

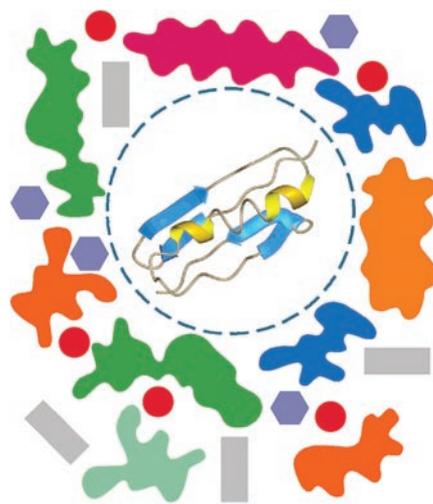
# Caging helps proteins fold

D. Thirumalai\*, Dmitri K. Klimov, and George H. Lorimer

Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742

How do proteins fold spontaneously? The quest to answer this question has led to significant developments on theoretical, experimental, and computational fronts in the last decade (1–7). A combination of approaches has provided a detailed understanding of the nature of pathways and the transition states that the polypeptide chain encounters as it traverses the rugged energy landscape. Even as our understanding of *in vitro* protein folding at infinite dilution has advanced, it has become urgent to address two additional issues of biological interest. (i) *In vivo* folding is not always a spontaneous event. A subset of proteins may require molecular chaperones. In an illuminating article in this issue of PNAS, Takagi *et al.* (8) provide a detailed study of five model proteins of varying native state architecture confined to cylindrical nanopores, which are meant to mimic the cavity of GroEL. Of all the molecular chaperones, the GroEL/GroES system from *Escherichia coli*, which assists folding of a fraction of cytosolic proteins, is the best understood. GroEL, a cylindrical barrel, consists of two heptameric rings stacked back-to-back giving rise to an unusual sevenfold symmetry about the axis of the cylinder (9). The annealing action of the chaperonin machinery (GroEL/GroES) is complex and involves large allosteric domain movements in GroEL in response to binding of the substrate protein (SP), ATP, and GroES (10, 11). The presence of a central cavity has led to the proposal that GroEL merely offers a protective environment in which SP folds as it would in infinite dilution. Indeed, for an undetermined duration out of the total of 10 sec of the chaperonin cycle, SP experiences confinement in the cylindrical cavity, whose maximum volume is  $\approx 175,000 \text{ \AA}^3$ . However, the annealing action of GroEL is due to the changes in the inner lining of the cavity that are triggered by SP, ATP, and GroES binding, which implies that GroEL plays an active role in the rescue of SP. (ii) Even proteins that fold spontaneously in cells do so only in a milieu that contains other biological macromolecules that serve as crowding agents (12). The volume fraction of the crowding agents (ribosomes, RNA, polysaccharides, etc.) can be in the 20–30% range (13). To a first approximation, the role of crowding agents may be modeled as confining the protein to a restricted

space (Fig. 1). Let us assume that the crowding agents merely exclude a fraction of the available volume to the polypeptide chain. If this is the case, then it is clear that the polypeptide chain would, with high probability, be found in a region free of the crowding agents; i.e., folding would take place in a restricted space (Fig. 1). To qualitatively understand *in vivo* folding it is important to develop a conceptual picture of how the principles that govern *in vitro* folding in infinite dilution change when other factors (molecular chaperones and crowding effects) are taken



**Fig. 1.** Schematic sketch of a protein in a crowded cellular environment. The crowding agents (ribosomes, RNAs, proteins, lipids, ions, etc.) are denoted by different shapes and colors. The polypeptide chain is effectively confined to a cavity, which is shown as a sphere. Confinement induces substantial native structure in proteins as indicated by partially formed  $\beta$ -strands and  $\alpha$ -helix (17).

into account. For this purpose, folding in nanopores can be a useful caricature of the more complex situations that the polypeptide chain encounters under cellular conditions.

Few *in vitro* experiments have monitored directly the effect of confinement on the thermodynamics of protein folding. Eggers and Valentine (14) showed that the melting temperature of  $\alpha$ -lactalbumin increases by nearly  $30^\circ\text{C}$  relative to the bulk value on encapsulation in the nanopores of silica glass. Inspired in part by these observations, theoretical and computational studies have begun to critically examine the effects of con-

finement on the thermodynamics and kinetics of protein folding (15–19).

For computational efficiency, Takagi *et al.* use Go-model  $\alpha$ -carbon representation of proteins and Langevin simulations (20) to monitor the kinetics of folding. The interactions between the walls of the cavity and the model proteins are purely repulsive, making the cavity inert. Folding in inert cavities is most directly relevant for *in vitro* experiments that probe confinement effects. In the inert cage the folding temperatures are markedly higher than in the bulk (8, 17). Thus, confinement increases the stability of the native states, in accord with previous theoretical studies (16). The increase in the native state stability arises because the conformations of the polypeptide chain with  $R_g/L > 1$  (where  $R_g$  is the radius of gyration of the protein and  $L$  is the characteristic length of the confining region) are disallowed. As a result, the free energy of the unfolded state, relative to the bulk,  $\Delta F_U$ , increases. Assuming that the native state free energy is unaffected by confinement, which is likely to be the case in inert pores as long as  $R_g^0/L < 1$  (where  $R_g^0$  is the native value of  $R_g$ ), it follows that the stability of the folded state in restricted spaces is higher than in the bulk. These arguments suggest that crowding agents (Fig. 1) should also promote collapse of the polypeptide chain and stabilize the native structure. Depletion of the available space by crowding agents causes chain collapse. Thus, the effect of crowding should be similar to folding in confined spaces as long as the volume fraction of the crowding agents is not too large.

The enhanced stability in the confined spaces is reflected in higher folding temperatures (8, 17). The substantial increase in the confinement-induced folding temperature comes at the expense of decreased cooperativity (see especially figure 3a in ref. 8). As  $L$  decreases, the transition to the folded state becomes broader. If the polypeptide chain in the unfolded state is approximated as a random coil, then the change in the free energy of the confined (to slits or cylinders) unfolded

See companion article on page 11367.

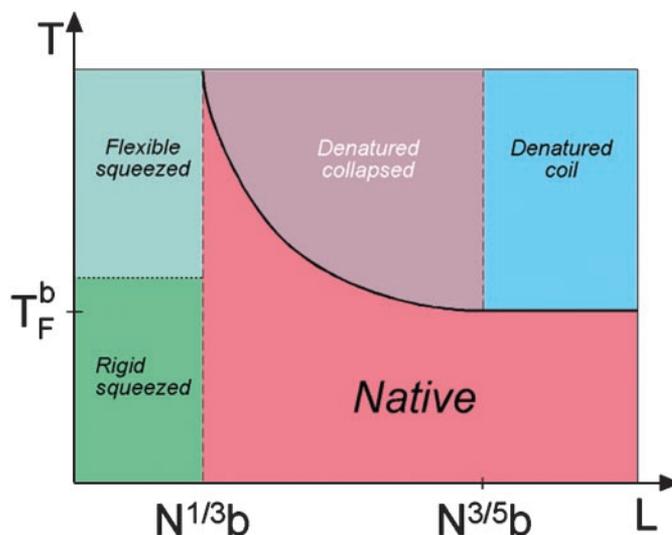
\*To whom correspondence should be addressed. E-mail: thirum@glue.umd.edu.

© 2003 by The National Academy of Sciences of the USA

state with respect to the bulk ( $L \rightarrow \infty$ ) is  $\Delta F_U \sim k_B T N (b/L)^{1/\nu}$ , where  $b$  is the distance between  $C_\alpha$  atoms ( $\approx 4 \text{ \AA}$ ),  $N$  is the number of amino acids, the Flory exponent  $\nu$  is defined from  $R_g \sim bN^\nu$ , and  $k_B$  is the Boltzmann constant (21). Assuming that the free energy change of the native state is negligible, Takagi *et al.* argue, in rough accord with the simulations, that the change in the folding temperature  $\Delta T_F = T_F - T_F^b \sim (R_g/L)^3$ , where they have naively assumed that  $R_g \sim bN^{1/3}$ . A more careful analysis (21) that accounts for the intramolecular interactions within a confined chain shows that the confinement free energy cost is  $\Delta F_U \sim k_B T (R_g/L)^{3/(3\nu-1)}$ . From this it follows that  $\Delta T_F \sim L^{-(15/4)}$ , a result that would not be inconsistent with the simulations of Takagi *et al.*

Not only does confinement in a finite-sized cylinder stabilize the protein, it also leads to modest (less than a factor of 10) rate enhancement relative to the bulk value (8, 17, 22). Because the unfolded state is destabilized in confined spaces, the average folding free energy barrier becomes smaller, thus enhancing the folding rates,  $k_F$ . Somewhat surprisingly, Takagi *et al.* find that the rate enhancement is greater for proteins with large fraction of long-range tertiary contacts. In the bulk, the entropy penalty for establishing long-range contacts is large. As a result,  $k_F$  for proteins with a large fraction of tertiary long-range contacts are usually smaller than for those with predominantly local tertiary contacts. In confined spaces, it is likely that a few of the long-range tertiary contacts are forced to form. The decrease in the entropic costs in the cylindrical nanopores explains the greater increase in  $k_F$  for proteins with more complex topology.

Confinement of proteins to a pore results in collapse of the chain. Consequently, the search for the native conformation takes place only within the ensemble of compact structures. From this argument it would follow that as  $L$  decreases, the folding rate should monotonically increase. Surprisingly, the simulations show that as  $L$  becomes less than a characteristic value ( $\approx bN^{1/3}$ ), the folding rates start to decrease even for  $L$  values, at which the native state is not compromised (8, 17). However, the folding mechanism changes as the confinement-induced folding rates become less than their peak value. For  $L \approx bN^{1/3}$ , the polypeptide chain may get squeezed so much that chain reconfiguration, which may be necessary to cross the free energy barrier, becomes difficult. In this range of  $L$ ,  $k_F$  is expected to decrease. The arguments rationalizing the nonmonotonic dependence of  $k_F$  on  $L$  apply only if the nanopores are inert.



**Fig. 2.** Idealized phase diagram of confined two-state proteins as a function of temperature  $T$  and cavity size  $L$ . Localization of the substrate protein in GroEL is an example of confinement. The cavity is characterized by a single length scale  $L$ . For  $L > N^{3/5}b$  ( $N$  is the number of amino acids and  $b$  is the monomer size) there are only two states, namely, native and denatured coil. For  $N^{1/3}b < L < N^{3/5}b$  the native state remains intact, whereas confinement induces collapse of the unfolded polypeptide chain (denoted by denatured collapsed state). The thick black line indicates phase transition between denatured and native states and is given by  $T \sim L^{-(15/4)}$ . For  $L < N^{1/3}b$  the native state is perturbed. At high temperatures the structures are squeezed and flexible, whereas at low temperatures they are rigid. The boundary between the two phases for  $L < N^{1/3}b$  is only suggestive. The folding temperature in the bulk ( $L \rightarrow \infty$ ) is  $T_F^b$ .

The key findings (the enhanced stability and folding rates) of the confinement-induced folding may be rationalized by using simple ideas. However, full understanding of confinement-induced folding, at the molecular level, requires structural characterization of the denatured state ensemble (DSE) and the transition state ensemble (TSE). For a range of  $L$  values considered in the simulations, the conformations belonging to the native basin of attraction (NBA) are not greatly affected by confinement (8, 17). Thus, the profound effect of confinement on folding must arise because of significant changes in the DSE and, perhaps, the TSE. Simulations of  $\beta$ -hairpin confined to a spherical cavity (17) suggest that there is considerable structure, perhaps even native-like, in the DSE of proteins in restricted spaces. The changes in the TSE on confinement have not been directly probed by either experiments or simulations. Because the DSE itself has many native-like characteristics, it is tempting to speculate that the average structure in the TSE must have enhanced native content as compared to the bulk case. This conclusion is consistent with the limited analysis of the TSE performed by Takagi *et al.* Detailed structural characterization of the DSE and the TSE might explain how nature manipulates the plasticity of both the DSE and TSE

to fold proteins spontaneously in the crowded cellular environment.

Confinement of SP to a restricted region with the volume  $\sim L^3$  can lead to phases that are absent in the bulk. For optimized protein sequences, folding in the bulk can be characterized as a cooperative all-or-none transition. With  $L$  as an additional variable, one can map the phase diagram for two-state folders in the  $(T, L)$  plane. We envision a rich phase diagram for confined two-state folders (Fig. 2) as temperature  $T$  and  $L$  are varied. Besides the familiar native and denatured states, additional phases emerge as  $L$  varies. So far, simulations have been used to characterize only the structural features of the denatured collapsed phase (Fig. 2). Remarkably, even at modest confinements the denatured collapsed states have significant long-range structure (17). More importantly, the structure persists as the extent of denaturation increases, a behavior that is drastically different from what is observed in the bulk (17). The nature of the predicted phases in Fig. 2, besides the denatured collapsed states, has not yet been characterized. The advances in nanotechnologies, which can be used to produce pores of arbitrary shapes, will enable us to fully probe the nature of the phases of polypeptide chains in confined spaces. Such experiments will unravel the precise way nature may have utilized the many confinement-induced phases to keep proteins happy in cages!

1. Dill, K. A. & Chan, H. S. (1997) *Nat. Struct. Biol.* **4**, 10–19.
2. Onuchic, J. N., Luthey-Schulten, Z. A. & Wolynes, P. G. (1997) *Annu. Rev. Phys. Chem.* **48**, 545–600.
3. Thirumalai, D., Klimov, D. K. & Woodson, S. A. (1997) *Theor. Chem. Acc.* **1**, 149–156.
4. Dinner, A., Sali, A., Smith, L., Dobson, C. & Karplus, M. (2000) *Trends Biochem. Sci.* **25**, 331–339.
5. Eaton, W. A., Munoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R. & Hofrichter, J. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359.
6. Mirny, L. & Shakhnovich, E. (2001) *Annu. Rev. Biophys. Biomol. Struct.* **30**, 361–396.
7. Daggett, V. & Fersht, A. R. (2003) *Trends Biochem. Sci.* **28**, 18–25.
8. Takagi, F., Koga, N. & Takada, S. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 11367–11372.
9. Xu, Z. & Sigler, P. B. (1999) *J. Struct. Biol.* **124**, 129–141.
10. Thirumalai, D. & Lorimer, G. H. (2001) *Annu. Rev. Biophys. Biomol. Struct.* **30**, 245–269.
11. Horovitz, A., Fridmann, Y., Kafri, G. & Yifrach, O. (2001) *J. Struct. Biol.* **135**, 104–114.
12. Minton, A. P. (2000) *Curr. Opin. Struct. Biol.* **11**, 34–39.
13. Ellis, R. J. (2001) *Curr. Opin. Struct. Biol.* **11**, 114–119.
14. Eggers, D. K. & Valentine, J. S. (2001) *J. Mol. Biol.* **314**, 911–922.
15. Betancourt, M. R. & Thirumalai, D. (1999) *J. Mol. Biol.* **287**, 627–644.
16. Zhou, H.-X. & Dill, K. (2001) *Biochemistry* **40**, 11289–11293.
17. Klimov, D. K., Newfield, D. & Thirumalai, D. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 8019–8024.
18. Gorse, D. (2001) *Biopolymers* **59**, 411–426.
19. Friedel, M., Sheeler, D. J. & Shea, J.-E. (2003) *J. Chem. Phys.* **118**, 8106–8113.
20. Honeycutt, J. D. & Thirumalai, D. (1992) *Biopolymers* **32**, 695–709.
21. Grosberg, A. Y. & Khokhlov, A. R. (1994) *Statistical Physics of Macromolecules* (AIP Press, New York).
22. Todd, M. J., Lorimer, G. H. & Thirumalai, D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 4030–4035.