Topological Requirement for the Nucleus Formation of a Two-State Folding Reaction. First Successful Blind Φ-Values Predictions - Activation Domain of Human Procarboxypeptidase

Roumen A. Dimitrov
dimitr@rpi.edu, roumend@hotmail.com
Department of Biology, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, USA

Keywords: protein folding, nucleus formation, Φ-values, transition state

Abstract

The folding problem of two-state small monomeric proteins is reduced to the question of how the folding nucleus at the transition state (TS) is formed from the ensemble of rapidly interconverting partly structured conformations in the denatured state. It is shown that in the denatured state the folding is energetically favored by certain highly fluctuating nucleation regions (α-helices and(or) β-hairpins), which in the experiments based on site directed mutagenesis are revealed by their high Φ-values. In the TS the folding is favored by the packing of these nucleation regions together with some other portions of the polypeptide chain, thus leading to a broad distribution of the Φ-values. The packing process results in a nucleus with native like-topology, approximately correctly formed secondary structures and loop regions with different degrees of order. This native-like nucleus is separated from all other folding alternatives by high free energy barrier. The calculations of the free energy of the folding nucleus are based on: 1) statistical mechanics of a linear cooperative systems; 2) A self-consistent molecular mean field theory previously developed for electrostatic interactions (Dimitrov and Crichton, 1997) and 3) A lattice model based on packing of idealized α-and(or) β-secondary structures. Blind computer simulations successfully predicted the experimental Φ-values of the activation domain of human procarboxypeptidase (Ada2H) (unpublished data at the time of simulations of L. Serrano EMBL, Heidelberg). From the calculated Φ-values it follows that the key residues for the folding of ada2H domain are ILE 23 from the amino end of the second α-helix and ILE 15 from the forth β-strand. Together with LEU 26, from the amino end of the second α-helix, and ALA 52 and VAL 54, from the first β-strand, they constitute the proposed nucleation site.

1 Introduction

Characterization of the refolding properties of a number of simple monomeric proteins has shown that the distribution of the Φ-values is strongly dependent on the topology of the final folded form of the protein. The influence of other factors, such as equilibrium stability of the native state and the chain length, are not apparent (Plaxco *et al.*, 1998). Therefore, the important question is to understand how the topological restrictions on the folding pathway result in the experimentally observed broad distribution of the Φ-values. None of the current theoretical models addresses this subject properly. Thus, in the nucleation-condensation mechanism (Fersht, 1997) the TS is seen as an uniformly expanded form of the native state with a large diffusive nucleus. It is composed of both neighboring residues in local secondary structure and long-range tertiary interactions. However, topological properties of the TS are not specified. In the funnel picture (Bryngelson, 1995), the TS is structurally degenerated. It is represented by a diffuse ensemble of states, distributed over the top of a broad free energy barrier. The nucleus is delocalized over the sequence and tertiary contacts. Finally, in the nucleation growth mechanism (Abkevich, 1994) the topological dependence applies to the accumulation of a kinetic intermediate with a native-like overall fold.

2 Methods and Results

Our approach is based on the existence of a high free energy gap at the TS level. The role of this gap is to increase the population of conformational states with productive interactions along the folding pathway. Thus, at its lowest free energy state, the TS is dominated by conformations with native-like topology, approximately correctly formed secondary structures and flexible loops. The population of all other folding alternatives, which include both changes in topology and secondary structures, are strongly reduced. A lattice model is introduced which takes into account the mutual packing of the secondary structures. Conformational states of the polypeptide chain are described by the fluctuations of the lengths and location of the secondary structures along the sequence and in the lattice. A theoretical approach, based on the statistical mechanics of a linear cooperative system and a self-consistent molecular field theory, is developed for the calculation of the free energy of the TS. Φ-Values are determined from the expression

$$\Phi^{\mathrm{calculated}} = \frac{\Delta F_{\mathrm{TS}}^{\mathrm{calculated}}}{\Delta\Delta F_{\mathrm{experiment}}},$$

where $\Delta\Delta F_{\mathrm{experiment}} = \Delta F_{\mathrm{experiment}}^{\mathrm{mutant}} - \Delta F_{\mathrm{experiment}}^{\mathrm{wild-type}}$ and $\Delta F_{\mathrm{TS}}^{\mathrm{calculated}} = F_{\mathrm{TS}}^{\mathrm{mutant}} - F_{\mathrm{TS}}^{\mathrm{wild-type}}$ are the perturbations of the free energy of the TS and that of the unfolding free energy upon mutation. Calculations were carried out with different sets of parameters for the secondary structure formation. In all cases the mutations of only hydrophobic residues were considered. Experimental data for the free energy of denaturation together with the sequence and coordinates for the Ada2H domain as well as the hydrophobic residues, which had been mutated, was kindly provided to us by L. Serrano EMBL, Heidelberg (unpublished data at the time of simulations). Our prediction is that ILE23 and ILE15 are the key residues for the folding of the Ada2H domain. It is connected with the stability of the N-terminus of the .α-helix 2 and its packing against the β-sheet. In particular ILE15 from the β-strand 1 and ALA52 and VAL54 from β-strand 4 form a small hydrophobic pocket, that packs against the ALA26, while the ALA23 is screened by the short proline rich connection between β-strand 3 and β-strand 4.

References

[1] Abkevich, V. I., Gutin A. M., and Shakhnovich E. I., Specific nucleus as the transition state for protein folding: evidence from the lattice model, *Biochemistry*, 33:10026–10036, 1994.

[2] Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G., Funnels, pathways and the energy landscape of protein folding: a synthesis, *Proteins Struct. Funct. Genet.*, 21:167–195, 1995.

[3] Dimitrov, R.A., and Crichton, R.R., Self-consistent field approach to protein structure and stability. I. pH Dependence of electrostatic contribution, *Proteins Struct. Funct. Genet.*, 27:576–596, 1997.

[4] Dimitrov, R.A., Self-consistent field approach to protein structure and stability, Ph. D. thesis, Published by Grafisch Service Centrum, Nude 40, Wageningen, The Netherlands, 1998.

[5] Fersht, A.R., Nucleation mechanisms in protein folding, *Curr. Opin. Struct. Biol.*, 7:3–9, 1997.

[6] Plaxco, K.W., Simons, K.T., and Baker, D., Contact order, transition state placement and the refolding rates of single domain proteins, *J. Mol. Biol.*, 277: 985–994, 1998.