Structure of invisible intermediate of a protein

Kalyan S. Chakrabarti

Dynamics is important in structural biology. Dynamics provides the essential link in understanding all the wonderful things that a protein is capable of, i.e. its function as well as the folding of a random polypeptide chain to a precise and beautiful structure. A necessary condition for understanding the dynamics of a protein molecule is to identify the non-native structures of the protein. These non-native structures can be visualized as valleys in the rugged conformational energy landscape of the protein. Protein molecule executes random motion among these valleys propelled by thermal energy, spending time in a valley it finds itself in dictated by the Boltzmann distribution. In this light it can be appreciated that there is considerable interest in determining the non-native structure of proteins in recent times[1–4]. The minor forms of proteins, which have very short lifetime, is hard to detect by structural techniques, e.g. X-ray crystallography or Nuclear Magnetic Resonance (NMR) spectroscopy. Atomistic information about minor states at equilibrium can be obtained from NMR spectroscopy. The NMR active nuclei, e.g. protons, have a characteristic resonance frequency depending on the chemical environment of the proton. Exchange of the nuclei between different chemical environments, which can be due to conformational transition in the protein molecule, causes broadening of the NMR resonance. This broadening of NMR signal can be suppressed by inverting the magnetization of a NMR active nucleus by applying a 180° pulse. A new technique in NMR spectroscopy utilizes the Carr-Purcell-Meiboom-Gill (CPMG) pulse-train, i.e. a series of successive 180° pulses, which can characterize motion in millisecond time-scale. This is the biologically important time-scale as enzyme catalysed reaction steps happen in this time scale, which is much slower than the time it takes for the actual chemical step to occur. Moreover, this time-scale is well outside the range at which atomistic molecular dynamics simulation of a protein in solution is possible at the present computational power.

As a result, the CPMG relaxation dispersion experiments are unique in that it provides information about population of minor state, the rate of exchange, and the chemical shift of the minor state at equilibrium (Box 1).

Korzhnev *et al*.[5] have calculated a protein folding intermediate using NMR spectroscopy. This technically demanding feat has been possible due to developments in two different fronts. Developments in NMR experimentations and labelling schemes have made the relaxation dispersion experiments possible. Perhaps more importantly, the development in computational methods for calculating structures from backbone chemical shifts has been a major contributing factor. The computational methods utilize a structural database to identify and predict the structure of short stretches of amino acids, which are then stitched together to solve the structure of the full protein. The program CS-Rosetta[6] has been used previously for determining the native state structure of proteins less than 120 residues in length[6]. The authors have expanded the scope of this program by incorporating residual dipolar couplings (RDCs), which are a form of orientational restraints, in the structure calculation[7]. For the present structure calculation the authors have modified the scoring function to take into account the difference between the experimental residual dipolar couplings and the dipolar couplings predicted from the structure[5].

Solving the structure of an invisible folding intermediate of proteins is a tour-de-force in structural biology (Box 2). Korzhnev *et al*. have solved the invisible excited state of 71 residues long FF domain from HYPA/FBP11 using recent advances in protein NMR spectroscopy and computational methods. The chemical shift dispersion data was collected in isotropic condition and in fractionally aligned state. The isotropic dispersion data fits to a two state $I \leftrightarrow N$ model and yielding the population, rate and the chemical shift of the excited state. The dispersion in the fractionally aligned state yielded RDCs that provided experimental data to calculate the structure. This intermediate has an equilibrium population of 2–3% and millisecond lifetime.

The protein was known to fold in two phases, the first phase being the formation of the intermediate from an unfolded state in microseconds, and the second slow phase is the transition of the intermediate to the native structure in milli-

Box 1.

Relaxation dispersion experiments in NMR spectroscopy can provide information about the exchange component in chemical equilibrium. The exchange component is given by $R_2(\nu_{CPMG}) = R_2(\nu_{CPMG} = \infty) + p_A p_B \Delta\omega^2 / k_{ex}$, (Figure 1), where $R_2(\nu_{CPMG} = \infty)$ is the intrinsic transverse relaxation rate, p_A and p_B are the populations of the two states, $\Delta\omega$ is the difference in chemical shift between the two states and $k_{ex} = k_{for} + k_{rev}$ is the rate of chemical exchange. The chemical shifts of the minor species (B) can be used to find the structure of state B using the program CS-Rosetta.

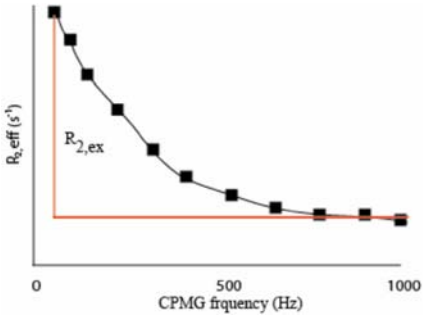

Figure 1. An example of relaxation dispersion profile for a backbone amide proton (squares) along with the best fits to a two-site exchange model.

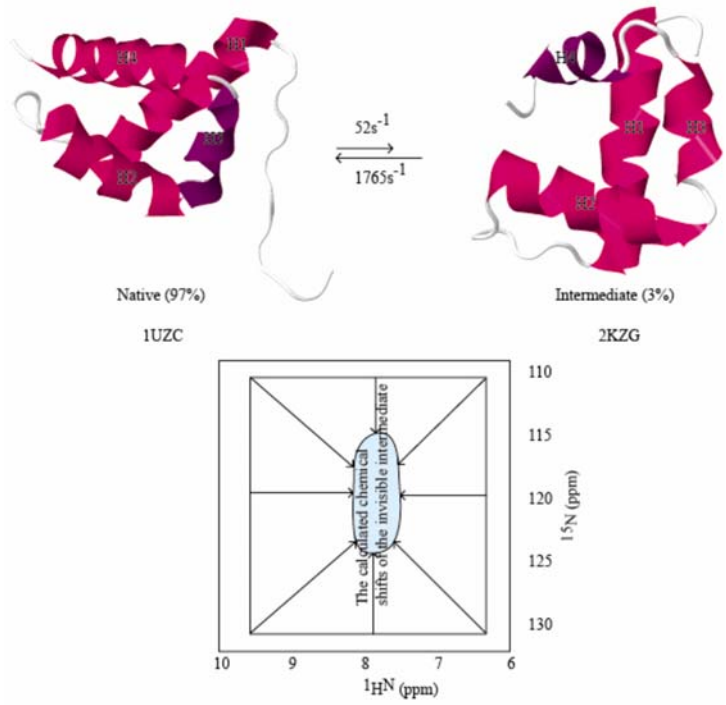

Figure 2. Comparison of the native (1UZC) and intermediate (2KZG) structures of the WT FF domain. The population and the rates at 30°C are also shown. The intermediate structure retains significant amount of secondary structural elements. However, the reconstructed NMR spectrum (from the calculated chemical shift values) shows the chemical shift dispersion characteristic of unfolded proteins (below). The native protein has chemical dispersion characteristic of a folded protein.

Box 2.

The science of protein folding has been dominated by the Levinthal's paradox and the presence of folding intermediates as a logical way out of the paradox. The Levinthal's paradox states that since there are an astronomical number of conformations available for a random polypeptide chain, an unfolded protein will take very long time if it samples the myriad conformations in an unbiased manner. The solution of the paradox lies in the fact that there are defined pathways, and intermediates, by which proteins fold. At the mechanistic level there are three dominant contenders for the protein folding mechanism. Framework model proposes that native secondary structures form first and then they diffuse until colliding with each other to form the tertiary structure (diffusion–collision model). The second model states that the secondary structure formation starts from nucleation center(s) bearing characteristic amino acid signature (nucleation model). In the third model hydrophobic side-chains collapse the random polypeptide chain which then rearrange to form the secondary structures directed by the native like tertiary contacts (hydrophobic collapse model)[8].

second timescale. Previous Φ-value analysis of the protein by comparing the rate of folding of the protein containing particular residues mutated has shown the existence of a stable core in the folding intermediate of the protein. Korzhnev *et al.* have done extensive validation of the intermediate structure due to the difficulty in seeing the intermediate state using any other method. The structure is compatible with the input structural restraints and the structure has the same features when calculated without the residual dipolar couplings. Moreover, the authors have changed the protein, e.g. deleted the carboxy-terminal part of the protein, to stabilize the intermediate structure. Lastly, the structure of the folding intermediate rationalizes the results from the Φ-value analysis.

The intermediate structure has some interesting features that will shed light on the folding of proteins. Though the structure resembles the native state of the protein, there are important differences towards the carboxy-terminal end of the protein, viz. there is a non-native helix in the intermediate that prevents formation of the native helix. These non-native interactions have to be broken before the protein can fold to its native form and this might be the rate-limiting step of folding, as this step is two orders of magnitude slower than the intermediate forming step. The non-native interactions are predicted by thermodynamic analysis when the Φ-value data are taken into account.

An interesting facet of the excited state is that the secondary structural elements are similar to those of the native protein even though the chemical shifts in the Hetero-Nuclear-Single-Quantum-Coherence (HSQC) spectrum would indicate severe loss of secondary structural elements[5] (Figure 2). This might necessitate a reevaluation of the common wisdom about the ^{1}H–^{15}N chemical shift distributions, especially when applied to non-native structures of proteins. This work throws open the gate for a thorough mapping of the protein conformational energy landscape by probing minor non-native structures.

1. Eisenmesser, E. Z., Bosco, D. A., Akke, M. and Kern, D., *Science*, 2002, **295**, 1520–1523.
2. Li, P., Martins, I. R., Amarsinghe, G. K. and Rosen, M. K., *Nat. Struct. Mol. Biol*., 2008, **15**, 613–618.
3. Boehr, D. D., McElheny, D., Dyson, J. H. and Wright, P. E., *Science*, 2006, **313**, 1638–1642.
4. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. and Kay, L. E., *Nature*, 2004, **430**, 586–590.
5. Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R. and Kay, L. E., *Science*, 2010, **329**, 1312–1316.
6. Shen, Y. *et al*., *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 4685–4690.
7. Tjandra, N. and Bax, A., *Science*, 1997, **278**, 1111–1114.
8. Fersht, A., *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman and Company, New York, 1999.

Kalyan S. Chakrabarti is in the Department of Biochemistry, Brandeis University, Waltham, MA-02453, USA.
e-mail: kalyan@brandeis.edu