How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations

Fumiko Takagi*, Nobuyasu Koga†, and Shoji Takada*†‡

*PRESTO, Japan Science and Technology Corporation, and †Department of Chemistry, Faculty of Science, Kobe University, Rokkodai, Nada, Kobe 657-8501, Japan

Edited by Peter G. Wolynes, University of California at San Diego, La Jolla, CA, and approved July 9, 2003 (received for review April 3, 2003)

How the *Escherichia coli* GroEL/ES chaperonin assists folding of a substrate protein remains to be uncovered. Recently, it was suggested that confinement into the chaperonin cage itself can significantly accelerate folding of a substrate. Performing comprehensive molecular simulations of eight proteins confined into various sizes L of chaperonin-like cage, we explore how and to what extent protein thermodynamics and folding mechanisms are altered by the cage. We show that a substrate protein is remarkably stabilized by confinement; the estimated increase in denaturation temperature ΔT_f is as large as ≈60°C. For a protein of size R_0, the stabilization ΔT_f scales as $(R_0/L)^\nu$, where $\nu \approx 3$, which is consistent with a mean field theory of polymer. We also found significant free energy cost of confining a protein, which increases with R_0/L, indicating that the confinement requires external work provided by the chaperonin system. In kinetic study, we show the folding is accelerated in a modestly well confined case, which is consistent with a recent experimental result on ribulose-1,5-bisphosphate carboxylase-oxygenase folding and simulation results of a β hairpin. Interestingly, the acceleration of folding is likely to be larger for a protein with more complex topology, as quantified by the contact order. We also show how ensemble of folding pathways are altered by the chaperonin-like cage calculating a variant of ϕ value used in the study of spontaneous folding.

The *Escherichia coli* GroEL/ES chaperonin is the best-characterized molecular chaperone that assists *in vivo* protein folding (1, 2). The cylindrical structure of GroEL complex and its conformational change upon binding to ATP and GroES have been experimentally determined (1, 3, 4). The ATP-dependent chaperonin cycle has been studied, and how these structural changes are coupled with substrate binding and release has been elucidated (5–7). Many protein-engineered GroEL molecules were used to identify residues and/or segments that are important for substrate binding, ATP hydrolysis, and so on (8, 9). With all of these, machinery of the chaperonin was reasonably well uncovered.

On the other hand, how substrate folding is assisted by the chaperonin is less understood. There are at least two different, but not mutually exclusive, scenarios regarding this issue (10, 11). The first "Anfinsen cage" model indicates that the chaperonin provides a passive cage that separates a substrate protein from other macromolecules, removing the danger of aggregation (10). In the other "iterative annealing" scenario, a substrate protein is mechanically forced to unfold upon binding to GroEL and it folds upon transfer into the chaperonin cavity or release from GroEL. This cycle is repeated until a substrate reaches the native state (12–14, 38). Both Anfinsen cage and mechanical unfolding effects may be present in reality. Here, we address yet another factor that can assist substrate folding. Using an engineered chaperonin system that inhibits the chaperonin cycle, Brinker *et al.* (15) showed that confinement of unfolded protein alone accelerates folding inside the cage. Thus, simply by guiding into the cage, the chaperonin may have an "active" role of refolding a substrate. Confinement effects on protein stability and folding kinetics were recently studied theoretically by Zhou and Dill (16) and Klimov *et al.* (17).

In this article, we extensively study properties of a protein molecule confined in a simple model of a chaperonin-like cage. There is enough evidence that a fraction of substrates can fold within the chaperonin chamber. Once the GroES is bound, the inner wall of the chaperonin chamber is known to be largely hydrophilic, and thus it is unlikely that substrate proteins strongly interact with inner wall atoms via hydrophobic interactions. Moreover, the chaperonin can assist folding of various proteins in a nonspecific manner, and thus any specific interaction between the substrate and the inner wall is not crucial. Thus, a primary effect of caging may be to restrict conformational motion of a protein into a small volume. Physically, a polymer confined into a small volume exhibits characteristic behavior caused by change in its conformational entropy (16–18). We therefore take a minimal model for the chaperonin cage, that is, a cylindrical box with no attractive force between a substrate and the box. We note that this caging effect must exist on top of any other factors that may assist protein folding. Including all possible factors into a model may lead to a slightly more realistic model, but also complicates understanding of the roles of each factor. In this article we focus on one: the confinement effect.

The substrate protein model we use here is one of the standard models (19). It is well buttressed by the recently developed theory of spontaneous protein folding, which greatly deepened our understanding of proteins over the last few decades. Natural proteins have evolved their sequences so that interactions at the native structure are almost perfectly optimized (20, 21). This leads to the view that the protein energy landscape has a funnel-like global shape toward the native structure (22, 23). Technically, this funnel-like landscape can easily be realized by the so-called Go-like models (19, 24, 25). Many simple Go-like models were proposed and used for folding study (19, 25). Surprisingly enough, these simple models explained quite well experimental observation of folding pathways and rate constants qualitatively and sometimes quantitatively. We use here one of these models that was well characterized (19, 26, 27) as a model protein, to which we add the chaperonin-like cage.

In this article, we compare in detail protein folding with and without the chaperonin-like cage via molecular simulations. We first describe the simulated system. The chaperonin-like cage is modeled as a cylindrical box that restricts a substrate protein molecule. Only nonspecific repulsive force is considered between the chamber wall and protein amino acids. We then show that a protein in the cage is thermodynamically more stable than that without the cage, as anticipated earlier (16, 17). This finding

This paper was submitted directly (Track II) to the PNAS office.

See commentary on page 11195.

‡To whom correspondence should be addressed. E-mail: stakada@kobe-u.ac.jp.

is consistent with the experimental result of encapsulated proteins in silica matrix (28). This stabilization is surprisingly large when the substrate protein is of comparable size to the cage scale. An estimated temperature shift is as large as ≈60°C. We found that, for all proteins studied, the shift in the folding transition temperature ΔT_f caused by confinement into size L cage scales in a universal manner, $\Delta T_f \sim (R_0/L)^\nu$, where $\nu \approx 3$ and R_0 is the characteristic size of a substrate. We then study folding kinetics and pathways, where we found that the chaperonin cage does not affect the folding of small protein, accelerates that of modestly large protein, and decelerates maximally large protein. Interestingly, the acceleration effect is likely to be more prominent for proteins that have more complex topology. We also compute site-resolved ensemble of folding pathways, showing that the pathways are significantly affected by confinement.

Model and Methods

The model of a substrate protein we use is that of Clementi *et al.* (19), where the protein chain is represented only with C_α atoms of every amino acid residue. C_α atoms represented as spherical beads are connected via virtual bonds. Go-like energetic bias toward the native structure is added to both local and nonlocal interactions. Local interactions provide energetic bias in the local geometry to that of the native structure. For the nonlocal interactions, attractive interactions are introduced only for the pairs of amino acids that are in contact in the native structure, which makes global folding energy surface funnel-like. The rest of the pairs have only repulsive interactions. The chaperonin-like cage is a hard wall that provides repulsive force to protein amino acids when they come close to the wall.

The effective energy, V, at a protein conformation Γ is comprised of two contributions, $V = V_{\text{protein}} + V_{\text{cage}}$; V_{protein} is for the intraprotein interactions and is the same as the energy function of Clementi *et al.* (19). V_{cage} is for the interaction between protein amino acids and the chaperonin cage and is purely repulsive. Explicitly, V is given as

$$\begin{aligned} V(\Gamma,\Gamma_0) = {} & \sum_{\text{bonds}} K_r(b_i - b_{0i})^2 + \sum_{\text{angles}} K(\theta_i - \theta_{0i})^2 \\ & + \sum_{\text{dihedrals}} \{K^{(1)}[1-\cos(\phi_i - \phi_{0i}^{(1)})] + K^{(3)}[1-\cos(\phi_i - \phi_{0i}^{(3)})]\} \\ & + \sum_{i<j-3}^{\text{native contact}} 1\left[5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10}\right] + \sum_{i<j-3}^{\text{non-native contact}} 2\left(\frac{C}{r_{ij}}\right)^{12} \\ & + \sum_i K_{\text{cage}}\left[\left(\frac{C}{2d_i}\right)^4 - 2\left(\frac{C}{2d_i}\right)^2 + 1\right]H(C/2 - d_i), \end{aligned} \quad [1]$$

where all but the last term correspond to V_{protein} and the last term is the V_{cage}. In the equation, b_i, is the virtual bond length between two adjacent amino acid C_αs, namely, $b_i = |\mathbf{b}_i| = |\mathbf{r}_{i+1} - \mathbf{r}_i|$, where $\mathbf{r}_i$ is the position of the ith amino acid. θ_i stands for the angle between two adjacent virtual bonds $\mathbf{b}_i$ and $\mathbf{b}_{i+1}$. ϕ_i represents the ith dihedral angle around the ith bond $\mathbf{b}_i$. The first three terms provide interactions local along the chain, whereas the fourth and fifth terms are interactions between pairs that are distant along the chain. In the latter, $\Sigma^{\text{native contact}}$ means that the summation is taken over only pairs that are close in space at the native structure (the precise definition is given below). These pairs are called the native contacts. r_{ij} is the distance $|\mathbf{r}_j - \mathbf{r}_i|$ between the ith and jth amino acids. Parameters with the subscript 0 are the constants, of which values are taken from the corresponding variables in the native structure. Thus, we note that each term in the first four terms gives the lowest energy in the native structure. In the caging potential V_{cage}, d_i is the distance between the chaperonin wall and the ith amino acid. $H(x)$ is the Heaviside function, namely, $H(x) = 1$ for $x > 0$ and $H(x) = 0$ for $x < 0$. Thus, a positive energy is enforced when amino acids are within $C/2$ distance from the wall. Throughout the article, we use $K_r = 100.0$, $K_\theta = 20.0$, $K_\phi^{(1)} = 1.0$, $K_\phi^{(3)} = 0.5$, $\varepsilon_1 = 0.18$, and $\varepsilon_3 = 0.18$ for intraprotein interactions, which are the same values as those used in Koga and Takada (27). For the caging potential, $K_{\text{cage}} = 100.0$, and $C = 4.0$ Å are used for all proteins studied. We also note that the same unit is used both for energy and temperature and thus the Boltzmann constant $k_B = 1$.

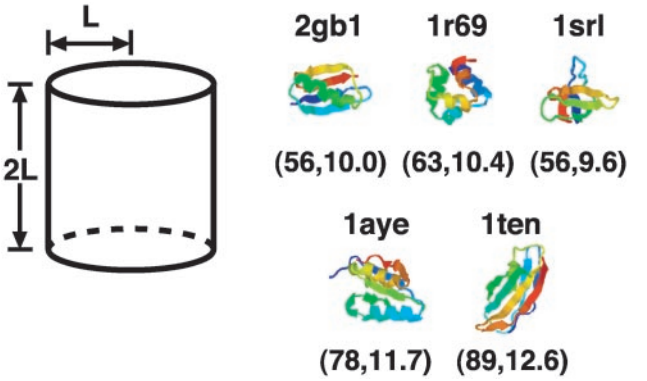

Fig. 1. Cartoon of chaperonin-like cage and proteins studied. (*Left*) The chaperonin-like cage is modeled as a cylindrical box with a characteristic length L, in which a folding protein molecule is confined. (*Right*) The native structures of five proteins studied (of eight); they are, from the top left to the bottom right, protein G, 434 repressor, src SH3 domain, Ada2h, and fibronectin type 3. The Protein Data Bank codes are given above the structures, and the number of residues and radii of gyration in the native structures are shown in parentheses.

Protein dynamics is simulated by the Langevin equation at a constant temperature T,

$$m_i\dot{\mathbf{v}}_i = \mathbf{F}_i - \gamma_i\mathbf{v}_i + \boldsymbol{\xi}_i, \quad [2]$$

where $\mathbf{v}_i$ is the velocity of the ith bead, a dot represents the derivative with respect to time t and thus $\mathbf{v}_i = \dot{\mathbf{r}}_i$. $\mathbf{F}_i$ and $\boldsymbol{\xi}_i$ are systematic and random forces on the ith bead, respectively. The systematic force $\mathbf{F}_i$ is derived from the effective energy V as usual, $\mathbf{F}_i = -\partial V/\partial \mathbf{r}_i$. The white and Gaussian random forces ξ_i satisfy $\langle \xi_i \rangle$ and its variance is $\langle \xi_i(t)\xi_j(t')\rangle = 2\gamma_i k_B T\delta_{ij}\delta(t - t')\mathbf{1}$, where the bracket denotes the ensemble average and $\mathbf{1}$ is a 3×3-unit matrix. For numerical integration of the Langevin equation, we use an algorithm by Honeycutt and Thirumalai (29). We use $\gamma_i = 0.25$, $m_i = 10.0$, and the finite time step $\Delta t = 0.2$.

We define that ith and jth amino acids are in the "native contact" set if one of the nonhydrogen atoms in the jth amino acid is within 6.5-Å distance from one of nonhydrogen atoms in the ith amino acid at the native structure Γ_0. For a given protein conformation, Γ, we define that the native contact between i and j is formed if the distance r_{ij} is $<1.2\, r_{ij0}$. We then use a standard measure of the nativeness, $Q(\Gamma)$, for a given protein conformation Γ, defined as the ratio of numbers of formed native contacts at Γ to those at the native structure Γ_0.

In the article, we use eight relatively small proteins (<100 residues) that cover a wide variety of protein topology, including α proteins, α/β proteins, and β proteins. For five of eight proteins (depicted in Fig. 1 where the four-character codes are those of the Protein Data Bank), various cage size is explored; they are IgG-binding domain of protein G (ID code 2gb1, α/β protein), 434 repressor (ID code 1r69, α protein), src SH3 domain (ID code 1srl, β protein), Ada2h (ID code 1aye, α/β protein), and tenascin fibronectin type 3 (ID code 1ten, β

protein). Three other proteins are simulated with a fixed size of cage to investigate topological effects on folding: chymotripsin inhibitor 2 (ID code 1coa, α/β protein), Sso7d (ID code 1bnz, β protein), and ACBD (ID code 2abd, α protein).

Confinement Effects on Thermodynamics

For each protein with or without the chaperonin cage, the first step is to use the bisection method to find out the approximate folding transition temperature, T_f. We start with folding simulations at a very low temperature (T_{low}) and a high temperature (T_{high}). Each trajectory contains $>10^7$ molecular dynamics steps, which are much longer than the folding/unfolding time scales. Typically we observe at T_{low} that the protein quickly reaches the native state with high Q values ($Q > 0.5$) and stays in the native state for most of the simulated time, indicating that the temperature is below T_f. While at T_{high}, the protein stays in the denatured state with a low Q score ($Q < 0.5$) showing that the temperature is above T_f. Next, we take the midtemperature $T_{new} = (T_{low} + T_{high})/2$ and monitor Q values in the repeated simulation. If the protein is in the native (denatured) state more than half the time, we then replace T_{low} (T_{high}) with T_{new}. The same procedure is iterated nine times to narrow down the range of T_f. The final T_{new} is expected to be close to T_f. All of these sampled data are used in the weighted histogram method (WHAM) (30) to obtain thermodynamics quantities, such as the heat capacity $C_v = \lfloor \langle E^2 \rangle - \langle E^2 \rangle \rfloor k_B T^2$, where E is the total energy and the free energy curve $F(Q) = -k_B T \log P(Q)$. The peak position in $C_v(T)$ along T is identified as the folding transition temperature T_f.

A representative time course of $Q(t)$ for the protein G is depicted in Fig. 2 *Top*, where the left is the case without the chaperonin cage (the bulk system), and the right is with the cage of size $L = 20$ Å. These are the result at T_f^0, the transition temperature of the bulk system. The bulk protein exhibits sharp two-state transitions between the native (high Q values) and the denatured states (low Q values), whereas the confined protein stays most of the time in the native state at T_f^0. The scattered plots in the (Q,R_g) plane at the same condition are given in Fig. 2 *Middle*, where R_g is the radius of gyration of a protein. We see clearly that, without the cage (the bulk system), the denatured state ensemble has the R_g distribution mostly >15 Å, whereas the caged protein has limited denatured conformations with $R_g < \approx 15$ Å. Because of this reduced conformational entropy, the denatured state of caged proteins has higher free energy than that of the bulk protein. The free energy curves at the corresponding conditions are presented in Fig. 2 *Bottom*. As expected, for the bulk system (Fig. 2 *Bottom Left*) the free energy curve has double minima with the equal free energy in the native and the denatured states, whereas the denatured state of the caged protein becomes unstable (Fig. 2 *Bottom Right*). Therefore, at the same temperature, the chaperonin-caged protein has larger native stability than the protein without the cage. Naturally, the folding transition temperature is increased by caging into the chaperonin chamber (17). At the transition temperature T_f ($T_f > T_f^0$) of the confined system ($L = 20$ Å), the protein exhibits the two-state transition, but it is less sharp than the bulk case (data not shown). Also in the confined system, the denatured state has higher Q values than that of the bulk system (data not shown), indicating that the denatured state of the confined system possesses significant residual order (17). We comprehensively performed the same type of simulations for five proteins in Fig. 1 for various cage sizes L, 14 Å $< L <$ 100 Å as well as the bulk and three more proteins with $L = 25$ Å and the bulk system.

Now, we consider more quantitatively to what extent the protein is stabilized by confinement. For that purpose, we compute the folding transition temperature T_f for five proteins in various sizes L of the cage. As mentioned above, T_f is

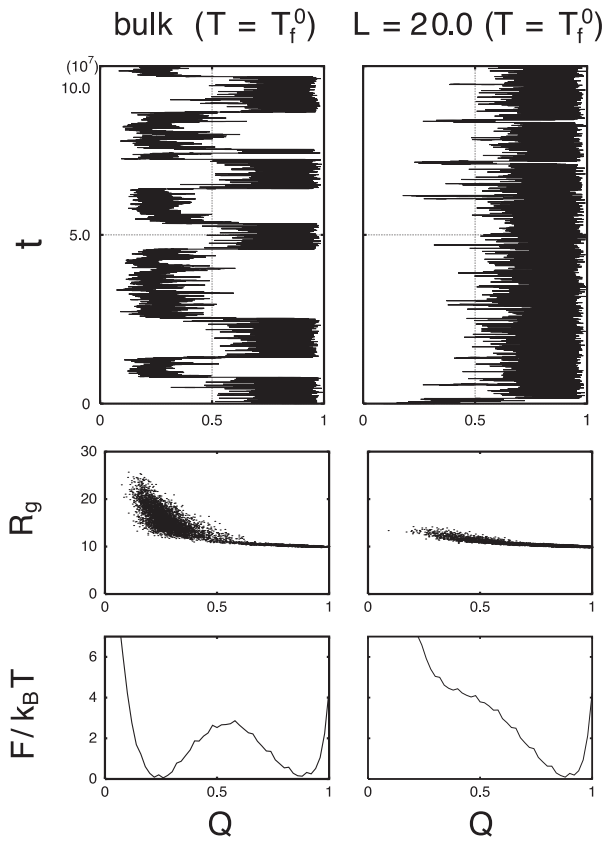

Fig. 2. Folding time course and distribution of the nativeness measure Q both for unconfined (*Left*) and confined (*Right*) protein G (Protein Data Bank code 2gb1) at the transition temperature of unconfined system T_f^0. (*Top*) Time series $Q(t)$ are plotted for representative trajectories. (*Middle*) Scattered plots in the (Q, R_g) plane are drawn from the same trajectories. (*Bottom*) The free energy profile $F(Q)$ as a function of Q is plotted. The confinement induces restriction to the denatured distribution as seen (*Middle Right*), which makes the denatured state less stable (*Bottom Right*) (at $T = T_f^0$).

identified by the peak in the heat capacity C_v curve as a function of T. Fig. 3*a* illustrates the heat capacity as a function of T in the case of protein G. We see that the peak in C_v shifts to higher temperature as the cage size decreases. The same effect was found for a β hairpin confined in a spherical pore studied by computer simulation somewhat similar to the present one (17). Here, the estimated temperature increase is "dramatically" large; for example, the protein G in the cage of size 15 Å has the transition temperature $\approx$20% higher than that without the cage, which corresponds to the increase of $\approx$60°C assuming that T_f^0 is of the order of $\approx$350 K. Similar amount of increases in T_f for proteins in a small cage is obtained from all five proteins studied (see Fig. 3*b* for details). Experimentally, α-lactalbumin confined in a silica matrix showed the denaturation temperature elevated by $\approx$30°C (28). We also note that the peak in C_v-T plot becomes broader as the cage size decreases. The denatured state in the smaller cage is made of more compact non-native structures. These compact non-native structures possess residual native-like partial order more significantly than the bulk proteins (data not shown). This results in weaker and less prominent phase transition, as seen in Fig. 3*a*.

In Fig. 3*b*, we plot the folding temperatures T_f of five proteins in the chaperonin cage of size L against $N^{3/5}/L$. Note that both axes are in logarithmic scale. We see clearly, for well-confined cases, that the data lie on a straight line, indicating the power low dependence of the temperature shift with $N^{3/5}/L$. The linear regression of the data points $N^{3/5}/L > 0.25$ leads to the scaling,

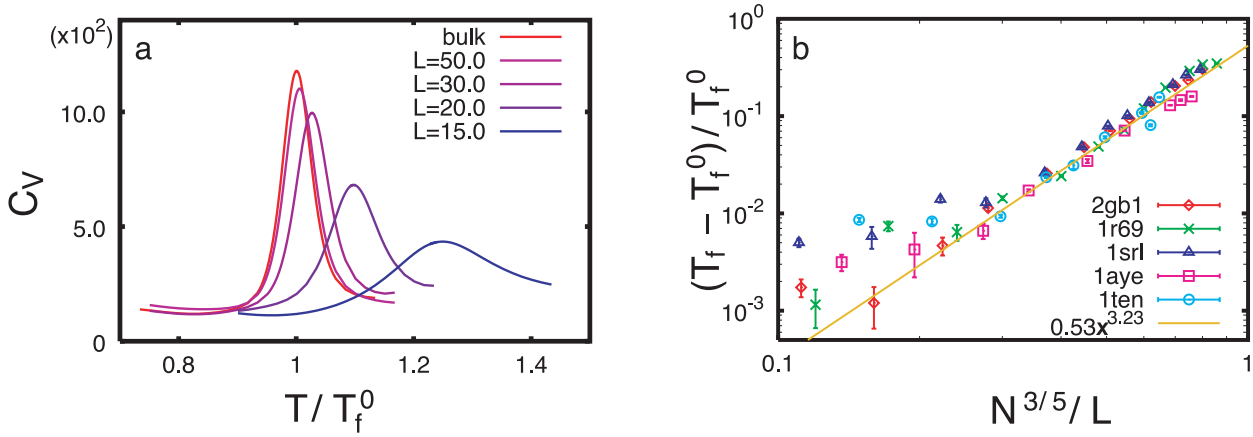

Fig. 3. Change in the thermodynamic stability of a chaperonin-caged protein. (*a*) The heat capacity C_v of protein G (Protein Data Bank ID code 2gb1) as a function of temperature T for several different cage sizes L [$L = \infty$ (bulk), 50, 30, 20, and 15 Å]. As the cage size decreases, the folding transition temperature T_f is increased and the peak becomes broader. (*b*) The relative change in the folding transition temperatures $(T_f - T_f^0)/T_f^0$ caused by confinement plotted against $N^{3/5}/L$ for five proteins studied. Both axes are in the logarithmic scale. Symbols are defined in the figure. The straight line is obtained by the linear regression.

$T_f - T_f^0 \sim (N^{3/5}/L)^{3.25\pm0.09}$. We emphasize it is remarkable that proteins with completely different topology and sizes exhibit this universal scaling. For a cage size much larger than the protein native size, i.e., the left side in Fig. 3*b*, the data deviate from the line. Here, the temperature shift becomes very small, and thus inherent numerical error would be significant. Also, for such a large box, translational motion of a protein molecule in the cage comes into play, and the translational entropy could be the source of the deviation.

This scaling rule can be interpreted with a simple polymer model. At the folding transition temperature T_f^0, the free energies of the native F_n and the denatured F_d states of a bulk protein are equal to each other by definition: $F_n = F_d$. The native state has only negligible conformation entropy, and thus we set the native entropy at zero: $F_n = E_n$. In the denatured state, we assume the residual interaction is negligible and thus only entropic contribution $-T_f^0 S_0$ remains. Combining these relations, we get $F_n = E_n = -T_f^0 S_0 = F_d$. If the denatured state were approximated as the random coil without self-avoiding interactions, i.e., the ideal chain in the polymer theory (16, 18), the confinement to the characteristic length L would induce free energy change of $\Delta F \sim T(R_0/L)^2$, where R_0 is the radius of the denatured protein (18). The resulting folding temperature can be estimated through $F_n = E_n = -T_f\lfloor S_0 - (R_0/L)^2 \rfloor = F_d$. These equations together give an estimate of the change in the transition temperature as $(T_f - T_f^0)/T_f^0 \sim (R_0/L)^2$ (16). Instead of the ideal chain, if we use a mean field approximation to the excluded volume interactions (18), we get an estimate of $\Delta F \sim T(R_0/L)^3$ and therefore $(T_f - T_f^0)/T_f^0 \sim (R_0/L)^3$. Here, R_0 should scale as the "real" chain with the volume interaction, for which the Flory theory gives an approximate scaling $R_0 \approx N^{3/5}$, where N is the number of amino acids here. The scaling estimated by the simulation $T_f - T_f^0 \sim L^{-3.25}$ is fairly close to that by the mean field theory. We note that scaling with respect to L is characterized here, but not the scaling on the chain length N. Simulated proteins here span only a very limited range of the chain length, and therefore scaling with N is not the issue here. We can only be sure that for a fixed L, $T_f - T_f^0$, increases with N.

We next compare the free energy of the confined system with that of the bulk system. Because confinement reduces conformational entropy of the protein, we expect increase in the free energy. The free energy change ΔF upon confinement into the size L cage can precisely be calculated by, for example, the free energy perturbation technique, but it can easily be estimated as $\sim -k_B T \log P_{<L}$, where $P_{<L}$ is the probability for a bulk protein to have conformations that can be fit in the size L cage. Fig. 4 shows the estimate of the free energy change for five proteins at TT_f^0 as a function of $N^{3/5}/L$. We see that the free energy increases with $N^{3/5}/L$. It grows drastically above $N^{3/5}/L \approx 0.7$, which would correspond to the maximal size of the substrate protein (≈60 kDa) that can enter into the chaperonin. We emphasize here that substrate proteins favor the bulk phase rather than inside the cage and thus external work is necessary for a substrate protein to be transferred in the cage. Therefore, the chaperonin cycle inherently requires some source of free energy. In the real chaperonin system, this is provided by ATP hydrolysis. More precisely, the binding affinity of GroEL with ATP and GroES provides work for a substrate to enter into the cage. The chaperonin cycle includes the release of the nucleotide and GroES from GroEL, for which the ATP hydrolysis is the prerequisite. If the free energy cost of confining a protein exceeds the work provided by the binding of ATP and GroES, the substrate does not preferably enter into the cage.

Confinement Effects on Folding Kinetics and Pathways

Here, we investigate folding mechanisms of a substrate protein in the chaperonin-like cage. First, we look into the folding rate constants k_f. For all proteins studied, with various sizes L of the

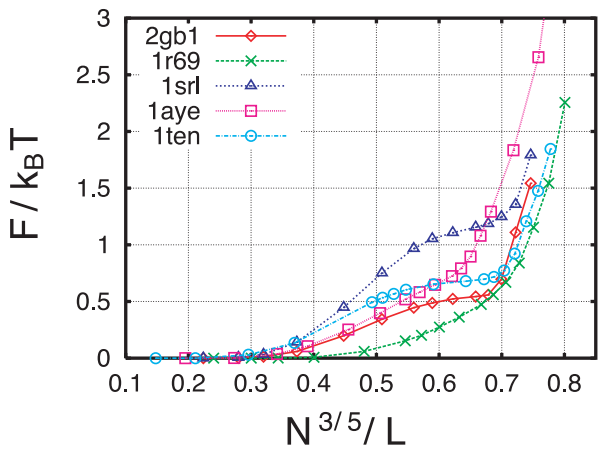

Fig. 4. Estimated free energy cost of confining a protein molecule into the chaperonin-like cage plotted against $N^{3/5}/L$. Results for five proteins are depicted, and symbols used are the same as those in Fig. 3.

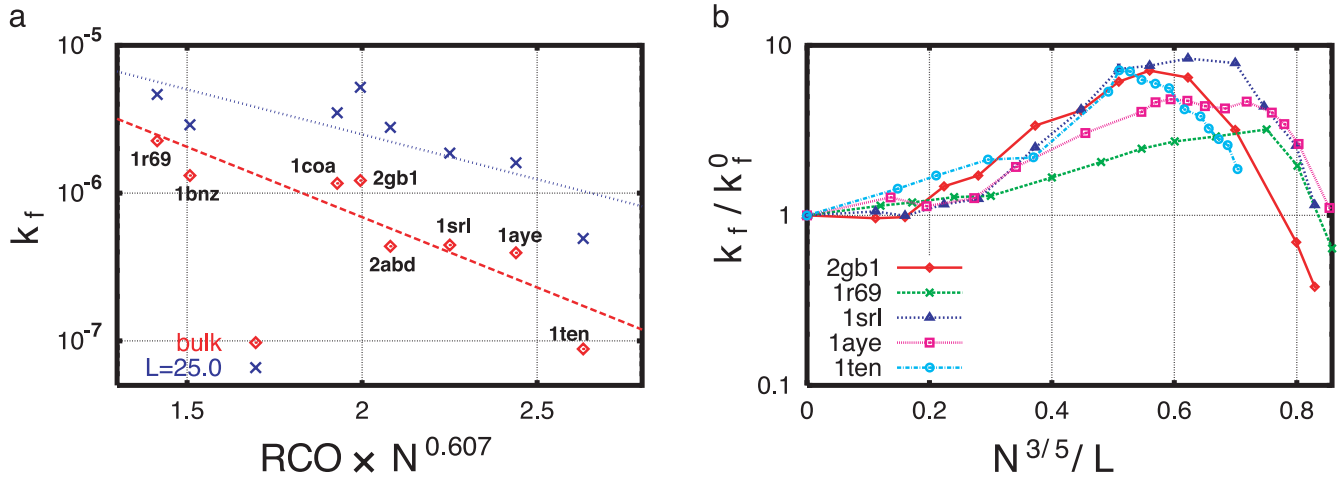

Fig. 5. (*a*) Folding rate constants k_f of eight proteins unconfined and confined in the chaperonin cage against the relative contact order (RCO) $\times N^{0.607}$ (slope: -2.19 ± 0.43 for bulk, -1.39 ± 0.45 for $L = 25$ Å). (*b*) Folding rate constant k_f as a function of $N^{3/5}/L$ for five proteins. Each k_f is normalized by the folding rate constant in the bulk system k_f^0.

cage, we repeat folding simulations 100 times from completely random structures at T_f^0, the folding temperature of the bulk system. The time at which the protein reaches the native structure ($Q > Q_n$, where Q_n is the free energy local minimum) for the first time is averaged over to get the first passage time τ_f for folding. The rate constants k_f (L, T_f^0) for folding in the size L cage at T_f^0 are estimated as the inverse of the first passage time τ_f, as usual.

Fig. 5*a* plots the calculated folding rate constants k_f with (the size $L = 25$ Å) and without the chaperonin cage against relative contact order (RCO) $\times N^{0.607}$, because it was shown previously, for the current simulation model, that the logarithm of the rate of spontaneous folding is best correlated with RCO $\times N^{\nu}$, where ν is $\approx 0.607 \pm 0.18$. Here RCO is a measure of native topological complexity introduced by Plaxco *et al.* (31). We see that, for all eight proteins studied, the folding is accelerated by the chaperonin-like cage. In the same way as the spontaneous folding, folding rates are correlated with the topological complexity. More interestingly, the slope in Fig. 5 is smaller in the case of confined proteins, which indicates that acceleration by the confinement is likely to be more prominent for proteins with more complex topology. *In vivo*, it was reported that proteins with α/β topology and more than two domains have higher propensity to binding to GroEL (32). Remarkably, our results suggest that folding of these proteins are more accelerated by confinement.

Next, we look at how the folding time depends on the size of cage L. Fig. 5*b* plots the ratio k_f (L, T_f^0)/k_f^0 against $N^{3/5}/L$. We see that the folding is accelerated in a modestly small cage ($N^{3/5}/L \approx 0.5$), whereas the folding rate is retarded in a very small cage size L, therefore giving a cage size at which the folding rate is maximal. The same tendency was found in a simulated β hairpin (17). As clarified above, with a fixed temperature, the native state becomes more stable as L decreases. This process makes the folding free energy barrier smaller, or downhill, and thus folding is accelerated. However, in a too small cage, the peptide chain cannot easily make global reconfiguration, leading to glassy dynamics. This clearly makes folding dynamics slower. It is interesting that the fastest folding is attained at roughly the same value of $N^{3/5}/L$. Recently, using an engineered chaperonin that inhibits the cycle, Hartl's group (15) showed that confinement itself accelerates the folding of ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCo), but not that of rhodanese. RuBisCo is 50 kDa, which is slightly below the limit size of substrates that enter into the chaperonin cage (≈60 kDa) and thus RuBisCo folding may correspond to near the peak in Fig. 5*b*. On the other hand, rhodanese is much smaller (33 kDa) and thus its folding may not be significantly affected. We also note a slight difference in situation between the current simulation and the experiment; the former focuses on the substrate folding in the cage, whereas in the latter folding is initiated by adding GroES and ATP analog.

We briefly investigate how ensemble of folding pathways is altered by the confinement. To this end, we compute the site-resolved nativelike order $q_i(Q)$, which monitors the degree of formed native interactions near the amino acid i along the folding reaction coordinate Q (27). The $q_i(Q)$ at the transition state ensemble can be viewed as a theoretical counterpart of the ϕ value that was developed by Fersht in the study of spontaneous folding (33). Here, we focus on the case of src SH3 domain folding because, for this protein, the current Go model was proven to give fairly consistent results with experiments (19, 34). In the spontaneous folding of SH3 domain, the distal β hairpin

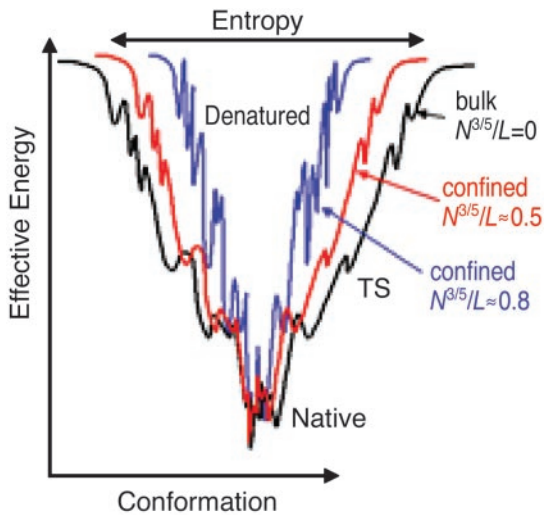

Fig. 6. Schematic view of protein folding funnel of bulk and confined proteins. The conformational entropy of the denatured state decreases with the confinement factor $N^{3/5}/L$, which leads to a steeper funnel. For the confinement factor larger than $N^{3/5}/L \approx 0.7$, the energy landscape becomes increasingly more rugged, and thus folding becomes slower.

(the third and the fourth strands, around residues 40 in Fig. 7, which is published as supporting information on the PNAS web site, www.pnas.org). Residues (numbered from 1 to 56) are known to be a part ordered earlier (see Fig. 7*a*). When confined into the size $L = 15$ Å cage, folding nucleus is somewhat delocalized to the distal β hairpin and a loop between the first and the second strands, called the RT loop. See *Supporting Text*, which is published as supporting information on the PNAS web site, for more details.

Discussions and Conclusion

We have studied in detail how folding of a substrate protein is assisted simply by putting it into a chaperonin-like cage. In thermodynamics, the confinement into the cage reduces primarily conformational entropy of the denatured protein, which leads to relative stabilization of the native state. With a fixed size of the cage, the stabilization is larger for a larger substrate. From Fig. 2 *Bottom*, we can deduce the effective energy $\langle E(Q)\rangle$ and the entropy $S(Q)$ as a function of Q, which is plotted in Fig. 8, which is published as supporting information on the PNAS web site. We clearly see that for a given E higher than the native energy the entropy S is reduced by confinement. This E-S curve corresponds to the slope of the funnel shape because the latter is the schematic illustration of the relation between E and S. From kinetic study, we saw that the substrate protein dynamics is slowed down at a cage size comparable to that of the protein. In the funnel perspective, this corresponds to increased ruggedness on the slope of the funnel. Combining these, we get a coherent view of protein folding in the chaperonin-like cage. Fig. 6 illustrates how folding funnel is altered by confinement. As $N^{3/5}/L$ increases, the slope of the funnel becomes larger, which corresponds to the increase in the folding transition temperature. On the other hand, the ruggedness of the energy landscape drastically increases above a critical value of $N^{3/5}/L$ (in the current model, $\approx$0.7) leading to slower folding.

In this article, we analyzed properties of a protein encapsulated in the chaperonin-like cage. As in the iterative annealing model, binding to the GroEL and transfer into the chaperonin cavity may induce mechanical unfolding and refolding of a substrate. Previous computer simulations showed that iterative cycles of hydrophilic and hydrophobic environment can accelerate sampling of protein conformation by once unfolding the protein when the environment is hydrophobic (13, 35). We stress that this effect and the confinement effect we studied here is not at all contradictory. Indeed, it is possible to combine these two effects in the context of the present simulation framework. For example, as in ref. 13, we can dynamically change the nature of the inner wall of the chamber to look at the effect of iterative annealing on top of the caging effect. Related to this is potential importance of non-native interactions in chaperonin-assisted folding. We can take into account non-native hydrophobic interactions in the current model although they are not included in the present work.

Another aspect not taken into account here is how the chaperonin can reduce protein aggregation, the aspect of the Anfinsen cage model. In particular, the living cells are crowded with biomolecules, and the macromolecular crowding crucially increases propensity to aggregation (36, 37). Technically, dealing with aggregation problem by molecular simulations is difficult. A simple mesoscopic simulation showed indeed that aggregation can be reduced by adding molecular chaperones (39). We also mention that a protein molecule confined in a cage resembles that in crowding solution. Enhancement of the folding rate by confinement is parallel to the prediction that macromolecular crowding accelerates folding unless aggregation occurs (36, 37).

It may be the case that the Anfinsen cage effect, the iterative annealing effect, and the static confinement effect studied here all are present in reality, and together they explain how protein folding is assisted by the chaperonin.

We thank Akira R. Kinjo for helpful discussions. This work was in part supported by a grant-in-aid for scientific research from the Ministry of Education, Science, Sports, and Culture of Japan and by the Advanced Computational Technology program of Japan Science and Technology Corporation.

1. Sigler, P. B., Xu, Z., Rye, H. S., Burston, S. G., Fenton, W. A. & Horwich, A. L. (1998) *Annu. Rev. Biochem.* **67,** 581–608.
2. Hartl, F. U. & Hayer-Hartl, M. (2002) *Science* **295,** 1852–1858.
3. Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994) *Nature* **371,** 578–586.
4. Xu, Z., Horwich, A. L. & Sigler, P. B. (1997) *Nature* **388,** 741–750.
5. Weissman, J. S., Kashi, Y., Fenton, W. A. & Horwich, A. L. (1994) *Cell* **78,** 693–702.
6. Weissman, J. S., Hohl, C. M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H. R., Fenton, W. A. & Horwich, A. L. (1995) *Cell* **83,** 577–587.
7. Rye, H. S., Burston, S. G., Fenton, W. A., Beechem, J. M., Xu, Z., Sigler, P. B. & Horwich, A. L. (1997) *Nature* **388,** 792–798.
8. Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. (1994) *Nature* **371,** 614–619.
9. Chatellier, J., Buckle, A. M. & Fersht, A. R. (1999) *J. Mol. Biol.* **292,** 163–172.
10. Weber, F., Keppel, F., Georgopoulos, C., Hayer-Hartl, M. K. & Hartl, F. U. (1998) *Nat. Struct. Biol.* **5,** 977–985.
11. Thirumalai, D. & Lorimer, G. H. (2001) *Annu. Rev. Biophys. Biomol. Struct.* **30,** 245–269.
12. Todd, M. J., Lorimer, G. H. & Thirumalai, D. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 4030–4035.
13. Betancourt, M. R. & Thirumalai, D. (1999) *J. Mol. Biol.* **287,** 627–644.
14. Shtilerman, M., Lorimer, G. H. & Englander, S. W. (1999) *Science,* **284,** 822–825.
15. Brinker, A., Pfeifer, G., Kerner, M. J., Naylor, D. J., Hartl, F. U. & Hayer-Hartl, M. (2001) *Cell* **107,** 223–233.
16. Zhou, H. X. & Dill, K. A. (2001) *Biochemistry* **40,** 11289–11293.
17. Klimov, D. K., Newfield, D. & Thirumalai, D. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 8019–8024.
18. de Gennes, P. G. (1979) *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca, NY).
19. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000) *J. Mol. Biol.* **298,** 937–953.
20. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7524–7528.
21. Go, N. (1983) *Annu. Rev. Biophys. Bioeng.* **12,** 183–210.
22. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) *Proteins* **21,** 167–195.
23. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995) *Science* **267,** 1619–1620.
24. Taketomi, H., Ueda, Y. & Go, N. (1975) *Int. J. Pept. Protein Res.* **7,** 445–459.
25. Takada, S. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11698–11700.
26. Clementi, C., Jennings, P. A. & Onuchic, J. N. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5871–5876.
27. Koga, N. & Takada, S. (2001) *J. Mol. Biol.* **313,** 171–180.
28. Eggers, D. K. & Valentine, J. S. (2001) *Protein Sci.* **10,** 250–261.
29. Honeycutt, J. D. & Thirumalai, D. (1992) *Biopolymers* **32,** 695–709.
30. Kumar, S., Bouzida, D., Swendesen, R. H., Kollman, P. A. & Rosenberg, J. M. (1992) *J. Comput. Chem.* **13,** 1011–1021.
31. Plaxco, K. W., Simons, K. T. & Baker, D. (1998) *J. Mol. Biol.* **277,** 985–994.
32. Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F. & Hartl, F. U. (1999) *Nature* **402,** 147–154.
33. Fersht, A. R. (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York).
34. Grantcharova, V. P., Riddle, D. S., Santiago, J. V. & Baker, D. (1998) *Nat. Struct. Biol.* **5,** 714–720.
35. Fukunishi, H., Watanabe, O. & Takada, S. (2002) *J. Chem. Phys.* **11,** 9058–9067.
36. Minton, A. P. (2000) *Curr. Opin. Struct. Biol.* **10,** 34–39.
37. Kinjo, A. R. & Takada, S. (2002) *Phys. Rev. E* **66,** 031911.
38. Gulukota, K. & Wolynes, P. G. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9292–9296.
39. Kinjo, A. R. & Takada, S. (2003) *Biophys. J.,* in press.