

Lecture 1

Introduction to Protein Structure and NMR

Today we first talked about the protein structure, and then presented some principles in protein structure determination using NMR technique.

1 Protein Structure

Proteins consist of linear polymers of amino acids, and generally different sequences of amino acids have different three-dimensional protein structures. The following figure (Figure 1) shows the structure of a prototypical amino acid, where R stands for any of 20 amino acid side chains.

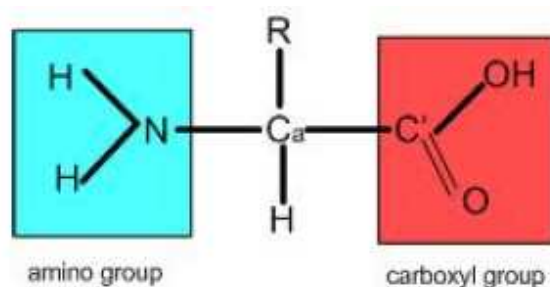


Figure 1: The structure of an amino acid.

Through a reaction of the corresponding amino and carboxyl groups, amino acids form peptide bonds. We refer to the atoms in the peptide bond as the peptide backbone. The characteristics of the peptide bond is shown in Figure 2. The peptide bond is ideally, although not in reality, planar and rigid, and the rotation is only allowed by the ϕ ($C_{\alpha} - N$) and ψ ($C_{\alpha} - C'$) torsion angles. Given a sequence of n amino acids, we are interesting in finding all torsion angles ($\phi_1, \psi_1, \phi_2, \psi_2, \dots, \phi_n, \psi_n$) for all peptide bonds to determine the three-dimensional structure of the protein backbone.

2 Structure Determination of Proteins with NMR Spectroscopy

Some NMR spectra can give us some biophysical (frequency) signals, which represent geometry information about the protein structure, such as the distance and angle information. Determination of the three-dimensional structure (ϕ and ψ angles) of protein backbone can be regarded as a graph realization or distance geometry problem, based on known distance

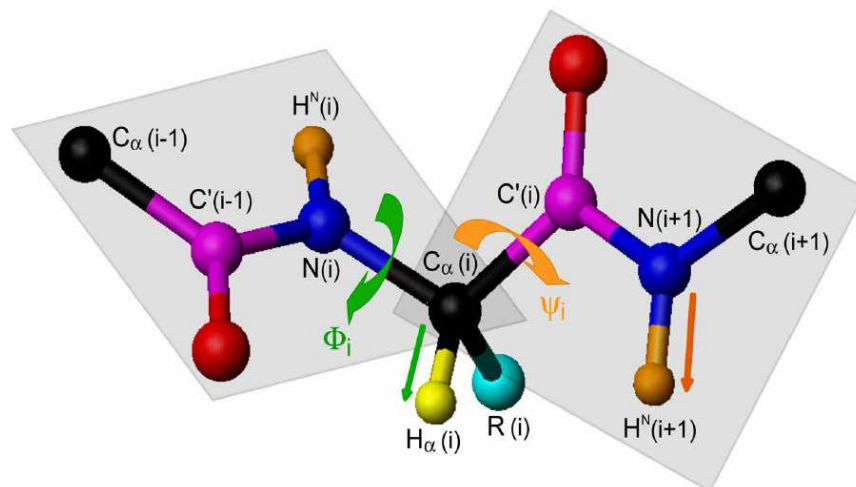


Figure 2: The ϕ and ψ torsion angles of peptide backbones.

or angle constraints. This problem can be solved easily if the *exact* distance constraints between *all* pairs of atoms are known. However, the NMR data unavoidably is uncertain and may have missing data. In fact, the noise in NMR spectra consists of the main complexity in our computational tasks. Also, distances can only be measured between pairs of protons less than 6 Å apart.

Nuclear Overhauser Enhancement (NOE). The NOE is a consequence of dipole-dipole coupling between different nuclear spins [1], and is used to measure the distance between atoms in proteins. The relation between the observed NOE intensity f and the distance r is ideally $f \propto r^{-6}$.

NMR Assignment Problem. The NMR resonance assignment problem is to find the mapping from each observed frequency (resonance) in NMR spectra to each NMR-active nucleus in the protein. More precisely, let S denote the set of atoms, and Ω denote the set of observed resonances. Physics gives us $f : S \rightarrow \Omega$. We have the inverse problem: to compute f^{-1} , where f^{-1} is called the *assignment*. We will talk about the resonance assignment problem in the next lecture.

Distance Geometry Problem. Let $d_{ij} = \|x_i - x_j\|$ where x_1, x_2, \dots, x_n denote the coordinates for atoms a_1, a_2, \dots, a_n in the protein sequence. We can represent all d_{ij} by a matrix, called the *distance matrix*, where each item intercrossed at the i th row and the j th column corresponds to d_{ij} . Now the distance geometric problem is to find coordinates¹ x_1, x_2, \dots, x_n such that the distance constraints in the given distance matrix are all satisfied. We classify the distance geometry problem into several cases according to different formats of distance matrices:

¹which are also related to the ϕ and ψ angles in the conformation of protein backbone.

- (1) **Problem with *All Exact Distances*.** If the exact distances between all pair of atoms are known, the problem can be solved in $O(n^3)$ time by using Singular Value Decomposition method, where n denote the number of atoms. More recently, Dong and Wu [2] have shown that this problem can be solved in linear time by using a “Geometric Build-Up” algorithm.
- (2) **Problem with *Sparse Sets of Distances*.** Given only subset of distances, the distance geometric problem becomes difficult to solve. Saxe [3] has shown that this problem is (strongly) NP-hard.
- (3) **Problem with *All But some Error Distances*.** This problem has been examined by Berger, Kleinberg and Leighton [5]. They mainly considered the approximation solution of this problem. Please refer to their paper [5] for more details.
- (4) **Problem with *Interval Bounds of Distances*.** Because of the noise from experimental data, we may only get some lower and upper bounds on the distances. It has been demonstrated such a problem is still hard to solve. More and Wu [4] considered the ε -approximation ² of this problem. They showed that if ε is smaller than a certain value, finding an ε -optimal solution to the distance geometric problem is still NP-hard.

Scalar (Spin-Spin) Coupling Constant ${}^3J_{HNHA}$. The spin-spin coupling constant ${}^3J_{HNHA}$ provides the torsion angle ϕ information, which is different to the regular secondary structures, such as α helix, β sheet, etc. The correspondence between ${}^3J_{HNHA}$ and ϕ can be found from empirically-derived Karplus curve, as shown in Figure 3. More specifically, the following values of ${}^3J_{HNHA}$ in some regular secondary structures are obtained [1]:

- (1) Right-handed alpha helix, $\phi = -57$, ${}^3J_{HNHA} = 3.9$ Hz
- (2) Right handed 3.10 helix, $\phi = -60$, ${}^3J_{HNHA} = 4.2$ Hz
- (3) Antiparallel beta sheet, $\phi = -139$, ${}^3J_{HNHA} = 8.9$ Hz
- (4) Parallel beta sheet, $\phi = -119$, ${}^3J_{HNHA} = 9.7$ Hz
- (5) Left-handed alpha helix, $\phi = 57$, ${}^3J_{HNHA} = 6.9$ Hz

The above data provide some hint for identifying regular secondary structures based on the measurements of ${}^3J_{HNHA}$.

References

- [1] Kurt Wuthrich. NMR of Proteins and Nucleic Acids. 1986.

²The solution is called ε -optimal if distance errors in the solution are within a tolerance of ε .

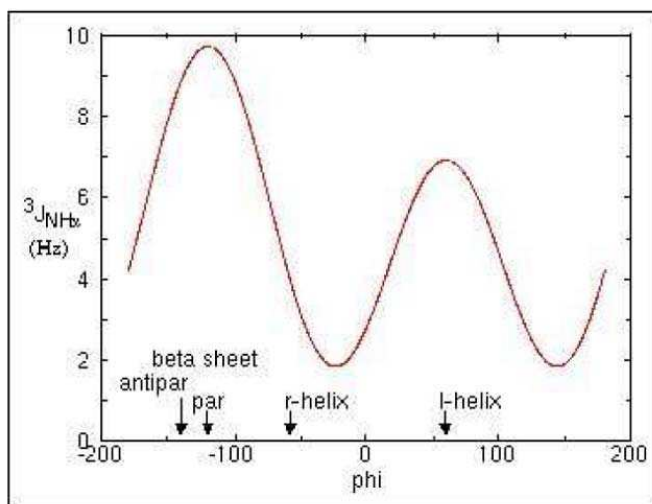


Figure 3: The Karplus relationship between the three-bond coupling constant ${}^3J_{HNHA}$ and the torsional angle ϕ . [<http://www.cryst.bbk.ac.uk/PPS/>]

- [2] Q. Dong and Z. Wu. A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances. *Journal of Global Optimization*, 22: 365-375, 2002.
- [3] J.B. Saxe. Embeddability of weighted graphs in k -space is strongly NP-hard. *Proceedings of the 17th Allerton Conference on Communications, Control, and Computing*, 480-489, 1979.
- [4] Jorge More and Zhijun Wu. ε -optimal solutions to distance geometry problems via global continuation. *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding*, P. M. Pardalos, D. Shalloway, and G. Xue, eds., pages 151-168, American Mathematical Society, 1996.
- [5] Bonnie Berger, Jon Kleinberg and Tom Leighton. Reconstructing a three-dimensional model with arbitrary errors. *Journal of the ACM (JACM)*, 46(2): 212 - 235, 1999.