Protein Folding in Nano-Sized Cylinders *

XU Wei-Xin(徐维新), WANG Jun(王骏), WANG Wei(王炜)**

National Laboratory of Solid State Microstructure, Department of Physics, Nanjing University, Nanjing 210093

(Received 3 August 2004)

The folding of a model protein confined in a nano-sized cylinder is studied by the off-lattice Gō-like model. The entropy and anisotropy effects of confinement on thermodynamics and dynamics for folding are investigated. Our results show that due to reduction of the search on conformations, the folding rate can be sped up and the thermodynamic stability is enhanced at the cost of the decrease of folding cooperativity. In addition, it is found that these are shape-dependent. Folding is optimized in a cylinder with an appropriate shape when the volume is fixed. This is probably related to the shape of the protein molecule. Furthermore, our results also suggest that there is an orientational transition for the protein molecule following the variation of the radius of cylinder.

PACS: 87.15.Cc, 87.15.Aa, 87.14.Ee, 87.64.Aa

Starting from the work by Anfinsen,[1] the *in vitro* folding of soluble proteins has been widely investigated.[2−6] These studies have made considerable progress in understanding the mechanism of *in vitro* folding. The folding behaviour in cells is more complex because of the presence of lipids, carbohydrates, and other biological molecules. The high concentrations of various molecules and the interactions between them make the cellular environments rather crowded.[7,8] Consequently, the folding *in vivo* generally experiences a heterogenous confined space which geometrically restricts the conformational space of the protein molecules.[9] For example, the channel of ribosome looks like a narrow tube, and the chaperonin has a chamber with comparable radius and length. Biologically, some molecular systems, such as chaperones, are believed to be helpful in the formation of the unique native structures for proteins. These molecules may create totally different pathways of the *in vivo* folding.

Previously, some theoretical works have pointed out that the confined space can limit the movement of protein molecules, which reduces the conformational entropy and produces some intriguing effects on the thermodynamics and kinetics for proteins.[10,11] The proteins are even put into nano-sized porous materials which have more diverse local shapes to study the confinement effect on protein folding. However, the mechanism for such processes is still not very clear. What is the role of the shape of confined space for the folding? Do different kinds of confinement act similarly for the folding of proteins? These are interesting questions and helpful for understanding the physics of the *in vivo* folding. The self-organization of polymeric molecules in confined space is also of technical interest in nano-scale science. To understand the thermodynamic and kinetic features of polymer in confined space is a topic with much physical and biological interest.

In this work, we concentrate on the protein folding in a cylinder to model the heterogeneous environment. The heterogeneity of the confined space is made with various radii and heights of the cylinders. It is found that the folding rate is maximized for certain values of radii and heights for the cylinder, not the case for equal radius and height. Further simulations indicate that the sensitivity of folding to the heights of cylinder depends on the cylinder radius. There are two different stages related to the axial compression. A detailed analysis on the structure and orientation of protein molecule illustrates that the onset of these stages is attributed to different settlement of the nonspherical substrate protein. These results imply that confined space can change the folding processes and the features of folding, which may help the cells to organize their functions more efficiently.

The model used in this Letter is shown in Fig. 1. The protein is modelled as connected beads with Gō-type interaction.[12] The importance of monomer types is considered elsewhere.[13−15] Native contacts are defined when a pair of non-hydrogen atoms between two residues ith and jth (with $|j-i|>4$) are in contact, namely the distance between the pair of atoms is smaller than 5 Å. Other parameters are mentioned in Ref. [12]. In cell, co-factors associated with protein folding such as the channel of ribosome looks like a narrow tube, and the best characterized GroEL/ES molecular chaperone has a cylindrical structure. Therefore, the constraint of environment is modelled as a cylinder-shape chamber.[11] The height and diameter of the cylinder are $2H$ and $2R$, respectively (as shown in Fig. 1). A repulsion interaction

* Supported by the National Natural Science Foundation of China under Grant Nos 90103031, 10074030, and 10204013, and the Nonlinear Project of the National Key Basic Research Special Foundation (NKBRSF) of China.

** Corresponding author. E-mail: wangwei@nju.edu.cn

between residues and the wall exists when their distances are less than $\sigma_0 = 2$ Å, the radius of hard core of residue, which is realized with a Lennard-Jones form,

$$V_r = \sum_i 50\Big[\Big(\frac{\sigma_0}{r_i}\Big)^4 - 2\Big(\frac{\sigma_0}{r_i}\Big)^2 + 1\Big]\Theta(\sigma_0 - r_i)\,, \quad (1)$$

where r_i is the distance between the ith residue and the wall, and $\Theta(x) = 1$ when $x > 0$, and 0 otherwise. The Langevin dynamics is simulated using a leapfrog algorithm. The details of such an algorithm can be found in Ref. [16]. The thermodynamic quantities are calculated using the WHAM algorithm.[12,17] In the computational simulations, initial states are obtained from the ensemble that was equilibrated at $2.4T_f^0$ (the folding transition temperature in bulk). Note that all the lengths are scaled by a unit $a_0 = 3.8$ Å which is the length of pseudo-bonds between the C_α atoms. The protein chymotrypsin inhibitor-2 (CI2, PDB code: 1COA), a well studied two-state protein, is chosen as the substrate protein. Geometrically, the gyration radius of CI2 along the three main axes is 9.68 Å, 9.14 Å and 7.41 Å, respectively. The features of the two long axes make the protein CI2 look like an oblate ellipsoid.

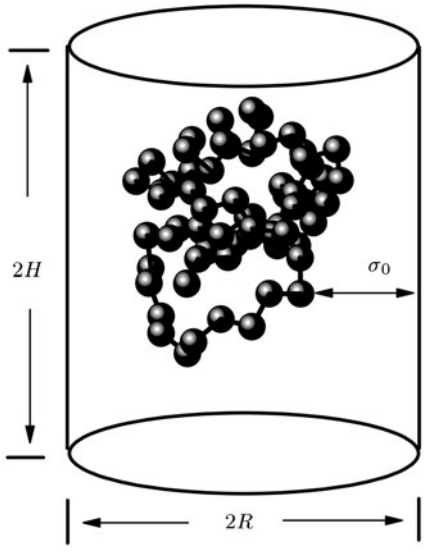

Fig. 1. Schematic diagram for substrate protein folding in a cylinder. Here $2R$ denotes the diameter and $2H$ denotes the height. There is a repulsive interaction when the distance between residues and wall is smaller than σ_0.

We first consider cylinders with equal radii and heights. In these cases, there is no orientational favourite for protein molecules due to the equal length scales in different directions. The thermodynamic stability of the substrate protein is depicted by the folding transition temperature T_f which is the peak temperature of the heat capacity, C_v. The higher the T_f value is, the more stable the protein is. In a small chamber of cylinder, the stability is found to be enhanced by confinement. The folding temperature T_f decreases monotonically (see the dot-dashed line in Fig. 2(b)). This indicates that the thermodynamic stability is enhanced compared with that of the bulk case, which is due to the reduction of the denatured state ensemble. This is consistent with former experimental findings[18] and theoretical results.[11]

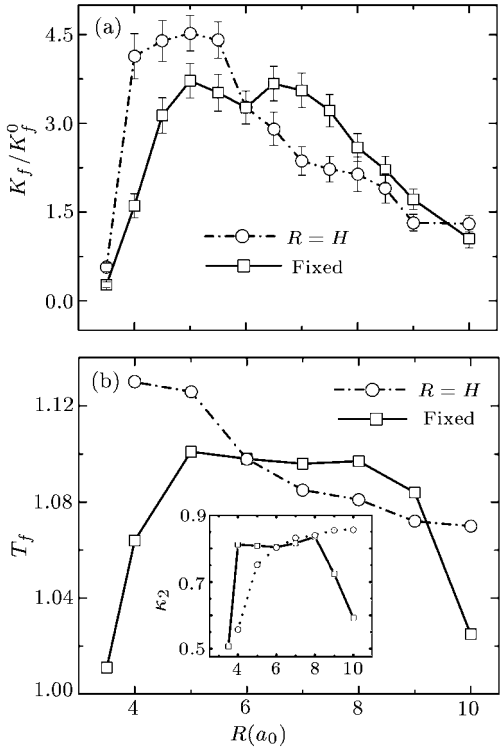

Fig. 2. The folding kinetics and thermodynamics of the protein in cylinders with equal R and H (dot-dashed line) and in cylinders with constant volume (solid line). (a) K_f and K_f^0 denote the folding rates in cylinder and in bulk. The folding rate is obtained over 100 trajectories which are simulated at T_f^0 (the folding transition temperature in bulk). (b) T_f versus R in the same cases as those for (a). The inset shows the corresponding folding cooperativity.

Another interesting effect which has not been studied before is that the folding cooperativity, κ_2, decreases as the cylinder shrinks (see the dot-dashed line in the inset in Fig. 2(b)). Here, κ_2 is defined as $2T_f\sqrt{k_B C_v(T_f)}/\Delta E_{\rm tot}$ where $\Delta E_{\rm tot}$ is the total change of energy between folded state and unfolded state.[19] Following with the shrink of cylinder, the height of barrier decreases severely. The folding in small-sized cylinders works more like a down-hill process, and a weak cooperativity appears. At the same time, due to the decrease of barrier, the folding kinetics is faster in this kind of down-hill type landscape, except for a very small cylinder which makes the protein unfoldable. As a remark, the foldability is greatly enhanced in cylinders.

When the values of the height and radius of a cylinder are different, the protein experiences a fold-

ing with spatial anisotropy. The folding can be different sometimes. As a simple example, we fix the volume of the cylinder, and vary the radius. A series of cylinders with different shapes, thin or wide, are created. The folding features in these cylinders are shown in Fig. 2. Except for the cases with small radius or heights, the folding temperature varies little. This might be due to almost the same effect on the denatured states of proteins by volume-fixed cylinders. More interestingly, the folding rate shows two peaks following the change of radii. One is almost consistent with the isotropic case at $R = 5$, and the other happens with a larger radius, $R = 6.5$. That is, the optimal situations happen in a thin cylinder ($R < H$) or a wide one ($R > H$). The isotropic case ($R = H$) is not the most favoured. In particular, the existence of the twin optima is quite out of intuition. This phenomenon may be attributed to the shape of substrate molecule. In a small space, the molecules generally locate their long axis in the largest dimension of the cylinder, which can gain the largest translational entropy. In our simulations, following the variations of radii, the molecules may change their orientations to fit the cylinder. In some compatible cases, the optimum may appear. Since the molecule is not spherical, there may be two favourite directions, one with large radius and the other with large height. For the isotropic condition, the radius $R = 6.5$ is in some degree larger than the optimal value. The constraint on molecule would deviate from the best situation. Therefore, there is not a peak of the folding rate at $R = 6.5$ for the isotropic cases. This is what we have observed.

To further investigate the effect of shape of the cylinder on the folding processes, the folding for various shapes and sizes of cylinders is simulated. The folding rates for various (R, H) pairs are shown in Fig. 3. It is found that the fastest folding corresponds to the cases $R \neq H$ as the volumes of cylinders are fixed, which indicates that proteins prefer to space with diverse spatial scales, rather than the $R = H$ case. The variation of the folding rate versus H has one peak for a fixed R, but there is an obvious shift of the peak for the cases with $R < 4.5$ and those with $R > 4.5$ (as shown in Fig. 3(b)). The peaks are at around the location $H = 4.794$ for $R < 4.5$ and $H = 3.730$ when $R > 4.5$. These observations indicate that the protein has different behaviour for various (R, H)s of cylinders. The equilibrium structures of proteins in various cases are also analysed, which suggests that the changes of folding behaviour for different radii may be attributed to the specific orientation of molecules. As the value of R is small and the value of H is large, the substrate protein has to locate itself along the axial direction to gain more translational entropy in the cylinder (see the upper sketch map in Fig. 3(b)). Therefore, the height of cylinder corresponding to the peak is large. That is, the substrate protein is sensitive to the confinement effect along the axial direction. Meanwhile, for a large R (> 4.5), the substrate protein can place both the long axes in the $x - y$ plane in the cylinder (see the lower sketch map in Fig. 3(b)). The compression along the axial direction would concern a short length scale of protein molecule. Consequently, the substrate protein is insensitive to the confinement along the axial direction. Note that the double-peak behaviour mentioned above, which is indicated as the dot-dashed line in Fig. 3(a), could be attributed to the two-stage feature. The twin peaks shows the interplay of the two stages. Quantitatively, the molecular shape can be detected partly from the variation of folding rate. H_P, which is the cylinder height corresponding to the peak of folding rate versus H for fixed R, shows the limit of compression along axial direction. It may be related to the size of molecule along this direction. Here, the ratio of average H_P for the cases with $R > 4.5$ to that for $R < 4.5$ is 0.778. This is consistent with the ratio of short axis of CI2 to its long axis (0.765), which further supports our picture for the folding in anisotropic cylinders. That is, there is an orientational transition

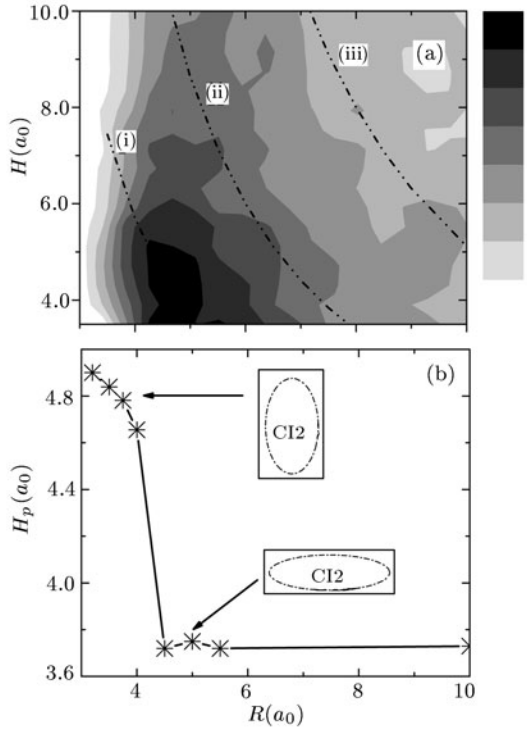

Fig. 3. (a) The contour map of folding rate for various R and H. The grey bar denotes the folding rate (the darker the bar is, the faster the folding is). The three dot-dashed lines correspond to the cases with R^2H equal to (i) 4.5^3, (ii) 6.0^3, and (iii) 8.0^3, respectively. (b) The cylinder height corresponding to the peak of folding rate. In the sketch maps, rectangle denotes the cylinder and ellipse denotes the protein.

for the protein molecule following the variation of the radius of cylinder.

In conclusion, we have studied the folding of protein in cylinders. Confinement produces a stable and fast folding, but a weak cooperativity. More interestingly, the proteins show different kinds of behaviour which depend on the radii and heights of cylinders. The optimal height shows a transition from the cases with radius $R > 4.5$ to those of $R < 4.5$. The comparison between the ratios of the optimal heights in the two regions and the data of molecular geometry suggests that the shape of the molecule is the basic reason of the different folding behaviour.

References

[1] Anfinsen C B 1973 *Science* **181** 223
[2] Dill K A et al 1995 *Protein Sci.* **4** 561
[3] Thirumalai D and Woodson S A 1996 *Acc. Chem. Res.* **29** 433
[4] Mirny L A, Abkevich V and Shakhnovich E L 1996 *Fold. Des.* **1** 103
[5] Otzen D E, Itzhaki L S and Fersht A R 1994 *Proc. Natl. Acad. Sci.* USA **91** 10422
[6] Fersht A R 1995 *Proc. Natl. Acad. Sci. U.S.A.* **92** 10869
[7] Minton A P 2000 *Curr. Opin. Struct. Biol.* **11** 34
[8] Ellis R J 2001 *Curr. Opin. Struct. Biol.* **11** 114
[9] Minton A P 1992 *Biophys. J.* **63** 1090
[10] Klimov D K, Newfield D and Thirumalai D 2002 *Proc. Natl. Acad. Sci. U.S.A.* **99** 8019
[11] Takagi F, Koga N and Takada S 2003 *Proc. Natl. Acad. Sci. U.S.A.* **100** 11367
[12] Clementi C, Nymeyer H and Onuchic J N 2000 *J. Mol. Biol.* **298** 937
[13] Qin M, Wang J and Wang W 2003 *Chin. Phys. Lett.* **20** 1883
[14] Liang H J and Wang Y Y 2002 *Chin. Phys. Lett.* **19** 1382
[15] Qin M, Wang J, Tang Y and Wang W 2003 *Phys. Rev.* E **67** 061905
[16] Hoang T X and Cieplak M 2000 *J. Chem. Phys.* **112** 6851
[17] Ferrenberg A M and Swendsen R H 1989 *Phys. Rev. Lett.* **63** 1195
[18] Eggers D K and Valentine J 2001 *Protein Sci.* **10** 250
[19] Kaya H and Chan H S 2000 *Proteins* **40** 637