Biochemistry

 Volume 40, Number 38 September 25, 2001

New Concepts

Stabilization of Proteins in Confined Spaces†

Huan-Xiang Zhou*,‡ and Ken A. Dill*,§

Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104, and Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94118

Received June 5, 2001; Revised Manuscript Received July 26, 2001

ABSTRACT: We present theory showing that confining a protein to a small inert space (a "cage") should stabilize the protein against reversible unfolding. Examples of such spaces might include the pores within chromatography columns, the Anfinsen cage in chaperonins, the interiors of ribosomes, or regions of steric occlusion inside cells. Confinement eliminates some expanded configurations of the unfolded chain, shifting the equilibrium from the unfolded state toward the native state. The partition coefficient for a protein in a confined space is predicted to decrease significantly when the solvent is changed from native to denaturing conditions. Small cages are predicted to increase the stability of the native state by as much as 15 kcal/mol. Confinement may also increase the rates of protein or RNA folding.

A protein inside a confined space will be stabilized by folding forces different from those for proteins in "bulk" solutions, i.e., away from any confining walls. In particular, some expanded configurations of the unfolded chain will not be allowed inside the confined space, due to excluded volume. Our interest in this problem is partly motivated by recent experimental work (*1*−*3*). Wei et al. (*1*) have developed techniques for encapsulating proteins in mesoporous host materials with tunable pore sizes, while Kumar and Chaudhari (*2*) have developed techniques for immobilizing proteins in the spaces between two parallel layers of α-zirconium. Of particular interest is the work of Eggers and Valentine (*3*), who encapsulated proteins in the pores of a silica glass and found that the melting temperature of α-lactalbumin was increased by as much as 32 °C. In this paper, we consider the folding equilibrium of proteins inside cages of various sizes and shapes. Figure 1 illustrates the restricting effect of the cage on the configurational space of the native (N) and unfolded (U) states.

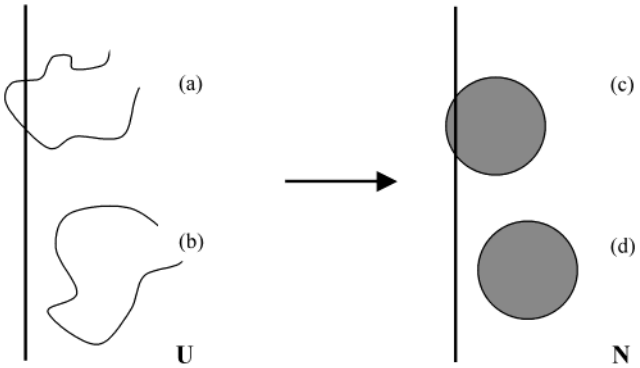

FIGURE 1: Effect of a wall or boundary (thick vertical line) on the configurations of a protein. In the unfolded state U shown at the left, configuration a is not viable since the protein chain would pass across the steric boundary. Configuration b, however, is viable. In the native state N shown at the right, the protein is modeled as a sphere, so there is only a single configuration. The native protein cannot be located in position c but can be located in position d.

† This work was supported in part by the National Institutes of Health (Grant GM58187 to H.-X.Z. and Grant GM34993 to K.A.D.).

* To whom correspondence should be addressed. H.-X.Z.: phone, (215) 895-2716; fax, (215) 895-5934; e-mail, hxzhou@einstein.drexel.edu. K.A.D.: phone, (415) 476-9964; fax, (415) 502-4222; e-mail, dill@maxwell.ucsf.edu.

‡ Drexel University.

§ University of California.

10.1021/bi0155504 CCC: $20.00
Published on Web 08/29/2001

The idea that confinement may shift the equilibrium between two states of a molecule with different shapes was first investigated by Minton (*4*). Minton found that a molecule that switches between a spherical shape and a cubic shape when confined in a cage will have an equilibrium constant significantly different from that in bulk solution because the volumes accessible to the two shapes are very different. Minton (*5*) also considered the effect of volume exclusion by crowding on the folding equilibrium of a protein. Here blocking of expanded configurations of the unfolded chain is due to the presence of other stable macromolecules at high concentrations (as opposed to the walls of a cage). The effect was found to be relatively small ($\Delta\Delta G \sim -2$ kcal/mol at room temperature with high cosolute concentrations).

We assume that the protein in the unfolded state can be modeled as a random-flight Gaussian chain. Let the number of residues in the protein be N and the effective bond length be b. The probability density $G(\mathbf{R}, \mathbf{R}_0, n)$ that the chain starting at position $\mathbf{R}_0$ will end at position $\mathbf{R}$ after n residues satisfies a diffusion equation (*6*−*8*):

$$\frac{\partial G(\mathbf{R}, \mathbf{R}_0, n)}{\partial n} = \frac{b^2}{6}\nabla^2 G(\mathbf{R}, \mathbf{R}_0, n) \qquad (1)$$

We treat n as a continuous variable even though it is discrete [a good approximation when $N \gg 1$ (*9*); see further discussion below]. n plays the role of time in the diffusion equation. Since the chain cannot cross the cage wall, we have $G(\mathbf{R}, \mathbf{R}_0, n) = 0$ at the wall (*6*−*8*). The partition function for the ensemble of chain configurations with every residue restricted to be within the cage is (*8*)

$$Z_U = \int d\mathbf{R}\, d\mathbf{R}_0\, G(\mathbf{R}, \mathbf{R}_0, N) \qquad (2)$$

where the integration is over the interior cage volume. In eq 2, it is implicitly assumed that the probability density for the first segment ($n = 0$) is uniform within the cage.

Folding between Two Parallel Walls

We first consider a denatured protein chain confined between two infinite parallel walls at $z = 0$ and $z = s$. This is a model of the experimental system devised by Kumar and Chaudhari (*2*). Solution of the diffusion equation subject to absorbing boundary conditions at the walls gives (*6*, *10*)

$$Z_U = \frac{8s}{\pi^2}\sum_{k=1,3,5,\ldots}\frac{1}{k^2}\exp(-\pi^2 k^2/4\beta s^2) \qquad (3)$$

where $\beta = {}^3/_2 Nb^2$. When the wall separation s is much larger than the average dimension of the chain ($\sim N^{1/2}b$), the exponential function in eq 3 approaches 1 and we have Z_U(2 walls) $\approx s$. Thus, in the limit of large cage size, the chain freely explores the space between the two walls. Therefore, the partition function equals the volume of the cage. In the opposite limit where $s \ll N^{1/2}b$, only the first term in the summation is important and Z_U(2 walls) $\approx (8s/\pi^2)\exp(-\pi^2 Nb^2/6s^2)$.

In the random-flight model, the residues are assumed to be point particles. Hence, each residue can sample every point between the two walls. The finite size of the residue could readily be accounted for by moving the location of each wall inward by a distance ($\sim b$) equal to the effective size of the residue. We neglect this small correction here. It should also be noted that the treatment of n as a continuous rather discrete variable introduces an outward shift of the boundary (*11*). The neglect of the finite residue size and the treatment of n as a continuous variable act in opposite ways on the size of the allowed space; for simplicity, we neglect both effects.

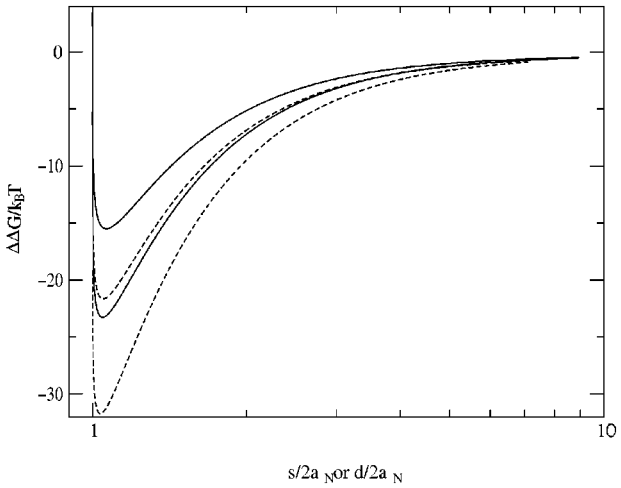

FIGURE 2: Effect of confinement on the folding free energy as a function of the cage size. The top two curves are for a cubic cage, and the bottom two curves are for a spherical cage. Solid curves are for $N = 100$, and dashed ones are for $N = 200$. The radius of the protein in the native state (a_N) was given by $3.73N^{1/3}$ (*13*). Cage size (in units of $2a_N$) is given on a log scale.

We model the native state as a sphere of radius a_N. The center of the protein is excluded from a boundary layer of that thickness. The partition function for the native protein is just the effective volume ($V_N = s - 2a_N$) of the cage (*4*). The cage changes the difference in free energy between the native and unfolded states by

$$\Delta\Delta G = -k_B T \ln(V_N/Z_U) \qquad (4)$$

where k_B is the Boltzmann constant and T is the absolute temperature.

Folding Inside a Cube or a Sphere

For a protein confined in a cube with side length s, the partition function is now the product of three independent terms, each of the type described above, representing x, y, and z components (*7*). Hence, the change in folding free energy is 3 times that in the two-wall situation.

For an unfolded chain confined to a sphere with diameter d, the partition function is (*6*, *10*)

$$Z_U = \frac{d^3}{\pi}\sum_{k=1,2,3,\ldots}\frac{1}{k^2}\exp(-\pi^2 k^2/\beta d^2) \qquad (5)$$

The effective volume V_N for the native state is $\pi(d - 2a_N)^3/6$.

Figure 2 shows $\Delta\Delta G$ as a function of the cage size. When the cage decreases to the size of the native state, V_N decreases more steeply than Z_U. Hence, $\Delta\Delta G$ exhibits a minimum (corresponding to maximal stabilization) when the cage is just slightly larger than the native protein. In this range of cage sizes, only the first term in the summation of eq 3 or

5 is important. Thus, we have $\Delta\Delta G/k_BT(\text{cube}) = -\pi^2Nb^2/2s^2 - 3\ln(1 - 2a_N/s) + 3\ln(8/\pi^2)$ and $\Delta\Delta G/k_BT(\text{sphere}) = -2\pi^2Nb^2/3d^2 - 3\ln(1 - 2a_N/d) + \ln(6/\pi^2)$. Setting the derivative with respect to s or d to zero, we obtain the cage size for maximal stabilization $s_M/2a_N$ and $d_M/2a_N = 2[1 - (1 - \gamma)^{1/2}]/\gamma \approx 1 + \gamma/4$, where $\gamma = 12(2a_N)^2/\pi^2Nb^2$ for a cube and $9(2a_N)^2/\pi^2Nb^2$ for a sphere. The maximal change in folding free energy $\Delta\Delta G_M/k_BT \approx -6/\gamma - 3\ln(\gamma/4) + C$, where $C = 3 + 3\ln(8/\pi^2)$ for a cube and $3 + \ln(6/\pi^2)$ for a sphere. When $N = 200$, we have $\Delta\Delta G_M = -21.6k_BT$ and $-31.7k_BT$, or -12.8 and -18.8 kcal/mol, respectively, at room temperature when the effective bond length (b) for the unfolded state is chosen to be 8 Å. This choice of bond length is justified by hydrodynamic measurements on the unfolded states of a large number of proteins (*12*). A smaller bond length would lead to less stabilization by the confinement. However, even when the effective bond length is 5 Å, $\Delta\Delta G_M = -2.4$ and -4.4 kcal/mol for the cube and sphere, respectively. It is also interesting to note that even when the cage size is 6 times that of the native protein, the confinement still provides stabilization of the native state on the order of k_BT.

Analysis of the Eggers−Valentine Experiment

Confinement is predicted to decrease the folding free energy and lead to an increase in the melting temperature. The folding free energy in bulk solution for many small proteins has been described by (*14*)

$$\Delta G_0 = -1.54N + 4.35 \times 10^{-3}NT - \Delta C_p[T - 385 - T\ln(T/385)] \quad (6)$$

where ΔC_p is the specific heat upon unfolding. For α-lactalbumin [which has 123 (=N) residues], Eggers and Valentine (*3*) found a melting temperature of ~343 K (corresponding to $\Delta G_0 = 0$) in bulk solution. This allows us to determine the unfolding specific heat change (ΔC_p) to be 2.48 kcal mol^{-1} K^{-1}. If the folding free energy is decreased by $15k_BT$ due to confinement (see Figure 2), the melting temperature will become 374 K. Hence we predict an increase of 31 °C, which agrees with the measurements of Eggers and Valentine.

Folding Inside a Cylinder

For an unfolded chain inside a cylinder with diameter d and infinite length, the partition function is (*6*, *10*)

$$Z_U = \pi d^2 \sum_{k=1,2,3,\ldots} \frac{1}{x_k^2} \exp(-x_k^2/\beta d^2) \quad (7)$$

where x_k are the roots of $J_0(x)$, the Bessel function of the first kind of order zero.

The central cavity of the GroE chaperonin is roughly cylindrical (*15*). It is blocked at the top by the bound GroES and at the bottom by the C-termini of the subunits of GroEL. An important caveat is that the interior of GroEL may not be inert (the protein may bind to the chaperonin walls). Hence, the theory is just intended as an indication of the confinement component of the free energy, since the interactions surely require a more detailed model. If the two ends of the cylinder are capped by flat surfaces with separation h, the partition function becomes

$$Z_U(\text{capped cylinder}) = Z_U(\text{infinite cylinder})Z_U(2\text{ walls}) = \frac{8d^2h}{\pi} \sum_{k=1,2,3,\ldots} \frac{1}{x_k^2} \exp(-x_k^2/\beta d^2) \times \sum_{k=1,3,5,\ldots} \frac{1}{k^2} \exp(-\pi^2k^2/4\beta h^2) \quad (8)$$

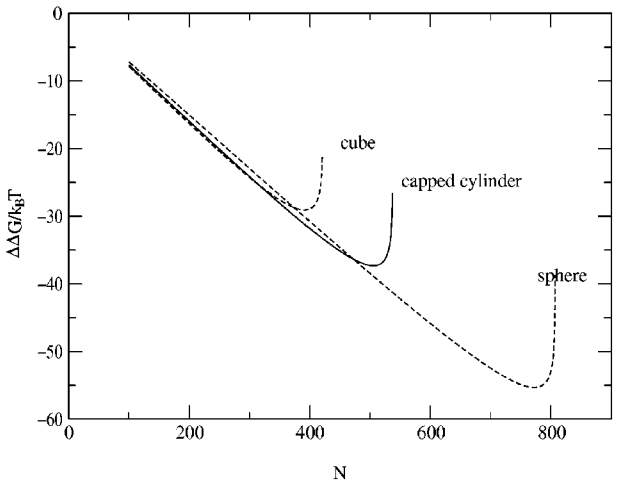

FIGURE 3: Effect of confinement on the folding free energy as a function of the chain length of the polypeptide. The effective bond length for the unfolded state b was 8 Å. The radius of the protein in the native state (a_N) was given by $3.73N^{1/3}$. The dimensions of the cages were as follows: $s = 55.9$ Å for the cube, $d = h = 60.9$ Å for the cylinder, and $d = 69.4$ Å for the sphere. These give the same cage volume of 175 000 Å^3.

Figure 3 shows how the folding free energy depends on the protein chain length within a capped cylinder with a volume of 175 000 Å^3 (roughly the size of the central cavity in the GroE system). It appears that there is an optimal chain length for which the stabilization by the confinement is maximized. The maximal stabilization is as much as 22 kcal/mol. It is interesting to note that the set of newly synthesized polypeptides, with molecular masses between 10 and 55 kDa, that are observed to pass through the GroE system in vivo (*16*) falls within the range of chain lengths with negative $\Delta\Delta G$ values in the capped cylinder cage. The upper limit for chain length can be estimated by setting $2a_N$ equal to the cylinder diameter d (=height h), and the result is $N = 544$. In bacteriophage T4, an analogue of GroES, Gp31, has a larger central cavity and thus can accommodate larger proteins (*17*). Compared to a capped cylinder, a sphere has a wider "operating" range of chain lengths (see Figure 3), simply because the latter can better accommodate the spherical native protein. A cube, on the other hand, has an even narrower operating range.

Confinement Effect of a Single Wall

Even when a protein approaches a single planar wall, it can affect the equilibrium between the native and unfolded states (*18*). Consider the situation in which the unfolded chain has its center of mass a distance s away from the wall. A

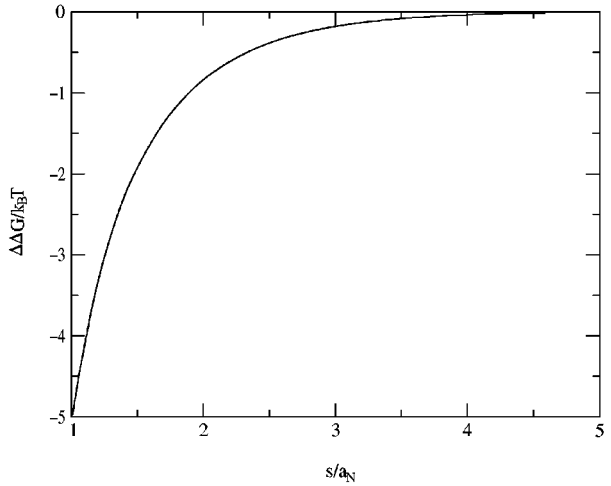

FIGURE 4: Change in the folding free energy due to the presence of a boundary wall. The distance s from the center of mass in either the unfolded or native state is measured in units of the radius of the native protein. The result for $N = 200$ is shown.

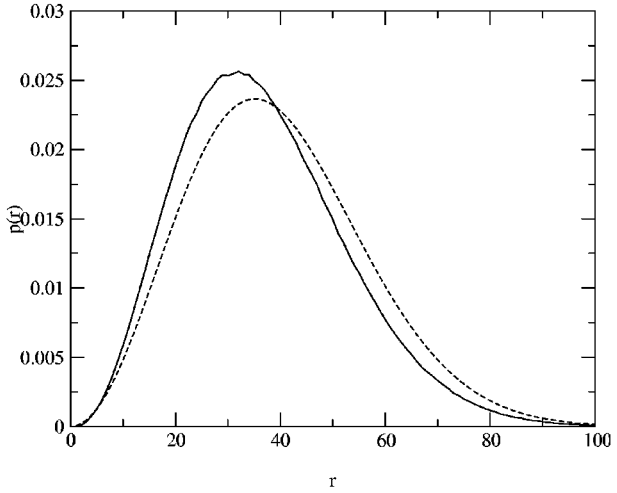

FIGURE 5: Distribution of distances (in angstroms) between the two N-termini of a coiled coil. The two C-termini are disulfide-bonded (with a distance l of 5 Å) and fixed to a planar wall ($z_0 = 0$). The solid curve shows the result of eq 12 with $N = 40$, whereas the dashed curve is the distribution for a Gaussian chain with $N = 80$.

fraction of the configurations will be eliminated by the presence of the wall. Let the allowed fraction be f_U. Provided that the center of mass is more than a radius a_N away from the wall, the protein will not lose configurational entropy in the native state. The change in folding stability by the wall is thus

$$\Delta\Delta G = k_B T \ln(f_U) \tag{9}$$

which applies when $s > a_N$.

Because of the constraint on the center of mass position (which involves all the residue positions), we are unaware of an analytical solution for f_U. However, f_U can be easily obtained by sampling configurations of a Gaussian chain. When the center of mass for each configuration was moved to the origin, f_U was calculated as the fraction of configurations with all the N residues having z coordinates greater than $-s$. Figure 4 shows the corresponding free energy change for the folding equilibrium of a protein with 200 residues. Near the wall, the native state is stabilized by 3 kcal/mol.

If the first residue (with $n = 0$) in the protein chain is tethered to the wall, the distribution of the last residue (with $n = N$) in the unfolded state is affected by the steric exclusion by the wall. The probability density for the last residue is (*19*)

$$G(\mathbf{R}, \mathbf{R}_0, N) = \left(\frac{\beta}{\pi}\right)^{3/2}\{\exp[-\beta(z - z_0)^2] - \exp[-\beta(z + z_0)^2]\} \exp[-\beta(x^2 + y^2)] \tag{10}$$

where $\mathbf{R}_0 = (0, 0, z_0)$ is the position of the first residue. Single-molecule techniques now offer opportunities to directly probe the probability density $G(\mathbf{R}, \mathbf{R}_0, N)$. In recent experiments, Hochstrasser et al. (*20*, *21*) attached the disulfide-bonded C-termini of the coiled coil GCN4-p1 to a surface. Suppose that the C-terminus of one chain is fixed at $\mathbf{R}_0 = (0, 0, z_0)$. Then in the unfolded state, the N-terminus will be distributed according to eq 10. If the C-terminus of the other chain is located at $\mathbf{R}'_0 = (l, 0, z_0)$, the corresponding N-terminus will be distributed according to

$$G'(\mathbf{R}', \mathbf{R}'_0, N) = \left(\frac{\beta}{\pi}\right)^{3/2}\{\exp[-\beta(z' - z_0)^2] - \exp[-\beta(z' + z_0)^2]\} \exp\{-\beta[(x' - l)^2 + y'^2]\} \tag{11}$$

The distance r between the two N-termini will have a distribution

$$p(r) = \int d\mathbf{R}\, d\mathbf{R}'\, \delta(|\mathbf{R} - \mathbf{R}'| - r) G'(\mathbf{R}, \mathbf{R}_0, N) G'(\mathbf{R}', \mathbf{R}'_0, N)/Q \tag{12}$$

where $Q = [\mathrm{erf}(\beta^{1/2} z_0)]^2$ (a normalization factor).

In Figure 5, we compare the distribution $p(r)$ with that of an unperturbed chain in solution. It can be seen that $p(r)$ shifts to smaller r values. In a confined space, the two N-termini are more likely to find each other than they would in bulk solution. As such, the folding of the coiled coil may be accelerated by confinement.

Partition Coefficient in Confined Space

As noted above, confinement shifts the folding equilibrium toward the native state. Now we reverse the logic. How does the partitioning of a protein into a confined space depend on solvent conditions? Shifting the solvent toward denaturing conditions expands the chain, leading to a diminished propensity to partition into confined spaces. The partition function for the unfolded chain in a confined space with volume V is Z_U (see eq 2). In the bulk solution, the partition function for the chain occupying the same volume will be V. The corresponding quantities for the folded chain are $V_N \exp(-\Delta G_0/k_B T)$ and $V \exp(-\Delta G_0/k_B T)$, respectively, where ΔG_0 is the folding free energy in bulk solution. The partition coefficient is thus

$$K = [Z_U + V_N \exp(-\Delta G_0/k_B T)]/[V + V \exp(-\Delta G_0/k_B T)] \tag{13}$$

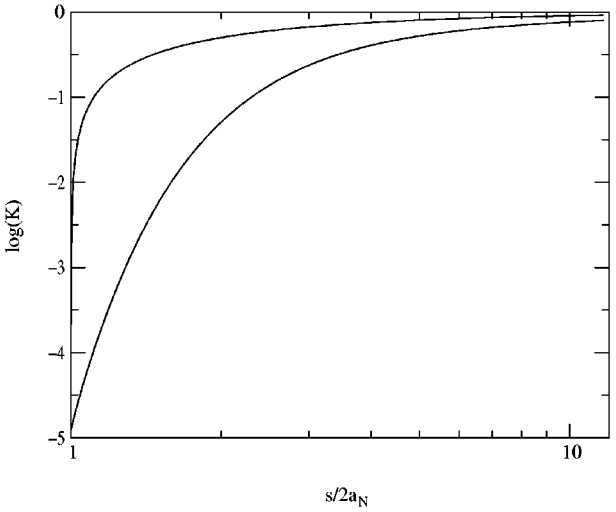

FIGURE 6: Partition coefficient for a protein in a space between two parallel walls. The top curve is for a native solvent condition ($\Delta G_0 = -10k_BT$). The bottom curve is for a denaturing solvent condition ($\Delta G_0 = 10k_BT$). Adding denaturant leads to chain expansion, which leads to a tendency for the chain to avoid the confined space.

We note that Casassa (*6*) many years ago considered the partition coefficient of the unfolded chain (equivalent to eq 13 without the second terms in both the numerator and denominator).

Figure 6 shows the predicted partition coefficient for a protein with 200 residues (N) between two parallel walls. For native solvent conditions, where $\Delta G_0 = -10k_BT$, the partition coefficient is 0.5 at a wall separation s of $4a_N$. When the solvent becomes denaturing with a ΔG_0 of $10k_BT$, the partition coefficient of the protein into the cage decreases by a factor of 10.

Our simple model here treats the denatured protein as a random-flight chain, neglecting intrachain excluded-volume interactions (i.e., steric clashes) among the residues [see, e.g., Pappu et al. (*22*)]. Our treatment, therefore, is most applicable to chains having relatively expanded denatured states.

In summary, we have shown that confined spaces should increase the folding stabilities of proteins. The theoretical predictions presented here appear to be in accord with recent experiments (*3*). Proteins may have increased stabilities inside the pores within chromatography columns, Anfinsen cages, the interiors of ribosomes, or regions of steric occlusion inside cells. To the extent that confinement shifts the folding equilibrium toward the native state, it may also speed the folding process (provided that the folding process does not involve the transient unfolding of a misfolded compact intermediate).

ACKNOWLEDGMENT

We thank Yen Wei for inspiring us to think about the problem addressed in this paper.

REFERENCES

1. Wei, Y., Xu, J., Feng, Q., Dong, H., and Lin, M. (2000) *Mater. Lett. 44*, 6−11.
2. Kumar, C. V., and Chaudhari, A. (2000) *J. Am. Chem. Soc. 122*, 830−837.
3. Eggers, D. K., and Valentine, J. S. (2001) *Protein Sci. 10*, 250−261.
4. Minton, A. P. (1992) *Biophys. J. 63*, 1090−1100.
5. Minton, A. P. (2000) *Biophys. J. 78*, 101−109.
6. Casassa, D. F. (1967) *Polym. Lett. 5*, 773−778.
7. Edwards, S. F., and Freed, K. F. (1969) *J. Phys. A 2*, 145−150.
8. Doi, M., and Edwards, S. F. (1986) *The Theory of Polymer Dynamics*, Clarendon Press, Oxford, U.K.
9. Dill, K. A., and Bromberg, S. (2002) *Molecular Driving Forces*, Garland Publishing.
10. Crank, J. (1975) *The Mathematics of Diffusion*, Clarendon Press, Oxford, U.K.
11. Jaeckel, A., and Dayantis, J. (1994) *J. Phys. A 27*, 7719−7731.
12. Tanford, C. (1968) *Adv. Protein Chem. 23*, 121−282.
13. Vijayakumar, M., and Zhou, H.-X. (2000) *J. Phys. Chem. B 104*, 9755−9764.
14. Creighton, T. E. (1993) *Proteins*, W. H. Freeman and Co., New York.
15. Xu, L., Horwich, A. L., and Sigler, P. B. (1997) *Nature 388*, 741−750.
16. Ewalt, K. L., Hendrick, J. P., Houry,, W. A., and Hartl, F. U. (1997) *Cell 90*, 491−500.
17. Hunt, J. F., van der Vies, S. M., Henry, L., and Deisenhofer, J. (1997) *Cell 90*, 361−371.
18. Dill, K. A., and Alonso, D. O. V. (1988) in *Protein Structure and Protein Engineering* (Huber, R., and Winnacker, E. L., Eds.) Vol. 39, pp 51−58, Springer-Verlag, Berlin.
19. DiMarzio, E. A. (1965) *J. Chem. Phys. 42*, 2101−2106.
20. Jia, Y., Talaga, D. S., Lau, W. L., Lu, H. S. M., DeGrado, W. F., and Hochstrasser, R. M. (1999) *Chem. Phys. 247*, 69−83.
21. Talaga, D. S., Lau, W. L., Roder, H., Tang, J., Jia, Y., DeGrado, W. F., and Hochstrasser, R. M. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 13021−13026.
22. Pappu, R. V., Srinivasan, R., and Rose, G. D. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 12565−12570.

BI0155504