A Simple Semiempirical Model for the Effect of Molecular Confinement upon the Rate of Protein Folding

Manajit Hayer-Hartl[‡] and Allen P. Minton*[,§]

Department of Cellular Biochemistry, Max Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany, and Laboratory for Biochemistry and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, U.S. Department of Health and Human Services, Bethesda, Maryland 20892-0830

Received August 7, 2006; Revised Manuscript Received August 29, 2006

ABSTRACT: A simple two-state model for the dependence of the rate of folding of a polypeptide confined within a spherical cavity upon the size of the cavity relative to that of the polypeptide is presented. A general prediction of the model is that decreasing the size of the cavity will increase the rate of refolding until the cavity becomes only slightly larger than the native state of the protein, at which point a further decrease in cavity size decreases the rate of refolding. The model qualitatively accounts for the behavior of several previously published simulations of folding within a cavity, as well as recently reported experimental measurements of the relative rate of refolding of each of five proteins encapsulated within wild-type and mutant GroEL−GroES complexes that have been engineered to provide internal cavities with similar surface composition and varying volume.

There is growing appreciation that the fluid phase of many cellular compartments is dispersed within a matrix comprised of membranes and/or structural fibers in close juxtaposition (see, for example, refs *1* and *2*). Soluble proteins within the fluid phase may be confined to elements of volume, generically termed "pores", that are bounded in one, two, or three dimensions by surfaces separated by distances not much greater than the characteristic dimension of the confined macromolecule. The effect of confinement upon the chemical potential and equilibrium properties of confined molecules has been explored using simplified statistical−thermodynamic models assuming regular shapes for both the confined molecule and pore (*3*). More recently, statistical−thermodynamic theories (*4*, *5*) have been proposed to account for the observed enhancement of the stability of proteins encapsulated in sol−gel glasses (*6*) and reverse micelles (*7*). The effect of confinement in spherical and cylindrical cavities upon the rate of folding of model polypeptides and proteins has recently been studied via Brownian dynamics simulation (*8*−*11*). It was reported that as the radius of the confining cavity decreased monotonically, the rate of folding increased, reached a maximum value, and decreased. The model presented here permits a simple rationalization of such behavior.

Very recently, Tang et al. (*12*) have reported results of measurements of relative folding rates of several proteins encapsulated within a cavity formed by the chaperonin complex of GroEL and GroES proteins. Data were obtained for folding within the wild-type chaperonin complex and within mutants engineered to provide larger and smaller cavities. The refolding rate of each protein exhibits a maximum at a cavity volume that varies systematically with the molecular weight of the protein. The model developed here is shown to account semiquantitatively for the observed dependence of the folding rate on cavity size with a minimum of adjustable parameters.

Description of the Model for the Effect of Confinement upon the Rate of Refolding

The model is based upon two mechanistic assumptions.

(1) The ensemble of non-native states, collectively termed U, is presumed to exist in thermodynamic equilibrium with a single transition state, T, the conversion of which to the native state is defined to be the rate-limiting step for refolding. This model can account for the simple exponential refolding kinetics observed for a variety of small proteins (Table 18.1 of ref *13*).

(2) It is assumed that confinement of an unfolded protein affects the equilibrium between fully unfolded conformations and the transition state, but not the rate of conversion of the transition state to the native state. Although there is no direct evidence of this assumption, it is consistent with classical transition-state kinetic theory, according to which the forward rate depends only upon the relative free energies of the reactant(s) and transition state and is independent of the free energy of the product(s).

Given these two assumptions, we may estimate the effect of cavity size upon refolding rate constant as follows. According to simple transition-state theory, the rate constant for refolding in a constant-volume system is given by

$$k_{UN} = \alpha \exp(-\Delta A^{\ddagger}/RT) \qquad (1)$$

where α is a preexponential factor and $\Delta A^{\ddagger}$ is the activation

* To whom correspondence should be addressed: Building 8, Room 226, National Institutes of Health, Bethesda, MD 20892-0830. E-mail: minton@helix.nih.gov. Telephone: (301) 496-3604. Fax: (301) 402-0240.

[‡] Max Planck Institute for Biochemistry.

[§] National Institutes of Health.

10.1021/bi061597j CCC: $33.50
Published on Web 10/13/2006

(Helmholtz) free energy for refolding, which is assumed to depend upon cavity size, shape, and composition. The ratio of rate constants for refolding of a protein encapsulated in cavities 2 and 1 is then

$$\frac{k_{UN}(2)}{k_{UN}(1)} = \exp\left(-\frac{\Delta A_2^{\ddagger} - \Delta A_1^{\ddagger}}{RT}\right) \quad (2)$$

where quantities subscripted with 1 or 2 denote a property dependent upon encapsulation in the associated cavity. A thermodynamic cycle may be constructed according the scheme depicted in Figure 1. It follows from the thermodynamic cycle indicated in this figure that

$$\frac{k_{UN}(2)}{k_{UN}(1)} = \exp\left(-\frac{\Delta A_{tr,T} - \Delta A_{tr,U}}{RT}\right) \quad (3)$$

$$= \exp\left[-\frac{(\Delta E_{tr,T} - \Delta E_{tr,U}) - T(\Delta S_{tr,T} - \Delta S_{tr,U})}{RT}\right] \quad (4)$$

where $\Delta A_{tr,X}$, $\Delta E_{tr,X}$, and $\Delta S_{tr,X}$ represent the free energy, (potential) energy, and entropy changes, respectively, accompanying the hypothetical transfer of X from cavity 1 to cavity 2.

Equation 4 may be simplified to

$$\ln\left[\frac{k_{UN}(2)}{k_{UN}(1)}\right] = -\Delta\Delta E - \frac{\Delta S_{tr,U}}{R} + \frac{\Delta S_{tr,T}}{R} \quad (5)$$

where $\Delta\Delta E \equiv (\Delta E_{tr,T} - \Delta E_{tr,U})/RT$. The entropy of transfer of X between cavity 1 and cavity 2 may be expressed as the difference between the entropies of confinement of X in each of the two cavities:

$$\Delta S_{tr,X} = \Delta S_{c,X}(2) - \Delta S_{c,X}(1) \quad (6)$$

The model so far is entirely thermodynamic. At this point, we introduce the following structural assumptions to facilitate a statistical−thermodynamic estimate of the values of entropic changes associated with changes in cavity volume.

(1) U is represented as an ensemble of self-avoiding random walks, with a mean end-to-end distance in bulk solution given by (*14*)

$$h_o(\text{Å}) = 5.7n^{0.58} \quad (7a)$$

where n denotes the number of amino acid residues, given by

$$n \approx \frac{MW}{111} \quad (7b)$$

where MW is the molecular weight of the refolding polypeptide and 111 is taken as the mean MW per residue (*15*). Equations 7a and 7b provide a reasonably good empirical description of the molecular weight dependence of the hydrodynamic properties of a series of unfolded proteins (*14*).

(2) The transition state T is assumed to be largely structured and nearly as compact as the native state. We therefore model it as an effective hard sphere, with volume v_T.

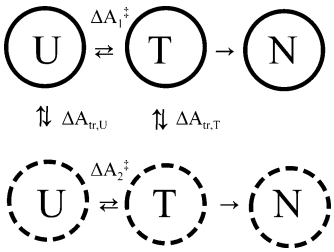

FIGURE 1: Schematic of a model for the effect of the change in cavity size upon refolding kinetics. The solid circles represent confinement within cavity 1, and the dashed circles represent confinement within cavity 2.

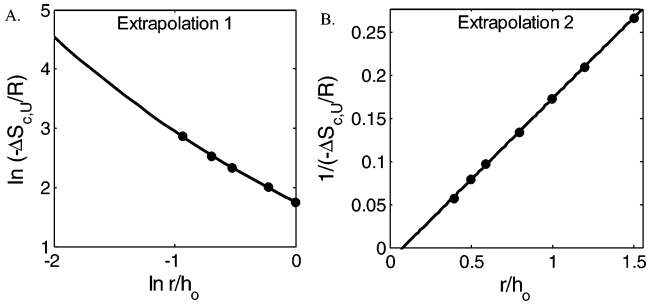

FIGURE 2: Two extrapolations of results of JD (*16*) to small values of r/h_o: (left) transformed "data" and eq 8 and (right) transformed data and eq 9.

(3) Confining cavities are assumed to be rigid and spherical in shape.

It is recognized that these assumptions are oversimplifications. Nevertheless, if the overall transition-state kinetic model summarized by eq 5 is realistic, estimates of the entropy of confinement and the resulting estimate of the effect of confinement upon folding rate should exhibit the correct qualitative and possibly even semiquantitative dependence upon cavity volume. We will discuss consequences of oversimplification at the end of this report.

Model for the Dependence of $\Delta S_{c,U}$ on Cavity Size

Jaeckel and Dayantis (*16*), hereafter called JD, performed Monte Carlo calculations of the entropy of confinement of a self-avoiding random walk within a spherical cavity of radius r. Their results were presented as a table of calculated $-\Delta S_{c,U}/R$ values as a function of r/h_o, where R is the molar gas constant, r the radius of the spherical cavity, and h_o the mean-square end-to-end distance of the unconfined self-avoiding walk, for values of r/h_o varying between 0.4 and 4.[1] We apply the model presented here to the case of proteins encapsulated in cavities that are not much larger than the native conformation of the protein, with values of r/h_o as small as 0.15. In the absence of an analytical theory for the dependence of configurational entropy upon cavity size, we propose two independent empirical methods for extrapolating the results of JD to r/h_o values of <0.4 and compare the results obtained using each.

Extrapolation 1. The tabulated results of JD are plotted in the form of $\ln(-\Delta S_{conf}/R)$ as a function of $\ln(r/h_o)$ in Figure 2A. Data points corresponding to the smallest five values of

[1] JD arbitrarily set the calculated value of $\Delta S_{c,U}/R$ to 0 at $r/h_o = 4$, so the tabulated values should be considered relative rather than absolute.

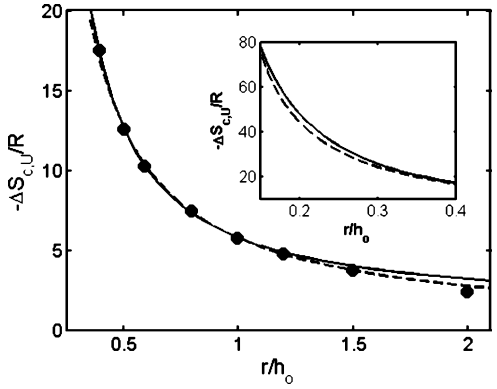

FIGURE 3: Entropy of confinement of U as a function of scaled cavity size. Symbols represent values from the calculations of JD (*16*). The solid curve was calculated using eq 8. The dashed curve was calculated using eq 9. The inset shows extrapolation of both functions to small values of r/h_o.

$\ln(r/h_o)$ were least-squares fit with a quadratic equation to yield

$$\ln(-\Delta S_{c,U}/R) = 1.755 - 0.987 \ln(r/h_o) + 0.203[\ln(r/h_o)]^2 \quad (8)$$

and the resulting functional dependence is plotted together with the transformed data in Figure 2A.

Extrapolation 2. The tabulated results of JD are plotted in the form of $(-\Delta S_{conf}/R)^{-1}$ values as a function of r/h_o in Figure 2B. Data points corresponding to the smallest seven values of r/h_o were fit with a linear equation to yield

$$(-\Delta S_{c,U}/R)^{-1} = -0.01506 + 0.18845(r/h_o) \quad (9)$$

and the resulting functional dependence is plotted together with the transformed data in the Figure 2B.

In Figure 3, the values of $-\Delta S_{c,U}/R$ tabulated by JD for $0.4 \leq r/h_o \leq 2$ are plotted as a function of r/h_o, together with the dependence of $-\Delta S_{c,U}/R$ upon r/h_o calculated according to eqs 9 and 10. It may be observed that although the methods of extrapolation are dissimilar, estimates of the value of $-\Delta S_{c,U}/R$ at small r/h_o values obtained from the two equations are quite similar, suggesting that these estimates are reasonably robust.

Model for the Dependence of $\Delta S_{c,T}$ on GroEL Cavity Size

The entropy of confinement of a rigid sphere of radius r_T in a spherical cavity of radius r is given by (*3*)

$$\frac{\Delta S_{c,T}(r)}{R} = 3 \ln\left(1 - \frac{r_T}{r}\right) \quad (10)$$

Qualitative Estimate of the Effect of Confinement on Refolding Rate

While electrostatic and/or hydrophobic interactions between the residues of a confined protein and the walls of the confining cavity are certainly possible, and even likely (*11*), the primary objective of this work was to ascertain how the rate of refolding is affected by confinement per se, that is to say, the entropic consequences of confinement. For this purpose, we set the energetic quantities $\Delta E_{tr,U}$, $\Delta E_{tr,T}$, and hence $\Delta\Delta E$ in eqs 5 and 6 equal to zero. Then the free energy

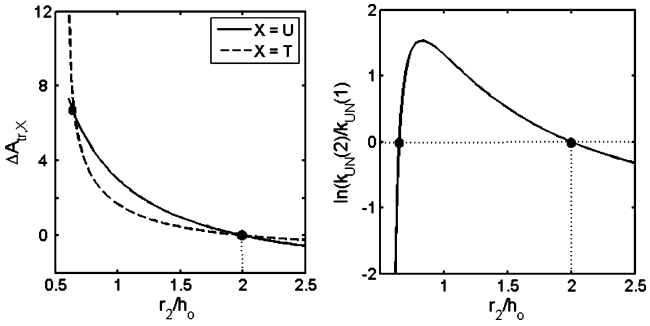

FIGURE 4: Results of a sample model calculation: (left) calculation of the free energy of transfer of U (solid line) and T (dashed line) from a cavity with an r_1 of $2h_o$ to a cavity of with arbitrary value of r_2 ($r_T = 0.6h_o$) and (right) logarithm of the refolding rate in a cavity with an arbitrary r_2 relative to that in a cavity with an r_1 of $2h_o$.

of transfer of species X from a cavity of radius r_1 to a cavity of radius r_2 is simply given by

$$\Delta A_{tr,X} = -T[\Delta S_{c,X}(r_2) - \Delta S_{c,X}(r_1)] \quad (11)$$

For a given ratio of r_T/h_o, one may use eqs 6, 8 or 9, and 10 to calculate the values of $\Delta S_{c,U}$ and $\Delta S_{c,T}$ associated with a change in cavity radius from r_1/h_o to r_2/h_o. Then eq 11 is used to calculate $\Delta A_{tr,U}$ and $\Delta A_{tr,T}$, and eq 3 is used to calculate $k_{UN}(2)/k_{UN}(1)$. The results of a sample calculation are plotted in Figure 4. While quantitative results vary widely for different combinations of the input parameters (r_1/h_o and r_T/h_o), the following qualitative features seem to be inherent in the model.

(1) At large cavity sizes, $A_{c,U}$ increases more rapidly with a decreasing r than does $A_{c,T}$. This is because in large cavities, U is on average significantly less compact than T and is more sensitive to boundary exclusion effects. Confinement-induced stabilization of N (the native state) relative to U is also attributed to differences between the compactness of the two states (*4*, *5*).

(2) At sufficiently small cavity sizes ($r \rightarrow r_T$), $A_{c,T}$ increases more rapidly with a decreasing r than does $A_{c,U}$. This is because T is assumed to be rigid (structured), and its free energy of confinement is accordingly calculated to diverge as v approaches v_T (*3*); on the other hand, U retains some internal conformational variability even when tightly confined.

It follows from these two observations that in the absence of a significant size-dependent energy of interaction between the confined protein and the cavity boundary surface, the folding rate constant must increase with a decreasing cavity size in cavities that are sufficiently large, reach a maximum rate at some intermediate cavity size, and then decrease as the cavity size decreases further. Qualitatively similar behavior was observed in Brownian dynamics simulations of the folding of several model polypeptides and proteins within spherical and cylindrical cavities (*8*−*11*).

Dependence of the Rate of Refolding within the Chaperonin Cavity upon Cavity Size

Very recently, Tang et al. (*12*) reported the measurement of the relative rates of refolding of several globular proteins within the central cavity of several GroEL−GroES complexes, in which subunits of GroEL have been mutated by

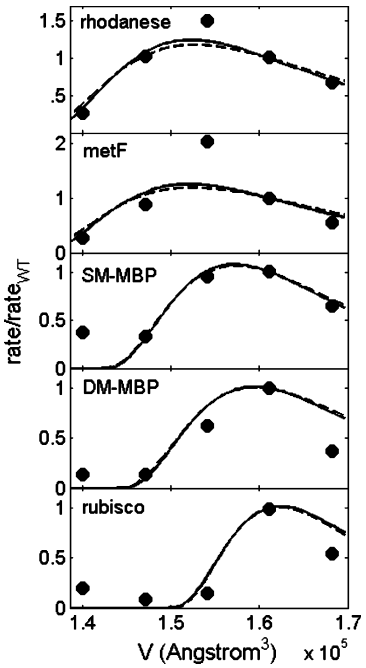

FIGURE 5: Data of Tang et al. (*12*) (●) and best-fit dependence of the folding rate relative to that in the wild-type chaperonin complex ($v = 1.61 \times 10^6$ Å^3), calculated using eq 3 and either eq 8 (solid line) or eq 9 (dashed line), and the parameters given in Table 1.

the addition or deletion of a terminal repeat sequence to vary cavity size in a controlled manner. They reported that each protein exhibited a maximum refolding rate at an intermediate cavity size, with slower refolding rates in smaller and larger cavity sizes, and that the cavity size corresponding to the maximum folding rate increased with an increasing molecular weight of the confined protein. We have attempted to model the data of Tang et al. (*12*) using the semiempirical theory described here. The chaperonin cavity is treated as a sphere with a radius corresponding to volumes associated with the wild type and deletion or additions of the GGM repeat sequence.[2] For a specified protein molecular weight, the independent variable is cavity volume, and the dependent variable is the ratio of folding rate in a cavity of arbitrary volume to that in the cavity of the wild-type chaperonin complex. Model calculations were performed using both extrapolation formulas (eqs 8 and 9) to estimate $\Delta S_{c,U}$ as a function of cavity size, and least-squares fitting of the model to the data was accomplished by variation of a single parameter, v_T ($=4\pi r_T^3/3$), particular to an individual confined protein species. The best fit dependence of the relative folding rate upon cavity volume for five proteins is plotted together with the data in Figure 5, and best-fit values of v_T are given in Table 1.

Discussion of Results

The model presented here is based upon the fundamental assumptions that the rate of refolding within the cavity is limited by the rate of conversion of a unique transition state to fully folded product and that the transition state may be treated as if it exists in quasi-equilibrium with the ensemble of non-native states. The accelerating effect of confinement in large cavities is simple to rationalize in this context: in this limit, a decrease in the radius of the cavity destabilizes the larger ensemble of unfolded molecules relative to the more compact transition state, hence lowering the free energy of activation for folding. The decelerating effect of confinement in very small cavities is rather more subtle. The model presented here suggests that in the limit of small cavities, a decrease in the radius of the cavity destabilizes the rigid transition state relative to the compressed but still malleable ensemble of unfolded molecules, hence increasing the free energy of activation for folding.[3]

Table 1: Best-Fit Values of v_T Obtained by Modeling the Data Presented in Figure 5 As Described in the Text

protein	MW	best-fit v_T ($\times 10^6$ Å^3)[a]
rhodanese	33000	1.34
metF	33000	1.34
singly mutated maltose-binding protein (SM-MBP)	41000	1.42
doubly mutated maltose-binding protein (DM-MBP)	41000	1.44
RuBisCo	50000	1.50

[a] Best-fit values obtained using either eq 8 or 9 to calculate $\Delta S_{c,U}$ were identical within $\pm 0.01 \times 10^6$ Å^3.

In view of the many simplifying approximations underlying the quantitative model presented here, and the very small number of variable parameters (only one for each confined protein), the degree of agreement between the calculated best fit and the measured rates is surprisingly good. The model successfully captures the observation that the rate of refolding goes through a maximum with an increasing cavity volume and that as the molecular weight of the refolding protein increases, the optimal cavity size increases and the width of the optimal peak narrows. By restricting the use of the model to interpretation of experimentally observed differences between rates of refolding in cavities of similar composition but modestly different sizes, we find the probability that the assumption of entropy dominance (i.e., the assumption that $\Delta E_{tr,U}$ and $\Delta E_{tr,T}$ are negligible) is seriously in error is minimized.[4]

The most noticeable discrepancy between model and experiment is the model prediction that the rate of refolding should tend toward zero with a decreasing cavity volume, whereas the measured rates do not go to zero but approach a small value (10−40% of the wild-type rate). This systematic discrepancy is likely due to the simplifying assumption that both the cavity and T are rigid, which is surely not the

[2] We did not model the dependence upon volume changes associated with addition of GGA repeats, as these rates appeared to increase with an increasing numbers of GGA repeats, indicating significant interaction of the repeat sequence with the confined polypeptide over and above steric repulsion.

[3] An anonymous reviewer has suggested as an alternative that the deceleration may be attributed to a compression-induced loss of mobility of segments of the unfolded protein, which becomes so severe in the limit of high compression that the transition state may no longer be treated as if it were in quasi-equilibrium with the unfolded states. While this hypothesis is reasonable, it does not appear to be as readily quantifiable as the current model, and moreover, it is not obvious that it can account, as the current model does, for the observation that refolding kinetics remain two-state (within experimental uncertainty) even in the smallest cavities (*12*).

[4] Additional modeling of the data without constraining $\Delta\Delta E$ to be equal to zero did not significantly improve the goodness of fit, and the absolute value of the best-fit $\Delta\Delta E$ so obtained was much less *RT* for all five confined proteins (data not shown).

case. As the size of T approaches that of the cavity and the free energetic penalty of confinement in a rigid cavity becomes extremely great, one would expect either T, the cavity, or both to deform to better accommodate T, minimize the overall free energy, and permit refolding, albeit at a rate slower than that in the wild-type cavity.

ACKNOWLEDGMENT

We thank Prof. F.-U. Hartl (Max Planck Institute for Biochemistry) and Dr. P. McPhie (National Institutes of Health) for helpful discussions and Dr. McPhie for careful review of preliminary drafts of the manuscript.

REFERENCES

1. Hirokawa, N. (1991) Molecular architecture and dynamics of the neuronal cytoskeleton, in *The Neuronal Cytoskeleton* (Burgoyne, R., Ed.) pp 5−74, Wiley-Liss, New York.
2. Nickell, S., Kofler, C., Leis, A., and Baumeister, W. (2006) A visual approach to proteomics, *Nat. Rev. Mol. Cell Biol. 7*, 225−230.
3. Minton, A. P. (1992) Confinement as a determinant of macromolecular structure and reactivity, *Biophys. J. 63*, 1090−1100.
4. Zhou, H. X., and Dill, K. A. (2001) Stabilization of proteins in confined spaces, *Biochemistry 40*, 11289−11293.
5. Zhou, H. X. (2004) Protein folding and binding in confined spaces and in crowded solutions, *J. Mol. Recognit. 17*, 368−375.
6. Eggers, D., and Valentine, J. (2001) Molecular confinement influences protein structure and enhances thermal protein stability, *Protein Sci. 10*, 250−261.
7. Shastry, M., and Eftink, M. (1996) Reversible thermal unfolding of ribonuclease T1 in reverse micelles, *Biochemistry 35*, 4094−4101.
8. Klimov, D. K., Newfield, D., and Thirumalai, D. (2002) Simulations of β-hairpin folding confined to spherical pores using distributed computing, *Proc. Natl. Acad. Sci. U.S.A. 99*, 8019−8024.
9. Takagi, F., Koga, N., and Takada, S. (2003) How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations, *Proc. Natl. Acad. Sci. U.S.A. 100*, 11367−11372.
10. Cheung, M. S., Klimov, D., and Thirumalai, D. (2005) Molecular crowding enhances native state stability and refolding rates of globular proteins, *Proc. Natl. Acad. Sci. U.S.A. 102*, 4753−4758.
11. Xu, W.-X., Wang, J., and Wang, W. (2005) Folding behavior of chaperonin-mediated substrate protein, *Proteins 61*, 777−794.
12. Tang, Y., Chang, H., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M., Hartl, F., and Hayer-Hartl, M. (2006) Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein, *Cell 125*, 903−914.
13. Fersht, A. (2002) *Structure and Mechanism in Protein Science*, Freeman, New York.
14. Goldenberg, D. P. (2003) Computational simulation of the statistical properties of unfolded proteins, *J. Mol. Biol. 326*, 1615−1633.
15. Berezovsky, I., Kirzhner, A., Kirzhner, V., Rosenfeld, V., and Trifonov, E. (2003) Protein sequences yield a proteomic code, *J. Biomol. Struct. Dyn. 21*, 317−325.
16. Jaeckel, A., and Dayantis, J. (1994) Statistics of confined self-avoiding walks: II. Entropy and pressure of confinement, *J. Phys. A: Math. Gen. 27*, 7719−7731.

BI061597J