Mechanisms of protein folding

Viara Grantcharova*, Eric J Alm†, David Baker† and Arthur L Horwich‡

The strong correlation between protein folding rates and the contact order suggests that folding rates are largely determined by the topology of the native structure. However, for a given topology, there may be several possible low free energy paths to the native state and the path that is chosen (the lowest free energy path) may depend on differences in interaction energies and local free energies of ordering in different parts of the structure. For larger proteins whose folding is assisted by chaperones, such as the *Escherichia coli* chaperonin GroEL, advances have been made in understanding both the aspects of an unfolded protein that GroEL recognizes and the mode of binding to the chaperonin. The possibility that GroEL can remove non-native proteins from kinetic traps by unfolding them either during polypeptide binding to the chaperonin or during the subsequent ATP-dependent formation of folding-active complexes with the co-chaperonin GroES has also been explored.

Addresses
*Center for Genomics Research, Harvard University, Cambridge, MA 02138, USA
†Department of Biochemistry and Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA
‡Department of Genetics and Howard Hughes Medical Institute, Yale School of Medicine, New Haven, CT 06510, USA

Current Opinion in Structural Biology 2001, 11:70–82

0959-440X/01/$ – see front matter

Abbreviations

AcP	acylphosphatase
Ada2h	activation domain of procarboxypeptidase
CO	contact order
EDTA	ethylenediamine tetra-acetic acid
GFP	green fluorescent protein
MDH	malate dehydrogenase
Rubisco	ribulose-1,5-bisphosphate carboxylase-oxygenase
SH	Src homology
TFE	trifluoroethanol

Introduction

Two aspects of protein folding mechanisms are considered in this review: recent insights into the folding behavior of small two-state folding proteins and the action of the chaperonin GroEL in assisting the folding of larger proteins.

Folding of small proteins

The past several years have witnessed a rapid increase in the amount of experimental data on the folding of small single-domain proteins. Comparison of results on sets of both homologous and unrelated proteins has provided considerable insight into the determinants of the folding process. In this part of the review, we present simple models that incorporate recent experimental findings and appear to capture the broad outlines of the folding process. An important feature of these models is that the folding free energy landscape is dominated by the trade-off between the unfavorable loss in configurational entropy upon folding and the gain in attractive native interactions; non-native interactions are assumed not to play a significant role. As will be discussed first, recent results suggest a picture in which several different routes through the free energy landscape with roughly equivalent free energy barriers can be consistent with the overall topology (low-resolution structure) of a protein and sequence changes can, by lowering or raising one barrier relative to another, produce significant changes in the transition-state ensemble without large changes in the folding rate. Because our recent articles have probably overly emphasized the role of native state topology [1–3], we shall subsequently focus our attention on several examples that illustrate how variations in local free energies of ordering can modulate the folding process.

We begin by considering a zeroth order model in which all native interactions in a protein are equally favorable (i.e. homogeneous contact model). In such a model, the free energy cost of forming different contacts in a protein depends solely on the entropic cost of restricting the chain to allow the contact. This entropic cost increases with increasing sequence separation between the interacting residues, as more of the chain must be constrained in order to form the contact. When many of the contacts in a protein are between residues distant in the primary sequence, a large portion of the chain must be ordered before even a few favorable contacts can form, leading to a large folding free energy barrier. Conversely, when interacting residues are close in the protein sequence, the entropic cost of chain ordering is partially compensated by the formation of contacts earlier in the folding process, leading to a smaller folding free energy barrier. Therefore, in this very simple model, one expects proteins with most of their contacts between residues close in the sequence to fold faster than proteins with contacts between residues distant in the sequence.

Several years ago, we found such a relationship between folding rate and the average sequence separation between contacting residues (the contact order — CO) [1]. A considerable number of proteins have been studied in the interim period and an updated version of the plot, encompassing all the two-state folding proteins that have been kinetically characterized (Table 1), shows an even stronger correlation between CO and rate of folding (Figure 1a). The correlation is particularly remarkable because of the very wide variation in the folds and functions of these proteins. It suggests that the low-resolution structure or topology of a protein is a major determinant of the trade-off between configurational entropy loss and formation of attractive interactions, as suggested by the simple model described in the previous paragraph. The correlation also

supports the assumption that non-native interactions play a relatively minor role in shaping the folding process as, unlike native interactions, they are not expected to be related to the native structure.

In the simple zeroth order model discussed above, increasing uniformly the strength of all interactions clearly reduces the free energy barrier to folding (the unfavorable entropy of ordering is better compensated by the formation of the more favorable interactions) and the folding rate increases. Thus, for a given protein, reducing the strength of the favorable interactions (i.e. reducing stability) is expected to reduce the folding rate. Indeed, there is a nearly linear correlation between folding rate and stability for a given protein upon changes in solution conditions, most notably upon the addition of denaturant. Also, within a protein family, more stable proteins generally fold more rapidly than less stable proteins [4,5]. However, the correlation between stability and folding rate for proteins with different folds is much weaker than that between CO and folding rate, consistent with the dominant role of native state topology in determining folding rates [2].

Interestingly, there is a better correlation between the folding rate and the relative CO (average sequence separation divided by chain length) than between the folding rate and the absolute (unnormalized) CO (compare Figure 1a,b). This is somewhat unexpected as the entropic cost of contact formation is a function of the absolute CO, rather than of the relative CO, and simple models of the sort discussed above predict relationships with the absolute CO. If the improved correlation with the relative CO is borne out by further experimental data over the next several years, it may be necessary to consider models in which there is a renormalization that removes the dependence on the absolute length of the protein. An alternative possibility is that, for the proteins in this set, the stability increases with increasing length and dividing by the length accounts for the effect of stability on the folding rate, albeit in a somewhat indirect way.

We frequently encounter two questions about the contact order/folding rate correlation. First, given that the entropic cost of closing a loop in a protein is proportional to the logarithm of the loop length, shouldn't folding rates be more closely correlated to the logarithm of the CO? As shown in Figure 1c, because of the limited range of the CO values, the relationship between folding rates and log CO is nearly indistinguishable from that between folding rates and CO. Second, as the magnitude of the entropic barrier to folding depends on the CO of the folding transition-state ensemble, why is there a correlation between folding rates and the CO of the native structure? The correlation suggests that the CO of the native structure is, in turn, correlated with that of the transition-state ensemble; this is not surprising given that a reasonable fraction of the native structure is usually formed in the transition-state ensemble and that contact lengths tend to be relatively consistent within particular protein structures (in an all-helical protein, the contact lengths are consistently shorter than in a parallel β-sheet protein, for example).

Table 1

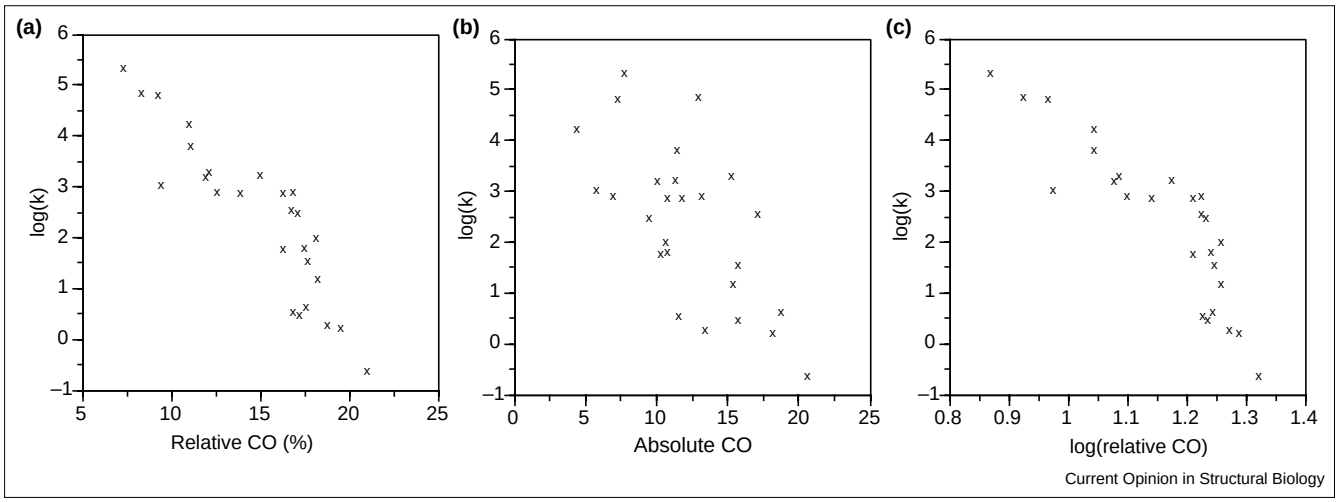
Rates of folding for two-state folding proteins.

Protein*	Log(k_f)†	CO‡ (%)	ΔG_u (kcal/mol)	Length§ (residues)	Temperature (C°)
Cyt-B_{562} [62]	5.30	7.47	10.0	106	20
Myoglobin	4.83#	8.50	8.4	154	25
λ-repressor [63]	4.78	9.37	5.6	80	20
PSBD [64]	4.20	11.20	2.2	41	41
Cyt-c [65]	3.80#	11.22	8.2	104	23
Im9 [66]	3.16	12.07	6.6	85	10
ACBP [67]	2.85	13.99	8.2	86	25#
Villin 14T [68]	3.25	12.31	9.8	126	25
N-term L9 [69]	2.87	12.74	4.5	56	25
Ubiquitin [70]	3.19	15.11	7.2	76	25
CI2 [71]	1.75	16.40	7.6	64	25
U1A [72]	2.53	16.91	9.9	102	25
Ada2h [73]	2.88	16.96	4.1	79	25
Protein G [74]	2.46	17.30	4.6	56	25
Protein L [75]	1.78	17.62	4.6	62	22
FKBP [76]	0.60	17.70	5.5	107	25
HPr [77]	1.17	18.35	4.7	85	20
MerP [78]	0.26#	18.90	3.4	72	25
mAcP [79]	–0.64	21.20	4.5	98	28
CspB [4]	2.84	16.40	2.7	67	25
TNfn3 [80]	0.46	17.35	5.3	92	20
TI I27 [80]	1.51	17.82	7.5	89	25
Fyn SH3 [5]	1.97	18.28	6.0	59	20
Twitchin [80]	0.18	19.70	4.0	93	20
PsaE(a)	0.51	17.01	1.57	69	22
Sso7d(b)	3.02	9.54	5.93	63	25

*A nonhomologous set of simple, single-domain, non-disulfide-bonded proteins that have been reported to fold via two-state kinetics under at least some conditions. Reported data and representative members of homologous families selected as previously described [1]. †Extrapolated folding rates in water. May differ from true folding rate in water (e.g. cyt-c, protein G, ubiqutin and others) due to 'roll-over' at low denaturant concentrations. ‡Calculated as previously described [1]. §Length of protein in residues from first structured residue to last. May differ from number of residues in construct characterized. #As reported previously in [2]. (a) P Bowers, D Baker, unpublished data. (b) L Serrano, personal communication.

In the simple zeroth order model, protein topology is the single most important determinant of the folding process because it determines the sequence separation and spatial arrangement of the contacting residues. Indeed, simple computational models based on the homogeneous contact picture have done reasonably well at capturing many of the overall features of protein folding rates and mechanisms [6–9]. However, there are now a number of examples in which differences in local free energies of ordering have a significant influence on the folding mechanism, particularly in cases in which several different pathways are equally consistent with the structure because of symmetry (see below). These differences may arise, for example, from particularly unfavorable local conformations that either are important for functional reasons or are compensated in the final folded structure

Figure 1



Correlation between the logarithm of the folding rate and (a) relative CO, (b) absolute CO and (c) log(relative CO).

by particularly favorable nonlocal interactions. Incorporation of these differences leads to a model in which the order of events in folding depends both on the overall topology and on the relative free energy of ordering different parts of the chain. Given two possible routes to the native state, which involve forming contacts between residues equally distant along the chain, the lowest free energy route is that involving the formation of the lowest free energy substructures. Such a model produces considerably better predictions of the folding rate and of the dominant features of the structure of the folding transition-state ensemble than the simple zeroth order model (see Figures 2 and 3; E Alm, A Morozov, D Baker, unpublished data).

Experimentally, the distribution of structure in the folding transition state can be determined by measuring the effect of mutations throughout the protein on the folding and unfolding rate [10]. Fersht's Φ value notation is a convenient way to summarize such data; a Φ value of one indicates that the interactions made by a residue are as ordered in the transition state as in the native state, whereas a Φ value of zero indicates that the interactions are not formed in the transition state [11]. Table 2 summarizes the general properties of the folding transition states studied so far using this kind of analysis. The following focuses on several recent examples that highlight the interplay between the native state topology and variations in local free energies of ordering in determining the folding mechanism (this is not a comprehensive summary of recent advances in protein folding studies).

GCN4 and λ repressor

The GCN4-p1 coiled coil is a particularly simple system for the detailed examination of the effects of topology and local structural propensity on the distribution of structure in the transition-state ensemble. The rate-limiting step in folding involves the association of two monomers to form a dimer in which hydrophobic residues are partially buried, but the helices are not completely formed. The C-terminal region of the helix exhibits higher helix propensity and mutations in that region have larger effects on the folding rate than mutations in the N terminus [12,13]. Interestingly, the effect of mutations on the folding rate can be altered by manipulating the helix propensity throughout the helix with the help of additional mutations. For example, once the N terminus of the helix is stabilized by two alanine substitutions, a subsequent mutation at the C terminus has a relatively small effect on folding, and when the C terminus is destabilized by a glycine substitution, a subsequent mutation at the N terminus has a much larger effect on folding than in the wild-type protein [12]. Thus, whereas in the wild-type protein the rate-limiting step appears to involve primarily the association of C-terminal portions of the two helices [13], association of the N-terminal regions can nucleate folding if the N terminus is stabilized or the C terminus is destabilized. Such malleability is expected given the symmetry of the helix — it appears that the rate-limiting step involves the pairing of helical regions of the two monomers, but whether these are C-terminal or N-terminal depends on the details of the sequence and can be perturbed by mutations that alter the helix propensity. However, when the symmetry is broken by connecting the N termini of the helices with a covalent cross-link, the portions of the helices adjacent to the (N-terminal) cross-link are largely formed and the C-terminal regions are largely disrupted in the transition state, regardless of the intrinsic helical propensities [12]. Therefore, in this system, local structural biases have some influence on the transition state when multiple folding routes are equally consistent with the overall topology because of symmetry (the dimeric

Table 2

Folding transition states characterized by mutational analysis.

Protein	Fold	Number of residues	Number of mutants	Transition state (TS) characteristics
λ Repressor	α helix	80	8	Some helices are more structured in the TS than others; multiple folding pathways were postulated because of the dramatic effect of single mutations and temperature on TS structure [14,15]
ACBP	α helix	86	26	Terminal helices come together in the TS, while the rest of the protein is involved in non-native interactions; conserved hydrophobic residues are important in the TS [67]
GCN4 coiled coil	α helix			TS for coiled-coil formation is different when the two helices are cross-linked and when they form a dimer [12,13]
Monomer		72	3	
Dimer		36/36	3	
src SH3 domain	β barrel	57	57	TS is structurally polarized, with part of the protein fully formed and the rest fully disordered; TS is conserved among distant sequence homologs [3,22]
α-Spectrin SH3 domain		62	17	
PsaE	β barrel	69	18	These proteins are structural homologs of the SH3 domain, but do not exhibit the same TS (P Bowers, D Baker, unpublished data; L Serrano, personal communication; Q Yi, D Baker, unpublished data)
Sso7d		63	24	
Simplified SH3		56	5	
src SH3 circ	β barrel	57	14	Circularization (circ) makes the TS more delocalized, whereas cross-linking (cross) of the distal hairpin leaves it unchanged [25]
src SH3 cross		57	9	
Spectrin SH3 perm1	β barrel	62	7	Permutation at the distal hairpin, but not at the RT loop, causes a shift in the structure of the TS [81]
Spectrin SH3 perm2		62	8	
TNfn3	β sandwich	92	48	Structurally polarized: a ring of core residues from the central β strands forms the folding nucleus, while the terminal strands are disordered [82]
Ada2h	α/β	81	15	The topology of this fold allows several different TSs, depending on which helix is more structured [19–21,73]
AcP	(βαββαββ)	98	26	
U1A		102	13	
S6		101	?	
Protein L	α/β	62	70	The symmetric topology of the protein allows for two possible TSs, depending on which hairpin is more stable; stabilizing the opposite hairpin leads to a switch in the transition state (protein G_Nu); ([16,17]; S Nauli, B Kuhlman, D Baker, unpublished data)
Protein G	(ββαββ)	57	19	
Protein G_Nu		57	4	
CI2	α/β	64	150	Delocalized TS, with most of the interactions only partially formed [71]
CI2 circ	α/β	64	11	Circularization (circ), circular permutation (perm) and fragmentation (frag) do not change the delocalized TS [83]
CI2 perm		64	11	
CI2 frag		40/24	23	
FKBP	α/β	107	34	[76,84]
CheY	α/β	129	34	[85]
p13suc1	α/β	113	57	[86]
Arc repressor	α/β	53	44	Delocalized TS [87]

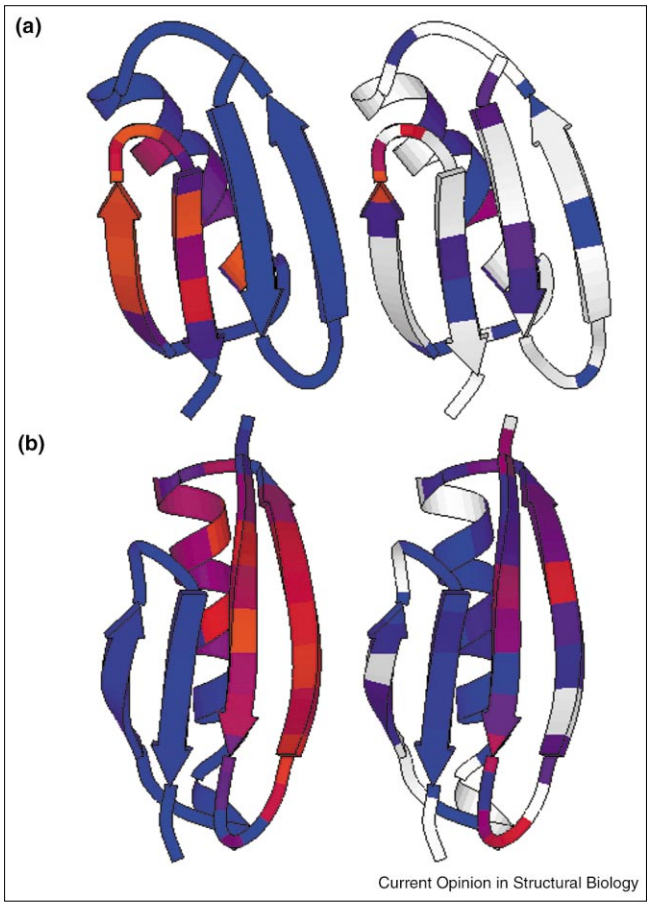
case). However, when the topology strongly favors one particular route to the native state because of the reduced entropic cost of forming more local interactions (the monomeric case), secondary structure propensities are of little consequence.

The λ repressor, another all α-helical protein, has also been postulated to fold by a number of pathways, depending on the intrinsic stability of each helix. Both point mutations [14] and temperature [15] have been shown to significantly change structure in the transition state.

Protein L and protein G

Protein L and protein G are structural homologs, but have little detectable sequence similarity. Both proteins consist of an α helix packed across a four-stranded sheet formed by two symmetrically disposed β hairpins. Remarkably, the symmetry of the fold is almost completely broken during folding: in protein L, the first hairpin is formed and the second disrupted at the rate-limiting step in folding, whereas in protein G, the second hairpin is formed and the first is disrupted [16,17] (Figure 2). Thus, despite the small size (~60 residues) of the two proteins and their topological

Figure 2



Folding transition states of (a) protein G and (b) protein L. Left, predicted phi values; right, experimental phi values. The color scheme is continuous from red (Φ = 1; structured in the transition state) to blue (Φ = 0; unstructured in the transition state). Sites not probed experimentally are indicated in white. Graphics were generated with MOLSCRIPT [88]. Predicted phi value distributions were obtained from the highest free energy configurations along the lowest free energy paths between the unfolded and native states, as described in [6], except that additional terms representing hydrogen bonding and local sequence/structure preferences were included in the free energy function. The second β hairpin is favored by the computational model for protein G, because of an extensive hydrogen-bond network, and the first hairpin is favored by the model for protein L, because the second β turn has considerable torsional strain (three consecutive residues with positive phi angles).

symmetry, there is a definite hierarchy to structure formation. The characterization of the two transition states suggests that the lowest free energy route to the native state for this fold involves formation of one of the two β hairpins; however, the choice of hairpin is determined by factors beyond native state topology. Interestingly, with the addition of hydrogen bonding and sequence- and structure-dependent local free energies of ordering, the simple computational model described above [6] recapitulates the experimentally observed symmetry breaking (Figure 2).

The correspondence between the predicted and experimentally determined phi values suggests that the hairpin formed at the rate-limiting step is the one with the lowest free energy of formation. To test this hypothesis, computational protein design methods [18] have recently been used to specifically stabilize the first β hairpin of protein G, which, as noted above, is not formed in the transition state in the wild-type protein. A redesigned protein G variant with a more optimal backbone conformation and sequence in the first hairpin folds 100-fold faster than the wild-type protein. Subsequent mutational analysis shows that the first β hairpin, rather than the second β hairpin (as in the wild-type), is formed in the transition state in the redesigned protein (S Nauli, B Kuhlman, D Baker, unpublished data). Likewise, following stabilization by redesign of the second hairpin of protein L, which contains three consecutive residues with positive phi angles in the wild-type structure, and destabilization of the first hairpin, the second hairpin was found to be better formed in the folding transition-state ensemble than the first turn (D Kim, B Kuhlman, D Baker, unpublished data). These switches in folding mechanism highlight the differences local free energies of ordering can have when the overall topology has considerable symmetry.

AcP, Ada2h, U1A and S6

The folding transition states of four proteins with the ferredoxin-like fold (two helices packed against one side of a five-stranded β sheet) have been characterized. The folding transition states of Ada2h (activation domain of procarboxypeptidase) and AcP (acylphosphatase) are similar, despite the low sequence similarity (13%) between the two proteins and variations in the length of the secondary structural elements [19,20]. In both cases, the overall topology of the protein appears to be already specified in the transition state, but the second α helix and the inside strands of the β sheet with which it interacts appear to be more ordered than the rest of the polypeptide chain. The characterization of two other members of this structural family, however, revealed an alternative nucleus with preferential structure around helix 1: U1A nucleates in helix 1 and S6 nucleates in both helices [21]. The topology appears to allow several roughly equivalent folding pathways: the choice of the dominant pathway may be determined by the detailed packing and orientation of structural elements. Proteins with this fold also exhibit a pronounced movement of the transition state from 20% to 80% native (as judged by the burial of surface area) with increasing concentration of denaturant. Remarkably, given the variation in the transition-state structure, the folding rates of these proteins are highly correlated with the CO over an approximately 4000-fold range of folding rates. Furthermore, changing the CO can significantly change the folding rate: a circular permutant of U1A with CO lower than that of the wild-type protein folds considerably faster (M Oliveberg, personal communication).

SH3 domain fold

SH3 family

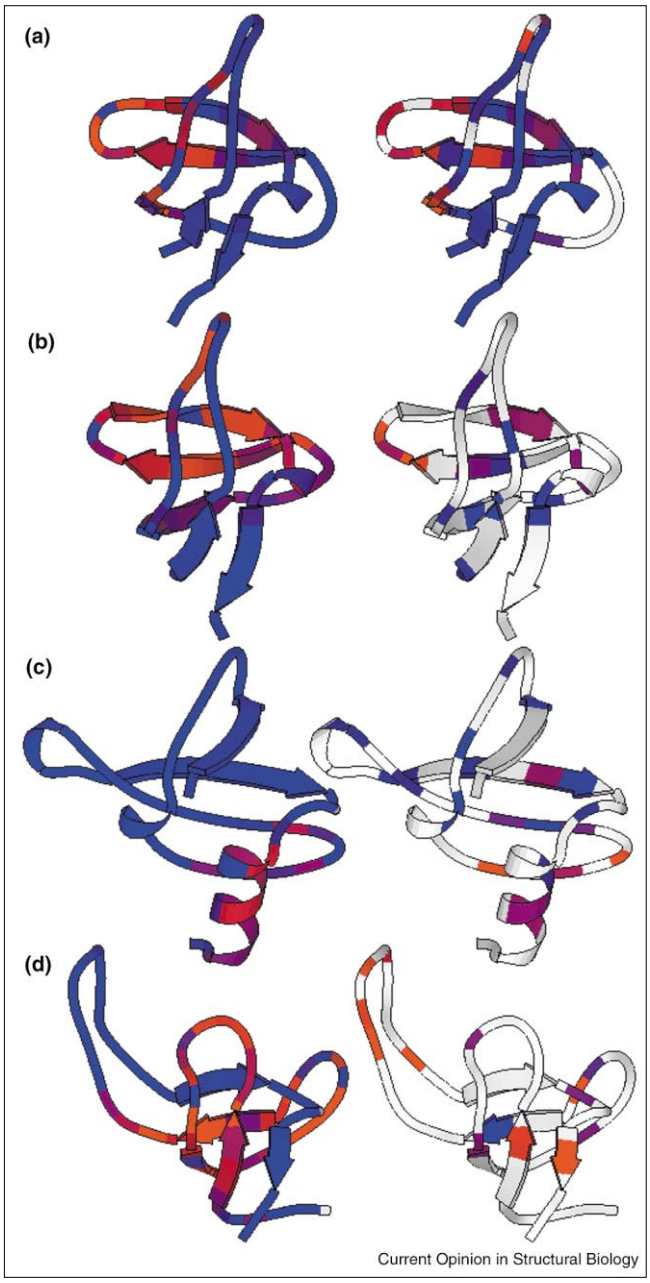
The homologous src and α-spectrin SH3 domains exhibit very similar transition states [3,22–24], despite the low

sequence identity (36%) (Figure 3a,b). Stabilizing mutations [23] and changes in pH [22] do not seem to affect the structure of the transition state of the α-spectrin SH3 domain. In the case of the src SH3 domain, stabilization of local structure by hairpin cross-linking and global stabilization by sodium sulfate do not alter the placement of the transition state along the reaction coordinate [25]. It appears, then, that SH3 domains allow quite large variations in sequence and experimental conditions with no change to the transition state, probably because there are no alternative structural elements that can be sufficiently stabilized to become folding nuclei. On the other hand, modifying the topology of the protein can significantly change the free energy landscape to favor alternative routes for folding. Circularization of the src SH3 domain causes the delocalization of structure in the transition state [25]. Circular permutation experiments on the α-spectrin SH3 domain also changed the transition state [26]. Connecting the wild-type termini with a small peptide linker and introducing a cut in the distal hairpin resulted in a shift in the structure of the transition state towards the n-src loop and the hairpin formed by the old termini; the former distal hairpin was completely disordered at the rate-limiting step. Therefore, shifts in transition-state structure can occur when formerly distant elements are covalently linked to reduce the entropic cost of their interaction. Drastic mutagenesis, which weakens the interaction energies throughout the protein, can also change the transition state. For example, a sequence-simplified mutant of the src SH3 domain made predominantly of five amino acids (isoleucine, lysine, glutamic acid, alanine and glycine) was found to have a more delocalized transition state (distal hairpin is not fully formed); the interactions stabilizing the wild-type SH3 transition state may not be strong enough in the simplified mutant to overcome the loss in entropy and residues from other parts of the protein may have to participate (Q Yi, D Baker, unpublished data).

SH3 structural analogs

The characterization of SH3 structural analogs has shown that transition-state structure is not always conserved in proteins with similar topologies. PsaE [27], a photosystem protein from cyanobacteria, has a large loop insertion at the distal hairpin (13 amino acids), making it entropically more costly to form stabilizing interactions. As a result, its transition state is more delocalized than that of the src SH3 domain, with well-ordered residues found in the distal hairpin, as well as in the N and C termini (P Bowers, D Baker, unpublished data) (Figure 3d). Sso7d, a DNA-binding protein from *Sulfolobus solfataricus* [28], has a significantly different transition state from that of the src and α-spectrin SH3 domains. The n-src loop and the C terminus (which is a helix in Sso7d, instead of a β strand) are the most structured in the transition state, whereas the distal hairpin is only weakly ordered (R Guerois, L Serrano, personal communication) (Figure 3c). This is in contrast to the src and α-spectrin SH3 transition states, in which the distal hairpin is completely ordered. In the SH3 domains and in Sso7d, the contiguous three-stranded sheet is formed but, in one case, the diverging turn interacts with it, whereas in the other case, it is the C-terminal helix. This difference may reflect variations in the free energies of forming the structural elements: in the SH3

Figure 3



Folding transition states of proteins with the SH3 fold: **(a)** src SH3 domain, **(b)** spectrin SH3 domain, **(c)** Sso7d and **(d)** PsaE. Left, predicted phi values (see legend to Figure 2); right, experimental phi values. The color scheme is described in the legend to Figure 2. The distal loop is favored over the n-src loop by the computational model for the src SH3 domain because it has more extensive hydrogen bonding, whereas the equivalent of the distal loop is disfavored by the model for Sso7d because it contains five glycine residues that are costly to order.

domains, the distal loop hairpin is well packed and the n-src loop is irregular, whereas in Sso7d, the opposite is the case — the equivalent of the distal hairpin contains five consecutive glycine residues (which are likely to be functionally important). With the inclusion of hydrogen bonding and sequence- and structure-dependent local free energies of ordering, the simple computational model described above [6] produces phi values very similar to those observed experimentally for the SH3 domains and Sso7d. Similar results were very recently published by Guerois and Serrano (R Guerois, L Serrano, unpublished data; see Now published).

In summary, folding transition-state structure is conserved more highly within the SH3 sequence superfamily than among SH3 analogs. The SH3 topology, then, although not as obviously symmetric as the protein L/protein G topology, still allows several alternative folding routes. The prevalence of one route over the other depends on the details of the structure. This may, in part, be due to the fact that functional constraints lead to the conservation within, but not between, superfamilies of portions of protein structures with unusual local features (the irregular n-src and RT loops in the SH3 domain, for example, are involved in proline-rich peptide binding) with higher free energies of formation. These features partially determine which of the pathways consistent with the native state topology is actually chosen.

The GCN4 and protein G experiments, together with the comparisons of transition-state structures in the AcP and SH3 families, suggest a picture in which several different 'pathways' with roughly equivalent free energy barriers can be consistent with the overall topology. Sequence changes can, by lowering or raising one barrier relative to another, produce significant changes in the transition-state ensemble without large changes in folding rate. Consistent with this picture, our most recent models of the folding process produce considerably more accurate predictions of folding transition-state structures when local free energies of ordering based on sequence-dependent backbone torsion angles and local hydrogen bonding terms are included. We anticipate considerable synergy between theory and experiment, and an important role for computational protein design methods in the further elucidation of the mechanisms of protein folding during the next few years.

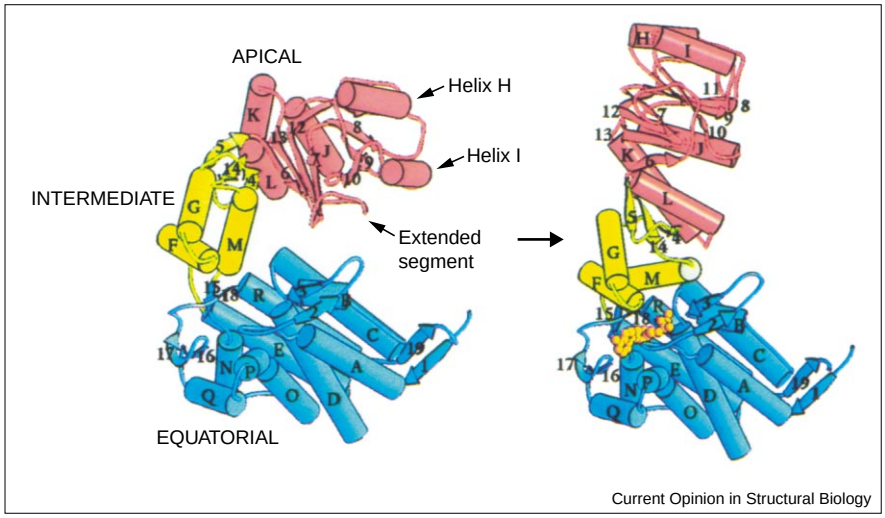
GroEL–GroES-assisted folding

How do the foregoing simple concepts apply to chaperone-assisted folding? In small proteins, the largest free energy barriers to folding involve the formation of particularly nonlocal portions of protein structures and regions with particularly unfavorable local energetics. It seems possible, therefore, that larger proteins containing such features may be particularly dependent on chaperones for suppressing alternative off-pathway misfolding/aggregation. Kinetic bottlenecks caused by unfavorable local structures or high contact order regions may tilt the kinetic competition between on- and off -pathway reactions in favor of the latter. It should be emphasized, however, that non-native interactions are likely to play a greater role in the folding of larger proteins simply because the increased size of the protein increases the probability of low free energy non-native interactions. Chaperones act on such non-native states in the first instance by binding the hydrophobic surfaces that are exposed, preventing these surfaces from 'wrongful interactions' that lead to multimolecular aggregation. Binding may, in some cases, be associated also with at least partial unfolding, as discussed below for GroEL. Release from the chaperones, in many cases driven by ATP binding (not hydrolysis), then allows the substrate polypeptide a chance to fold. Uniquely, in the case of the chaperonin ring class of chaperones, polypeptide is released into an encapsulated chamber where folding proceeds in isolation. In the case of the bacterial chaperonin, GroEL, this is mediated by ATP/GroES binding, which is associated with rigid-body movements of the GroEL intermediate and peptide-binding apical domains of the bound ring [29] (see Figure 4). The 60° elevation and 90° twisting of the apical domains act to remove the hydrophobic peptide-binding sites away from the central cavity, releasing polypeptide into this GroES-encapsulated space. Because the character of the wall of the cavity is switched from hydrophobic to hydrophilic as the result of the rigid-body movements, it may influence the released polypeptide to fold in this space because burial of exposed hydrophobic surfaces and exposure of hydrophilic surfaces, features of the native state, will be energetically favored.

Both cryo-EM reconstructions [30] and high-resolution crystal structures have resolved the rigid-body domain movements of the GroEL–GroES machinery itself during the reaction cycle [29,31] (see Figure 4). In addition, there are dynamic fluorescence and kinetic studies indicating, respectively, rapid release of bound polypeptide into the central cavity upon ATP/GroES binding ($t_{1/2}$ ~1 s) and productive folding inside the GroEL–GroES cavity [32–34]. However, the exact effects of the various states and transitions of the GroEL–GroES machinery during the reaction cycle on the conformation of polypeptide substrates are not well understood because, as ensembles of unstable non-native states, the substrates are much less accessible to structural study, particularly in the presence of the megadalton GroEL ring structure. Thus, our 'view' of what is happening to substrate proteins themselves during the GroEL–GroES reaction is poorly resolved. At this point, the study of stringent substrates, which are dependent on the complete system to reach their native form and are unable to productively fold without it, seems valuable for identifying and characterizing the full range of steps in the reaction that are critical to producing the native state. Nevertheless, there can also be value to studying nonstringent substrates, particularly those whose nonchaperoned folding is well described, because folding behavior can be compared in the presence and absence of

Figure 4

Rigid-body movements of a GroEL subunit attendant to ATP/GroES binding. Rigid-body rotations about the top and bottom of the intermediate domain redirect the peptide-binding surface of the apical domain, composed of helices H and I and an underlying extended segment, from a position facing the central cavity (lying to the right of the subunit) to a new position facing out of the page. The binding of peptides in the groove between helices H and I, through contacts with resident hydrophobic sidechains, has been observed (see text). Although the involvement of the extended segment of the apical domain in polypeptide binding has been indicated by mutational studies, a structural basis for such interaction remains undefined (adapted from [29]).



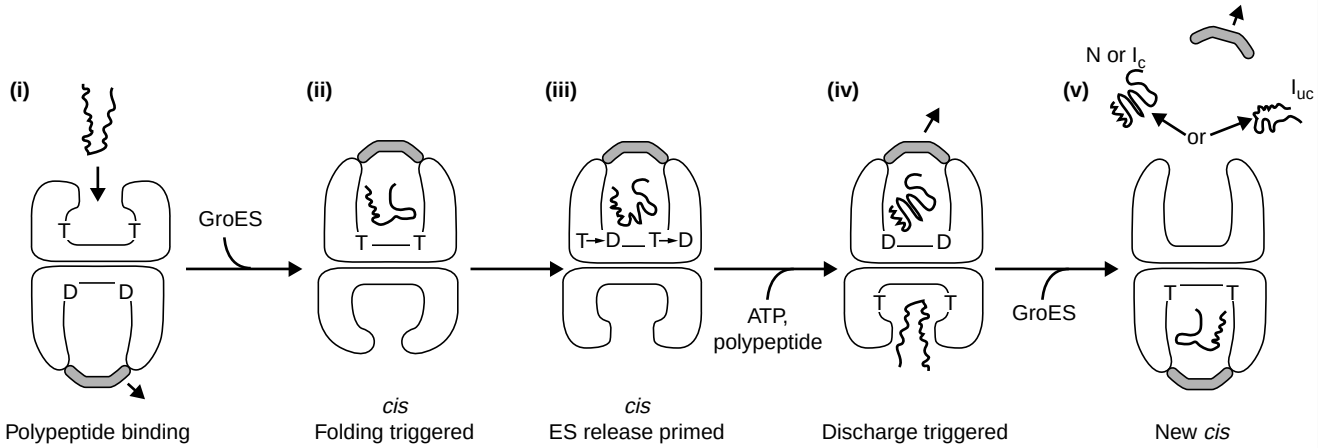
chaperonin. Even small peptides may, to some extent, simulate the behavior of a region of polypeptide chain, at least in binding to GroEL.

Binding to GroEL – potential unfolding action

There are definable points in the GroEL–GroES reaction cycle (Figure 5) at which major actions on polypeptide substrates have been considered likely to occur. One is the step of polypeptide binding to an open GroEL ring (which, under physiological conditions, would be the open ring of a GroEL–GroES–ADP asymmetric complex) [35] (see Figure 5). Binding may be associated with at least partial unfolding of a substrate protein, which is potentially a means for removing a non-native form from a kinetic trap. This could occur through either or both of two mechanisms, one catalytic, in which GroEL lowers the energy barriers between various non-native states, the other thermodynamic, in which GroEL preferentially binds less-folded states without affecting the transition states between the various conformations. The best evidence to date for a catalytic unfolding action associated with binding comes from a hydrogen-deuterium exchange experiment showing that GroEL in catalytic amounts can globally unfold the 6 kDa protein barnase [36]. Whether GroEL can exert similar effects on larger proteins, including those that form stable binary complexes with it, remains unclear. A number of exchange studies carried out with stable binary complexes of such proteins as α-lactalbumin [37], human dihydrofolate reductase [38,39] and Rubisco (ribulose-1,5-bisphosphate carboxylase-oxygenase) [40•] indicate that these proteins do not become globally exchanged while bound to GroEL, exhibiting modest levels of amide proton protection that are, in some cases, localized (but see, however, [41,42], which showed that cyclophilin and a chemically denatured β-lactamase, respectively, were completely exchanged while bound). In the case of Rubisco, it was possible to examine the protein both while in a metastable intermediate state in solution and after becoming bound to GroEL [40•]. In this case, a high degree of protection from exchange was observed for a small number of amide protons both in the metastable intermediate in solution and in the binary complex with GroEL. Thus, whatever the nature of this secondary structure(s), it appears to be resistant to the unfolding action associated with GroEL binding. Some proteins, however, may nevertheless be subject to catalyzed unfolding at a local level during the process of binding to GroEL.

The thermodynamic mechanism for unfolding in the presence of GroEL involves the greater affinity of GroEL for less-folded states among an ensemble of conformers that are in equilibrium with each other [43]. This would effectively shift the equilibrium by mass action toward the less-folded states. Perhaps the best evidence supporting an action of this sort comes from study of an RNase T1 mutant that populates two non-native states, one more structured than the other [44]. In the presence of GroEL, the less-folded state became more populated, without alteration of the microscopic rate constants between the two states, arguing for a thermodynamic effect (see also [42,45,46] for descriptions of such effects on β-lactamase, dihydrofolate reductase and barstar). Such partitioning between non-native states has yet to be demonstrated for stringent substrates, although the ability of GroEL to inhibit the production of off-pathway aggregates of malate dehydrogenase (MDH) has been kinetically modeled to such a mechanism. In the model, GroEL favors binding of MDH monomers and shifts an equilibrium of low-order aggregates of MDH toward this state [47]. Clearly, the ability to resolve different conformational states within an ensemble of substrate proteins, both unbound and GroEL-bound, using spectroscopic techniques, for example, will be necessary to better characterize the behavior of an open GroEL ring toward its substrates. Both catalytic and thermodynamic mechanisms could be

Figure 5



GroEL–GroES reaction cycle. Non-native polypeptide is bound in the open (*trans*) ring of an asymmetric GroEL–GroES–ADP (D) complex via hydrophobic interactions with the surrounding apical domains (panel i). Binding of ATP (T) and GroES to the same ring as the polypeptide produces large rigid-body movements in the subunits of the ring, elevating and twisting the hydrophobic binding surface away from the bound polypeptide, releasing it into the encapsulated and now hydrophilic *cis* chamber where folding commences (panel ii). After 8–10 s, ATP hydrolysis occurs in the seven subunits of the folding-active ring, relaxing the affinity of the ring for GroES and 'priming' it for release (panel iii). At the same time, *cis* hydrolysis produces an allosteric adjustment of the *trans* ring that allows rapid entry of ATP and non-native polypeptide (panel iv). The arrival of ATP triggers allosteric dissociation of the *cis* ligands (panel v); the binding of non-native polypeptide serves to accelerate the rate of this departure by 30–50-fold. Note that the polypeptide can be ejected in either a native form (N), a form committed to reaching the native state in the bulk solution (I_c) or an uncommitted non-native state (I_{uc}) that can be rebound by chaperonin. The relatively slow binding of GroES to the new ATP/polypeptide-bound ring orders the formation of the next folding-active GroEL–GroES complex (panel v). Thus, GroEL alternates rings back and forth as folding-active, expending the ATP of one ring to simultaneously initiate a new folding reaction, while dissociating the previous one from the opposite ring. As discussed in the text, polypeptide binding in an open GroEL ring (panels i and iv) may be associated with an action of unfolding. The step of ATP/GroES binding may also produce forced mechanical unfolding (panels ii and v).

operative, depending on the particular substrate and its position on the landscape. Finally, although the binding of substrate proteins is usually thought of as redirecting off-pathway states, there seems no reason to exclude that, in at least some cases, GroEL could recognize on-pathway intermediates, which could also receive kinetic assistance as a result of recruitment to the GroEL–GroES cavity.

Both catalytic and thermodynamic unfolding mechanisms could be enabled by the ability of the multiple surrounding GroEL apical domains to interact with a substrate protein. Such multivalent binding was recently indicated by an experiment with covalent GroEL rings bearing various numbers and arrangements of binding-proficient and binding-incompetent apical domains [48•]. A minimum of three consecutive proficient domains was required for efficient binding of a stringent substrate protein. In agreement, an accompanying experiment employing cysteine cross-linking between a bound substrate protein and a GroEL ring observed cross-links with multiple GroEL apical domains.

Translating binding action back to structure – what does GroEL recognize?

Ultimately, it would be desirable to translate the foregoing actions associated with chaperonin binding into structural terms. Lacking, however, any high-resolution information on the structure of a substrate protein bound to GroEL, we can only extrapolate from a variety of different types of experimental information, which, in the past year, has been derived from proteomic, biochemical, spectroscopic and crystallographic studies. At the level of binding to individual apical domains, a crystallographic study observed that a dodecamer peptide, selected for its high affinity for an isolated apical domain, associated with it as a β hairpin, both in a co-crystal with an isolated apical domain and in one with full occupancy of the apical domains of the GroEL tetradecamer [49•]. In these structures, one strand of the hairpin contacted the apical domain at a position between the two α helices (H and I) facing the central cavity (see Figure 4). A host of hydrophobic contacts were formed between tryptophan and phenylalanine residues in the peptide and hydrophobic sidechains in the two α helices; these helices had been previously implicated in polypeptide binding by a mutagenesis study [50] and by a previous crystallographic study of an apical domain [51]. In the latter study, similar topology and contacts were observed between an extended N-terminal tag segment of one monomer found lying in the groove between these two α helices in a neighboring monomer in the asymmetric unit. In the dodecamer study, it was additionally noted that, compared with the unoccupied isolated apical domain crystal structure, in which a

number of regions, including the channel-facing ones, were found to differ somewhat in positioning between monomers in the asymmetric unit, the conformations of the isolated domains with peptide bound became virtually identical. This suggests that there is a structural plasticity to the apical binding surface that accommodates the variety of substrates and that, upon contact with a particular substrate, optimizes contacts with it.

Lest it seem that only β strands can associate with the GroEL apical domain, two different NMR studies re-examined an N-terminal 13-residue peptide from the substrate rhodanese that is known to form an α helix in the intact native protein. This peptide had been observed through transfer NOE effects to adopt an α-helical structure upon association with intact GroEL [52]. In the first of the new studies, the same transfer NOE effects were observed when the peptide was incubated with an isolated GroEL apical domain, and chemical shift changes could be observed that localized to the same two cavity-facing α helices (H and I) [53]. In the second study, carried out with intact GroEL, D and D,L chiral forms of the same peptide were observed to bind as well as the original L form [54]. Whereas the D form could form a left-handed helix in TFE, the D,L form did not form α helix. This suggested that the hydrophobic content of the peptides was more critical to binding than adoption of a particular secondary structure. Two dodecameric α-helical peptides with the same composition were also compared, observing that one with hydrophobic sidechains clustered on one side of the predicted helix opposite hydrophilic sidechains (amphiphilic character) bound more strongly than another peptide interspersing hydrophobic sidechains with hydrophilic sidechains. This suggested that a contiguous hydrophobic surface is the feature in a substrate favoring its recruitment to GroEL. In a third study, a series of 14-residue peptides that exhibited α-helical character in solution was examined [55]. In this case also, those peptides with amphiphilic character were found to bind most strongly to GroEL, some with submicromolar affinity.

Thus, GroEL appears able to recognize both major secondary structural elements, so long as hydrophobic surface is presented. It remains curious, however, that, where examined, recognition appears to occur through the same two apical α helices without recognizable participation of an underlying extended segment (amino acids 199–209; see Figure 4) that also bears hydrophobic residues, mutation of which abolishes polypeptide binding. Thus, the question remains as to whether this segment participates directly in binding. Notably, the H and I α helices also form the major contacts with the GroES mobile loop (itself in an extended state), also through hydrophobic interactions, after elevation and twisting of the apical domains [29] (see Figure 4). Thus, binding through these two α helices may be an energetically favored mode, although polypeptide and GroES binding occur at two very different points in space.

Both major secondary structural elements figure together in a proteomic study identifying several dozen proteins from *Escherichia coli* that could be co-immunoprecipitated with anti-GroEL antiserum upon cell lysis in EDTA (to inhibit nucleotide-driven dissociation) [56]. Whether any of these are stringent substrates, that is, dependent on GroEL–GroES for proper folding, remains to be seen, but of this collective of bound species, where a structure of the native form was available, the topology favored was αβ, with two or more domains. Thus, it seems plausible that GroEL multivalently binds individual α and β units through exposed hydrophobic aspects that will be buried together in the native state. This potentially stabilizes the individual domains against inappropriate intermolecular or even intramolecular interactions until ATP/GroES-driven release directs an optimal chance for correct association within the molecule, while it is confined to the *cis* cavity. A direct illustration of such putative action comes from a study of the folding of four-disulfide hen lysozyme, composed of an α and β domain, in the presence of GroEL [57]. The open GroEL ring accelerated the rate of acquisition of the native state by 1.3-fold, without affecting the rate or mechanism of domain folding. Rather, GroEL accelerated the slower step of proper docking of the two domains, presumably by binding one or both individual domains and disfavoring or reversing non-native contacts.

ATP/GroES-driven release of GroEL-bound substrate into the central cavity – potential unfolding action

The action of ATP/GroES binding on polypeptide conformation, associated with release into the GroEL–GroES cavity, has been of major interest. An earlier study of the substrate Rubisco, examining its tryptophan fluorescence anisotropy, observed a rapid drop ($t_{1/2}$ ~1 s), followed by a slow rise correlating with production of the native state [32]. The nature of the fast phase had been a mystery, but an exchange experiment with tritium-labeled Rubisco has begun to address this [40•]. A metastable intermediate of this protein exhibited 12 highly protected amide tritiums both in solution and while bound to GroEL. When ATP and GroES were added, all but two of the tritiums were exchanged by 5 s, the earliest time examined. The elevation and twisting of the apical domains, driven by ATP/GroES binding to a polypeptide-bound ring, were proposed to produce a stretching of substrate between the apical domains before complete release into the cavity. Such a mechanism would couple the energy of ATP/GroES binding to a forced unfolding action. But the deprotection observed does not seem fully accountable only by a stretching action exerted on molecules becoming encapsulated in the *cis* ring. Consider the experimental observation that GroES binds randomly to either of the two GroEL rings of a Rubisco–GroEL binary complex to form two different asymmetric complexes: approximately 50% *cis* ternary complexes and approximately 50% *trans* ternary complexes, the latter with GroES on the ring opposite the polypeptide-bound one. Thus, one would expect

that, at a time (here, 5 s) less than that of a single turnover (~10 s), only about half of the tritiums should be deprotected, corresponding to those of the Rubisco molecules that had become encapsulated in *cis*. Yet nearly all were deprotected, suggesting that molecules in the *trans* ring must likewise have been deprotected. Previous studies have indicated that the *trans* ring of a *cis* complex in ATP has no significant affinity for non-native Rubisco [35], thus suggesting that any deprotection of Rubisco bound on that ring must be associated with its release into the bulk solution. Perhaps there is also a twisting action on that ring, attended by unfolding during release. Alternatively, simple release without unfolding may be sufficient to produce deprotection if, for example, the protection derives from association of the substrate with the GroEL cavity wall (either through direct hydrogen bond formation or via steric shielding of amide protons). Thus, more needs to be learned about whether forced unfolding is really occurring in this case, whether it is a general aspect of the chaperonin mechanism and whether substrate polypeptides bound in *trans* are somehow also affected. Furthermore, it remains to be demonstrated whether such an action is required for productive Rubisco folding.

In a further experiment, the kinetics of tritium exchange of the metastable Rubisco intermediate was examined in the presence of substoichiometric concentrations of GroEL–GroES. The observed rate of decay indicated that molecules whose tritiums had been exchanged were subsequently being released from *cis* complexes in non-native forms that competed with the remaining pool of still tritium-labeled Rubisco molecules for binding to GroEL [40•]. This reflects, as established by earlier studies, the occurrence of multiple rounds of binding and release of non-native polypeptide from GroEL during a productive folding reaction, underscoring the trial-and-error process of achieving the native state, as opposed to a process in which non-native forms remain at GroEL until productive folding occurs (see Figure 5). Indeed, in a stoichiometric reaction, only a few percent of Rubisco molecules reach native form in what corresponds to any given round of folding at chaperonin. Addition of 'trap' versions of GroEL, able to bind but not release non-native forms, rapidly halts a reaction, with non-native substrate physically accumulated at the trap (e.g. [58,59]). Such observations also reflect on the model for forced unfolding, indicating that, in and of itself, even if it occurs, such an action is not sufficient for producing the native state; otherwise, multiple rounds would not be required.

By contrast, when a stable, long-lived (>100 min) *cis* complex is formed between SR1, the single-ring version of GroEL, and GroES, it produces nearly 100% recovery of native Rubisco inside the *cis* cavity. This indicates a major role, if not a dominant one, for the encapsulated *cis* space in producing the native state (see also [60,61]). Furthermore, as suggested by kinetic studies with MDH, non-native molecules expelled into the bulk solution during a normal folding reaction with wild-type GroEL (where the lifetime of a *cis* complex is ~10 s) can form low-order aggregates on a short time-scale [47], in part explaining why such released forms fail to achieve the native state in the bulk solution. In contrast, MDH molecules held in a stable *cis* complex (inside SR1–GroES) are forestalled from such aggregation and are productively folded essentially quantitatively [32].

Productive folding in the GroEL–GroES cavity

Although features of the GroEL–GroES cavity that favor productive folding have been identified from crystallographic study, the path inside it that a protein takes to the native state is unknown. Does this chamber simulate an infinite dilution condition? Perhaps it can for smaller polypeptides, but the physical dimensions argue for close confinement of larger substrates like Rubisco. Experimentally, even in its native state, the smaller protein GFP appeared to be tumbling into the walls of this space, with a rotational correlation time of 42 ns, instead of the 12 ns observed in solution [33]. Perhaps such confinement presents limits to the conformational space that can be explored by non-native forms, limiting their folding trajectory. Clearly, a comparison of folding in this *cis* cavity with folding at infinite dilution would be instructive and might be possible using single-molecule techniques.

Conclusions

In sum, then, for both the folding of small two-state folding proteins and chaperonin action on larger ones, some basic outlines of mechanism are now available. Yet it seems likely that there will be still other basic mechanistic principles concerning these reactions that lie as yet unrecognized.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Plaxco KW, Simons KT, Baker D: Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 1998, 277:985-994.
2. Plaxco KW, Simons KT, Ruczinski I, Baker D: Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry* 2000, 39:11177-11183.
3. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D: Experiment and theory highlight role of native state topology in SH3 folding. *Nat Struct Biol* 1999, 6:1016-1024.
4. Perl D, Welker C, Schindler T, Schroder K, Marahiel MA, Jaenicke R, Schmid FX: Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat Struct Biol* 1998, 5:229-235.
5. Plaxco KW, Guijarro JI, Morton CJ, Pitkeathly M, Campbell ID, Dobson CM: The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry* 1998, 37:2529-2537.
6. Alm E, Baker D: Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc Natl Acad Sci USA* 1999, 96:11305-11310.
7. Debe DA, Goddard WA III: First principles prediction of protein folding rates. *J Mol Biol* 1999, 294:619-625.
8. Galzitskaya OV, Finkelstein AV: A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc Natl Acad Sci USA* 1999, 96:11299-11304.

9. Muñoz V, Eaton WA: A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 1999, 96:11311-11316.

10. Fersht AR, Matouschek A, Serrano L: The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 1992, 224:771-782.

11. Fersht AR: Characterizing transition states in protein folding: an essential step in the puzzle. *Curr Opin Struct Biol* 1995, 5:79-84.

12. Moran LB, Schneider JP, Kentsis A, Reddy GA, Sosnick TR: Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking. *Proc Natl Acad Sci USA* 1999, 96:10699-10704.

13. Zitzewitz JA, Ibarra-Molero B, Fishel DR, Terry KL, Matthews CR: Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J Mol Biol* 2000, 296:1105-1116.

14. Burton RE, Huang GS, Daugherty MA, Calderone TL, Oas TG: The energy landscape of a fast-folding protein mapped by Ala→Gly substitutions. *Nat Struct Biol* 1997, 4:305-310.

15. Myers JK, Oas TG: Contribution of a buried hydrogen bond to lambda repressor folding kinetics. *Biochemistry* 1999, 38:6761-6768.

16. Kim DE, Fisher C, Baker D: A breakdown of symmetry in the folding transition state of protein L. *J Mol Biol* 2000, 298:971-984.

17. McCallister EL, Alm E, Baker D: Critical role of beta-hairpin formation in protein G folding. *Nat Struct Biol* 2000, 7:669-673.

18. Kuhlman B, Baker D: Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci USA* 2000, 97:10383-10388.

19. Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM: Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat Struct Biol* 1999, 6:1005-1009.

20. Taddei N, Chiti F, Fiaschi T, Bucciantini M, Capanni C, Stefani M, Serrano L, Dobson CM, Ramponi G: Stabilisation of alpha-helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding. *J Mol Biol* 2000, 300:633-647.

21. Ternstrom T, Mayor U, Akke M, Oliveberg M: From snapshot to movie: phi analysis of protein folding transition states taken one step further. *Proc Natl Acad Sci USA* 1999, 96:14854-14859.

22. Martinez JC, Serrano L: The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* 1999, 6:1010-1016.

23. Martinez JC, Pisabarro MT, Serrano L: Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat Struct Biol* 1998, 5:721-729.

24. Grantcharova VP, Riddle DS, Baker D: Long-range order in the src SH3 folding transition state. *Proc Natl Acad Sci USA* 2000, 97:7084-7089.

25. Grantcharova VP, Baker D: Circularization changes the folding transition state of the src SH3 domain. *J Mol Biol* 2001, in press.

26. Viguera AR, Jimenez MA, Rico M, Serrano L: Conformational analysis of peptides corresponding to beta-hairpins and a beta-sheet that represent the entire sequence of the alpha-spectrin SH3 domain. *J Mol Biol* 1996, 255:507-521.

27. Falzone CJ, Kao YH, Zhao J, Bryant DA, Lecomte JT: Three-dimensional solution structure of PsaE from the cyanobacterium ***Synechococcus*** sp. strain PCC 7002, a photosystem I protein that shows structural homology with SH3 domains. *Biochemistry* 1994, 33:6052-6062.

28. Baumann H, Knapp S, Lundback T, Ladenstein R, Hard T: Solution structure and DNA-binding properties of a thermostable protein from the archaeon ***Sulfolobus solfataricus***. *Nat Struct Biol* 1994, 1:808-819.

29. Xu Z, Horwich AL, Sigler PB: The crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex. *Nature* 1997, 388:741-750.

30. Roseman AM, Chen S, White H, Braig K, Saibil HR: The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell* 1996, 87:241-251.

31. Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB: The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* 1994, 371:578-586.

32. Rye HS, Burston SG, Fenton WA, Beechem JM, Xu Z, Sigler PB, Horwich AL: Distinct actions of cis and trans ATP within the double ring of the chaperonin GroEL. *Nature* 1997, 388:792-798.

33. Weissman JS, Rye HS, Fenton WA, Beechem JM, Horwich AL: Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell* 1996, 84:481-490.

34. Ranson NA, Burston SG, Clarke AR: Binding, encapsulation and ejection: substrate dynamics during a chaperonin-assisted folding reaction. *J Mol Biol* 1997, 266:656-664.

35. Rye HS, Roseman AM, Chen S, Furtak K, Fenton WA, Saibil HR, Horwich AL: GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell* 1999, 97:325-338.

36. Zahn R, Perrett S, Stenberg G, Fersht AR: Catalysis of amide proton exchange by the molecular chaperones GroEL and SecB. *Science* 1996, 271:642-645.

37. Robinson CV, Gross M, Eyles SJ, Ewbank JJ, Mayhew M, Hartl FU, Dobson CM, Radford SE: Conformation of GroEL-bound alpha-lactalbumin probed by mass spectrometry. *Nature* 1994, 372:646-651.

38. Gross M, Robinson CV, Mayhew M, Hartl FU, Radford SE: Significant hydrogen exchange protection in GroEL-bound DHFR is maintained during iterative rounds of substrate cycling. *Protein Sci* 1996, 5:2506-2513.

39. Goldberg MS, Zhang J, Sondek S, Matthews CR, Fox RO, Horwich AL: Native-like structure of a protein-folding intermediate bound to the chaperonin GroEL. *Proc Natl Acad Sci USA* 1997, 94:1080-1085.

40. • Shtilerman M, Lorimer GH, Englander SW: Chaperonin function: folding by forced unfolding. *Science* 1999, 284:822-825.
Evidence from tritium exchange that Rubisco is 'stretched on the rack' in the act of ATP/GroES binding to a Rubisco-bound GroEL ring.

41. Zahn R, Spitzfaden C, Ottiger M, Wuthrich K, Pluckthun A: Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. *Nature* 1994, 368:261-265.

42. Gervasoni P, Gehrig P, Pluckthun A: Two conformational states of beta-lactamase bound to GroEL: a biophysical characterization. *J Mol Biol* 1998, 275:663-675.

43. Zahn R, Pluckthun A: Thermodynamic partitioning model for hydrophobic binding of polypeptides by GroEL. II. GroEL recognizes thermally unfolded mature beta-lactamase. *J Mol Biol* 1994, 242:165-174.

44. Walter S, Lorimer GH, Schmid FX: A thermodynamic coupling mechanism for GroEL-mediated unfolding. *Proc Natl Acad Sci USA* 1996, 93:9425-9430.

45. Clark AC, Frieden C: The chaperonin GroEL binds to late-folding non-native conformations present in native ***Escherichia coli*** and murine dihydrofolate reductases. *J Mol Biol* 1999, 285:1777-1788.

46. Bhutani N, Udgaonkar JB: A thermodynamic coupling mechanism can explain the GroEL-mediated acceleration of the folding of barstar. *J Mol Biol* 2000, 297:1037-1044.

47. Ranson NA, Dunster NJ, Burston SG, Clarke AR: Chaperonins can catalyse the reversal of early aggregation steps when a protein misfolds. *J Mol Biol* 1995, 250:581-586.

48. • Farr GW, Furtak K, Rowland MB, Ranson NA, Saibil HR, Kirchhausen T, Horwich AL: Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. *Cell* 2000, 100:561-573.
Functional and physical evidence that polypeptide substrates are bound by multiple consecutive apical domains of an open GroEL ring.

49. • Chen L, Sigler PB: The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell* 1999, 99:757-768.
Exogenously bound dodecamer peptide in a binary complex with the GroEL apical domains forms hydrophobic contacts and a few hydrogen bonds.

50. Fenton WA, Kashi Y, Furtak K, Horwich AL: Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* 1994, 371:614-619.

51. Buckle AM, Zahn R, Fersht AR: A structural model for GroEL-polypeptide recognition. *Proc Natl Acad Sci USA* 1997, 94:3571-3575.

52. Landry SJ, Gierasch LM: The chaperonin GroEL binds a polypeptide in an alpha-helical conformation. *Biochemistry* 1991, 30:7359-7362.

53. Kobayashi N, Freund SM, Chatellier J, Zahn R, Fersht AR: NMR analysis of the binding of a rhodanese peptide to a minichaperone in solution. *J Mol Biol* 1999, 292:181-190.

54. Wang Z, Feng H, Landry SJ, Maxwell J, Gierasch LM: Basis of substrate binding by the chaperonin GroEL. *Biochemistry* 1999, 38:12537-12546.

55. Preuss M, Hutchinson JP, Miller AD: Secondary structure forming propensity coupled with amphiphilicity is an optimal motif in a peptide or protein for association with chaperonin 60 (GroEL). *Biochemistry* 1999, 38:10272-10286.

56. Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU: Identification of ***in vivo*** substrates of the chaperonin GroEL. *Nature* 1999, 402:147-154.

57. Coyle JE, Texter FL, Ashcroft AE, Masselos D, Robinson CV, Radford SE: GroEL accelerates the refolding of hen lysozyme without changing its folding mechanism. *Nat Struct Biol* 1999, 6:683-690.

58. Weissman JS, Kashi Y, Fenton WA, Horwich AL: GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* 1994, 78:693-702.

59. Todd MJ, Viitanen PV, Lorimer GH: Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding. *Science* 1994, 265:659-666.

60. Beissinger M, Rutkat K, Buchner J: Catalysis, commitment and encapsulation during GroE-mediated folding. *J Mol Biol* 1999, 289:1075-1092.

61. Grallert H, Buchner J: Analysis of GroE-assisted folding under nonpermissive conditions. *J Biol Chem* 1999, 274:20171-20177.

62. Wittung-Stafshede P, Lee JC, Winkler JR, Gray HB: Cytochrome b562 folding triggered by electron transfer: approaching the speed limit for formation of a four-helix-bundle protein. *Proc Natl Acad Sci USA* 1999, 96:6587-6590.

63. Ghaemmaghami S, Word JM, Burton RE, Richardson JS, Oas TG: Folding kinetics of a fluorescent variant of monomeric lambda repressor. *Biochemistry* 1998, 37:9179-9185.

64. Spector S, Raleigh DP: Submillisecond folding of the peripheral subunit-binding domain. *J Mol Biol* 1999, 293:763-768.

65. Mines GA, Pascher T, Lee SC, Winkler JR, Gray HB: Cytochrome c folding triggered by electron transfer. *Chem Biol* 1996, 3:491-497.

66. Ferguson N, Capaldi AP, James R, Kleanthous C, Radford SE: Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* 1999, 286:1597-1608.

67. Kragelund BB, Osmark P, Neergaard TB, Schiodt J, Kristiansen K, Knudsen J, Poulsen FM: The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. *Nat Struct Biol* 1999, 6:594-601.

68. Choe SE, Matsudaira PT, Osterhout J, Wagner G, Shakhnovich EI: Folding kinetics of villin 14T, a protein domain with a central beta-sheet and two hydrophobic cores. *Biochemistry* 1998, 37:14508-14518.

69. Kuhlman B, Luisi DL, Evans PA, Raleigh DP: Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. *J Mol Biol* 1998, 284:1661-1670.

70. Khorasanizadeh S, Peters ID, Butt TR, Roder H: Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* 1993, 32:7054-7063.

71. Itzhaki LS, Otzen DE, Fersht AR: The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 1995, 254:260-288.

72. Silow M, Oliveberg M: High-energy channeling in protein folding. *Biochemistry* 1997, 36:7633-7637.

73. Villegas V, Martinez JC, Aviles FX, Serrano L: Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J Mol Biol* 1998, 283:1027-1036.

74. Smith CK, Bu Z, Anderson KS, Sturtevant JM, Engelman DM, Regan L: Surface point mutations that significantly alter the structure and stability of a protein's denatured state. *Protein Sci* 1996, 5:2009-2019.

75. Scalley ML, Yi Q, Gu H, McCormack A, Yates JR III, Baker D: Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry* 1997, 36:3373-3382.

76. Main ER, Fulton KF, Jackson SE: Folding pathway of FKBP12 and characterisation of the transition state. *J Mol Biol* 1999, 291:429-444.

77. van Nuland NA, Meijberg W, Warner J, Forge V, Scheek RM, Robillard GT, Dobson CM: Slow cooperative folding of a small globular protein HPr. *Biochemistry* 1998, 37:622-637.

78. Aronsson G, Brorsson AC, Sahlman L, Jonsson BH: Remarkably slow folding of a small protein. *FEBS Lett* 1997, 411:359-364.

79. van Nuland NA, Chiti F, Taddei N, Raugei G, Ramponi G, Dobson CM: Slow folding of muscle acylphosphatase in the absence of intermediates. *J Mol Biol* 1998, 283:883-891.

80. Clarke J, Cota E, Fowler SB, Hamill SJ: Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. *Structure* 1999, 7:1145-1153.

81. Viguera AR, Serrano L, Wilmanns M: Different folding transition states may result in the same native structure. *Nat Struct Biol* 1996, 3:874-880.

82. Hamill SJ, Steward A, Clarke J: The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J Mol Biol* 2000, 297:165-178.

83. Otzen DE, Fersht AR: Folding of circular and permuted chymotrypsin inhibitor 2: retention of the folding nucleus. *Biochemistry* 1998, 37:8139-8146.

84. Fulton KF, Main ER, Daggett V, Jackson SE: Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J Mol Biol* 1999, 291:445-461.

85. López-Hernández E, Serrano L: Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold Des* 1995, 1:43-55.

86. Schymkowitz JW, Rousseau F, Irvine LR, Itzhaki LS: The folding pathway of the cell-cycle regulatory protein p13suc1: clues for the mechanism of domain swapping. *Structure* 2000, 8:89-100.

87. Milla ME, Brown BM, Waldburger CD, Sauer RT: P22 arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry* 1995, 34:13914-13919.

88. Kraulis PJ: MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 1991, 24:946-950.

Now published

The work referred to in the text as (R Guerois, L Serrano, unpublished data) has now been published:

89. Guerois R, Serrano L: The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J Mol Biol* 2000, 304:987-982.