

FOLDING SIMULATIONS FOR A THREE-HELIX BUNDLE PROTEIN

LONGHUA HU, JUN WANG and WEI WANG

*National Laboratory of Solid State Microstructure
and Institute of Biophysics, Nanjing University, Nanjing 210093, P. R. China*

Received 9 August 2002

The folding thermodynamics and kinetics of a three-helix bundle protein with off-lattice model are studied using Monte Carlo simulations over a wide range of temperatures. By varying the strength of the dihedral angle potential, the effect of dihedral angle in folding is examined and it is demonstrated that the dihedral angle potential has much influence on protein folding thermodynamics and kinetics.

1. Introduction

It is well known that proteins can fold spontaneously into their unique native states, and the folding may be a cooperative process controlled both thermodynamically and kinetically. However, a completely understanding for the folding mechanism, one of the most challenging problems in biophysics, is still under extensively studying, and its solution is of fundamental importance for the prediction of the three dimensional structures of proteins. Thermodynamic control means protein native structure is the global minimum in free energy and is supported by the famous work of Anfinsen¹ and some other works later. With the development of experimental methods and theoretical models, a picture of energy landscape for folding kinetics has been proposed, and the folding of the protein chains follows the diffusion over different states or different conformations on the rough funnel with a free energy gradient toward the native structure. That is, the folding occurs as a multi-pathway kinetics of “water” flows along different routes down mountainsides.^{2,3} Thus, a common acceptable scenario for protein folding may be that first the polypeptide chain collapses from a random coil state to a rather compact one, then explores a set of compact structures and finally finds the native conformation with a minimum free energy.

So far, the computer simulation based on various kinds of simplified models for protein molecules is still the main method for protein folding studies since the all-atomic interaction simulation is limited by the speed of the computer at the present time.⁴ In this work, we study a three-helix protein folding process using an off-lattice representation of the protein with simplified interactions between the amino acids.

Using Monte Carlo simulation method, the thermodynamics and kinetics of the folding of the model chain are studied. Over a wide range of temperatures, the effect of the dihedral angle potential on folding is examined and it is demonstrated that the dihedral angle potential has much influence on protein folding thermodynamics and kinetics.

2. Model and Methods

The protein chain studied here consists of 46 amino acids, which comes from a real protein (see the insert of Fig. 1).⁵ From the database, we know that this small protein has a bundle consisting of three helices separated by helix-breaking prolines in the turn regions. Since this three-helix bundle protein is small and simple, it is expected to be a representative of the typical fast folding helical protein. Thus it has been extensively studied both experimentally^{6,7} and theoretically through various modeling such as the all-atom,^{8,9} the lattice,^{3,10} and the off-lattice modeling.^{11,12} Here, in our simulation, we utilize the off-lattice Monte Carlo method, and we expect that the exploration of the thermodynamics and kinetics of this protein will provide some detailed understanding of the folding of such a small helical protein.

We use the coarse-grained continuum representations for proteins in which only the positions of the C_α atoms are retained. Residues in the three-helix protein are simply represented as single beads centered in the C_α positions. Adjacent beads are strung together into a polymer chain by means of bond and angle interaction, while the geometry of the native state is encoded in the dihedral angle potential and a non-bond potential.^{12,13} As argued by a number of authors that this type of model accurately reproduces the three-dimensional structure of the protein.^{12,13} The potential energy includes the contributions from bond length and bond angle potential, the dihedral angle potential and nonbonded potential, respectively,¹³

$$V_{tot} = V_{bond} + V_{angle} + V_{dihedral} + V_{non-bond}. \quad (1)$$

The bond length interaction is represented with a harmonic potential $V_{bond} = (1/2)K_r(r - r_0)^2$. Here $K_r = 150\varepsilon/r_0^2$, and $r_0 = 3.8 \text{ \AA}$ is the average length between two beads. The bond angle potential V_{angle} is the bond angle deformation energy among three successive residues, and is given by a harmonic potential $V_{angle} = (1/2)K_\theta(\theta - \theta_0)^2$ where $K_\theta = 20\varepsilon/\text{rad}^2$ is the force constant, and $\theta_0 = 105^\circ$ or 1.836 rad is the equilibrium bond angle, which is close to the average virtual bond angles in α -helices in real proteins. $V_{dihedral}$ is the dihedral angle potential involving three successive bonds. For dihedral angles in the helical region, the energy function is represented by a standard potential consisting of three minima, one corresponding to the *trans* state, and the two *gauche* states,

$$V_{dihedral} = A_\phi(1 - \cos \phi) + B_\phi(1 + \cos(3\phi)) + C_\phi[1 + \cos(\phi + \pi/4)]. \quad (2)$$

The dihedral angle potential in the turn region is represented by

$$V_{dihedral} = D_\phi[1 + \cos(3\phi)] \quad (3)$$

where $A_\phi = B_\phi = C_\phi = 1.2\varepsilon$, $D_\phi = 0.2\varepsilon$. The small value of parameter D_ϕ makes the barrier in the turn region be shallow so that the turn is sufficiently flexible to adopt any conformation. This in turn facilitates the formation of loops.^{13,14}

Two nonbonded residues (i, j) form a contact if their distance is less than^{13,14} 6.5 Å. In the study of globular proteins, the topology of the native state is encoded in the contact map between the pair (i, j) of nonbonded residues that are in contact. The crucial interaction potential between nonbonded residues (i, j) is taken to be a modified Lennard-Jones 12-10 potential,

$$\Gamma(i, j)[5(d_{ij}/r_{ij})^{12} - 6(d_{ij}/r_{ij})^{10}] + 5\Gamma_1(i, j)(d_{ij}/r_{ij})^{12}. \quad (4)$$

It contains the non-local native interaction and a short-range repulsive term for the non-native pairs. Here d_{ij} is the distance between residues (i, j) in native state and r_{ij} is the distance between (i, j) in folding process. The parameter $\Gamma(i, j)$ and $\Gamma_1(i, j)$ are set as follows: if residues (i, j) form a native contact, $\Gamma(i, j) = 1$ and $\Gamma_1(i, j) = 0$. If residues (i, j) form a non-native contact, $\Gamma(i, j) = 0$ and $\Gamma_1(i, j) = 1$. Therefore, protein has preference to form native contact.

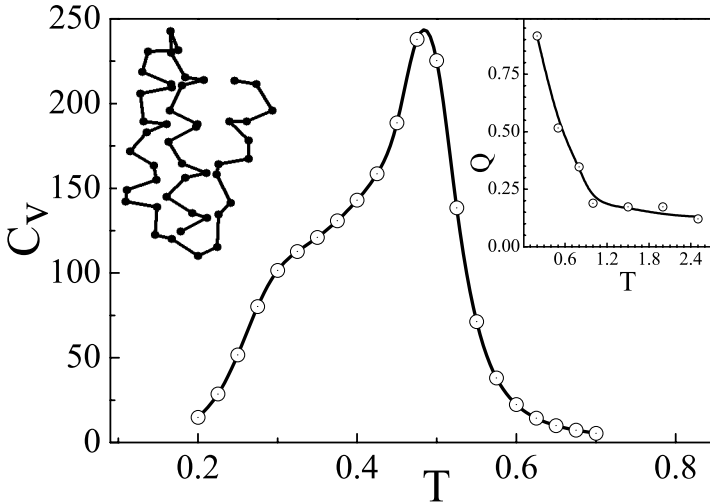


Fig. 1. Temperature dependence of the specific heat. The right inset shows the fraction of native contacts Q versus the temperature. the left panel gives the model structure of the three-helix bundle protein (residues of 10-55 of the B domain of *Staphylococcus aureus* Protein A (PDB Code: 1BBD)).

In this work, the method of off-lattice Monte Carlo simulation with Metropolis algorithms has been applied. The bead is allowed for local deformation, and for

each attempt, only one bead is allowed to move. The bond length is allowed to vary freely between $(1 - \gamma)r_0$ and $(1 + \gamma)r_0$, here we chose $\gamma = 0.05$. r_0 is the average bond length in native structure.

3. Results and Discussions

Let us report the folding thermodynamics. Our simulations show that the model chain folding undergoes two stages. The chain first collapses to a rather compact state, and then follows a folding transition. We calculate the specific heat $C_V = 1/K_B T^2 (\langle E^2 \rangle - \langle E \rangle^2)$ where K_B is the Boltzmann constant and T the temperature and E the energy of the chain. From Fig.1, we can see that there is a peak around $T \simeq 0.5$ and a sholder around $T \simeq 0.35$. The peak relates to the collapses transition from the extended coil state to the compact state of which the chain has an ensemble of compact three conformations. To clearly show the folding processes of the model chain, we use the fraction of native contacts Q as a reaction coordinate which is defined as the ratio of the number of native contacts, $Q(t)$, during folding to its value, Q_N , at the native state, i.e., $Q = Q(t)/Q_N$. Thus the value of Q represents the degree of similiarity between a certain state during folding and the native state. If in a stat the model chain has no native contact, then one has $Q = 0$, and if the chain is in its native state, then one has $Q = 1$. In the insert of Fig.1, we show the fraction of the native contacts Q as a function of temperature T . It is seen that for $T > 0.5$, the Q value is about only 0.2, and after that temperature there is increasing in the Q value. Thus, $T = 0.5$ real means a transition. This transition should also be connected with a pronounced change in the average potential energy as a function of temperature, which implies a peak in the specific heat. It is worthy to noting that the folding transition is of first order.

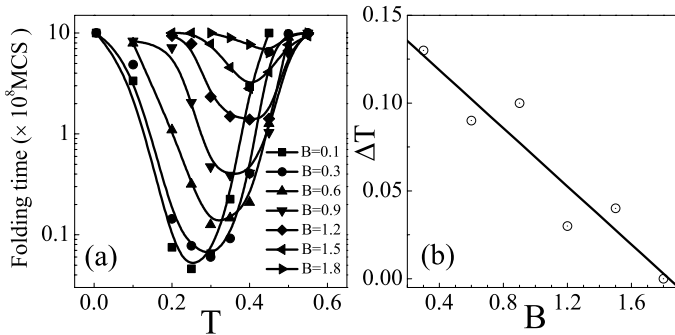


Fig. 2. (a). Folding time as a function of temperature for different values of the strength B . (b). ($\Delta T = T_f - T_{min}$) versus the value of the strength B .

We now report the folding kinetics of the model chain. The folding time (or mean folding time τ_f) is taken to be the mean first passage time (MFPT) averaged over an ensemble of random initial conditions. The first passage time for a chain

starting from an unfolded initial state is the total number of the Monte Carlo steps at which the model chain reaches to the native state. If the chain can not get to its native state within 1×10^9 MC steps, the first passage time is set as 1×10^9 MC steps. In Fig.2(a), we plot the folding time versus the temperature. We can see that a parabolic dependence of the folding time on the temperature. This form of the folding time has been observed in many simulations, for example in Ref.[15]. Both in the high and low temperature regions, the folding is slow, and in the intermediate region, the folding of the model chain is fast. However, it is noted that the microscopical folding processes are different for these two slow folding regions, respectively. In the high temperature region, the free energy is dominated by the the conformational entropy, which contributes to the folding barrier, thus the native state is hard to reach. While at low temperature, the energy landscape is rather rough, the folding of the chain is a downhill process.

In various local interactions of the residues in protein, the local conformation of the protein is determined mainly by the dihedral angle interaction. In order to study how the dihedral angle affects the folding behavior, we change the strength of the dihedral angle interaction and then observe how the folding rate changes with this strength. We choose to change the strength B . This quantity controls the difference of the energies between the *gauche* and the *trans* states, that is, affects the locally conformational preference of the backbone. For several different values of B , we obtain the folding time as shown in Fig.2. From Fig.2, we can see that as the strength of B increases the well-folding behavior region becomes narrow, and at the same time the folding becomes slow. In addition, the folding transition temperature increases too. That is, the folding behavior of the model chain becomes worse. From a physical point of view, the reason for such a behavior can be interpreted as follows. Since the dihedral angle potential is a function of the dihedral angle for different strengths of B . The potentials (easily seen from the figures of the function), we have seen that the potential curve has barriers. In order to get to the native state, the model chain must overcome the barriers. When the strength B becomes large, the barriers become large. Thus it is hard for the model chain to overcome the high barriers, which leads to the chain can not get to its native state and the folding becomes slow.

There are two characteristic temperatures T_f and T_{min} for the description the folding behavior. As mentioned above T_f is the folding temperature, and relates to the thermodynamics stability. It can be defined as the temperature at which the probability to occupy the native state is about 0.5. T_{min} corresponds to the fast folding temperature at the bottom of the U-shape curve in Fig.2(a), which means an optimal folding for the chain and it relates only to the kinetics. As argued in many works, the difference between T_f and T_{min} characterizes the goodness of the folding of the chain.^{15,16} In general, T_f should be greater than T_{min} , and a larger difference between them means a better folder from the kinetic aspect. Fig.2(b) shows the difference of these two temperatures ($\Delta T = T_f - T_{min}$) as a function of the strength B . From Fig.2(b), we can see that the folding property becomes good

as the value of the strength B increases, which is consistent with our results shown in Fig.2(a). Finally, we note that all points in this work related an average over 50 running, and the error bar is slight large. However, the main conclusion does not change if more times of running are averaged.

4. Conclusion

In conclusion, using Monte Carlo simulation method, we have studied the folding behavior of off-lattice three-helix bundle model chain. This small model chain shows typical folding behaviors of the proteins. We have also seen the effect of the dihedral angle potential on the folding features of the chain. We found that the folding kinetics has close relation with the dihedral angle potential, which is mainly due to different free energy landscapes for different dihedral angle potentials. Proteins with a very rough free energy landscape and high energy barrier between native state and the random coiled state are hard to fold.

Acknowledgements

We acknowledge the support from the National Natural Science Foundation (Grant No. 90103031, No. 10074030, and No. 10021001).

References

1. C. B. Anfinsen, *Science* **181**, 223-230 (1973).
2. K. A. Dill and H. S. Chan, *Nat. Struct. Biol.* **4**, 10 (1997).
3. H. S. Chan and K. A. Dill, *Proteins: Strut. Funct. Genet.* **30**, 2 (1998).
4. Y. Duan and P. A. Kollman, *Science* **282**, 740 (1998).
5. H. Gouda, *et al.*, *Biochemistry* **31**, 9665 (1992).
6. S. P. Bottomley, *et al.*, *Protein Eng.*, **7**, 1463 (1994).
7. Y. Bai, A. Karimi, J. Dyson, and P. Wright, *Protein Sci.* **6**, 1449 (1997).
8. Z. Guo, C. L. Brooks III, and E. M. Boczko, *Proc. Natl. Acad. Sci. USA* **94**, 10161 (1997).
9. E. M. Boczko and C. L. Brooks III, *Science* **269**, 393 (1995).
10. A. Kolinski, W. Galazka, and J. Skolnick, *J. Chem. Phys.* **108**, 2608 (1998).
11. Y. Q. Zhou and M. Karplus, *Nature* **401**, 400, (2002).
12. Y. Q. Zhou and M. Karplus, *Proc. Natl. Acad. Sci. USA* **94**, 14429 (1997).
13. Z. Guo and D. Thirumalai, *J. Mol. Biol.* **263**, 323 1996.
14. D. K. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **97**, 2544 (2000).
15. N. D. Socci and J. N. Onuchic, *J. Chem. Phys.* **101**, 1519 (1994).
16. R. Miller, C. Danko, M. J. Fasolka, A. C. balazs, H. S. Chan, and K. A. Dill, *J. Phys. Chem.* **96**, 768 (1992).