Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity

Article *in* Journal of Molecular Biology · May 1999

DOI: 10.1006/jmbi.1999.2591 · Source: PubMed

2 authors, including:

Marcos R Betancourt
City University of New York - BMCC
28 PUBLICATIONS **982** CITATIONS

SEE PROFILE

Article No. jmbi.1999.2591 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 287, 627–644

JMB

Exploring the Kinetic Requirements for Enhancement of Protein Folding Rates in the GroEL Cavity

Marcos R. Betancourt[1] and D. Thirumalai[1,2]*

[1]*Institute for Physical Science and Technology, University of Maryland, College Park MD 20742, USA*

[2]*Department of Chemistry and Biochemistry, University of Maryland, College Park MD 20742, USA*

**Corresponding author*

The chaperonin system, GroEL and GroES of *Escherichia coli* enable certain proteins to fold under conditions when spontaneous folding is prohibitively slow as to compete with other non-productive channels such as aggregation. We investigated the plausible mechanisms of GroEL-mediated folding using simple lattice models. In particular, we have investigated protein folding in a confined environment, such as those offered by the GroEL, to decipher whether rate and yield enhancement can occur when the substrate protein is allowed to fold within the cavity of the chaperonins. The GroEL cavity is modeled as a cubic box and a simple bead model is used to represent the substrate chain. We consider three distinct characteristic of the confining environment. First, the cavity is taken to be a passive Anfinsen cage in which the walls merely reduce the available conformation space. We find that at temperatures when the native conformation is stable, the folding rate is retarded in the Anfinsen cage. We then assumed that the interior of the wall is hydrophobic. In this case the folding times exhibit a complex behavior. When the strength of the interaction between the polypeptide chain and the cavity is too strong or too weak we find that the rates of folding are retarded compared to spontaneous folding. There is an optimum range of the interaction strength that enhances the rates. Thus, above this value there is an inverse correlation between the folding rates and the strength of the substrate-cavity interactions. The optimal hydrophobic walls essentially pull the kinetically trapped states which leads to a smoother the energy landscape. It is known that upon addition of ATP and GroES the interior cavity of GroEL offers a hydrophilic-like environment to the substrate protein. In order to mimic this within the context of the dynamic Anfinsen cage model, we allow for changes in the hydrophobicity of the walls of the cavity. The duration for which the walls remain hydrophobic during one cycle of ATP hydrolysis is allowed to vary. These calculations show that frequent cycling of the wall hydrophobicity can dramatically reduce the folding times and increase the yield as well under non-permissive conditions. Examination of the structures of the substrate proteins before and after the change in hydrophobicity indicates that there is global unfolding involved. In addition, it is found that a fraction of the molecules kinetically partition to the native state in accordabce with the iterative annealing mechanism. Thus, frequent ''unfoldase'' activity of chaperonins leading to global unfolding of the polypeptide chain results in enhancement of the folding rates and yield of the folded protein. We suggest that chaperonin efficiency can be greatly enhanced if the cycling time is reduced. The calculations are used to interpret a few experiments on chaperonin-mediated protein folding.

Keywords: protein folding; chaperonin; iterative annealing mechanism; kinetic partitioning; Anfinsen cage

Abbreviations used: DAC, dynamic Anfinsen cage; SR1, single-ring mutant; 3D, three-dimensional; MFPT, mean first passage time; MCS, Monte Carlo step; rb, rigid body.

E-mail address of the corresponding author: thirum@Glue.umd.edu

0022-2836/99/130627–18 $30.00/0

Introduction

It is believed that, at least in *Escherichia coli*, a significant fraction of proteins fold spontaneously to their native state (Lorimer, 1996) as outlined by Anfinsen (1973). This is remarkable because one might argue that the heavy traffic in cellular environment, due to the presence of various macromolecules, might lead to strong interactions between unfolded polypeptide chains with exposed hydrophobic residues. The external factors could, in principle, alter the folding kinetics sufficiently such that aggregation could interfere with monomeric folding. Nevertheless, it appears that many *in vivo* proteins are unable to reach the functional folded state spontaneously. It has been shown that molecular chaperones assist in the folding of such proteins by preventing aggregation.

In recent years a number of studies have focussed on the subclass of proteins referred to as chaperonins, which rescue misfolded substrate proteins and assist them in an ATP-dependent manner to reach the native conformation. The best studied of the chaperonin family is the GroE machine consisting of two proteins, GroEL and GroES. The crystal structures of unliganded GroEL (Braig *et al*., 1994), GroES (Hunt *et al*., 1996) and the complex GroEL/GroES/ADP (Xu *et al*., 1997) have provided considerable insight into the mechanisms of the chaperonin action. The crystal structure of GroEL (Braig *et al*., 1994) shows that it is a double-ring oligomer consisting of back-to-back seven-member rings. It has an overall cylindrical structure divided into two non-connected cavities in which the substrate protein can be sequestered. The co-chaperonin GroES (Hunt *et al*., 1996) is composed of seven subunits and serves as a dome on the cylindrical structure in the presence of nucleotides. A remarkable feature, that clearly has mechanistic implications, is that upon binding of GroES and ATP the volume of the cavity can nearly double. The large conformational fluctuations in this machinery can accommodate substrate proteins of varying sizes, which is one of the reasons the GroEL/GroES system is responsible for the rescue of many unrelated families of proteins.

The mechanistic actions of the GroE machine in facilitating protein folding are beginning to emerge thanks to considerable experimental work (Viitanen, 1992; Fenton & Horwich, 1997; Netzer & Hartl, 1998; Richardson *et al*., 1998) and some progress in theoretical understanding (Thirumalai, 1994; Gulukota & Wolynes, 1994; Chan & Dill, 1996; Sfatos *et al*., 1996, Cui *et al*., 1996). The current view of the cycle, representing the concerted action of GroEL/GroES, is schematically sketched in Figure 1. The process starts with GroEL in a configuration with a hydrophobic interior near the entrance of its cavity (apical domain) which binds the misfolded protein. The favorable interaction of the substrate protein with GroEL occurs because of the attraction between the exposed hydrophobic residues of the misfolded protein and the residues comprising the apical domain. Seven ATP molecules and GroES bind to GroEL, producing a significant conformational change in GroEL that buries its hydrophobic interior and exposes a more hydrophilic surface to the substrate protein. As the ATP hydrolyzes in a quantified manner (Todd *et al*., 1994) into ADP, the substrate protein may continue to fold in the interior of GroEL-GroES. Once the hydrolysis of ATP in the *cis*-ring is complete, the substrate protein is prepared for release (Rye *et al*., 1997). This occurs when ATP is bound to the *trans*-ring which results in the release of GroES, the substrate protein and the seven ADP molecules from the *cis*-ring. This constitutes one ATP cycle. For simplicity, we have not considered the possible formation of the symmetric (football) complex.

Although the underpinnings of the overall scheme shown in Figure 1 are based on several sound experiments (Weissman *et al*., 1994; Todd *et al*., 1994) and theoretical arguments (Todd *et al*., 1996), several outstanding questions remain. For example, the number of binding/release cycles needed to fold the substrate protein is known only in a few cases. The processes by which a chaperonin could accelerate the folding kinetics are not well understood. Clearly one cannot expect universal answers to such questions which could be applicable to all substrate proteins under a range of conditions. General arguments based on the energy landscape perspective of folding suggest that one should expect diverse mechanisms for chaperonin assisted folding (Todd *et al*., 1996). The precise mechanism will depend on the size of the protein, external condition, strength of substrate-chaperonin interactions, etc. Some suggestions concerning the alterations in the energy landscape of the substrate protein due to interactions with the cavity have been made using simple models. It has been proposed that the chaperonins unfold misfolded proteins giving them another chance to fold (Jackson *et al*., 1993). This actually can take place in several absorption cycles by the chaperonins in a process similar to iterative annealing (Todd *et al*., 1996). Using lattice models, it has been also asserted that a hydrophobic environment surrounding a protein reduces the energy barriers along the folding pathways allowing a faster folding transition (Chan & Dill, 1996).

Several aspects of the iterative annealing mechanism sketched in Figure 1 are to a large extent accepted, but many issues remain. One of the most critical is: what is the extent of folding of the substrate protein while it is sequestered underneath GroES in the GroEL cavity? (steps (2) and (3) in Figure 1). There are contradictory answers to this in the literature (Weissman *et al*., 1994; Todd *et al*., 1994; Hartl, 1996). Simple arguments based on size distribution of proteins alone suggest that pathways followed by substrate proteins can be diverse. Thus, even though the GroE machinery acts as a ''one-size-fits-all'' type system, there can be many mechanisms in the rescue of proteins

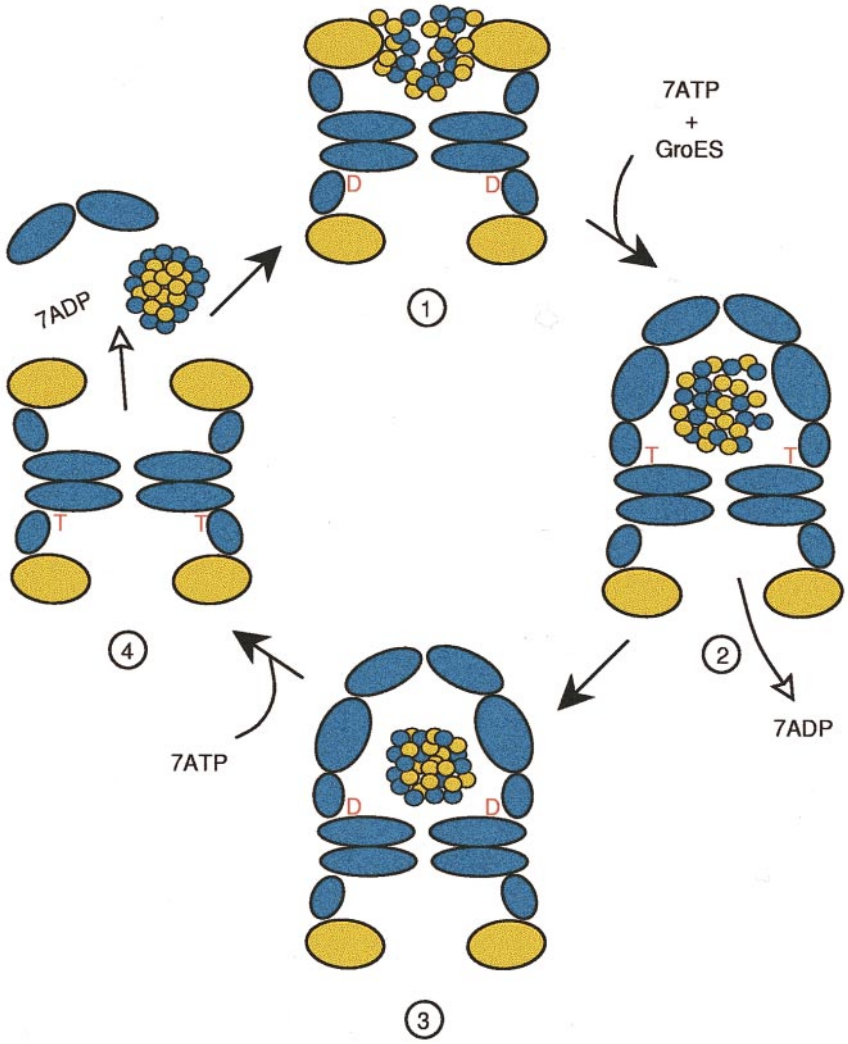

Figure 1. Schematic representation of one cycle in the GroEL-GroES mediated folding of a protein. The substrate protein is encapsulated in the cavity of GroEL (step (1)). This is followed by the addition of ATP and GroES that leads to nearly doubling of the volume of the interior of the GroEL cavity. The dynamic fluctuations caused by the increase in volume may enable the sequestered protein to fold while still being associated with GroEL. The hydrolysis of ATP in the *cis*-ring takes place in a quantified fashion (step (3)). This is followed by the binding of ATP to the *trans*-ring which primes the release of GroES and the substrate protein from the *cis* side. For simplicity we have only shown the asymmetric complex. The role of symmetric (football) complex is discussed by Schmidt *et al*. (1994).

depending on their size and intrinsic kinetic constraints. Two extreme models can be envisaged. (1) The folding is complete while the protein is still associated with the chaperonin in the central GroEL cavity. This is the Anfinsen cage model (Ellis, 1994). (2) The encapsulated substrate protein is unfolded by GroE machinery and released into solution (Weissman *et al*., 1994), where it undergoes kinetic partitioning (Guo & Thirumalai, 1995) with a fraction reaching the native state while the remaining fraction gets trapped in a manifold of misfolded states. This cycle of binding and release is repeated several times until sufficient yield of the folded protein is obtained. It should be emphasized that even the proponents of the Anfinsen cage model suggest that there is a dynamic interaction between the GroEL and the protein that leads, at least, to partial unfolding (Hartl, 1996), an event that is necessary for transition from misfolded states to the native conformation (Camacho & Thirumalai, 1993; Chan & Dill, 1994).

Our objective is to explore the various limits of the dynamic Anfinsen cage (DAC) model using simple lattice models. This work is partially inspired by the experiments by Weissman *et al*. (1996), who showed that, when the release of GroES is blocked by non-hydrolizable ATP analog or a designed single-ring mutant SR1, folding of the sequestered Rhodanase goes to completion. This suggests that for some substrate proteins it is possible to reach the native conformation while still being associated with GroEL. Thus, it is worth investigating the parameters for efficient refolding in a confined space. In the spirit of minimal model representations of proteins we investigated the possibility of folding in a DAC for a variety of substrate–wall interactions. The model protein consists of a three-dimensional (3D)-lattice chain of single bead residues with contact interactions simulating the non-bonded energies between the amino acid residues. The chaperonin is modeled as a cubic box with a length L so that the interior volume is L^3. The interior walls of the cavity interact with the protein through a contact potential just like the ones between the residues. Although this is an extremely simplified model of a chaperonin, it allows one to extensively study the general kinetic and thermodynamic effects of confined folding and the effects of a surrounding environment with a variable degree of hydrophobicity.

The simulations are performed for various conditions and chaperonin properties. First, the effect of the chaperonin size on the folding time is investigated at various temperatures when the cavity walls present a hydrophilic (neutral) environment to the substrate protein. Following that, the chaperonin interior wall hydrophobicity is gradually increased and its effects on folding are examined. The results show that a confined environment with a moderate degree of hydrophobicity can actually increase the yield and decrease the folding time over a range of temperatures. Finally, the chaperonin wall hydrophobicity is allowed to oscillate periodically in order to simulate the cycles of ATP binding and hydrolysis and the possibility of several protein encapsulation/release events by the chaperonin.

Results

Spontaneous folding

In order to assess the effects of folding in the DAC, it is useful to compare several thermodynamic and kinetic folding properties in the absence of confinement. We designed two sequences, a 16-mer which is kinetically a poor folder and 25-men which folds relatively rapidly. The nationale for sequence selection is given in Methods section. The results of the calculations of the thermodynamic properties for the 16-mer and the 24-mer are shown in Figure 2. The plots show the mean folding time (τ_F), the native state probability and $1-\chi$, where χ is the average overlap function, as function of temperature. The overlap function is defined as (Camacho & Thirumalai, 1993):

$$\chi = 1 - \frac{1}{(N-1)(N-2)} \sum_{i \neq j, j \pm 1} \delta(r_{ij} - r_{ij}^{o}) \qquad (1)$$

where r_{ij} is the distance between residues i and j and r_{ij}^{o} correspond to the native conformation. For both sequences the mean folding time reaches a minimum at a temperature close to the collapse temperature T_θ, which corresponds to the maximum of the specific heat (Camacho & Thirumalai, 1993). They both have similar collapse temperatures with values of $T_\theta \simeq 0.22$ for the 16-mer and $T_\theta \simeq 0.24$ for the 24-mer. At the collapse temperature, the 16-mer folds about an order of magnitude faster than the 24-mer, measured as arbitrary units. However, the 24-mer is a faster folder under equivalent stability conditions of the native state. Equivalent stability conditions can be established by measuring the folding time at the folding temperature T_f for each chain. If T_f is defined as the temperature for which probability of being in the native state (single microstate) is 0.5, then T_f is 0.18 for the 24-mer and only 0.053 for the 16-mer. At their respective folding transition temperatures the 16-mer folds 400 times slower, i.e. $\tau_f = 5.6 \times 10^7$ mcs for the 16-mer and $\tau_f = 1.4 \times 10^5$ mcs for the 24-mer. It has been shown (Camacho & Thirumalai, 1993; Klimov & Thirumalai, 1996) that an indication of the kinetic accessibility of the native state is given by the thermodynamic quantity $\sigma = 1 - T_f/T_\theta$, which is close to zero for rapidly folding sequences. For the 24-mer, $\sigma = 0.25$, and for the 16-mer, $\sigma = 0.71$. Thus, dramatic difference in the folding times between the 16-mer and the 24-mer is expected. Alternatively, if T_f is defined as a the temperature for which the fluctuations of the overlap function are maximum then $\sigma = 0.0$ for the 24-mer and $\sigma = 0.28$ for the 16-mer. Either way, it is clear that the selected 16-mer is a worse folder than the 24-mer.

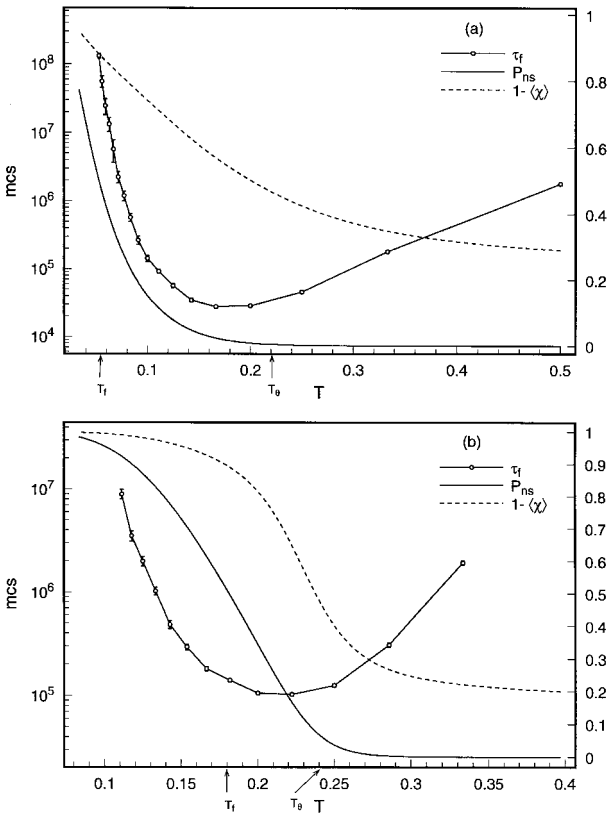

Figure 2. The stability of the native state and the mean folding time *versus* temperature for the (a) 16-mer and (b) 24-mer in the absence of confinement. The native stability is measured by the native state probability (continuous line) and by $1-\chi$, where χ the average overlap function $1-\langle\chi\rangle$ (broken line). The folding time-scale is on the left axes while the native state stability scale in on the right. Each folding time data point was calculated from an average of 400 different trajectories.

The stability of the native state for both sequences is marginal around T_f. If higher stability values are required, the temperature must be lowered resulting in considerably large values of τ_f even for the 24-mer. For example, in order to obtain a native state stability of 85% with the 24-mer, the temperature must be $T = 0.13$ which will require a folding time of approximately 10^6 mcs, or one order of magnitude higher than the

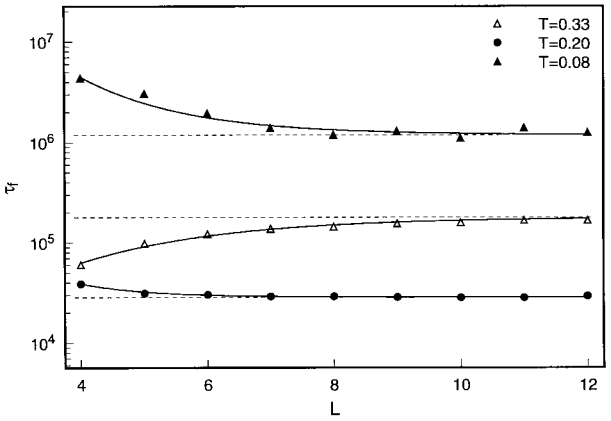

Figure 3. Effect of a neutral confining cavity on the folding time for the 16-mer. L is the available space in each direction inside the cavity so that the volume of the box is L^3. The minimum size of the cavity in which the sequence can fold is $L = 4$. Three temperatures are used, $T = 0.08$, 0.20, 0.33, corresponding to below, near and above the collapse temperature, $T_\theta = 0.22$. The broken lines represent the folding time for spontaneous folding at the three temperatures. For each data point, 1600 trajectories where used. The error bars are as large as the symbols.

one at T_f. For the 16-mer, the time required to obtain a significant native state yield is much larger. It is of interest to investigate whether the presence of a chaperonin could accelerate the folding rates under high stability conditions even for a good sequence. Thus, with the choice of these two sequences, the effect of confinement and other characteristics of chaperonins under permissive and non-permissive conditions can be investigated.

Folding in a neutral confinement: passive Anfinsen cage

Here, we study the effects of confinement on the folding of a protein in a neutral cavity, i.e. one in which the interaction between the wall and the residues is a hard core potential. The calculations are shown for the 16-mer only. Similar qualitative results are obtained for the 24-mer. This mode, in which the chaperonin is a static Anfinsen cage, was originally thought to be sufficient to explain assisted folding (Ellis & Hartl, 1996). The kinetic effects of confinement are shown in Figure 4 for the 16-mer. Here, we have plotted the folding time as a function of the cavity size L for three different temperatures: above, at and below the collapse transition temperature, T_θ. The simulations are carried out by selecting an ensemble of conformations at random and letting them collapse until they are sufficiently small to fit into the cavity. The folding time is measured from this point until the native state is found.

At a temperature ($T = 0.33$) higher than the collapse temperature ($T_\theta = 0.20$), a neutral confinement reduces the folding time (up to a factor of 2) as the cavity size is reduced. When a protein is confined, the number of conformations is reduced to the ones that fit in the cavity. One might argue that the reduction in the conformational space should speed up folding. As a result, we expect that at high temperatures where inter-conversion between the accessible conformations should be facile, confinement should increase the rate of folding. At ''high'' temperatures the polypeptide chain can also undergo large conformational fluctuations. This, when combined with the reduction in the conformations due to confinement, explains the decrease in folding time (for $L \lesssim 6$, see Figure 3). Such a decrease in τ_f is observed until $L \approx 4$. Note that for $L = 3$ the folding time becomes infinity because the native state, which spans the $3 \times 3 \times 3$ volume, becomes inaccessible from most conformations.

Near the collapse temperature ($T = 0.20$), the behavior is similar, but at these temperatures the folding time is not significantly affected by the chaperonin. Because the collapsed conformations are more stable at this temperature, the benefits of reducing the conformation space by confinement are less important. This temperature becomes the turning point in the ability to decrease the folding time by confinement alone. Below the collapse temperature ($T = 0.08$), at which the probability of being in the native state is 40 %, the folding time strictly increases as the cavity size is reduced. In this case the protein becomes increasingly trapped in the low energy minima, and confinement further prevents the necessary conformational fluctuations required to make transitions to the native state. The increase in folding time at this temperature (compared to the case when $L = \infty$) becomes appreciable for $L \approx 6$ and is only four times larger

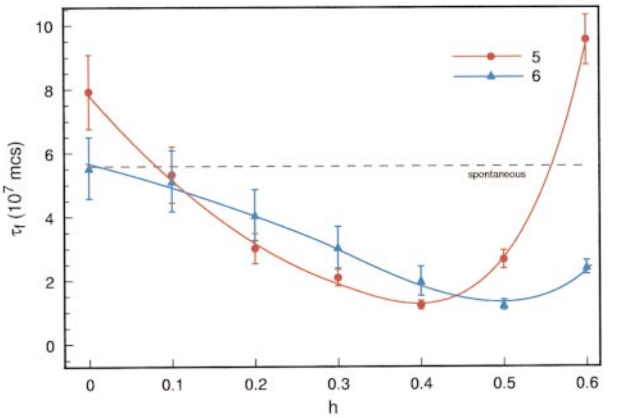

Figure 4. The folding time as a function of the strength of the hydrophobic interaction between the walls of the cavity and the sequestered polypeptide chain. The circles correspond to $L = 5$ while the triangles are for $L = 6$. In both the cases the temperature is 0.053. For comparison the spontaneous folding time is shown as broken lines. This Figure shows that for a given cavity size there is an optimum value of the hydrophobic interaction that produces the maximum in the folding strength.

for $L = 4$. Therefore, the effects of the chaperonin on the kinetics can be assumed to be small for $L \geqslant 4$ under stable conditions. For the 24-mer, the increase in folding time at low temperatures and $L = 4$ is much larger.

These results suggest that if the GroEL cavity merely acts as a static Anfinsen cage, as was originally thought (Ellis, 1994), then one cannot expect significant enhancement in the folding rates found in experiments if aggregation effects are excluded (Federov & Baldwin, 1997; Nieba-Axmann *et al.*, 1997). On the contrary, our calculations show that at low temperatures ($T < T_f$) when the native conformation is stable, the folding time increases compared to the time needed for spontaneous folding. Since in many experiments, especially those performed under non-permissive conditions, rate enhancement is observed, it follows that the GroEL/GroES apparatus has a dynamic role in its interaction with the substrate proteins. In what follows we show that this is indeed the case.

Folding in a hydrophobic cavity: folding rate is a maximum for optimal strength of substrate-wall interactions

It is known that the interior walls of a chaperonin offer a variable degree of hydrophobicity to the substrate protein depending on the presence of ATP and the binding of GroES. Thus, it is instructive to study the effect of folding in a hydrophobic cavity where the strength of the hydrophobic interaction can be tuned. In the following simulations, the wall hydrophobicity strength h (see Eq. (9) in Methods) is gradually increased from zero and its effects on folding are obtained. Once again, we only show the results for the slow folding 16-mer.

In Figure 5, we show the mean folding time as function of h for two cavity sizes at the folding temperature, $T_f = 0.053$. For comparison, the mean folding time for spontaneous folding is also shown. The plot clearly shows that τ_f decreases somewhat below its spontaneous value as the hydrophobicity is moderately increased. The folding time reaches a minimum at an optimum value of h, which for the cases studied is $h = 0.4$ for $L = 5$ and $h = 0.5$ for $L = 6$. As h is increased further, τ_f increases above its spontaneous value. The increase in folding times has also been noted in an earlier study (Chan & Dill, 1996). These results imply that for optimal values, the wall hydrophobicity can lower the barriers between local energy minima by decreasing their energies. Our results are consistent with earlier arguments that suggest that only for optimal interactions between the substrate protein and GroEL the folding state is obtained (Thirumalai, 1994; Orland & Thirumalai, 1997). In addition, experiments also reveal that the stronger the effective substrate-GroEL interaction, the slower the refolding rate is (Itzhaki *et al.*, 1995). In general, if the hydrophobic interaction between the protein and the wall is too strong then the substrate would be strongly pinned to the walls of the cavity. As a result, the rearrangements necessary to reach the native conformation, i.e. global unfolding of the polypeptide chain, become difficult and leads to an increase in folding time. In the limit of small h the interactions with the walls are negligible, and in this case the folding should be similar to that seen in a neutral Anfinsen cage. For $h \to 0$ there is no considerable reduction of the energy barrier arising from unfolding, and as a result the folding time is unaffected or increases moderately compared to spontaneous folding. These arguments are reflected in the results presented in Figure 5.

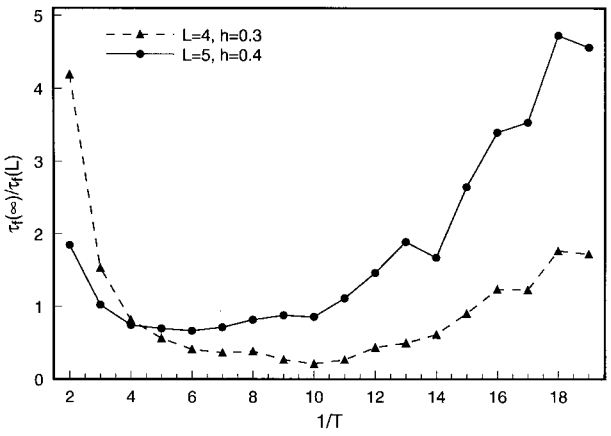

Figure 5. The ratio of spontaneous folding time to the folding time in a hydrophobic cavity as a function of the inverse temperature for the 16-mer. The circles are for $L = 5$ and $h = 0.4$ whereas the triangles represent the case with $L = 4$ and $h = 0.3$. The folding times in the hydrophobic cavity are less than the spontaneous folding times, both at high and low temperatures.

To further explore the effects of folding in the presence of hydrophobic walls we calculated the time dependence of the population of the native state, $f_N(t)$. We find that, in general, the pool of molecules that reach the native conformation in a moderate time scale ($\lesssim 10^8$ mcs) is not aided by being confined in a hydrophobic cavity. However, the hydrophobic walls enhance the efficiency of slow folding molecules. Thus, we can suggest that whenever slow folding events dominate, i.e. the partition factor ϕ is small (Todd *et al.*, 1996), the rate of folding can be enhanced in the presence of moderate hydrophobic walls. We expect the partition factor to be small under non-permissive conditions.

Temperature dependence of folding times in a hydrophobic cavity

The preceding results suggest that the efficiency gain by folding in a hydrophobic confinement should increase as the temperature is lowered. We plot in Figure 5 the ratio of spontaneous to confined mean folding time as a function of the inverse temperature for several conditions. We have plotted $1/T$ in the abscissa to emphasize the effect at low temperatures. Three regimes can be

observed from this plot. At high temperatures, τ_f is reduced in the hydrophobic confinement, consistent with the previous results of folding in a neutral ($h = 0$) cavity. For temperatures around the collapse temperature, τ_f is larger in the hydrophobic confinement. In this regime, folding is dominated by fast folding events and the hydrophobic cavity is not as effective in speeding up folding. At low enough temperatures, τ_f is once again smaller in the hydrophobic cavity and the gain in efficiency seems to improve as the temperature is further decreased. These results support the previous observation that folding in a hydrophobic confinement assists the slow folding events, which become more significant at low temperatures.

Confining a substrate protein in a hydrophobic environment alters its thermodynamics in a temperature-dependent manner. In particular, it is useful to compare the probability of being in the native state ($P_{\rm ns}$) as a function of temperature for the 16-mer in three situations: the bulk, a neutral cavity and a hydrophobic cavity. In a neutral cavity, the native state stability is higher than the bulk stability, particularly at higher temperatures. This is a result of the conformational entropy loss due to confinement. When the cavity walls are hydrophobic ($h = 0.4$ at $L = 5$ in this case) the native state is destabilized above $T \simeq 0.37$. $P_{\rm ns}$ increases as the temperature decreases until it reaches a maximum near a temperature of $T = 0.13$. Below this temperature, $P_{\rm ns}$ drops rapidly indicating that some non-native conformations have replaced the native conformation as the most stable state. These non-native conformations expose hydrophobic residues that, when in contact with the wall, lower the overall energy below that of the native conformation (Minton, 1995). However, the loss of identity of the native state as the lowest energy state is not required for a reduction in the folding time. The 24-mer sequence, for example, maintains its native state identity for all values of h and T, and also shows a reduction in folding time. What seems to be required is the destabilization of the native state, as τ_f drops below its spontaneous value near the temperature for which the stability rapidly decreases ($T = 0.13$). In general, the wall hydrophobicity interactions destabilize the native state by pulling it apart and smoothes the energy landscape roughness with an overall effect similar to that of increasing the temperature. Stabilization of the native state is expected only after the protein has been released from the chaperonin.

The analysis of the thermodynamics and kinetics in a confined hydrophobic cavity suggest that whenever the polypeptide chain is stuck in one of the low energy non-native conformation with exposed hydrophobic residues, the walls can pull the protein apart. This allows for a partial unraveling or unfolding of the chain that in effect reduces the barrier to the native conformation. This ''unfoldase'' activity observed here is associated with the GroEL cavity as well. Our results also show that if the wall-pinning energy is sufficiently large then the substrate/GroEL complex would be very stable. Under these conditions we expect a decrease in folding rate as compared to folding under permissive conditions. This is in accordance with earlier studies by Chan & Dill (1996).

Cycling the hydrophobicity of chaperonins: repeated global unfolding facilitates folding

In the previous section we examined the effects of protein folding in a confined cavity with a fixed degree of hydrophobicity. However, the environment presented to a substrate protein by a chaperonin is one in which the hydrophobicity changes from high to low values during an ATP cycle (Fenton & Horwich, 1997). In other words, during one cycle the cavity changes from being able to bind the substrate protein to a situation when the binding is not favored. In some cases one requires multiple rounds of binding and release until sufficient yield of the folded protein is obtained (Weissman *et al.*, 1994; Todd *et al.*, 1994). The effect of changing hydrophobicity can be mimicked by a simplified model in which the hydrophobicity of the confining cavity varies during the cycle time t_c. We divide the time period t_c into two subintervals. During one subinterval, t_p, the walls are hydrophilic (P) and for the remainder of the cycle ($t_c - t_p$) the cavity is hydrophobic (H). We examine different ratios of the subintervals:

$$R_p = t_p/(t_c - t_p) \tag{2}$$

The purpose of this model is to investigate the effects on protein folding of confinement with a varying hydrophobicity, ignoring other effects such as GroES binding and unbinding, and the GroEL volume changes upon binding of ATP-GroES and ejection of the substrate-protein upon ATP hydrolysis in the *trans*-ring. Some of these effects will be studied using a more elaborate model in a future publication.

The effect of hydrophobic cycling on the kinetics can be seen by calculating τ_f as a function of the cycling time t_c. The results are shown in Figure 6 for the 16-mer. For this calculation, the chaperonin cavity size is $L = 5$ and the hydrophobic interaction strength between the wall and the substrate protein is varied between $h = 0.0$ and $h = 1.0$ at the folding temperature ($T_f = 0.053$). These are the extreme hydrophobic changes where the character of the chaperonin interior walls varies between being fully hydrophilic to being fully hydrophobic. The ratio of time intervals of each hydrophobic cycle, namely, R_p, is set to three different values; 1, 9 and $\frac{1}{9}$. We have chosen to vary R_p because it is not known experimentally the duration over which the GroEL remains in an off or on state. It can be seen in Figure 6 that for all three cases, the folding time shows a remarkable decrease as the cycling time is decreased. When R_p is either 1 or $\frac{1}{9}$, the fold-

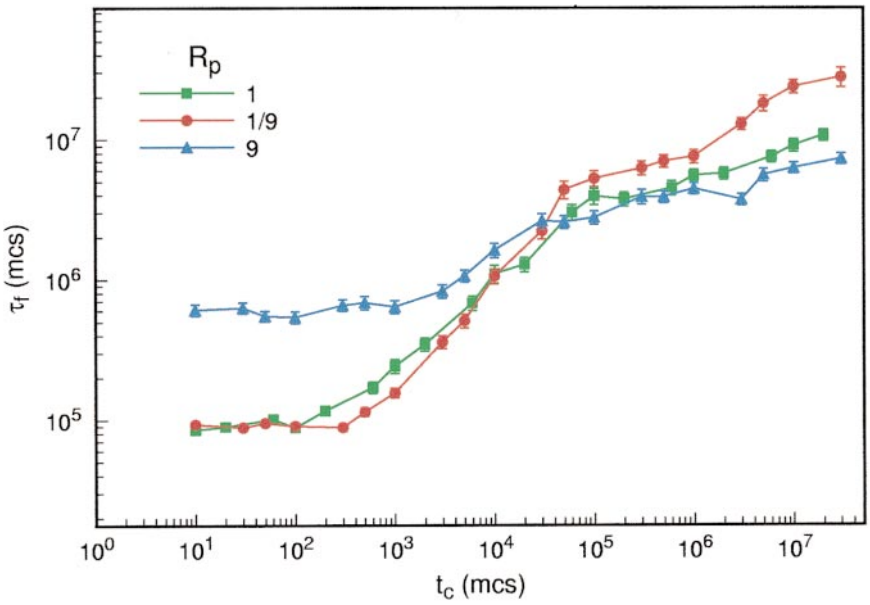

Figure 6. Dependence of the folding time τ_f for the 16-mer on the hydrophobic cycling time t_c at three R_p time interval ratios (see equation (2)). The chaperonin cavity size is $L = 5$ and its hydrophobicity varies between $h = 0.0$ and $h = 1.0$. For comparison, the spontaneous folding time is $\tau_f = 5.6 \times 10^7$ mcs. This Figure clearly shows that repeated cycling, which leads to multiple global unfolding, leads to dramatic decrease in the folding times.

ing time decreases by nearly three orders of magnitude at the smallest t_c, as compared to the spontaneous folding time ($\tau_f = 5.6 \times 10^7$ mcs). When R_p is 9, τ_f also decreases but only by two orders of magnitude. These results clearly demonstrate that there is a considerable rate enhancement at high hydrophobic cycling rates. Repeated changes from hydrophobic to hydrophilic character allow for considerable fluctuations in the conformation of the substrate protein. Such fluctuations lead to global unfolding (see below), and destabilization of the local minima. This enables the polypeptide chain to surmount barriers rather readily making the transition to the native state facile.

In the previous calculations, the effective strength of the interaction between the hydrophilic residues of the substrate protein and the wall changes from neutral ($h = 0.0$) to maximal attraction ($h = 1.0$). It is likely that in practice the changes are not that large. The effects of varying degree of favorable interaction between the exposed hydrophobic residues of the substrate protein and the wall can be examined by allowing the value of h to go from h_{low} to h_{high} during a cycle. The folding time depends on the precise choice of h_{low} and h_{high}. For a cycling time of $t_c = 100$ mcs and $R_p = 1$, τ_f is plotted in Figure 7 as a function of the hydrophobic change, $\Delta\ h = h_{high} - h_{low}$, and their average, $\overline{h} = (h_{high} + h_{low})/2$. The ordinate represent $\overline{h}$ while points with the same Δh are connected through the same lines. This plot shows that, for a given T and L, the folding time is a minimum for the largest values of Δh as long as it is $0.4 < \overline{h} < 0.5$, and decreases as Δh increases. The variation which yields the smallest folding time corresponds to $h_{low} = 0.0$ and $h_{high} = 1.0$, i.e. the extreme values used previously. Notice that as Δh decreases towards zero, the optimal $\overline{h}$ value approaches the optimal value for fixed hydrophobicity, i.e. $h = 0.4$, shown in Figure 4. This Figure and Figure 6 clearly show that in addition to frequent cycling between on and off states, it is prudent to induce maximal changes in the interaction between the substrate and the cavity.

Temperature dependence of folding in a cavity with hydrophobic cycling

The rate enhancement obtained by chaperonin assisted folding with hydrophobic cycling can be

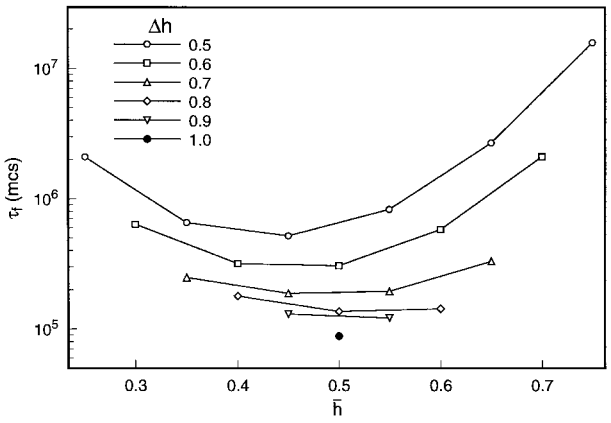

Figure 7. Mean folding time as a functions of the fluctuating chaperonin hydrophobicity. The hydrophobicity varies between $\overline{h}$ $\Delta h/2$ in a time of $t_c = 100$ mcs and for $R_p = 1$. The chaperonin size is $L = 5$ and the simulation temperature is $T = 0.053$. The error bars are as large as the symbols. Maximum variations in the fluctuations of the wall hydrophobicity lead to the fastest folding.

better appreciated by calculating τ_f as a function of temperature, with and without the chaperonin. The results are presented for the 16-mer and the 24-mer, in Figure 8. The chaperonin properties for both protein folding simulations are selected to be the same, i.e. $L = 5$, $h_{low} = 0.0$, $h_{high} = 1.0$, $R_p = 1$ and $t_c = 100$ mcs. These are the optimal conditions (leading to a minimum in the folding time) for the 16-mer obtained at $T_f = 0.053$. The same conditions are used for the 24-mer, because in reality the same chaperonin can assist many different proteins to fold. This is the case here as well because identical alterations in the properties of the cavity lead to improvement of folding of both model proteins in spite of their difference in length, sequence and folding ability (see Figure 8). At high temperatures (above T_θ), the folding time is smaller with the chaperonin; near T_θ, the folding time is smaller for spontaneous folding; and at low temperatures the chaperonin assisted folding is again more efficient. These results for the temperature are similar to that obtained for the case of fixed hydrophobicity. The main difference when the hydrophobicity oscillates is that at low temperatures, which are the relevant temperatures for reasons of stability and hence constitute non-permissive conditions, the decrease in folding time is significantly larger than when the hydrophobicity is fixed. The magnitude of the decrease in folding time increases dramatically as the temperature is lowered.

Inherent structure of the substrate protein before and after hydrophobic cycling shows evidence for global unfolding

One of the unresolved questions is the state of the substrate protein before and after being released from the cavity. Upon release of the GroES from the *cis* side, due to hydrolysis of ATP in the *trans*-ring, the polypeptide chain may be in a state that is fully committed to folding or may undergo kinetic partitioning. The course followed by the protein depends primarily on the size of the protein, but can also be affected by external conditions. The model considered here allows us to shed some light on this issue by following the state of the protein prior to change in the nature of the walls. It should be emphasized that the change in the hydrophobicity of the walls in computer simulations is instantaneous, whereas the relaxation time for volume expansion in GroEL cavity when ATP and GroES are added can be quite long. Nevertheless, our model can give qualitative insights into the nature of conformations of the substrate upon the change in the wall hydrophobicity.

We can monitor the conformational changes when the cavity goes from the on-state (hydrophobic interior) to an off-state (hydrophilic wall). Such conformational changes are most conveniently mapped by computing the inherent structure of the substrate chain. If the iterative annealing mechanism (Todd *et al.*, 1996) is operative then one expects that, upon changing the hydrophobicity of the cavity, the polypeptide chain will be committed to the native conformation with probability p. On the other hand, with probability $(1 - p)$ the chain will be trapped in a manifold of non-native states. For substrate proteins with sufficiently large ($p \geqslant 0.7$) one expects that one cycle of binding and release by GroE machinery is sufficient. This might be the case for Rhodanase. If p is small, as appears to be the case for Rubisco (Todd *et al.*, 1994, 1996), several iterations of binding and release will be needed to obtain sufficient yield.

The examination of the inherent structures, which correspond to that found in the local minima of a set of instantaneous conformations before and after the change in character of the hydrophobicity of the wall, allows us to estimate the degree to which the polypeptide chain is committed to the native conformation. Because of the stochastic nature of the Monte Carlo method, even at zero temperature, the inherent structure is represented as a probability function of some parameter representing the degree of folding. The inherent structure is parameterized in terms of

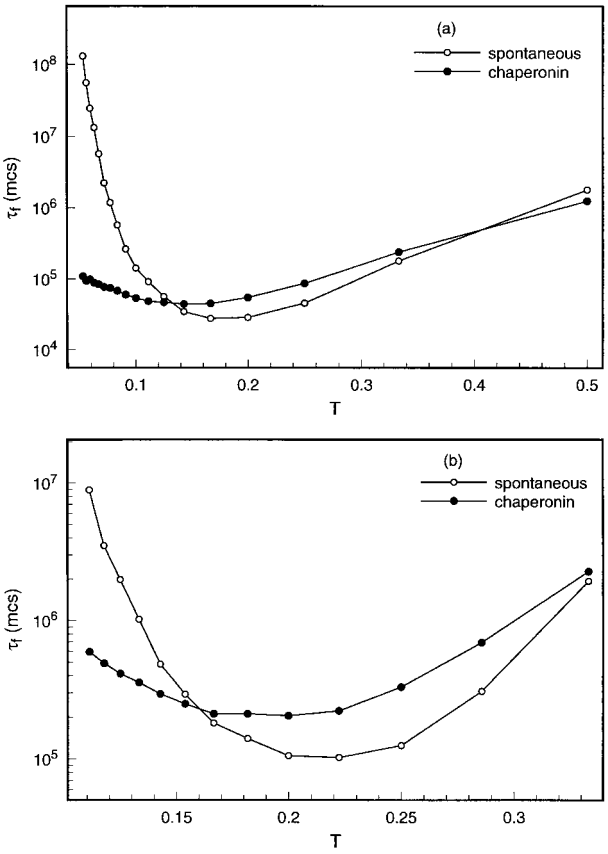

Figure 8. The dependence of the mean folding time on temperature with and without the chaperonin for (a) the 16-mer and (b) the 24-mer. The chaperonin conditions are $L = 5$ and $t_c = 100$ mcs for hydrophobicities fluctuating between $h = 0.0$ and $h = 1.0$ at $R_p = 1$. It is remarkable that the chaperonin-mediated folding continues to be efficient when spontaneous folding time becomes exceedingly long ($T \lesssim 0.1$).

the overlap, χ, which gives a direct measure of the degree of similarity between the given conformation and the native state. In order to observe the changes produced by hydrophobic cycling, we obtain the inherent structure right before and after a change in chaperonin hydrophobicity. The inherent structure is calculated by instantaneously changing the temperature to zero, fixing to and letting the conformation evolve to an energy minimum we show the inherent structure for the 16-mer (1) before and after a $h = 0.0$ (off-state) to $h = 1.0$ (on-state) transition and (2) before and after a $h = 1.0$ (on-state) to $h = 0.0$ (off-state) transition, for $T = 0.053$ and $L = 5$. The cycling time is selected as $t_c = 10^4$ mcs to allow for significant inherent structural changes after each hydrophobic change. Figure 9(a) shows that, right before a change from $h = 0.0$ to $h = 1.0$, a fraction (about 3 %) of the conformations is committed to folding ($\chi = 0$), while most of the conformations are likely to fall in misfolded ($\chi \approx 0.08$) or more unfolded ($\chi \geqslant 0.3$) conformations. Note that those misfolded conformations have many elements in common with the native state, and some may be thought of as being native-like. After the hydrophobicity changes to $h = 1.0$, the probability shifts to mostly unfolded conformations. This shows that upon change of the nature of the wall, the polypeptide chain undergoes nearly global unfolding. In other words, the chief mechanism operative in the GroEL-mediated folding (whether it takes place inside the cavity or not) is that chaperonin acts as an unfoldase (Todd *et al.*, 1996; Corrales & Fersht, 1996). For a hydrophobic change from $h = 1.0$ to $h = 0.0$, the inherent structure remains almost unchanged, with high prob-

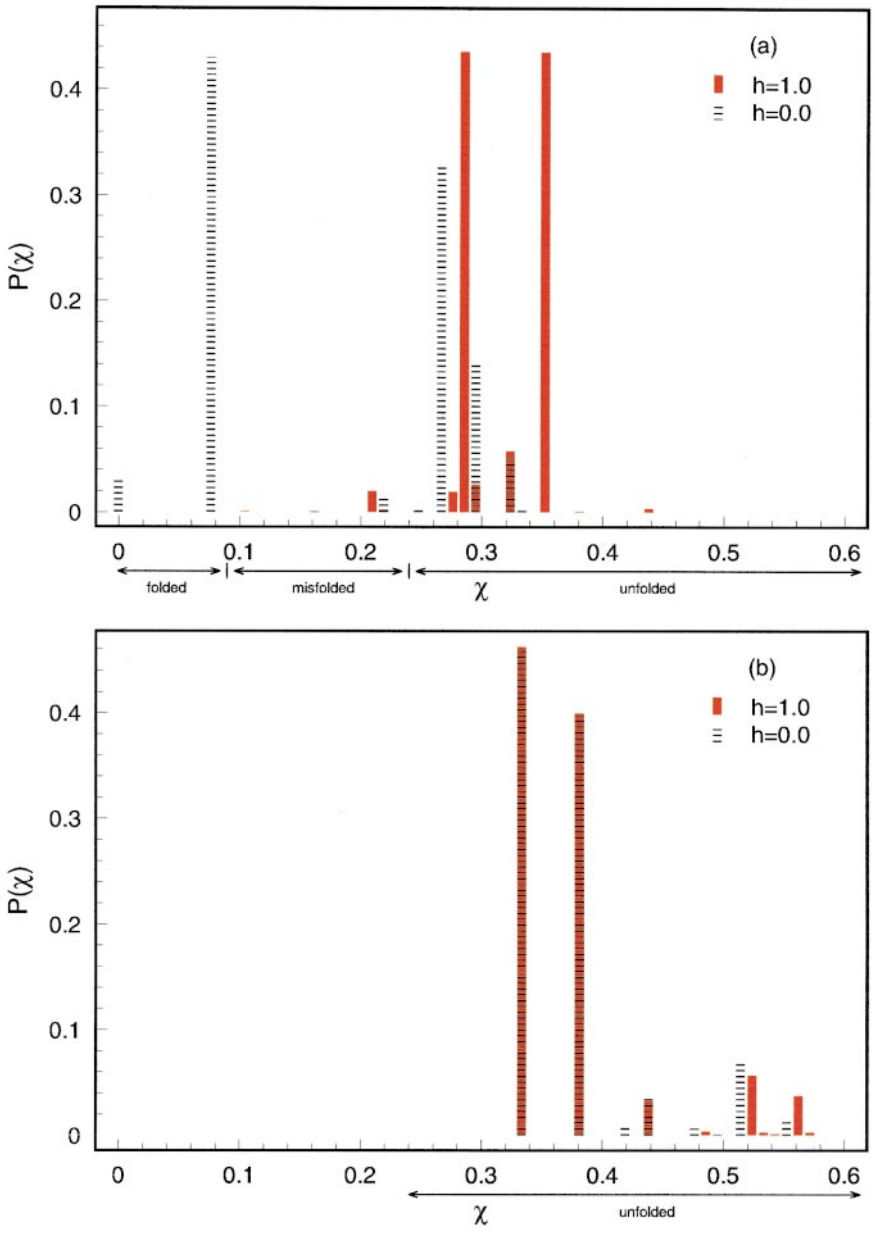

Figure 9. Inherent structure for the 16-mer in a chaperonin with hydrophobic cycling. The inherent structure was calculated before and after the hydrophobicity changed from (a) $h = 0.0$ to $h = 1.0$ and from (b) $h = 1.0$ to $h = 0.0$. The cycling time was set to $t_c = 10^4$ mcs, with a simulation temperature of $T = 0.053$ and $L = 5$. (a) Upon change of the wall from hydrophobic to hydrophilic, a fraction of molecules reach the native state while the remaining are either in misfolded conformations or in globally unfolded states.

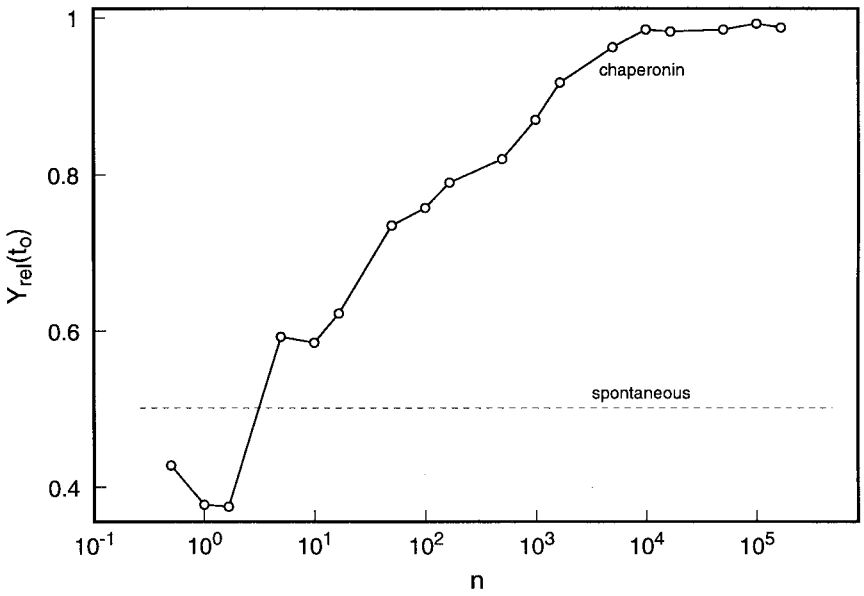

Figure 10. Relative yield *versus* hydrophobic cycling time for the 16-mer. The observation time is fixed at $t_o = 5.0 \times 10^5$ mcs which is almost six times as long as the folding time obtained at $t_c = 100$ mcs. The broken line shows the relative yield for spontaneous folding. The yield continues to increase upon repeated cycling which is in qualitative accord with the prediction of the iterative annealing mechanism. The temperature is held at $T = 0.053$ and $L = 5$.

abilities around the unfolded conformations (see Figure 9(b)). While the hydrophobicity is at $h = 0.0$, a fraction of the protein ensemble is able to fold to the native state while the rest get trapped in non-native conformations. When the hydrophobicity changes to $h = 1.0$, the fraction of molecules that fail to fold will be unfolded by the high hydrophobicity and will be given another chance to reach the native state after the hydrophobicity changes back to $h = 0.0$. This picture is in accord with the predictions of IAM, which was suggested as the mechanism by which chaperonin assisted folding works (Todd *et al.*, 1996). It is interesting to note that the same calculations for high cycling frequency give an inherent structure with a great portion of the probability concentrated around the native state. Therefore, high cycling frequency smooths the energy landscapes and produces trajectories that nucleate directly to the native state.

Influence of hydrophobic strength on the release of substrate protein from the walls of the cavity

The hydrophobic interactions play a fundamental role in the binding and unbinding of the substrate protein to the chaperonin. A chaperonin can recognize and bind non-native residues because they have a higher number of exposed hydrophobic residues than native conformations, which usually have most of the hydrophobic residues in the interior. For this reason, proteins in the native state can be easily released from the chaperonin. The binding affinity of the substrate protein to the chaperonin walls can be obtained by calculating the probability of de-pinning of the substrate protein from a wall for different conditions of wall hydrophobicities h and degree of conformation similarity with the native state χ. First, as a reference, the probability of the protein to come off a wall is calculated for $h = 0.0$, which is relatively large for all conformations and decay slightly as χ increases. The decay occurs because conformations with higher χ values have larger sizes and can extend from one wall to the other, restricting their movement off the walls. For $h = 1.0$, we found that the unfolded conformations are less likely to come off a wall. On the contrary, folded conformations have a higher probability of detaching from the walls, even higher than for $h = 0.0$. These observations imply that the hydrophobicity of the wall shifts the equilibrium to partially or fully unfolded states which once again confirms the crucial role of protein unfolding in association with GroEL. If during the ATP cycle the chaperonin changes immediately to its hydrophobic configuration after the GroES release, then the protein should have a greater probability of exiting the chaperonin in its native state.

Yield of the native state as a function of iterations

In order to probe the dependence of the yield of the native conformation as a function of the number of iterations, n, we define a dynamic relative yield, $Y_{rel}(t_o)$, which is the fraction of molecules that have folded in time smaller that the observation time t_o. The yield of the native state, which depends on t_o, is:

$$f_N(t_o) = Y_{rel}(t_o) P_{ns} \tag{3}$$

where P_{ns} is the thermodynamic probability of being in the native state in the absence of the chaperonin. As $t_o \to \infty$ $Y_{rel} \to \infty$, and $f_N(t_o \to \infty)$ is the thermodynamic fraction of molecules in the native state. The yield in equation (3) gives the fraction of native states reached inside the cavity under the assumption that upon reaching the native conformation they exit the cavity. The number of iterations n is:

$$n = t_o/t_c \quad (4)$$

In Figure 10 we show Y_{rel} as a function of n. As anticipated, the yield increased as n increases. According to the simpler version of IAM, the yield after n interactions of binding and release is given by:

$$f_N \simeq [1 - (1 - \phi)^n] \quad (5)$$

where ϕ is the kinetic partition factor. The above equation is only valid when t_h is relatively large. Nevertheless, we find that the value of ϕ extracted is not inconsistent with a more direct estimate based on the computation of the inherent structures.

The effect of the repeated cycles of changes in the hydrophobicity of the cavity is most vividly illustrated in Figure 11 which shows the fraction of native states $f_N(t_o)$ as a function of t_o. In this experiment the cycling time is fixed at $t_c = 100$ mcs. The higher temperature is $T_f = 0.024$ whereas the lower temperature corresponds to the condition $P_{ns} \simeq 0.87$. In both instances there is a dramatic increase in the yield compared to the spontaneous folding. At $T = 0.024$, the yield of the native state nearly reaches 0.9 in $t_o \lesssim 10^6$ mcs whereas the value of $f_N(t_o)$ for the spontaneous case is less than 0.2 even at $t_o \approx 7 \times 10^6$ mcs. Thus, the repeated annealing cycles has its most dramatic effect under non-permissive conditions when spontaneous folding is not-efficient.

Consistency with experiments and predictions

Despite the simplicity of the models used to investigate the effects of various confining environments on the refolding of proteins, there are a number of implications for experiments that are worth pointing out. We argue that certain global characteristics of refolding in the presence of GroEL/GroES are captured by the studies described here.

Folding rates and strength of binding between substrate and GroEL

Itzhaki *et al.* (1995) noted that by studying the refolding of GroEL-associated CI2 and 32 mutants that there is an inverse correlation between folding rates and the strength of interactions with the chaperonin. The stronger such interactions are, the slower the folding rate is. This is essentially consistent with our findings (see Figure 4). We have further shown that the maximum folding rates are found when the attractive interactions are around optimal values. Further decrease in the binding strength would be tantamount to having the substrate protein in a passive Anfinsen cage which does not (in general) lead to rate enhancement (see Figure 3). It would be interesting to devise mutations that show retardation in folding rates upon weakening of the substrate-GroEL interactions below the optimal value.

Substrate proteins can fold in a dynamic Anfinsen cage

The fundamentally important question of whether efficient refolding can occur while the substrate is sequestered in the central cavity was addressed using a number of experiments by Weissman *et al.* (1996). By using a designed single ring mutant of GroEL, referred to as SR1, they demonstrated that Rhodanase can refold while still

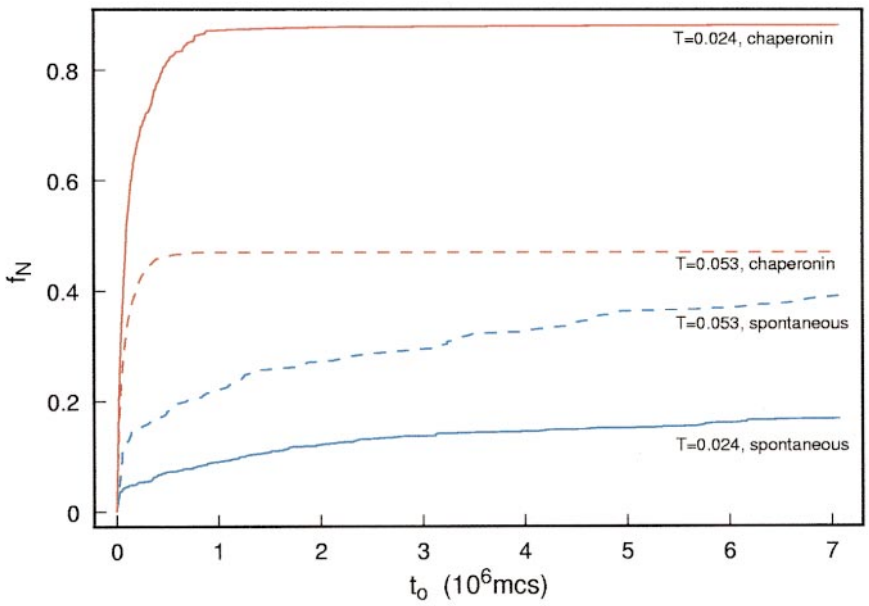

Figure 11. Dependence of yield (see equation (3)) on the observation time t_o for the 16-mer. The cycling time is fixed at $t_c = 100$. The blue lines are for spontaneous folding at two temperatures whereas the red lines correspond to assisted folding at the same temperatures. The yield for assisted folding is considerably greater than for spontaneous folding which implies that the repeated cycling does rescue those structures that are unable to reach the native state spontaneously on reasonable timescales.

being associated with SR1-GroES complex. The efficient refolding inferred from the activity of Rhodanase requires ATP, while the process does not occur with ADP. The kinetics for the wild-type GroEL and the substrate protein in association with the SR1-GroES-ATP complex are similar. These results suggest that upon adding ATP to SR1-GroES complex there is sufficient conformational change in the central cavity that the nature of interaction between SR1 and the substrate protein is altered. This could lead to a destabilization of the state in which Rhodanase is trapped which in turn leads to the efficient refolding. This interpretation emerges from our model of folding in a hydrophilic cavity with minimal recycling. We found (see Figure 7) that even moderate changes in the substrate-cavity interaction can alter the energy landscape favorably to lead to an enhancement in the folding rates.

Helping proteins unfold

Three experiments clearly show that upon change in the wall characteristics the substrate protein unfolds partially, if not globally. Zahn *et al.* (1996) used a hydrogen exchange labeling method to investigate the protection kinetics of barnase in the presence and absence of the molecular chaperones GroEL and SecB. They discovered that both the chaperones rapidly accelerates the exchange of highly protected proteins. In addition, they showed that collapse of the unfolded states to an intermediate containing protected amide proteins was considerably slowed in the presence of SecB or GroEL. These observations are consistent with the proposal that GroEL helps the protein globally, which in effect reduces the barriers to the native state.

In a more recent report, Nieba-Axmann *et al.* (1997) have investigated plausible structural fluctuations in GroEL-bound Cyclophilin A (CypA) using amide-proton exchange measurements probed by NMR spectroscopy. They find that in the absence of nucleotides (ATP or ADP) and GroES the folding is extremely slow with the rate being 1.4×10^{-3} min^{-1} at $T = 6\,^{\circ}$C. We will assume that the spontaneous folding rate is approximately the same as folding in a passive (no nucleotide or GroES) cage. Upon addition of ADP the rate increases by a factor of about 2.5, while the addition of ATP leads to a factor of three enhancement in the rate. However, when GroES is added the rate increases by a factor of about 14 at $T = 6\,^{\circ}$C and is independent of the nature of the nucleotide (ATP or ADP). The refolding rate increases by a factor of nearly 30 at $T = 30\,^{\circ}$C in the presence of ADP and GroES. The yield reaches nearly 100 % in the presence of GroES. The near independence of the refolding process on the nucleotides suggest that there is no cycling of GroES involved in the release of native CypA from the GroEL-CypA complex.

The data on the CypA folding rates gives qualitative support for results presented here. It is known that upon binding of ATP and GroES, the apical domain moves inward exposing an unfavorable cavity to the substrate protein. We assume that this can be modeled by a change in the wall hydrophobicity. In order to compare directly with the experiments on CypA we performed the following computation. We started with the 16-mer in a mildly hydrophobic cavity ($h = 0.4$). The chain was allowed to fold for nearly one-tenth of the expected folding time. Then the value of h was raised to 1.0 and maintained at this value for an equal time. After this, h was set to 0.4 and the folding was allowed to complete. We find that the folding rate increases by almost a factor of ten, relative to spontaneous folding. If the addition of GroES results in a change on wall hydrophobicity, then our model would be consistent with the GroEL-GroES mediated folding of CypA.

More importantly, Nieba-Axmann *et al.* (1997), showed directly that protons that are protected from hydrogen exchange in the native state of CypA in the absence of GroEL become much less protected when it is bound to the chaperonin. The protection factor changes by nearly two to three orders of magnitude. This suggests that the equilibrium of CypA has shifted markedly from compact native state to more globally unfolded states, which typically have low protection factors. Thus, GroEL acts dynamically to unfold the substrate proteins in the course of assisted folding. Nieba-Axmann *et al.* (1997) suggest that efficient GroEL-mediated folding can be triggered by repeatedly unfolding the protein. These observations are reflected in our findings as well. In the case of confined folding subject to hydrophobic cycling we find that just after the cavity offers a predominantly hydrophilic environment to the substrate protein, globally unfolded states ($\chi > 0.3$) become populated (see Figure 9). The equilibrium shifts to these states until the hydrophobicity is reversed. In addition, we find that both the rates and yield can be substantially enhanced if the cavity undergoes frequent cycling between hydrophobic and hydrophilic environment. Both these observations support the findings in CypA refolding mediated by GroEL.

G.H. Lorimer and co-workers (personal communication) have explored the function of GroEL in mediating the folding of Rubisco using hydrogen-trition exchange. Remarkably, they found under non-permissive folding conditions, that the complete exchange kinetics of protected proteins occurs in a time-scale of 13 seconds. This experiment clearly shows that within one turnover of the GroEL, partial unfolding of the substrate proteins takes place.

Relating folding rates to the equilibrium between on and off-states of GroEL

One of the important predictions of our work is illustrated in Figure 6. For a given cycling time, the decrease in the GroE-mediated folding time is

greatest if R_p is less than or equal to one. The quantity R_p measures the equilibrium between H (on-state) and P (off-state) of GroEL. This equilibrium constant has been measured and can be changed chemically (Horovitz, 1998). Since R_p is directly linked to the degree of cooperativity in the allosterictransitions in the GroEL subunits, our calculations predict that there has to be a direct link between the chaperonin-mediated rates and cooperativity of the H → P transitions. This prediction is experimentally testable (A. Horovitz, personal communication).

The results presented here and suggestions from the experiments on barnase, CypA and Rubisco show that if the rate of on and off-states of GroEL can be increased one should see a substantial increase in the GroEL-assisted folding of proteins. The natural cycling time for *E. coli* is roughly 15–20 seconds. This time is typically larger than the times needed for kinetic partitioning to occur. A key unresolved question is why is t_c in *E. coli* GroEL considerably larger than the times needed for all other processes that involve the substrate proteins? It may be a biological constraint that is imposed so that GroEL can perform other functions besides the processing of proteins. Alternatively, the relaxation time for large conformational fluctuations, volume expansion from 85,000 Å^3 to 175,000 Å^3 in the cavity size, can be large. In any event our results suggest that if t_c is decreased (chemically or by mutations), the rates and yields of refolding can be increased, especially for substrate proteins that require multiple rounds of binding and release.

Discussion

The results presented here suggest that protein folding in a confined environment, such as that provided by a molecular chaperonin like GroES-GroEL, can in fact speed up the folding process, and not just serve as a mechanism to keep the proteins from aggregation. In other words, even when only one round of binding and release is involved, GroEL-GroES acts as a dynamic Anfinsen cage. In the simple minimal models studied here, we have been able to systematically study the effects on folding created by several environmental factors presented to the substrate protein by the chaperonin. These factors are the confinement, a constant hydrophobic environment, and dynamically changing hydrophobic environment.

The folding kinetics are not significantly affected by confinement alone. The transitions for a collapsed protein *en route* to the native state generally involves inter conversion from one nearly compact conformation to another. The conformational changes are only large enough for this process to occur, so the confining cavity does not need to be much larger than the protein. In our case, we found that the confining cavity needs to be around one lattice space (the length of one bond) larger than the native state size, for the protein to fold without a significant decrease in the folding time. However, at low temperatures, when the native state is stable, the following time increases upon confinement. At temperatures higher than the collapse temperature, confinement also stabilizes the protein and decreases its folding time by reducing the available conformation space. The volume of the cavity even in this case is considerably larger than that occupied by the substrate protein.

If the walls of the confining cavity are moderately hydrophobic, then the folding time (compared to the time for spontaneous folding) can be reduced under non-permissive folding conditions. That is, at temperatures higher or lower than the collapse temperature, the folding time is significantly decreased by letting a fraction (less than half) of the confining cavity interior walls be hydrophobic. However, around the collapse temperature, the spontaneous folding time is minimal and confinement in a purely hydrophobic cavity does not reduce the folding time. Our calculations indicate that a moderately hydrophobic confinement assists the protein by lowering the barriers (or smoothing the energy landscape) as pointed out by Chan & Dill (1996). However, if the confining walls are too hydrophobic, it leads to enhanced substrate-cavity interactions and folding becomes inefficient. Real chaperonins are mostly hydrophilic when GroES is bound to GroEL. By visual inspection of the GroEL-GroES complex (Xu *et al.*, 1997), the number of hydrophobic residues in the cavity interior walls is between 20 % and 40 %, which is in the range of the optimal values obtained here.

The finding that purely hydrophobic cavity can refold a chain under non-permissive conditions gives support to the suggestion that mini-chaperones can have function in a manner similar to GroEL. Fersht and co-workers (Buckle *et al.*, 1997) have argued, using the X-ray structure of the apical domain of GroEL (residues 191-376), that by binding to exposed hydrophobic regions of substrate protein, they essentially facilitate unfolding. This event is, as we have shown, crucial to smoothing the energy landscape so that folding to the native conformation is feasible. These arguments would also imply that the precise structure adopted by mini-chaperones is not crucial. The key characteristic may be that the mini-chaperone retains an overall hydrophobic character so that several proteins can be processed. Mutational studies will be helpful in deciphering the extent to which the hydrophobic character can be altered while still ensuring chaperonin activity.

A cavity with periodically varying hydrophobicity can substantially reduce the folding time. Such an environment is a better model for a chaperonin whose structural changes expose the substrate protein to changes in hydrophobicity rather than to a constant hydrophobicity. A number of these chaperonin cycles are believed to take place as the substrate protein continue to be encapsulated and

ejected by the chaperonin until it folds (Weissman *et al.*, 1994; Todd *et al.*, 1996). The optimal folding times are obtained for high hydrophobic cycling rates. The resulting folding times can be many orders of magnitude faster than the ones corresponding to spontaneous folding, in particular as the temperature is decreased. It is tempting to speculate that if it were possible to increase the rate of conformational changes of GroEL, leading to frequent changes in hydrophobicity, one could further increase the folding rate of proteins. This prediction is, in principle, amenable to experimental test.

Methods

Model

The polypeptide chains in our simulations are modeled as a linear connected object of single-bead residues restricted to a 3D cubic lattice. This model has been commonly used in computational studies of *in vitro* protein folding (Dill *et al.*, 1995). The interactions between the residues are taken to be contact pair interactions which are non-zero only when the distance between two residues is one lattice space, a. If we denote the conformation of the chain as $\vec{r}_i$, ($i = 1, 2, \ldots, N$), with N being the number of residues, then its energy is given by:

$$E(\{r_i\}) = \sum_{i<j} B_{ij}\Delta(|r_i - r_j| - a) \tag{6}$$

In equation (6) the matrix elements B_{ij} represent the contact interaction between the residues i and j, and:

$$\Delta(x) = \begin{cases} 1 & x = 1 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The interactions between beads in the C^{α} representation of the polypeptide chain are mimicked using popular statistical potentials representing the contact interactions between the various amino acid residues. The contact interaction potentials between the amino acid residues, given by the matrix elements B_{ij} in equation (6), are obtained from a modification of the Miyazawa-Jernigan (MJ) potential (Miyazawa & Jernigan, 1996). The chaperonin machinery is promiscuous in its binding to a variety of substrate proteins in its non-native forms, and the interaction between GroEL and the substrate protein occurs by hydrophobic interaction (Viitanen *et al.*, 1992). Since the major driving force for protein folding and for the recognition of the substrate protein by GroEL primarily involve hydrophobic interactions (Lin *et al.*, 1995) it is useful to use potentials that reflect this underlying property. The advantage of using statistical potentials is that they correlate well with the experimentally determined hydrophobicities (Jernigan & Bahar, 1996).

Here, we modify the MJ parameters to account for some apparent discrepancies. The elements of the modified MJ potentials are obtained by setting the interactions between solvent molecules and between the residues and the solvent molecules to zero. As in the HP model (Dill *et al.*, 1995), we expect that the energies between the hydrophilic residues should be approximately zero. However, the MJ parameters give large negative energies between most hydrophilic residues and between like-charged residues. We modify the MJ parameters by assuming that threonine (T) and the solvent interact similarly with other residues and define the energy between the two types of residues as:

$$E_{ij} = M_{ij} + M_{\mathrm{TT}} - M_{i\mathrm{T}} - M_{\mathrm{T}j}, \tag{8}$$

where M_{ij} are the MJ parameters, $i, j = 1, \ldots, 20$ and $E_{i0} = 0$ for $i, j = 0, 1, \ldots, 20$, and 0 represents the solvent. Notice that $E_{iT} = 0$ for all residues i. Threonine was selected as the reference residue because it gives the best correlation between experimental hydrophobicities and $E_{ii}/2$. A detailed description of this model can be found elsewhere (Betancourt & Thirumalai, 1998).

The central cavity of the chaperonin is modeled as a cubic box of length L. The interaction between the walls of the cavity and the ith residue of the substrate protein is obtained by assuming that the chaperonin interior can be modeled as a hydrophobic surface. The diversity of hydrophobic interaction can be varied by letting, for simplicity, one specific residue chosen as leucine (L) to describe the wall character. Thus, the interactions between the wall and the substrate protein is:

$$E_c = \sum_i hE_{\mathrm{L}i} \tag{9}$$

where $0 \leqslant h \leqslant 1$ gives the strength of the interaction, and $E_{\mathrm{L}i}$ is the contact interaction between the ith residue of the substrate protein and the wall. The total energy of an encapsulated substrate protein is given by the sum of equations (6) and (9).

The energies E_{ij} and temperature T are given in units of RT_o, where T_o represents some constant reference temperature in Kelvin. Because of the approximations involved in the calculation of the energies E_{ij}, the temperature T_o is not necessary the ambient temperature. Therefore, the absolute scale of our energies and temperatures is somewhat arbitrary. Length is measured in units of lattice spacing a (see equation (6)).

Description of the simulations

The dynamics of folding are simulated using the standard Monte Carlo method by Metropolis *et al.* (1953). We allow for all types of one, two or three simultaneous residue moves that are locally displaced for a minimum number of lattice sites (Betancourt, 1998). The protein folding time (or mean folding time τ_f) is taken to be the mean first passage time (MFPT) averaged over an ensemble of random initial conditions. The first passage time for a given initial member of the denatured ensemble is the first time the unfolded polypeptide chain reaches the native state. The unit of time is the Monte Carlo step (mcs) defined as each attempt to make a move by the algorithm whether successful or not. This is necessary to ensure detailed balance.

The standard Metropolis Monte Carlo method has to be modified in the presence of the confining walls. The conformations of the encapsulated polypeptide chain can undergo substantial fluctuations due to interactions with the walls. In some instances topological restrictions can arise that pin certain positions of the chain against the wall. In this case, the substrate protein could remain in a metastable conformation for unphysically long periods causing most of the transitions to fail. Highly correlated multi-particle moves may be necessary to release the chain from such frustrated conformations. The usual local moves, such as the ones employed here, may prove inadequate. From a kinetic point of view, hard repulsive collisions tend to push the protein away from the walls enabling the polypeptide chain to escape from such

metastable states. These effects can be simulated by adding a transition rate representing a protein rigid body displacement in the direction opposite to the collision due to interactions with the wall. In order to mimic these effects, we introduce rigid body (rb) moves that are generated by interactions with the walls at a transition rate given by:

$$K_{\mathrm{rb},i} = K^{o}_{\mathrm{rb},i} e^{-\max(\Delta E_i, 0)/k_b T} \tag{10}$$

where $K^{o}_{\mathrm{rb},i}$ is the transition attempt rate in the ith direction ($i = 1, 2, 3$) and ΔE_i is the change in energy caused by the rb move, due to the wall in i. Thus in the presence of the chaperonin, a Monte Carlo step now consists of the regular residue (or residues) move followed by a rb move. The rb moves have to be accepted first for $K^{o}_{\mathrm{rb},i}$, and then for the energy change ΔE_i due to the chaperonin walls.

The determination of $K^{o}_{\mathrm{rb},i}$ is somewhat arbitrary and is calculated as follows. When the substrate protein is in contact with a wall in the ith direction and a residue move attempt fails because of the presence of that wall (a collision), we set $K^{o}_{\mathrm{rb},i} = \pm 1$, where the sign corresponds to the direction opposite to the wall. This is followed by a rb move attempt, which if it fails, $K^{o}_{\mathrm{rb},i}$ is set to zero, but if it succeeds then $K^{o}_{\mathrm{rb},i}$ is calculated from the total time spent near that wall (w_i) and the number of collisions against that wall (c_i), i.e.:

$$K^{o}_{\mathrm{rb},i} = \pm \frac{c_i}{w_i} \tag{11}$$

While the protein is not in contact, $K^{o}_{\mathrm{rb},i}$ remains constant until another collision with a wall occurs. If the protein collides with a wall due to a rb move, the transition attempt rate is changed from $K^{o}_{\mathrm{rb},i}$ to $-qK^{o}_{\mathrm{rb},i}$, where $0 \leqslant q \leqslant 1$ which we set to $q = 0.9$. This represents a ''semi-elastic'' collision introduced to increase the convergence rate of the detail balance for $K^{o}_{\mathrm{rb},i}$, which is only satisfied on an average due to opposite wall collisions. The value of $q = 0.9$ was selected so that the protein has a high probability of being pushed away from the walls, if the change in energy allows it. If after a rb collision the protein fails to bounce off the wall, $K^{o}_{\mathrm{rb},i}$ is set to zero and it is later regenerated by residue collision as previously described. The appeal of this method for obtaining $K^{o}_{\mathrm{rb},i}$ is that the rb motions originate from the residue own moves and their interactions with the walls.

The protein thermodynamics, with or without the chaperonin, are also obtained from Monte Carlo simulations using a similar move set to the one used in the dynamics simulations, or from full enumeration if possible. We utilize an implementation of the multiple histogram algorithm (Ferrenberg & Swendsen, 1989) to obtain the probability distribution in energy and in any given order parameter. The details are described elsewhere (M.R.B. & D.T., unpublished results).

Sequence selection

It is suspected that majority of the proteins fold spontaneously (Lorimer, 1996; Netzer & Hartl, 1998) even in cellular conditions. This implies that the crowded conditions in the cell do not significantly alter the pathways encountered in the course of *in vitro* folding. The hallmark of chaperonins is that they permit efficient folding of substrate proteins to occur under non-permissive conditions, where spontaneous folding is hardly detectable. In terms of the kinetic partitioning mechanism (Guo & Thirumalai, 1995) the partition factor, i.e. the fraction of the initial population of molecules that fold rapidly without being trapped in discernible intermediates is very small. Thus, to explore chaperonin-induced folding mechanisms it is necessary to choose protein-like sequences that are not optimized so that their folding times are considerably longer than the typical observation times. We are also interesting in investigating the effect of chaperonins in cases where the protein folds quite readily.

With the above criteria in mind, two sequences are chosen. The sequences are 16 and 24 residues long. The sequence with 16 residues, or 16-mer, is shown in Figure 12(a) and is designed to have a unique energy ground state, referred to as the native state. This sequence is not well optimized so that the rate of folding is expected to be slow. This allows us to test whether the chaperonin model can improve the folding properties of proteins under ''non-permissive'' conditions. The

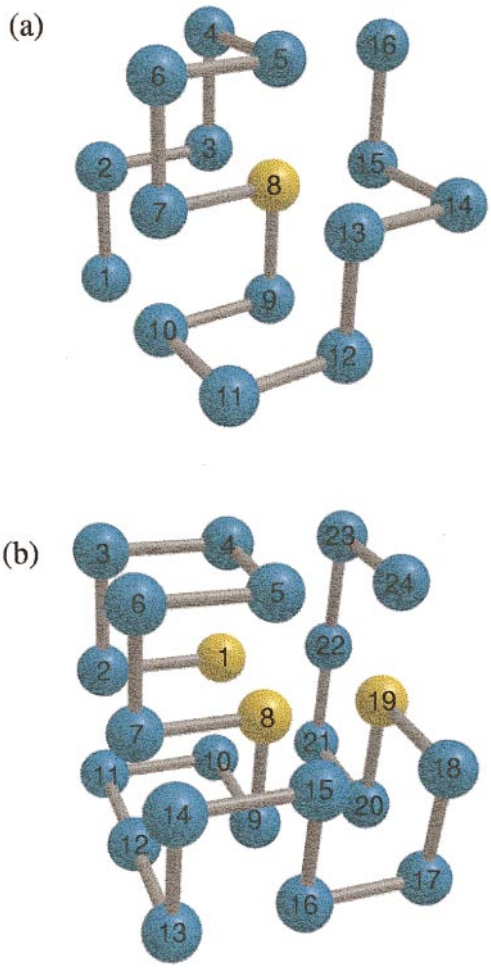

Figure 12. Native structure for the (a) 16-mer and (b) 24-mer. The yellow residues represent the hydrophobic residues and the blue residues the hydrophilic ones. The sequence for the 16-mer is (D R A H Y H D F R R H D A H A P) and its native state energy is −4.72. This sequence produces a unique native state with bad folding properties. For the 24-mer, the sequence is (C D R G A E K V D H H N Q Q A K E A V R H A G A) with a native state energy of −8.69. This sequence produces a unique native state with good folding properties.

sequence is limited to 16 residues so that better statistics on the slow folding protein can be obtained. This seems to be the smallest chain length for which the ''core'' is completely surrounded by other residues and the native state is characterized by a unique set of residue contacts. The core residue is hydrophobic (Phe) and the surface residues are mostly hydrophilic. The surface residues that are in contact with the core residues (Tyr, Ala) have intermediate hydrophobicities in our model (Betancourt & Thirumalai, 1998). It is important to have the hydrophobic residues in the core because the ability of chaperonins to distinguish between folded and unfolded conformations depends on it.

The second sequence, a 24-mer, is designed to be an efficient folder. Simulations with this model allow one to determine the effect of confinement on a protein that folds spontaneously, i.e. without any chaperonins. The 24-mer sequence was obtained with a design algorithm described elsewhere (M.R.B. & D.T., unpublished results). The native conformation and sequence for the 24-mer is shown in Figure 12(b). The native state of the 24-mer also seems to be unique, as determined by performing global search for lowest energy conformation by Monte Carlo methods. The arrangement of the residues resulted in a core consisting of a hydrophobic residue and the surface of mostly hydrophilic residues. Only two hydrophobic residues lie in the surface of the structure but these are located on the center of faces and in contact with many other residues, including the core residue, so they could almost be considered core residues.

Acknowledgments

We thank G. Lorimer and A. Horovitz for several stimulating discussions and providing results on Rubisco prior to publication. We thank A. Horovitz for bringing a few references to our attention. M.R.B is a National Science Foundation Minority Postdoctoral Fellow. This work was supported in part by a grant from the National Science Foundation through grant number NSF CHE96-29845.

References

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science,* **181**, 223-230.

Betancourt, M. R. (1998). Smoothing the landscapes of protein folding: insights from a minimal model. *J. Chem. Phys.* **109**, 1545-1155.

Betancourt, M. R. & Thirumalai, D. (1998). Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.* **8**, 361-369.

Buckle, A. M., Zahn, R. & Fersht, A. R. (1997). A structural model for GroEL-polypeptide recognition. *Proc. Natl Acad. Sci. USA,* **94**, 3571-3575.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature,* **371**, 578-586.

Camacho, C. J. & Thirumalai, D. (1993). Kinetics and thermodynamics of folding in model proteins. *Proc. Natl Acad. Sci. USA,* **90**, 6369-6372.

Chan, H. S. & Dill, K. A. (1994). Transition states and folding dynamics of proteins and heteropolymers. *J. Chem. Phys.* **100**, 9238-9257.

Chan, H. S. & Dill, K. A. (1996). A simple model of chaperonin-mediated protein folding. *Proteins: Struct. Funct. Genet.* **24**, 345-351.

Corrales, F. J. & Fersht, A. R. (1996). Toward a mechanism for GroEL-GroES chaperone activity: an ATPase-gated and -pulsed folding and annealing cage. *Proc. Natl Acad. Sci. USA,* **93**, 4509-4512.

Cui, Y., Chen, R. S. & Wong, W. H. (1998). Molecular dynamic simulation of chaperonin-mediated protein folding. *J. Protein Chem.* **17**, 377-380.

Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995). Principles of protein folding - a perspective from simple exact models. *Protein Sci.* **4**, 561-602.

Ellis, R. J. (1994). Molecular chaperones: opening and closing the Anfinsen cage. *Curr. Biol.* **4**, 633-635.

Ellis, R. J. & Hartl, F. U. (1996). Protein folding in the cell: competing models of chaperonin action. *FASEB J.* **10**, 20-26.

Federov, A. N. & Baldwin, T. O. (1997). GroE modulates kinetic partitioning of folding intermediates between alternative states to maximize the yield of biological active protein. *J. Mol. Biol.* **268**, 712-723.

Fenton, W. A. & Horwich, A. L. (1997). GroEL-mediated protein folding. *Protein Sci.* **6**, 743-760.

Ferrenberg, A. M. & Swendsen, R. H. (1989). Optimized Monte Carlo data analysis. *Phys. Rev. Letters,* **63**, 1195-1198.

Gulukota, K. & Wolynes, P. G. (1994). Statistical mechanics of kinetic proof reading in protein folding *in vivo. Proc. Natl Acad. Sci. USA,* **91**, 9292-9296.

Guo, Z. & Thirumalai, D. (1995). Protein folding: nucleation mechanism, time scales, and pathways. *Biopolymers,* **36**, 83-102.

Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. *Nature,* **381**, 571-580.

Horovitz, A. (1998). Structural aspects of GroEL function. *Curr. Opin. Struct. Biol.* **8**, 93-100.

Hunt, J. F., Weaver, A. J., Landry, S. J., Gierasch, L. & Deisenhofer, J. (1996). The crystal structure of the GroES co-chaperonin at 2.8 Å resolution. *Nature,* **379**, 37-49.

Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). Nature and consequences of GroEL-protein interactions. *Biochemistry,* **34**, 14581-14587.

Jackson, G. S., Staniforth, R. A., Halsall, D. J., Atkinson, T., Holbrook, J. J., Clarke, A. R. & Burston, S. G. (1993). Binding and hydrolysis of nucleotides in the chaperonin catalytic cycle: implications for the mechanism of assisted protein folding. *Biochemistry,* **32**, 2554-2563.

Jernigan, R. L. & Bahar, I. (1996). Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* **6**, 195-209.

Klimov, D. K. & Thirumalai, D. (1996). Criterion that determines the foldability of proteins. *Phys. Rev. Letters,* **76**, 4070-4073.

Lin, Z., Schwartz, F. P. & Eisenstein, E. (1995). The hydrophobic nature of GroEL-substrate binding. *J. Biol. Chem.* **270**, 1011-1014.

Lorimer, G. H. (1996). A quantitative assessment of the role of the chaperonin proteins in protein folding *in vivo. FASEB J.* **10**, 5-9.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. & Teller, A. H. (1953). Equation of state calculations

by fast computing machines. *J. Chem. Phys.* **21**, 1087-1092.
Minton, A. P. (1995). Confinement as a determinant of macromolecular structure and reactivity. II. Effects of weakly attractive interactions between confined macrosolutes and confining structures. *Biophys. J.* **68**, 1311-1322.
Miyazawa, S. & Jernigan, R. L. (1996). Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623-644.
Netzer, W. J. & Hartl, F. U. (1998). Protein folding in the cytosol: chaperonin-dependent and -independent mechanisms. *Trends Biochem. Sci.* **23**, 68-73.
Nieba-Axmann, S. E., Ottinger, M., Wüthrich, K. & Plückthun, A. (1997). Multiple cycles of global unfolding of GroEL-bound cyclophilin A evidenced by NMR. *J. Mol. Biol.* **271**, 803-818.
Orland, H. & Thirumalai, D. (1997). A kinetic model for chaperonin assisted folding of proteins. *J. Phys.* **7**, 553-560.
Richardson, A., Landry, S. J. & Georgopoulos, C. (1998). The ins and outs of a molecular chaperone machine. *Trends Biochem. Sci.* **23**, 138-143.
Rye, H. S., Burston, S. G., Fenton, W. A., Beechem, J. M., Xu, Z., Sigler, P. B. & Horwich, A. L. (1997). Distinct actions of *cis* and *trans* ATP within the double ring of the chaperonin GroEL. *Nature,* **388**, 792-798.
Schmidt, M., Rutkat, K., Rachel, R., Pfeifer, G., Jaenicke, R., Viitanen, P., Lorimer, G. H. & Buchner, J. (1994). Symmetric complexes of GroE chaperonins as part of the functional cycle. *Science,* **265**, 656-659.
Sfatos, C. D., Gutin, A. M., Abkevich, V. T. & Shakhnovich, E. I. (1996). Simulations of chaperone-assisted folding. *Biochemistry,* **35**, 334-339.
Thirumalai, D. (1994). Theoretical perspective on in vitro and in vivo protein folding. In *Statistical Mechanics, Protein Structure, and Protein-Substrate Interactions* (Doniach, S., ed.), pp. 115-134, Plenum, New York.
Todd, M. J., Viitanen, P. V. & Lorimer, G. H. (1994). Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding. *Science,* **256**, 659-666.
Todd, M. J., Lorimer, G. H. & Thirumalai, D. (1996). Chaperonin facilitated folding: optimization of rate and yield by an iterative annealing mechanism. *Proc. Natl Acad. Sci. USA,* **93**, 4030-4035.
Viitanen, P. V., Gatenby, A. A. & Lorimer, G. H. (1992). Purified chaperonin 60 (GroEL) interacts with the non-native states of a multitude of *Escherichia coli* proteins. *Protein Sci.* **1**, 363-369.
Weissman, J. S., Kashi, Y., Fenton, W. A. & Horwich, A. L. (1994). GroEL-mediated folding proceeds by multiple rounds of binding and release of non native forms. *Cell,* **78**, 693-702.
Weissman, J. S., Rye, H. S., Fenton, W. A., Beechem, J. M. & Horwich, A. L. (1996). Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell,* **84**, 481-490.
Xu, Z., Horwich, A. & Sigler, P. B. (1997). The crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex. *Nature,* **388**, 741-750.
Zahn, R., Perrett, S., Stenberg, G. & Fersht, A. R. (1996). Catalysis of Amide proton exchange by the molecular chaperones GroEL and SecB. *Science,* **271**, 642-645.

Edited by A. R. Fersht

(Received 12 October 1998; received in revised form 1 February 1999; accepted 1 February 1999)