

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/49642075>

Protein Folding and Confinement: Inherent Structure Analysis of Chaperonin Action

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · DECEMBER 2010

Impact Factor: 3.3 · DOI: 10.1021/jp107257b · Source: PubMed

CITATIONS

3

READS

10

2 AUTHORS, INCLUDING:



Amandeep K Sangha

Vanderbilt University

10 PUBLICATIONS 99 CITATIONS

[SEE PROFILE](#)

Protein Folding and Confinement: Inherent Structure Analysis of Chaperonin Action

Amandeep K. Sangha[†] and Tom Keyes*

Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States

Received: August 2, 2010; Revised Manuscript Received: October 25, 2010

A coarse-grained model of the action of a chaperonin cage of tunable hydrophobicity, h , upon a protein with the possibility of misfolding is studied with inherent structure (IS) analysis and statistical temperature molecular dynamics (STMD) simulation. Near the folding temperature, the equilibrium properties of the system may be understood in terms of <10 IS. The known phenomenon of an optimal cage hydrophobicity for productive folding, found at $h = 0.25$, is seen to arise from a striking suppression of the occupations of IS in the misfolding funnel, which in turn arises from a decrease in translational entropy due to confinement to the region of the cage wall. The kinetics of folding is correspondingly fastest at $h = 0.25$, where a minimum is found in the h -dependent barrier height. While true kinetics is determined by conventional MD, it is shown that the accelerated dynamics of STMD provide a valuable quantitative perspective.

I. Introduction

Folding of a chain of amino acids to the correct three-dimensional structure is essential for the biological activity of the cell. Proteins which do not fold spontaneously or are prone to misfolding can form aggregates which endanger the organism. Consequently, large cylindrical proteins, called chaperones, have evolved to assist slow-folding proteins in finding their native structures. The most extensively studied is the GroEL-GroES system, the chaperonin, which is found to help about 85 proteins, mostly with α/β and $\alpha + \beta$ folds.¹ Chaperonin consists of a double ring of GroEL, with hydrophobic patches on its opening, and a single ring of GroES that acts like a cover for the GroEL cavity. The hydrophobic patches capture a non-native protein via its exposed hydrophobic amino acid residues. Subsequent binding of ATP to the open GroEL ring causes the non-native protein to unfold. ATP and GroES binding results in conformational changes of the GroEL ring, and the unfolded protein is released into the folding chamber of the GroEL-GroES complex, where it attains its native structure. The inner surface of the GroEL ring is observed to change from hydrophobic to hydrophilic during the process of assisted protein folding.

To understand the mechanism of chaperonin action, several experimental,^{2–11} theoretical, and computational^{12–27} studies have been conducted. Two different mechanisms have been proposed, active and passive. In the latter, chaperonin simply acts as an Anfinsen's cage to isolate the protein, thus preventing aggregation.^{3,4,23,28} The folding route remains the same with and without GroEL-GroES.

In the active mechanism of the iterative annealing model,^{12–15} ATP-driven cycles of repeated binding and release of the non-native protein from the chaperonin inner surface help to overcome the barriers on the routes to the native structure. It has been proposed²⁴ that interaction with the chaperonin creates an alternative fast-folding pathway and that the solvent confined inside the cavity exhibits an increased hydrophobic effect, the driving force for protein folding.²⁷ It has also been shown that

the productive folding inside the chamber is proportional to the extent of unfolding of the non-native protein.¹⁰

Confinement alone can significantly influence protein stability and folding pathways. Several studies have been reported using computer simulations on off-lattice model proteins in various types of confining potentials.^{17,18,25,29–31} Confinement limits the number of accessible unfolded conformations, as chains with larger radii of gyration are suppressed. The result is to reduce the conformational entropy of the unfolded state and to stabilize the native state, reducing the folding time. The effect is sensitive to the cage size and the topology of the protein.¹⁸ The folding time to the native structure decreases partly because of faster dynamics of unfolded chains.³¹

The role of attractive cavity–protein interactions^{13,22,24,25} must be considered along with simple confinement, and the effect of caging upon a protein which is not a simple two-state folder is nontrivial. Thus, the precise mechanism by which chaperonin helps proteins fold correctly is still somewhat unclear.

In this article, we study a protein plus chaperonin model developed by Shea et al.²⁴ The chaperonin is a spherical cavity of tunable hydrophobicity, denoted h , i.e., there is a tunable cavity–protein attraction. The off-lattice, 27-residue protein has competing folding and misfolding funnels in bulk (no cage), and we seek to understand the influence of the cage upon the competition. Despite its small size, the energy landscape of the model protein is sufficiently rough to present a sampling challenge, so the statistical temperature molecular dynamics (STMD) enhanced sampling algorithm^{32,33} is employed to compute equilibrium averages.

First, we investigate the thermodynamics of protein folding under different environmental conditions within the chaperonin cavity. Thermodynamic properties change nonmonotonically with h . The method of inherent structure (IS) analysis^{33,35–37} is applied, exploiting, in particular, our finding³² that a small (<10) number of IS attain finite occupation probabilities below the collapse temperature. The combination of STMD and IS analysis provides more reliable sampling of and a new way to understand the mechanism of chaperonin action. In agreement with other studies,^{12,13,22,24,25} we find that a slightly hydrophobic cavity enhances folding. At the optimal hydrophobicity, the occupation

* To whom correspondence should be addressed. E-mail: keyes@bu.edu.

[†] Present Address: UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831.

TABLE 1: Parameter, ε , in the Nonlocal Interaction Term of Potential Function Depends upon the Nature of the Two Monomers Involved

	B	L	N
B	ε_h	$7/12\varepsilon_h$	$2/3\varepsilon_h$
L	$7/12\varepsilon_h$	$1/6\varepsilon_h$	$1/4\varepsilon_h$
N	$2/3\varepsilon_h$	$1/4\varepsilon_h$	$1/3\varepsilon_h$

TABLE 2: Parameter, Λ , in the Nonlocal Interaction Term of Potential Function Depends upon the Nature of the Two Monomers Involved^a

	B	L	N
B	1	0	0
L	0	-1	0
N	0	0	0

^a Attractive interactions occur only between hydrophobic monomers with $\Lambda = 1$.

of the low-lying IS states belonging to the misfolding funnel is strongly suppressed and the folding kinetics is fastest.

II. Models and Methods

Protein Plus Chaperonin Model. The “BLN”³⁸ model protein/chaperonin is due to Shea et al.^{21,22,24} A 27-residue protein is represented by a chain of beads, which may be hydrophilic (L), hydrophobic (B), or neutral (N); the sequence is (LB)₃N₂(BL)₃N₃(B₂L₂)₂BL. The native fold is an α/β sandwich. The protein is designed to have competing folding and misfolding funnels. The potential function is

$$V = \sum_{\text{bonds}} \frac{K_b}{2}(r - \sigma)^2 + \sum_{\text{angles}} \frac{K_\theta}{2}(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} -A \cos^2\left(\frac{\phi - \phi_\alpha}{2}\right) - B \cos^6\left(\frac{\phi - \phi_\beta}{2}\right) + \sum_{i,j>3} 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \Lambda \left(\frac{\sigma}{r_{ij}}\right)^6 \right] \quad (1)$$

where $K_b = 100\varepsilon_h/\sigma^2$, $K_\theta = 13.33\varepsilon_h/\text{rad}^2$, $\theta_0 = 1.8326$ rad, $\phi_\alpha = 1.0$ rad, and $\phi_\beta = \pi$ rad. A and B are different for different regions of the chain. For α -helices: $A = 6\varepsilon_h$, $B = 5.6\varepsilon_h$. For β -sheets: $A = 5.6\varepsilon_h$, $B = 6.0\varepsilon_h$. For turns: $A = 0$, $B = 0$.

To moderately facilitate the formation of misfolded states, torsion angles in the turn regions are turned off. The resulting misfolded conformation differs from the native structure in the orientation of the β -hairpin. In the nonlocal interaction term, ε and Λ are monomer dependent, as shown in Tables 1 and 2.

Chaperonin is modeled as a spherical cavity formed from closely spaced beads, with surface density of $1/\sigma^2$. Each monomer in the wall exerts a Lennard-Jones force of the form $4\varepsilon_h[(\sigma/r)^{12} - h_i(\sigma/r)^6]$ on each monomer in the protein. The most general cavity potential for each monomer is

$$U_{\text{cage}}(r) = 4\varepsilon_h \frac{\pi R}{r} \left(\frac{1}{5} \left[\left(\frac{\sigma}{r-R}\right)^{10} - \left(\frac{\sigma}{r+R}\right)^{10} \right] - \frac{h_i}{2} \left[\left(\frac{\sigma}{r-R}\right)^4 - \left(\frac{\sigma}{r+R}\right)^4 \right] \right) \quad (2)$$

where r is the distance of peptide monomer from the center of a sphere of radius R and h_i depends on the nature of the monomer. In the version of the potential used herein, hydro-

phobic monomers experience an attractive interaction with the cage, $h_B = h$, while neutral and hydrophilic monomers have a repulsive interaction only, $h_L = h_N = 0$. A spherical cavity of radius $R = 6\sigma$ is used for all simulations.

In the BLN model, the mutual attraction of hydrophobic beads implicitly represents the effect of a hydrophilic solvent. While hydrophilic residues in reality contain charges and dipoles which interact favorably with the solvating waters, the BLN attraction of the hydrophobic beads leads to the same end result. Hence, we refer to the $h = 0$ case as hydrophilic. As h is increased, the attraction between the hydrophobic beads and the cage partially counteracts their mutual attraction, so the environment is called hydrophobic, meaning more so than the solvent implicitly represented by the protein potential alone. The model is very simple: our goal is to apply IS analysis and STMD sampling to the existing Shea model, not to develop a better potential.

The order parameter used to monitor folding and determine the folding temperature, T_f , is the structural overlap, Q , defined as

$$Q = \frac{1}{N} \sum_{i,j>i+3}^n \Theta[\zeta - |r_{ij} - r_{ij}^0|] \quad (3)$$

where N is the normalization constant, r_{ij} and r_{ij}^0 are the distances between beads i and j in a given conformation and in the native state of the protein, $\zeta = 0.3$ is the closeness of beads required to count as overlap, and Θ is the Heaviside step function. The native state has $Q = 1$ and a conformation with $Q = 0$ has no similarity to the native state.

Simulation Algorithms. Statistical temperature molecular dynamics (STMD)³² and Langevin dynamics are used to compute equilibrium average and kinetic quantities, respectively. STMD samples a flat-energy distribution, and Boltzmann averages are obtained by reweighting the configurations in the trajectory.

All the folding simulations are performed at the bulk folding temperature, $T_f = 0.18$, expressed in standard dimensionless BLN units. Since folding times are measured starting from unfolded states, initial configurations are taken from an ensemble of open-chain conformations generated at a high temperature of 0.5. The required low- Q conformations could be obtained at any temperature above the collapse temperature, but we simply follow the procedure of ref 24. In determining folding times, the simulation is terminated (first passage achieved) when the structural overlap reaches $Q = 0.7$. The friction coefficient used for all folding simulations is $0.5\Delta t^{-1}$, where Δt is the Langevin simulation time step.

Inherent Structure (IS) Analysis. The inherent structure method of Stillinger and co-workers is the basis of a complete thermodynamic formalism and also allows one to view the dynamics of continuum systems as a series of transitions between distinct IS states.^{35–37,39} The IS are the local minima of the potential energy, and a “state” is comprised of all the configurations belonging to the basin of attraction of an IS; the method is a natural clustering algorithm. Recently, IS theory has been applied to model proteins by Kim et al., providing some novel perspectives on folding vs misfolding.^{32–34} We employ the IS method to study the effects of confinement on protein folding at different levels of hydrophobicity of the chaperonin chamber.

IS are identified by applying conjugate-gradient minimization to configurations taken from STMD trajectories. A configuration is said to be minimized if the squared gradient of the energy

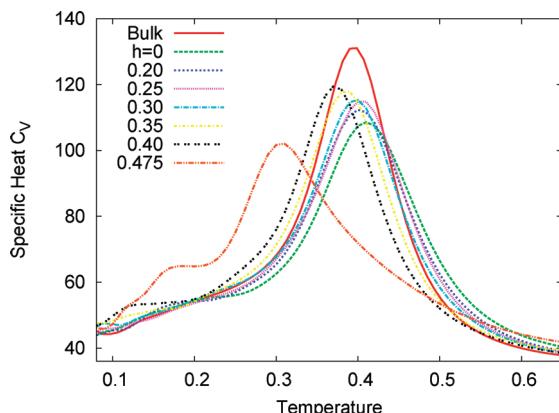


Figure 1. Collapse temperature from the peak in C_V , T_θ , varies with the environment of the chaperonin cavity.

changes by less than 1.0×10^{-11} upon taking further steps. The IS with the lowest potential energy is called the ground state, ISO, the second lowest IS is the first excited state, IS1, and so on. The perfect native structure is found in ISO, but at ambient temperature the “native state” consists of a range of configurations, which may lie in the basin of attraction of ISO or in that of other low-lying IS in the folding funnel, sharing the basic native fold. The occupation probabilities, $p_i(T)$, of these low-lying IS states are calculated below for a range of temperature that includes the collapse and folding temperatures, T_θ and T_f .

The $p_i(T)$ are essential ingredients of our analysis of model proteins. Above T_θ , the p_i are infinitesimal, as is the case in liquids, with a very large number of states each making a very small contribution to the averages. As T is decreased below T_θ , the p_i of the lowest-lying IS attain finite values, as discussed in ref 33. The relative occupations of the “excited states” vs that of the ground state, p_0 , provide a good indication of whether folding will be preferred over misfolding, or vice versa.

To calculate the $p_i(T)$, each configuration in the trajectory is indexed according to the IS to which it maps. Then reweighting³³ is performed using the number belonging to the index as the observable.

Dynamics may be formulated as transitions among the IS, as has been done in liquids. However, for this model protein, some of the open-chain conformations minimize to the ground state, ISO. Thus, one must consider intrabasin folding as the shortest possible pathway, and its existence supports the idea that heterogeneity in the denatured state is significant, i.e., that certain denatured configurations fold much faster than others.⁴⁰

III. Simulations and Analysis

Thermodynamics of Protein Folding under Confinement.

The nonmonotonic dependence of the collapse and folding temperatures upon the hydrophobicity of the chaperonin cavity is shown in Figures 1 and 2 and Table 3.

The collapse temperature is determined from the peak in the specific heat. In bulk, $T_\theta = 0.398$. Confinement in a hydrophilic cavity reduces the entropic advantage of open-chain conformations and the protein collapses at a slightly higher temperature of $T_\theta = 0.41$. The transition is also not so sharp as in bulk, because unfolded conformations with a restricted radius of gyration are less different from the collapsed states. Upon changing the surface of the cavity to hydrophobic, T_θ decreases to 0.40, remains the same as the hydrophobicity, h , increases from 0.20 to 0.30, and decreases again with a further increase in h . The decrease in T_θ at the higher hydrophobicity values

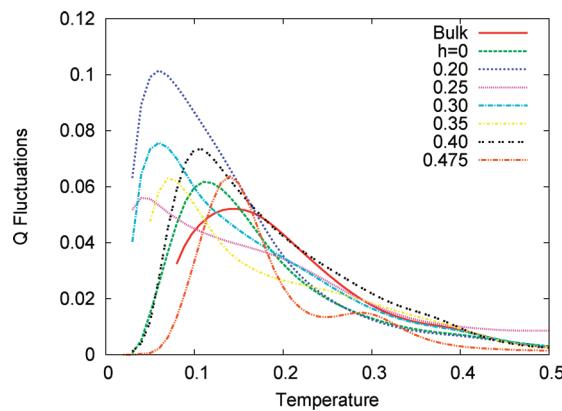


Figure 2. Folding temperature determined from the peak in the Q fluctuations, T_f , varies with the environment of the chaperonin cavity.

TABLE 3: Collapse and Folding Temperatures under Different Environmental Conditions^a

Hydrophobicity (h)	T_θ	$T_f(Q)$	$T_f(\text{wells})$
Bulk	0.398	0.14	0.18
0	0.41	0.11	0.14
0.20	0.40	0.06	0.12
0.25	0.40	0.04	0.23
0.30	0.40	0.06	0.21
0.35	0.39	0.07	0.24
0.40	0.37	0.11	0.19
0.475	0.31	0.14	—

^a The collapse temperature, T_θ , is obtained from the peak in the heat capacity. The folding temperature from equal population of the double wells, $T_f(\text{wells})$, is higher than that from the peak in the Q fluctuations, $T_f(Q)$.

occurs because the extended conformations are stabilized by strong interactions with the wall.

The folding transition connects the native state and collapsed, non-native configurations, as measured by the order parameter. Folding temperatures may be obtained from the peak in the order parameter fluctuations, $\langle \delta Q^2 \rangle(T)$, or from the condition that the low- Q and high- Q wells in the double-well free energy, $F(Q)$, have equal populations, yielding temperatures denoted $T_f(Q)$ and $T_f(\text{wells})$, respectively.

In bulk, $T_f(\text{wells}) = 0.18$. Moving (Table 3) from bulk to a hydrophilic cage to a hydrophobic environment, $T_f(\text{wells})$ first decreases and then increases to a plateau value ~ 0.23 for $0.35 \geq h \geq 0.25$. In all cases, $T_f(Q) < T_f(\text{wells})$, with the largest difference for $h = 0.25$, which is also the optimal hydrophobicity for folding. Thus, we have $\langle \delta Q^2 \rangle(T)$ increasing as T is decreased below $T_f(\text{wells})$, while, in the simplest case of fluctuations arising from interwell transitions, the equi-population condition is also the condition for maximal fluctuations. For the optimal $h = 0.25$, $\langle \delta Q^2 \rangle(T)$ has no sharp peak but rises almost linearly (Figure 1) as T is decreased from $T_f(\text{wells})$ before turning over at $T = 0.04$.

The key to understanding the h dependence of the T_f is the realization that the 27-mer has folded, misfolded, and unfolded states: it is a 3-state folder, not 2-state, and care is required in analysis based upon a single order parameter. The high- Q well in $F(Q)$ contains the contributions from the native, folded states and the low- Q well has both misfolded and unfolded states. The misfolded states differ substantially in Q from the native state but are only slightly higher in energy. At low T , the low-energy, misfolded states increase the population of the low- Q well above what it would be with unfolded states only.

Thus, it is immediately clear that optimal foldability at intermediate h implies maximal $T_f(\text{wells})$ at intermediate h : with

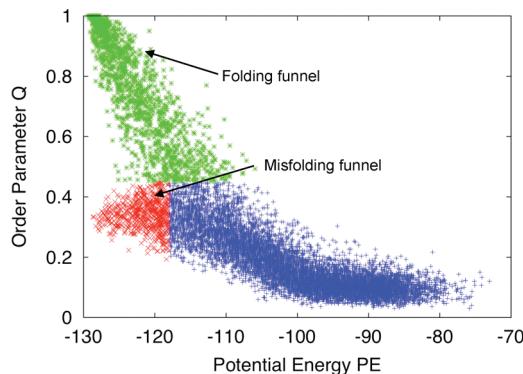


Figure 3. Scatter plot of order parameter, Q , and potential energy, PE, shows the folding and misfolding funnel of the model protein.

the misfolded states suppressed, the temperature at which the high- Q well has one-half the population will be maximal. A high $T_f(\text{wells})$ directly indicates the beneficial action of the chaperonin. Furthermore, the highest $T_f(\text{wells})$ is closest to T_θ . Having T_f close to T_θ is a standard criterion for a good folder, although the condition is generally expressed with $T_f(Q)$.

The behavior of $T_f(Q)$ is more subtle. Why is $T_f(Q) < T_f(\text{wells})$, that is, why do the δQ fluctuations continue to increase as T is decreased below the equal-well-population temperature? An explanation may be found with IS theory. The fluctuation peaks at low temperatures must arise from transitions between misfolded and folded conformations. Recall that^{32,33} below T_θ , the occupation probabilities of the lowest-lying IS become finite. Figure 4 shows the summed occupation probabilities for IS with significant contributions to the folding and misfolding funnels. At the lowest T , only these IS are occupied and the partition function in IS thermodynamics^{35–37} is a finite sum.

At $h = 0.30$ and 0.35 , the summed probabilities have evident local minima and maxima, respectively, at temperatures $\approx T_f(Q)$. Thus, the peak in $\langle \delta Q^2 \rangle(T)$ is associated with a low- T local minimum in the population difference of folded and misfolded states. The resulting fluctuations are larger than those at $T_f(\text{wells})$ because the separation of the populations in Q is larger and more distinct. The nontrivial behavior of the low- Q population is a consequence of the low energy of the misfolded states.

Less well-defined features may be seen in the total IS occupation probabilities at $h = 0.25$ and 0.40 , also at temperatures $\approx T_f(Q)$. The diffuse structure of $\langle \delta Q^2 \rangle(T)$ at $h = 0.25$ is a direct indication of the suppression of the misfolded states, which hence cannot contribute to a sharp peak.

Even higher hydrophobicity favors the non-native states via stronger interactions between the exposed hydrophobic beads of the misfolded states and the wall, thereby lowering $T_f(\text{wells})$ to 0.19 at $h = 0.40$. Correspondingly, $T_f(Q)$ rises, as there is no low- T maximum in misfolded state occupation when it is already stabilized at higher T .

Energy Landscape and Inherent Structures. A two-dimensional scatter plot of sampled configurations vs order parameter, Q , and potential energy, PE, was obtained from the STMD trajectory in bulk. Figure 3 shows the energetically competing folding and misfolding funnels, confirming that the model protein is prone to misfolding in bulk. The conformations with high PE and low Q are the unfolded ensemble (blue). These conformations can evolve directly to the folded ensemble with low PE and high Q (the folding funnel, green) or can get trapped in the misfolded states with PE comparable to folded states but with low Q (the misfolding funnel, red). Once the protein enters the misfolding funnel, it must unfold to enter the folding funnel.

This two-dimensional scatter plot provides an alternative to a disconnectivity graph for viewing the multifunnel shape of the energy landscape.

The chaperonin can promote productive folding in two different ways: (1) by directly increasing the probability that open-chain conformations first reach the folding funnel, escaping a visit to the misfolded states, or (2) by helping a misfolded protein to unfold, from which point it can fold correctly. In both cases, a low probability of misfolded states would indicate an enhancement in productive folding. This probability is a function of both the hydrophobicity of the chaperonin cavity and the temperature.

Here we exploit IS analysis to illuminate the effect of hydrophilic and hydrophobic environments on the folding of the model protein. The significant p_i values for the different environments are shown in Figures 5–10.

The ground state, IS0, with the ideal native fold, varies somewhat with the confinement conditions. There is no change going from bulk to hydrophilic confinement in terms of visualized conformation and order parameters such as Q , the radius of gyration, R_g , and the IS energy. With increasing h , IS0 becomes slightly distorted, with a small rmsd from its bulk configuration. The IS energy decreases with increasing h , primarily due to stronger interactions with the wall. In the following, the ground state for each value of h is considered to define the native state for those conditions. It is expected that the confined native configurations will acquire the bulk native state structure after their escape from the chaperonin cavity.

Folding and IS Occupation Probabilities in Bulk. The ground state, IS0, which is of course a folding state, has the highest occupation for all temperatures below T_θ , but p_1 and p_3 are considerable and IS1 and IS3 belong to the misfolding funnel. Thus, the probability of misfolding in bulk is non-negligible. Below $T = 0.18$, p_1 and p_3 fall rapidly and p_0 rises to unity.

Folding and IS Occupation Probabilities Inside a Hydrophilic Cavity ($h = 0$). For a temperature range of $0.4–0.23$, misfolding state IS3 has a higher occupation than any other state, including IS0. Thus, hydrophilic confinement induces a “misfolding interval”, which was proposed in ref 33 as an IS-based measure of misfoldability. As discussed earlier, with the misfolded states slightly more compact than the folded, hydrophilic caging increases the probability to misfold. It seems that there are two main routes, along IS1 and IS3, to the misfolding funnel from the unfolded ensemble.

Folding and IS Occupation Probabilities Inside a Hydrophobic Cavity. To see the effect of a hydrophobic environment on protein folding inside the chaperonin cavity, the hydrophobicity is varied from 0.20 to 0.525 .

At $h = 0.20$, the low-lying bulk states split up into multiple states with different IS energy but the same order parameter value. For example, the ground state in bulk (and at $h = 0$) splits into IS0, IS1, and IS2. The misfolding IS1 in bulk becomes IS3, IS4, IS5, and IS6. IS7 at $h = 0.20$ corresponds to IS2 in bulk, and IS3 in bulk splits into IS8, IS9, and IS10.

One might imagine that the splitting of the states simply involves partitioning the basin of attraction of the bulk IS into several sub-basins, with no change in the occupation probability of the original region of configuration space. At $h = 0.20$, this is approximately the case down to $T \approx 0.15$. While Figure 7 shows that two misfolding states now have probabilities that exceed p_0 over a misfolding temperature interval, the summed probabilities of the folding and misfolding funnels are similar to those in the hydrophilic cage, Figure 4. However, below the

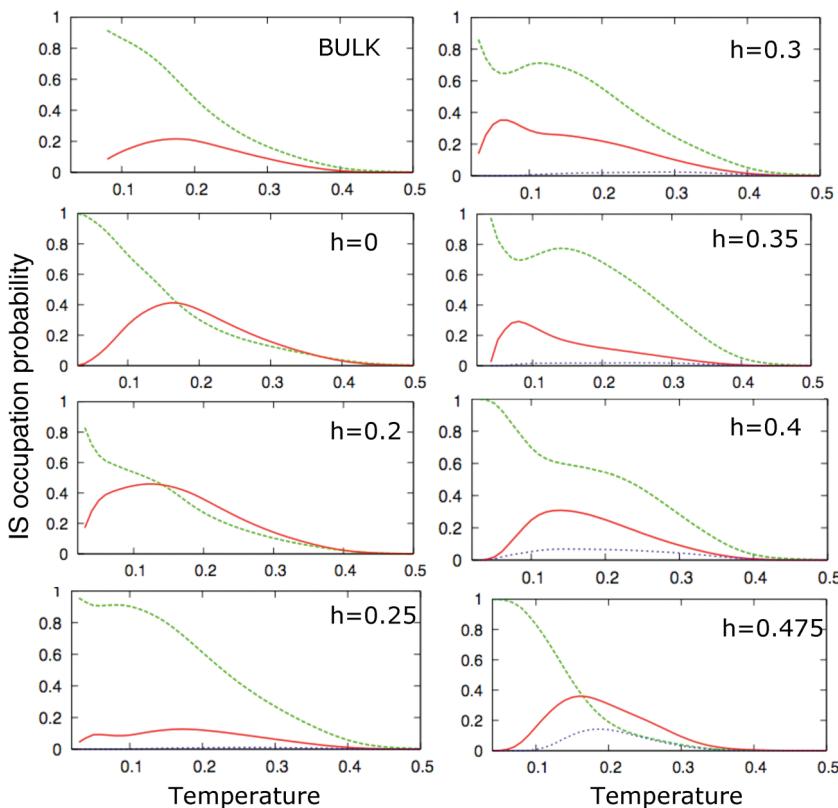


Figure 4. Summed occupation probabilities of the states belong to the folding (green) and misfolding (red) funnel. The blue line represents the sum of the occupation probability of low-lying states with a Q value close to 0.5.

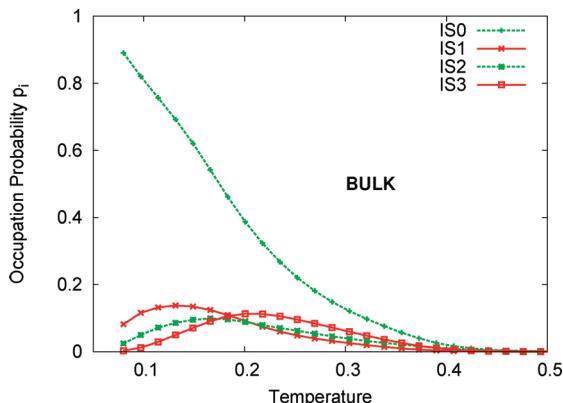


Figure 5. IS occupation probabilities in bulk. Red states belong to the misfolding funnel and green states to the folding funnel. Considerable occupation of IS1 and IS3 shows that the protein is susceptible to misfolding.

misfolding interval, the summed folding probability increases less strongly than at $h = 0$; so, overall, misfoldability has increased.

A dramatic reversal of the trend toward misfolding occurs with an increase in hydrophobicity to 0.25, Figure 8. The occupation probabilities for the ground state and for the two “daughter” IS of the ground state in bulk now strongly exceed the p_i of all the misfolding states. The occupation of the misfolding states, IS3, IS4, IS9, and IS10, the main routes to the misfolding funnel, becomes very small. Correspondingly, the summed probabilities, Figure 4, show a powerful suppression of misfolding. While it is now well established that folding is optimal at intermediate hydrophobicity, we submit that IS analysis has provided a particularly clear picture of the phenomenon.

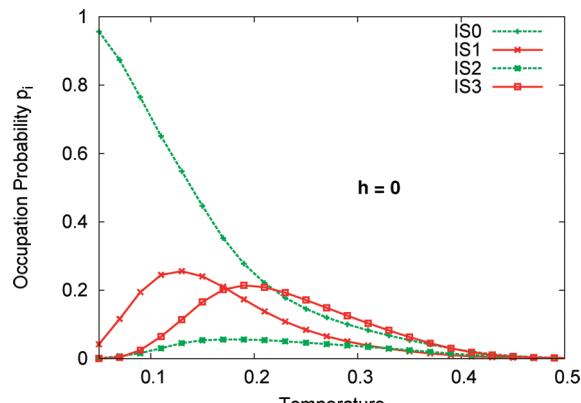


Figure 6. IS occupation probabilities inside the hydrophilic cage, $h = 0$. The temperature range of 0.22–0.40 represents the misfolding interval, where occupation of the misfolding state, IS3, is higher than that of the ground state, IS0, indicating that misfolding is favored.

With a further increase in hydrophobicity, to $h = 0.30$ and 0.35, misfolding again becomes competitive with folding, Figures 9 and 10. Also, the distortion of the ground state increases substantially, due to stronger interactions with the wall. Except for the very lowest T , the first two excited folding-funnel states have occupations exceeding that of the ground state.

Starting from $h = 0.40$ onward (p_i not shown), in contrast to the low- h case, the order parameter value of the IS is not sufficient to determine if the state is in the folding or misfolding funnel. The second excited state, with $Q = 0.456$, is quite similar to the ground state. However, without looking at the structure itself, it is difficult to say that this state belongs to the folding funnel. In addition, IS8, with $Q = 0.576$, belongs to the misfolding funnel.

At $h = 0.475$, the configurations of the significant IS change from those found in bulk and at lower h : the orientation of the

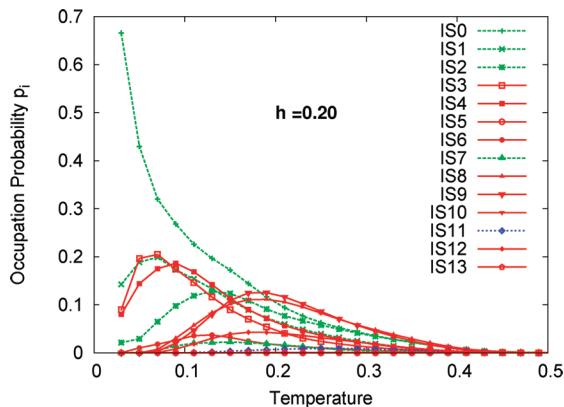


Figure 7. IS occupation probabilities inside a hydrophobic cage, $h = 0.20$. The protein–wall interaction splits the low-lying IS into multiple states with different IS energies but the same order parameter, Q . Misfolding states (red) have higher occupancy than the ground state, indicating that misfolding is preferred.

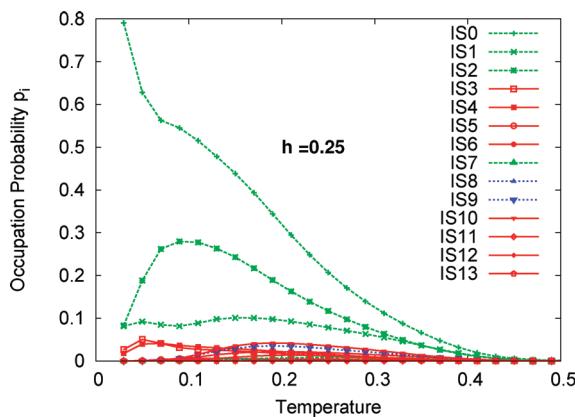


Figure 8. IS occupation probabilities inside a hydrophobic cage, $h = 0.25$. The two major routes to the misfolding funnel, via IS3, IS4 and IS9, IS10, are suppressed at this optimal hydrophobicity. The occupation of the ground state is higher than that of all the excited states at all temperatures. Folding is strongly favored at this optimal hydrophobicity.

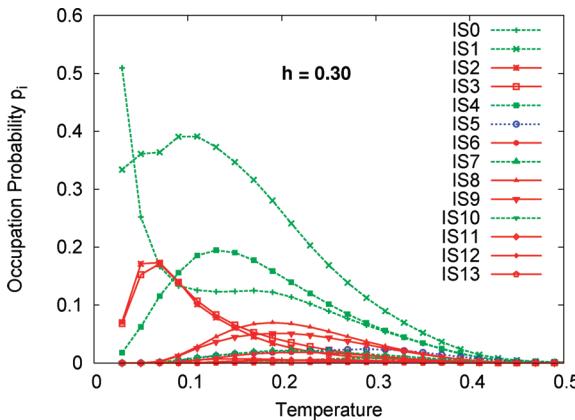


Figure 9. IS occupation probabilities inside a hydrophobic cage, $h = 0.30$. The occupations of higher excited states belonging to the folding funnel, IS1 and IS4, are higher than that of the ground state. The occupations of the misfolding states, p_2 , p_3 , p_9 , and p_{10} , have again become significant as the hydrophobicity exceeds $h = 0.25$. Some non-native states with low IS energy (blue) have also become populated.

β -hairpin strand in IS0 changes slightly; IS1 with $Q = 0.45$ is similar to IS0 in bulk; IS2 with $Q = 0.676$ is actually a highly misfolded structure, with the orientation of the β -hairpin strand opposite to that of IS0 in bulk. Again, it is difficult to say, from either Q or R_g , if the IS belongs to the folding or misfolding

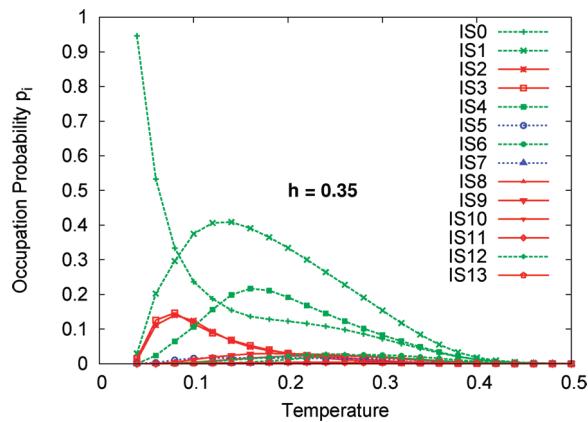


Figure 10. IS occupation probabilities inside a hydrophobic cage, $h = 0.35$. As with $h = 0.30$, some excited states from the folding funnel have higher occupation than the ground state.

funnel without examining the structure. A misfolding interval of $T = 0.21$ – 0.38 again appears at $h = 0.475$.

At $h = 0.525$ the interaction with the wall so stabilizes the misfolding states that the protein takes only unfolded and misfolded conformations; no correctly folded structure is reached, and STMD does not converge.

Because the splitting of the IS with increasing h makes the p_i analysis complicated for intermediate to high hydrophobicity values, the summed occupation probabilities of all states belonging to the folding and misfolding funnels in Figure 4 are the best IS-based indicators of foldability. It is very clear that $h = 0.25$ is optimal for folding.

Exploiting the Raw STMD Trajectory. Equilibrium averages may be rigorously calculated with STMD but only after reweighting the configurations. STMD dynamics is accelerated, and the times are not the true times of canonical MD. Nevertheless, we now show that qualitatively correct conclusions about the effects of the cage–protein interactions on thermodynamics and kinetics may be drawn from the raw trajectory. The time in this section refers to STMD time, and we refer to Figure 11.

Simple confinement makes the misfolded state more stable than the native state, because the misfolded structure is more compact than the native structure. Corresponding to those thermodynamic observations, going from bulk to hydrophilic confinement, the STMD trajectory exhibits an increase in unfolded \rightarrow misfolded transitions relative to the number of unfolded \rightarrow folded transitions.

Thermodynamics indicated that a slightly hydrophobic cavity, with $h = 0.20$, also favors misfolding as compared to bulk, and the trajectory shows that the protein makes many visits to the misfolding funnel before reaching the native state, which it quickly exits.

The trajectory clearly shows that $h = 0.25$ is the optimal hydrophobicity. Visits to the misfolding funnel become far less frequent. Most of the time, the unfolded protein follows a direct path to the native state, avoiding the misfolding funnel. A similar, but weaker, pattern favorable to folding is seen at $h = 0.30$.

Even higher hydrophobicity, with $h > 0.30$, strengthens the interactions between the cavity wall and the protein. The non-native ensemble is stabilized by its lower energy. The protein spends more time in the unfolded ensemble. As a result, at $h = 0.35$ and 0.40 , the occupation probabilities of IS with low order parameter values start increasing. At $h = 0.475$, the interaction is so strong that the protein remains in the unfolded and

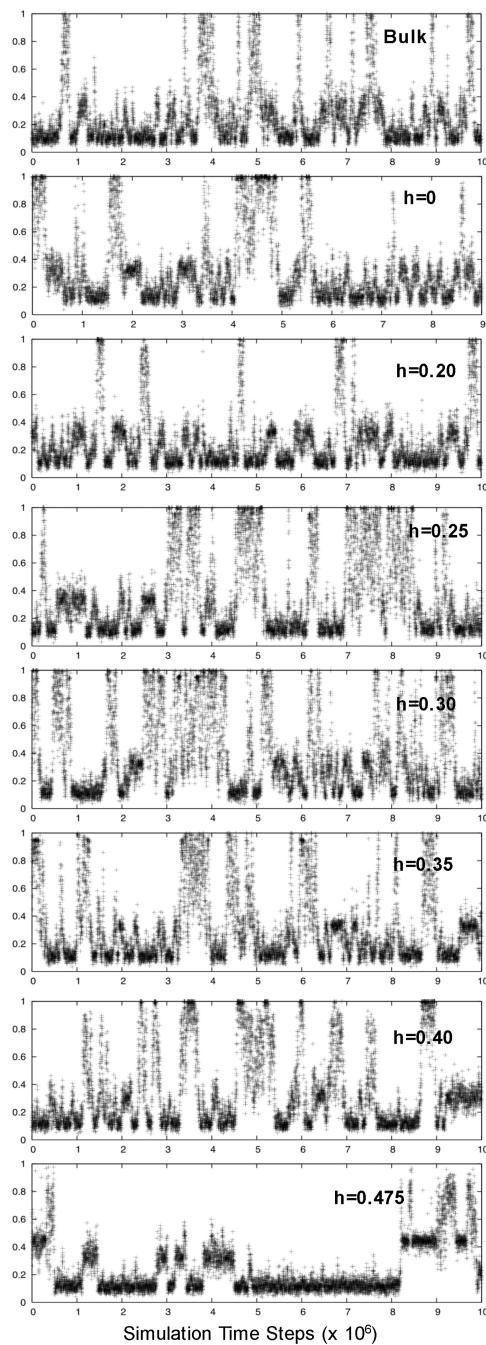


Figure 11. Order parameter, Q , along STMD trajectories at different hydrophobicities. High Q represents the correctly folded conformations; $Q \approx 0.3$ corresponds to the conformations in the misfolding funnel.

misfolded ensemble for very long times. Thus, at these high hydrophobicities, the native state is less stable.

Now consider two probability distributions calculated from the raw STMD trajectory: those of the root-mean-square displacement (rmsd) from the native state and of the radius of gyration, R_g , of the minimized configurations.

Distribution of rmsd. The probability distribution of the rmsd from the native structure is shown for various hydrophobicities in Figure 12. rmsd values of 1.0–1.5 correspond to the misfolded state, while values of 0–0.5 correspond to the native structure. The long tail on the bulk distribution function, corresponding to open-chain conformations, is missing under confinement.

In agreement with our prior findings, the strength of the misfolded peak relative to the folded peak increases with

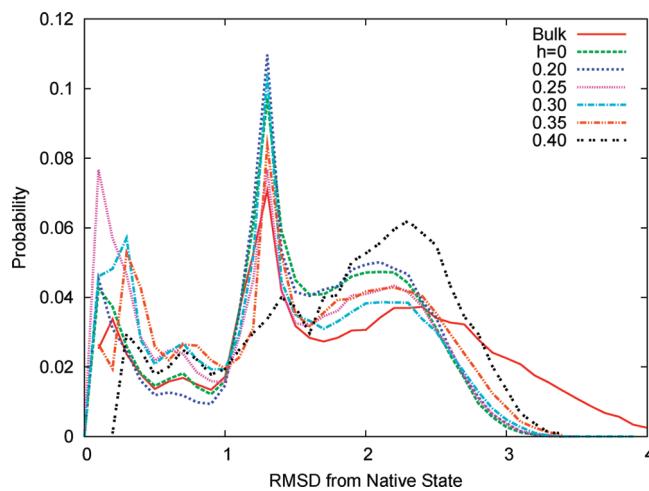


Figure 12. Probability distribution of the rmsd from the native state from unweighted STMD trajectories at various hydrophobicities. rmsd in the ranges 0–0.5 and 1.0–1.5 correspond to native and misfolded structures, respectively.

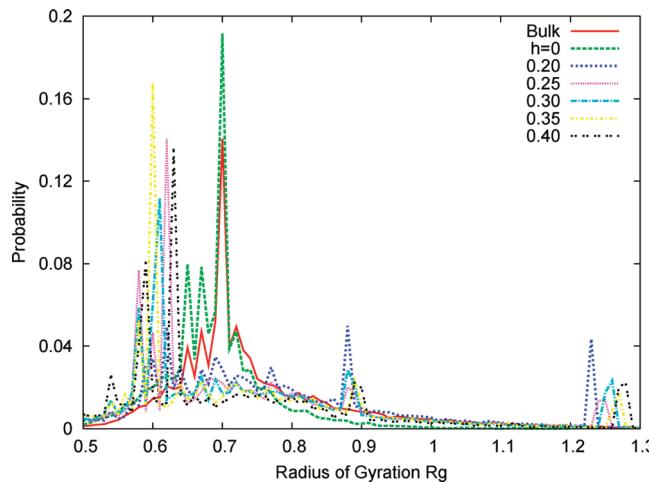


Figure 13. Probability distribution of R_g of the IS obtained by minimizing all the configurations from STMD trajectories at various hydrophobicities.

confinement and with increasing h up to 0.20 but drops sharply at $h = 0.25$. The primary effect of a further increase in h is to increase the population of unfolded configurations with rmsd in the 2–3 range, which is unfavorable to folding. Thus, it is seen how an optimum hydrophobicity emerges.

Distribution of R_g of Sampled Inherent Structures. Each configuration in the trajectory is mapped to its corresponding IS. The distribution of R_g calculated from these IS is shown in Figure 13. In bulk and in hydrophilic confinement, the peak at $R_g = 0.697$ corresponds to the folded ensemble while the peaks at $R_g = 0.64, 0.67$, and 0.72 arise from the misfolded states. Within the hydrophobic environment, the misfolded states are seen in the two distinct peaks at larger R_g of 0.9 and 1.2 – 1.3 , indicating that their structure has become more open and chain-like. Once again, the misfolded state probability is smallest for $h = 0.25$. The folded conformations, with R_g in the range 0.5 – 0.662 , become more compact with increasing h .

Why Are the Misfolded States Suppressed at $h = 0.25$? Here we sketch a first step toward a physical and structural explanation of the very strong suppression of the misfolded states at $h = 0.25$. The occupation probability of i th IS depends upon its energy, ε_i , and vibrational free energy, $f_{\text{vib},i}$

$$p_i(T) = e^{-\beta(\varepsilon_i + f_{\text{vib},i}(T))}/Z(\beta) \quad (4)$$

where $Z(\beta)$ is the configurational integral and the vibrational free energy is determined^{35–37} by the partial configurational integral over the basin of attraction. Referring the p_i to p_0 in the same environment

$$p_i/p_0 = e^{-\beta(\Delta\varepsilon_i + \Delta f_{\text{vib},i})} \quad (5)$$

where $\Delta\varepsilon_i = (\varepsilon_i - \varepsilon_0)$ and $\Delta f_{\text{vib},i} = (f_{\text{vib},i} - f_{\text{vib},0})$. Some $\Delta\varepsilon_i$ in various environments are shown in Figure 14. Taking the difference in relative occupation probabilities between two different environments

$$\ln\left(\frac{p'_i}{p'_0}\right) - \ln\left(\frac{p_i}{p_0}\right) = -\beta(\Delta\Delta\varepsilon_i + \Delta\Delta f_{\text{vib},i}) \quad (6)$$

where $\Delta\Delta\varepsilon_i = \Delta\varepsilon'_i - \Delta\varepsilon_i$, etc.

In bulk, overall translation and rotation are irrelevant. However, these coordinates become significant in the presence of the chaperonin cage. Corresponding to the division of the protein coordinates into internal, translational, and rotational

$$f_{\text{vib},i} = f_{\text{vib},i}^{\text{int}} + f_{\text{vib},i}^{t+r} \quad (7)$$

We now consider optimal $h = 0.25$ as the “initial” unprimed environment and $h = 0.20$ as the “final” primed environment and seek to understand a strong increase in the occupation of the misfolded states with $\ln(p'_i/p'_0) - \ln(p_i/p_0) > 0$. Simulation yields $\Delta\Delta\varepsilon_i \approx 0$, which also suggests $\Delta\Delta f_{\text{vib},i}^{\text{int}} \approx 0$. Changing the protein–wall interaction by a small $\Delta h = 0.05$ is more likely to influence $f_{\text{vib},i}^{t+r}$, which is an explicit protein–wall quantity, than the internal $f_{\text{vib},i}^{\text{int}}$.

Thus, we look to $f_{\text{vib},i}^{t+r}$, which has energetic and entropic contributions. We will not present a complete theory in this paper, but we will attempt to learn as much as possible about the effect of the chaperonin cavity with the translational, vibrational entropy

$$S_{\text{vib},i}^t = -k_B \int 4\pi r^2 dr w_i(r) \ln(w_i(r)) \quad (8)$$

where $w_i(r)$ is the probability distribution of the center of mass for configurations mapping to the i th IS.

Figure 15 shows $\Delta\Delta S_{\text{vib}}^t$ for the two most important misfolded states. While the data are noisy and more averaging would be desirable, its positive value proves that the entropy associated with the distribution of the protein center of mass in the cage favors misfolding away from $h = 0.25$ and, conversely, opposes misfolding at $h = 0.25$. Clearly, translational entropy is one of the mechanisms of the suppression of the misfolded states at $h = 0.25$.

The physical picture that emerges is as follows. The misfolded states are somewhat more compact than the folded states, so they are favored by translational entropy in the hydrophilic, $h = 0$, cavity, and energetic effects are irrelevant. As h is increased, two competing effects come into play: misfolded states, with exposed hydrophobic residues, develop a favorable energy of interaction with the wall and a corresponding unfavorable entropy of being preferentially near the wall. The entropic effect dominates at low h , leading to the strong

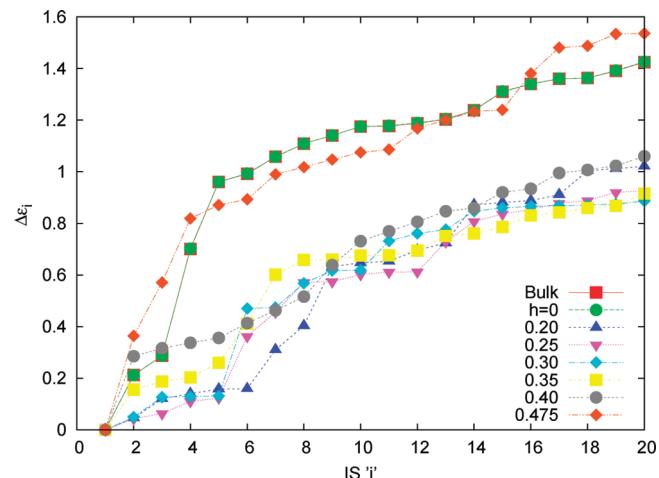


Figure 14. Difference of the IS energy from the ground-state energy, $\Delta\varepsilon_i$, for the first 20 IS in bulk, hydrophilic cage, and hydrophobic cages of varying h .

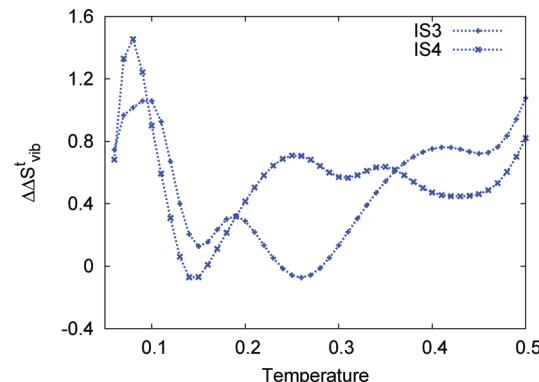


Figure 15. $\Delta\Delta S_{\text{vib}}^t$ for the misfolded states, IS3 and IS4, in chaperonin cage with $h = 0.20$ and $h' = 0.25$.

suppression of the misfolded states at $h = 0.25$, and then the energetic effect dominates, making misfolding more likely for $h > 0.25$.

In short, the optimal hydrophobicity is that which produces the maximum preference of the misfolded states for the region close to the cavity wall, lowering their translational entropy, without being large enough to substantially lower their energy.

Kinetics of Folding under Confinement. In previous sections, the phenomenon of an optimal hydrophobicity of $h = 0.25$ has been established with equilibrium methods. To investigate related behavior for kinetics, we run Langevin dynamics simulations at the bulk folding temperature, $T = 0.18$, under different environmental conditions. Folding is monitored via the decay of the fraction of unfolded trajectories. As expected, Figure 16 shows that at intermediate–long times folding is fastest at $h = 0.25$.

At very short time, the number of unfolded trajectories decays fastest in hydrophilic confinement, $h = 0$. However, this protein is designed to misfold and, under simple confinement, quickly collapses to the misfolded structure, which is more compact than the native structure. Thus, the fast decay does not truly represent folding and, in any case, does not persist to longer times more indicative of productive folding.

For a slightly hydrophobic cavity with $h = 0.20$, folding is even slower than that inside the hydrophilic cavity. At higher hydrophobicity values of $h > 0.25$, folding slows down because of the stronger wall–protein interactions. The slowest folding is for $h = 0.475$. This trend of folding kinetics is well described

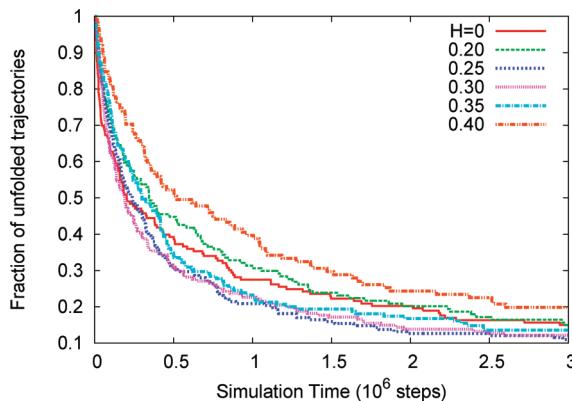


Figure 16. Fraction of unfolded trajectories vs time at $T = 0.18$. Fastest folding occurs at $h = 0.25$. Very high hydrophobicity lead to slow folding.

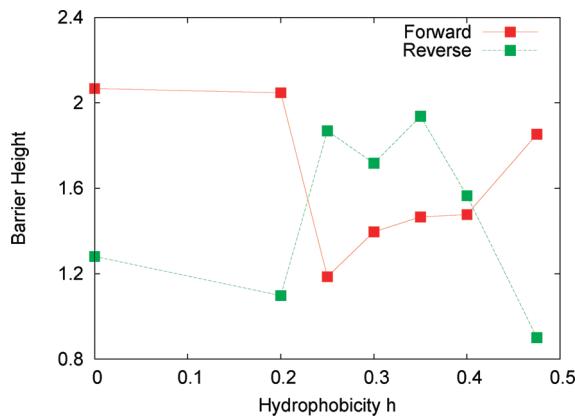


Figure 17. Free energy differences from unfolded ensemble to transition state (forward) and from folded ensemble to transition state (reverse) at different environmental conditions and $T = 0.18$, the folding temperature in bulk.

in terms of the barrier heights for folding and unfolding, along the one-dimensional free energy, $F(Q)$, at $T = 0.18$. Figure 17 shows the barrier heights to the forward (non-native \rightarrow native) and reverse (native \rightarrow non-native) directions. The barrier to folding is lowest at $h = 0.25$.

IV. Conclusions

Chaperonin acts^{5,6} by enclosing proteins which might misfold in a cage of varying hydrophobicity. Here we studied a coarse-grained model²⁴ due to the Shea group of a 27-mer with a significant misfolding funnel in bulk in a cage of beads with tunable bead–protein interactions, characterized by the hydrophobicity parameter, h . The Shea model and related ones,^{13,25} are known to exhibit an optimal h for productive folding, determined by kinetics and two-dimensional free energy diagrams.

Our primary contribution is to demonstrate that inherent structure (IS) analysis^{35–37} is a powerful tool to understand the effects of a changing environment on a small protein. In large part this work exploits our previous finding^{32–34} that below the collapse temperature the occupation probabilities of the low-lying IS of model proteins become finite, allowing a description of these continuum systems in terms of a small number of natural states.

Since the IS can be classified as belonging to the folding or misfolding funnels, their occupations provide an excellent window on foldability. We previously proposed that a tendency to misfold could be identified by the presence of a “misfolding

interval”, a temperature range in which the occupation of misfolding states exceeds that of the ground state.

The hydrophilic and slightly hydrophobic chaperonin cavities were found to possess misfolding intervals. Moving from $h = 0.20$ to $h = 0.25$, the optimal hydrophobicity, the misfolding interval disappears and the misfolding states are strikingly suppressed. At higher hydrophobicity, the misfolding states emerge again but the stronger interaction of the protein with the cavity wall both splits the bulk IS into daughter states and distorts their structure, so that it is difficult to discuss productive folding.

We attempted to explain the existence of the optimal hydrophobicity, presumably an essential element of chaperonin action. In the cavity, overall translation and rotation of the protein, which is irrelevant for thermodynamics in bulk, becomes significant. As h is increased from zero, the exposed hydrophobic residues of misfolded proteins are attracted to the wall, which is energetically favorable and entropically unfavorable. The entropic effect dominates at small h , and the energetic effect dominates at large h , leading to maximal suppression at intermediate h , which turns out to be $h = 0.25$ in this case.

Although the protein is only a coarse-grained 27-mer, the energy landscape, with two funnels, is complex enough to benefit from enhanced sampling, and we employed our STMD algorithm in calculating all equilibrium quantities. The characteristic collapse and folding temperatures, T_θ and T_f , were obtained as a function of h . STMD sampling is probably why some of our results differ from those of the Shea group.²⁴ The folding temperature may be identified with equal population of the low- Q and high- Q wells in the free energy or from the peak in the δQ fluctuations.

With unfolded, misfolded, and folded states, the “quasichemical” picture of this protein should be 3-state; it is not a simple “2-state folder”. It was explained how misfolded states with low energy but also low Q influenced the T dependence of the folding temperatures. As T is decreased below T_θ , misfolded \leftrightarrow folded transitions cause the Q fluctuations to keep growing and cause $T_f(Q) < T_f(\text{wells})$ at all h .

STMD yields equilibrium averages rigorously, but only after the configurations are reweighted,³⁴ and we have not yet discovered a transformation between accelerated STMD time and true time. Nevertheless, we demonstrated that qualitatively correct conclusions could be quickly drawn from the raw STMD trajectory. At the optimal $h = 0.25$, unfolded states quickly evolve into folded states, avoiding the misfolding funnel.

The utility of the IS “states” for thermodynamics suggests that kinetics be formulated as transitions among the IS, and this approach is under development. We observed, however, that some unfolded configurations belong to the basin of attraction to the ground state, so the possibility of intrabasin folding will also have to be included. In fact, the special unfolded configurations belonging to ISO do fold quickly. Thus, IS theory provides a natural realization of the “topomer search model”.⁴⁰

Acknowledgment. This work was supported by NSF grant CHE-0833605 and has benefitted from discussions with Dr. Jaegil Kim.

References and Notes

- (1) Hartl, F. U.; Hayer-Hartl, M. *Nat. Struct. Mol. Biol.* **2009**, *16*, 574.
- (2) Langer, T.; Pfeifer, G.; Martin, J.; Baumeister, W.; Hartl, F.-U. *EMBO J.* **1992**, *11*, 4757.
- (3) Weissmann, J. S.; Rye, H. S.; Fenton, W. A.; Beechem, J. M.; Horwich, A. L. *Cell* **1996**, *84*, 481.
- (4) Hayer-Hartl, M. K.; Weber, F.; Hartl, F. U. *EMBO J.* **1996**, *15*, 6111.

- (5) Brinker, A.; Pfeifer, G.; Kerner, M.; Naylor, D.; Hartl, F.-U.; Hayer-Hartl, M. *Cell* **2001**, *107*, 223.
- (6) Weber, F.; Keppel, F.; Georgopoulos, C.; Hayer-Hartl, M. K.; Hartl, F. U. *Nat. Struct. Biol.* **1998**, *5* (11), 977.
- (7) Horst, R.; Fenton, W. A.; Englander, S. W.; Wuthrich, K.; Horwich, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20788.
- (8) Apetrii, A. C.; Horwich, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17351.
- (9) Chapman, E.; Farr, G. W.; Fenton, W. A.; Johnson, S. M.; Horwich, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19205.
- (10) Lin, Z.; Madan, D.; Rye, H. S. *Nat. Struct. Mol. Biol.* **2008**, *15*, 203.
- (11) Tang, Y. C.; Chang, H. C.; Roeben, A.; Wischnewski, D.; Wischnewski, N.; Kerner, M. J.; Hartl, F. U.; Hayer-Hartl, M. *Cell* **2006**, *125*, 903.
- (12) Tod, M. J.; Lorimer, G. H.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4030.
- (13) Betancourt, M. R.; Thirumalai, D. *J. Mol. Biol.* **1999**, *287*, 627.
- (14) Thirumalai, D.; Lorimer, G. H. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 245.
- (15) Stan, G.; Thirumalai, D.; Lorimer, G. H.; Brooks, B. R. *Biophys. Chem.* **2003**, *100*, 453.
- (16) Klimov, D. K.; Newfield, D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8019.
- (17) Zhou, H.-X.; Dill, K. A. *Biochemistry* **2001**, *40*, 11289.
- (18) Takagi, F.; Koga, N.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11367.
- (19) Ping, G.; Yuan, J. M.; Vallieres, M.; Dong, H.; Sun, Z.; Wei, Y.; Li, F. Y.; Lin, S. H. *J. Chem. Phys.* **2003**, *118*, 8042.
- (20) Thirumalai, D.; Klimov, D. K.; Lorimer, G. H. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11195.
- (21) Freedal, M. D.; Sheeler, D. J.; Shea, J.-E. *J. Chem. Phys.* **2003**, *118*, 8106.
- (22) Baumketner, A.; Jewett, A.; Shea, J.-E. *J. Mol. Biol.* **2003**, *332*, 701.
- (23) Ellis, R. J. *Curr. Biol.* **2003**, *13*, R881.
- (24) Jewett, A. I.; Baumketner, A.; Shea, J.-E. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13192.
- (25) Lu, D.; Liu, Z.; Wu, J. *Biophys. J.* **2006**, *90*, 3224.
- (26) Ojeda, P.; Garcia, M. E.; Londono, A.; Chen, N. Y. *Biophys. J.* **2009**, *96*, 1076.
- (27) Lucent, D.; Vishal, V.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10430.
- (28) Agard, D. *Science* **1993**, *260*, 1903.
- (29) Zhou, H. X. *J. Mol. Recognit.* **2004**, *17*, 368.
- (30) Rathore, N.; Knotts IV, T. A.; de Pablo, J. J. *Biophys. J.* **2006**, *90*, 1767.
- (31) Mittal, J.; Best, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20233.
- (32) Kim, J.; Straub, J. E.; Keyes, T. *Phys. Rev. Lett.* **2006**, *97*, 050601.
- (33) Kim, J.; Keyes, T. *J. Phys. Chem. B* **2007**, *111*, 2647.
- (34) Kim, J.; Straub, J. E.; Keyes, T. *J. Chem. Phys.* **2007**, *126*, 135101.
- (35) Stillinger, F. H.; Weber, T. A. *Phys. Rev. A* **1983**, *28*, 2408.
- (36) Stillinger, F. H.; Weber, T. A. *Science* **1984**, *225*, 983.
- (37) Stillinger, F. H.; Weber, T. A. *J. Chem. Phys.* **1984**, *81*, 5095.
- (38) Honeycutt, J. D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3526.
- (39) Keyes, T.; Chowdhary, J. *Phys. Rev. E* **2001**, *64*, 032201. Keyes, T.; Chowdhary, J. *J. Phys. Chem. B* **2004**, *108*, 19786.
- (40) Makarov, D.; Plaxco, K. *Protein Sci.* **2002**, *12*, 17.

JP107257B