Coupling between allosteric transitions in GroEL and assisted folding of a substrate protein

George Stan*†‡, George H. Lorimer†§, D. Thirumalai†§, and Bernard R. Brooks*†

*Laboratory of Computational Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892; and §Biophysics Program, Institute for Physical Science and Technology, and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742

Edited by Harold A. Scheraga, Cornell University, Ithaca, NY, and approved March 27, 2007 (received for review January 22, 2007)

Escherichia coli **chaperonin, GroEL, helps proteins fold under nonpermissive conditions. During the reaction cycle, GroEL undergoes allosteric transitions in response to binding of a substrate protein (SP), ATP, and the cochaperonin GroES. Using coarse-grained representations of the GroEL and GroES structures, we explore the link between allosteric transitions and the folding of a model SP, a *de novo*-designed four-helix bundle protein, with low spontaneous yield. The ensemble of GroEL-bound SP is less structured than the bulk misfolded structures. Upon binding, which kinetically occurs in two stages, the SP loses not only native tertiary contacts but also experiences a decrease in helical content. During multivalent binding and the subsequent ATP-driven transition of GroEL the SP undergoes force-induced stretching. Upon encapsulation, which occurs upon GroES binding, the SP finds itself in a "hydrophilic" cavity in which it can reach the folded conformation. Surprisingly, we find that the yield of the native state in the expanded GroEL cavity is relatively small even after it remains in it for twice the spontaneous folding time. Thus, in accord with the iterative annealing mechanism, multiple rounds of binding, partial unfolding, and release of the SP are required to enhance the yield of the folded SP.**

native state yield | soft nanomachine | force-induced stretching

The *Escherichia coli* chaperonin GroEL, which helps substrate proteins (SPs) fold (1–4), has two heptameric rings that are stacked back to back. The two rings function out of phase with respect to one another like a two-stroke motor (5). During the reaction cycle (Fig. 1), GroEL is driven through a series of large-scale domain movements that are triggered by binding of ATP and the cochaperonin, GroES (6–8). The SP is recognized by GroEL in the T state (step 1 in Fig. 1). ATP binding triggers a transition to the R state (step 2 in Fig. 1). The equilibrium $T \leftrightarrow R$ transition is followed by the binding of the cochaperonin GroES, which results in doubling of the volume of the GroEL cavity and encapsulation of the SP (step 3 in Fig. 1). Subsequently, ATP hydrolysis results in the transition to the R'' state (step 4 in Fig. 1). The dual coordinated function of the two heptameric rings is required for the release of the SP and GroES from the cis ring (step 5 in Fig. 1). The cycle is completed by the release of ADP so as to regenerate the SP acceptor state (step 6 in Fig. 1). However, to understand the fate of the SP as GroEL undergoes the ligand-induced allosteric transitions (9–11) it suffices to consider SP folding in a single ring (Fig. 2).

Although the outlines of the reaction cycle of the ATP-consuming GroEL–GroES nanomachine are well documented (12, 13), the relationship between the allosteric transitions and SP folding, has not been fully explored (14, 15). In part, this is because of the presence of a number of competing time scales that escalates the inherent complexity of the problem. Even in the absence of GroES there are at least four distinct time scales to consider. *In vitro* folding of most SPs, with low spontaneous yield, can be described by the kinetic partitioning mechanism (16, 17) according to which a fraction (Φ) of molecules reaches the native state rapidly on time scale τ_{fast}. The remaining fraction ($1 - \Phi$) is kinetically trapped in one of the many competing basins of attraction (CBAs) from which the transition to the native state occurs on a slower time τ_{slow}. Even if there is only one dominant CBA there are two time scales that characterize spontaneous folding, namely, τ_{fast} and τ_{slow}. The equilibrium T ↔ R transition (step 2 in Fig. 1) is controlled by two time scales, namely, $\tau_{T\to R}$ and $\tau_{R\to T}$. Depending on the values of the four global time scales (τ_{fast}, τ_{slow}, $\tau_{T\to R}$, and $\tau_{R\to T}$) a number of distinct scenarios can emerge (12). As a result, the link between SP folding and GroEL allostery, even in the absence of GroES, has been difficult to sort out (14, 15).

Here, we use coarse-grained models to follow the fate of a model four-helix bundle protein (18), with small Φ, as it interacts with the chaperonin system during one hemicycle by using coarse-grained models for GroEL, GroES, and the SP. The SP is the *de novo*-designed four-helix bundle whose folding characteristics have been modeled in an earlier study (19). Upon binding to GroEL the SP experiences a stretching force that results in an overall extension of the polypeptide chain. The structures of the ensemble of misfolded conformations in the bulk are substantially different from those that are bound to GroEL. When GroES binds, the SP is encapsulated in the GroEL cavity in which folding can occur in a protected environment. Surprisingly, we find the yield of the folded SP is only $\approx 5\%$ ($\approx\Phi$) on the time scale $\approx 0.2\ \tau_F$, where $\tau_F \approx \Phi\tau_{fast} + (1 - \Phi)\ \tau_{slow}$. Our results show that the stochastic nanomachine GroEL participates actively by "mechanically" unfolding SPs, which puts it in a different part of the energy landscape from which it folds by the kinetic partitioning mechanism. The small yield even after encapsulation suggests, in accord with the iterative annealing mechanism (IAM) (17), that multiple rounds of binding and release are often required for sufficient yield.

Results and Discussion

Misfolded SPs in the Bulk Have Partial Secondary Structures but Lack Stable Native Tertiary Contacts. The bulk (in the absence of GroEL) folding simulations are carried out by quenching from an initial temperature of $T = 500$ K to $T = 300$ K. Shortly after the temperature quench, the polypeptide chains become compact. The distribution of the radius of gyration $P(R_g)$ shows two peaks (Fig. 3*a*). The one at $\langle R_g \rangle \simeq 12$ Å is close to the radius of gyration of the native structure ($R_g^N = 11.4$ Å), whereas the broader peak at $\langle R_g \rangle \simeq 16$ Å represents an ensemble of misfolded conformations. From the distribution of the structural overlap function [Eq. **1** in

Author contributions: D.T. designed research; G.S. and B.R.B. performed research; G.S. and B.R.B. contributed new reagents/analytic tools; G.S., G.H.L., and D.T. analyzed data; and G.S., G.H.L., and D.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: SP, substrate protein; IAM, iterative annealing mechanism; PDB, Protein Data Bank.

†To whom correspondence may be addressed. E-mail: george.stan@uc.edu, glorimer@umd.edu, thirum@glue.umd.edu, or brb@nih.gov.

‡Present address: Department of Chemistry, University of Cincinnati, P.O. Box 210172, Cincinnati, OH 45221.

This article contains supporting information online at www.pnas.org/cgi/content/full/0700607104/DC1.

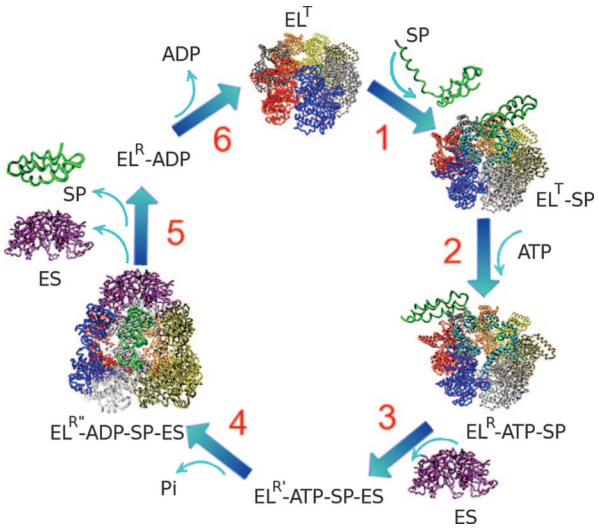

Fig. 1. Reaction hemicycle of GroEL. The six steps in the hemicycle are explained in the text. EL and ES stand for GroEL and GroES, respectively. Molecular images were created by using VMD (44) and POV-ray (www.povray.org).

supporting information (SI) *Text*] of the misfolded structures computed by using only the coordinates of residues in the helical regions of the SP (Fig. 3*b*), we obtain an average of $\langle \chi_H \rangle = 0.3$. The relatively small value of $\langle \chi_H \rangle$ shows that misfolded conformations have a large number of native intrahelical contacts. An example of a misfolded structure is shown in Fig. 3*a*. From the $P(R_g)$ (Fig. 3*a*) and the values of χ, we conclude that the secondary (helical) structures are at least partially formed even when the polypeptide chain is misfolded. By contrast, there are only a few stable native interhelical contacts in misfolded conformations, which is reflected in an average of $\langle \chi \rangle = 0.82$ (Fig. 3*b*). Thus, the misfolded conformations have native-like secondary structures and are stabilized by several incorrect long-range contacts.

GroEL Has a High Affinity for Expanded Denatured SP States. When the ensemble of structures containing a mixture of folded and misfolded conformations is presented to GroEL only the polypeptide chains that are in non-native conformations bind to GroEL (20, 21), because the SP recognition motifs are inaccessible in the native state (22, 23).

The ensemble of protein substrates that binds to GroEL has a lower native content in the initial conformation than the ensemble of unbound chains. The distribution of R_g in the initial conformations shows a bimodal distribution (Fig. 3*c*). Analysis of the SP radius of gyration, R_g, in the initial states of binding simulations reveals that GroEL has a high affinity for expanded, non-native states. SP chains that bind to GroEL are predominantly presented to GroEL in expanded conformations (Fig. 3*c*). Chains that do not bind to GroEL are most likely to have a native-like R_g in the initial conformation. These results are in accord with recent NMR experiments (24) that showed that only random-coil states of dihydrofolate reductase are captured by GroEL.

The simulations also show that $P(R_g)$ of the bound states is much broader than that of the initial misfolded ensemble (Fig. 3*c*). Surprisingly, there is a loss in the overall helix content in the SP upon binding to GroEL (see SI Fig. 6). These results show that the GroEL-bound ensemble of structures is disordered and is largely devoid of tertiary interactions (Fig. 3*d*). Thus, the very act of binding indeed results in chain expansion that is also consistent with all-atom simulation results (25).

To the extent the polypeptide chain can be treated as a locally stiff polymer, the stretching must be caused by a "mechanical" force that

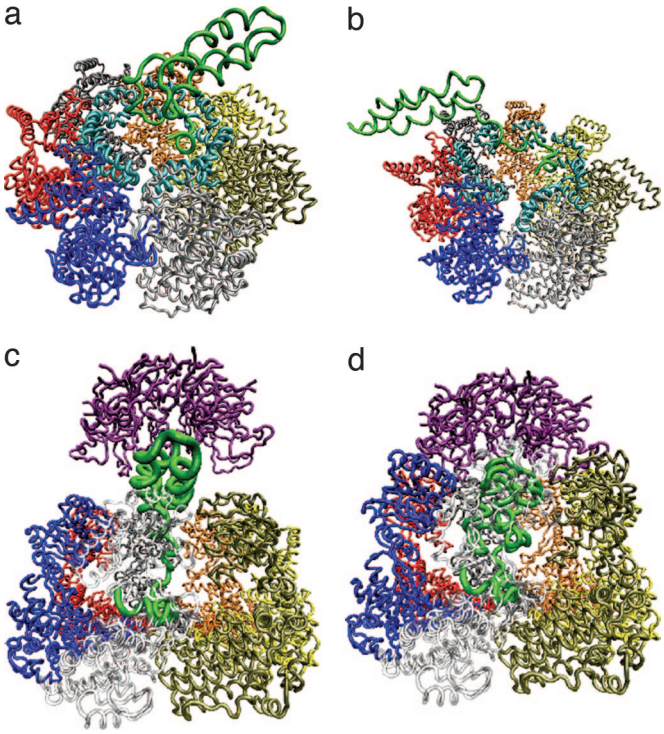

Fig. 2. Snapshots during the GroEL hemicycle. (*a*) *T* state (PDB code 1AON). (*b*) *R* state (PDB code 2C7E). (*c*) Intermediate state in the $T \rightarrow R''$ transition (the GroEL cavity is 2/3 open). (*d*) R'' state (PDB code 1OEL). Individual GroEL subunits are indicated by distinct colors, and the SP is in green. The seven single-ring GroEL binding sites are in cyan.

the SP feels because of its multivalent binding to GroEL. The force on the SP is computed in terms of the virial function that gives the work done on the SP as a result of chain expansion (Fig. 4). The average stretching force ($\bar{\mathfrak{F}} \approx |W|/\Delta R$, where ΔR is the extension of the SP) estimated from the red line in Fig. 4 is $\approx$ 20–40 pN. Fig. 4 shows that each molecule experiences different magnitude, which indicates the dynamic heterogeneity in the SP–GroEL interaction.

Yield of Spontaneous SP Folding Is Low. To establish that $T = 300$ K represents nonpermissive conditions for the SP we obtained the folding kinetics of the four-helix bundle by computing the first passage times for 1,000 molecules. The fraction of molecules that have not folded at time t obeys the kinetic partitioning mechanism and is given by

$$P_u(t) = \Phi e^{-\frac{t}{\tau_{fast}}} + (1 - \Phi) e^{-\frac{t}{\tau_{slow}}},$$

where $\Phi \approx 0.06$, $\tau_{fast} = 1.8$ ns, and $\tau_{slow} = 164$ ns. At $T = 300$ K, 244 of 1,000 molecules (native yield of 24%) folded in 30 ns (SI Table 1), which we set as the time for the hemicycle. The remaining (76%) trajectories result in misfolded conformations that are kinetically trapped (Fig. 3*b*). Thermal fluctuations induce transitions from misfolded conformations to the native state, with low probability, resulting in a slow monotonic decrease of the radius of gyration of the SP (Fig. 5*a*) and the structural overlap function, χ (green curve in Fig. 5*b*). Thus, unassisted folding of the SP shows that $T = 300$ K mimics the nonpermissive condition.

SP Binding and Release Requires an Optimal Range of GroEL–SP Interaction Strength. To investigate the effect of the GroEL–SP interaction strength on SP binding, we performed a number of simulations (see SI Table 1) for different values of the strength of the SP–GroEL interaction, λ (see *Methods*). The minimum value for which SP capture by GroEL (the binding criterion is defined in

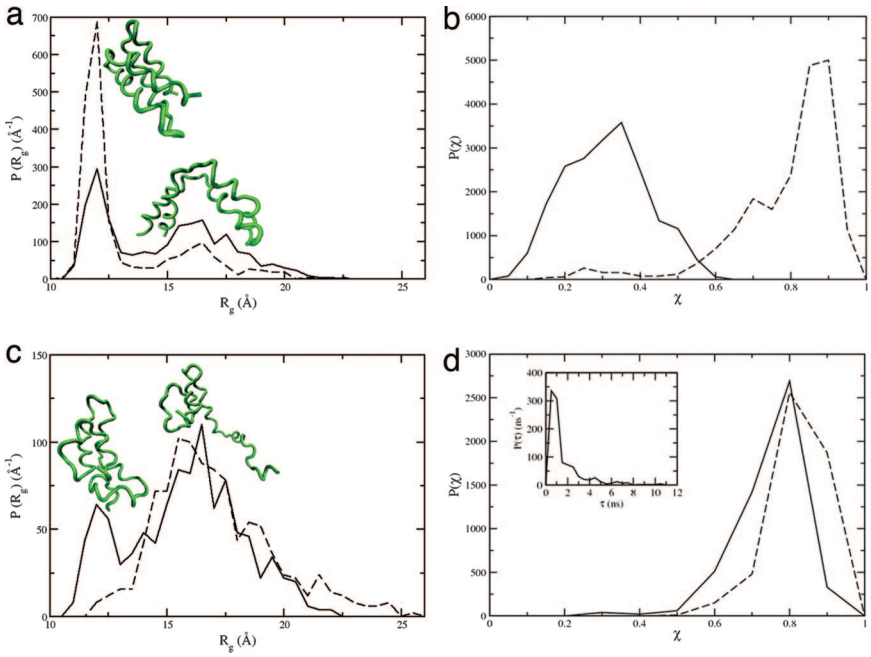

Fig. 3. Distribution functions characterizing the global structural features of the SP at 300 K. (*a*) Probability distribution $P(R_g)$ of the radius of gyration (R_g) of the SP in the bulk. The $P(R_g)$ for the misfolded conformations, obtained at 7.5 ns after the temperature quench from 500 K to 300 K, is given by the solid line, and the dashed lines are for the ensemble of folded states obtained after 30 ns. The peak at $\langle R_g \rangle \simeq 12$ Å in the solid line is a mixture of native and misfolded conformations that are targets for GroEL. (*b*) Distribution of the structural overlap function $P(\chi)$ in the bulk for the ensemble of misfolded conformations whose $P(R_g)$ is shown by the solid line in *a*. The solid line represents the overlap function for intrahelical contacts, and the dashed curve corresponds to interhelical contacts. (*c* and *d*) The same as in *a* and *b* except that the solid lines and dashed lines show distributions before and after binding to GroEL with $\lambda = 1$, respectively. In computing $P(R_g)$ and $P(\chi)$ only those molecules that bind to GroEL are included.

Methods) is observed is $\lambda = 0.35$ and for $\lambda = 0.5$ SP binds to multiple GroEL binding sites. Large values of λ (>1) enhance the yield of the binding reaction SP + GroEL → SP · GroEL. However, the yield of the SP · GroEL complex increases more at small values of λ than for larger ones, which shows that GroEL has a binding capacity for SPs. Thus, the requirement that the SP binding be strong enough to effect capture of non-native SPs but not so strong as to preclude its release is satisfied. This conclusion is in accord with earlier experimental (26) and theoretical (14, 27–31) studies, which showed that efficient processing of SPs requires optimal strength of the SP–GroEL interaction.

SP Capture Kinetics Involves Multiple Time Scales and Mechanisms. The overall capture kinetics computed from the distribution of binding times (see Fig. 5*a Inset*) is best described by using a sum of two exponentials (see *Methods*). The binding rates indicate a fast ($k_1 = 1.7$ ns^{-1}, $\tau_1 = 0.6$ ns) and a slow ($k_2 = 0.4$ ns^{-1}, $\tau_2 = 2.5$ ns) capture by at least one GroEL subunit. Strong binding to multiple subunits also occurs on multiple time scales with rates $k_1^{strong} = 1.3$ ns^{-1}, $k_2^{strong} = 0.3$ ns^{-1}. The observation of multiple time scales in SP capture suggests distinct modes of recognition that are a function of the nature of conformation of the polypeptide chain. Analysis of the binding-induced conformational changes shows that capture is fast for bulk conformations that have large solvent-accessible binding regions. These conformations are readily captured by the bulky hydrophobic residues in the H and I helices of GroEL. By contrast, the molecules that are recognized on a longer time scale require significant changes in the bulk conformations to increase the solvent accessibility of regions. Thus, only those conformations that have their interface regions oriented to complement the SP binding sites are rapidly ensnared. The contrasting capture mechanisms, namely, conformational selection (rapid binding) vs. induced-fit (slow binding), described in protein–ligand binding, indicates that GroEL efficiently recognizes those misfolded conformations that expose specific binding regions (32, 23). Capture of non-native dihydrofolate reductase by GroEL occurs on a millisecond time scale, whereas capture of native dihydrofolate reductase by GroEL occurs on a much longer time scale (20).

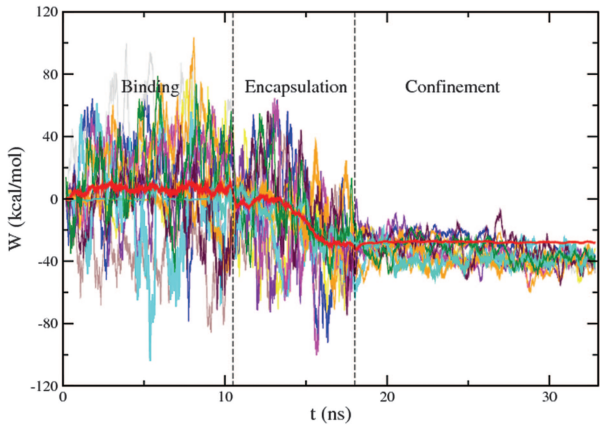

Fig. 4. Virial function for the SP helices (thick red line). The virial values represented in this graph are calculated as running averages of instantaneous virial values with a window of 0.4 ns. Thin curves represent virial functions for the 12 independent trajectories that result in SP folding. The average over all of the trajectories, regardless of whether they fold or not, is shown in red. Shown is heterogeneous behavior in individual molecules during the reaction cycle.

SP Unfolds upon Binding and During the *T* → *R* Transition. Fig. 3*c* shows the shift of the SP size distribution toward larger R_g in the binding process, which suggests that it must at least partially unfold. The shift in R_g to higher values is almost entirely caused by the unfolding of compact states with the initial R_g that is slightly larger than R_g^N (Fig. 3*c*). Unfolding of SP during binding to GroEL is also accompanied by an overall increase in the structural overlap function (Fig. 5*b*). Cooperative ATP binding at the seven GroEL subunits induces rigid-body motion of the apical domain of each subunit, resulting in a conformational transition from the nucleotide-free *T* state to the nucleotide-bound *R* state (step 2 in Fig. 1). During this transition, apical domains move up ≈10° and rotate counterclockwise ≈25° (33). To ascertain the effect of ATP binding on the SP conformational change we considered three values (see SI Table 1) of the strength of the SP–GroEL interaction (Eq. **3** in *Methods*) $\lambda = 1.0$, 0.75, and 0.5. The lowest value of λ results in the release of the SP from the binding sites. At the two larger values of λ, SP chains that are bound to the GroEL in the *T* state remain bound to the GroEL in the *R* state. For $\lambda = 0.75$ and 1, unfolding of the SP chain that begins during binding to the *T* state continues

BIOPHYSICS

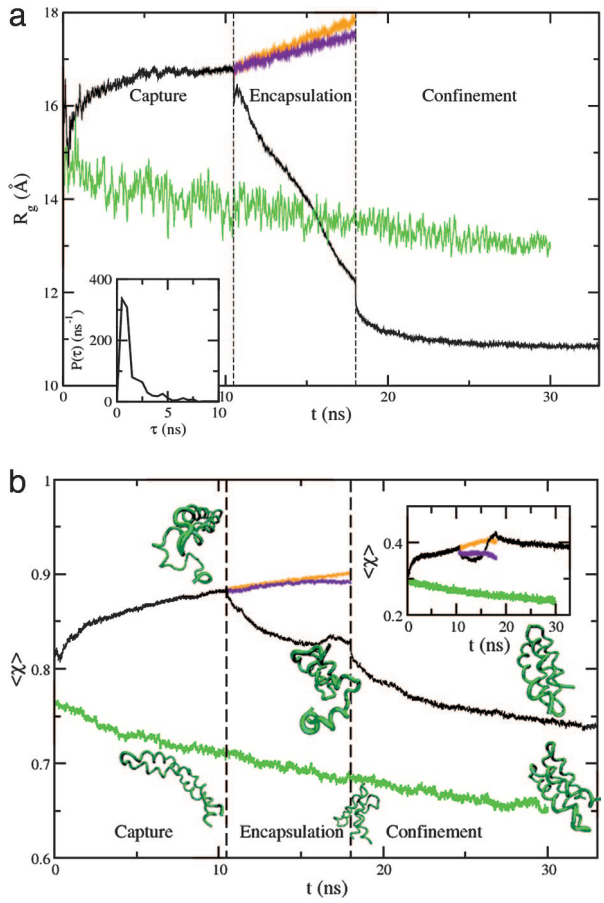

Fig. 5. Dynamical changes in the SP upon interaction with GroEL and GroES. Radius of gyration (*a*) and structural overlap function (*b*) of nonlocal (interhelical) contacts of the SP conformations during the GroEL cycle (black) and in the bulk (green). Orange (purple) curves correspond to changes in R_g and $\langle\chi\rangle$ during the $T \rightarrow R$ transition at $\lambda = 1.0$ (0.75). The various stages of the chaperonin cycle are delimited by discontinuous lines. (*a Inset*) The distribution of binding times. (*b Inset*) The structural overlap function for local (intrahelical) contacts. The results for the bulk is noisier because the SP undergoes larger structural fluctuations than in the GroEL cavity.

as the transition to the R state takes place (Fig. 5*a*). However, the extent of unfolding beyond what is observed in the process of SP recognition by GroEL is small. At both values of λ, we find only a ≈3% change in $\langle\chi\rangle$ and a ≈6% change in R_g. The changes in the R_g upon capture and the subsequent $T \rightarrow R$ transition is ≈3 Å (or ≈20% over the initial value). The SP undergoes further expansion upon $T \rightarrow R$ transition (Fig. 5*a*), which suggests that GroEL functions as an annealing machine from the capture stage. A similar expansion in the size of rhodanese was observed upon capture and subsequent $T \rightarrow R$ transition (25).

These results show that upon capture and in the subsequent $T \rightarrow R$ transition the SP is stretched. As a result, there is a net loss in the number of tertiary contacts compared with the bulk. In other words, under nonpermissive folding conditions, SPs are more compact in the absence of GroEL than in the bound state. Several previous experimental studies (34, 35), based on hydrogen-exchange NMR, and computational studies (14) have shown that SP unfolding is the mechanism used by GroEL to disrupt the non-native and native contacts in misfolded conformations.

GroEL Resolves Non-Native Interhelical Contacts. Misfolding of proteins in the bulk results in the formation of both local (intrahelical) and nonlocal (interhelical) contacts. The rearrangement of incorrect local contacts is a kinetically rapid process, whereas annealing of nonlocal contacts could require partial or global unfolding of the polypeptide chain. We have dissected, using the changes in $\langle\chi\rangle$, the type of non-native contacts that are annealed by SP capture. For comparison, we have also computed the time-dependent changes in $\langle\chi\rangle$ to monitor folding in the bulk (unassisted folding of the four-helix bundle). The unfolding action of GroEL results primarily in the disruption of contacts that stabilize interhelical contacts (Fig. 5*b*). Native intrahelical contacts are only slightly perturbed in the SP binding process (black curve in Fig. 5*b*). By contrast, a large number of native interhelical contacts (black curve in Fig. 5*b*) is broken when SP binds to GroEL. In the bulk, the number of native interhelical contacts increases monotonically under folding conditions (green curve in Fig. 5*b*). Thus, GroEL unfolds SP largely by unpacking the helices without significantly altering the local helical structures. Assisted folding occurs by the disruption of mismatched tertiary contacts between the helices in the four-helix bundle.

Taken together, these results give a coherent picture of assisted folding of the four-helix bundle. Upon capture by GroEL the secondary structure content of the polypeptide chain does not differ significantly from its value in the bulk. However, disruption of tertiary contacts, which is accompanied by the increase in the size of the polypeptide chain, shows that GroEL actively helps unfold the SP. As a result, the SP is in a different part of the energy landscape from which folding can commence. The simulation results are consistent with the IAM (17) that postulates that SP annealing occurs because unfolding puts the polypeptide chain in higher free-energy surface from which it can fold by kinetic partitioning. Thus, GroEL stochastically kicks the chain into different regions of the rough energy landscape, which is determined by the sequence and external conditions.

GroES Binding Encapsulates the SP and Helps Folding. The interaction of the cochaperonin with GroES leads to large-scale (step 3 in Fig. 1 and Fig. 2 *c* and *d*), fully coordinated, rigid body motions in all seven subunits of the GroEL ring that results in the doubling of the cavity volume. We model the $T \rightarrow R''$ transition (see *Methods*) by slowly deforming the GroEL structure between the end states by rigidly translating GroES from an infinite separation from GroEL (T state) to its location atop the GroEL cavity (R''). We perform encapsulation and confinement simulations starting from bound SP structures and by varying λ so that the SP is fully encapsulated in the cavity. Upon encapsulation the SP can fold in a predominantly hydrophilic environment. To mimic the finite time the SP spends in the GreoEL cavity we use 30 ns for the hemicycle.

If the SP folding were to occur in the bulk then the native yield of the SP would be 24% after 30 ns has elapsed. The arbitrary time of 30 ns for the duration of the hemicycle is much longer than τ_{fast} (1.8 ns), the time in which fast track molecules fold in the bulk. Confinement-induced stabilization of the native state and biasing of the denatured state ensemble toward the native state has been reported in previous studies using simple pore models (14, 30, 36–38). In our coarse-grained model, which corresponds to a moderate confinement ($R_g^{GroEL}/R_{g,Native}^{bulk} \simeq 2.3$), there is a net decrease in $\langle\chi\rangle$ during encapsulation and confinement (Fig. 5*b*). During the encapsulation stage, the native content of SP chains first increases ($\langle\chi\rangle$ decreases) (Fig. 5*b*). Toward the end of the encapsulation phase there is a slight increase in $\langle\chi\rangle$, which suggests a decrease in the native content (Fig. 5*b Inset*). Compared with bulk SP folding, confinement within the GroEL cavity macroscopically results in a smaller overall acquisition of the native content (green and black curves in Fig. 5*b*).

Yield of the Native SP Is Low in the One Round of the Reaction Cycle. If we use the criterion that, in the native state, $\langle\chi\rangle$ for all contacts (intrahelical and interhelical) should be <0.22 (see *Methods*), only ≈4% of the encapsulated chains reach the native state. The low yield in the GroEL cavity, in which the SP is not allowed to spend

an infinite time, can be understood by using the IAM. According to IAM the native yield after n iterations is $\Psi_N = 1 - (1 - \Phi)^n$. If $n = 1$, as is the case in SR1, then we expect from the bulk simulations of the SP that $\Psi_N = \Phi \approx 6\%$. Upon confinement the time scale for the slow folding τ_{slow} greatly increases because global unfolding that may be required to enable misfolded structures to reach the native state is prohibited. As a result, in $t = 30$ ns only fast track molecules (those that fold on the time scale τ_{fast}) can reach the native state, which explains the low yield during a single round of the hemicycle.

The small yield that is observed here is similar to that for the stringent substrate Rubisco for which ≈5% of the molecules fold per iteration of the GroEL reaction cycle (17). As in the case of stringent SPs, the four-helix bundle protein must undergo several iterations to enhance the yield of the native state. To confirm the above arguments, we simulated the second iteration of SP folding within the model GroEL cavity. At the outset of the second iteration, misfolded SP conformations released from the GroEL cavity after the first round are recaptured by the T state GroEL apical binding sites. The subsequent SP encapsulation and confinement, mimicking the second annealing cycle, result in a yield of ≈6% native SP, similar to the first round. Thus, after multiple iterations the yield of the native SP is significantly enhanced, which confirms the IAM.

The reason for the larger yield in the bulk, on the hemicycle time scale used in the simulations, is the following. Upon confinement, the SP cannot sample a number of conformations in which R_g exceeds the size of the cavity. As a result, confinement can greatly increase τ_{slow} (37). Hence, time scales much greater than the encapsulation time used in our simulations are needed to reach the native state when the SP is in a cavity. These arguments are vividly illustrated in Fig. 5*b* (black curves), which shows that a plateau in $\langle\chi\rangle$ is reached after ≃ 5 ns after confinement. Compared with bulk folding, trapped misfolded states in the GroEL cavity have larger free-energy barriers to the native state because of the restricted conformational space available (global unfolding of misfolded states occurs with low probability inside the cavity). Refolding in the GroEL cavity involves strong perturbation of the local (intrahelical) and long-range (interhelical) native contacts. As shown in Fig. 5*b*, black curves, corresponding to long-range (local) contacts, have a similar gradient during the $T \rightarrow R''$ transition (the encapsulation stage), whereas in confinement both curves reach plateau values within a few nanoseconds.

Alteration of the duration of the productive folding time relative to the total GroEL cycle time affects the yield of native SP. The GroEL cycle time ($= \tau_H + \tau_P$) can be approximately partitioned into two fundamental time scales, τ_H and τ_P (12). During τ_H, the SP interacts with the hydrophobic sites of the GroEL cavity, whereas during τ_P, the productive folding time, it interacts with mostly hydrophilic sites lining the cavity interior. To study the effect of changing $R_P = \tau_P/\tau_H$ on SP yield, we performed simulations corresponding to three values of τ_H, namely 5, 10.5, and 21 ns (see SI Table 1). We find that for the largest value of $R_P = 4.5$, the yield is 1.5%, whereas for $R_P = 2.2$ and 1.1, the yield is 4% and 3.2%, respectively.

Concluding Remarks

To examine the structural changes that take place in a SP as it interacts with GroEL during one round of the reaction cycle, we have simulated, using minimal models and Langevin simulations, the assisted folding of a model four-helix bundle protein. The majority of the model SP molecules misfolds at $T = 300$ K, making them susceptible to be captured by GroEL. The GroEL-bound SP is an ensemble of structurally diverse conformations that do not have native-like features. Both in terms of secondary and tertiary structures the bound SP is more disordered than in the bulk. On an average, the SP is stretched upon binding. The extent of folding upon encapsulation depends on the time the SP spends in the cavity, which is set by the load that the nanomachine experiences. The forced stretching of the SP followed by encapsulation supports the basic premise of the IAM of assisted folding.

In our simulations we chose the reaction cycle time τ_{RC} to be less than τ_{slow} but greater than τ_{fast}. Studies on folding in confined spaces show that τ_{slow} depends on the size of the cavity, becoming longer as the volume of the confined space becomes smaller (37). Unless the productive folding time (τ_P) in the cavity is comparable to the cavity-size-dependent τ_{slow} the native state yield in one hemicycle of assisted folding is expected to be small. Even when $\tau_P \approx \tau_{slow}$ the maximum yield in one iteration is expected to be $\Psi_N \approx (1 - \Phi)/e$ (12). For our model system this would result in recovery of 35% folded state. In the SR1 mutant $\tau_P = 300$ min (39); hence, experiments that use only SR1 do not give a true picture of GroEL-assisted folding. In reality τ_P depends on the load placed on the machine because GroEL reaction cycle time critically depends on the binding of SP and ATP to the trans cavity, whereas the SP is encapsulated in the cis cavity (J. Grason, J. Gresham, and G.H.L., unpublished data). By restricting ourselves to $\tau_{RC} < \tau_{slow}$ we have mimicked the expected loading conditions. To enhance the yield it is more efficient to subject the SP to multiple rounds of binding and release. The IAM and the present work show that to understand the efficiency of GroEL function it is necessary to not only vary the "loading" conditions but also measure the yield of the folded molecules as a function of ATP and SP concentration.

Methods

Structural Models for GroEL. We represent the T (40) [Protein Data Bank (PDB) code 1AON], R (41) (PDB code 2C7E), and R'' (6) (PDB code 1OEL) structures using their α-carbon atoms. The residues are classified by three letter codes, namely, hydrophobic (B), hydrophilic (L), and neutral (N). The hydrophobic residues are Cys, Phe, Ile, Leu, Trp, Val, Met, Tyr, and Ala; the hydrophilic residues are Gly, Asn, Thr, Ser, Gln, Asp, Glu, Lys, and His; and the Arg and Pro residues are neutral. Because electrostatic interactions are not explicitly used in our simulations, charged residues are classified as hydrophilic or neutral. To prevent the unphysical release of SP from the GroEL cavity through the equatorial region, the coordinates of 23 C-terminal amino acids of each GroEL subunit, not resolved in the crystal structures, are generated by using random dihedral angles. The SP size (≈8 kDa) is smaller than the typical stringent GroEL substrates (30–50 kDa). To obtain a GroEL–SP size ratio that is characteristic of stringent SPs, we scale down the coordinates of GroEL and GroES amino acids by a factor of 2.

SP Energy Function. We represent the energy of a conformation of the four-helix bundle (19), expressed in terms of the cartesian coordinates r_i ($i = 1, \ldots, M$), with $M = 73$, as:

$$V_{tot} = V_{NB} + V_{BL} + V_{BA} + V_{DA}, \qquad [1]$$

where V_{NB} is the nonbonded potential, V_{BL} is the bond-length potential, and V_{BA} and V_{DA} are the bond angle and the dihedral angle potentials, respectively. The nonbonded potential $V_{NB} = \Sigma_{ij} V_{ij}(r_{ij})$, where $V_{ij}(r_{ij})$ is:

$$V_{ij} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]. \qquad [2]$$

Here, $\epsilon_{ij} = 2.125$ kcal/mol and $\sigma_{ij} = 3.8$ Å (C_α–C_α distance) for $i = j = B$, $\epsilon_{ij} = 10^{-12}$ kcal/mol, and $\sigma_{ij} = 40.47$ Å for $i = \{L, N\}$ and $j = \{B, L, N\}$, and r_{ij} is the distance between residues. The small value of ϵ_{ij} for involving the L, N residues results in a purely repulsive interactions (see SI Fig. 7).

Chaperonin-SP Interaction. Nonbonded interactions between GroEL/S and the SP residues are given by

$$V_{G_i, H_j} = \lambda V_{H_i, H_j}, \quad [3]$$

where G (H) represent GroEL/S (SP) residues, $ij = \{B, L, N\}$, and λ is used to scale the intermolecular nonbonded interactions and depends on the allosteric state of GroEL. The value of λ is larger when GroEL is in the T state to facilitate capture of the SP. We decrease λ progressively as $T \rightarrow R''$ transitions take place to reflect the change in the nature of SP–GroEL interactions as the cavity lining changes from being hydrophobic (T state) to hydrophilic (R and R'' states).

Models for GroEL and GroES. GroEL and GroES residues are either constrained to fixed positions corresponding to the crystal structures or move along straight paths as the allosteric transitions take place (see discussion below). Such a procedure is similar in spirit to that used in the targeted molecular dynamics (25, 42).

SP Binding to the GroEL Cavity. The SP binding simulations are initiated by placing a polypeptide chain at a center-to-center separation of 43 Å from GroEL along the GroEL symmetry axis (Fig. 2). The SP is positioned closer to the apical domain end of GroEL. The GroEL residues are maintained at fixed positions corresponding to the T state. During the simulations, the center of mass of the SP is weakly restrained by a harmonic potential with a force constant of 10^{-3} kcal/(mol Å^2). The restraint is imposed to prevent large separations between SP and GroEL. A number of trajectories is generated (see SI Table 1) for the range of λ values.

Recognition of the SP is assumed to occur when the SP makes at least one contact with residues in the grove between both helices H (residues 234–243) and I (residues 257–268). In our model, a binding contact is defined as a maximum separation of 6 Å between a SP residue and both regions 230–245 and 255–272 in GroEL. The fraction of trajectories that result in SP chains bound to the GroEL cavity is $f_B = N_B/N_T$, where N_B is the number of trajectories that result in the SP being recognized by GroEL and N_T is the total number of trajectories.

Simulations of the $T \rightarrow R''$ Transition. Allosteric transitions during the $T \rightarrow R''$ transition, induced by ATP and GroES binding, are mimicked as follows. The T and R'' structures are aligned to minimize the rmsd between the GroEL equatorial domains. During the $T \rightarrow R''$ transition, GroEL residues are constrained to move at a constant velocity along a path connecting their T and R'' positions. GroES binding is modeled by translating GroES, also at a constant speed, along the GroEL/GroES symmetry axis from a position far away from GroEL to the top of the GroEL cavity (Fig. 2). The assumed ballistic motion of the GroEL residues is not unreasonable, because, to a large extent, the subdomains of GroEL move as rigid bodies between the various allosteric states. The initial position of GroES is obtained from the R'' conformation by displacing GroES by 23 Å away from GroEL along the GroEL/GroES symmetry axis. During the $T \rightarrow R''$ transition, chaperonin amino acids move at speeds on the order of 0.1–1 m/s (10^5 to 10^6 μm/s). Although these speeds are large, it is unlikely that the qualitative results on the dependence of SP folding on GroEL allostery, the focus of the present study, will be greatly affected. The constant velocity procedure has been implemented as a generalization of the constant velocity subroutine in the CHARMM program. As shown in SI Table 1, four encapsulation runs are initiated from each successful binding event observed for $\lambda = 1.0$.

The annealing function of GroEL is linked to the allostery-induced alterations in the nature of the SP–GroEL interactions. To mimic the chemical environment switch experienced by the SP in the transition from the mostly hydrophobic T state to the mostly hydrophilic interaction within the GroEL cavity in the R'' state we vary the λ parameter. The intermolecular interaction is scaled by $\lambda = 0.35$ for a set of hydrophobic residues that participate in the GroEL–GroES interface or are exposed to the external surface of GroEL in the R'' state. Specifically, these are the hydrophobic residues in the GroEL segments 140–170, 180–190, 231–272, 287–316, and 338–376, and the GroES segment 10–33. The optimal value of the λ parameter facilitates the observation of a reasonable number of SP unbinding and encapsulation events (SI Table 1) in the course of our simulations. The remaining interactions are unscaled ($\lambda = 1.0$).

GroEL-Assisted Protein Folding. To investigate the dynamics of the GroEL-assisted folding of the SP in the single ring we used low-friction Langevin simulations of the coarse-grained, off-lattice models (43). The low-friction dynamics results in small time scales for binding and folding.

This work is supported in part by National Institutes of Health Grant 1R01GM067851-01 and the Intramural Research Program of the National Institutes of Health, National Heart, Lung, and Blood Institute.

1. Viitanen PV, Gatenby AA, Lorimer GH (1992) *Protein Sci* 1:363–369.
2. Lorimer GH (1996) *FASEB J* 10:5–9.
3. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang H-C, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, Hartl FU (2005) *Cell* 122:209–220.
4. Chapman E, Farr GW, Usaite R, Furtak K, Fenton WA, Chaudhuri TK, Hondorp ER, Matthews RG, Wolf SG, Yates JR, *et al.* (2006) *Proc Natl Acad Sci USA*, 103:15800–15805.
5. Lorimer GH (1997) *Nature* 388:720–723.
6. Xu Z, Horwich AL, Sigler PB (1997) *Nature* 388:741–750.
7. Xu Z, Sigler PB (1998) *J Struct Biol* 124:129–141.
8. Horovitz A, Willison KR (2005) *Curr Opin Struct Biol* 15:646–651.
9. Keskin O, Bhahar I, Flatow D, Covell DG, Jernigan RL (2002) *Biochemistry* 41:491–501.
10. Chennubhotla C, Bahar I (2006) *Mol Sys Biol*, 10.1038/msb4100075.
11. Hyeon C, Lorimer GH, Thirumalai D (2006) *Proc Natl Acad Sci USA* 103:18939–18944.
12. Thirumalai D, Lorimer GH (2001) *Annu Rev Biophys Biomol Struct* 30:245–269.
13. Horwich AL, Farr GW, Fenton WA (2006) *Chem Rev* 106:1917–1930.
14. Betancourt MR, Thirumalai D (1999) *J Mol Biol* 287:627–644.
15. Yifrach O, Horovitz A (2000) *Proc Natl Acad Sci USA* 97:1521–1524.
16. Guo Z, Thirumalai D (1995) *Biopolymers* 36:83–102.
17. Todd MJ, Lorimer GH, Thirumalai D (1996) *Proc Natl Acad Sci USA* 93:4030–4035.
18. Ho SP, DeGrado WF (1987) *J Am Chem Soc* 109:6751–6758.
19. Guo Z, Thirumalai D (1992) *J Mol Biol* 263:323–343.
20. Viitanen PV, Donaldson GK, Lorimer GH, Lubben TH, Gatenby AA (1991) *Biochemistry* 30:9716–9723.
21. Walter S, Lorimer GH, Schmid FX (1996) *Proc Natl Acad Sci USA* 93:9425–9430.
22. Goloubinoff P, Christeller JT, Gatenby AA, Lorimer GH (1989) *Nature* 342:884–889.
23. Stan G, Brooks BR, Lorimer GH, Thirumalai D (2006) *Proc Natl Acad Sci USA* 103:4433–4438.
24. Horst R, Bertelsen EB, Fiaux J, Wider G, Horwich AL, Wüthrich K (2005) *Proc Natl Acad Sci USA* 102:12748–12753.
25. van der Vaart A, Ma J, Karplus M (2004) *Biophys J* 87:562–573.
26. Gray TE, Fersht AR (1993) *J Mol Biol* 232:1197–1207.
27. Cheung MS, Thirumalai D (2006) *J Mol Biol* 357:632–643.
28. Thirumalai D (1994) in *Statistical Mechanics, Protein Structure, and Protein Substrate Interactions*, ed Doniach S (Plenum, New York), pp 115–134.
29. Orland H, Thirumalai D (1997) *J Phys* 7:533–560.
30. Baumketner A, Jewett A, Shea JE (2003) *J Mol Biol* 332:701–713.
31. Jewett AI, Baumketner A, Shea JE (2004) *Proc Natl Acad Sci USA* 101:13192–13197.
32. Stan G, Brooks BR, Lorimer GH, Thirumalai D (2005) *Protein Sci* 14:193–201.
33. Ranson NA, Farr GW, Roseman AM, Gowen B, Saibil HR, Horwich AL (2001) *Cell* 107:869–879.
34. Zahn R, Perrett S, Stenberg G, Fersht AR (1996) *Science* 271:642–645.
35. Nieba-Axmann SE, Ottiger M, Wüthrich K, Plückthun A (1997) *J Mol Biol* 271:803–818.
36. Brinker A, Pfeifer G, Kerner MJ, Naylor DJ, Hartl FU, Hayer-Hartl M (2001) *Cell* 107:223–233.
37. Klimov DK, Newfield D, Thirumalai D (2002) *Proc Natl Acad Sci USA* 99:8019–8024.
38. Takagi F, Koga N, Takada S (2003) *Proc Natl Acad Sci USA* 100:11367–11372.
39. Weissman JS, Hohl CM, Kovalenko O, Kashi Y, Chen S, Braig K, Saibil HR, Fenton WA, Horwich AL (1995) *Cell* 83:577–587.
40. Braig K, Otwinowski Z, Hedge R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB (1994) *Nature* 371:578–586.
41. Ranson NA, Clare DK, Farr GW, Houldershaw D, Horwich AL, Saibil HR (2006) *Nat Struct Mol Biol* 13:147–152.
42. Ma J, Karplus M (1998) *Proc Natl Acad Sci USA* 95:8502–8507.
43. Veitshans T, Klimov D, Thirumalai D (1997) *Folding Des* 2:1–22.
44. Humphrey W, Dalke A, Schulten K (1996) *J Mol Graphics* 14:33–38.