doi:10.1016/S0022-2836(03)00929-X J. Mol. Biol. (2003) 332, 701–713

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Effects of Confinement in Chaperonin Assisted Protein Folding: Rate Enhancement by Decreasing the Roughness of the Folding Energy Landscape

A. Baumketner, A. Jewett and J. E. Shea*

Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106, USA

Chaperonins, such as the GroE complex of the bacteria *Escherichia coli*, assist the folding of proteins under non-permissive folding conditions by providing a cavity in which the newly translated or translocated protein can be encapsulated. Whether the chaperonin cage plays a passive role in protecting the protein from aggregation, or an active role in accelerating folding rates, remains a matter of debate. Here, we investigate the role of confinement in chaperonin mediated folding through molecular dynamics simulations. We designed a substrate protein with an α/β sandwich fold, a common structural motif found in GroE substrate proteins and confined it to a spherical hydrophilic cage which mimicked the interior of the GroEL/ES cavity. The thermodynamics and kinetics of folding were studied over a wide range of temperature and cage radii. Confinement was seen to significantly raise the collapse temperature, T_c, as a result of the associated entropy loss of the unfolded state. The folding temperature, T_f, on the other hand, remained unaffected by encapsulation, a consequence of the folding mechanism of this protein that involves an initial collapse to a compact misfolded state prior to rearranging to the native state. Folding rates were observed to be either accelerated or retarded compared to bulk folding rates, depending on the temperature of the simulation. Rate enhancements due to confinement were observed only at temperatures above the temperature T_m, which corresponds to the temperature at which the protein folds fastest. For this protein, T_m lies above the folding temperature, T_f, implying that encapsulation alone will not lead to a rate enhancement under conditions where the native state is stable ($T < T_f$). For confinement to positively impact folding rates under physiological conditions, it is hence necessary for the protein to exhibit a folding transition above the temperature at which it exhibits its fastest folding rate ($T_m < T_f$). We designed a protein with this property by reducing the energetic frustration in the original α/β sandwich substrate protein. The modified protein exhibited a twofold acceleration in folding rates upon encapsulation. This rate enhancement is due to a mechanistic change in folding involving the elimination, upon encapsulation, of accessible local energy minima corresponding to structures with large radii of gyration. For this protein, confinement hence plays more than the role of a passive cage, but rather adopts an active role, accelerating folding rates by decreasing the roughness of the energy landscape of the protein.

Keywords: chaperonin; protein folding; energy landscape; minimalist off-lattice model; computer simulation

*Corresponding author

Introduction

Molecular chaperonins, a class of large multi-domain proteins found in abundance in the cytosol, facilitate the folding of newly synthesized proteins under conditions where their spontaneous

Permanent address: A. Baumketner, Institute for Condensed Matter Physics, 1 Svientsitsky Str., Lviv 79011, Ukraine.

Abbreviations used: IAM, iterative annealing mechanism; ACM, Anfinsen cage model.

E-mail address of the corresponding author: shea@chem.ucsb.edu

folding is prohibited or slow.[1] The most thoroughly studied chaperonin system is the *Escherichia coli* GroEL/ES complex.[2] A protein in a non-native conformation can bind to the apical domain of one of the two rings of GroEL through non-specific hydrophobic interactions. Binding of ATP triggers the capping of GroEL by its co-chaperone GroES, doubling the size of the cavity[1] and displacing the substrate into a now hydrophilic inner cage. Subsequent hydrolysis of ATP molecules in the GroEL/ES complex and binding of ATP molecules to the opposite ring destabilizes the GroEL/ES complex and leads to its dissociation.[1] The substrate protein is then released in either the native state, in the event of successful folding within the cage, or in a non-native state. The non-native substrates can easily be recaptured at a later time, while the native, functional proteins can no longer be bound.

While the structure of the GroE complex is now well characterized,[1,3] the exact mechanism by which chaperonins help proteins to fold remains a subject of continuing debate. Two theories for chaperonin-mediated folding have been put forth: the iterative annealing mechanism (IAM)[4–7] and the Anfinsen cage model (ACM).[8–10] The IAM builds upon experimental observations that multiple rounds of ATP-driven GroEL/ES association and dissociation may be needed to fully convert the unfolded/misfolded proteins to the folded state under non-permissive folding conditions.[4,11] It is assumed that folding partitions between a fast and slow folding channel, and that efficient folding is hampered by the presence of long-lived (kinetically trapped) folding intermediates that contribute to the slow folding channel.[5] The IAM suggests that chaperonins actively unfold misfolded intermediates, possibly through mechanical force,[7] allowing them to re-partition into the folding channels. The net result of chaperonin cycling is acceleration in folding rates due to the transfer from the slow folding track to the faster route. Recent theoretical studies and computer simulations using lattice protein models[12–17] have lent support to the idea that cycling may enhance folding rates. The unfolding of the substrate may result from the rotation of the apical domain of GroEL upon binding of its co-chaperone GroES. It remains unclear, however, whether the mechanical deformation of the protein plays a significant role in the enhancement of folding rates observed in chaperonin mediated folding. While a large number of proteins show folding rate accelerations in the presence of GroEL, forced unfolding has only been clearly demonstrated for a single substrate,[7] leaving open the possibility that other factors may be responsible for chaperone activity.

The alternative theory for chaperonin assisted folding, the ACM, posits that inefficient folding in the cytosol is due to protein aggregation rather than to the formation of long-lived traps. Proteins that would otherwise fold unhampered in a dilute environment can encounter various obstacles in the crowded cellular milieu. Under conditions of high macromolecular concentration, inter-protein interactions become increasingly relevant[18] and proteins may irreversibly aggregate through mutually exposed hydrophobic facets. If this occurs, the folding process may be brought to a complete halt as chaperones are generally incapable of driving aggregate species to their monomeric native state.[19] Increasing experimental evidence points to the fact that protein aggregation rather than the formation of intermediates may be responsible for the observed slow folding kinetics of certain proteins.[20,21]

The ACM stipulates that chaperonins "rescue" proteins by binding them before they have a chance to aggregate. Once encapsulated, the substrate has the possibility of folding autonomously, sheltered from other proteins. Recent experiments by Hartl and co-workers[10] have examined a number of key issues related to the ACM. In particular, they demonstrated that not only does folding to the native state take place entirely within the chaperonin cavity, but also that ATP-driven chaperonin cycling is not required for efficient folding. Mutant non-cycling single-ring GroEL chaperonins (present at sufficient concentrations) were able to convert misfolded conformations into the native one at the same rate and with the same yield as did a cycling GroEL/ES complex. We note that it has been argued[22] that the folding of Rubisco in GroEL is a limiting case of the IAM in which only a single binding event is required to fold the protein. Another significant finding of the Hartl study is that encapsulation in the chaperonin cavity only yields a folding rate enhancement for certain proteins. Confinement in the GroEL cavity produced a fourfold rate acceleration compared to a bulk solution for Rubisco, while no rate enhancement was observed for the smaller protein rhodanese. The authors conjecture that folding rate enhancement of certain GroE substrates may be linked to the influence of the confined environment on the energy surface of the protein.[1,10] Specifically, they propose that encapsulation in the hydrophilic cavity of chaperonins can eliminate local energy minima thereby reducing the roughness of the corresponding free energy surface and accelerating folding rates. Little supporting evidence for this conjecture has however been put forth, as the Hartl hypothesis is extremely difficult to test experimentally.

Here, we examine the effect of encapsulation on protein folding by studying the folding of model proteins confined to hydrophilic cages. We consider a 27 residue off-lattice protein model designed to fold to an α/β sandwich, a typical motif of substrate proteins bound by the GroE chaperonin. The protein is represented by a bead-and-spring minimalist model[23] with the beads representing amino acids placed at the C^{α} positions and fixed bonds among the beads (Figure 1). The model contains three types of residues: mutually

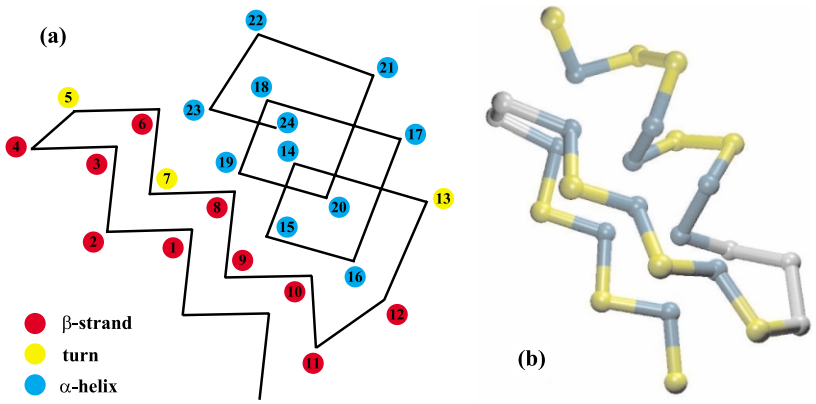

Figure 1. Native state of the α/β model considered in the present work in (a) schematic and (b) ball-and-stick representations. (a) Torsions are enumerated and color-coded by secondary structural element: β-strands are represented in red, turns in yellow and the α-helix in cyan. (b) Hydrophobic residues are colored in green, hydrophilic residues in yellow and neutral residues in gray.

attractive hydrophobic, repulsive hydrophilic and neutral residues. The hydrophilic environment of the chaperonin cavity was mimicked by confining the protein to a repulsive sphere of varying radius R.

Molecular dynamics simulations were performed at different temperatures and for different confining sphere radii to study the effect of confinement on the thermodynamics and kinetics of folding. The results of these simulations are presented in Results and Discussion. The effects of confinement on the folding energy landscapes are discussed and an explanation is provided for the rate enhancement observed for certain proteins upon encapsulation. Details of the model and the simulation methodology are given in Methods and Model.

Results and Discussion

Bulk *versus* cage folding

Simulations were performed in an extremely diluted regime (hydrophilic sphere radius $R = \infty$, referred to as the "bulk") as well as at five finite values of the cage $R/R_{max} = 1.44$, 1.55, 1.66, 1.88, 2.10. The quantity R_{max} denotes the maximum separation of any monomer from the center of mass of the native state of the model, i.e. the minimum radius of a cage that can accommodate the native conformation. Cage radii smaller that $1.44\,R_{max}$ ($R < 1.44\,R_{max}$) lead to substantial distortions of the native state and were not used in this study. For all cage radii considered, the RMSD distances among the native states did not exceed 0.1 Å.

Thermodynamics of folding

The effect of confinement of the chaperonin cage on the thermodynamic properties of the protein substrate were investigated through extended molecular dynamics simulations over temperatures ranging from elevated temperatures at which the protein is predominantly unfolded to lower temperatures at which the protein populates the native state. Details of the simulation methodology are given in Methods and Model.

The thermodynamics of folding were first characterized under bulk folding condition (i.e. in the absence of the cage). The protein is observed to exhibit a collapse transition (as determined from the fluctuations in the energy) at 300 K from a random coil state to a collapsed misfolded state. This is followed by a folding transition at 240 K from the compact misfolded state to the native state. The folding transition temperature was determined from the fluctuations in χ. The parameter χ measures the structural difference between a given conformation and the native state. For states very close to the native conformation, the structural overlap approaches unity while in fully unfolded conformations, χ is close to zero. The free-energy surface calculated at the

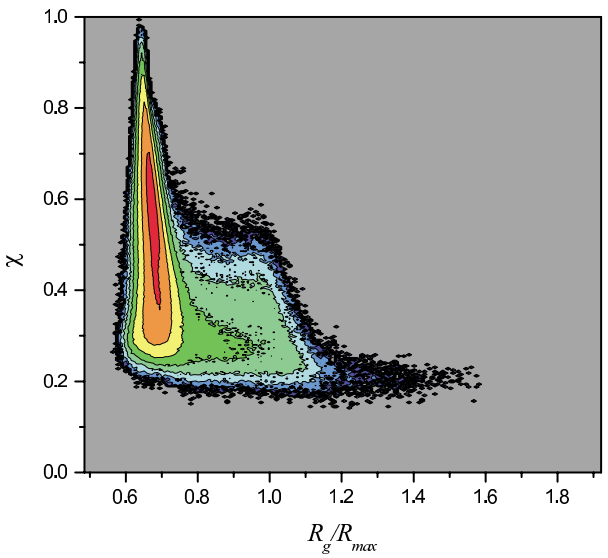

Figure 2. Free energy map of the α/β model in the bulk case as a function of the radius of gyration R_g and the structural overlap function χ at the folding temperature T_f. The regions colored in red correspond to the highest occupation probability. The free energy surface has an L-shape, indicative of a folding mechanism involving an initial collapse to a globular state, followed by a rearrangement to the native state.

folding temperature T_f is shown in Figure 2. It has a characteristic L-shaped structure[24] reflective of the two-stage folding mechanism involving an initial collapse to globular conformations and subsequent folding from the collapsed state. From initial expanded conformations ($R_g/R_{max} \sim 1$), the protein quickly collapses to states with $R_g/R_{max} \sim 0.7$ and stays there until folding is complete. As is clear from the relative populations of the states in the $R_g-\chi$ plane, folding almost entirely takes place within the collapsed state. The random coil conformations are seldom visited at the folding temperature T_f.

The nature of the collapse transition upon encapsulation can be gleaned from the behavior of the radius of gyration R_g of the substrate with temperature (Figure 3(a)). The size of the protein drops by a factor of nearly 2 from an initial value of $1.1\,R_{max}$ at 400 K when the temperature is reduced to $T \sim 250$ K. The region of fastest variation in R_g coincides with the position of the collapse temperature, T_c, computed from the maximum of the specific heat as a function of temperature. Figure 3(a) clearly shows the effects of confinement on the model's structure. As expected, the radius of gyration of the unfolded ensemble decreases with the cage radius R. Figure 3(b) displays the temperatures of the collapse and folding transitions as a function of the cage radius R. The collapse temperature increases gradually upon encapsulation from its bulk value of 300 K to 380 K for $R/R_{max} = 1.44$. The increasing values of T_c can be linked to the changes in the entropy of the unfolded ensemble brought about by confinement.[25] The temperature T_c is determined under the condition that the free energies of the random coil and collapsed states are equal. Considering the repulsive character of the confining potential one expects that the cage wall will not strongly influence the potential energies of either state. The entropy of the model, on the other hand, particularly for the coil-like, expanded conformations, will be affected by the presence of the cage. Since the primary role of confinement is to limit conformational freedom at high temperatures, the entropy of the random coil state is expected to drop when the cage confines are imposed. As a result, the smaller entropy gap between the collapsed and coil states in a confined environment will require higher temperatures to balance the energy gap (which is unaffected by the cage).

Interestingly, the folding temperature, T_f, computed from the maximum of the variation in χ, does not depend significantly on the confinement radius (Figure 3(b)) and remains nearly constant around its bulk value of $T_f = 240$ K. While this result may at first appear counter-intuitive, an explanation is provided by the free energy surface plotted as a function of χ and R_g at T_f (Figure 2). The design of the model ensures a hydrophobic collapse causing a dramatic reduction in the protein size in the initial stages of folding. The confining sphere can hence have no influence on the protein structure unless its radius is smaller than the size of the collapsed globule. Radii that are too small can, however, strongly distort the native state of the model and were not considered in this study.

Kinetics of folding

Details of the folding kinetics simulations can be found in Methods and Model. The folding time as a function of temperature and radius of confining sphere are plotted in Figure 4. The curves have a "V" shape resulting from two regions of strong growth in τ_f, at high and low temperatures. At high temperatures the statistical weight of the native conformation is low compared to that of expanded conformations of the unfolded ensemble, hence the long folding times. On the

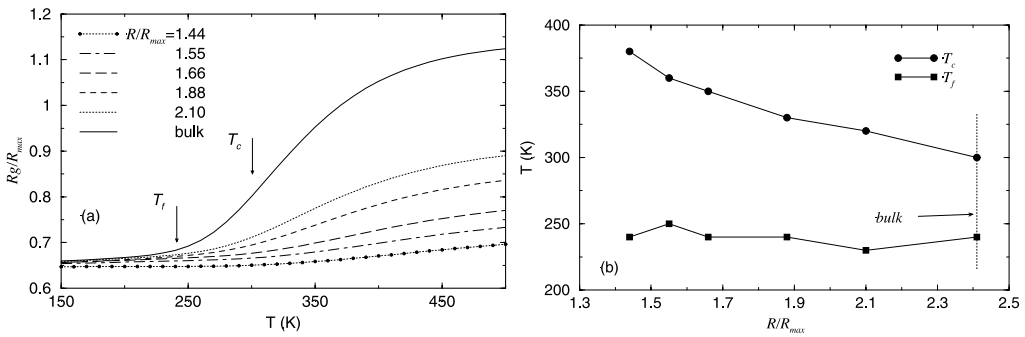

Figure 3. Thermodynamic functions of the α/β model. (a) Radius of gyration as a function of temperature plotted for several values of the confinement radius R. The bulk case is represented by a continuous line, while the radii of gyration for the encapsulated protein are represented by a broken line. The radius of gyration of the unfolded ensemble decreases as smaller cage radii are imposed. The collapse $T_c = 300$ K and folding $T_f = 240$ K temperatures for the bulk regime are indicated in the plot. (b) Collapse and folding temperatures as a function of the cage radius. R. The bulk values of the collapse and folding transitions are denoted by the vertical broken line. The collapse temperature, T_c, increases upon encapsulation as a result of the changes in entropy of the unfolded state. The folding temperature, on the other hand, does not change significantly with cage radii. This is a result of the folding mechanism that involves an initial collapse of the protein to a compact globular structure prior to the folding transition.

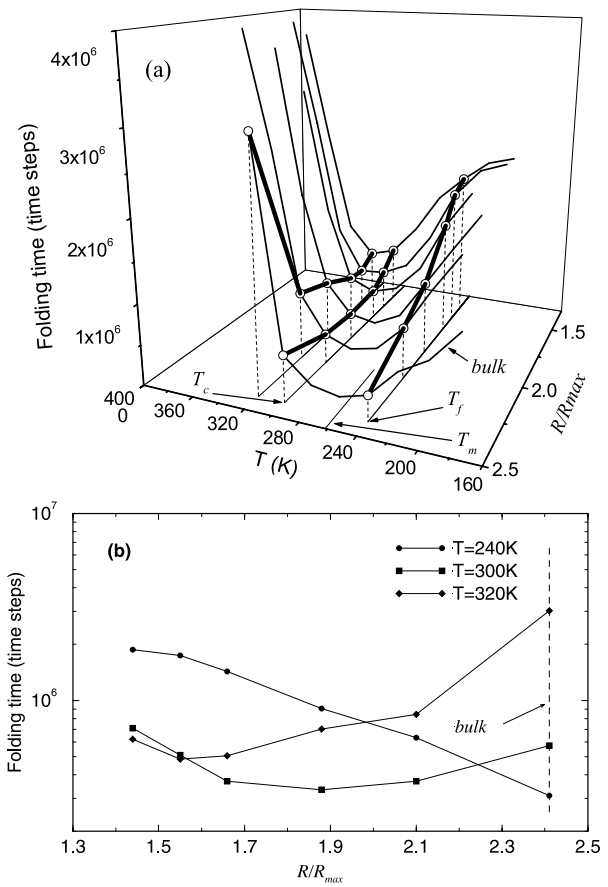

Figure 4. Folding time of the α/β protein model studied in the present work. (a) Folding time as a function of temperature T and the confinement radius R. The folding time curve corresponding to the bulk case is indicated as a reference point. Confinement does not affect the V-shape of the folding curve. The increased folding time at high temperatures is a result of the low statistical weight of the native state, whereas the long folding times at low temperatures come from the trapping of protein conformations in local energy minima. The bottom of the V corresponds to the minimum folding time temperature, T_m, at which the thermodynamic drive to the native state balances the trapping in local energy minima. The collapse temperature, T_c, the folding temperature, T_f, and the minimum folding time temperature, T_m, are indicated. For this protein, T_f lies below T_m. (b) Folding time for selected temperatures ($T = T_f = 240\,\mathrm{K} < T_m$, $T = 300\,\mathrm{K}$ and $T = 320\,\mathrm{K}$) as a function of the cage radius R. Bulk folding time values are indicated by the broken line. Encapsulation only yields folding rate enhancements at temperatures above T_m. Rate accelerations are hence not observed for temperatures at which the protein is thermodynamically stable ($T_f < T_m$ for all cage radii).

other hand, at low temperatures, the statistical weight of the native state is high but the protein dynamics are now dominated by transient trapping in local energy minima. The escape time from the minima, and hence the folding time, grows when the temperature is decreased. These two mutually opposing forces, the thermodynamic drive toward the native state and the slow structural dynamics due to local minima, balance each other at some intermediate temperature T_m where the protein folds faster than at any other temperature.[26–28] For our model in bulk conditions, this minimum folding temperature is $T_m = 260\,\mathrm{K}$.

It is clear from Figure 4(a) that confinement neither alters the "V" shape of the folding curve, nor significantly affects the minimal folding time. On the other hand, the minimum folding-time temperature, T_m, is strongly affected by the presence of the chaperonin cage. As the cage radius R decreases, T_m is monotonically shifted toward higher temperatures. From its value of 260 K in the regime of infinite dilution it gradually increases to 320 K when the model is confined to a cage of radius $R/R_{max} = 1.44$.

From a practical standpoint, it is important to understand how encapsulation in a chaperone cavity affects substrate folding times at biologically relevant temperatures. On the basis of the behavior of T_m with cage radius R, we predict that at temperatures below the minimal folding-time temperature ($T < T_m$), confinement will slow folding, regardless of the size of the cage. At low temperatures, compact structures tend to be favored over higher energy extended conformations and the protein spends most of its time trapped in collapsed misfolded states. Confinement does nothing to discourage the protein from adopting these compact misfolded conformations and in some cases may even inhibit escape from these traps. Hence, in the low temperature limit confinement is more likely to hinder than accelerate folding. We note that the possibility of rate deceleration in chaperone mediated folding has also been predicted in the context of the IAM.[5,15,29] On the other hand, at temperatures above T_m ($T > T_m$), we expect folding to be accelerated for R values above a certain threshold cage radius, and decelerated below. This scenario is illustrated in Figure 4(b), where the folding time is plotted as function of R at selected temperatures $T_f = 240\,\mathrm{K} < T_m$, $T = 300\,\mathrm{K}$ and 320 K. At the folding temperature, T_f, ($T_f < T_m$), the folding time inside the chaperonin cage is always longer than in the bulk. At higher temperatures (above T_m), an acceleration of folding rates is observed at cage radii ranging from $2.1\,R_{max}$ to $1.5\,R_{max}$. The precise value of the cage radius where folding enhancement is optimal depends on the temperature considered. The importance of temperature effects on the folding of frustrated sequences in confined environments has been noted in an earlier study.[15]

Since the folding temperature of this model lies below the minimum folding-time temperature ($T_f < T_m$), it is apparent that encapsulation alone in the GroE cavity will not lead to an enhancement in folding rates under physiological conditions ($T < T_f$). Confinement for this protein under experimental conditions where the native state is stable will play primarily a role in shielding it from aggregation in concentrated solutions. Clearly, a folding rate enhancement is only of

practical use in the cell at temperatures where the protein is in a thermodynamically stable state. In light of our observation that the folding temperature, T_f, does not change much upon encapsulation and that a folding acceleration can only be achieved for temperatures above T_m ($T > T_m$), the only possible way for a protein to exhibit a rate enhancement under biologically relevant conditions is, if its minimum folding-time temperature lies below the folding transition temperature ($T_m < T_f$).

This argument raises a number of questions of theoretical interest. First, how is the temperature of minimal folding time, T_m, related to the intrinsic characteristics of proteins (in particular to their energy landscapes)? Second, is it possible to design a potential energy function for a model protein such that its T_m is lower than its folding temperature T_f?

In what follows we show that the relationship between T_f and T_m can be related to the levels of frustration in the potential energy surface. In particular, we show that by constructing proteins with smoothed energy surfaces, the folding transition temperature can be moved above the minimal folding-time temperature.

Protein model with folding rate enhanced due to the confinement effects

At present, there are no exact analytical treatments capable of relating the characteristic transition temperatures of model proteins (T_c and T_f, for instance) to their potential energy functions. While we cannot establish an exact relationship between the energy function and the transition temperatures, we can, nonetheless qualitatively relate the transition temperatures to the underlying energy landscape. Studies by Wolynes[30,31] and Thirumalai[32–34] have shown that the location of the folding transition temperature (T_f) is intrinsically tied to the degree of frustration of the protein. Proteins with very frustrated sequences have small ratios of folding to glass transition temperatures (T_f/T_g) and large values of the parameter $\sigma = (T_c - T_f)/T_c$ which relates the collapse transition temperature and folding transition temperature. On the basis of the σ parameter,[35] reducing the level of frustration of the model should lead to a folding transition temperature that approaches more closely the collapse transition temperature. Since T_c will be mostly unaffected by a change in level of frustration of the model,[36,37] T_f must increase in value when the degree of frustration is reduced. We demonstrated this feature in an earlier work where we showed that in the Go-model limit (i.e. a protein with minimal energetic frustration), the folding and collapse transitions occur concurrently.[36] A second variable affected by the degree of frustration is the minimum folding time T_m. The location of T_m results from the subtle interplay between the thermodynamic drive to the native state and kinetic trapping. A reduction in the degree of frustration of the energy surface will cause the dynamical trapping mechanism to become dominant at a lower temperature, resulting in a lower value for the temperature T_m.

A common method to remove frustration levels in computer simulations of proteins is by designing Go-type models.[38–40] In a Go-type model, only pairs of hydrophobic residues that are in contact in the native state experience attractive interactions. Hydrophobic residues that participate in non-native contacts do not contribute to the potential energy, even if they approach each other at short distances. This allows a protein to avoid states that are structurally dissimilar to the native state due to wrong hydrophobic contacts, but that nonetheless have relatively low potential energies.

Here, we follow an alternative avenue to construct less frustrated protein models. We note that due to the spherical symmetry of the hydrophobic force, local minima created by wrong contacts correspond to structures with globular shapes. This means that some structural elements of the model that are designed to be extended in their lowest energy state, such as β strands for instance, will have to be bent somewhere in their middle in order to fit to the perfectly globular shape. Since this bending will come at the expense of the torsional part of the potential it should be possible to remove the local minima caused by wrong hydrophobic contacts by strengthening the constants of the torsion potential. The difference between this approach and the Go-style approach is that instead of elevating the energy of conformations with wrong contacts by canceling some of the contributions of the hydrophobic force, the energy of these conformations is raised due to the contributions of the torsion potential. The values of the torsional constants a and b for the modified model are given in Table 1 with those of the original model in parentheses. Details on the torsional potential can be found in Methods and Model.

Thermodynamics and kinetics of folding of the less frustrated protein model

The thermodynamics and kinetics of folding of the modified protein were investigated in the manner described for the original model. The less

Table 1. Parameters of the torsion potentials used for the model with smoothed energy surface

Constants	α-Helix	β-Strand	Turn
a (a.u.)	3.5×10^{-3} (1.0×10^{-3})	3.0×10^{-3} (1.5×10^{-3})	3.5×10^{-3} (1.5×10^{-3})
b (a.u.)	2.0×10^{-3} (2.0×10^{-3})	3.0×10^{-3} (2.0×10^{-3})	N/A

N/A, not applicable.

As in the original model the potentials for the 6th, 7th and 13th torsions were taken to be zero. Values for the original model are indicated in parentheses.

frustrated (modified) protein exhibits a similar folding mechanism to the original model, with an initial collapse followed by a rearrangement to the native state. The collapse temperature, T_c, is seen to increase upon encapsulation from a bulk value of 340 K to 440 K for a cage radius of $R/R_{max} = 1.45$, while the folding transition temperature T_f remains unaltered from its bulk value of 280 K.

The folding free energy surface in the $R_g - \chi$ plane is shown in Figure 5 at $T = T_f$. The free energy surfaces for the original (Figure 2) and modified (Figure 5) protein models both have similar L-shaped appearances. The only significant difference between the two is a higher population of the random coil ensemble in the less frustrated model. This is consistent with the fact that increasing the strength of the torsional potential should lessen the tendency of the model to collapse in a non-specific manner.

The kinetics of folding present a more striking difference between the two models. The mean first passage time computed in these simulations is shown in Figure 6(a) as function of temperature, T, and cage radius, R. The first apparent difference lies in the significantly shorter minimum folding time of 120×10^3 for the bulk modified model as opposed to 220×10^3 in the bulk original model. In addition, for all cage radii, the minimum folding-time temperature for the modified model is smaller than the folding temperature ($T_m < T_f$), in agreement with our earlier prediction. While the folding temperature is unaffected by encapsulation, the minimum folding-time temperature, T_m, increases as the cage radius is reduced.

The folding times at T_f are plotted as a function of cage radius in Figure 6(b). In contrast to the original model, a rate enhancement is observed at this temperature upon encapsulation. A maximum rate enhancement of nearly 1.6-fold is observed at $R/R_{max} = 2.07$. Interestingly, a similar enhancement rate was obtained in an earlier study of a Go-model β sheet model,[41] i.e. a model with highly reduced levels of frustration. In agreement with our original hypothesis, rate enhancements can be induced by confinement under physiological conditions when the model has the property that its minimal folding-time temperature T_m is smaller than the folding transition temperature T_f.

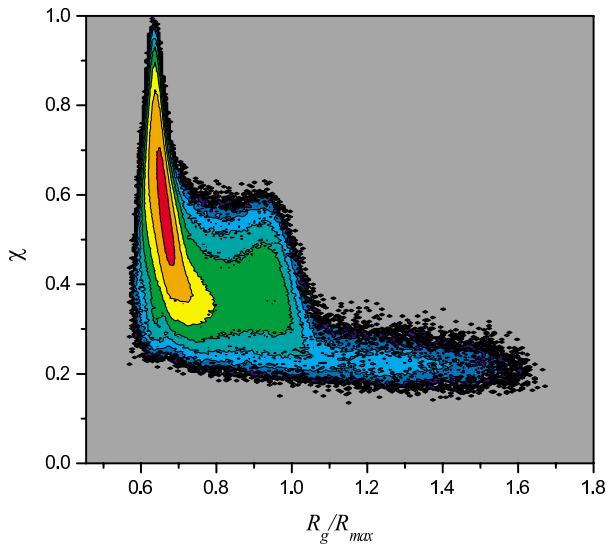

Figure 5. Free energy map of the less frustrated α/β model in the bulk case as a function of R_g and χ, calculated at the folding temperature T_f. The regions in the map colored red correspond to the highest occupation probability. The surface has a similar L-shape to the free energy surface of the original model (Figure 2), but the random coil ensemble is more populated.

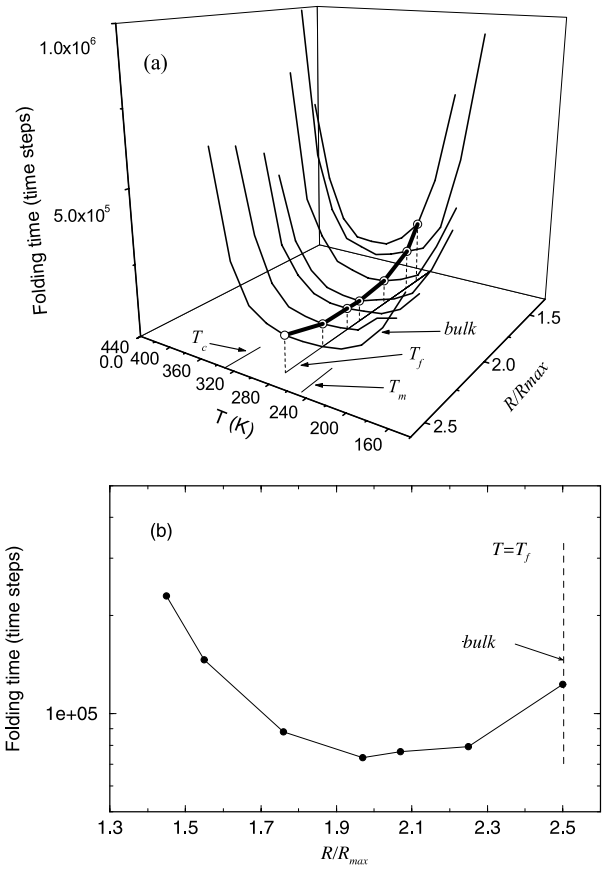

Figure 6. Folding time of the less frustrated α/β model. (a) Folding time as a function of temperature T and confinement radius R. The folding curve corresponding to the bulk regime is included as a reference. The locations of T_c, T_f and T_m are indicated on the plot. The less frustrated protein folds significantly faster than the original model ($\tau_f = 120 \times 10^3$ for the bulk modified model and $\tau_f = 220 \times 10^3$ for the original model). In the modified model, T_m is lower than the folding temperature T_f for all radii considered. (b) Folding time at the folding temperature T_f as a function of the cage radius R. In contrast to the original model, a folding rate enhancement is observed at this temperature. A maximum folding acceleration of 1.6-fold can be achieved of the present model through encapsulation.

Reducing the roughness of the energy landscape through encapsulation

In the previous section, we showed that we can

modify the potential energy function of the model in such a way that T_m becomes lower than T_f. Proteins with this property exhibit an increase in folding rates upon encapsulation. In order to understand the physical mechanism that leads to this rate enhancement, we investigated in detail the pathways adopted by the protein from the unfolded state to the native state. These pathways are computed as distribution over the conformations that the protein follows during the course of the folding simulations. Folding maps at T_f for the modified bulk protein and for the protein confined to a sphere of radius $R/R_{max} = 2.07$ (the radius at which maximal rate enhancement was observed) are shown in Figure 7(a) and (b), respectively. The maps were generated from conformations saved every 200 time-steps during the 500 folding simulations.

The folding maps (Figure 7(a) and (b)) differ from the free energy map (Figure 5) in the significant respect that the former corresponds to a non-equilibrium property while the latter is calculated under equilibrated conditions. The utility of the folding map lies in providing information on where the protein spends the majority of its time en route to the native state.

A closer look at the folding maps reveals a dramatic difference in the states favored by the protein in the bulk and confined cases. The folding map in the infinite dilution regime presents two maxima, a first one corresponding to the population of fairly extended states ($R_g/R_{max} = 1.2$, $\chi = 0.2$) and a second one associated with more collapsed states ($R_g/R_{max} = 0.9$, $\chi = 0.3$). These maxima disappear in the confined case and are replaced by a new maximum at low radius of gyration ($R_g/R_{max} = 0.7$, $\chi = 0.4$). Clearly, the protein no longer significantly populates conformations with large radius of gyration in the confined environment, and this in turn affects folding rates. In order to understand the impact of confinement on the folding rates of certain proteins, it is instructive to look at folding in the context of local potential energy minima. Since the energy required to break a hydrophobic contact exceeds kT, such potential energy minima are likely to present barriers that can trap the protein and thus decrease folding rates.

To locate where the protein spends most of its time during a folding event, we selected conformations at regular time intervals from folding simulations and generated the corresponding local minima by the steepest decent method. Maps of these local energy minima are present at the folding temperature, T_f, as a function of the radius of gyration, R_g, and the maximum extension radius, R^s_{max}, in Figure 8(a) and (b), respectively. These maps show both the location of the energy minima as well as minima which are preferentially visited during folding simulations.

The regions that are most frequently visited, i.e. those that impact folding times, are apparent in the clustering in Figure 8. In both Figure 8(a) and (b), the energies corresponding to the original model are represented in blue and those corresponding to the modified (less frustrated) model in yellow. The energy scale for the original model is given on the bottom axis and the energy scale for the modified model on the top.

In the bulk regime, the difference in the relevant minima for the original (frustrated) and modified (less frustrated) protein is already significant. The minima that serve as folding traps for the original model appear uniquely at near native R_g and hence correspond to compact structures. On the other hand, in the modified model, two major lustering regions are present at both high and low R_g. The most populated minima occur at energies and radii of gyration of $e = 0.02$, $R_g/R_{max} = 1.2$ and at $R_g/R_{max} = 0.9$. These cluster regions correspond to the minima observed in the folding map in Figure 7(a). The modified model in the bulk environment hence appears to visit local minima conformations that have a much more

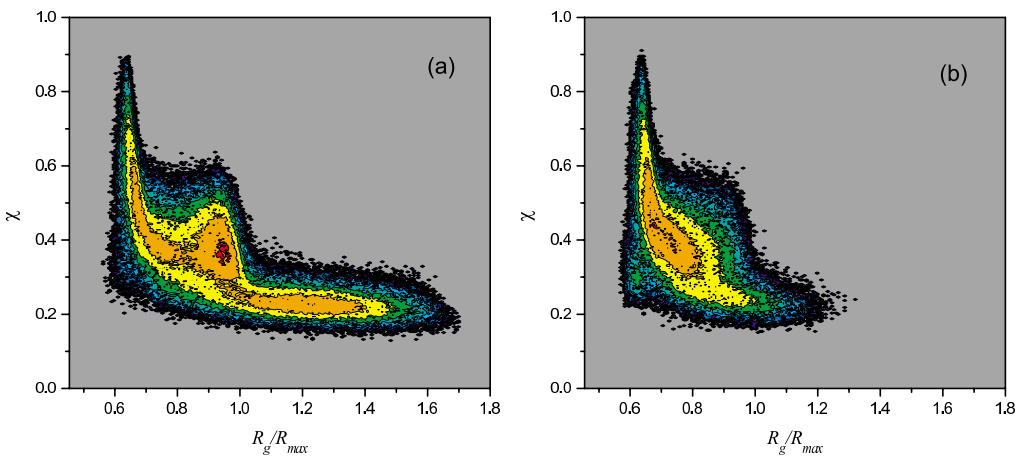

Figure 7. Folding map of the less frustrated α/β model calculated at the folding temperature T_f. In (a) the bulk conditions and (b) confined to the sphere of radius $R/R_{max} = 2.07$. Regions in red and orange correspond to those most frequently visited. Extended conformations are no longer visited in the confined case.

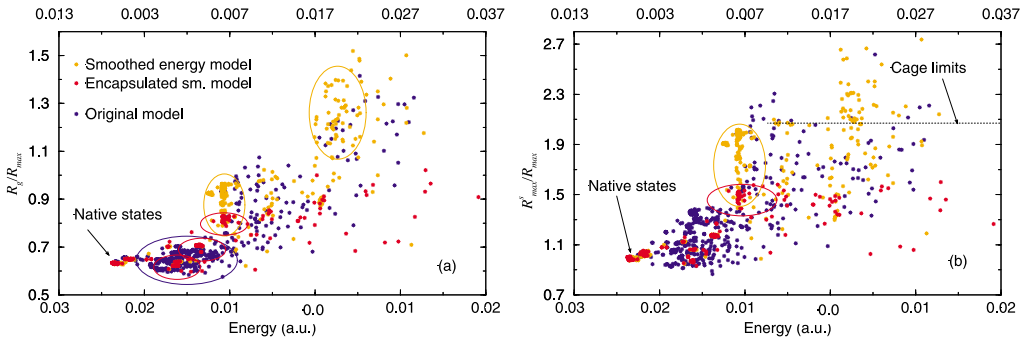

Figure 8. Maps of local potential energy minima visited during folding simulations of the studied α/β models as a function of (a) the radius of gyration R_g/R_{max} and (b) the maximum extension radius R^s_{max}/R_{max}. The energy scale listed on the upper *x*-axis corresponds to the modified model with reduced frustration and the energy scale on the lower *x*-axis to the original model. The energy minima visited by the original protein in the bulk case are colored blue and those visited by the modified protein in the bulk case in red. The minima sampled by the modified protein encapsulated in a cage of radii $R_g/R_{max} = 2.07$ are colored orange. For the modified model, the encircled regions in the plots coincide with the highest populated regions of the folding maps in Figure 7(a) and (b). The cage boundaries are indicated in (b) by the horizontal broken line. Once the cage boundaries are imposed, all minima with R^s_{max}/R_{max} greater than 1.7 (cage radius minus σ) are eliminated.

extended shape than those seen in the original model.

The minima visited by the modified protein confined to a sphere of radius $R_{max} = 2.07$ (the radius at which maximum rate enhancement is observed) are represented in red in Figure 8. The effects of the cage are dramatic: all minima present in the bulk that exceed a certain R_g value are eliminated. This effect can be seen more clearly when the minima are plotted in Figure 8(b) as a function of the maximum extension radius R^s_{max} of the protein conformation. This plot enables us to make a direct comparison between the conformations corresponding to each minima and the size of the cage. The cage boundaries are indicated in Figure 8(b) by a broken line. In contrast to the bulk original model (blue), a significant number of minima sampled by the bulk modified protein (yellow) correspond to conformations with R^s_{max}/R_{max} values close to or greater than that of the confining sphere. Once the cage limits are imposed, all minima with R^s_{max}/R_{max} values greater than 1.7 (i.e. the cage radius minus the distance σ at which the monomers can approach the sphere) are eliminated. We note that the local energy minima correspond to conformations of more or less globular shape with incorrect hydrophobic contacts. Strong local interactions may lead to conformations with a few monomers protruding from the globule. It is sufficient for the cage walls to strongly repel only one such monomer for the energy of the whole minimum to be elevated and the structure destabilized. The folding acceleration is hence achieved in the confined environment because a subset of local minima, which would otherwise be frequently visited, are now inaccessible. This mechanism of folding rate enhancement is consistent with the proposal by Hartl and co-workers[10] that the confines imposed by the chaperonin are able to reduce the roughness of the free energy surface of some proteins by eliminating energy minima that can serve as kinetic traps for folding.

Conclusions

Here, we examined the role of confinement in chaperone assisted protein folding and proposed a novel relationship between the degree of frustration of the substrate protein and the corresponding effect of encapsulation on its folding mechanism and rate.

We used a simplified representation of both the protein and the chaperone, which rendered the problem computationally tractable, but nonetheless allowed us to capture the essence of folding in a cage. The chaperonin cage was modeled by a repulsive sphere that mimicked the hydrophilic lining of the GroEL cavity post ATP and GroES binding. Experimental studies by Hartl have shown that certain proteins do not require multiple rounds of binding/release in order to exhibit rate enhancements.[10] Such proteins only exhibit a single binding event to the chaperonin and fold entirely inside its cage. Our study has focused on understanding how confinement affects the folding of proteins after this single binding event has occurred. We first considered a protein with a high-degree of frustration, arising from allowed non-specific hydrophobic interactions. This protein folded in a two-step manner, with an initial collapse to a compact non-native state at the collapse temperature, T_c, followed by a rearrangement to the native state at the folding temperature, T_f. The dependence of folding time on temperature was V-shaped for all cage radii considered, reflective of two regions of growth in folding time at high and low temperatures. Encapsulation did not affect the folding transition temperature of this protein, but shifted the minimum folding-time temperature, T_m, to higher values. In all instances, the folding transition temperature lied below T_m, implying that encapsulation would not lead to enhanced folding rates under conditions where the native state is thermodynamically stable.

In order to observe rate enhancement under biologically meaningful conditions, it is necessary for the protein to exhibit a folding transition above the temperature of minimal folding time. We designed a model with this property by increasing the value of the torsional potential of our original model. This modified protein hence had the same topology as the original protein, but a different interaction scheme. This design leads to a less frustrated folder by destabilizing conformations with incorrect local hydrophobic contacts. Analysis of maps of local energy minima revealed that minima that trapped the protein in the bulk regime were eliminated upon encapsulation. These results offer support to the recent hypothesis of Hartl that encapsulation increases substrate folding rates in GroEL by reducing the roughness of the energy landscape for folding.[10]

Methods and Model

Off-lattice protein model

The majority of the numerical studies on proteins confined to cavities[14,15] have involved Monte Carlo simulations of lattice protein models. While the reduced description of the protein, compared to fully atomic models, renders the system computationally tractable, these simulations are not well suited for kinetics studies. Indeed, kinetic sequences of events generated using lattice models depend very strongly on the adopted Monte Carlo move set.[42,43] In confined environments, the problem is further aggravated by the repulsive wall that can lock the lattice model into a single conformation thereby leading to breakdown of ergodicity. In order to deal with this situation, special move sets, which can translate the protein as a whole need to be introduced.[15] The influence of these specific move sets on the final folding results is not clear.

On the other hand, molecular dynamics simulations of off-lattice protein models are devoid of such move-set complications. Additionally, they offer a more realistic description of the protein in terms of the treatment of the excluded volume affect of the monomers.

Interactions among the monomers of the model

Here, we use a variant of the bead-and-spring protein model originally introduced by Honeycutt and Thirumalai.[23] Within the model, each entire amino acid residue is represented by a single, interacting bead. The distances between consecutive beads are kept constant. There are three types of residues in the model: hydrophobic (H), hydrophilic (P) and neutral (N). The hydrophobic residues separated by at least one residue along the sequence attract each other through the Lennard–Jones potential:

$$U_{\mathrm{HH}}(r) = 4e_{\mathrm{h}}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{1}$$

The neutral residues interact with the residues of any type by purely repulsive potential:

$$U_{\mathrm{Ni}}(r) = 4\frac{e_{\mathrm{n}} + e_{\mathrm{i}}}{2}\left(\frac{\sigma}{r}\right)^{12}, \quad i = H, N, P \tag{2}$$

To reproduce the experimentally observed presence of hydrophilic residues on the protein surface, we impose a stronger repulsion between two hydrophilic residues than between a hydrophilic and a hydrophobic residue:

$$U_{\mathrm{PP}}(r) = 4e_{\mathrm{p}}\left[\left(\frac{\sigma}{r}\right)^{12} + \left(\frac{\sigma}{r}\right)^{6}\right] \tag{3}$$

$$U_{\mathrm{PH}}(r) = 4\frac{e_{\mathrm{h}} + e_{\mathrm{p}}}{2}\left(\frac{\sigma}{r}\right)^{12} \tag{4}$$

In the above equation, the parameters measuring the strength of the interactions were taken to be $e_{\mathrm{h}} = 3\times10^{-3}$ a.u., $e_{\mathrm{n}} = 1\times10^{-3}$ a.u. and $e_{\mathrm{p}} = 5\times10^{-4}$ a.u. The parameter σ, which is responsible for the spatial extent of the potential, was taken to be equal to the length of peptide bonds (3.8 Å). Overall, the set of the hydrophobic/hydrophobic potentials described above is very similar to many other potential schemes found in the literature.[44]

Besides long range interactions operating among the monomers there are also two types of the short range interactions, bending and torsion potentials, intended to limit the local conformational freedom of the model. The bending potential is expressed in terms of the bond angle Θ formed by three consecutive beads along the chain:

$$U_{\mathrm{B}} = \frac{k}{2}(\Theta - \Theta_0)^2 \tag{5}$$

where the potential constant k was taken to be $4\times10^{-2}\times$ a.u./rad^2 and the equilibrium bond angle was $\Theta_0 = 150°$. Torsion potentials involve four monomers along the sequence. Three bonds that connect these monomers form two non-equivalent planes. The torsion potentials are taken to depend on the angle formed by these planes. Since our model contains various secondary structure elements, we designed the torsion potential in a way that one of these elements, either β-strand, α-helix or turn, are favored. For β-strand segments the potential was taken to be:

$$U_{\mathrm{T}} = a(1 + \cos(\phi) + b(1 + \cos(3\phi))) \tag{6}$$

where constants a and b control the relative energy of the three minima of this potential: one *trans* and two *gauche* states. For α-helical conformations the potential is desired to favor one of the *gauche* state and thus has to be asymmetrical. The asymmetry is achieved by adding a sine term to the potential:

$$U_{\mathrm{T}} = a(1 - \cos(\phi) + \cos(3\phi)) - b(\sin(\phi) - 1) \tag{7}$$

Finally, in turn regions instead of *trans* or *gauche* states conformations with 90° torsions should be favored. Accordingly, we used the following potential for these regions:

$$U_{\mathrm{T}} = \begin{cases} 2a\cos^2(\phi), & 0 < \phi \le \pi \\ 2a(2 - \cos^2(\phi)), & \pi < \phi \le 2\pi \end{cases} \tag{8}$$

Note that the above potential has only one minimum at $\phi = 90°$ and not at an equivalent point 270°. This is to prevent the formation of incorrectly "wound" mirror images of β sheets which lead to the creation of a deep potential energy minimum, competing with the native state conformation.

The ground state of the present model is shown in Figure 1 in schematic and graphical representations. The schematic representation enumerates all torsions in the model. For example, the torsion encompassing mono-

mers 1, 2, 3 and 4 is assigned number 1 and so on. Additionally all the torsions are color-coded by secondary structure element. In total there are three turn regions in the model, ten β-sheet regions and 11 α-helical regions. The turn regions are found at the locations where the two β-strands join and where the α-helix is connected with the β-sheet. Specific values adopted for the parameter a and b of the torsion potentials are summarized in Table 2. The sequence of hydrophobic, hydrophilic and neutral residues is shown in Figure 1(b). The β-sheet contains alternating H and P monomers while the α-helix is composed of pairs of HH or PP monomers. Neutral monomers are placed at the turn regions. The net result of the monomer arrangement is that the protein possesses two hydrophobic facets: one created by the β-sheet segment and the other one by the α-helix. In the native state the mutually attractive facets are stacked back to back but in unfolded states they should, at least partially, get exposed to the solvent and other hydrophobic agents in the solution.[2] The hydrophilic residues are found mostly on the surface of the model, in agreement with experiment.[45]

Confining potential

Upon encapsulation, the substrate protein appears to be exposed to a mostly hydrophilic environment[1,2] within the GroEL activity. Several types of the confining potentials have been proposed[41,46] to model this hydrophilic cavity. In particular, it was noted[46] that the interactions between the encapsulated protein and the surrounding cavity should be short-ranged in character. As long as the protein is completely inside the confining sphere there should be no additional force exerted on its monomers. Accordingly, a short-range exponential potential was developed for the protein–cage interaction. In another recent paper[41] power-law repulsive potentials were considered and it was found that as long as the potential remains short-ranged, the exact value of the exponent used did not affect the final results. Here, we assume that all the monomers on the chain experience $1/r^{12}$ repulsive interaction coming from every element on the surface of the cage. When integrated over the entire spherical surface this interaction yields a final expression for the protein–cage potentials:[41]

$$U_c(\vec{r}) = 4e_p \frac{\pi R}{5r}\left[\left(\frac{\sigma}{r-R}\right)^{10} - \left(\frac{\sigma}{r+R}\right)^{10}\right] \qquad (9)$$

Here $\vec{r}$ is the radius-vector of any monomer of the model, measured from the position of the center of mass, and R is the radius of the cage. The total energy of the protein confined to a sphere is composed of all the terms mentioned above.

Table 2. Parameters of the torsion potentials used in the original model

Constants	α-Helix	β-Strand	Turn
a (a.u.)	1.0×10^{-3}	1.5×10^{-3}	1.5×10^{-3}
b (a.u.)	2.0×10^{-3}	2.0×10^{-3}	N/A

N/A, not applicable.
In addition to the listed values, the potentials for the 6th, 7th and 13th torsions were taken to be zero.

Simulation details

The folding process of the substrate protein confined to a spherical cavity was investigated using molecular dynamics method. More specifically, the monomer positions and velocities are advanced following the Langevin dynamics algorithm.[47] The use of this dynamics protocol serves a dual purpose. First, the trajectory generated in the simulations is sampled in the canonical ensemble, which corresponds with the constant temperature conditions of real experiments. Second, stochastic terms in the algorithm help to emulate the influence of the solvent molecules on the protein dynamics. In the present simulations, the only free parameter of the algorithm, the friction coefficient, was chosen so that the monomers experience about one-tenth of the frictional force exerted on a small amino acid residue such as alanine in water at room temperature. This choice of the friction parameter allows a considerable acceleration of the kinetic simulations while not significantly affecting the folding mechanism.[48]

Thermodynamics and kinetic simulations

The protein coordinates at each simulated temperature were recorded every 100 time-steps and used to construct potential energy histograms as well as joint histograms of the radius of gyration R_g and structural overlap function χ[33,49] with the potential energy. These histograms served as input in the multiple weighted histogram method[50] to evaluate all thermodynamic quantities of interest as functions of temperature: mean potential energy, radius of gyration, and the structural overlap. Initial starting structures for the kinetic folding simulations were obtained from long simulations at $T =$ 700 K, a temperature well above the collapse and folding temperature. To ensure statistical independence 500 conformations were selected over sufficiently long-time intervals. These conformations are considered representative of the unfolded state ensemble of the model and were used to initiate folding simulations at a given temperature T and confining radius R. During the simulations, the structural overlap function was constantly monitored. At points where χ reached 0.9 the simulations were stopped and the time elapsed from their start, the first passage time, recorded. The folding time of the model τ_f is taken to be an average over all 500 pre-recorded first passage times. The folding simulations were carried out for a maximum of 5×10^6 time-steps. If the system did not reach its native state over that time its first passage time was taken to be τ_{max}. Thus, folding times obtained in this study should be considered as a lower bound on the actual folding time of the model.

Acknowledgements

We thank Miriam Friedel for helpful discussions. This work was supported by the NSF Career Award #0133504.

References

1. Hartl, F. U. & Hayer-Hartl, M. (2002). Molecular

chaperones in the cytosol: from nascent chain to folded protein. *Science*, **295**, 1852–1858.
2. Saibil, H. R., Horwich, A. L. & Fenton, W. A. (2002). Allostery and protein substrate conformational change during GroEL/GroES-mediated protein folding. In *Advances in Protein Chemistry*, vol. 59, Academic Press, London.
3. Xu, Z., Horwich, A. L. & Sigler, P. B. (1997). The crystal structure of the asymmetric GroEL/GroES-$(ADP)_7$ chaperonin comples. *Nature*, **388**, 741–750.
4. Todd, M. J., Viitanen, P. V. & Lorimer, G. H. (1994). Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding. *Science*, **265**, 659–666.
5. Todd, M. J., Lorimer, G. H. & Thirumalai, D. (1996). Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. *Proc. Natl Acad. Sci. USA*, **93**, 4030–4035.
6. Lorimer, G. (1997). Folding with a two-stroke motor. *Nature*, **388**, 720–723.
7. Shtilerman, M., Lorimer, G. H. & Englander, S. W. (1999). Chaperonin function: folding by forced unfolding. *Science*, **284**, 822–825.
8. Ellis, R. J. (1994). Opening and closing the Anfinsen cage. *Curr. Biol.* **4**, 633–635.
9. Ellis, R. J. (2001). Molecular chaperones: inside and outside the Anfinsen cage. *Curr. Biol.* **11**, R1038–R1040.
10. Brinker, A., Pfeifer, G., Kerner, M. J., Naylor, D. J., Hartl, F. U. & Hayer-Hartl, M. (2001). Dual function of protein confinement in chaperonin-assisted protein folding. *Cell*, **107**, 223–233.
11. Weissman, J. S., Kashi, Y., Fenton, W. A. & Horwich, A. L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell*, **78**, 693–702.
12. Gulukota, K. & Wolynes, P. G. (1994). Statistical mechanics of kinetic proofreading in protein folding *in vivo*. *Proc. Natl Acad. Sci. USA*, **91**, 9292–9296.
13. Chan, H. S. & Dill, K. A. (1996). A simple model of chaperonin-mediated protein folding. *Proteins: Struct. Funct. Genet.* **24**, 345–351.
14. Sfatos, C. D., Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. (1996). Simulations of chaperone-assisted folding. *Biochemistry*, **35**, 334–339.
15. Betancourt, M. & Thirumalai, D. (1999). Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. *J. Mol. Biol.* **287**, 627–644.
16. Gorse, D. (2001). Global minimization of an off-lattice potential energy function using a chaperone-based refolding method. *Biopolymers*, **59**, 411–426.
17. Gorse, D. (2002). Application of a chaperone-based refolding method to two-and threedimensional off-lattice protein models. *Biopolymers*, **64**, 146–160.
18. Minton, A. P. (2001). The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* **276**, 10577–10580.
19. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. (1989). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature*, **342**, 884–888.
20. Nawrocki, J., Chu, R.-A., Pannell, L. K. & Bai, Y. (1999). Intermolecular aggregations are responsible for the slow kinetics observed in the folding of Cytochrome *c* at neutral pH. *J. Mol. Biol.* **293**, 991–995.
21. Silow, M. & Oliveberg, M. (1997). Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl Acad. Sci. USA*, **94**, 6084–6086.
22. Stan, G., Thirumalai, D., Lorimer, G. H. & Brooks, B. R. (2003). Annealing function of GroEL: structural and bioinformatic analysis. *Biophys. Chem.* **100**, 453–467.
23. Honeycutt, J. D. & Thirumalai, D. (1992). The nature of folded states of globular proteins. *Biopolymers*, **32**, 695–709.
24. Guo, Z. & Brooks, C. L., III (1997). Thermodynamics of protein folding: a statistical mechanical study of a small all-β protein. *Biopolymers*, **42**, 745–757.
25. Zhou, H.-X. & Dill, K. A. (2001). Stabilization of proteins in confined spaces. *Biochemistry*, **40**, 11289–11293.
26. Socci, N. D. & Onuchic, J. N. (1994). Folding kinetics of proteinlike heteropolymers. *J. Chem. Phys.* **101**, 1519–1528.
27. Lee, C.-L., Stell, G. & Wang, J. (2003). First-passage time distribution and non-Markovian diffusion dynamics of protein folding. *J. Chem. Phys.* **118**, 959–968.
28. Cieplak, M., Hoang, T. X. & Li, M. S. (1999). Scaling of folding properties in simple models of proteins. *Phys. Rev. Letters*, **83**, 1684–1687.
29. Thirumalai, D. & Lorimer, G. H. (2001). Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 245–269.
30. Bryngelson, J. D. & Wolynes, P. G. (1989). Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem*. **93**, 6902–6915.
31. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem*. **48**, 545–600.
32. Kilmov, D. K. & Thirumalai, D. (1996). Criterion that determines the foldability of proteins. *Phys. Rev. Letters*, **76**, 4070–4073.
33. Veitshans, T., Kilmov, D. & Thirumalai, D. (1996). Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold. Des.* **2**, 1–22.
34. Klimov, D. K. & Thirumalai, D. (1998). Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J. Chem. Phys.* **109**, 4119–4125.
35. Camacho, C. J. & Thirumalai, D. (1993). Kinetics and thermodynamics of folding in model protein. *Proc. Natl Acad. Sci. USA*, **90**, 6369–6372.
36. Shea, J.-E., Nochomovitz, Y. D., Guo, Z. & Brooks, C. L., III (1998). Exploring the space of protein folding Hamiltonians: the balance of forces in a minimalist β-barrel model. *J. Chem. Phys.* **109**, 2895–2903.
37. Nymeyer, H., García, A. E. & Onuchic, J. N. (1998). Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl Acad. Sci. USA*, **95**, 5921–5928.
38. Shea, J.-E., Onuchic, J. N. & Brooks, C. L., III (2000). Energetic frustration and the nature of the transition state in protein folding. *J. Chem. Phys*, **113**, 7663–7671.
39. Cheung, M., García, A. E. & Onuchic, J. N. (2002). Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur

after the structural collapse. *Proc. Natl Acad. Sci. USA*, **99**, 685–690.
40. Shea, J.-E., Onuchic, J. N. & Brooks, C. L., III (1999). Exploring the origins of topological frustration: design of a minimally frustrated model of fragment b of protein a. *Proc. Natl Acad. Sci. USA*, **96**, 12512–12517.
41. Klimov, D. K., Newfield, D. & Thirumalai, D. (2002). Simulations of β-hairpin folding confined to spherical pores using distributed computing. *Proc. Natl Acad. Sci. USA*, **99**, 8019–8024.
42. Chan, H. S. & Dill, K. A. (1993). Energy landscapes and the collapse dynamics of homopolymers. *J. Chem. Phys.* **99**, 2116–2127.
43. Chan, H. S. (1997). Modelling protein folding by Monte Carlo dynamics: chevron plots, chevron roll-over and non-Arrhenius kinetics. In *Monte Carlo Approach to Biopolymers and Protein Folding*, World Scientific, Singapore.
44. Thirumalai, D. & Klimov, D. K. (1999). Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr. Opin. Struct. Biol.* **9**, 197–207.
45. Creighton, T. E. (1992). *Protein Folding*, W. H. Freeman and Co., New York.
46. Friedel, M., Sheeler, D. J. & Shea, J.-E. (2003). Effects of confinement and crowding on the thermodynamics and kinetics of folding of an off-lattice beta barrel model. *J. Chem. Phys.* **118**, 8106–8113.
47. Allen, M. P. & Tildesley, D. J. (1986). *Computer Simulation of Liquids*, Clarendon Press, Oxford.
48. Klimov, D. & Thirumalai, D. (1997). Viscosity dependence of the folding rates of proteins. *Phys. Rev. Letters*, **79**, 317–320.
49. Baumketner, A. & Hiwatari, Y. (2002). Diffusive dynamics of protein folding studied by molecular dynamics simulations of an off-lattice model. *Phys. Rev. ser. E*, **66**, 011905/1–011905/6.
50. Ferrenberg, A. M. & Swendsen, R. H. (1989). Optimized Monte Carlo data analysis. *Phys. Rev. Letters*, **63**, 1195–1198.

Edited by M. Levitt

(Received 17 March 2003; received in revised form 23 June 2003; accepted 11 July 2003)