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 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 $\Delta T_f$ is as large as $\sim60^\circ C$. For a protein of size $R_0$, the stabilization $\Delta T_f$ 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 cage, which is consistent with a recent experimental result on ribulose-1,5-bisphosphate carboxylase-oxygenase folding and simulation results of a $\beta$ 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 $\phi$ 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
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 \( \sim 60^\circ C \) We found that, for all proteins studied, the shift in the folding transition temperature \( \Delta T_f \) caused by confinement into size \( L \) cage scales in a universal manner, \( \Delta T_f \sim (R_0/L)^n \), where \( n = 3 \) 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 \( C_N \) atoms of every amino acid residue. \( C_N \) 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 \( \Gamma \) is comprised of two contributions, \( V = V_{\text{protein}} + V_{\text{cage}} \), for the intraprotein interactions and is the same as the energy function of Clementi et al. (19). \( V_{\text{cage}} \) is for the interaction between protein amino acids and the chaperonin cage and is purely repulsive. Explicitly, \( V \) is given as

\[
V(\Gamma, \Gamma_0) = \sum_{\text{bonds}} K(b_i - b_{i0})^2 + \sum_{\theta} K(\theta_i - \theta_{i0})^2 + \sum_{\text{dihedrals}}\{K^{(1)}[1 - \cos(\phi_i - \phi_{i0}^{(1)})] + K^{(2)}[1 - \cos(\phi_i - \phi_{i0}^{(2)})]\} + \sum_{ij} K_{\text{cage}} \left[ \left( \frac{C}{2d_i} \right)^4 - 2 \left( \frac{C}{2d_i} \right)^2 + 1 \right] H(C/2 - d_i),
\]

where all but the last term correspond to \( V_{\text{protein}} \) and the last term is the \( V_{\text{cage}} \). In the equation, \( b_i \) is the virtual bond length between two adjacent amino acid \( C_N \)s, namely, \( b_i = |b_i| = |r_{i+1} - r_i| \), where \( r_i \) is the position of the \( i \)th amino acid. \( \theta_i \) stands for the angle between two adjacent virtual bonds \( b_i \) and \( b_{i+1} \). \( \phi_i \) represents the \( i \)th dihedral angle around the \( i \)th bond \( b_i \). 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, \( \Sigma \) native contact means that the summation is taken only over 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 \( |r_i - r_j| \) between the \( i \)th and \( j \)th 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 \( V_{\text{cage}} \), \( d_i \) is the distance between the chaperonin wall and the \( i \)th 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 \( K_0 = 100.0 \), \( K_0 = 20.0 \), \( K_0^{(1)} = 1.0 \), \( K_0^{(2)} = 0.5 \), \( e_1 = 0.18 \), and \( e_3 = 0.18 \) 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 \( C = 4.0 \) \( \AA \) 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 \( k_B = 1.0 \).

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

\[
m_i \ddot{x}_i = F_i - m_i \gamma \dot{x}_i + \xi_i,
\]

where \( m_i \) is the mass of the \( i \)th bead, \( \dot{x}_i \) represents the derivative with respect to time \( t \) and thus \( \dot{x}_i = \frac{\partial}{\partial t} \). \( F_i \) and \( \xi_i \) are systematic and random forces on the \( i \)th bead, respectively. The systematic force \( F_i \) is derived from the effective energy \( V \) as usual, \( F_i = -\partial V/\partial r_i \). The white and Gaussian random forces \( \xi_i \) satisfy \( \langle \xi_i(t) \xi_i(t') \rangle = 2T\delta_{tt'} \delta(r_i - r_i') \), where the bracket denotes the ensemble average and \( T \) is a 3 \( \times \) 3-unit matrix. For numerical integration of the Langevin equation, we use an algorithm by Honecutt and Thirumalai (29). We use \( \gamma = 0.25 \), \( m_i = 10.0 \), and the finite time step \( \Delta t = 0.2 \).

We define that \( i \)th and \( j \)th amino acids are in the “native contact” set if one of the nonhydrogen atoms in the \( i \)th amino acid is within 6.5- \( \AA \) distance from one of nonhydrogen atoms in the \( j \)th amino acid at the native structure \( \Gamma_0 \). For a given protein conformation, \( \Gamma \), 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 \( \Gamma \), defined as the ratio of numbers of formed native contacts at \( \Gamma \) to those at the native structure \( \Gamma_0 \).

In the article, we use eight relatively small proteins (<100 residues) that cover a wide variety of protein topology, including \( \alpha \) proteins, \( \alpha/\beta \) proteins, and \( \beta \) 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, \( \alpha/\beta \) protein), 434 repressor (ID code 1r69, \( \alpha \) protein), src SH3 domain (ID code 1sr1, \( \beta \) protein), Ada2h (ID code 1aye, \( \alpha/\beta \) protein), and tenascin fibronectin type 3 (ID code 1ten, \( \beta \) protein).
position in C stays most of the time in the native state at temperature is below than half the time, we then replace depicted in Fig. 2 Top protein becomes unstable (Fig. 2). At the transition temperature natural stability than the protein without the cage. Naturally, the same temperature, the chaperonin-caged protein has larger for the bulk system (Fig. 2). For each protein with or without the chaperonin cage, the first confinement effects on thermodynamics, where the left is the case without the chaperonin cage of size 15 Å, the protein has limited denatured conformations with R_g <~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) 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 identified by the peak in the heat capacity C_v curve as a function of T. Fig. 3a 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 ~20% higher than that without the cage, which corresponds to the increase of ~60°C assuming that T_f is of the order of ~350 K. Similar amount of increases in T_f for proteins in a small cage is obtained from all five proteins studied (see Fig. 3b for details). Experimentally, α-lactalbumin confined in a silica matrix showed the denaturation temperature elevated by ~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. 3a.

In Fig. 3b, we plot the folding temperatures T_f of five proteins in the chaperonin cage of size L against N^{1/3}/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 law dependence of the temperature shift with N^{1/3}/L. The linear regression of the data points N^{1/3}/L > 0.25 leads to the scaling,
The free energy change of 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. 3b, 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 only entropic contribution $-T_0 S_n$ remains. Combining these relations, we get $F_n = F_d = -T_0 S_0 = F_d$. 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 $\Delta F \sim (R_0/L)^3$, where $R_0$ is the radius of the denatured protein (18). The resulting folding temperature can be estimated through $F_n = E_n = -T_0 S_0 = (R_0/L)^3$. These equations together give an estimate of the change in the transition temperature as $(T_f - T_f^0)/(T_f^0 - (R_0/L))^3$ (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 $(T_f - T_f^0)/(T_f^0 - (R_0/L))^3$. Here, $R_0$ should scale as the “real” chain with the volume interaction, for which the Flory theory gives an approximate scaling $R_0 \sim N^{3/5}$, where $N$ is the number of amino acids here. The scaling estimated by the simulation $T_f - T_f^0 \sim N^{-3.25}$ is fairly close to that by the mean field theory. We note that scaling with respect to $L$ is characterized here, 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 $\Delta F$ 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 $-k_BT \log P_{\perp L}$, where $P_{\perp L}$ 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 $T_f^0$ 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 ($\sim 60$ 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
cage, we repeat folding simulations 100 times from completely random structures at $T_f$, the folding temperature of the bulk system. The time at which the protein reaches the native structure ($Q > Q_c$, where $Q_c$ is the free energy local minimum) for the first time is averaged over to get the first passage time $\tau_1$ for folding. The rate constants $k_i (L, T_0^f)$ for folding in the size $L$ cage at $T_0^f$ are estimated as the inverse of the first passage time $\tau_1$, as usual.

Fig. 5a plots the calculated folding rate constants $k_i$ 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 $\text{RCO} \times N^r$, where $r$ 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 $\alpha/\beta$ 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. 5b plots the ratio $k_i (L, T_0^f)/k_i^0$ against $N^{3/5}/L$. We see that the folding is accelerated in a modestly small cage ($N^{3/5}/L = 0.5$), whereas in the latter folding 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 $\beta$ 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 ($\approx 60$ kDa) and thus RuBisCo folding may correspond to near the peak in Fig. 5b. 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 $\phi$ 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 $\beta$ hairpin
We clearly see that for a given substrate size, 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 the entropy \( S \). The schematic illustration of the relation between \( S \) and \( L \) is published as supporting information on the PNAS web site. We see clearly 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 size of the funnel. Combining these, we get a coherent view of protein folding in the chaperinon-like cage. Fig. 6 illustrates how a folding funnel is altered by confinement. As \( N^{\beta/L} \) 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 \( N^{\beta/L} \) (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 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.

Supporting Information

Site-Resolved Folding Pathways

Here, we show how protein folding pathways are altered by confining the chaperonin-like cage by using src SH3 domain as an example. Site-resolved folding pathways $q_i(X)$ are quantified here in the same way as Koga and Takada (1). Namely,

$$q_i(X) = \frac{\langle e_i \rangle_X - \langle e_i \rangle_D}{\langle e_i \rangle_N - \langle e_i \rangle_D},$$

where $e_i$ is the site-energy defined as a sum of interaction energies of the $i$th amino acid with any other residues. $\langle >_X$ denotes the average over the ensemble specified by X. Simply, $q_i(X)$ is close to zero when the interaction around the site $i$ is dissimilar to that of the native structure in the ensemble X, while it is close to unity if the native-like interaction is attained around the site $i$. In particular, $\phi_i = q_i(TS)$ is an analog of experimental $\phi$ value of site $i$ (2) under the assumption of the transition state theory.

Fig. 7 illustrates the site-resolved folding pathways of src SH3 domain without (Left) and with (Right) the chaperonin-like cage. For the bulk system, it was investigated that the distal $\beta$ hairpin, the third and the fourth strands, is formed before the transition state $Q \sim 0.4$ (1, 3). This finding is perfectly consistent with experimental $\phi$ value analysis (4). When this protein is put into the chaperonin-like cage of the size $L = 15$ Å, the ensemble of folding pathways is significantly altered so that the turn in the distal $\beta$ hairpin and the RT loop (a loop between the first and the second strands) is simultaneously formed earlier. The delocalization of the folding nucleus may occur because of the reduced conformational entropy in the denatured states.

The Energy-Entropy Relation in the Bulk and Confined Systems

From the simulated trajectories at $T_f^0$, we can calculate the average energy $\langle E(Q) \rangle$. Combining this with $F(Q)$, we can deduce the conformational entropy $S(Q)$ by the relation $TS(Q) = \langle E(Q) \rangle - F(Q)$. Fig. 7 plots $\langle E(Q) \rangle$ against $S(Q)$ for the case of src
SH3 domain. In the bulk system, the native, transition, and denatured states correspond to $E \approx 20$, $E \approx 70$, and $E \approx 120 \ k_B T$, respectively. We note the range of axes; comparing with the free energy profile such as Fig. 2 Bottom, the changes in energy and entropy upon folding are almost 2 orders of magnitude larger than the change in free energy. Namely, energy-entropy compensation is realized here.

Now, we see in Fig 8 that the conformational entropy in the denatured state is reduced by $\approx 20 \ k_B$ at the energy $\approx 120 \ k_B T$. This is remarkably large, considering that thermodynamic stability of a protein is typically of order $10 \ k_B T$. In terms of the energy landscape of a protein, this finding indicates that the confinement makes the folding funnel steeper than the corresponding bulk system.


