FOLDING OF PROTEINS IN PRESCULPTED FREE ENERGY LANDSCAPE

NGUYEN BA HUNG[*], TRINH XUAN HOANG[†1]

[*] *Vietnam Military Medical University, 104 Phung Hung, Ha Dong, Ha Noi, Viet Nam*
[†] *Center for Computational Physics, Institute of Physics, Vietnam Academy of Science and Technology, 10 Dao Tan, Ba Dinh, Ha Noi, Viet Nam*

Abstract. *Recent studies of the tube model of protein have indicated that the free energy landscape of proteins is presculpted by symmetry of the protein backbone and geometrical constraints played by the hydrogen bonds. In this study, we investigate the role of amino acid sequences in the folding of proteins. We consider two models that are differed by sequence specificity: the tube HP model with hydrophobic (H) and polar (P) sequences, and the tube Go model with native-centric contact potentials. Monte Carlo simulations are carried out for two sequences of length of 48 amino acids, whose ground states are a three-helix bundle and a GB1-like structure. The results show that folding in the Go model is more cooperative than in the HP model. In the HP model the collapse transition and the folding transition are separated, whereas in the Go model the two transitions coincide.*

I. INTRODUCTION

Proteins are biomolecules composed by one or more chains of amino acids. There are nearly 100,000 different types of proteins in the human body. Proteins are engaged in every process on which our lives depend. They perform a vast array of functions including catalyzing metabolic reactions, replicating DNA, transporting molecules from one location to another, and participation in our immune system. Biological activity is observed only when protein is found in its native state, a well-defined three dimensional structure uniquely determined by its sequence of amino acids [1]. The ability of protein to fold quickly to its native state has been a subject of intense research for several decades. A widely accepted view of current understanding is that the native state is the global energy minimum in a funnel-shaped energy landscape [2, 3] and folding is a down-hill process which is associated with a much reduced conformational space. A successful class of models to describe the folding kinetics are the off-lattice versions (see e.g. [4]) of the Go model [5]. Go-like models assign favorable interactions only to contacts in the native state and ignores the amino acid sequence. Such an over simplistic approach, while successful and elucidating, is surprising [6] and requires more understanding. Therefore, it is useful to reexamine Go model in light of more realistic models.

Recent works [8, 9] have shown that common attributes of proteins such as the tube-like symmetry [7] of the backbone, the energetic and geometrical constraints induced by the backbone-to-backbone hydrogen bonds, and the overall hydrophobic attraction given by the side-chains are primary determinants of protein native structures. These studies

---

[1]Electronic address: hoang@iop.vast.ac.vn

indicate that protein structures are located in the marginally compact phase of a tube-like homopolymer which is poised between the swollen phase and the compact phase. The free energy landscape of proteins is presculpted by geometry and symmetry, and contains just a small number of minima that correspond to a menu of folds. The role of amino acid sequence is to choose the structures from such a menu of predetermined folds.

In this study, we investigate how the folding mechanism is affected by the sequence of amino acids, given the presculpted free energy landscape. In particular, we try to answer the question how the hydrophobic-polar (HP) sequence influences the folding of protein. By comparing two models of pairwise interactions between amino acids, the HP model and the Go model, we elucidate the role played by the HP sequence.

II. MODELS AND METHODS

We consider the tube model of protein [8] with two of its variants for hydrophobic interaction. Amino acids are coarse-grained as single beads located at the positions of the C_α atoms, and are placed along the axis of a self-avoiding tube of thickness $\Delta = 2.5\mathring{A}$. The bead spacing along the chain is $3.8\mathring{A}$. For any triplet of beads, (i,j,k), one can draw a circle of radius R_{ijk} going through the positions of the beads. The tube constraint is imposed by requiring that $R_{ijk} \geq \Delta$ for every triplet (i,j,k) [11, 12]. Additionally, sterics requires that two non-consecutive C_α's cannot be closer than $4\mathring{A}$ from each other. The bond angle associated with three consecutive C_α atoms is constrained to stay between 82° and 148°. The energy of a chain conformation is given by:

$$E = E_{bending} + E_{hydrophobic} + E_{hbonds} \ , \tag{1}$$

where the three terms on the right hand side correspond to bending energy, hydrophobic energy and hydrogen bonding energy, respectively. The bending energy is equal to the sum of local bending penalties along the chain. A bending penalty energy $e_R = 0.3\epsilon > 0$ is applied when the local radius of curvature at a given bead is smaller than $3.2\mathring{A}$ (the unit ϵ corresponds to the energy of a local hydrogen bond). The hydrophobic energy is the total energy of all pairwise hydrophobic contacts between amino acids. A contact is formed when two non-consecutive beads are found within a distance of $7.5\mathring{A}$. Hydrogen bonds have to satisfy a set of distance and angular constraints [8] on the C_α's as found by a statistical analysis of PDB's native protein structures [9]. A local hydrogen bond is formed between residues that are separated by three peptide bonds along the chain, and is given an energy $-\epsilon$. A non-local hydrogen bond is given an energy of -0.7ϵ. Additionally, a cooperative energy of -0.3ϵ is given for each pair of hydrogen bonds that are formed by pairs of consecutive amino acids in the sequence.

Two models of hydrophobic interactions are considered. The first one is the HP model with two kinds of amino acids: hydrophobic (H) and polar (P). Only contacts between hydrophobic residues are favorable and are assigned an energy of $e_{HH} = -0.5\epsilon$ per contact. Contacts involving polar residues are given zero energy. The second one is the Go-like model [5] which assign an energy of e_G for a native contact, and 0 for a non-native contact. A native contact is the contact that is present in the native state. In order to compare the two model, we choose e_G such that the total hydrophobic energy of

the native state is the same in the two models. From here on, we call the first model 'the tube HP model' and the second model 'the tube Go model'.

A parallel tempering [13] Monte Carlo scheme is employed for obtaining the ground state as well as other equilibrium characteristics of the system. For each system, 20 to 24 replicas are considered, each evolving at its own selected temperature T_i. For each replica, the simulation is carried out with standard pivot and crankshaft move sets and the Metropolis algorithm for move acceptance. In a pivot move, one randomly chooses a bead i and rotates the shorter part of the chain (either from 1 through $i-1$ or from $i+1$ to N) by a small angle and about a randomly chosen axis that goes through the bead i. In a crankshaft move, two beads i and j are chosen randomly such that $|i-j| < 6$, and the beads between i and j are rotated by a small angle and about the axis that goes through i and j. In both move sets, the rotation angle is drawn randomly from a Gaussian distribution of zero mean and a dispersion of 4°. An attempt to exchange replicas is made every 100 MC steps. The exchange of replicas i and j is accepted with a probability equal to $p_{ij} = \min\{1, \exp[k_B^{-1}(T_i^{-1} - T_j^{-1})(E_i - E_j)]\}$, where k_B is the Boltzmann constant, and E_i and E_j are the energies of the replicas at the time of the exchange.

The specific heat of the system is given by:

$$C = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2} \ , \tag{2}$$

where $\langle \cdot \rangle$ denotes thermodynamic average. For a given protein conformation, the radius of gyration, R_g, is defined as the root mean square distance from the beads to the chain's center of mass:

$$R_g^2 = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{r}_{cm})^2 \ , \tag{3}$$

where the center of mass location is given by:

$$\mathbf{r}_{cm} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{r}_i \ . \tag{4}$$

The weighted multiple-histogram technique [14] is used to compute the thermodynamic averages of quantities considered.

III. RESULTS AND DISCUSSION

We have chosen two structures shown in Fig. 1 for studying in the tube HP model and tube Go model: a three helix bundle (Fig. 1a) and a GB1-like structure (Fig. 1b). They are ground state conformations of two HP sequences of length $N = 48$ beads, which have been studied in Ref. [10] in the tube HP model. The name GB1 is due to the fact that structure is similar to the B1 domain of protein G. In this study, we construct the tube Go model for the two structures in such a way that the total hydrophobic energy of each structure are the same in the two models. Note that in the tube HP model, the hydrophobic energy is contributed by only the H-H contacts, whereas in the tube Go model, it is contributed by all the contacts in the native state. Otherwise, both models
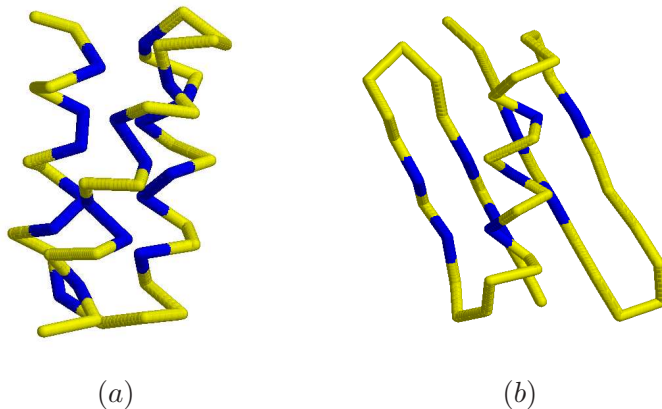
Fig. 1. Ground state conformations of two HP sequences considered in our study: a three-helix bundle (a) and a GB1-like structure (b). The hydrophobic (H) and polar (P) amino acids are shown in blue and yellow colors, respectively.
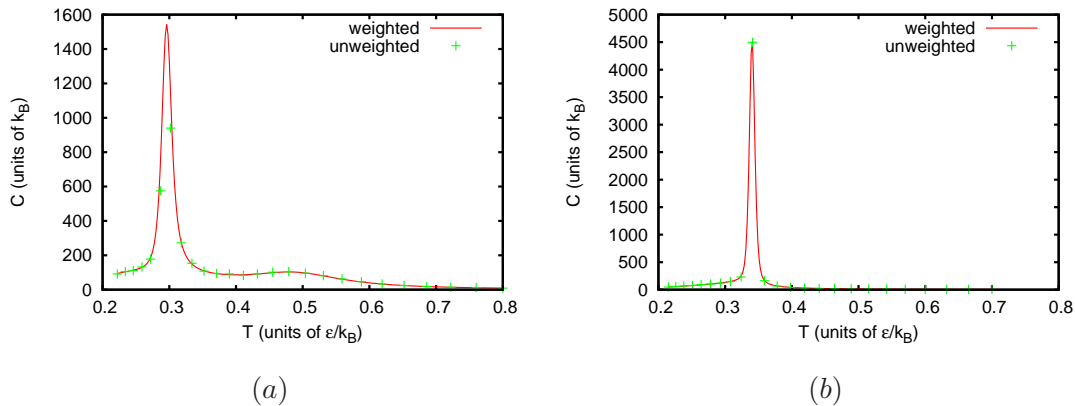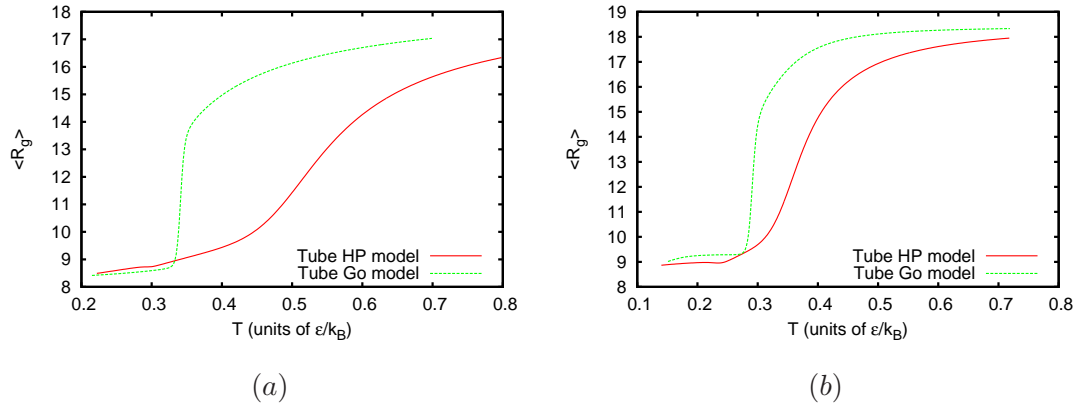

Fig. 2. Temperature dependence of the specific heat of the three-helix bundle in the tube HP model (a) and in the tube Go model (b). The data points (crosses) correspond to the specific heat calculated directly from the simulation runs, whereas the curve (solid line) is obtained with the weighted histogram method.

have the same energetic and geometrical constraints for the tube thickness, local bending and hydrogen bonding.

Parallel tempering simulations were carried out to obtain the ground state conformations as well as other equilibrium characteristics of the models. First, the simulations have confirmed that the two models have the same ground state as shown in Fig. 1. This indicates that both the HP sequence and the Go-like potentials provide sufficient bias towards the chosen native state structure. Note that, the tube constraint and hydrogen bonding provide no such bias but presculpt the free energy landscape so that the number of possible ground states is drastically reduced comparing to that of conventional polymers
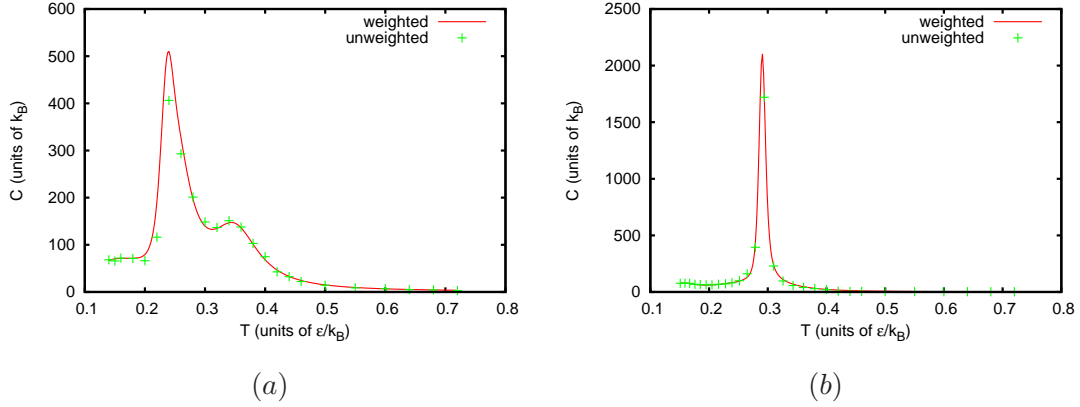
Fig. 3. Same as Fig. 2 but for the GB1-like structure.

Fig. 4. Temperature dependence of the averaged radius of gyration, $\langle R_g \rangle$, for the three-helix bundle (a) and the GB1-like structure (b) in the tube HP model (red) and tube Go model (green).

[8]. Design of a HP sequence that folds to a chosen ground state is relatively simple in this free energy landscape [10] and here, we have shown that Go-like potentials are also efficient.

We proceed to compare the thermodynamics of the two models. Fig. 2 and Fig. 3 show the temperature dependence of the specific heat for the three-helix bundle and the GB1-like structure, respectively. Both models display a sharp peak of the specific heat which signatures a cooperative folding transition [15]. The maximum of the specific heat, C_{max}, are roughly three times higher in the tube Go model comparing to the tube HP model. The temperature of the specific heat maximum, T_{max}, is also slightly higher in the tube Go model. These observations suggest that the tube Go model is significantly more cooperative than the tube HP model and the latter also yields a higher stability of the native state.

It can be seen that for the tube HP model there is a small shoulder on the right of the specific heat peak. The shoulder is more pronounced in the case of the GB1-like structure in the tube HP model, and not present in the tube Go model. This shoulder corresponds to a collapse transition which happens at higher temperature than the folding transition temperature. The situation is similar to the θ-transition of polymer in a bad solvent. In the tube Go model, the collapse and folding transitions coincide at temperature T_{max}.

The collapse transition can be seen in Fig. 4 which shows the sigmoidal shape of the temperature dependence of the mean radius of gyration, $\langle R_g \rangle$. The points of inflexion of the $\langle R_g \rangle (T)$ curves roughly correspond to the shoulder or the peak maximum in the specific heat for the tube HP model and the tube Go model, respectively. Note that collapse transition of the tube HP model occurs at a much higher temperature than in the tube Go model. Instead, the folding transition of the former occurs at a lower temperature than that of the latter.

IV. CONCLUSION

We have studied the folding process of proteins in two coarse-grained models that differ from each other by the pairwise interactions for the contacts between amino acids. Both models have common attributes of protein backbone such as the tube constraint, bending energy penalty and backbone-to-backbone hydrogen bonding. Such energetic and geometrical constraints have been shown to presculpt the free energy landscape of proteins with a few minima corresponding to protein-like structures. We have shown that the tube Go model has a higher folding cooperativity and a higher native state stability than the tube HP model. In the tube HP model the collapse transition and the folding transition happen at two different temperatures whereas in the tube Go model they appear at the same temperature. This finding allows us to conclude that even in the presculpted free energy landscape the folding process is strongly influenced by the sequence specificity. The HP sequence yields a more complex folding behavior than the Go model for pairwise interactions.

REFERENCES

[1] C. B. Anfinsen, *Science*, **181** (1973) 223-230.
[2] K. A. Dill, H. S. Chan, *Nat. Struct.Biol*, **4** (1997) 10-19.
[3] J. N. Onuchic, P. G. Wolynes, *Curr. Opin. Struct. Biol*, **14** (2004) 70-75.
[4] T. X. Hoang and M. Cieplak, *J. Chem. Phys.*, **113** (2000) 8319.
[5] N. Go and H. Abe, *Biopolymers*, **20** (1981) 991.
[6] D. Baker, *Nature* (London), **405** (2002) 39.
[7] A. Maritan, C. Micheletti, A. Trovato, J. R. Banavar, *Nature*, **406** (2000) 287-290.
[8] T. X. Hoang, A. Trovato, F. Seno, J. R. Banavar, and A. Maritan, *Proc. Natl. Acad. Sci. USA*, **101** (2004) 7960-7964.
[9] J. R. Banavar, T. X. Hoang, A. Maritan, F. Seno, A. Trovato, *Phys. Rev. E*, **70** (2004) 041905.
[10] T. X. Hoang, L. Marsella, A. Trovato, F. Seno, J. R. Banavar, A. Maritan, *Proc. Natl. Acad. Sci. USA* , **103** (2006) 6883-6888.
[11] O. Gonzalez, J. H. Maddocks, *Proc. Natl. Acad. Sci. USA*, **96** (1999) 4769.
[12] J. R. Banavar, O. Gonzalez, J. H. Maddocks, A. Maritan, *J. Stat. Phys*, **110** (2003) 35.

[13] R. H. Swendsen, J. S. Wang, *Phys. Rev. Lett* , **57** (1986) 2607.

[14] A. M. Ferrenberg, R. H. Swendsen, *Phys. Rev. Lett*, **63** (1989) 1195-1198.

[15] H. Kaya & H. S. Chan, *Phys. Rev. Lett*, **85** (2010) 4823-4826.