Last Class: Review

- Announcements
  1. **Final exam**: Open-book exam, **Tuesday (June 4th) from 2 pm to 5 pm (Beijing Time)**
  - 3 hours
  - in a campus computer room, but personal electronic devices are NOT allowed
  - You are allowed to search the internet for information but not to copy text directly. Write the answers ONLY on the booklet provided in your own words and always acknowledge the source(s) from which your answers are derived.

  2. No class on Thursday and in week 13
  - Schedule a meeting with me at anytime if you have questions; or
  - **Office hours**: 9 am – 5 pm on May 24th (Friday in week 13)

  3. Please use "20240509 tutorial of CWs" as the feedback of your CWs.
  - Marks of CW1 will be released by the end of this week.
  - Marks of CW2 will be released by the end of next week.

Please fill in the **MQ of BIO312** at your free time. It is available from **May 13th to May 26th, 2024.**

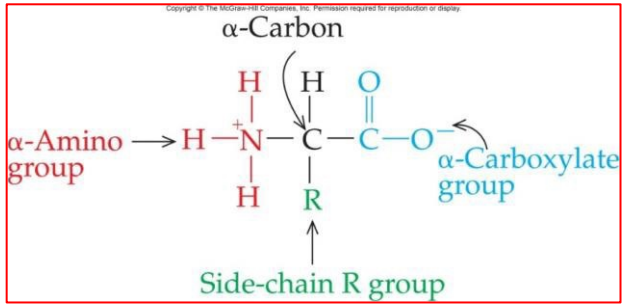
Your honest and constructive feedback is appreciated. Thanks.

Completion of all questionnaires will enable you eligible for a lucky draw at the end of this semester.

Learning Outcomes

1. Know the methods used for the extraction, enrichment and analysis of proteins
2. Discuss how proteomics-based approaches can be used to study fundamental and applied biological problems
3. Describe methods of analysis of post-translational modifications of proteins and implications for cell function
4. Understand how technologies such as nuclear magnetic resonance (NMR), X-ray crystallography and other physical methods can be used to determine the detailed fine structure of proteins
5. Explain how the knowledge of protein structure can be used to explain function

Protein Building Blocks



Nonpolar side chains

Glycine (G) Gly

Alanine (A) Ala

Valine (V) Val

Leucine (L) Leu

Isoleucine (I) Ile

Methionine (M) Met

Phenylalanine (F) Phe

Tryptophan (W) Trp

Proline (P) Pro

Polar side chains

Serine (S) Ser

Threonine (T) Thr

Cysteine (C) Cys

Tyrosine (Y) Tyr

Asparagine (N) Asn

Glutamine (Q) Gln

Electrically charged side chains

Acidic

Aspartate (D) Asp

Glutamate (E) Glu

Electrically charged side chains

Basic

Lysine (K) Lys

Arginine (R) Arg

Histidine (H) His

Methods for Protein Separation

Different sizes	Different solubility	Different charges	Different Ligand Binding
• Ultracentrifugation • Dialysis • Size exclusion chromatography • PAGE (SDS-PAGE or native PAGE)	• Salting out e.g. Ammonium sulfate	• Ion exchange chromatography • Electrophoresis	• Affinity chromatography

2D PAGE: based on different **pI** and **sizes**

2.Discuss how proteomics-based approaches can be used to study fundamental and applied biological problems

a. protein identification/presence

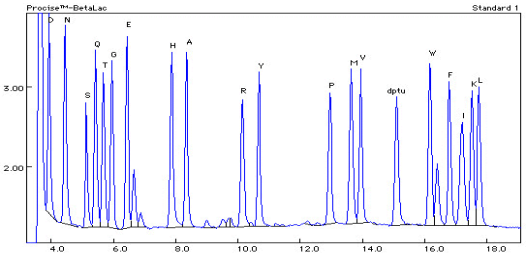
Strategies for Protein Identification

Works with 30 aa fragments (in practical)

Cut protein with cyanogen bromide (CNBr)
- cuts on carboxyl side of Met

Phenylisothiocyanate (PITC)
- binds to and releases N-terminal residue

Chromatography against known standards



Sample

One or two-dimensional Protein electrophoresis

Or muti-dimensional chromatography

Electroblotting

In gel digestion with proteases

CNBR/Skatole Cleavage

Mass Spectrometry

Mixed Peptide Edman Sequencing

Peptide Mass Fingerprinting

Peptide Sequencing

Database Searching

Protein Identification

Recap in Slide 9

Graves PR, Haystead TA. Molecular biologist's guide to proteomics. *Microbiol Mol Biol*

The Order of Amino Acids Matters

- How to assemble the fragments in Correct order?

1. Overlapping fragments cut by another reagent/protease

2. Design degenerate PCR primers based on the solved AA sequences, then amplify the gene or cDNA sequences. These can be translated into AA sequence

Edman sequencing results

Digestion with Trypsin

Leu-Asp-Glu-Trp-Gly-Val-Ile-Lys

Ala-Val-Ile-Leu-Ser-Glu-Ile-Lys

His-Thr-Val-Glu-Val-Arg

Digestion with endoprotease Glu-C

Trp-Gly-Val-Ile-Lys-Ala-Val-Ile-Leu-Ser-Glu

Ile-Lys-His-Thr-Val-Glu-Val-Arg

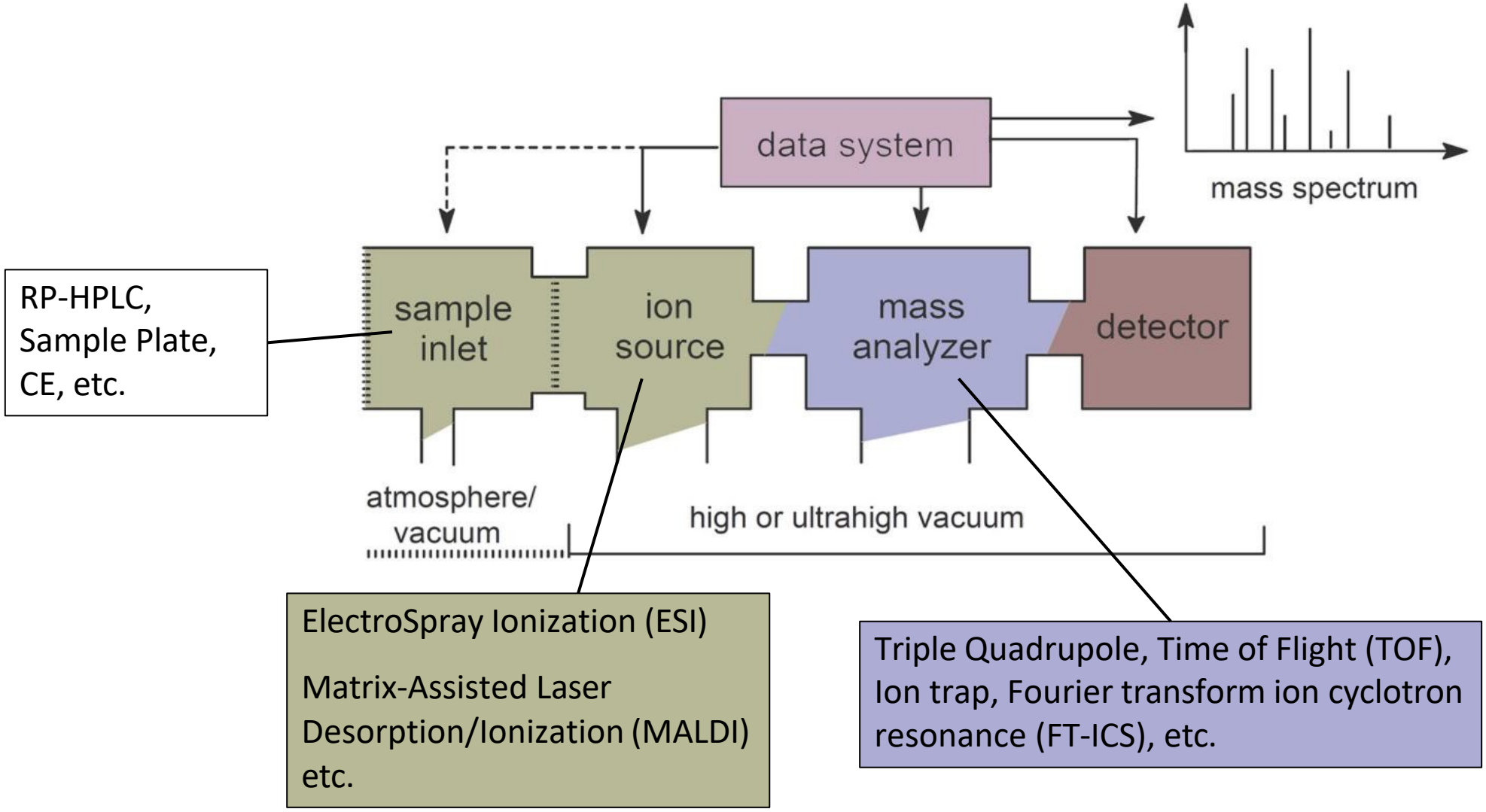
Overlap

Sequence deduced:

Leu-Asp-Glu-Trp-Gly-Val-Ile-Lys|Ala-Val-Ile-Leu-Ser-Glu-Ile-Lys|His-Thr-Val-Glu-Val-Arg

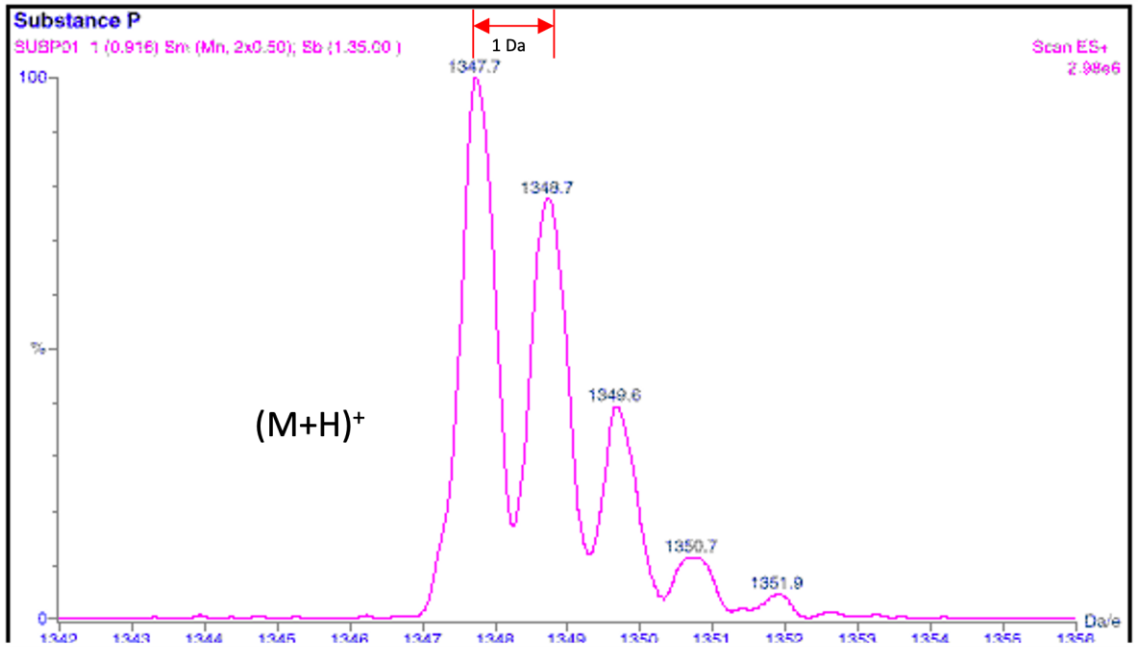
Trp-Gly-Val-Ile-Lys-Ala-Val-Ile-Leu-Ser-Glu|Ile-Lys-His-Thr-Val-Glu-Val-Arg

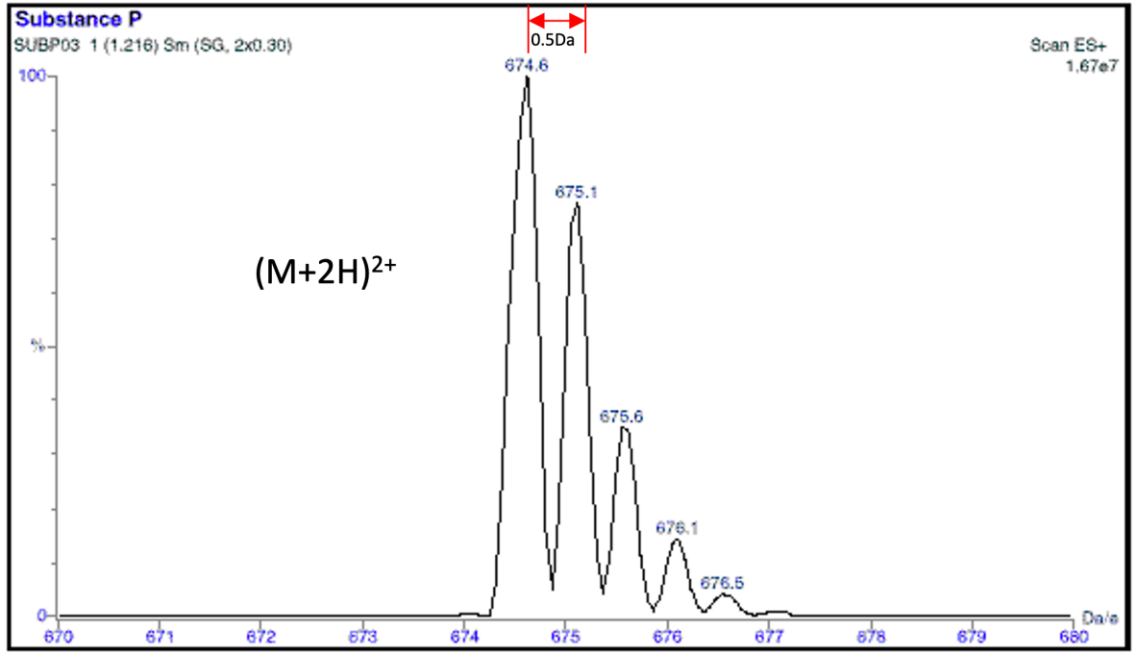
General Scheme of a Mass Spectrometer



- Mass spectrometry is an extremely **sensitive.**
- Samples are needed to be transformed into **ions**, for MS to work.
- Ions are delivered in the dilute gas phase

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





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

Isotopes of singly charged ions are separated by 1 Da

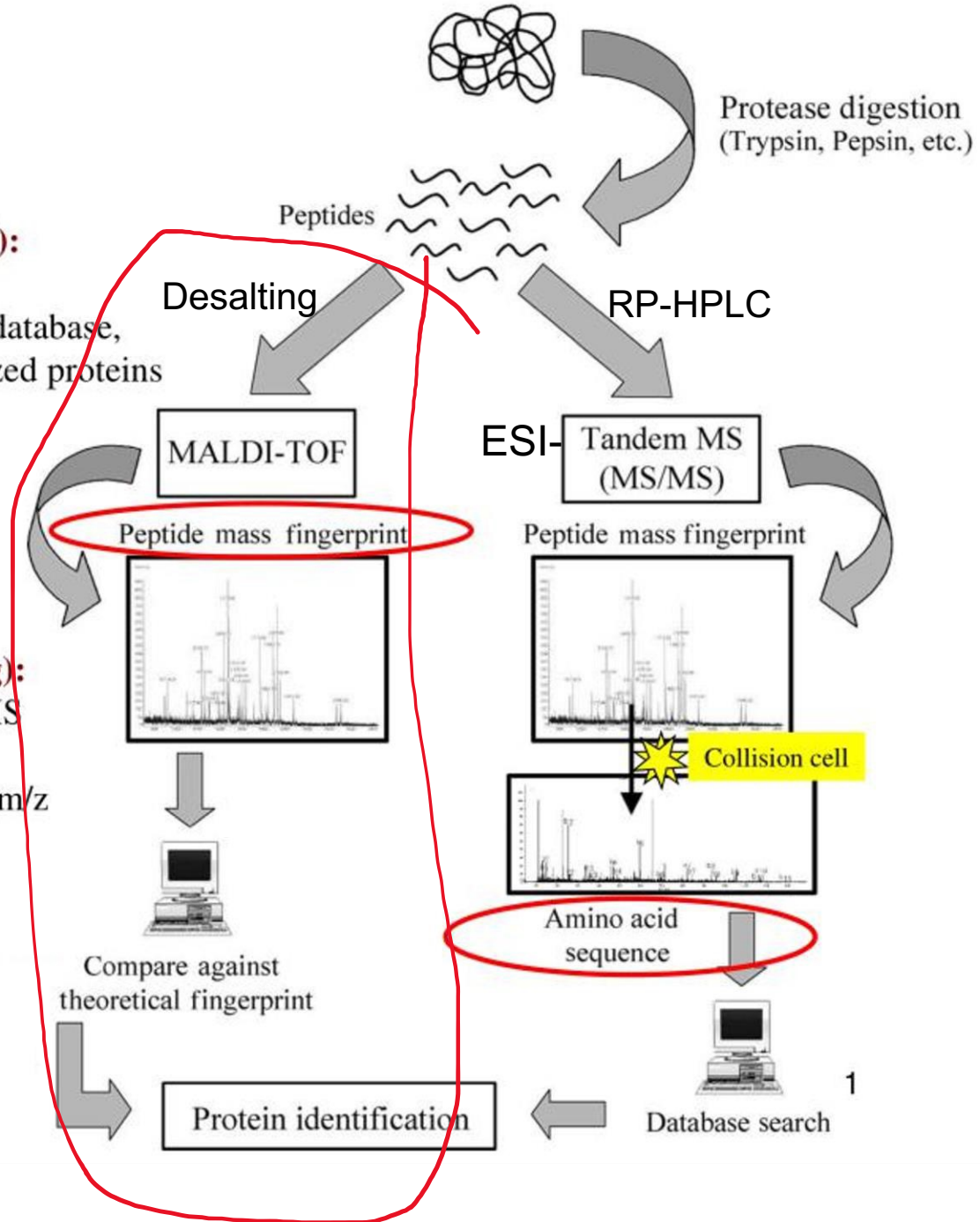
Isotopes of doubly charged ions are separated by 0.5 Da

MS-based Protein Identification

Single MS (peptide fingerprinting):
Identifies m/z of peptide only
Peptide id'd by comparison to database,
of predicted m/z of trypsinized proteins

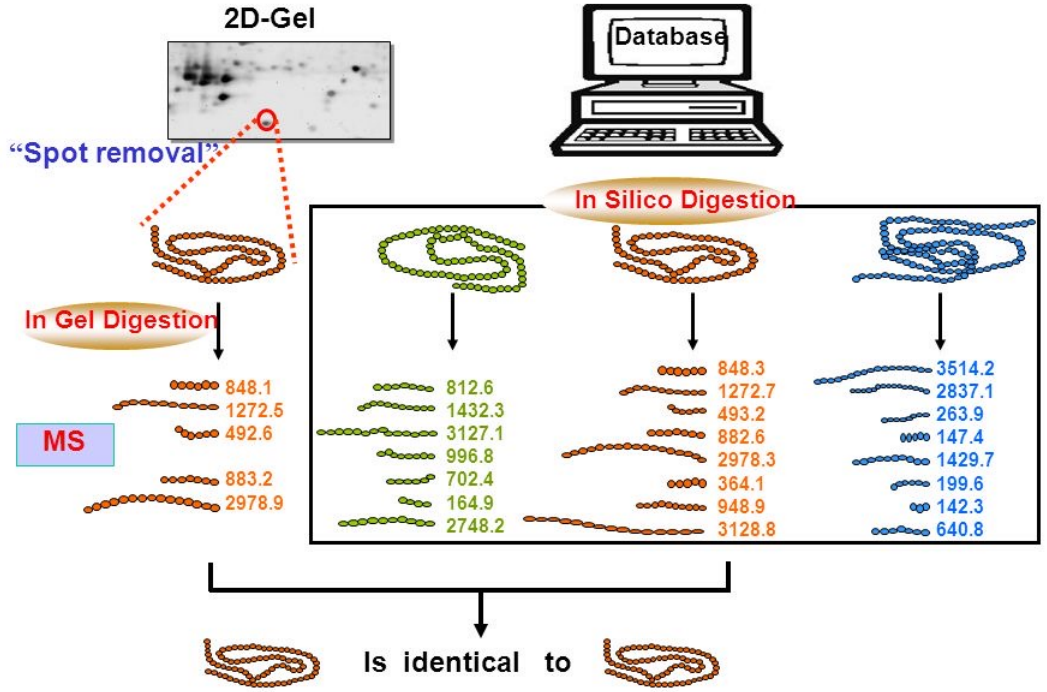
Tandem MS/MS (peptide sequencing):
Pulls each peptide from the first MS
Breaks up peptide bond
Identifies each fragment based on m/z

CID: collision induced dissociation



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).
  - In the presence of some buffer ions, sodium (M+ Na^+), potassium (M+K^+), and ammonium (M+ NH_4^+) adducts may be formed.

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.

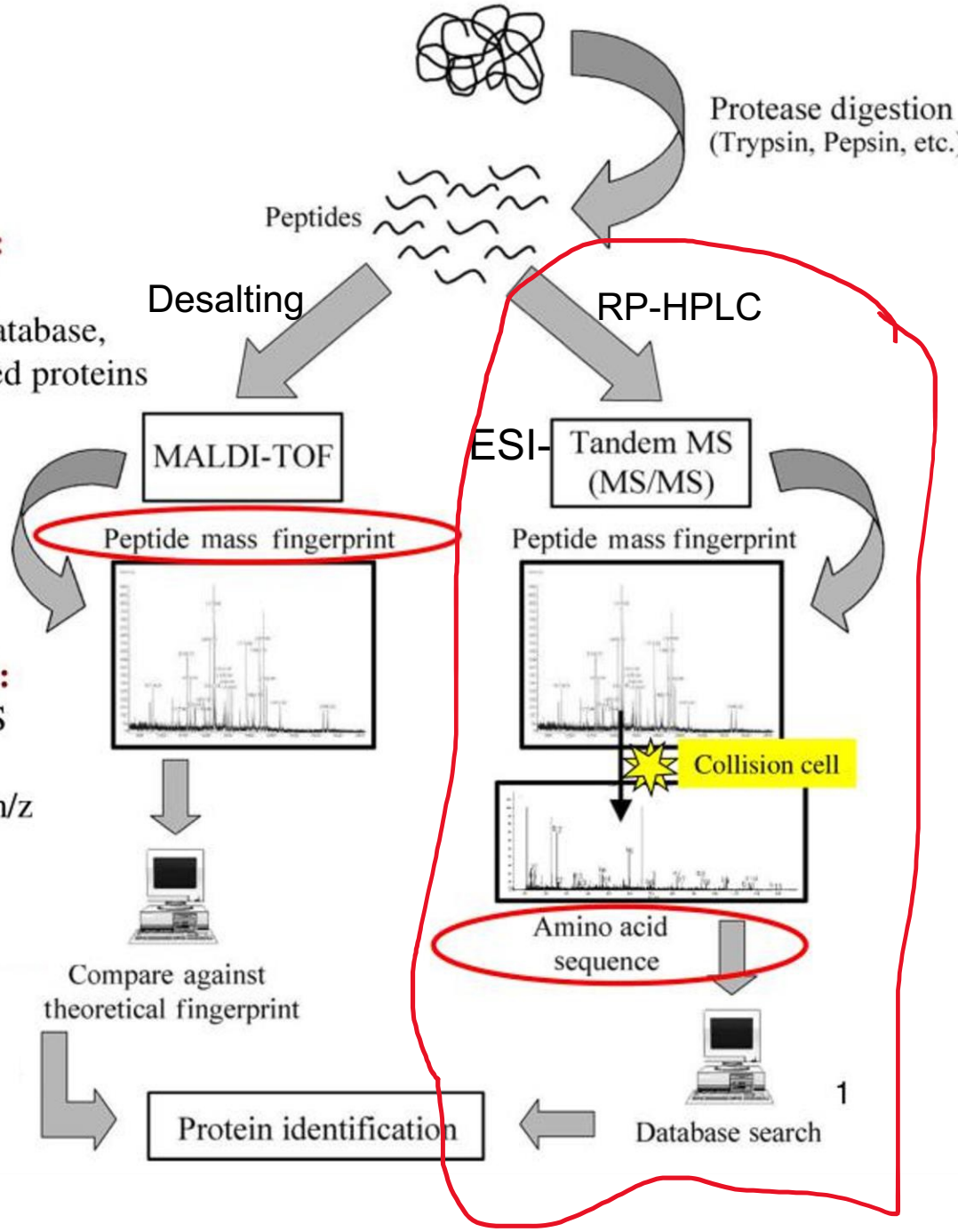
Non-specific cleavage, isobaric peptide (same mass but different order), etc.

MS-based Protein Identification

Single MS (peptide fingerprinting):
Identifies m/z of peptide only
Peptide id'd by comparison to database,
of predicted m/z of trypsinized proteins

Tandem MS/MS (peptide sequencing):
Pulls each peptide from the first MS
Breaks up peptide bond
Identifies each fragment based on m/z

CID: collision induced dissociation

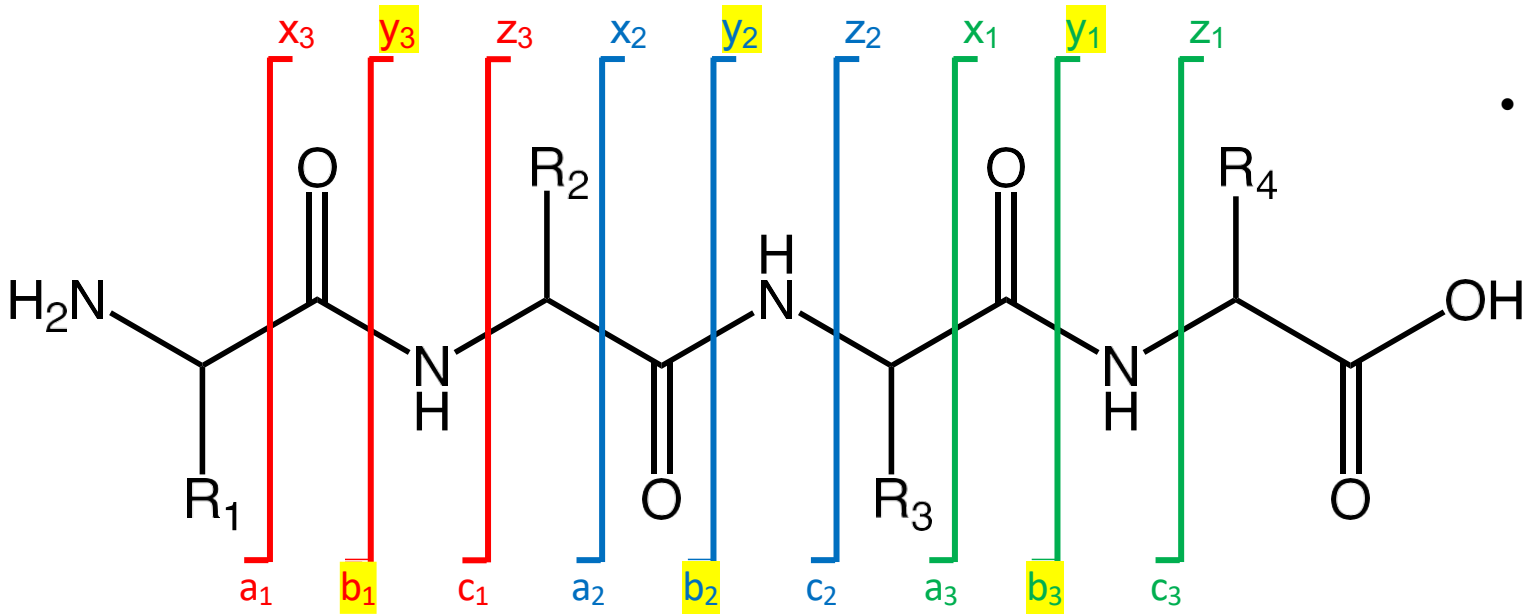


2.Discuss how proteomics-based approaches can be used to study fundamental and applied biological problems

b. Peptide sequencing

Nomenclature of Fragment Ions

Charge stays on *either* the 'left' (a, b, or c) or 'right' (x, y, or z) side of cleavage

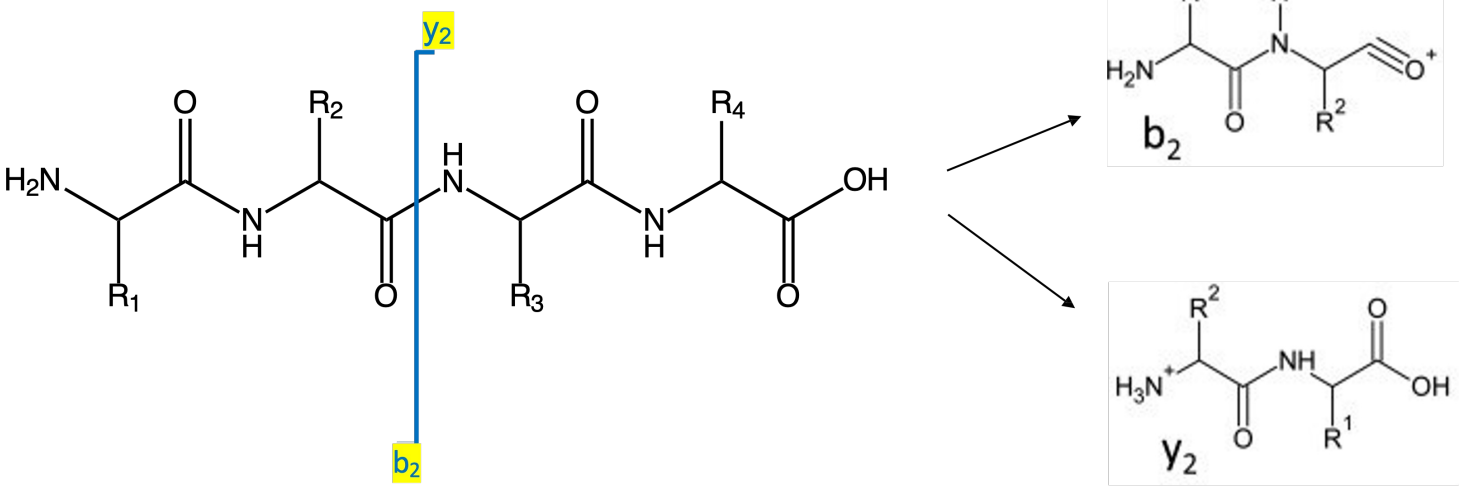


- Letter: Indicates the bond broken and the terminus contained in the fragment
- Number: Indicates the number of Cα in the fragment

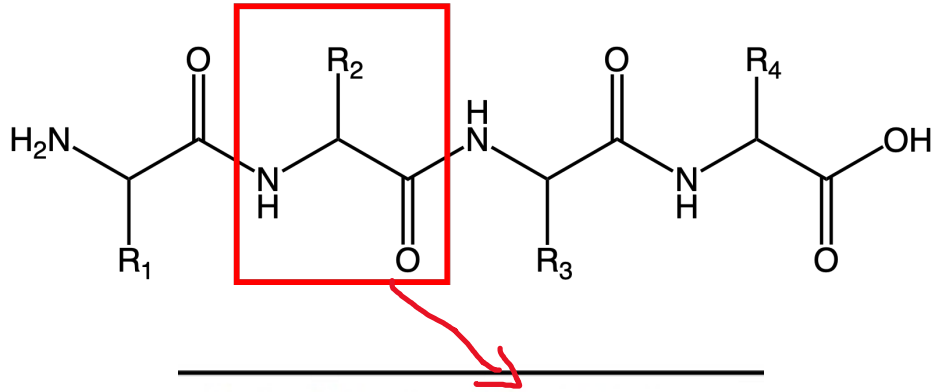
CID: mostly b ions, y ions

ECD/ETD: c ions and z ions

Peptides Fragment by CID



- Mass of b-ions = Σ (residue masses) + 1 (H^+)
- Mass of y-ions = Σ (residue masses) + 19 (OH + H + H^+)



Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

Residue Identity Determination based on b/y ions in MS2 spectrum

✓ **If there are <u>small m/z</u> ions corresponding to the b_1 or y_1 ion, but you don't know which:**

- unknown m/z – 1 gives you reasonable mass value →b1, N terminus residue
- unknown m/z – 19 gives you reasonable mass value →y1, C terminus residue

✓ **If there are <u>big m/z</u> ions corresponding to the b_{n-1} or y_{n-1} ion, but you don't know which:**

- M - y_{n-1} ion + 1 = mass of 1st residue on N terminus
- M- b_{n-1} ion - 17 = mass of 1st residue on C terminus

- After knowing the terminal residue(s), successively calculate Δmass .

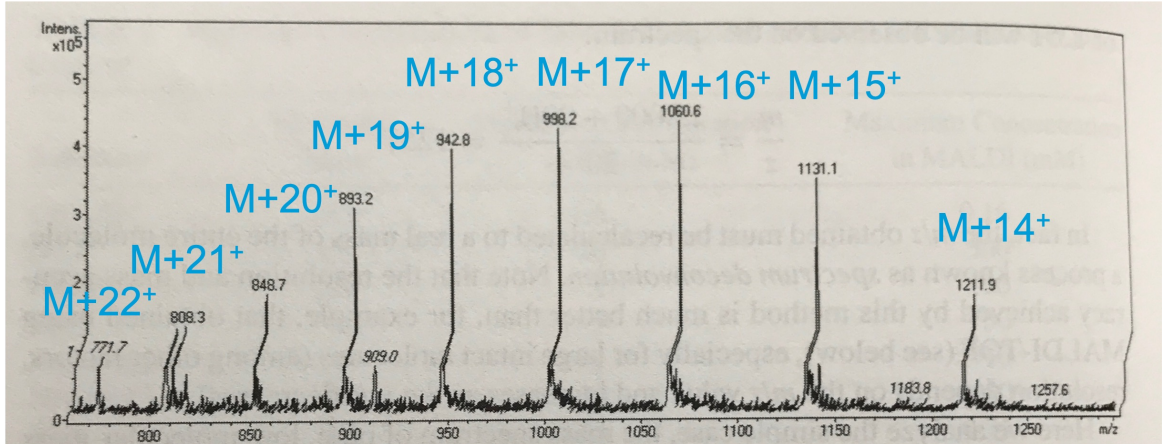
- A complementary b-y ion pair can be observed in multiply charged ions spectra.
  - **Mass of peptide = b_m+ y_{n-m} – 2**
  - n is the number of residues;
  - m is the number of residues on a fragment ion.

Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

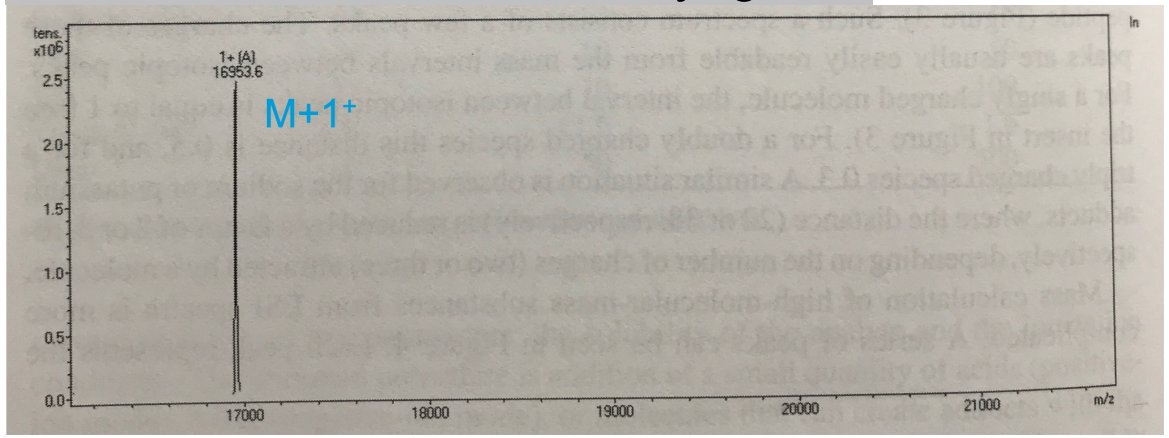
Peptide Mass Calculation

$$M = [M+H]^+ - 1$$

ESI-MS of horse heart myoglobin Multiple charged ions



MALDI-MS of horse heart myoglobin Singly charged ions



To deconvolute the myoglobin mass,

1. Calculate the charge

$$z_n = \frac{m_{n+1} - 1.0078}{m_n - m_{n+1}}$$

2. Calculate the mass of each ion
3. Average them with deviation

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

myoglobin MW = 16953.6 – 1 = 16952.6 Da

Reference Tables

- **Mass of b-ions** = Σ (residue masses) + 1 (H^+)

Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

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.

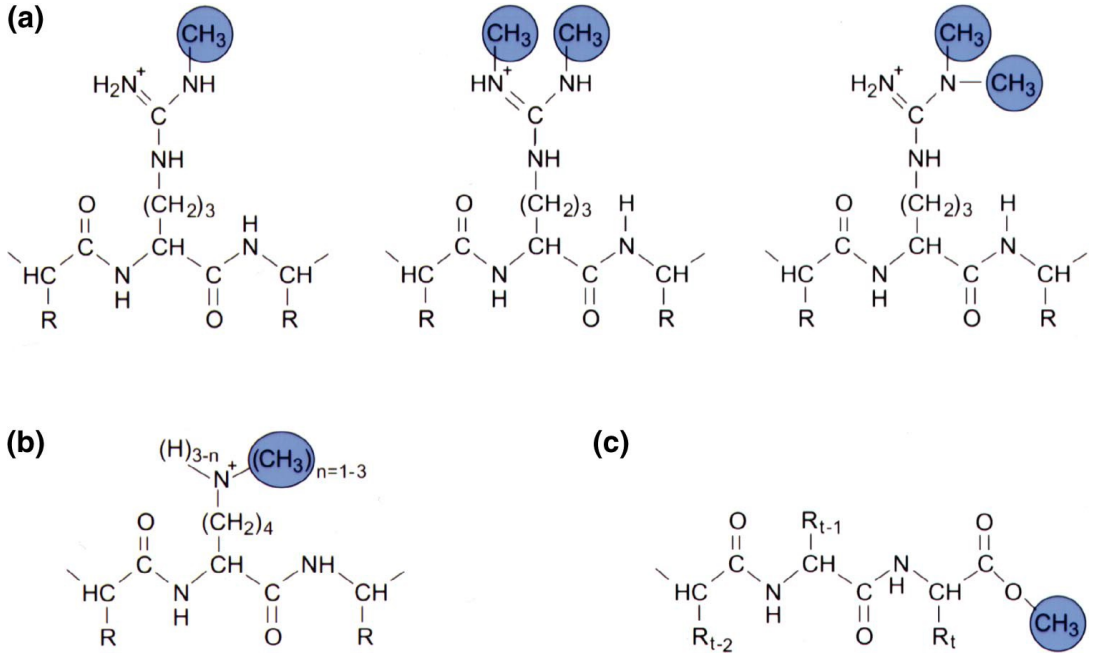
2.Discuss how proteomics-based approaches can be used to study fundamental and applied biological problems

3. Describe methods of post-translational modifications analysis and implications for cell function

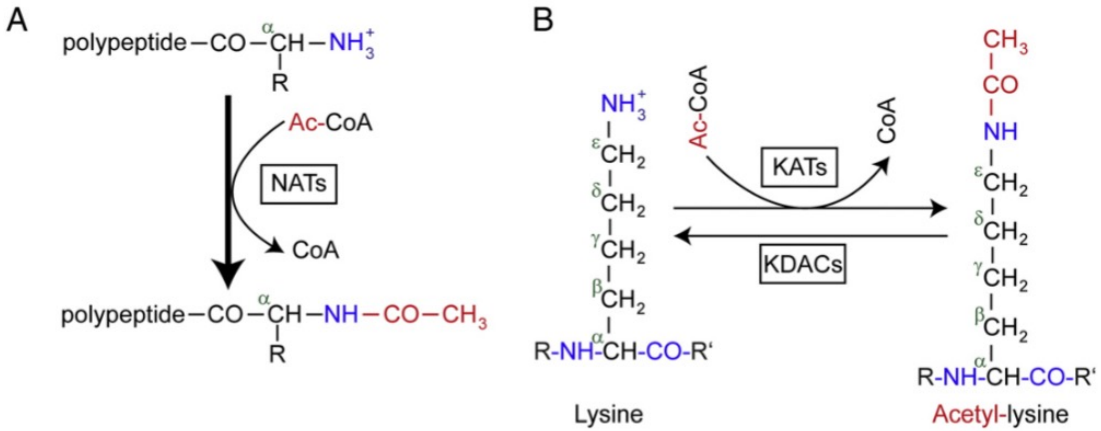
c. Identify protein/peptide with PTMs

PTMs

Methylation



Acetylation

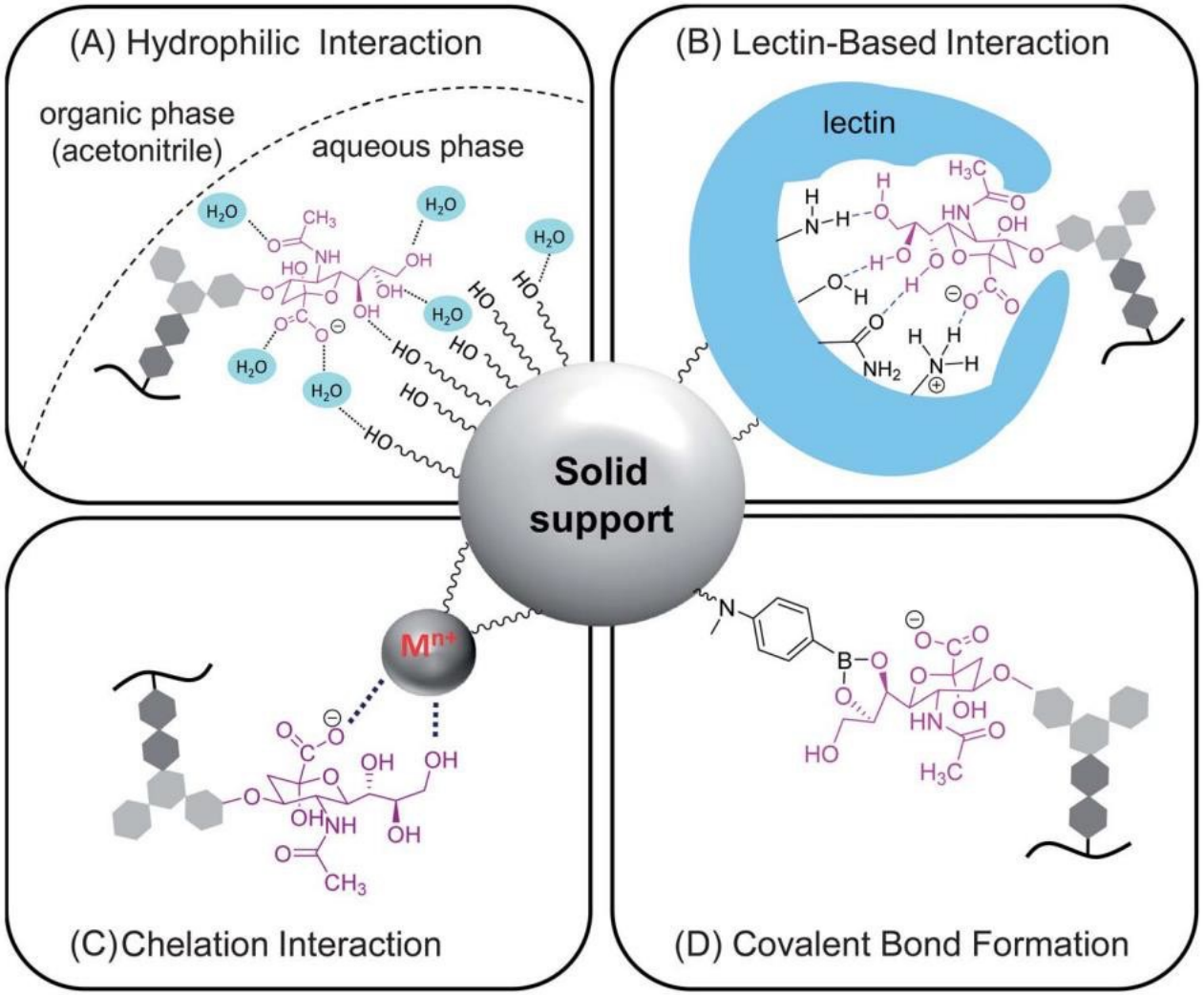


- Identified them by comparing the mass shift.

Methylation: +14 Da x n
Acetylation: +42 Da x n

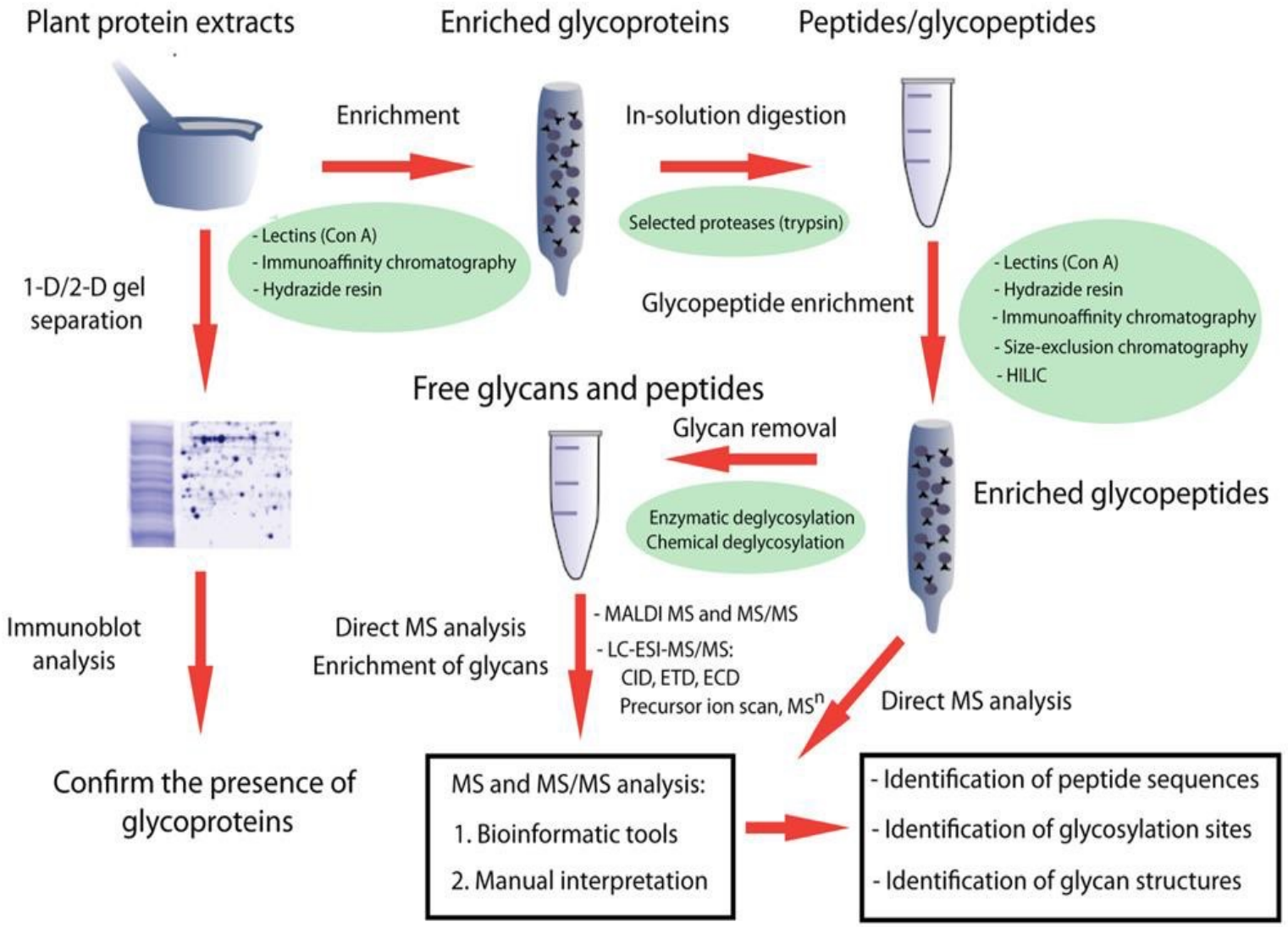
n is the number of modifications.

Enrichment of Glycopeptides



- Due to low abundance glycoproteins/glycopeptides usually need to be enriched
- **Lectin affinity chromatography**
  - Lectins are sugar-binding proteins
  - Several lectin-affinity procedures may be used one after the other to select different classes of glycan chains progressively.
- Unlike lectin affinity methods, enrichment **through covalent bonding** is a universal method because the binding is based on carbohydrate *cis-diol group* reactivity without discrimination between different glycan structures.

General Work-flow for the Full Analysis of Glycoproteins



- The full analysis of glycoproteins must involve characterization of both the peptide and glycan.
- For glycan analysis,
  - Stepwise degradations with specific reagents (e.g., *O*- or *N*-glycosidase) that reveal bond position and stereochemistry
  - Mixture separated by chromatography
  - Overall composition and analysis by GC, Mass Spec and NMR

Thannhauser, T. W., Zhang, S., & Rose, J. K. (2012). Analytical technologies for identification and characterization of the plant N-glycoproteome. *Frontiers in Plant Science*, *3*. https://doi.org/10.3389/fpls.2012.00150

Enrichment of Phosphoproteins

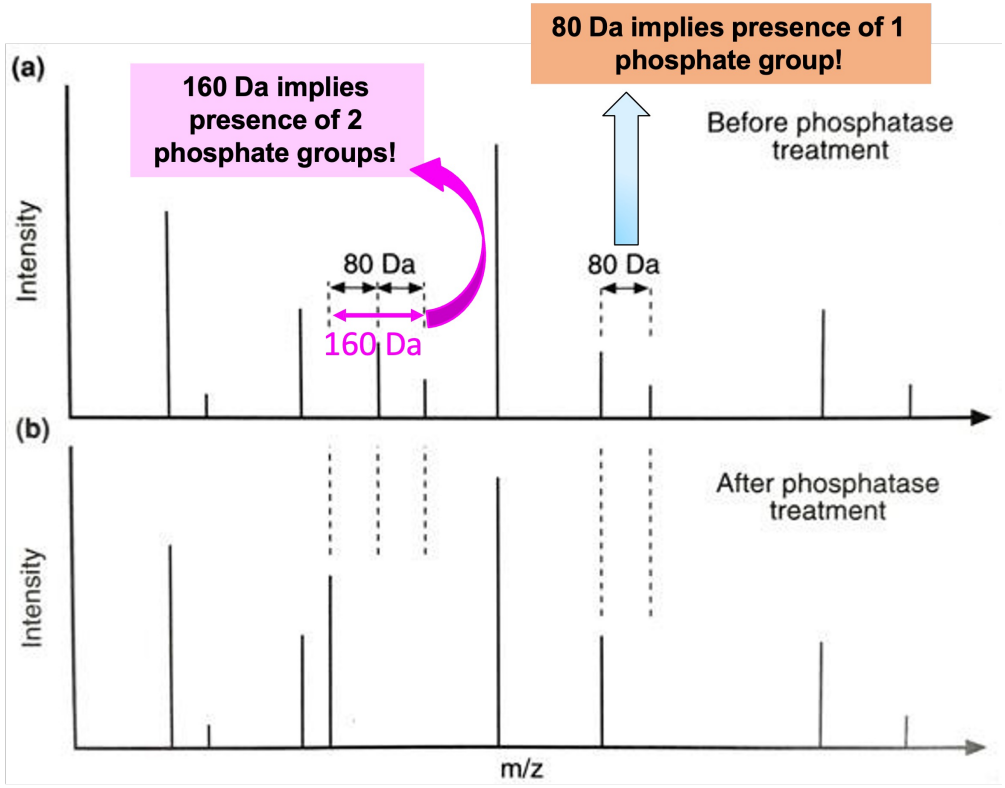
1. Antibodies (anti-phospho-Tyr)
   - limited in throughput and hard to automate
2. IMAC
   - Interactions between negatively charged phosphate groups and positively charged trivalent metal ions or TiO_2
   - Relatively low selectivity
3. SCX
   - Difference in the solution **charge state** of phosphorylated and non-phosphorylated peptides
   - Phophopeptides come out earlier than non-phosphorylated peptides
   - multiply phosphorylated peptides will be in the flow-through fraction
4. Chemical modification
   - β–elimination reaction➔addition of Biotin➔bind to immobilized streptavidin
     - Only for phospho-Ser and phospho-Thr, it doesn't work with phospho-Tyr
     - O-glycosylated Ser/Thr can also be derivatized
   - Carbodiimide condensation reaction➔Cystamine added-->binds to Iodoacetylated beads

MS for Phosphopeptides Detection[1]

1. MALDI-TOF MS combining with alkaline phosphatase treatment

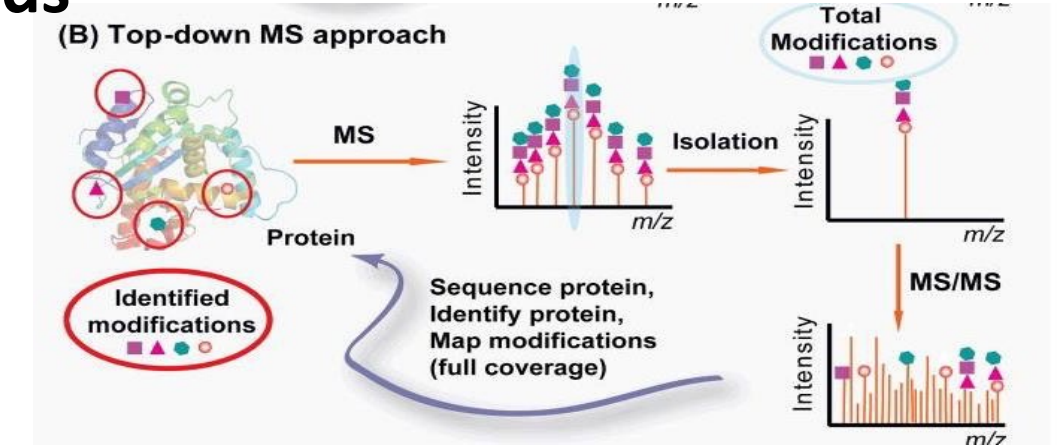
- Intact peptides: match against the theoretical peptides of known proteins.
- Phosphopeptides: compare the mass spectra for mass shift of 80 x N Da (N is the number of phosphoryl groups)

But the method cannot directly pinpoint the exact location of a phosphoryl group when there are multiple potential places for phosphorylation.



2. Electron-based dissociation methods (ECD/ETD)

- single proteins or relatively simple mixture
- ESI and trapped in FT-ICR or orbit trap mass spectrometer
- It leaves PTMs intact during fragmentation
- c and z ions

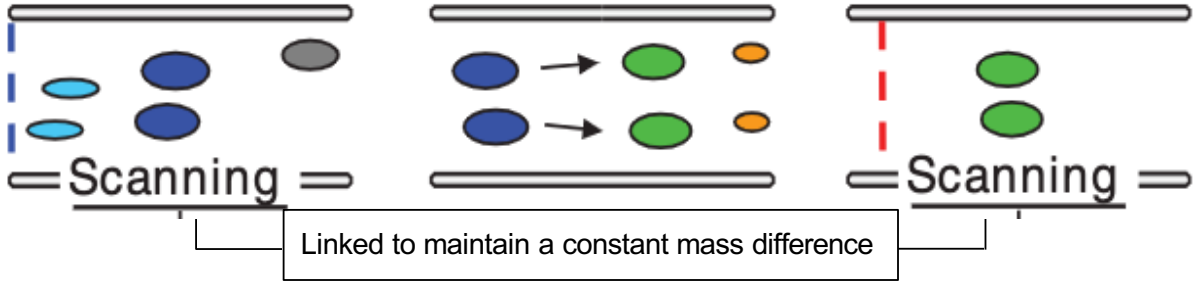


MS for Phosphopeptides Detection$_2$

3. Neutral loss ion scanning

QqQ MS (scanning mode in Q1 and Q3)

Detects neutral loss of 98 Da and 80 Da between precursor and fragment ions.



4. Precursor/Reporter Ion Scanning

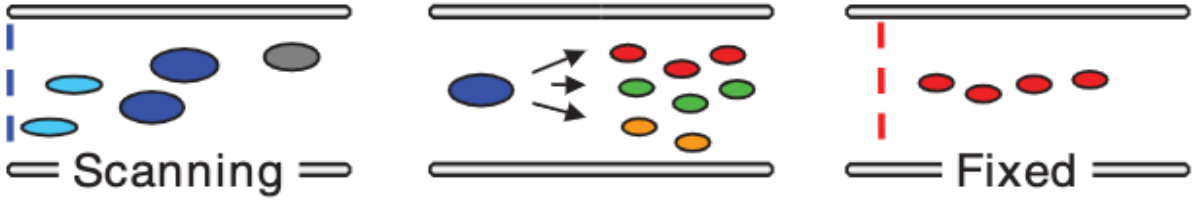
Q-q-TOF/Ion trap(q: Higher collision energies)

Negative ion mode to detect the presence of phosphopeptides (97, 79, and 63 Da).

Positive ion mode to determine the tandem mass spectra of identified phosphopeptides.

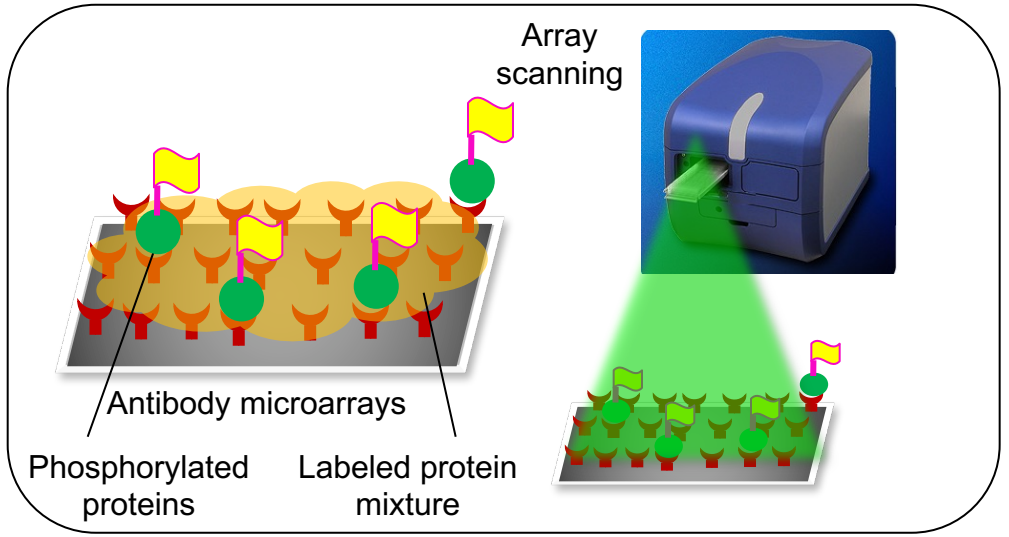
Less sensitive QqQ can be used to detect diagnostic fragment ions. (Q1: scanning; Q3: fixed)

216.043 Da in positive mode for p-Tyr peptides.



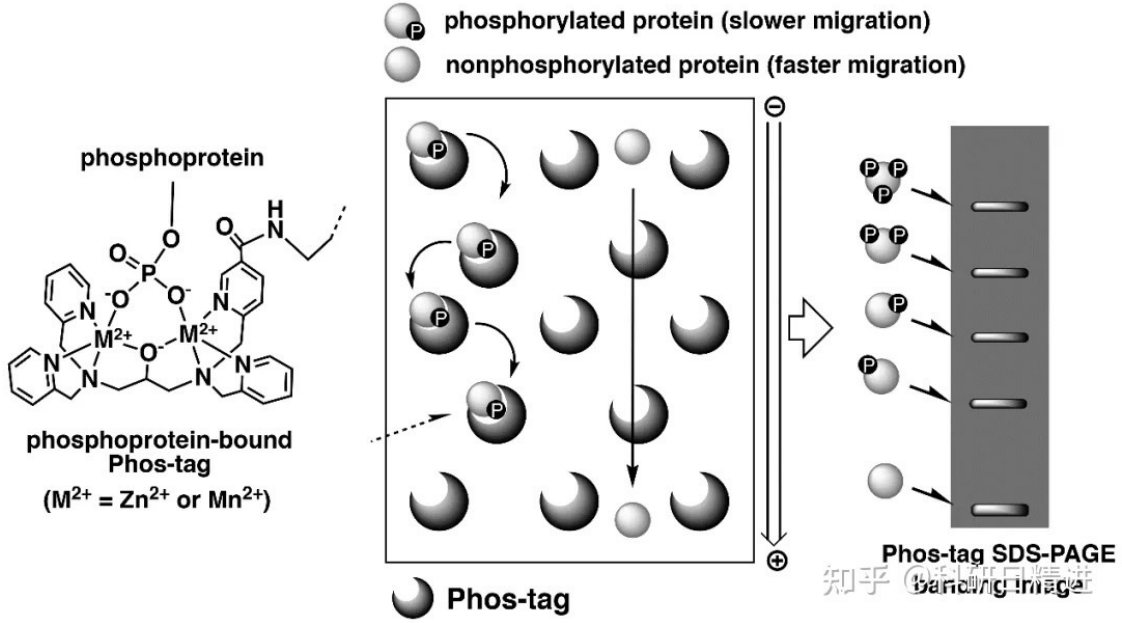
Other Phosphopeptides Detection Methods

5. Microarray-based detection



Anti-phospho serine, threonine and tyrosine antibodies immobilized on array

6. Gel-based detection with Phospho-tag



2.Discuss how proteomics-based approaches can be used to study fundamental and applied biological problems

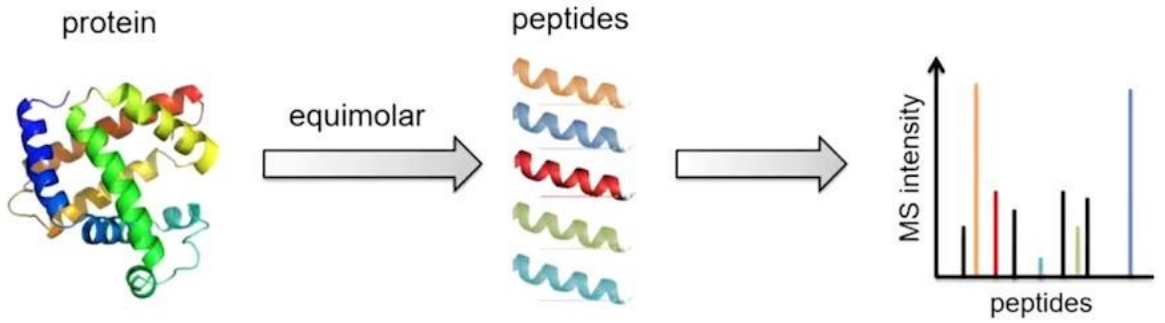
d. Quantification

- Identifying expression profiles in different biological processes
- Diagnosing certain diseases and cancer biomarkers
- Monitoring changes in certain biological process proteomes (time course)
- Studying protein interaction networks

Quantitative Proteomics

- simultaneously quantitation of level differences between many proteins in different samples

- There is a poor correlation between the amount of a peptide and the MS intensity in a single MS spectrum



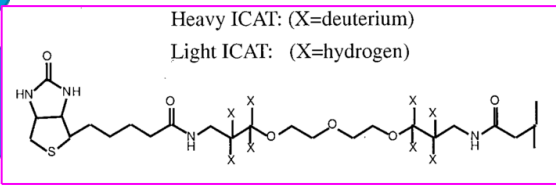
- MS intensity does **NOT** tell us peptide abundance directly. Many factors affect the MS intensity:
  - Peptide concentration
  - Day to day and long-term instrument reproducibility
  - Digestion efficiency
  - Recovery during sample preparation
  - **Ionization efficiency**
  - **Instantaneous matrix effects**

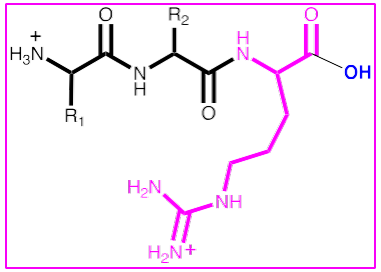
(Relative) Quantitative Proteomics

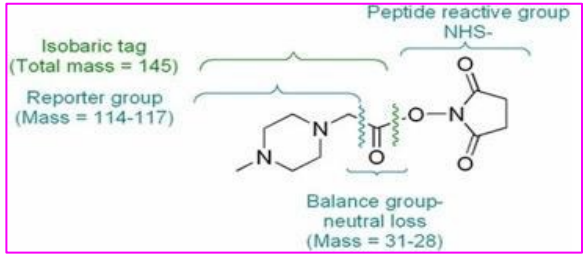
- In vivo labeling with isotope enriched media
- Cell culture

Quantitative proteomics

- 2-DE based methods
- MS based methods
  - Stable isotope labeling
    - MS1 quantification
      - Metabolic labeling (SILAC, ^{15}N)
      - Chemical labeling (ICAT, Dimethyl)
      - Enzymatic labeling (^{18}O)
    - MS2 quantification
      - Isobaric labeling (iTRAQ, TMT)
  - Label-free relative quantification







Compare the spectral count of the same protein or peak intensity of the same peptide.

- **Spectral Counting**: Count number of MS^2 spectra matching peptides for each protein
- **Peptide ion intensity**: Integration of MS^1 peak area in extracted ion chromatograms (XIC)

Label versus label free

Label (pros and cons)

- Stable isotopes are expensive and not suitable for clinical samples
- Limited number of samples

- But... less experimental variation if samples are mixed
- Even better the earlier workflows can be mixed...
- Improved quantitative precision and accuracy
- Improved confidence in peak identification

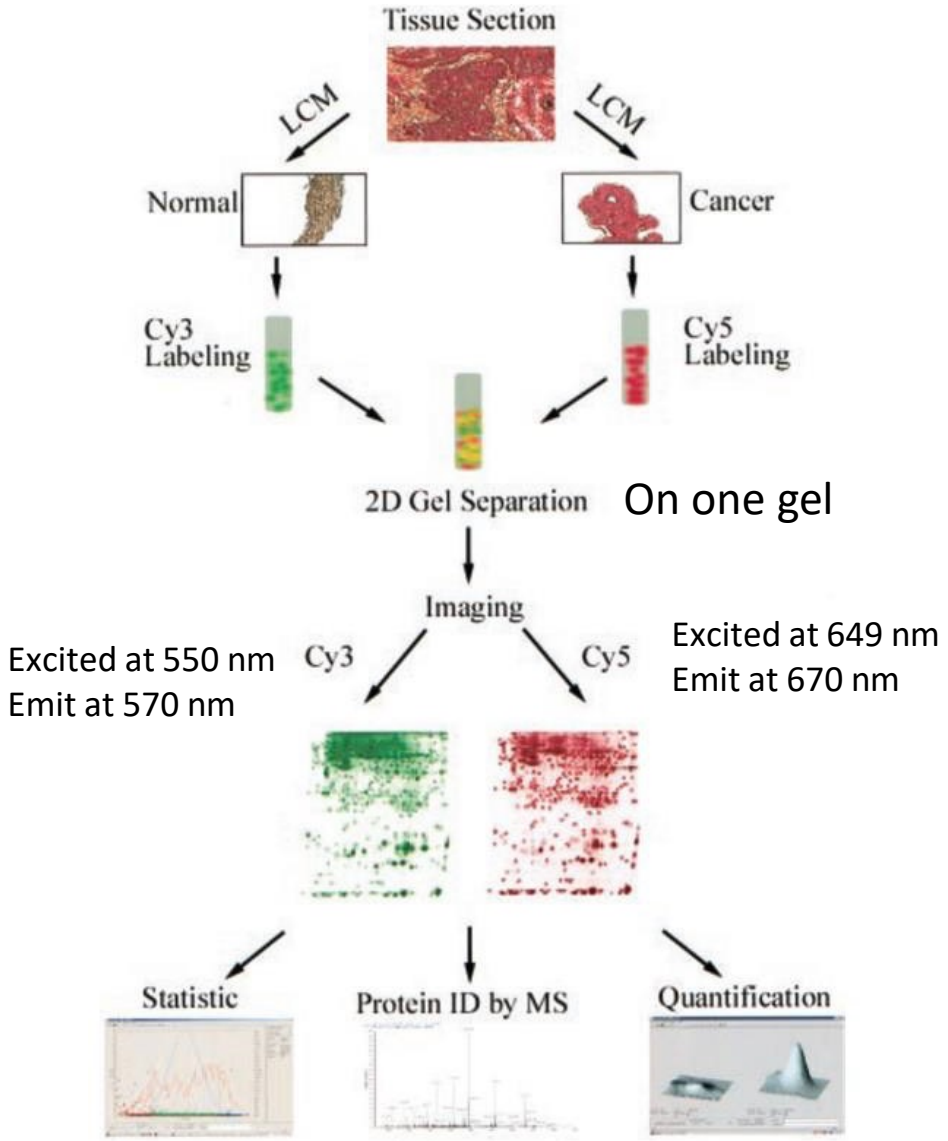
Label-free (pros and cons)

- Need lots of replicates to get statistical power
- So lots of time on MS instrument – therefore also can be expensive
- Fairly new technique therefore not enough high-quality published studies showing best practice

- No labelling needed
- No limit on the number of samples
- Applicable to any kind of samples

Electrophoretic Technique

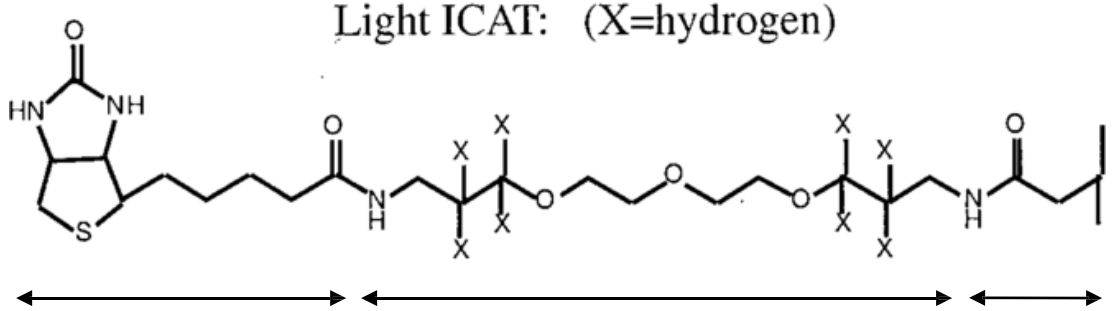
- It is often difficult to obtain reproducible separations by standard 2D PAGE.
- Quantitative proteomics utilizes *fluorescent dyes*
  - Ease of use
  - High sensitivity (1 ng)
  - Low background (not staining the gel)
  - Compatibility with MS
  - Broad range of linearity
- Differential gel electrophoresis (DIGE) is used in quantitative proteomics.



Labeling: ICAT (Isotope- Coded Affinity Tagging)

Heavy ICAT: (X=deuterium)

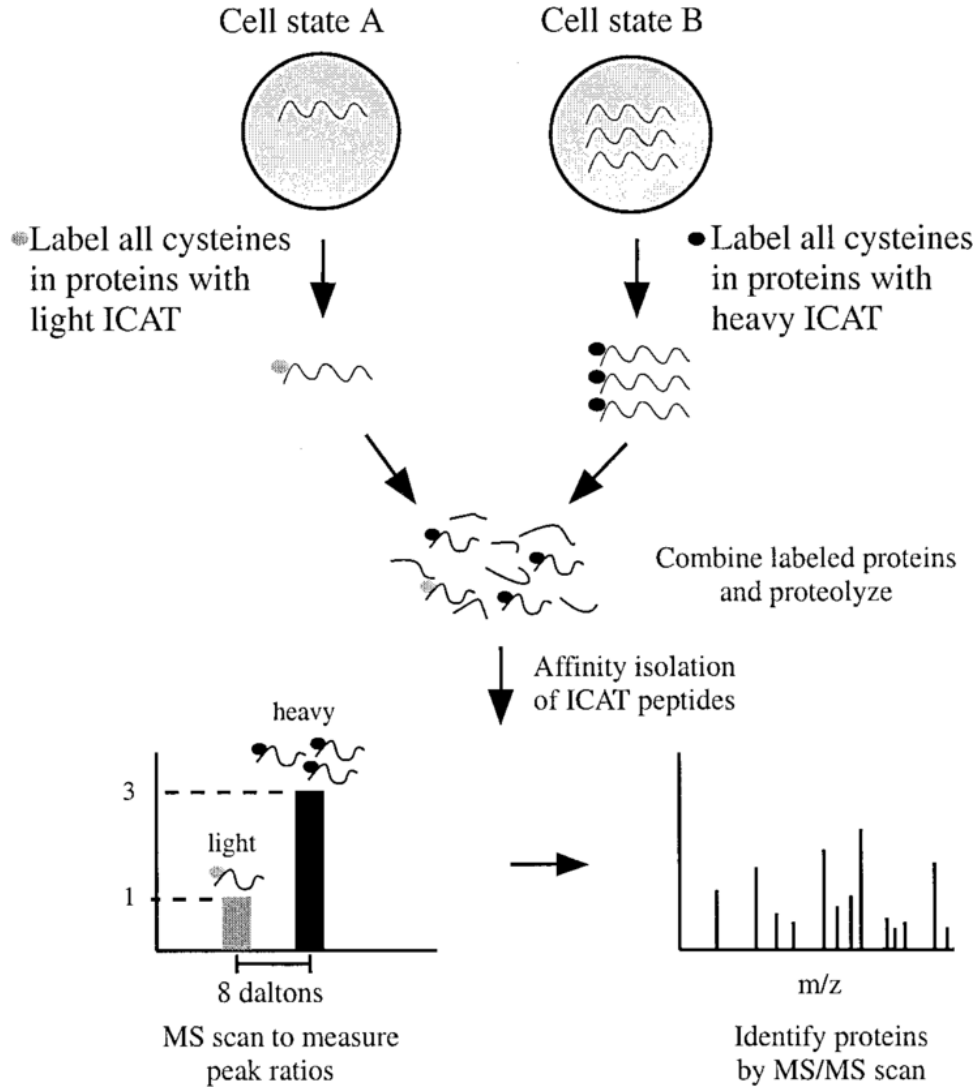
Light ICAT: (X=hydrogen)



Biotin moiety, which permits isolation of the labeled molecules by affinity chromatography (interact with avidin immobilized)

Linker containing light (X= H) or heavy (X=D) isotopes

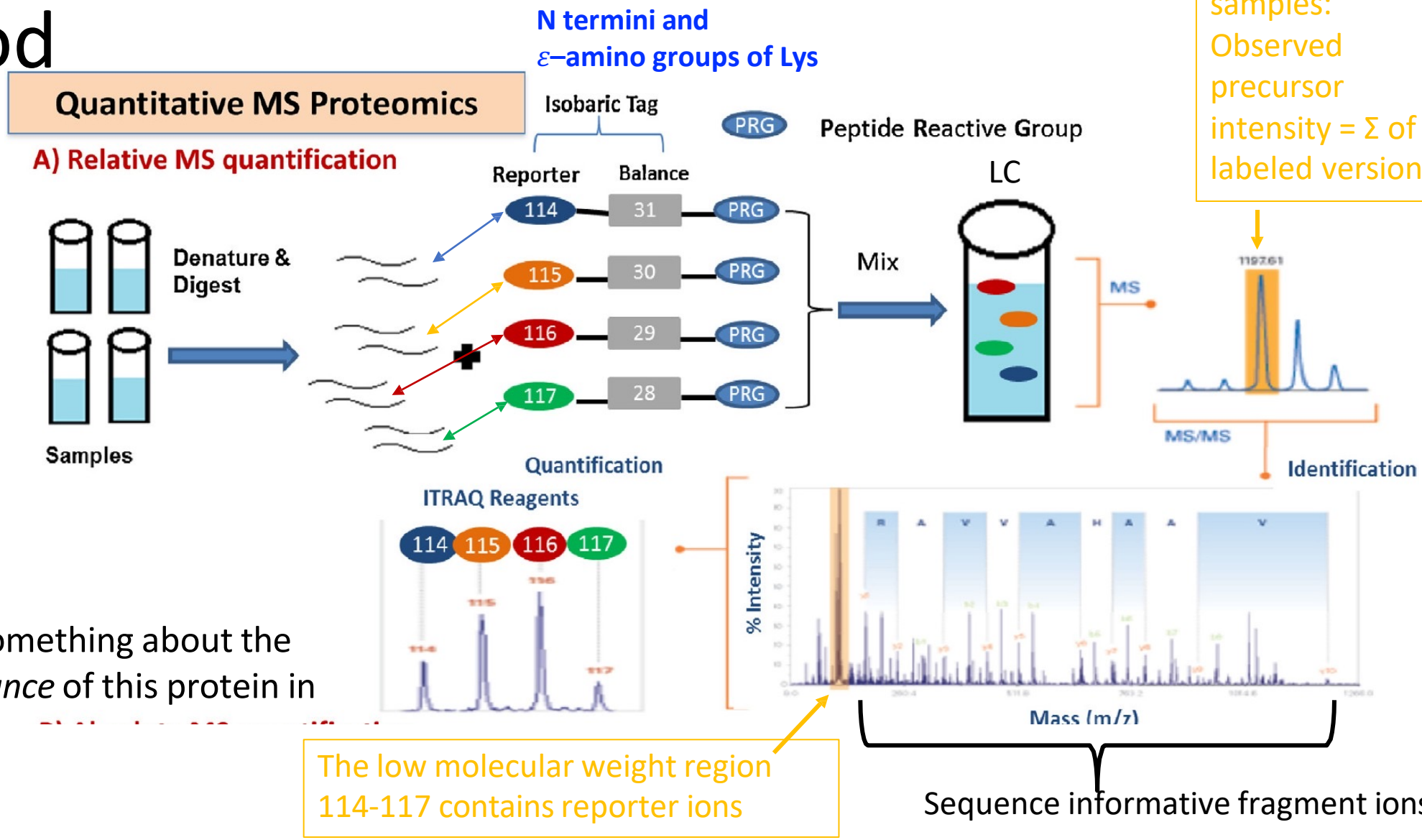
Iodoacetamide moiety, which binds to cysteine thiol group



Quantitate relative protein levels by measuring peak ratio

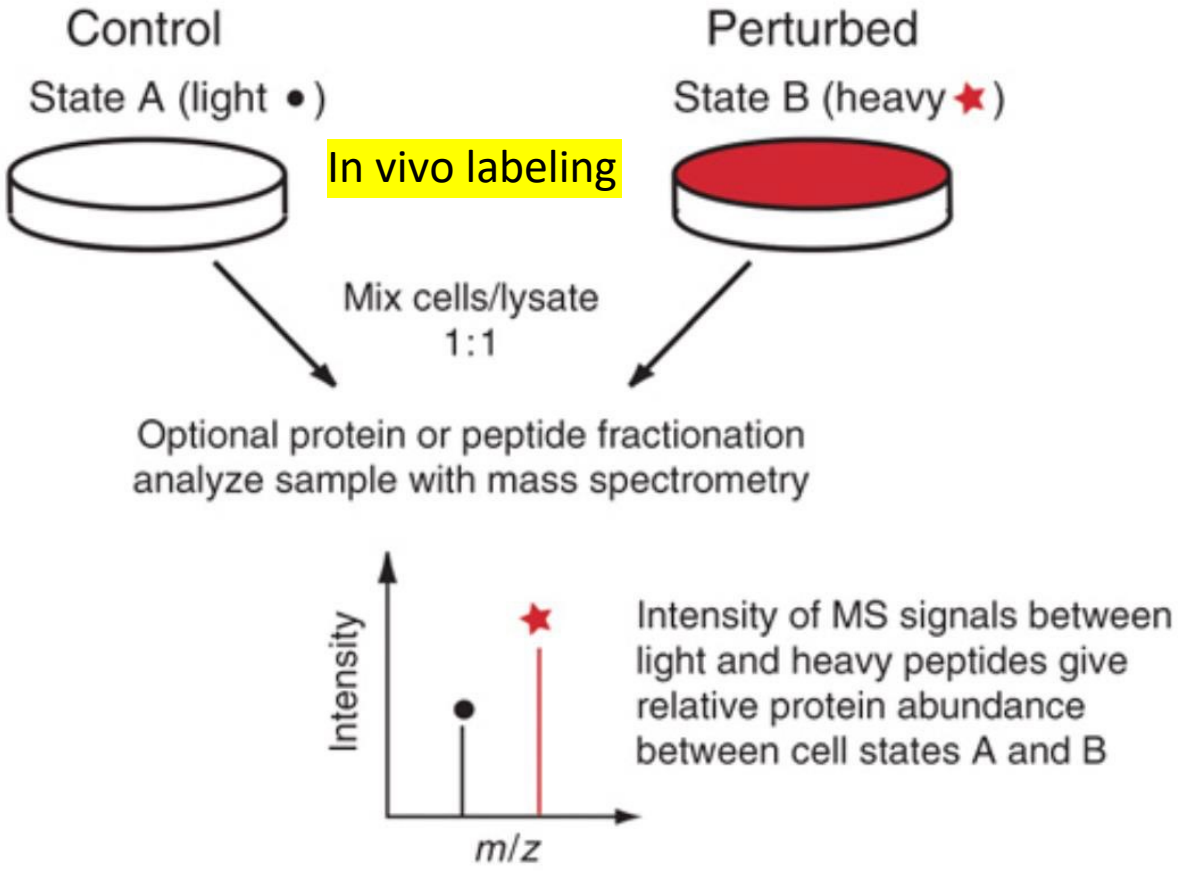
Identify peptide by sequence information

General Workflow with iTRAQ Method



- **Ratio** tells us something about the *relative abundance* of this protein in the 4 samples

SILAC (Stable Isotope Labeling with Amino Acids in Cell Culture)



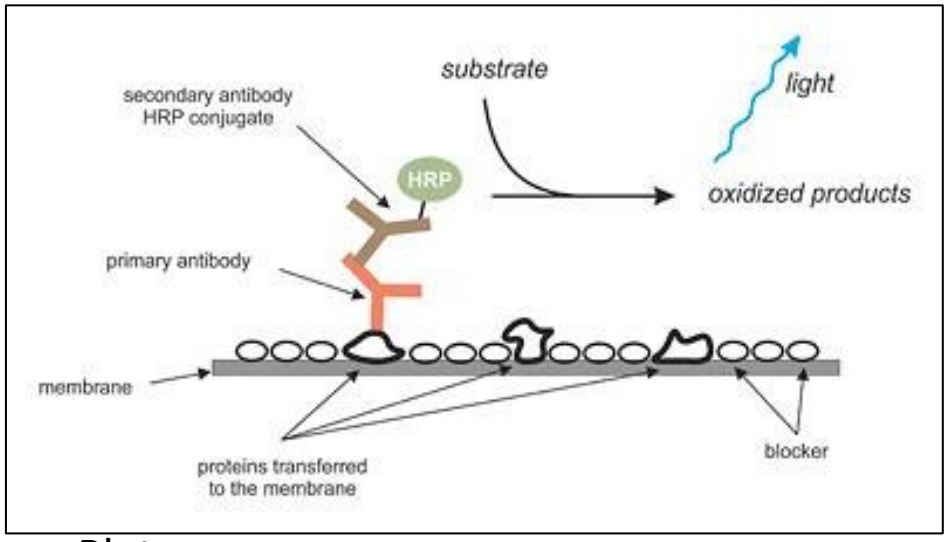
Drawbacks:

- Limited plex level (3 max)
- The method does not allow for the analysis of proteins directly from tissue.
- The stable-isotope enriched media are costly and may themselves affect cellular growth and protein production.
- The increase in nominal mass because of stable-isotope incorporation is not known until the sequence is determined.

2.Discuss how proteomics-based approaches can be used to study fundamental and applied biological problems

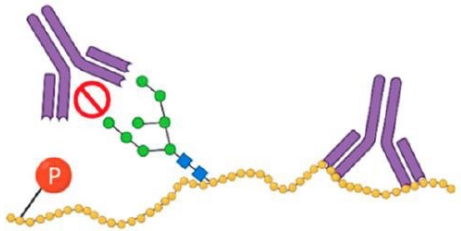
d. Targeted proteomics (identification and quantification)

Targeted Protein Quantitation: Affinity-based Approaches



Western Blot

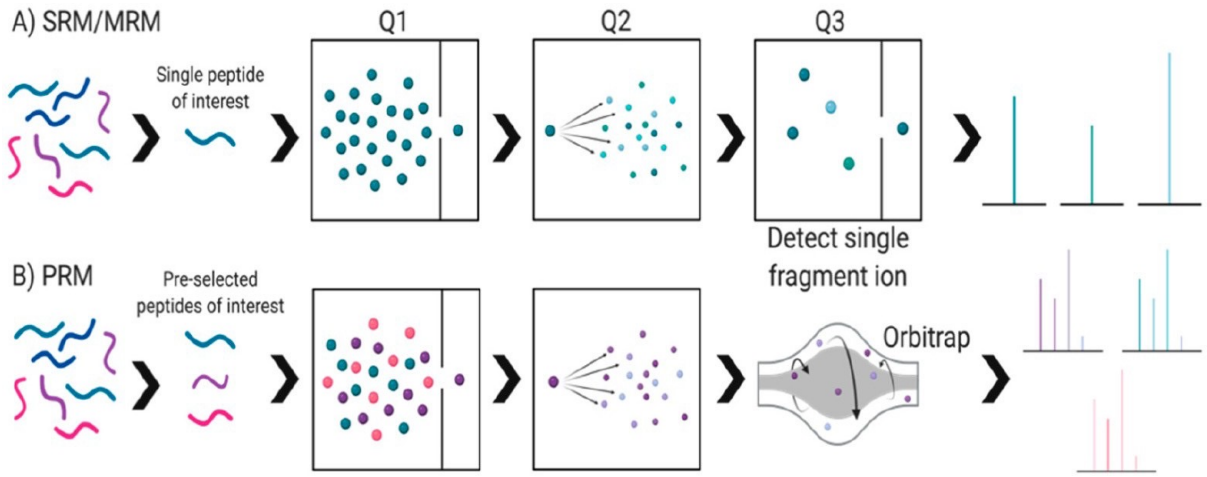
ELISA



- Use “high-quality” antibodies to quantify proteins (even low-level proteins)
- Drawbacks:
  - Antibodies do not exist for all proteins, vary widely in quality
  - Antibody validation may take months to years to validate
  - Lack throughput (quantitative capacity)
  - Fewer proteoforms–selective antibodies available
    - Antibodies generally identify amino acid epitopes on canonical protein sequences, so the PTMs or neighboring will be likely to block the epitope and prevent the binding/detection.

Targeted Protein Quantitation: MS-based Approaches

- **Selected Reaction Monitoring (SRM)/ Multiple Reaction Monitoring (MRM)**
  - Triple Quadrupole
  - Select transitions on both Q1 and Q3 prior to run



- **Parallel Reaction Monitoring (PRM)**
  - Quadrupole Orbitrap or Quadrupole TOF
  - Produce full MS2 spectra for each precursor

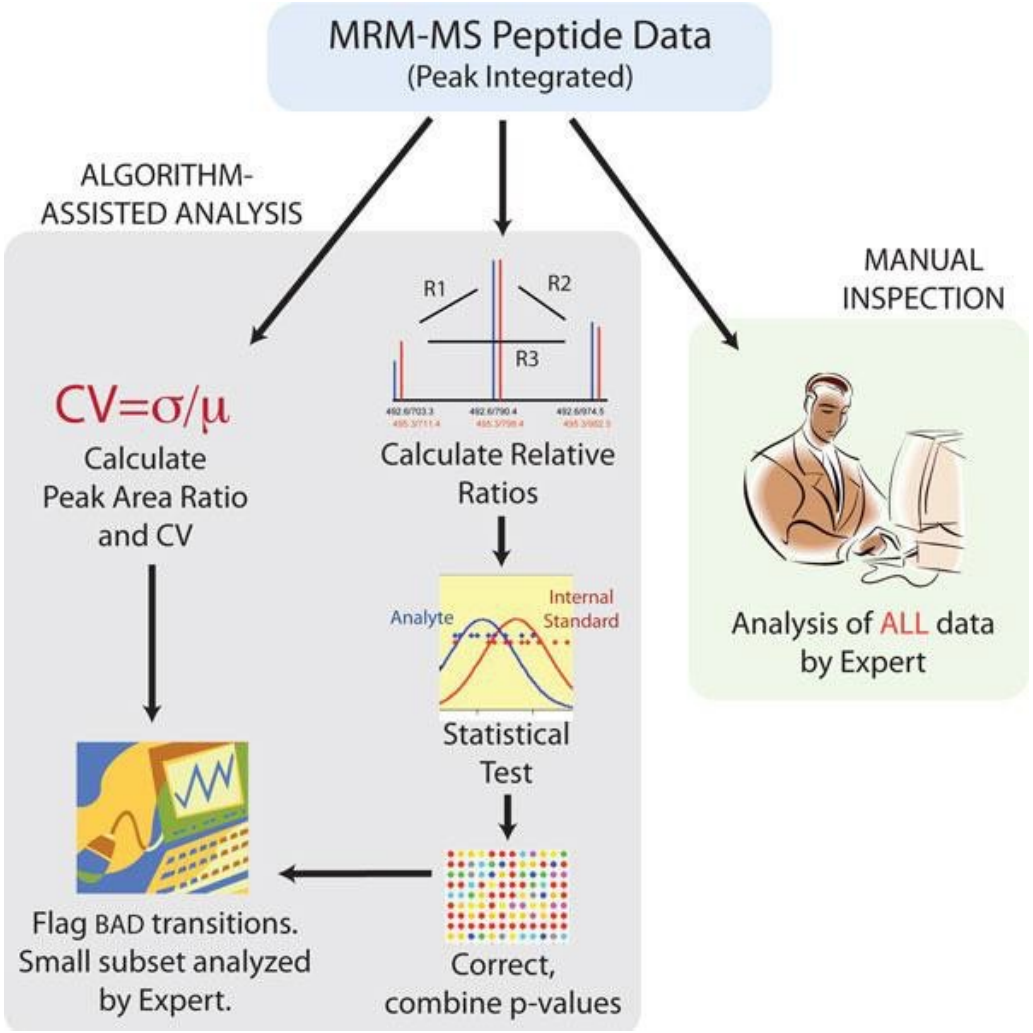
- **Data Independent Analysis (DIA) (as known as SWATH)**
  - Analyzes all peptide mass ranges within the window without pre-selection

Selection of Transitions

1. Selection of best representative (proteotypic) peptides per protein
    - Have a **unique** sequence
    - Good ionization efficiency
    - Consistently observed by LC-MS methods with a good signal-to-noise ratio
    - m/z within the range of the instrument
    - No missed cleavages
    - 8-25 amino acids
    - Not too hydrophilic (poorly retained) or hydrophobic (may stick to column)

2. Select transitions: Combination of precursor and fragment ions
    - sequence-specific to achieve maximum selectivity

Validating Transitions

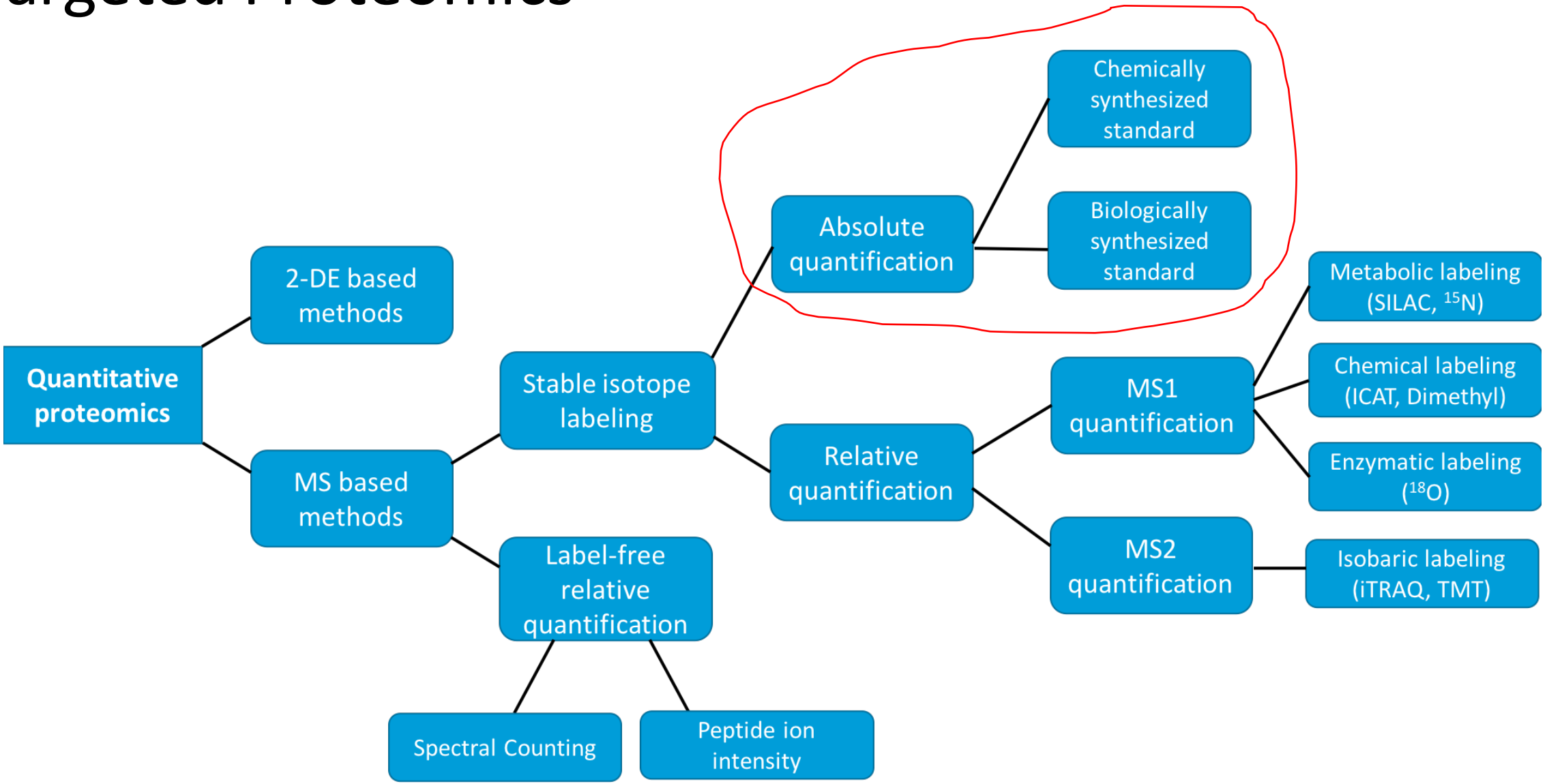


- AuDIT: Automated Detection of Inaccurate and imprecise Transitions

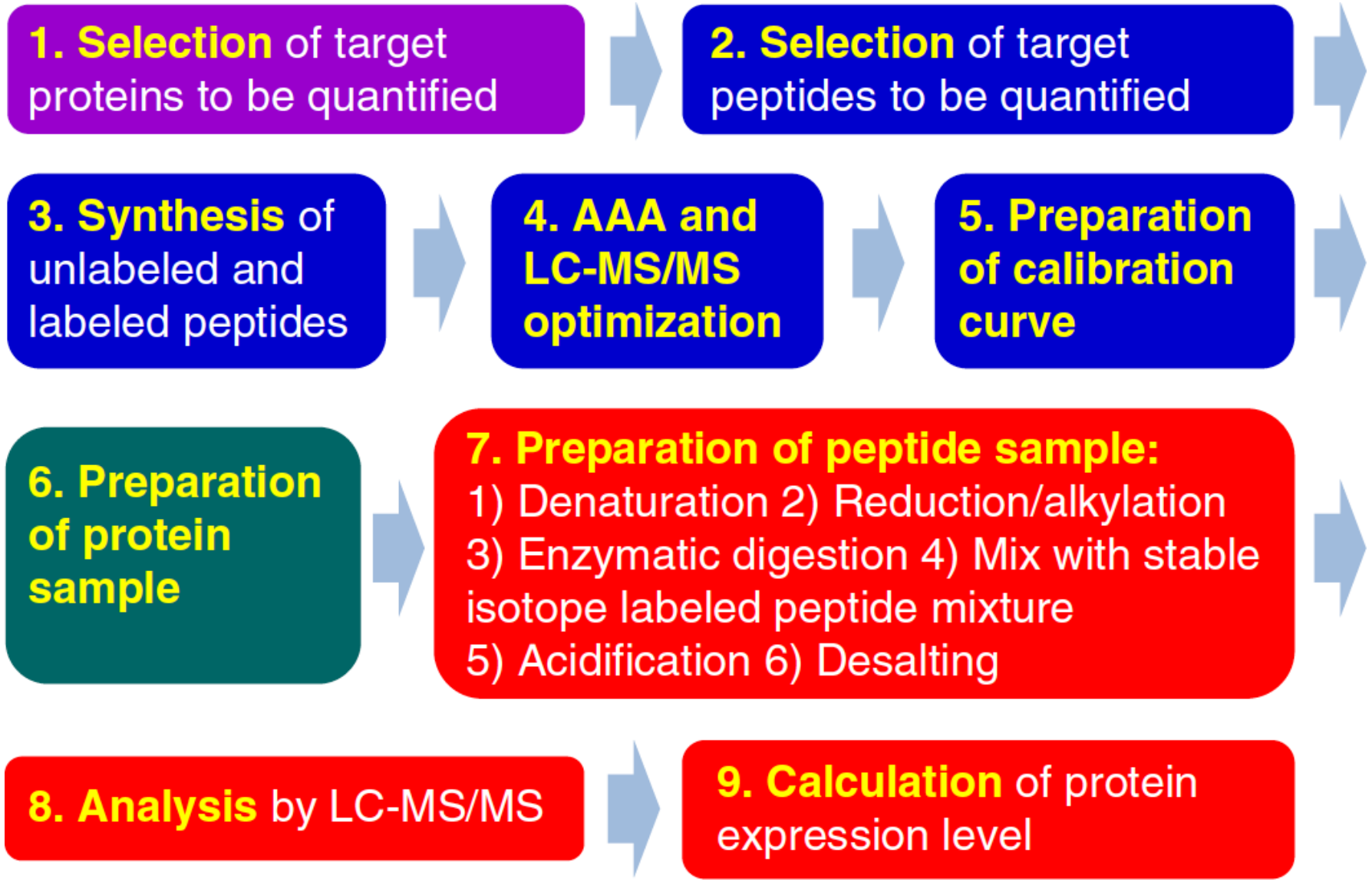
- Requirements:

1. Relative ratios of analytes over relative ratios of SIS (Stable Isotope Standard) close to 1.
2. CV measurements of replicate samples Usually, CV < 5-10% means a good reproducibility.

(Absolute) Quantitative Proteomics in Targeted Proteomics



Workflow of an SRM/MRM Experiment



Applications of SRM/MRM

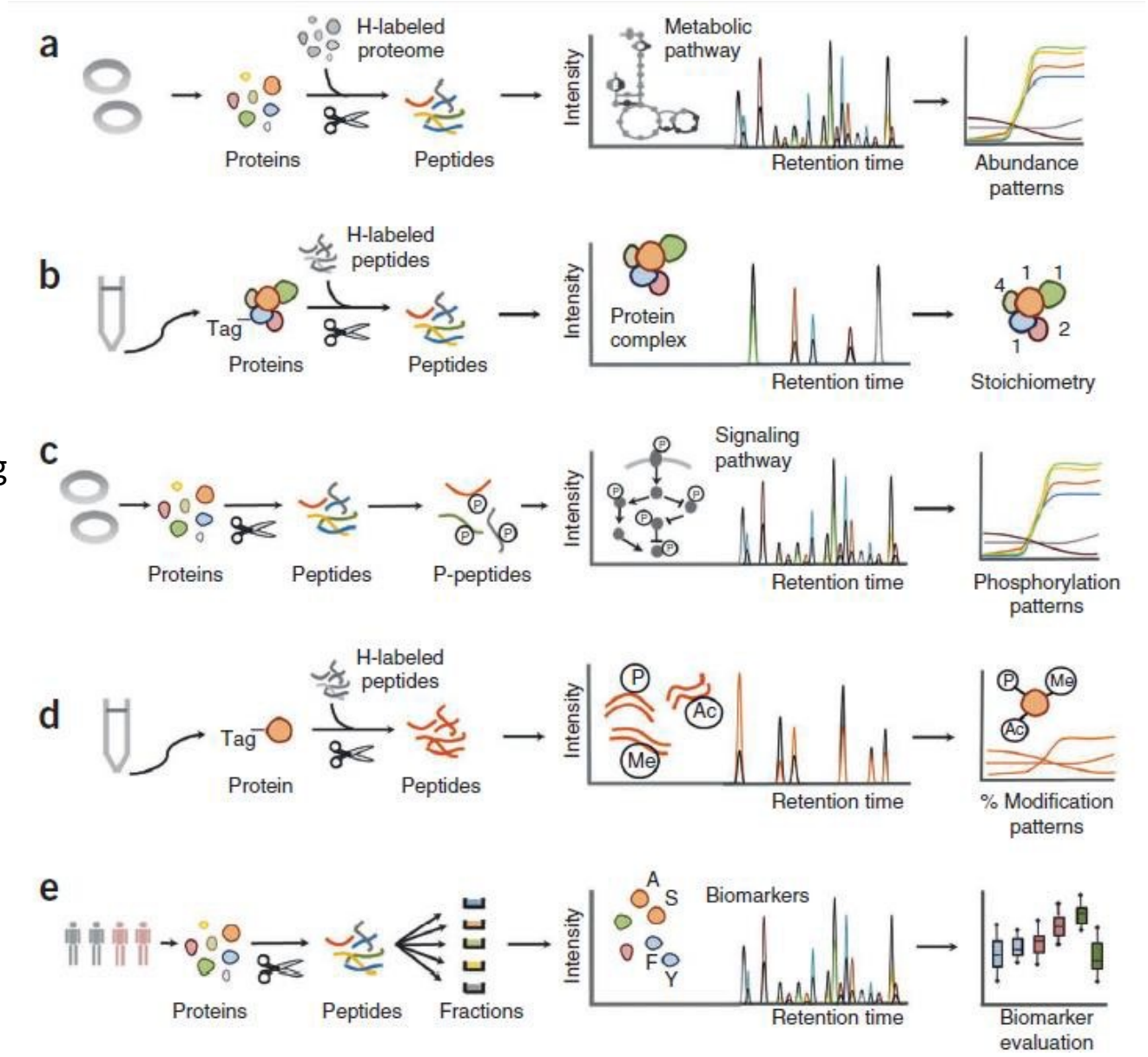
Measure protein abundance changes in a pathway to a series of perturbations

Protein complex subunit stoichiometry

Phosphorylation events in a signaling pathway are quantitatively analyzed by SRM over a time course

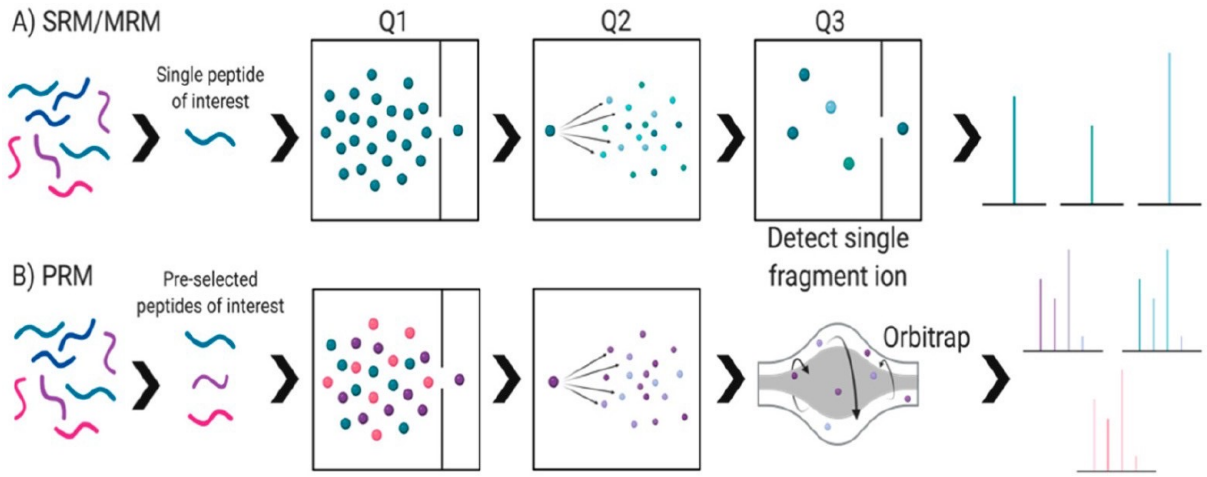
The cross-talk between different modifications of the same purified protein is analyzed

Candidate biomarkers can be verified by applying SRM assays



Targeted Protein Quantitation: MS-based Approaches

- **Selected Reaction Monitoring (SRM)/ Multiple Reaction Monitoring (MRM)**
  - Triple Quadrupole
  - Select transitions on both Q1 and Q3 prior to run



- **Parallel Reaction Monitoring (PRM)**
  - Quadrupole Orbitrap or Quadrupole TOF
  - Produce full MS2 spectra for each precursor

- **Data Independent Analysis (DIA) (as known as SWATH)**
  - Analyzes all peptide mass ranges within the window without pre-selection

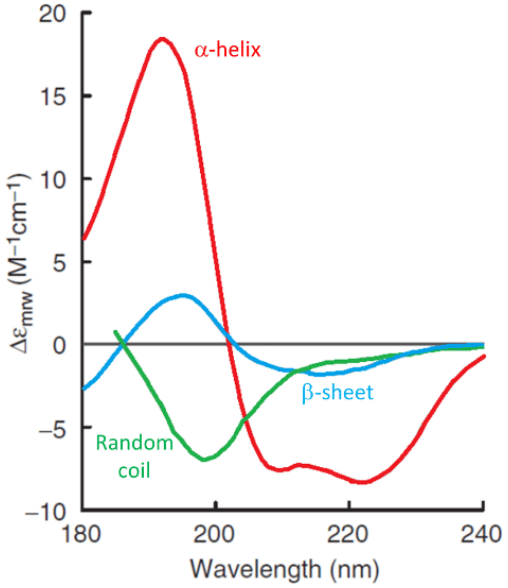
4. Understand how technologies such as nuclear magnetic resonance (NMR), X-ray crystallography and other physical methods can be used to determine the detailed fine structure of proteins

Protein Structure and Analytical Methods

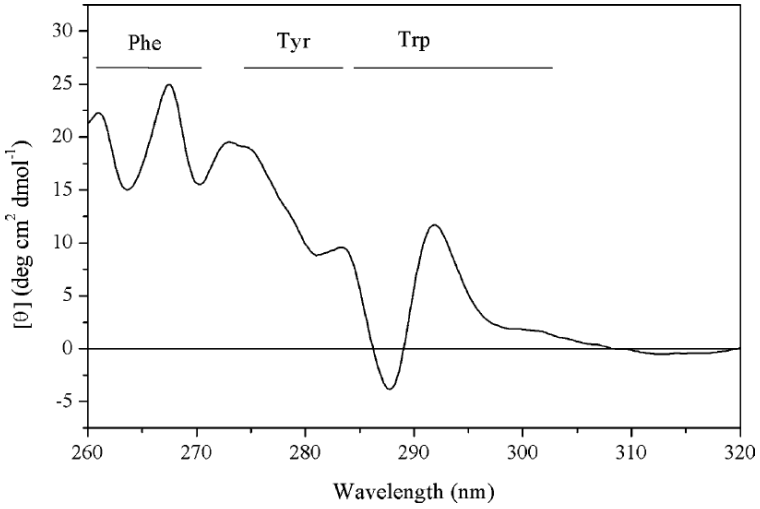
Structure	Information Gained	Analytical Techniques
Primary	Amino acid sequence	Chemical method (Edman Sequencing); Mass Spectrometry
Secondary	**Folding and composition of α helix, β sheet, random coils**	**Circular Dichroism (CD); Infrared Spectroscopy (IR)**
Tertiary	**Overall 3D structure/shape of a single polypeptide chain**	**X-ray crystallography; Nuclear Magnetic Resonance (NMR); Cryo-EM**
Quaternary	Overall 3D structure/shape of two or more polypeptide chains	Size Exclusion Chromatography (SEC);

2° Structure Determination: Circular Dichroism (CD)

- Different wavelength probe different levels of structure:
  1. **Far UV** (smaller wavelengths, 190-250 nm)
     - ➢ probing individual aa --> proportion of 2° structure in a protein
  2. **Near UV** (longer wavelengths, 260-320 nm)
     - ➢ probing larger chiral structure (e.g., helices, sheets)--> 3° structure information



- α helix : negative bands at 208 and 222 nm and positive band at 193 nm
- β sheet : negative band at 218 nm and positive band at 195 nm
- Random coil: negative band at 195 nm and low ellipticity above 210 nm



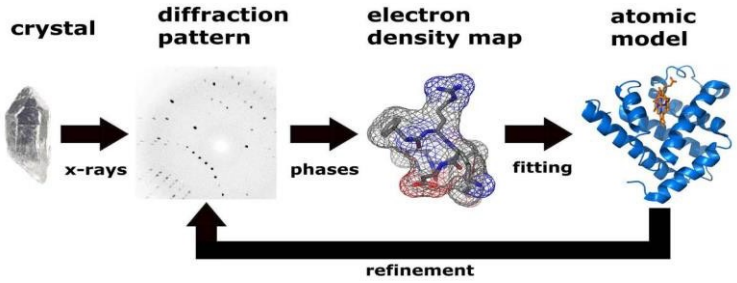
Applications of CD Spectroscopy

- Determine the 2^{o} structure content
- Comparison of wt vs mutant protein (structural integrity)
- Structural stability
  - Thermally
  - Chemically (e.g., Guanidinium chloride unfolding)
- Conformational changes due to ligand binding or protein interactions

3º Structure Determination

You need to know the **principles**, general **workflow**, **pros and cons** of the three techniques.

- X-Ray Diffraction (aka X-ray Crystallography)
  - Crystalize proteins (can be difficult)
  - Pummel them with X-rays
  - Reconstruct their shape from diffracted light



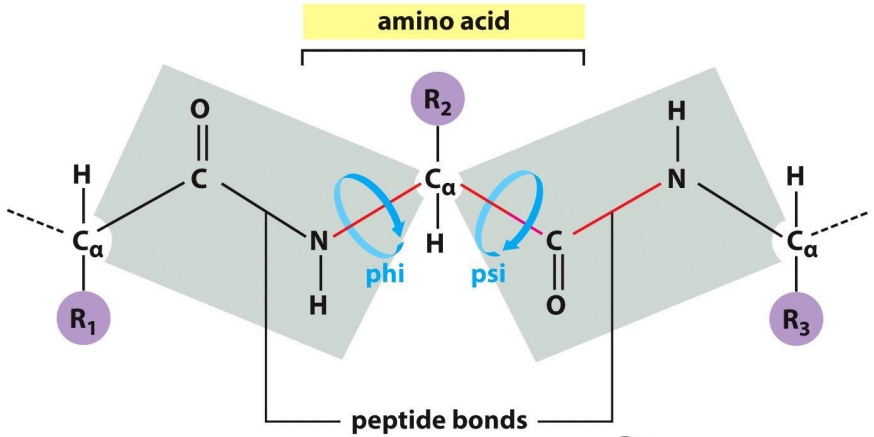
- NMR (Nuclear Magnetic Resonance Spectroscopy)
  - small proteins (<40 kD)
  - Isotope labelled

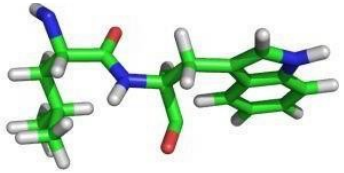
- Cryo-EM (Cryo-Electron Microscopy)
  - Larger molecules (>50 KD)
  - Does not require protein crystals
  - transforming the images into sharper molecular structures.

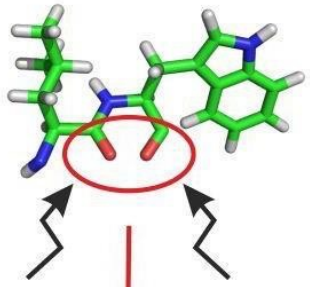
Things to Know for X-ray crystallization

1. What is X-ray Crystallography. Know the general process.
2. What are protein crystals? What's the common method used for making crystals? Why are protein crystals fragile?
3. Understand the protein phase diagram.
4. Phase problem. How to solve?
5. Resolution. (know high/low resolution)
6. Know how to assign atoms based on environment.
7. Occupancy; b factor
8. Understand Ramachandran Plot and know its application.
9. Common parameters used to judge the model.

Ramachandran Plot

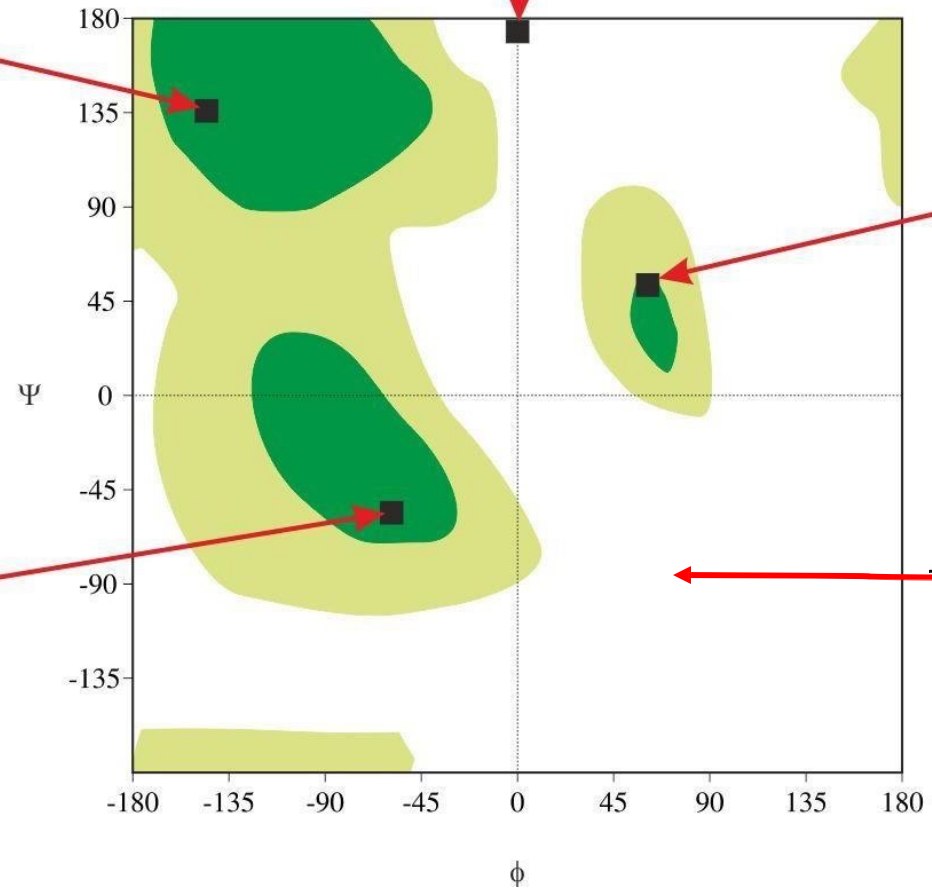


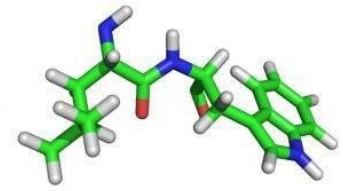

antiparallel β-sheet

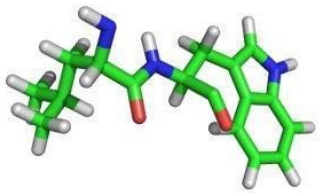

steric distortion

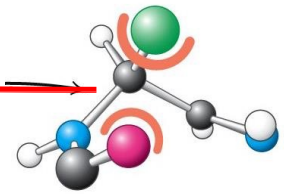
The **green/yellow regions** correspond to conformations where there are no steric clashes

White regions: sterically disallowed for all amino acids except glycine




left-handed α-helix


right-handed α-helix


($\phi = 90°, \psi = -90°$)
Disfavored

Things to know for NMR

Except principles, general workflow, pros and cons, you need to know

how to interpret 2D NMR spectrum (COSY, TOCSY, NOSEY).

- Refer to Lecture 10-1 and 10-2.

5. Explain how the knowledge of protein structure can be used to explain function

Protein-Ligand Binding: Constants: Review

$$P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL$$

[P]=concentration of free protein
[L]=concentration of free ligand

- $K_a = \frac{[PL]}{[P][L]} = \frac{k_{on}}{k_{off}}$ K_a is the association constant Units = M^{-1}

- $K_D = \frac{[P][L]}{[PL]} = \frac{k_{off}}{k_{on}}$ K_D is the dissociation constant Units = M

K_a is a direct measure of the strength of binding: the higher the value of Ka, the stronger the association.

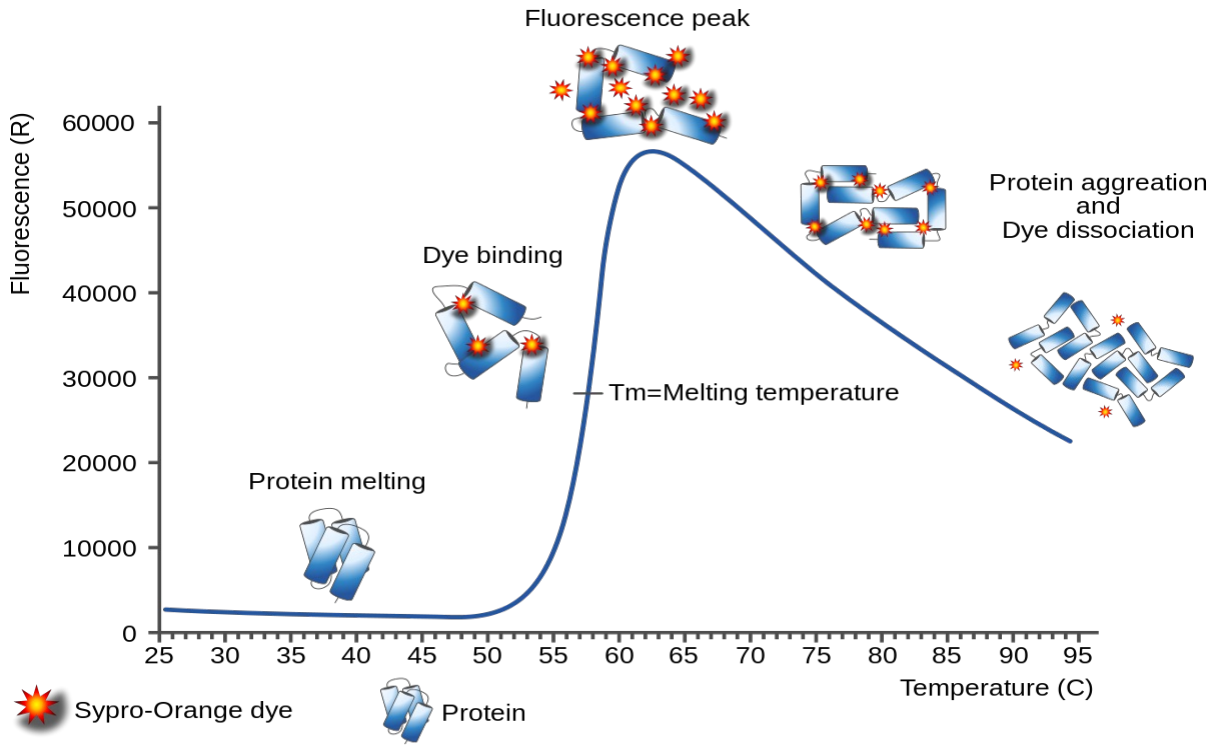
- Tight binding = High values of K_a and low values of K_D
- Weak/loose binding = Low values of K_a and high values of K_D

Measuring Protein-Protein/Ligand Interactions: Review

- **Thermal Shift Assay**
  - Increased thermal stability (T_m) upon ligand binding
- **Circular Dichroism (CD)**
  - Gain information about the 2^o and conformational (3^o) change of proteins
- **Fluorescence Resonance Energy Transfer (FRET)**
  - FRET can be used to study Protein-Protein/Ligand interactions *in vivo.*
  - e.g., CFP and YFP
- **Isothermal Titration Calorimetry (ITC)**
- **Surface Plasmon Resonance (SPR)**

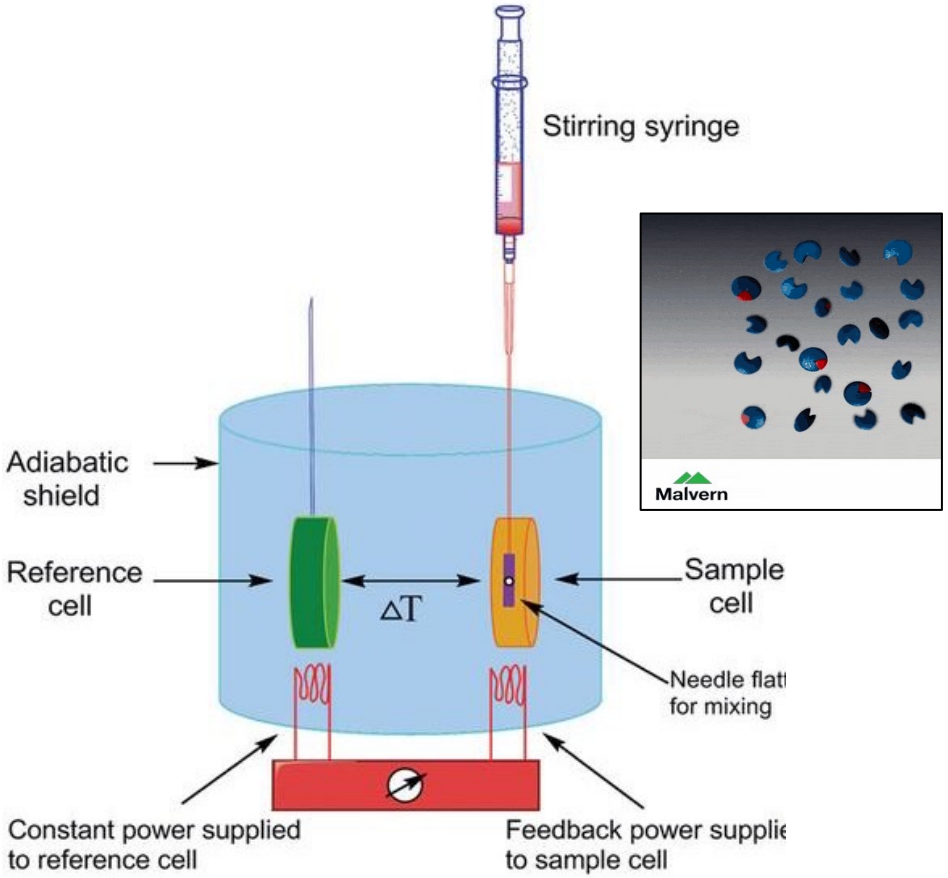
Thermal Shift Assay

- *Principle*: Protein hydrophobic core is buried. When the protein solution is heated up, the protein is denatured, and hydrophobic region is exposed to SYPRO™ orange fluorescent dye. The fluorescent signal can be determined in real time instrument.



ITC Principle

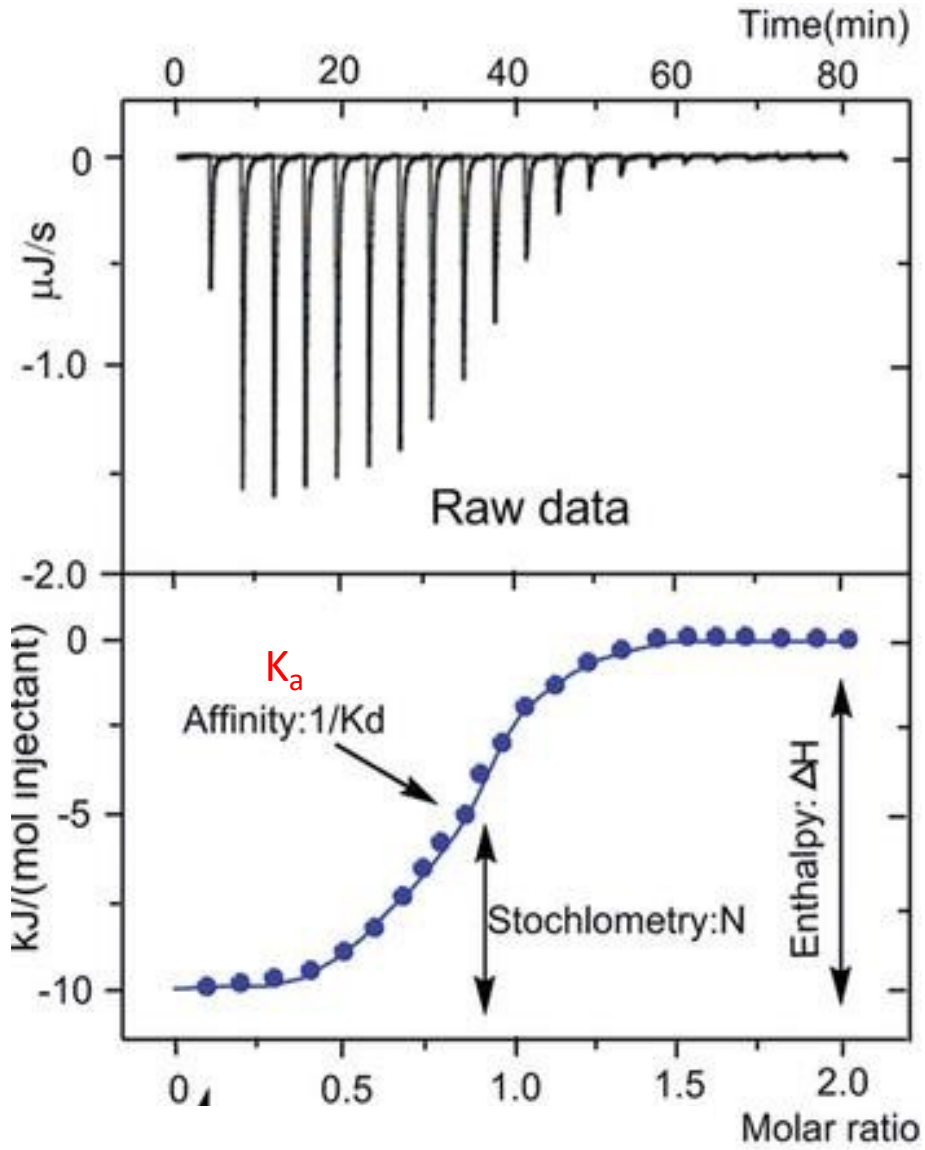
- ITC microcalorimeter measures **heat** released or absorbed during gradual titration of ligand into the sample cell.



$$\Delta G = -RT \ln K_a$$

$$= \Delta H - T\Delta S$$

ITC thermogram

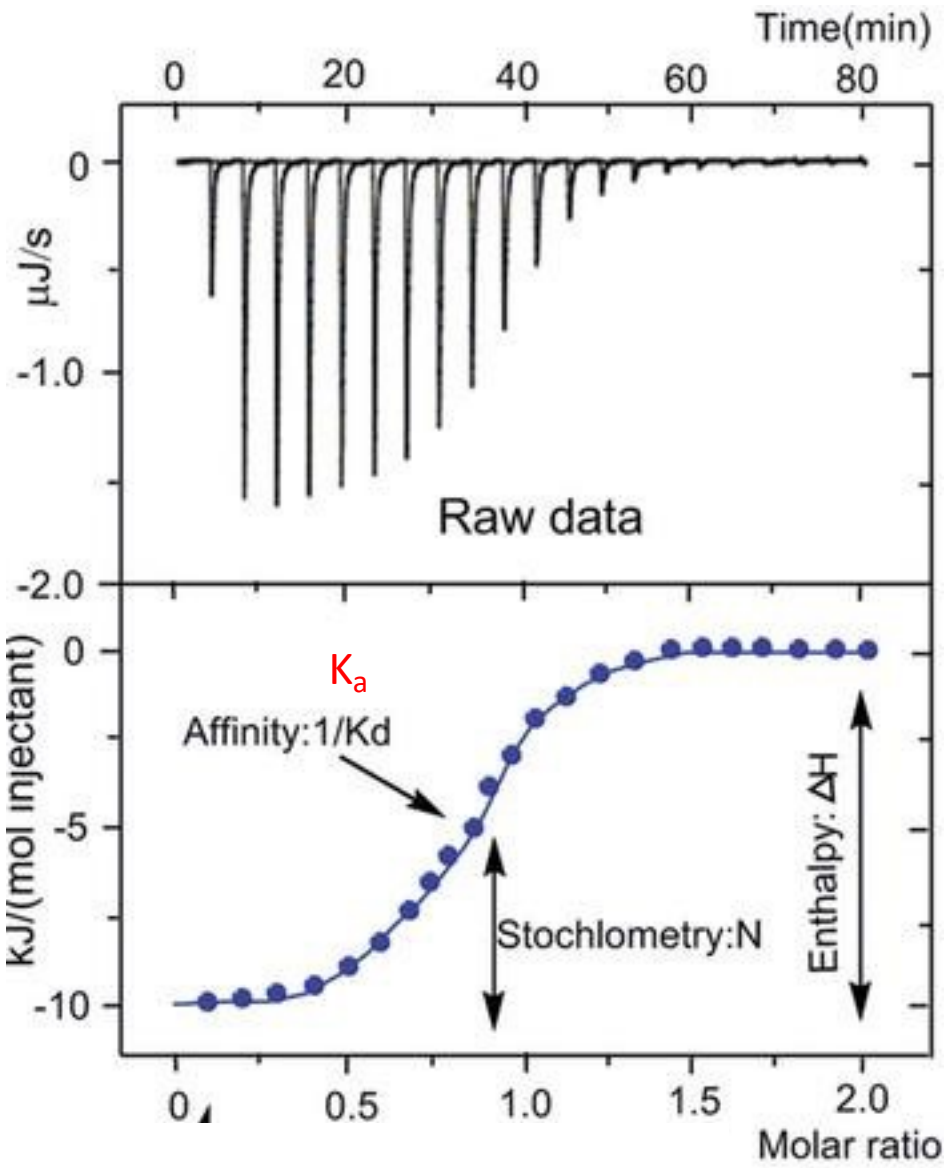


ITC Principle:

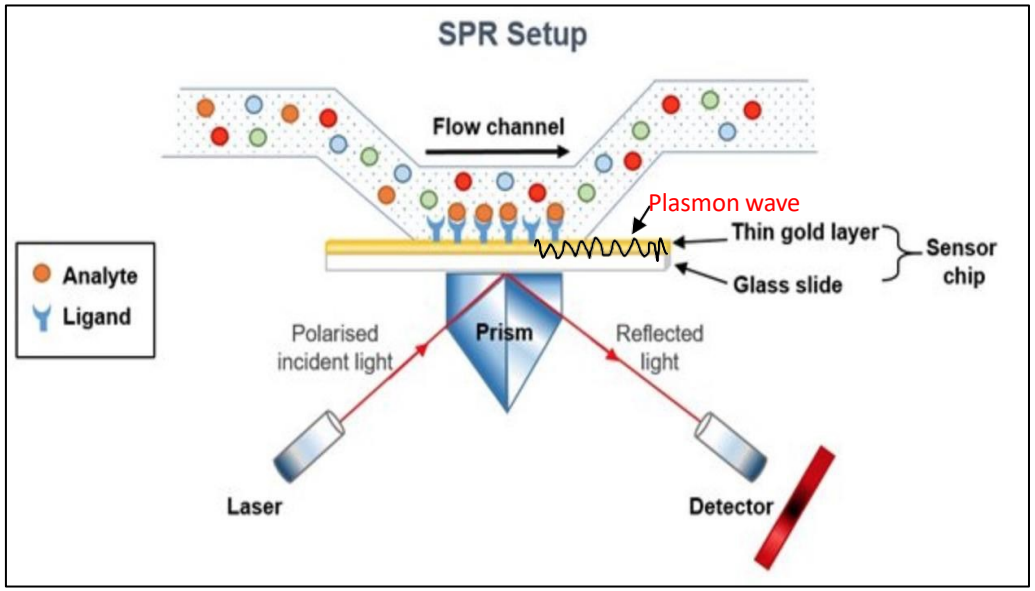
$\Delta G = -RT \ln K_a$

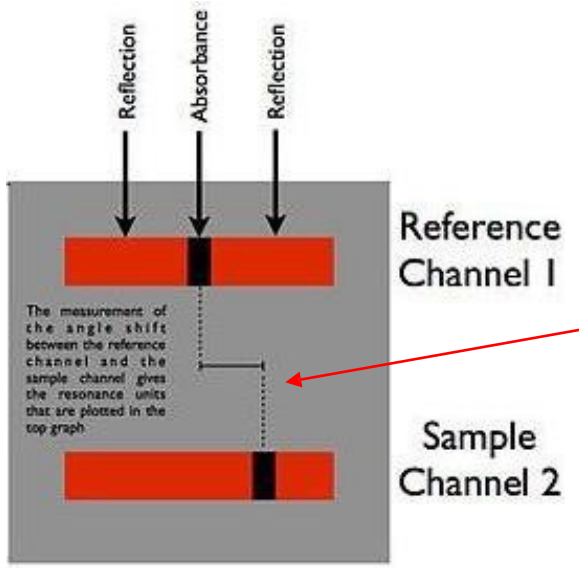
$= \Delta H - T\Delta S$

ITC thermogram

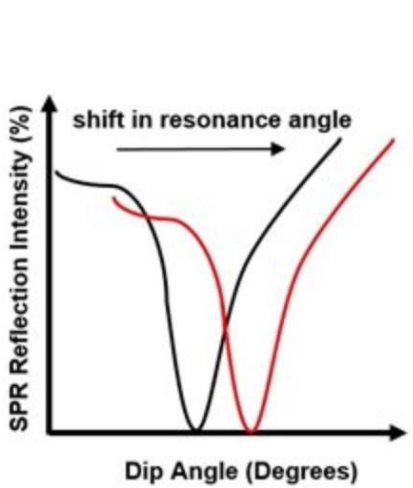


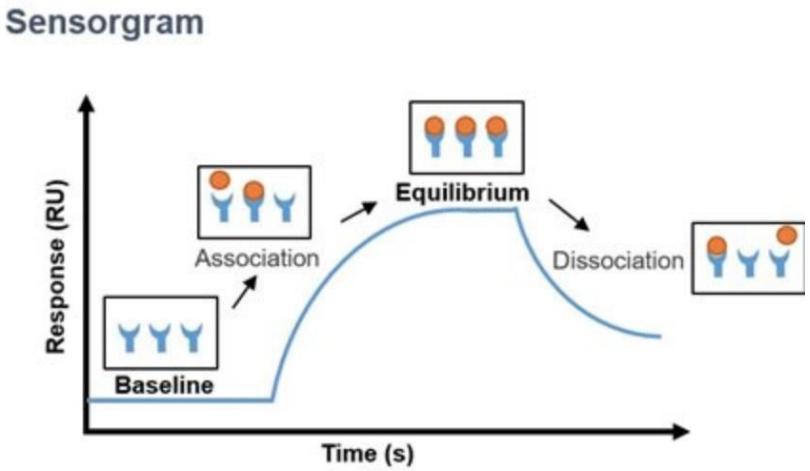
Surface Plasmon Resonance (SPR)

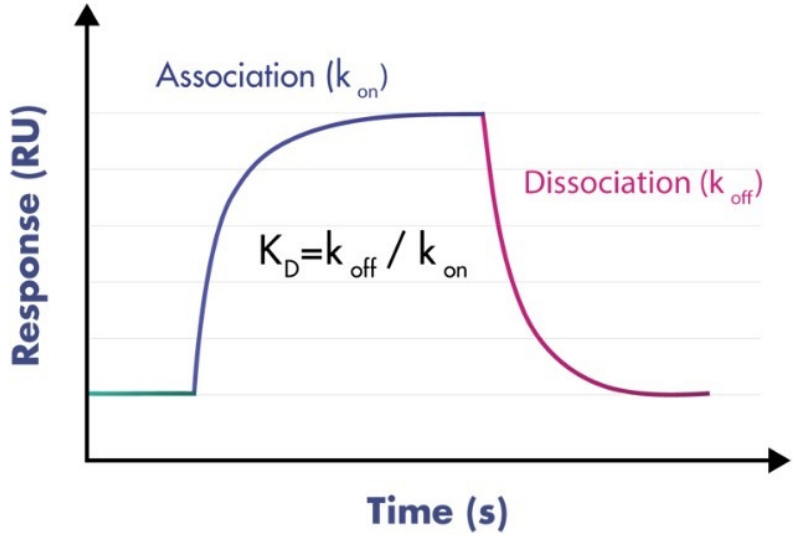


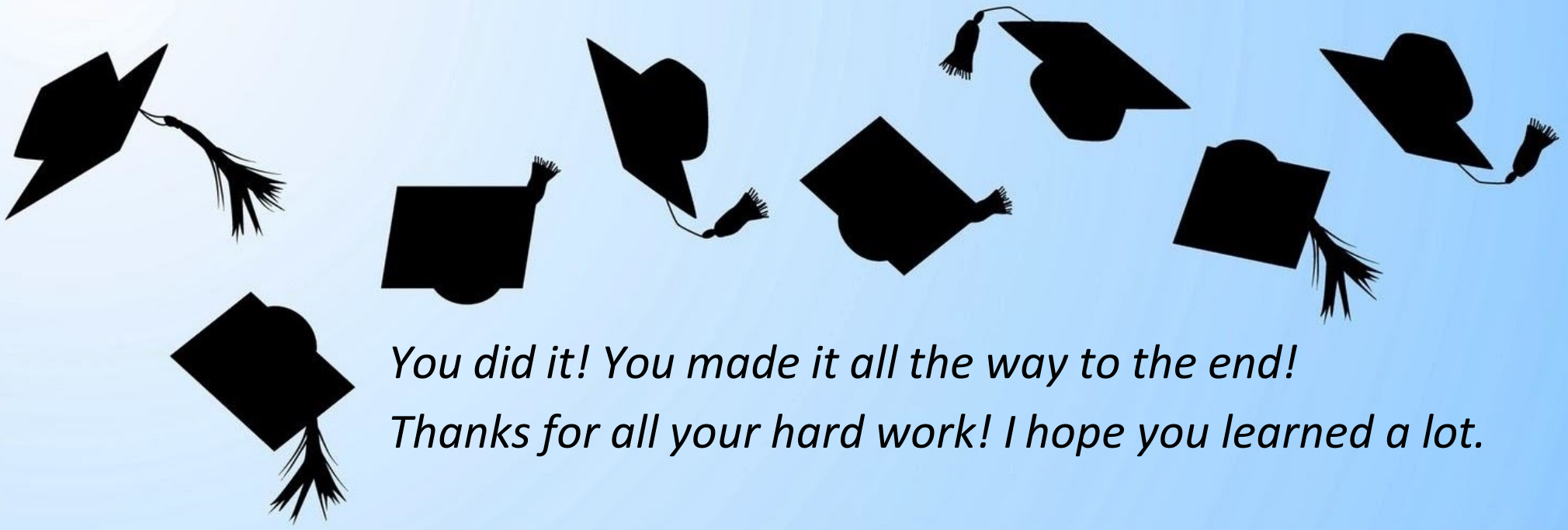


Angle of the reflection changed upon ligand binding









You did it! You made it all the way to the end!

Thanks for all your hard work! I hope you learned a lot.

Best of luck with your oral presentations and finals.

Congratulations on your graduation and best wishes for your next adventure! ☺

-Dr. Han