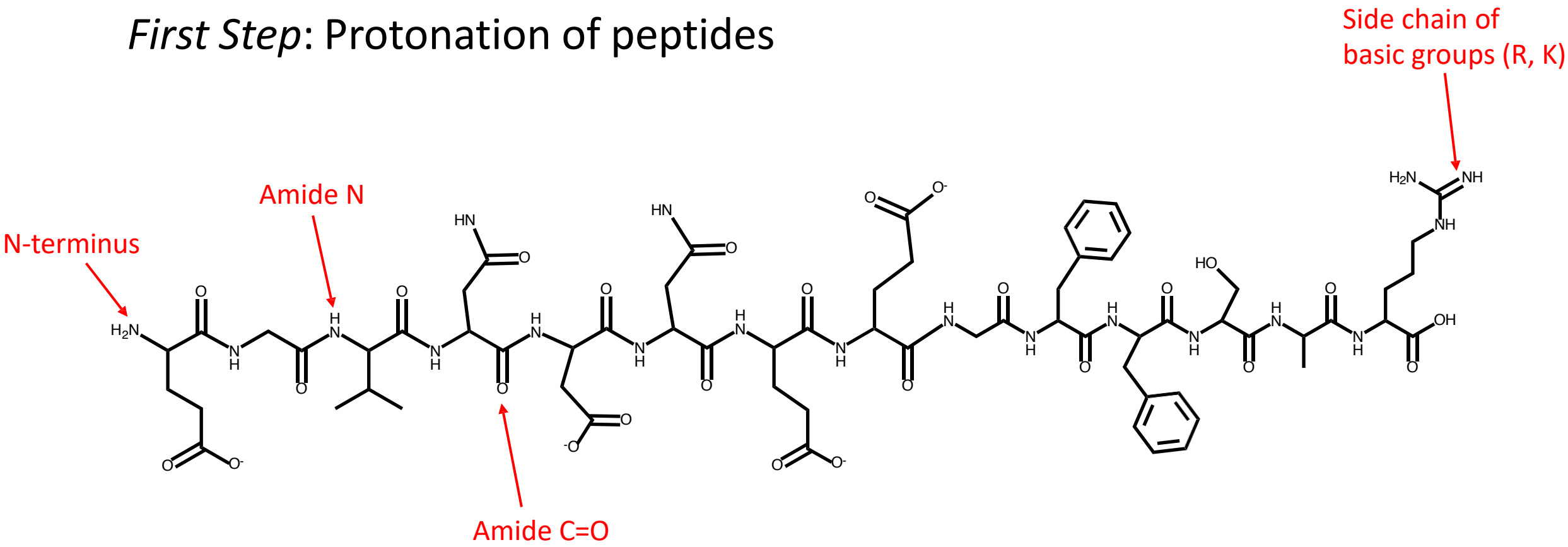
2. Fragment Ion Analysis$_2$





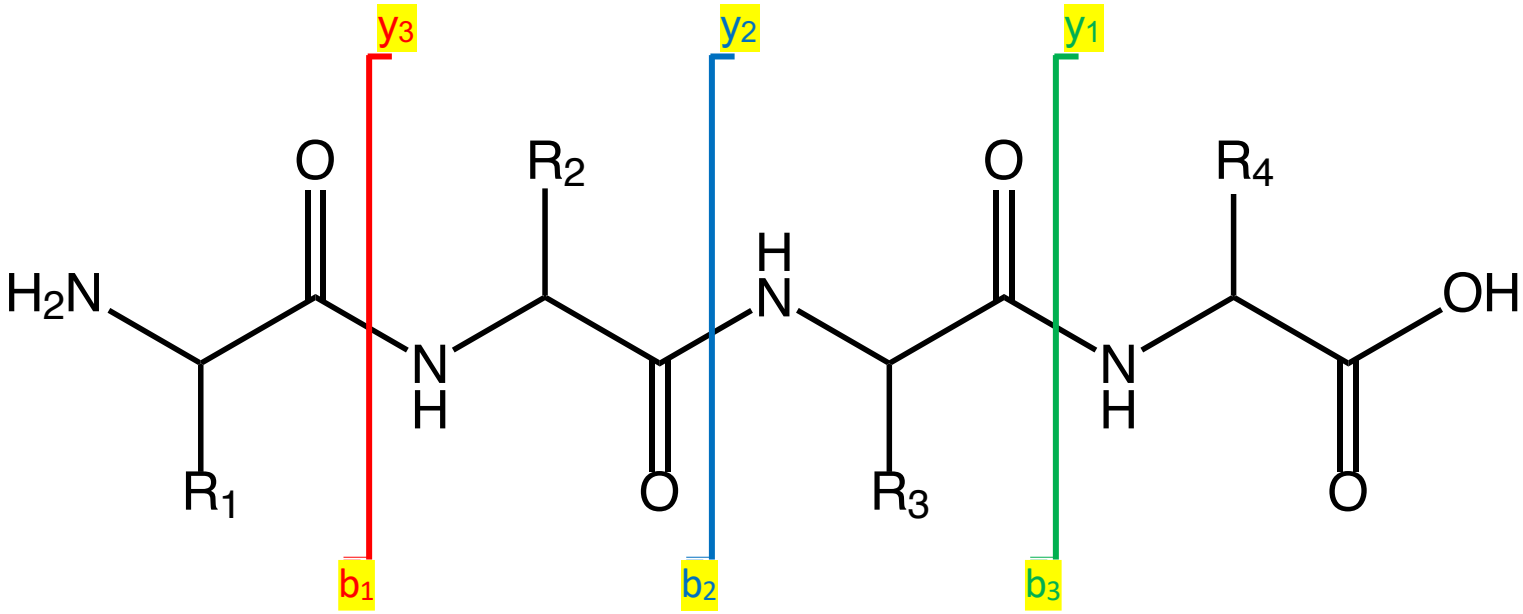
Peptides Fragment by CID

First Step: Protonation of peptides

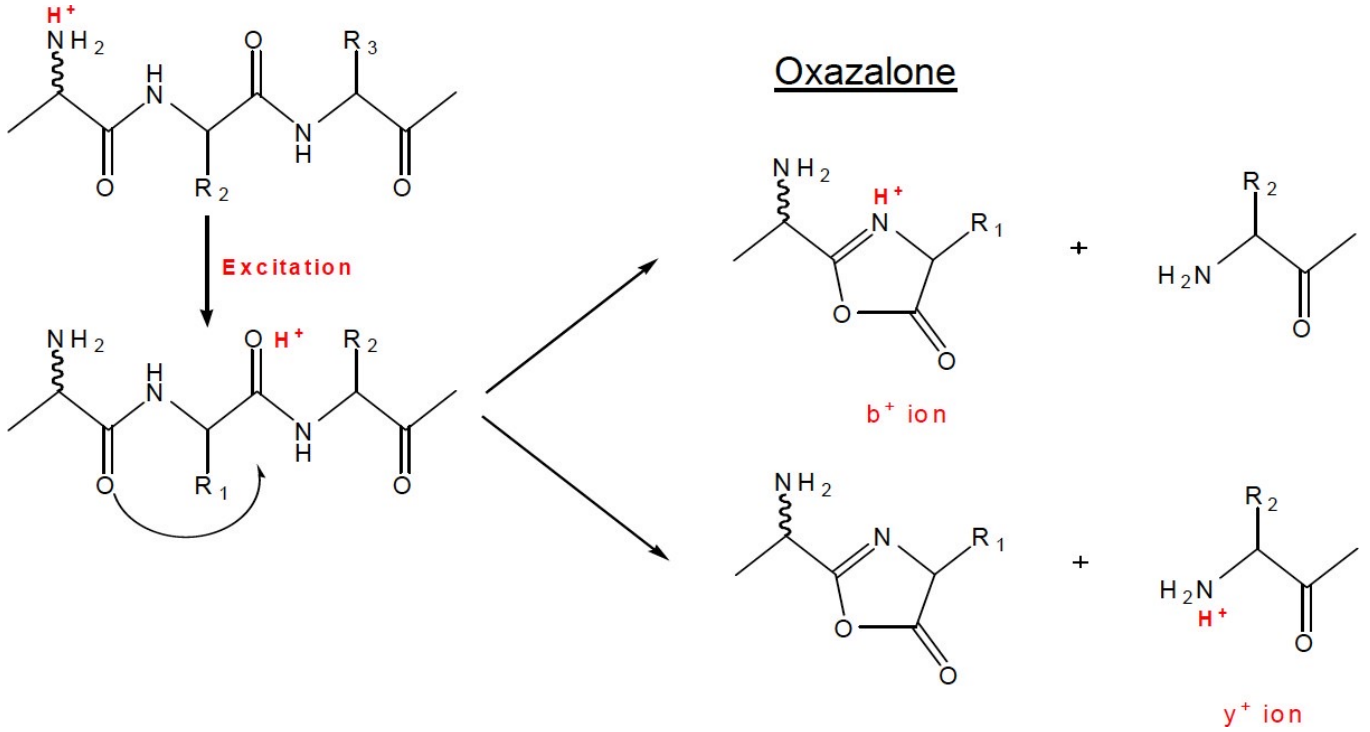


Peptides Fragment by CID

Second step: Cleavage along the CO-NH bond is most common, generating **b** and **y** ions



Peptides Fragment by CID



For a singly protonated peptide,

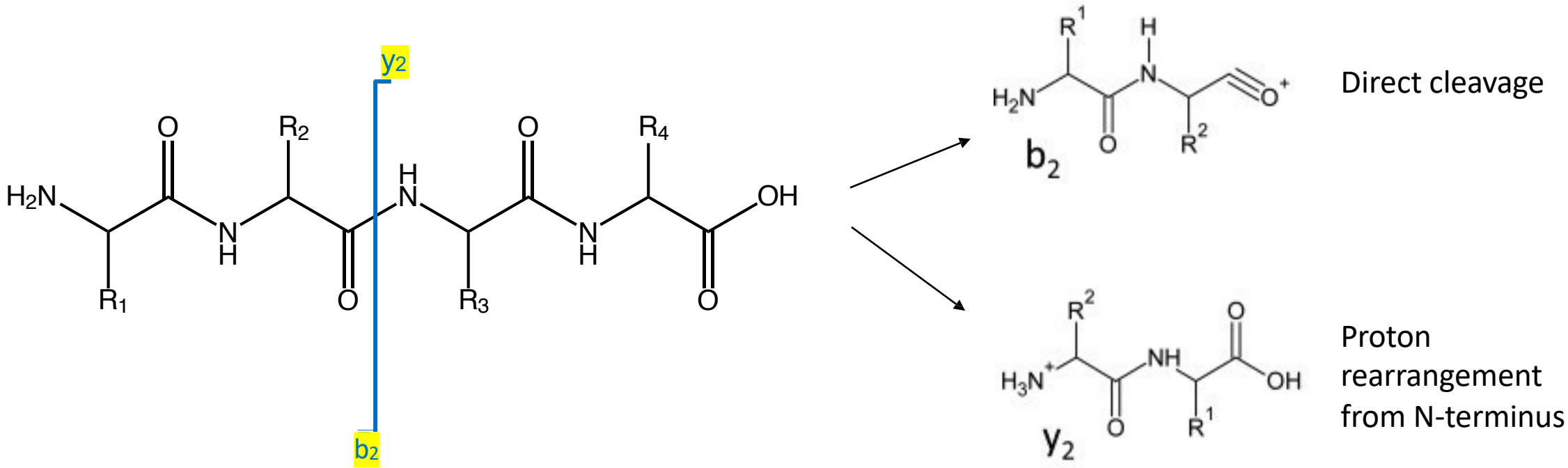
Singly charged N term ion ($+H^+$) and neutral C-term

OR

Neutral N term and Singly charged C-term ion ($+H^+$)

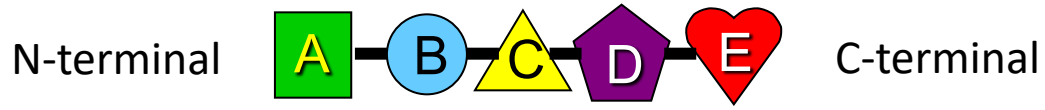
From Wysocki et al., J. Mass Spectrom. (2000) 35:1399

Peptides Fragment by CID

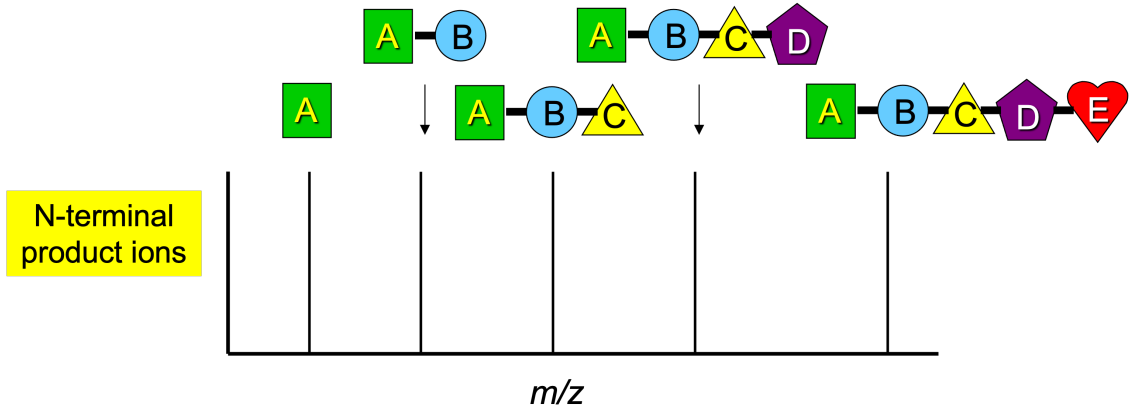


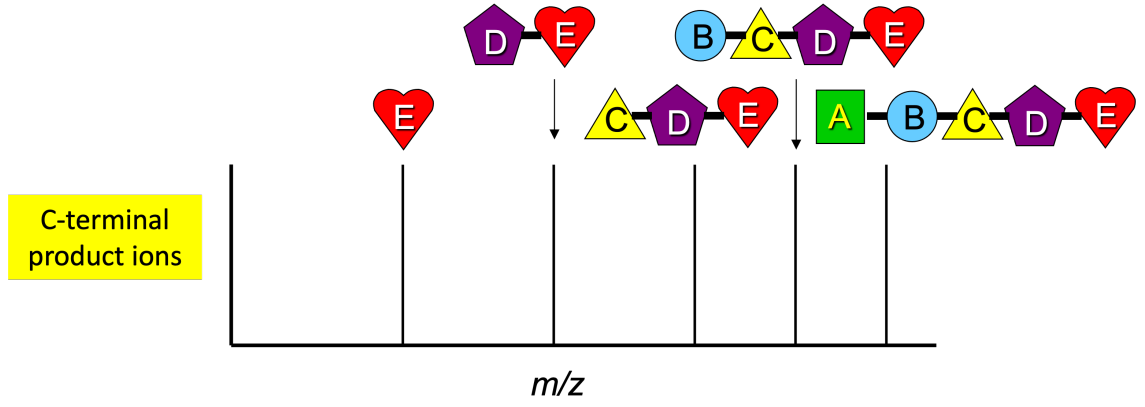
For a doubly protonated peptide, both N- and C-terminal fragments can be generated from a single dissociate event.

Peptide Sequencing



- Ideally, one can measure the spacings between product ion peaks to deduce the sequence
  - if each amide bond dissociates with equal probability
  - if only a single amide bond fragments for each molecule
  - if only C-terminal or N-terminal products ions are formed
- In reality, this is not the case…



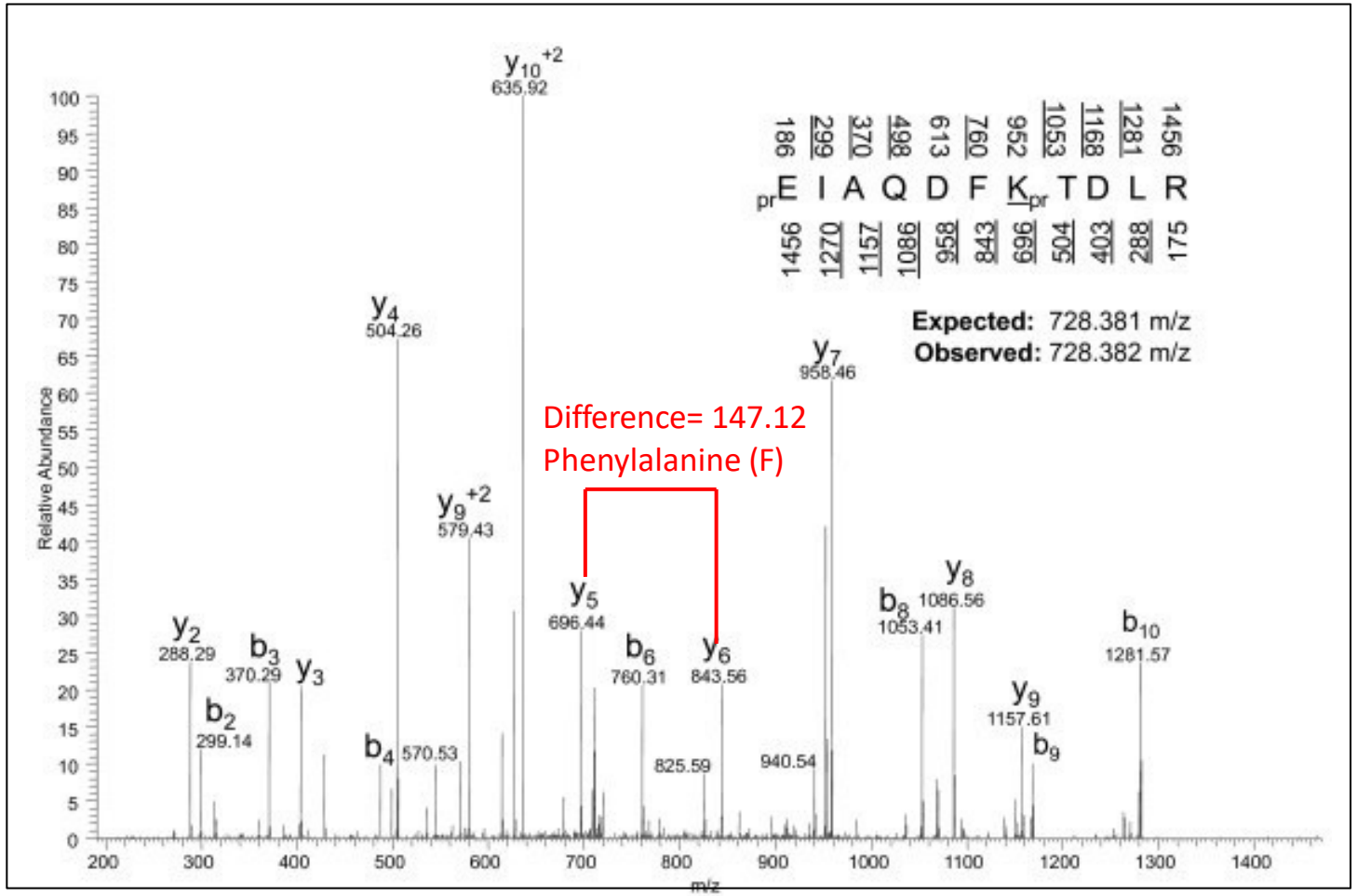


Fragmentation Results in a Peptide “Ladder”

Peptide: A-B-C-D-E

	b-ions	y-ions	
b_1^+	A	BCDE	y_4^+
b_2^+	AB	CDE	y_3^+
b_3^+	ABC	DE	y_2^+
b_4^+	ABCD	E	y_1^+

Mass Spectrum (Assignment of *b*- and *y*-ions)



- Mixture of b ions and y ions
- MS/MS of 2^+ charged tryptic peptides yield (often) 1^+ charged product ions (but 2^+ charged products can be observed as well)
- Not all b ions or y ions are visible

The mass of the precursor is 1454 (the observed ion was doubly charged)
728.382 x 2 - 2 = 1454.764 Da
Precursor ion ($M+2H^+$) is 1456.764 Da.

https://www.researchgate.net/profile/Rebecca-Levin-2/publication/49660823/figure/fig2/AS:214315966701573@1428108316420/Tandem-mass-spectrometry-MS-MS-spectrum-of-labeled-lysine-peptide-MS-MS-of-the-H3.png

Mass of amino acid fragment ion

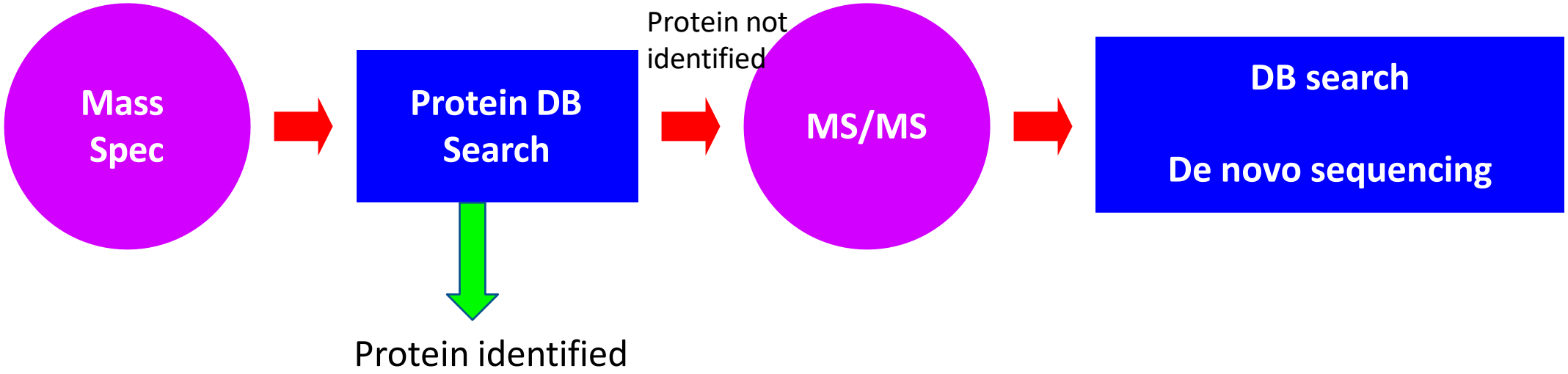
Name	3-letter code	1-letter code	Residue Mass	Immonium ion	Related ions	Composition
Alanine	Ala	A	71.03711	44		C_3H_5NO
Arginine	Arg	R	156.10111	129	59,70,73,87,100,112	$C_6H_{12}N_4O$
Asparagine	Asn	N	114.04293	87	70	$C_4H_6N_2O_2$
Aspartic Acid	Asp	D	115.02694	88	70	$C_4H_5NO_3$
Cysteine	Cys	C	103.00919	76		C_3H_5NOS
Glutamic Acid	Glu	E	129.04259	102		$C_5H_7NO_3$
Glutamine	Gln	Q	128.05858	101	56,84,129	$C_5H_8N_2O_2$
Glycine	Gly	G	57.02146	30		C_2H_3NO
Histidine	His	H	137.05891	110	82,121,123,138,166	$C_6H_7N_3O$
Isoleucine	Ile	I	113.08406	86	44,72	$C_6H_{11}NO$
Leucine	Leu	L	113.08406	86	44,72	$C_6H_{11}NO$
Lysine	Lys	K	128.09496	101	70,84,112,129	$C_6H_{12}N_2O$
Methionine	Met	M	131.04049	104	61	C_5H_9NOS
Phenyalanine	Phe	F	147.06841	120	91	C_9H_9NO
Proline	Pro	P	97.05276	70		C_5H_7NO
Serine	Ser	S	87.03203	60		$C_3H_5NO_2$
Threonine	Thr	T	101.04768	74		$C_4H_7NO_2$
Tryptophan	Trp	W	186.07931	159	11,117,130,132,170,100	$C_{11}H_{10}N_2O$
Tyrosine	Tyr	Y	163.06333	136	91,107	$C_9H_9NO_2$
Valine	Val	V	99.06841	72	44,55,69	C_5H_9NO

Mass of b2 ions in peptide fragmentation

	G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
G	115																	
A	129	143																
S	145	159	175															
P	155	169	185	195														
V	157	171	187	197	199													
T	159	173	189	199	201	203												
C	161	175	191	201	203	205	207											
I/L	171	185	201	211	213	215	217	227										
N	172	186	202	212	214	216	218	228	229									
D	173	187	203	213	215	217	219	229	230	231								
K/Q	186	200	216	226	228	230	232	242	243	244	257							
E	187	201	217	227	229	231	233	243	244	245	258	259						
M	189	203	219	229	231	233	235	245	246	247	260	261	263					
H	195	209	225	235	237	239	241	251	252	253	266	267	269	275				
F[b]	205	219	235	245	247	249	251	261	262	263	276	277	279	285	295			
R	214	228	244	254	256	258	260	270	271	272	285	286	288	294	304	313		
Y	221	235	251	261	263	265	267	277	278	279	292	293	295	301	311	320	327	
W	244	258	274	284	286	288	290	300	301	302	315	316	318	324	334	343	350	373

GG=N=114; GA=K/Q=128; GV=R=156; GE=AD=SV=W=186.

Protein Identification and Characterization Map



Databases

- Three components are required for database searching support of proteomics: MALDI or MS/MS data, the algorithms used to search protein databases, and the protein databases.
- A challenge for database searching is that these protein databases are constantly changing, making database search results potentially obsolete as new entries are added that better fit the MALDI or MS data.
    - Even as genomes are completed, there is still flux as new coding regions are identified and novel mechanisms of increased translational complexity are better understood, such as alternative splice products, RNA editing, and ribosome slippage leading to novel, unexpected translation products.

Some Representative Internet Sources for Protein Identification from Mass Spectrometric Data

Program	Web Address
BLAST	http://www.ebi.ac.uk/blastall/
Mascot	http://www.matrixscience.com/cgi/index.pl?page=/home.html
MassSearch	http://cbrg.inf.ethz.ch/Server/ServerBooklet/MassSearchEx.html
MOWSE	http://srs.hgmp.mrc.ac.uk/cgi-bin/mowse
PeptideSearch	http://www.narrador.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html
Protein Prospector	http://prospector.ucsf.edu/
Prowl	http://prowl.rockefeller.edu/
SEQUEST	http://fields.scripps.edu/sequest/