Experimental Techniques in Protein Structure Determination

Homayoun Valafar

Department of Computer Science and Engineering, USC

Two Main Experimental Methods

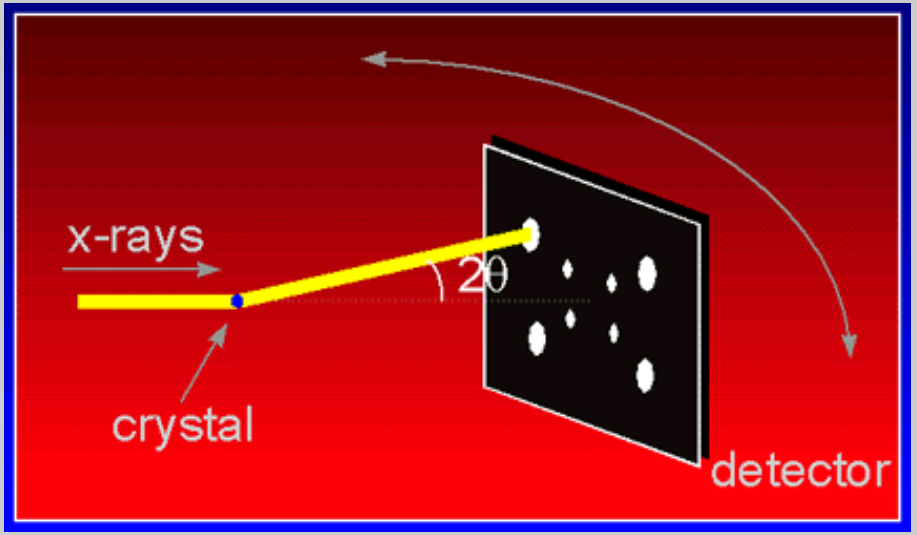
- X-Ray crystallography
- Nuclear Magnetic Resonance
- Some proteins are more challenging.
  - Membrane proteins
  - Complex carbohydrates.
  - Post-translationally modified proteins.
  - Multi subunit protein complexes.
  - Protein/Ligand complexes.
  - Dynamical proteins.

Protein Structure

- Drawings of the structures of proteins often convey the impression of a fixed, rigid structure, in which the side-chains of individual amino acid residues are locked into position.
- Nothing could be further from the truth.
- The changes that occur in the structure of hemoglobin when oxygen binds to the hemes are so large that crystals of deoxygenated hemoglobin shatter when exposed to oxygen.
- Further evidence for the flexibility of proteins can be obtained by noting that there is no path in the crystal structures of myoglobin and hemoglobin along which an O2 molecule can travel to reach the heme group.
- The fact that these proteins reversibly bind oxygen suggests that they must undergo simple changes in their conformation changes that have been called **breathing motions** that open up and then close down the pathway along which an O2 molecule travels as it enters the protein.
- Computer simulations of the motion within proteins suggests that the interior of a protein has a significant "fluidity," with groups moving within the protein by as much as 20 nm.

X-Ray Crystallography

- Principally, it is similar to medical X-ray imaging
- Automated analysis of the diffraction patterns yields structure
- Most critical part is the attainment of the crystal
- Even good quality crystals may not diffract well enough



X-ray Crystallography Summary

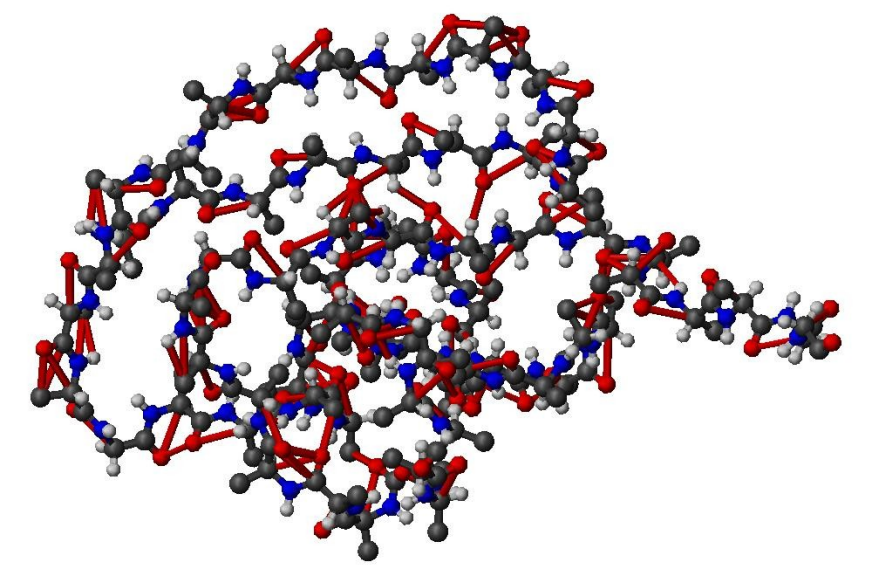
- Advantages:
  - Very routine and high potential for automation.
  - Fast structure determination. From data to structure under 6 hours.
- Disadvantages:
  - Crystal growth is a major unsolved problem.
  - Crystals maybe obtained at irrelevant conditions.
  - Semi-desiccated environment may influence structure.
  - Crystal packing forces may influence structure.
  - Aqueous proteins exist in a conformational ensemble. It can be argued that X-ray may filter for the conformation that crystallizes.
  - Glycosylation will normally render the protein hard to crystallize.
  - Multi-unit proteins are hard to crystallize.
  - Motion impedes crystallization $\leftrightarrow$ Crystallization perturbs motion.

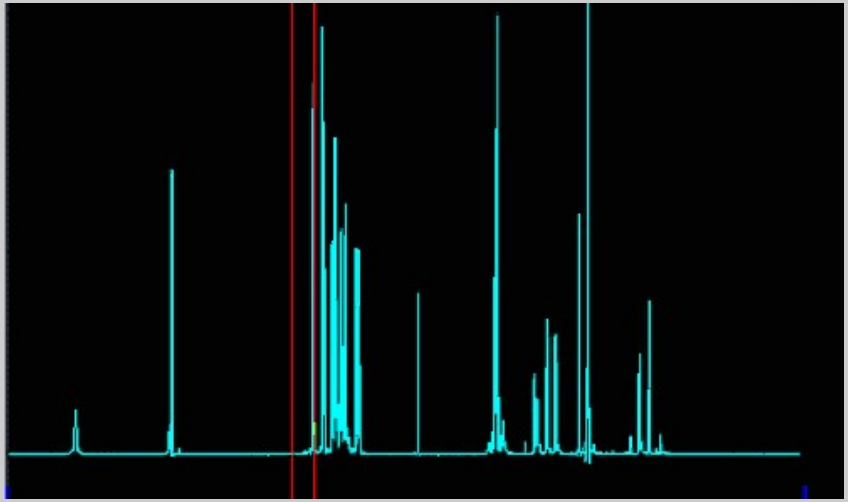
Nuclear Magnetic Resonance (NMR)

- Utilizes the magnetic property of nuclei
- ^{1}H, ^{15}N and ^{13}C are magnetically active
- Conventional NMR methods rely on the NOE (Nuclear Overhauser Enhancement) data
- NOE signal strength is related to the distance of two interacting nuclei by $1/r^{6}$(r is the distance)
- Can see only out to 5Å distance
- The limited range is usually useful for side chains

Structure Determination with NOE

- Collect as many NOEs as possible.
- Convert NOEs to distances.
- Use the distance restraints to fold the protein.





Restrained Optimization

- Experimentally collected data can be used as restraints.
- Restraints can be:
  - Distances
  - Dihedrals
  - Residual Dipolar Couplings
  - Relaxation
  - Electron density map (diffraction pattern)
  - Radius of Gyration
  - Diffusion
  - Etc…

Dihedral Restraints in XPLOR-NIH

- Example:

```
!!   2
assign (resid   1 and name c )   (resid   2 and name n )
       (resid   2 and name ca)   (resid   2 and name c )  1.0 -105.0 40.0 2
 !!   3
assign (resid   2 and name c )   (resid   3 and name n )
       (resid   3 and name ca)   (resid   3 and name c )  1.0 -115.0 30.0 2
```

- Make a table of these in a separate file
  - Typically named filename_dihe.tbl
- Include directives in your main Xplor script to load the file.

Distance Restraints in XPLOR-NIH

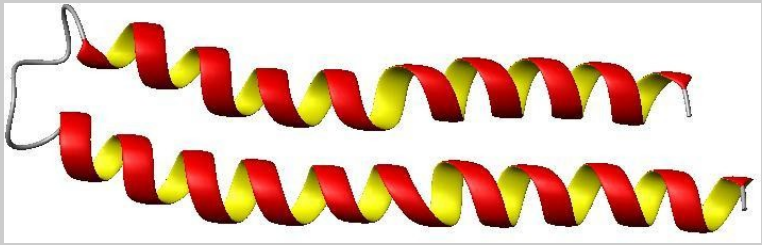
- Example:

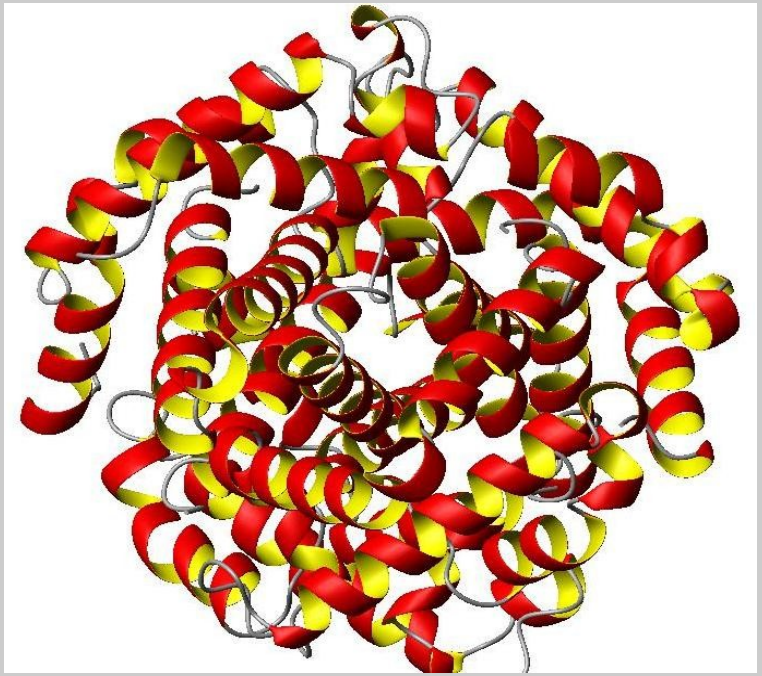
```
! CaH(i)-NH(j) long range
assign (resid   2 and  name HA   )(resid   18 and  name HN   ) 4.0 2.2 1.0
assign (resid   4 and  name HA   )(resid   18 and  name HN   ) 4.0 2.2 1.0
assign (resid   5 and  name HA   )(resid   52 and  name HN   ) 2.5 0.7 0.2
assign (resid   6 and  name HA   )(resid   16 and  name HN   ) 4.0 2.2 1.0
assign (resid   8 and  name HA   )(resid   14 and  name HN   ) 4.0 2.2 1.0
```

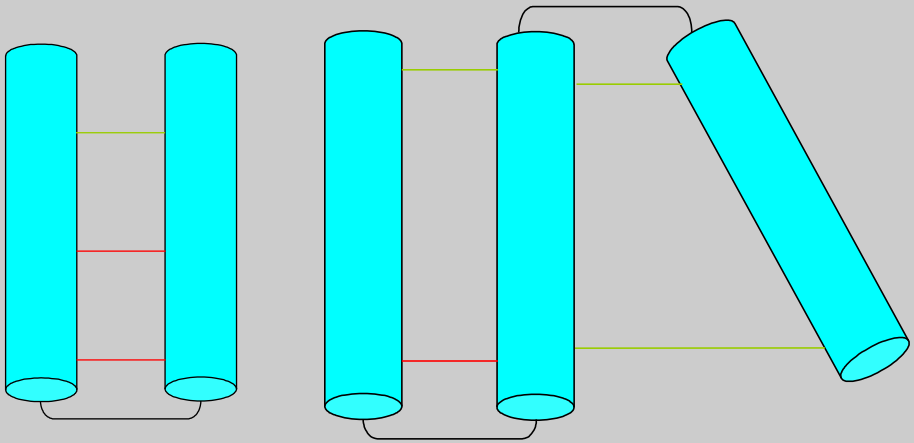
- Make a table of these in a separate file
  - Typically named filename_noe.tbl
- Include directives in your main Xplor script to load the file.

Consequences of Missing Data

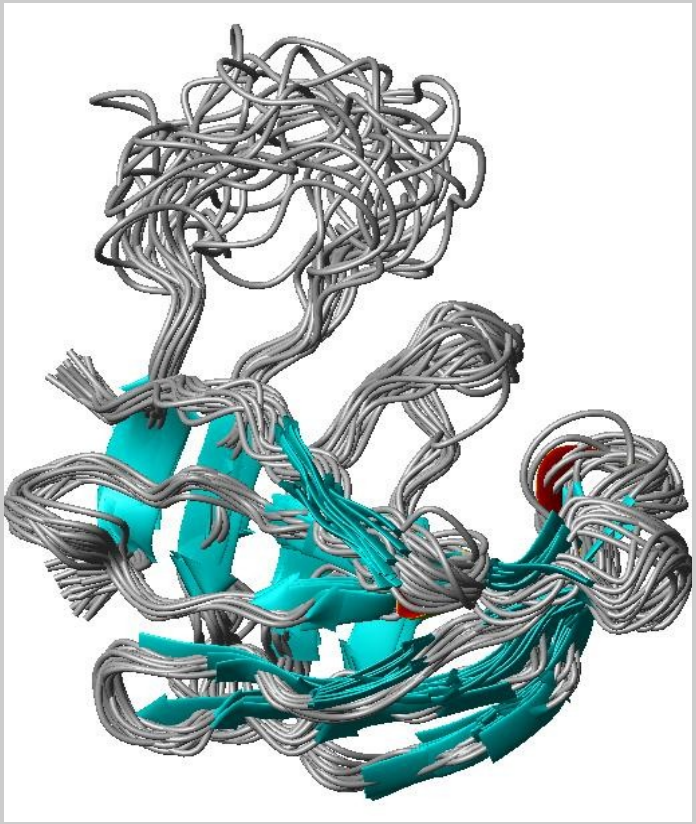
- How many distances are needed to determine the 2° structures?
- How many distance are needed to align the 2° structures (get the 3° structure) ?
- What are the consequences of missing data?
- Motion or just missing data.







Motion or Lack of Data?



1A57, INTESTINAL FATTY ACID BINDING PROTEIN

NOE Based NMR Summary

- Advantages
  - Protein is in its physiological environment.
- Disadvantages
  - Requires full isotope labeling of the protein.
  - Requires relatively high concentration of the sample.
  - Isotope labeling is expensive (time and money).
  - Very slow. From data to structure in months.
  - Acquisition of data can take up to months.
  - Consequence of missing data is unclear.
  - NOE is very susceptible to motion.

Residual Dipolar Couplings

- Uniform molecular tumbling reduces the average to zero.
- Uniform molecular tumbling can be perturbed using a crystalline solution.
- Non-uniform tumbling of the molecules at the interface of the crystalline solution will resurrect RDC.
- Crystalline solution can consist of Bicelle, Phage, cellulose fragments, PEG and etc…