Class Today

- Finish lecture 10
- CryoEM
- Lecture 11: Protein-ligand interactions

- Announcement
  1. Final exam: Open-book exam, **Tuesday (June 4th) at 2 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.

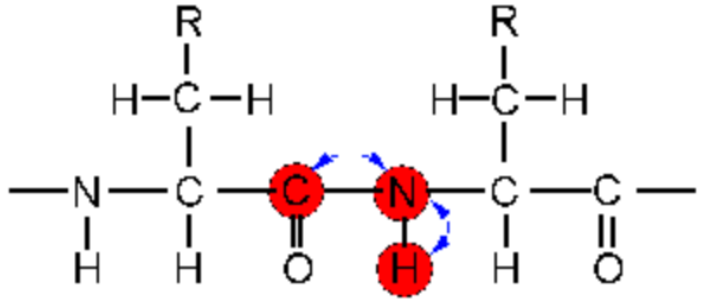
- Assignment:
  1. Coursework 2 is due on Tuesday (May 7th) at 5 pm

LECTURE 10-3: NUCLEAR MAGNETIC RESONANCE (NMR) SPECTROSCOPY

Triple Resonance Experiments: Review

HNCO

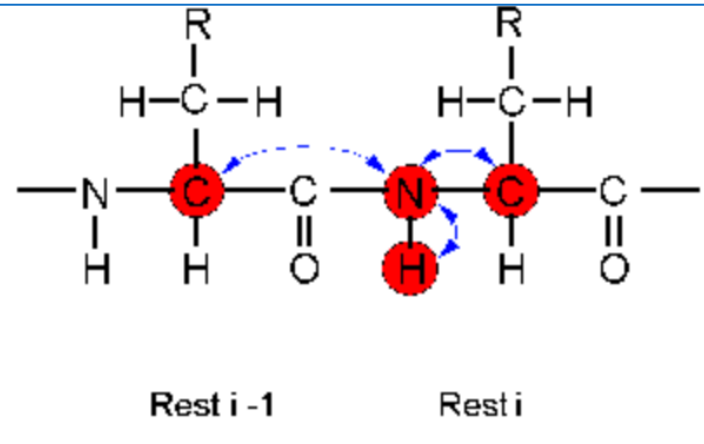
$^{1}H^{N}_{i}$ - $^{15}N_{i}$ - $^{13}CO_{i-1}$

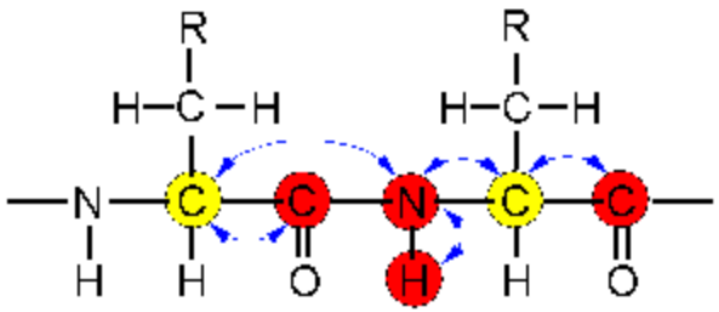


HNCA

$^{1}H^{N}_{i}$ - $^{15}N_{i}$ - $^{13}C^{\alpha}_{i}$

$^{1}H^{N}_{i}$ - $^{15}N_{i}$ - $^{13}C^{\alpha}_{i-1}$





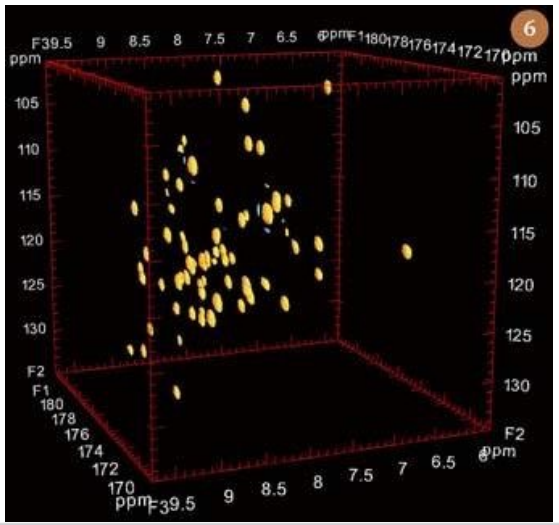
HN(CA)CO

$^{1}H^{N}_{i}$ - $^{15}N_{i}$ - $^{13}CO_{i}$

$^{1}H^{N}_{i}$ - $^{15}N_{i}$ - $^{13}CO_{i-1}$

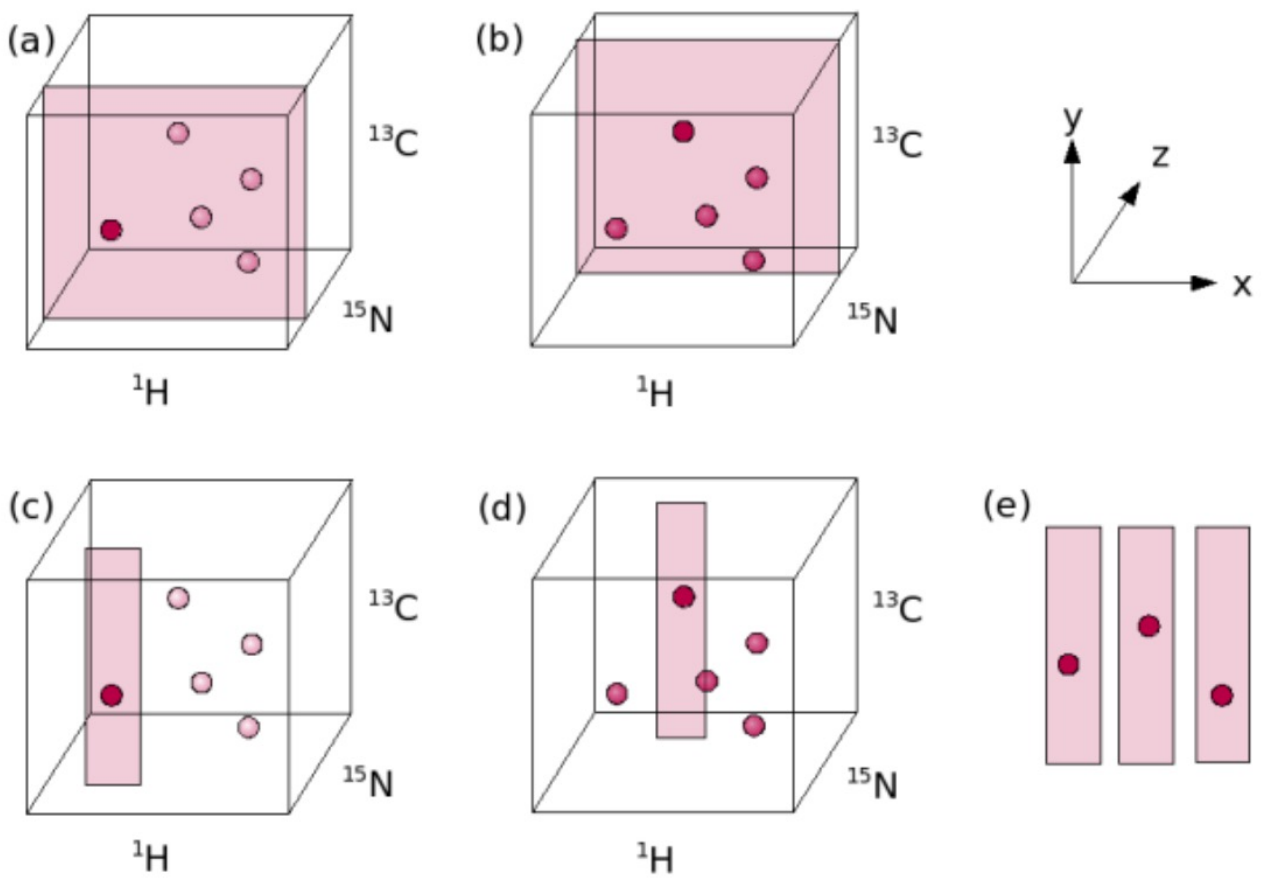
Protein backbone: HNCO, HNCA, HN(CO)CA

- 'HNCO‘ correspond to the chemical shift of the nitrogen atom to which this hydrogen is attached (labelled by ^{15}N), and to that of the adjacent C=O group in the peptide bond (labelled by ^{13}C).
- HNCO 'connects together' the proton, nitrogen and carbon resonances in a peptide bond.
- Thus, each of the peptide bonds in a protein appears individually as a small sphere in the HNCO experiment.
- HNCO, HNCA and HN(CO)CA, which when used together with HCNO allow the assignment of all the carbon, hydrogen and nitrogen resonances in the backbone of a protein.



Visualizing 3D Spectra: Review

- A common way of visualizing 3D spectra is as so-called **Strips**.
- If you start with an HSQC where each peak is defined by a ^{1}H and ^{15}N value, you can then pick out strips for each HSQC peak and then lay them next to one another for easy comparison (e). For protein backbone assignment this is particularly useful.



Workflow of Protein Structure Determination by NMR

Protein sample: ~0.5-1 mM or greater in ~0.5 mL solution
Stable for several days at desired temp.

^{15}N and/or ^{13}C, ^{2}H labeled protein

⇩

NMR spectroscopy

⇩

Sequence-specific & Total resonance **Assignment**

2^{o} structure identification

⇩

Determining conformational **constraints**

-**NOE**: distance restraints
-**J coupling**: dihedral angle
-Others: H-bonds

⇩

Tertiary long range distance **restraints**

⇩

Calculation of initial structure

⇩

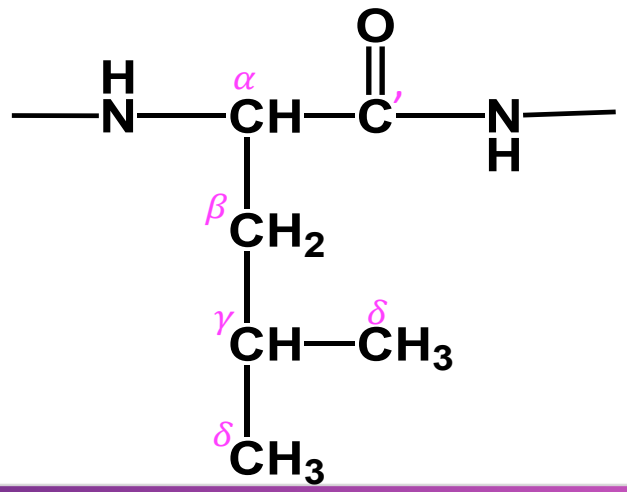
Structure refinement

Assignment of spectra

- The aim of the analysis of NMR spectra is to extract all available information about **interatomic distances** and **torsion angles**.
- Initial stage: Assign each resonance to a specific nucleus
  1. Backbone: C_α, C_β, C', N, H_N

     HSQC, CBCANH, CBCACONH, HN(CA)CO, HNCO, HN(CO)CA, HNCA
  2. Side chain: all others (especially CHs)

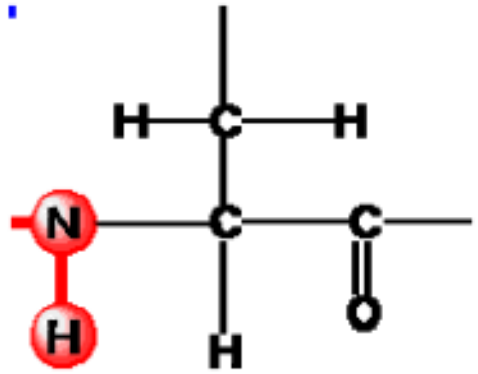
     TOCSY-HSQC, HCCCONH, CCCONH, HCCH-TOCSY



1. Backbone: **C_α, C_β,** C', **N**, **H_N**

   HSQC, CBCANH, CBCACONH,
   HN(CA)CO, HNCO, HN(CO)CA, HNCA

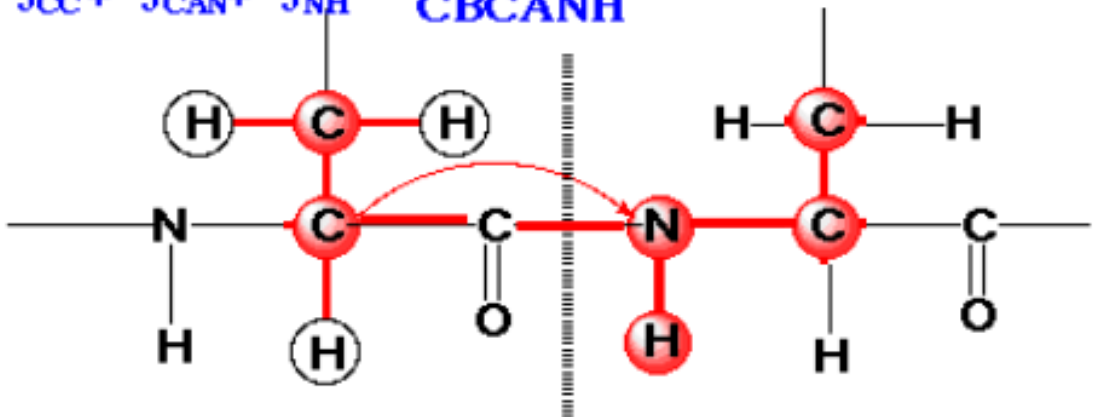
^1H-^{15}N HSQC



$^1J_{CH}$ + $^1J_{CC}$ + $^1J_{CAN}$ + $^1J_{NH}$

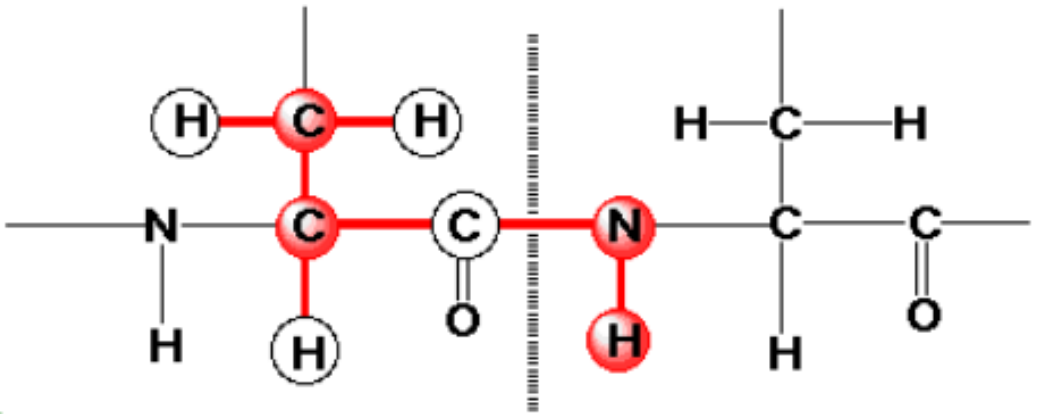
$^1J_{CH}$ + $^1J_{CC}$ + $^2J_{CAN}$ + $^1J_{NH}$

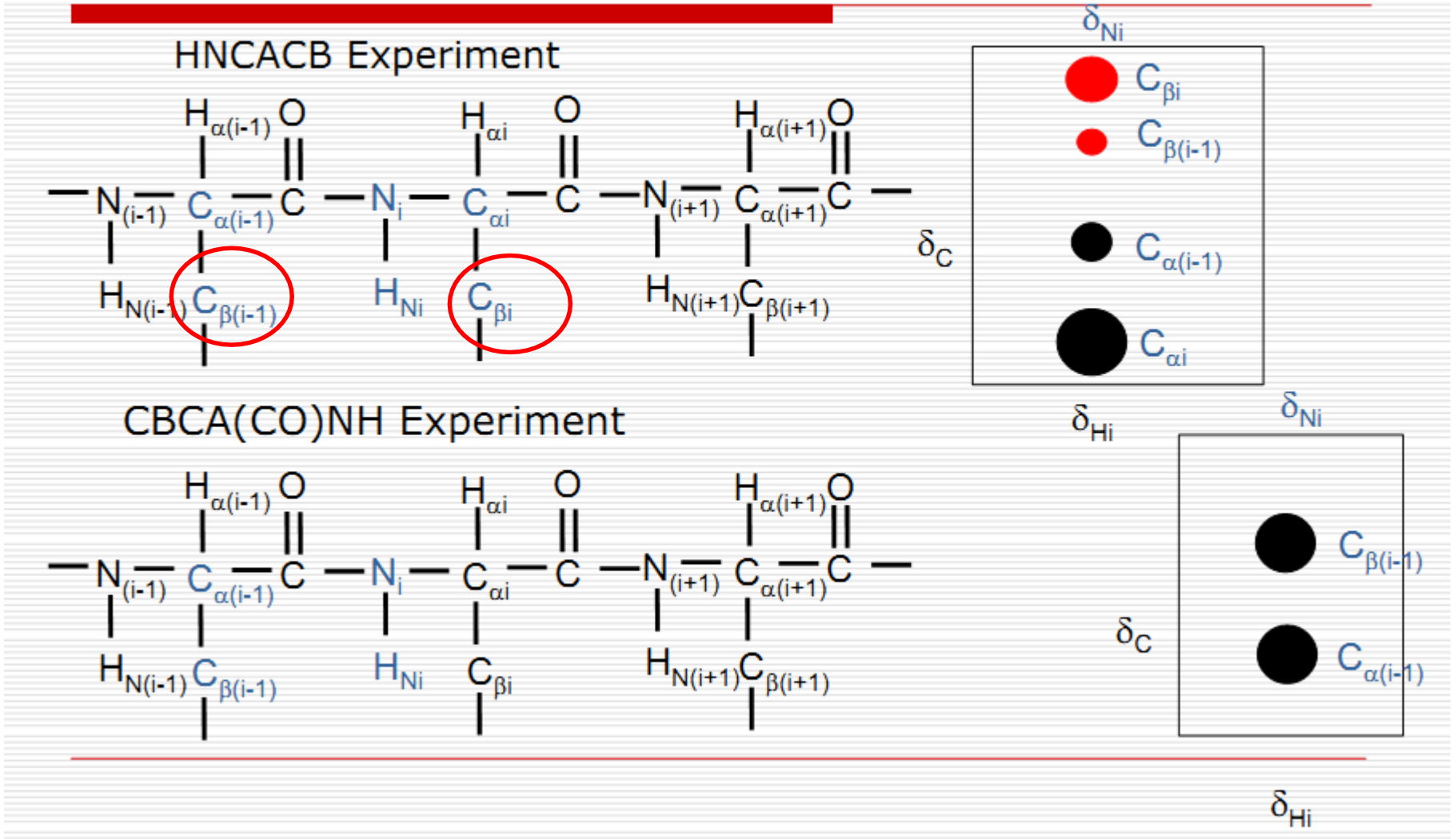
CBCANH



$^1J_{CH}$ + $^1J_{CC}$ + $^1J_{CCN}$ + $^1J_{NH}$

CBCA(CO)NH



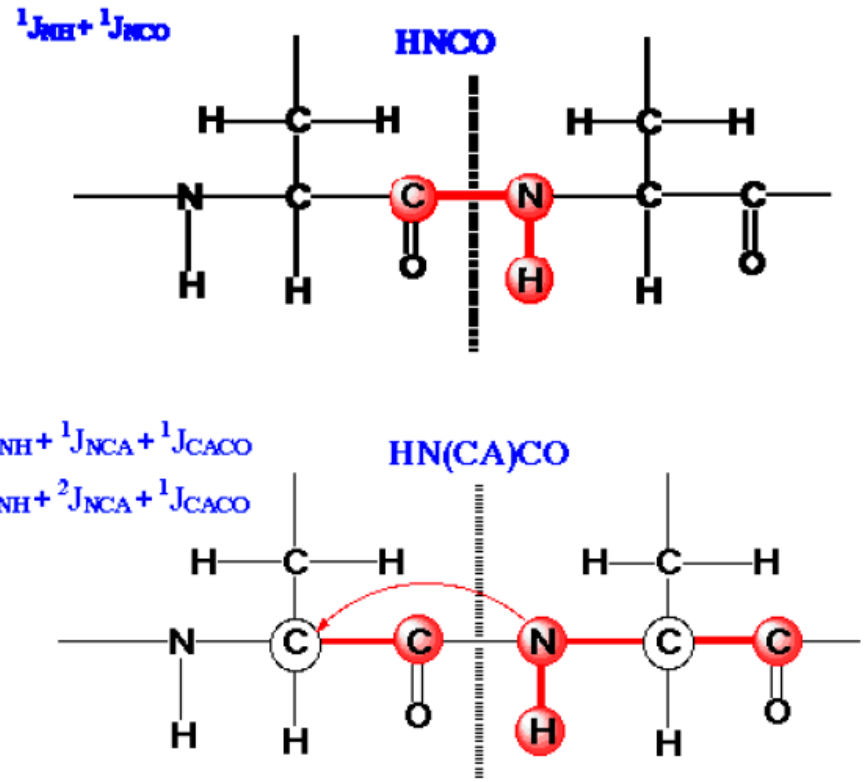


- Sequential assignment
- Chemical shifts of C_α, C_β, NH

- ^{13}C chemical shifts are amino acid specific

Backbone: Ca, Cb, C', N, NH

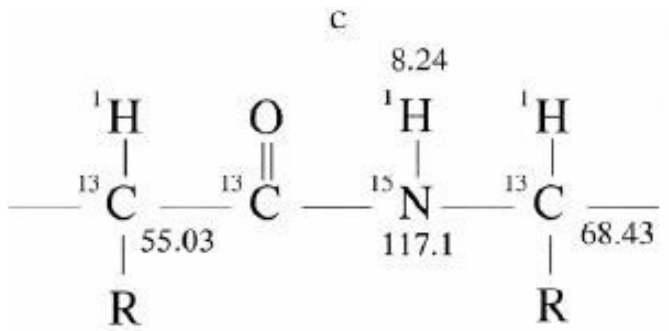
HSQC, CBCANH, CBCACONH, **HN(CA)CO, HNCO,** HN(CO)CA, HNCA



➔confirm the previous sequential assignment

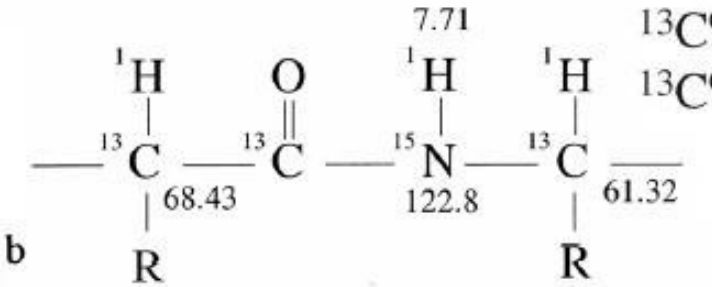
➔adding CO chemical shifts

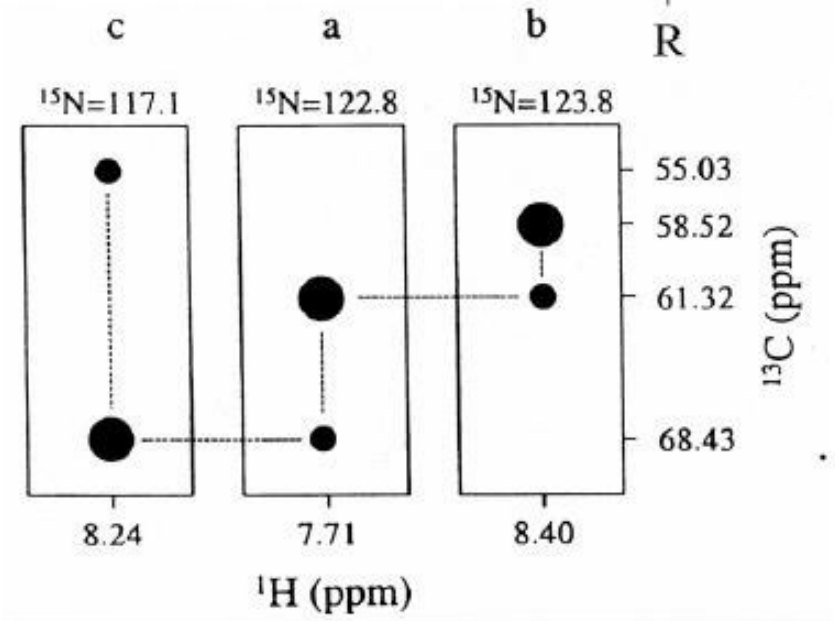
Linking the dots from every NMR

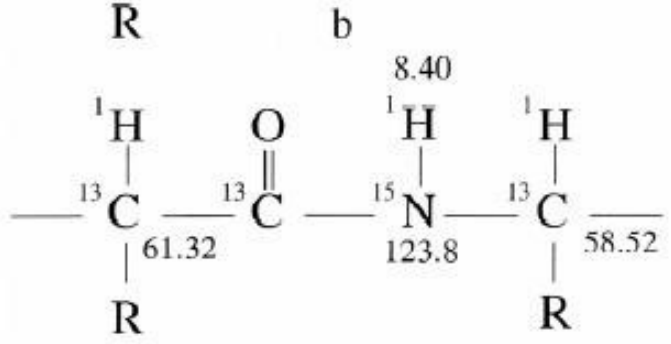


• link the correlated shifts numerically….

	c	a	b
$^{1}H^{N}_{i}$	8.24	7.71	8.40
$^{15}N_{i}$	117.1	122.8	123.8
$^{13}C^{\alpha}_{i}$	68.43	61.32	58.52
$^{13}C^{\alpha}_{i-1}$	55.03	68.43	61.32







…. or visually

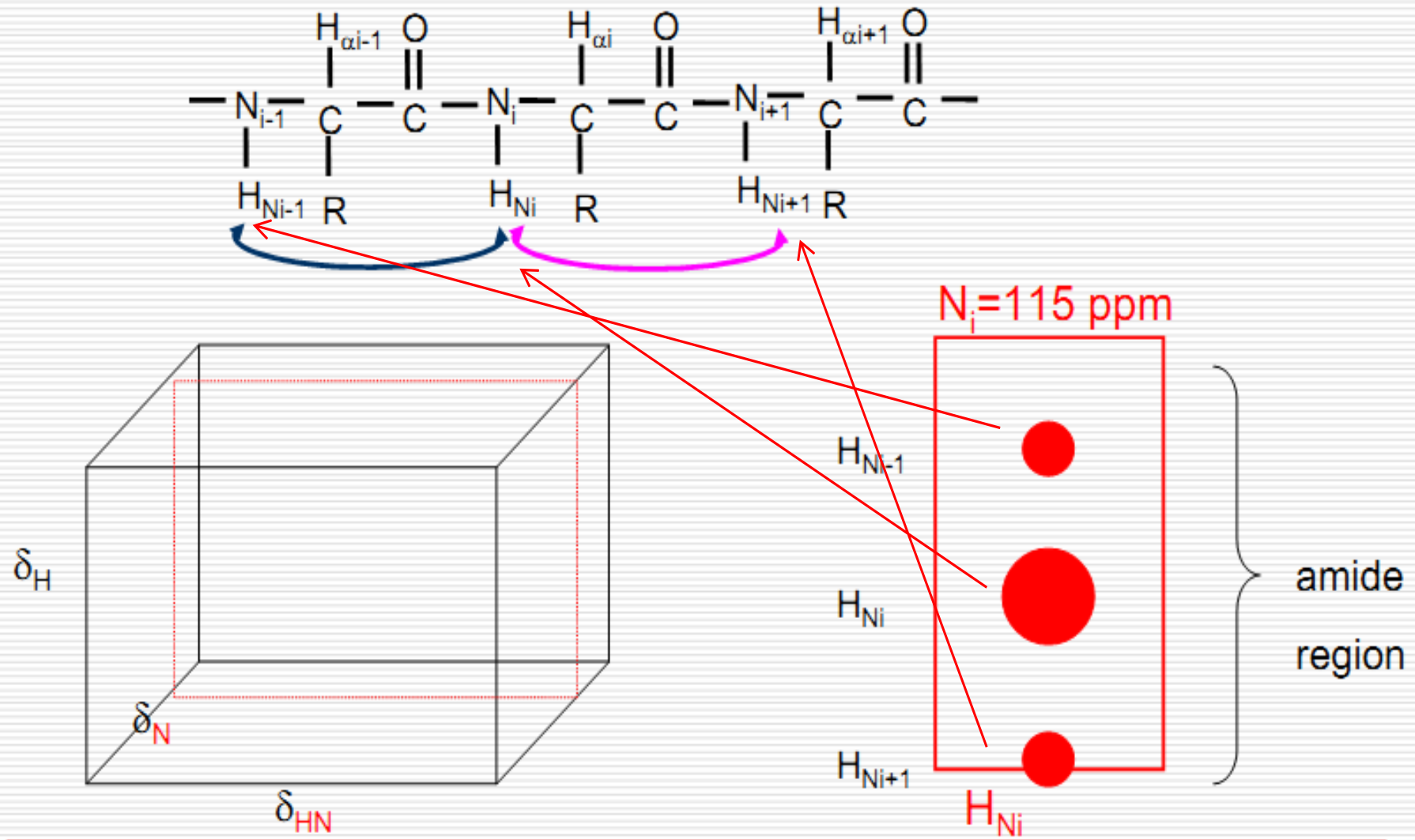
2. Side chain assignment:

- ^{1}H-^{15}N-^{1}H TOCSY-HSQC → based on backbone NH assignment
- ^{1}H-^{13}C-^{1}H: HCCH-TOCSY, HCCH-COSY → based on CH correlations
- ……

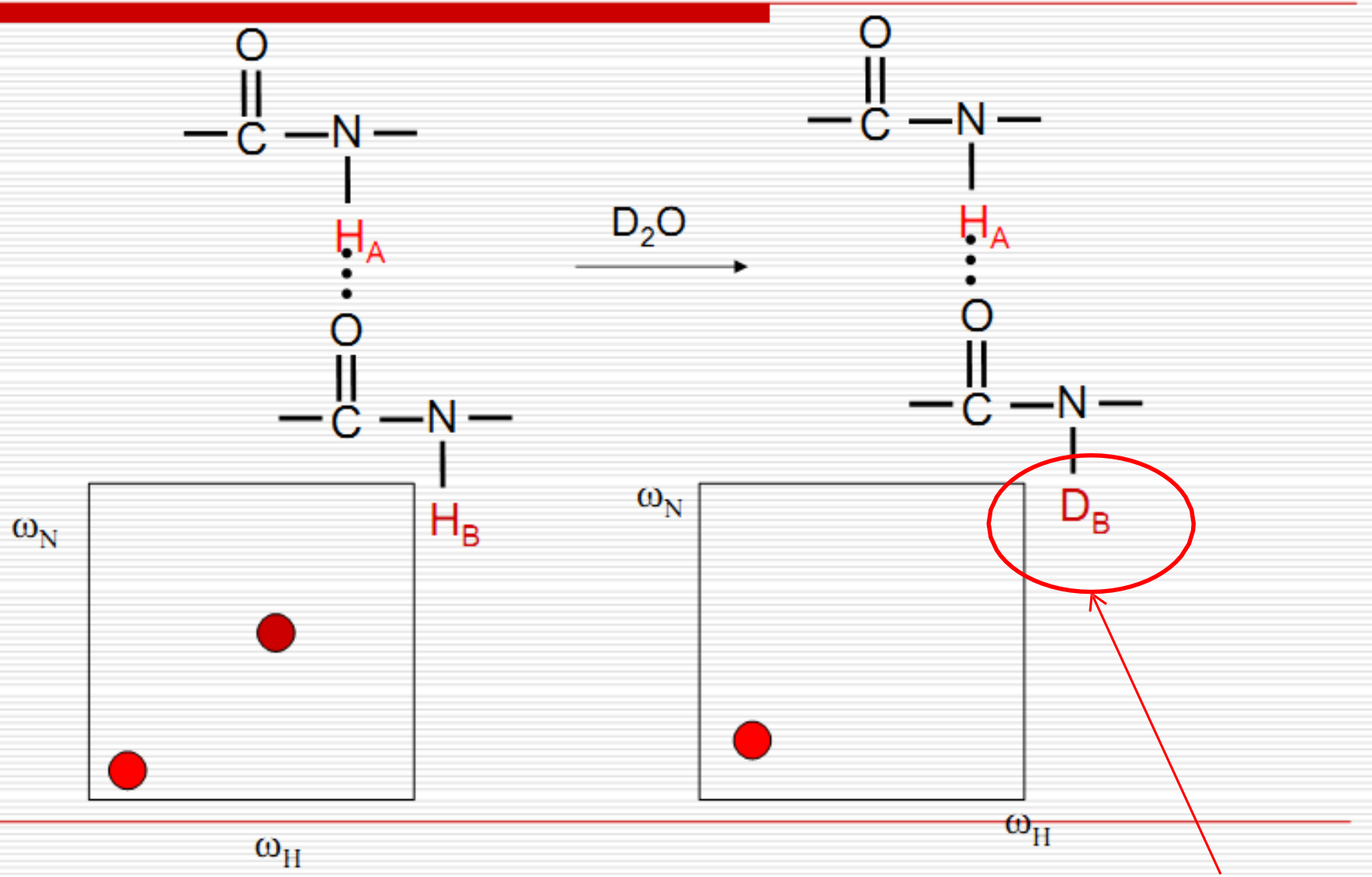
→ confirm the previous assignment

→ adding side chain H & C chemical shifts

15N Separated NOESY-HSQC



To find protons involved in H-bond



Replace H with D – signal disappears

Continued..

Combining all the results found from above experiments protein 3D structure can be determined.

Step-by-Step: How to Construct a 3D structure (not necessary in sequence)

(1) **Assign each resonance in the ^{1}H-NMR spectrum to a specific hydrogen nucleus** by COSY, which established 'connections' between protons made through covalent bonds.

(2) **Assign peptide backbone** (**carbon**) by HNCO, HNCA and HN(CO)CA: used together to allow the assignment of all the carbon, hydrogen and nitrogen resonances in the backbone of a protein (**Remember to label protein with C13**!).

(3) **Assign peptide backbone** (**amide**) by HNCO, HNCA, HN(CO)CA: (**label protein with N15**!)

(4) **Determine the side chains** by NOESY (probes connections through space): identifies protons which are brought into close proximity to one another even though these protons may be quite distant in the linear sequence

(5) Given sufficient NOE-derived '**distance constraints**', it then became possible to reconstruct the 3D shape of the protein.

(6) **Computer converts large number of NOE-derived distance restraints into energy terms**, and then **minimize the energy** of the system as a whole.

(7) Such an approach typically results in an **ensemble of structures**.

(8) These structures overlap in highly ordered regions of the interior of the protein, but there is less convergence in more disordered regions on the surface.

Secondary structure

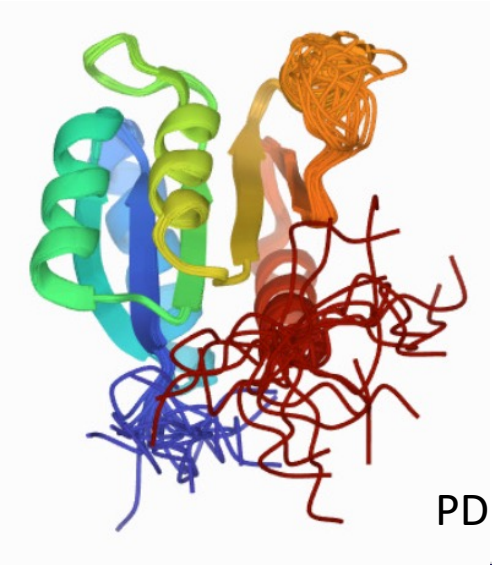
- the torsion angle phi calculated by the coupling constant $^{3}J\ H_N\text{-}H_\alpha$

right-handed alpha helix: phi = -57, $3JH_AH_N$ = 3.9 Hz
right handed 3.10 helix: phi = -60, $3JH_AH_N$ = 4.2 Hz
antiparallel beta sheet: phi = -139, $3JH_AH_N$ = 8.9 Hz
parallel beta sheet: phi = -119, $3JH_AH_N$ = 9.7 Hz
left-handed alpha helix: phi = 57, $3JH_AH_N$ = 6.9 Hz

- NOE: a number of short distances are fairly unique to 2^{o} structure elements
- Proton exchange

NMR Structures

- The NMR structure of a protein is presented as a bundle of conformers
- Each conformer presents a good solution to the NMR restraints
- Typically, a bundle of 20 conformers are deposited in the PDB



PDB: 2N9U

NMR Structures: calculation and refinement

- Solving a structure by NMR is a computational process where the experimental data constitute a set of restraints (e.g., this atom is connected to that one, these other two are <5 Å) and random structures are calculated and subjected to molecular dynamics simulations that take these restraints and other like known stereochemical parameters, bond lengths/angles) into account.

- Generally, 50 randomly generated configurations of the polypeptide are generated and subjected to restrained molecular dynamics (rMD) – the atoms are allowed “wander” and the energy of the conformation is calculated at each step.
- Any violations of the restrains (generated from data or chemical knowledge) incur an energy “penalty” in the rMD algorithm. When the calculation converges (E no longer changes), each of the 50 different models is scored. Those with gross errors are thrown out, and what is left is an ensemble of ~20 structures.
- NMR does not determine one single structure, since there are a range of possible conformations that satisfy all of the restraints.
  - Gives directly information about how the molecule acts in solution -- what parts of the molecule are mobile and what kind of range of motion these parts have.

Assessing the Quality of NMR Models

1. Data Quality
   - Number of restraints: the rule of thumb is 15-20 NOE restraints per residue. Even more important are the long range NOEs – a good NMR structure will have ≥3-5 restraints per residue
2. Model quality
   a. Root mean square deviation (RMSD) of the structure in the ensemble.
      a. Estimate of the precision of the model
      b. Values < 0.8 Å (backbone) and <1.5 Å (non-hydrogen) ⟹ high resolution
   b. Model geometry statistics
      a. There should be few/no violations of the restraints
      b. >90% of residues should be in the most favored region of the Ramachandran plot (with few/no outliers)
      c. Low deviation: <0.01 Å in bond length and <1.5 Å in bond angles
      d. Few/no unusual (strained) side chain rotamers

CRYO ELECTRON MICROSCOPY (CRYO-EM)

Cryo Electron Microscopy (Cryo-EM)

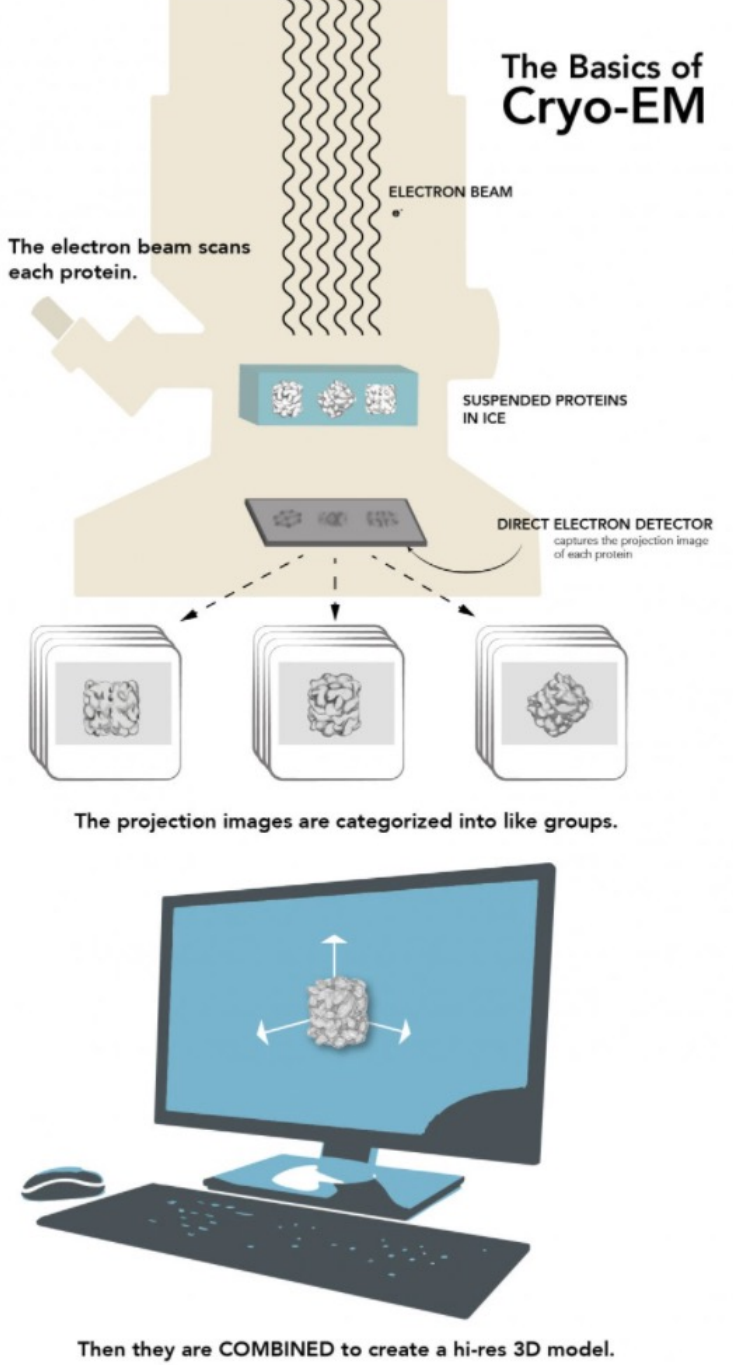
- Cryo-EM is extremely useful for getting structural information for proteins and large complexes that are too big (>50 kD) and/or too disordered for X-ray diffraction or NMR
- Cryo-EM doesn’t require protein crystals
- In the best cases, the techniques can get down to ~ 3 Å resolution

- 3D structure is the result of taking pictures of particles in a large number of random orientations
  - Particles are flash-cooled to give particles suspended in a thin layer of “vitreous ice” – no ordered, crystalline water lattice. This gives a field of dispersed, randomly-oriented molecules.

Cryo-EM

- Use beams of electrons rather than light to form images of tiny samples (proteins)
- Freeze samples to preserve the natural structure of biological specimens and reduce damage from the electron beam
- Generates 3-D images at nearly atomic resolution

Video: https://www.youtube.com/watch?v=BJKkC0W-6Qk



Summary

- There are three main research techniques for structural biology: single crystal X-ray diffraction (SC-XRD), nuclear magnetic resonance (NMR) and cryo-electron microscopy (Cryo-EM). However, there is no “all-purpose” method since all three of them offer unique advantages as well as limitations.

X-ray crystallography

- Advantages:
  - High atomic resolution
  - Not limited by the molecular weight of the sample
  - suitable for water-soluble proteins, membrane proteins as well as macromolecular complexes
- Disadvantages:
  - Must be crystallizable
  - Well-ordered single crystal
  - Represents the static form

NMR

- Advantages:
  - High resolution
  - provide information on a kinetic basis
- Disadvantages:
  - High purity
  - Difficult and expensive sample preparation
  - Difficult computational simulation
  - Water soluble sample
  - Below 40-50 kD

Cryo-EM

- Advantages:
  - Easy sample preparation
  - Small amount of sample
- Disadvantages:
  - Relative low resolution
  - >150 kD

LECTURE 11: OVERVIEW OF TECHNIQUES FOR PROTEIN-LIGAND INTERACTION

Measuring Protein-Protein/Ligand Interactions

- **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)**

Protein-Ligand Binding: Constants

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

[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.

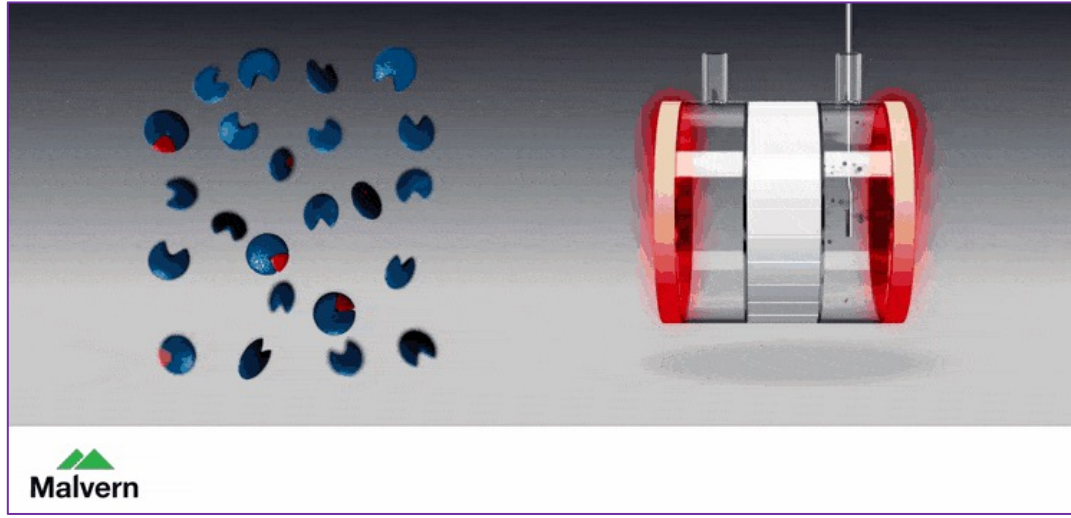
- Tight binding = High values of K_a and low values of K_D
- Weak binding = Low values of K_a and high values of K_D

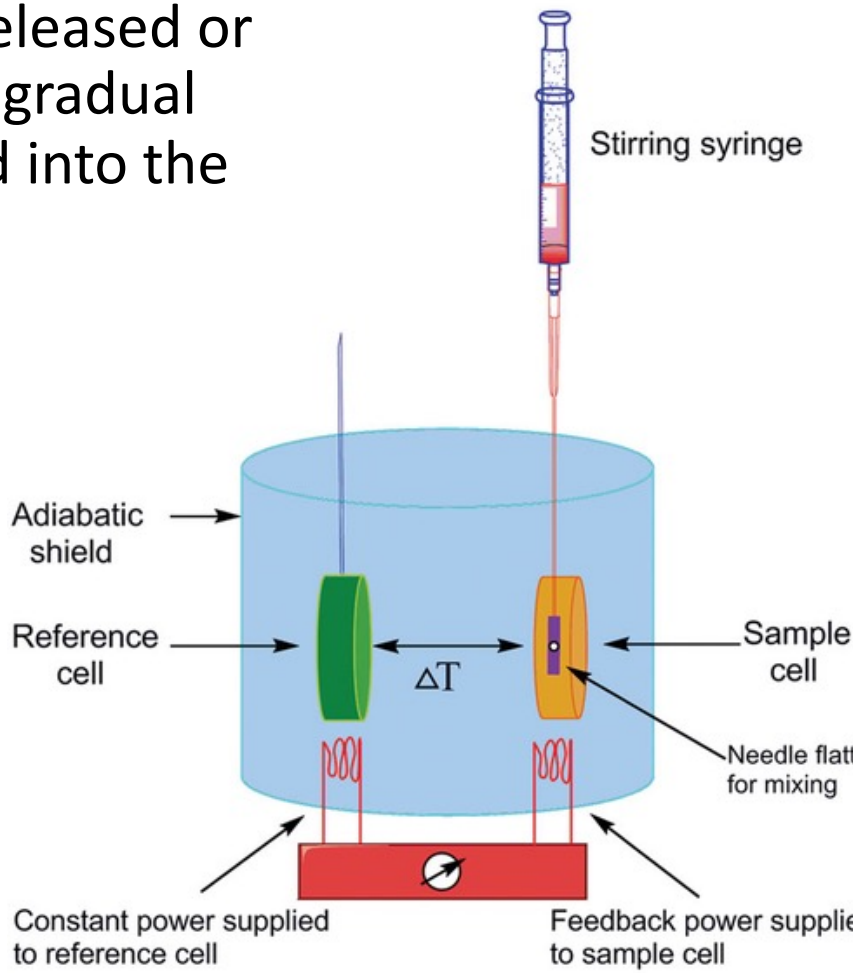
Isothermal Titration Calorimetry (ITC)

- ITC relies on the accurate measurement of heat changes caused by protein-protein/ligand interactions.
- ITC can be used to determine many protein-protein/ligand binding parameters in a single experiment:
  - Affinity constant (K_D)
  - Stoichiometry
  - Enthalpy (ΔH)
  - Entropy (ΔS)

ITC Principle

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

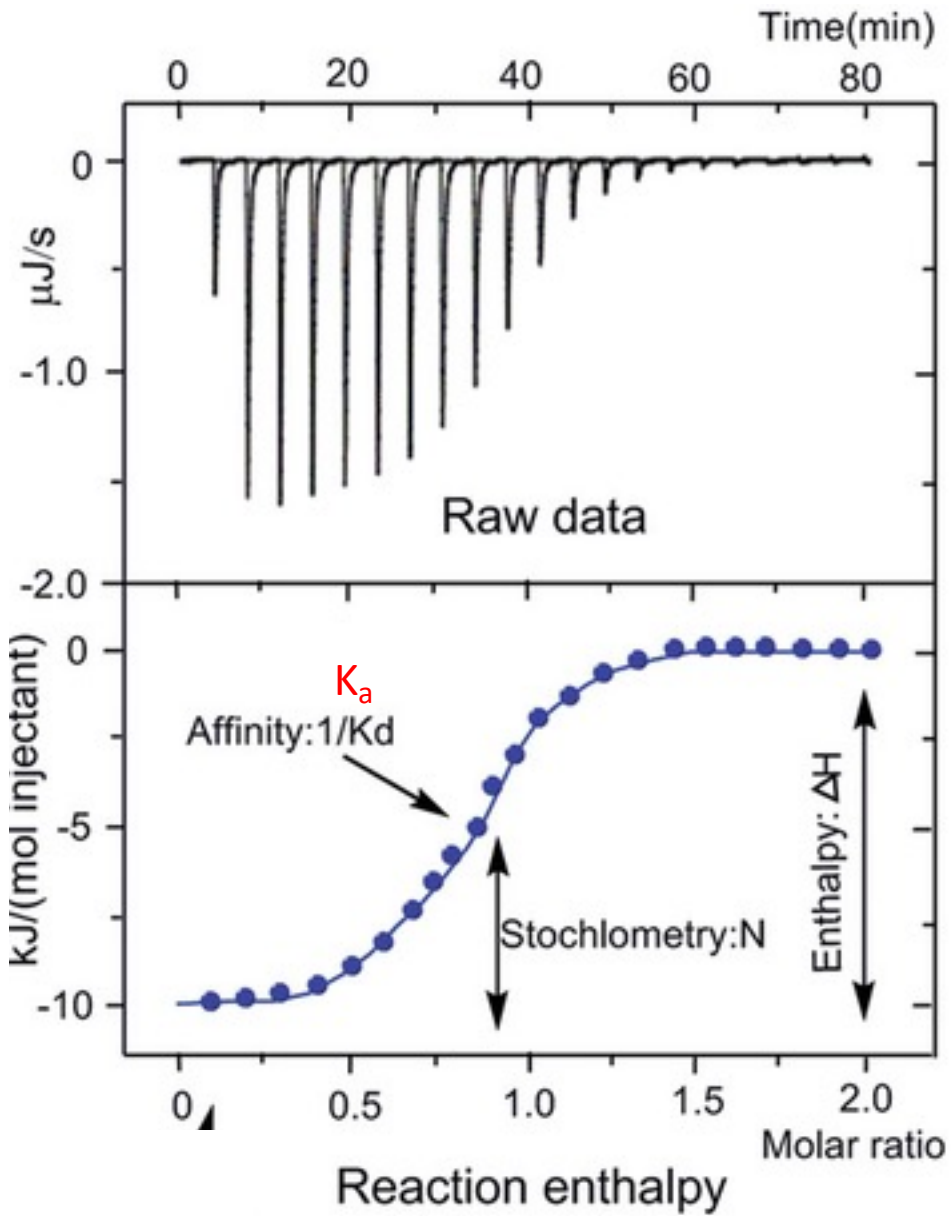




ITC Principle

$\Delta G = -RT \ln K_a$
$= \Delta H - T\Delta S$

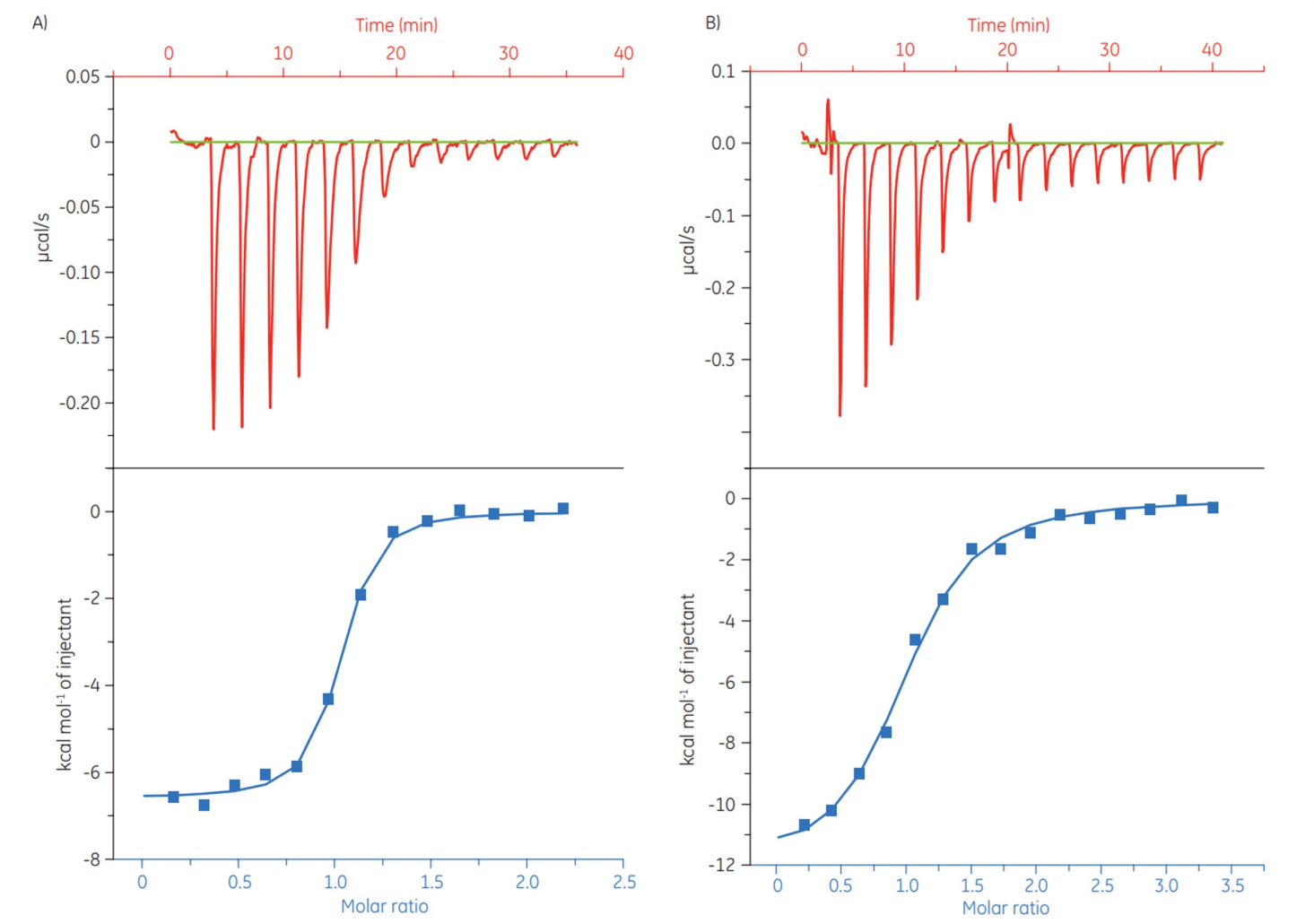
ITC thermogram

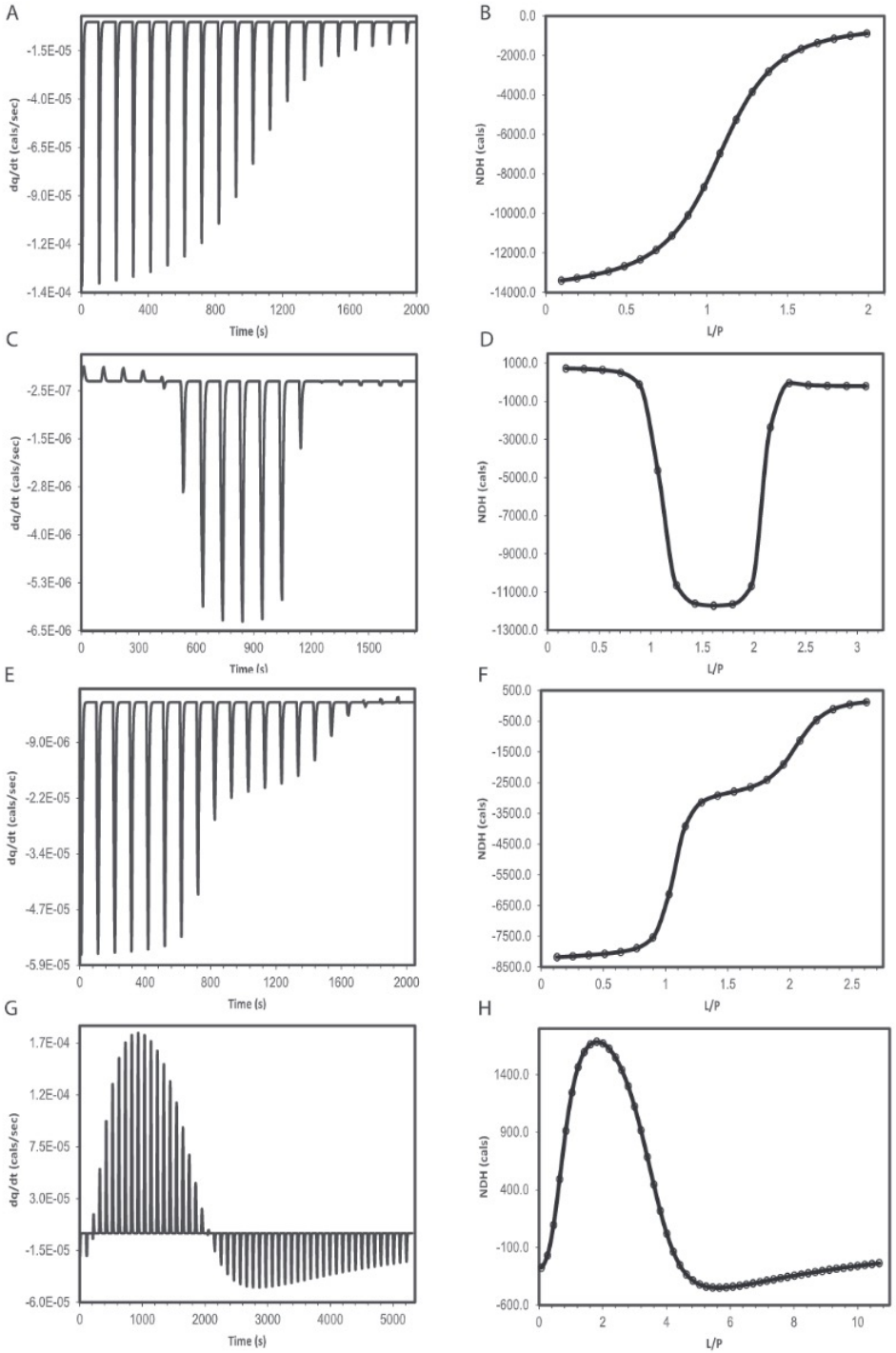


Poll Question: Which ligand binds to the protein more tightly?

A. Left

B. Right





single site binding

Two independent/parallel binding sites

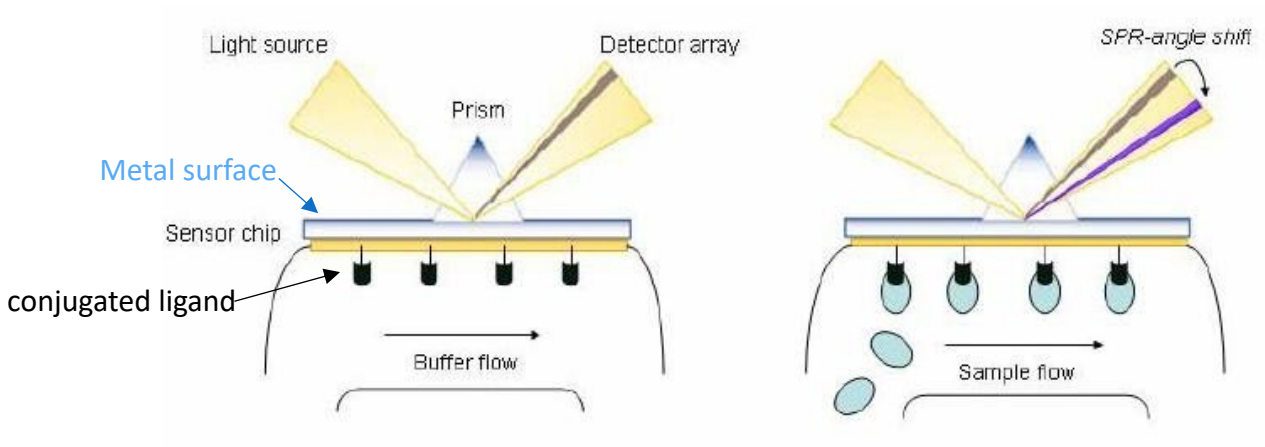
Two equivalent sequential binding sites

Four equivalent sequential binding sites

Surface Plasmon Resonance (SPR)

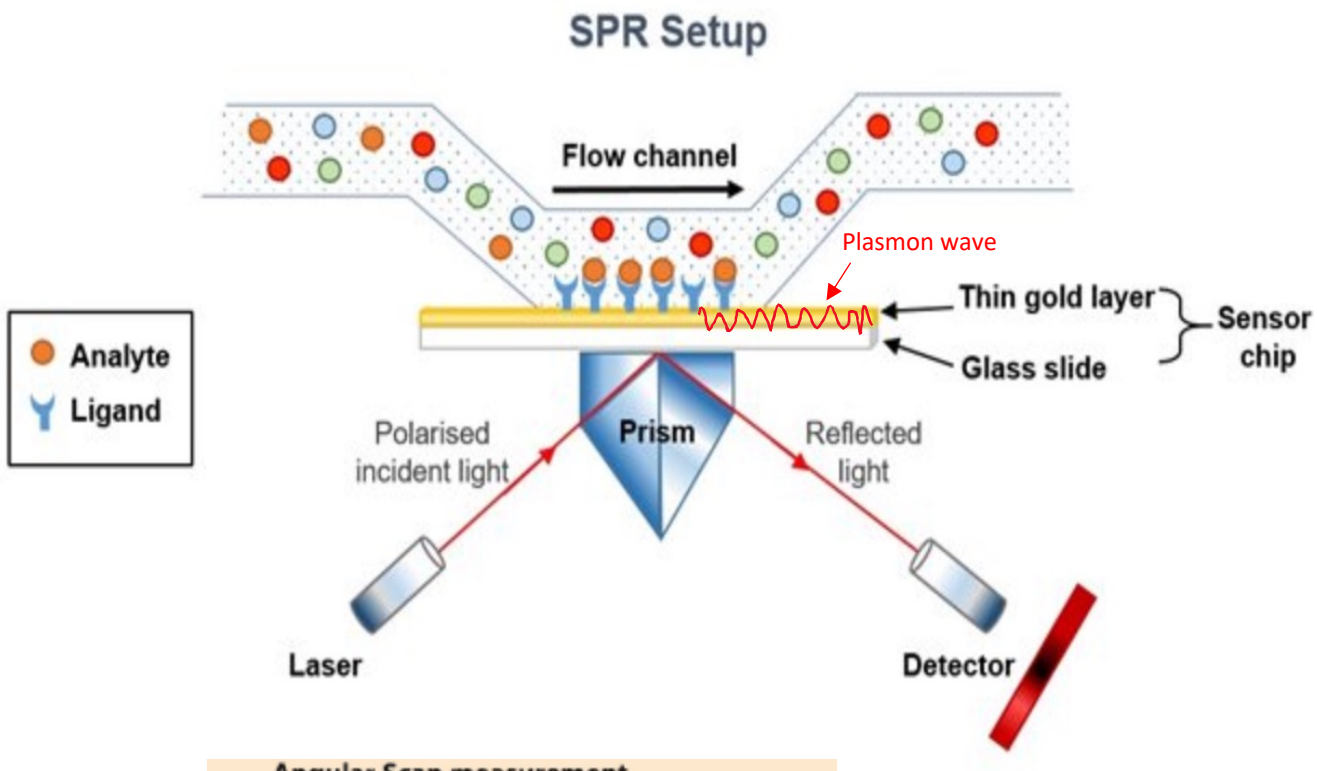
- **Biomolecular interaction** analysis

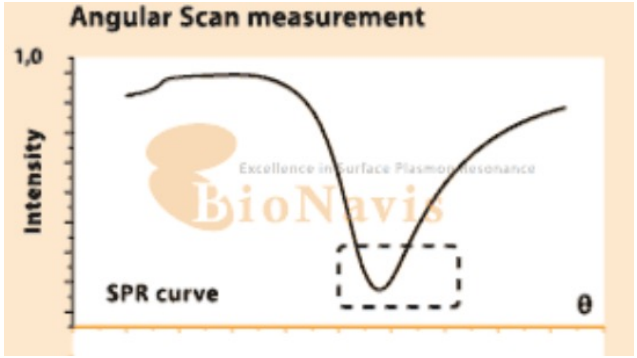
- Protein is immobilized onto surface
- Light is refracted onto thin metal layers
- Immobile protein **refractive index** changes when ligand is bound



- SPR detects changes in the refractive index at the surface of a sensor chip as a result of molecular mass changes of a target upon binding of the analyte.

SPR Principle

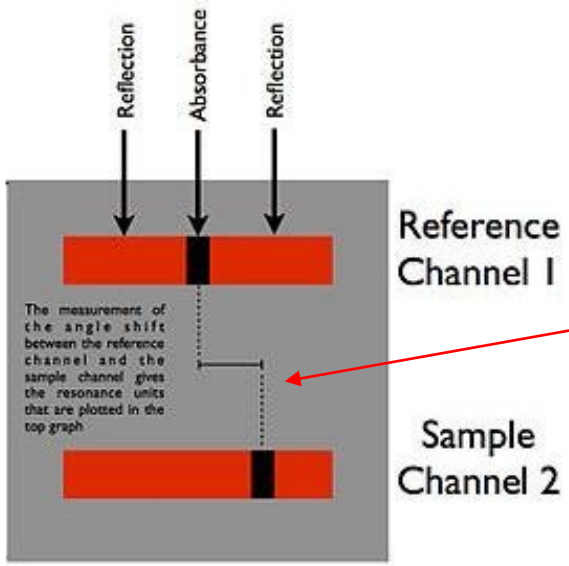




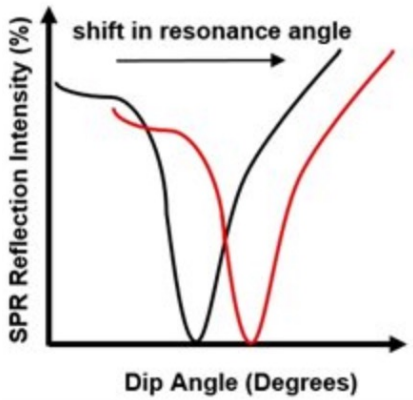
Angle of the reflection minimum (absorption maximum) is measured

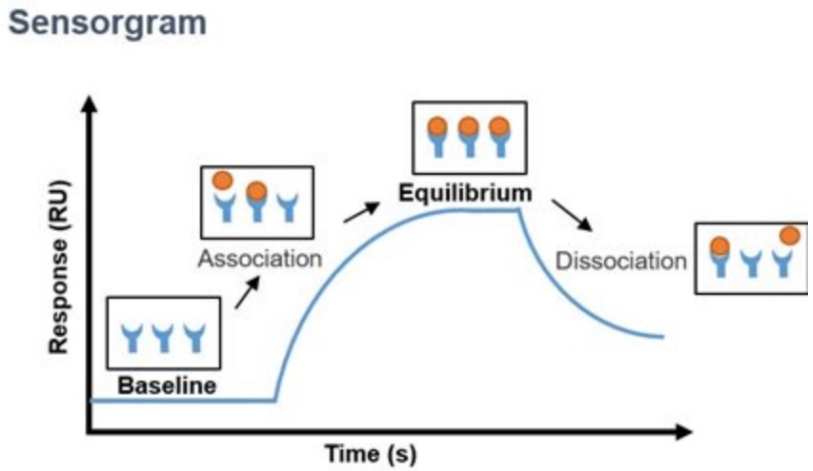
- A sensor chip with a gold coating is placed on a hemisphere (or prism).
- Polarized light shines from the light source on the sensor chip.
- Reflected light intensity is measured in the detector (disk).
- At a certain angle of incidence, excitation of surface plasmons occurs, resulting in a dip in the intensity of the reflected light.

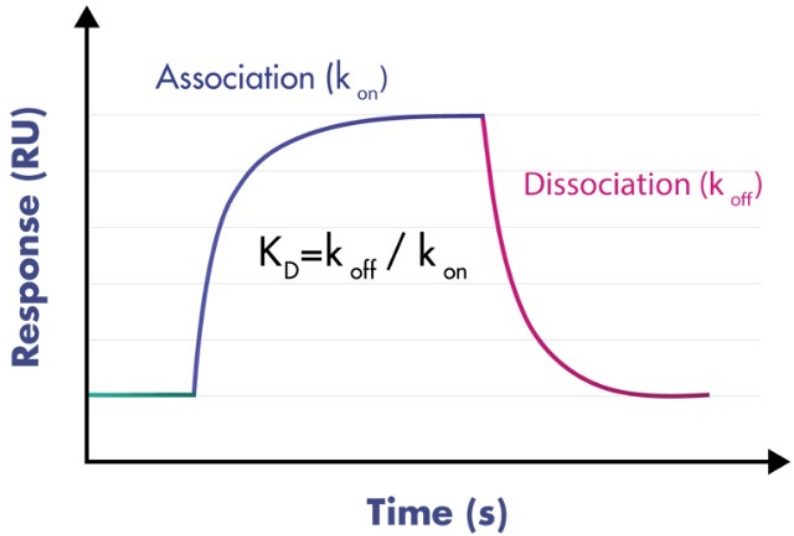
SPR Principle



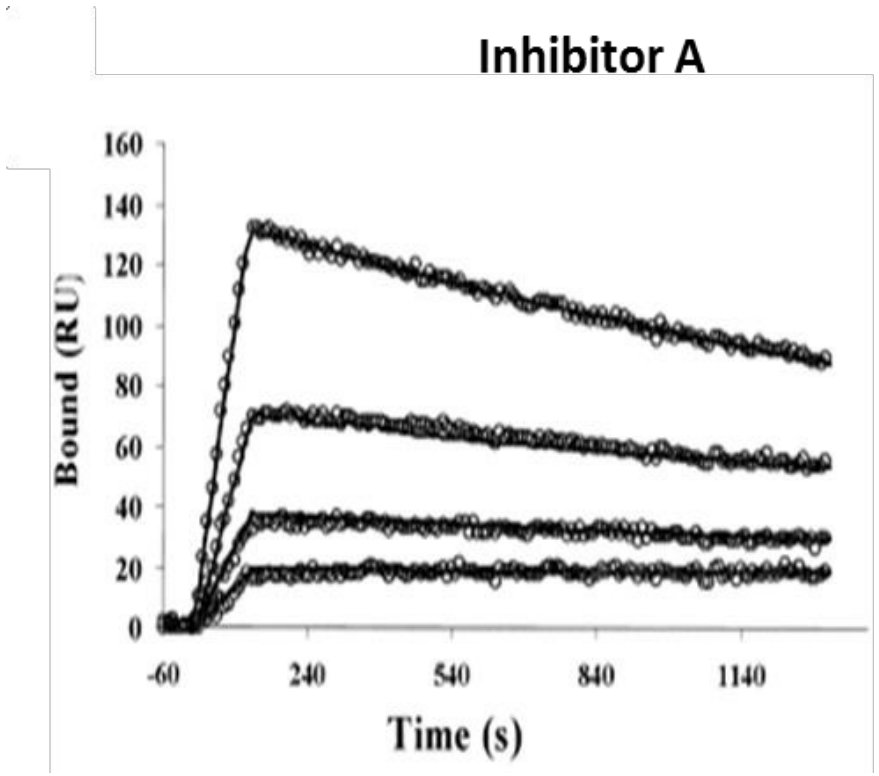
Angle of the reflection changed upon ligand binding

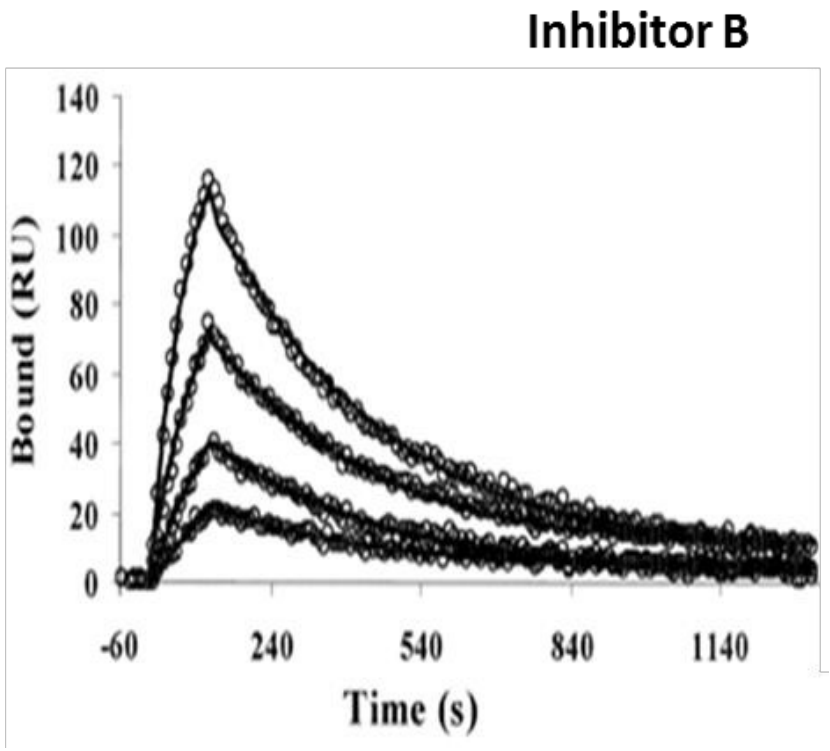






What can you tell from a SPR sensorgram?





Advantages

1. Very sensitive, pg-ng of protein required
2. Molecules > 25Da can be detected
3. Process is fast, < 2 h
4. Can differentiate different modes of binding: quick, slow, tight, etc.
5. Reproducible if carried out correctly
6. Sample can be recovered from machine (for MS)
7. Kinetic information (k_{on}, k_{off}, k_i) on binding, dissociation, affinity readily obtainable
8. Suitable for a variety of conditions (e.g., His tagged proteins, membrane proteins, untagged proteins)
9. Potentially label-free

Disadvantages

1. Expensive, equipment > CNY 5millions
2. Expensive to run, chips cost >CNY4000
3. Results often changed with different pH, buffers, temperature, etc.
4. Cannot differentiate specific and non-specific binding
5. Steric hindrance due to “coating”