IUBMB *Life*, 54: 213–221, 2002
Copyright © 2002 IUBMB
1521-6543/02 $12.00 + .00
DOI: 10.1080/15216540290114496

Perspective

Phenomenological Perspectives on the Folding of β/α-Barrel Domains Through the Modular Formation and Assembly of Smaller Structural Elements

Sankar Maiti,[1] Manni Luthra-Guptasarma,[2] and Purnananda Guptasarma[1]
[1] *Institute of Microbial Technology, Sector 39 A, Chandigarh 160 036, India*
[2] *Department of Immunopathology, Postgraduate Institute of Medical Education and Research, Chandigarh 160 012, India*

Summary

The β/α-barrel motif was once considered to be a single protein domain. In recent years, however, it has been shown to consist of smaller substructures displaying the ability to fold autonomously. Here we review the current status of experimental findings concerning the motif's folding behavior in the light of what is currently known about (a) the relative rates of formation of helices and sheets in proteins, in general, and (b) the peculiarities of topology and architecture of the motif, in particular, to develop a detailed phenomenological understanding of how β/α-barrels might form through the modular folding and assembly of substructures.
IUBMB *Life*, 54: 213–221, 2002

Keywords Beta sheet formation; folding intermediates; helix-strand assembly; hierarchical assembly; kinetically regulated folding; multi-path folding; protein substructures.

INTRODUCTION

The topological motif known as the β/α-barrel has commonly tended to be regarded as a single structural domain (*1–4*). A domain is required, by definition, to be capable of folding autonomously through a two-state transition involving only the native (N) and unfolded (U) states of its polypeptide chain (*5–9*), since folding through the formation and assembly of observable substructures implies a level of autonomy of behavior that could entitle substructures within a so-called 'domain' to be themselves called domains. Intriguingly, whereas two β/α-barrels, namely triosephosphate isomerase (TIM) (*10*, *11*), and aldolase (*12*), do indeed 'appear' to fold through apparent two-state mechanisms, other β/α-barrels such as the alpha subunit of tryptophan synthase (*13–22*), N-(5′-phosphoribosyl)-anthranilate isomerase (*23*, *24*), and indole-3-glycerol phosphate synthase (*25*), clearly fold through non–two-state mechanisms involving populated intermediate states that have been attributed to the autonomous folding of N-terminal segments of their polypeptide chains. One β/α-barrel from the latter group, the alpha subunit of tryptophan synthase (α-TS), turns out, in fact, to be cleavable into proteolytic fragments that retain both native-like secondary structure, as well as substantial tertiary structure (*16*). Observations with α-TS have given rise to serious doubts about whether the β/α-barrel motif truly constitutes a single unit of folding (*26–28*). Only in recent times, however, have such doubts come to occupy biochemical centrestage, with the demonstration of the autonomous refolding of recombinant half-barrel-sized fragments of both α-TS (*19–22*), and an enzyme from the histidine biosynthesis pathway, HisF (*29*). Further support for folding of β/α-barrels through substructures comes from evidence of sequence homology, as well as of structural homology, between the two halves of the HisF β/α-barrel (*30*), indicative of a possible ancestral gene duplication event involving a half-barrel. Thus it is now clear that the β/α-barrel is not a single domain at all, but rather a structure that arises through the formation and assembly of smaller structures—which, for the moment at least appear to be minidomains that are the size of half barrels.

However, many questions remain, e.g., regarding how some β/α barrels such as TIM and aldolase appear to fold seamlessly without detectable intermediates, or about whether all β/α-barrels fold through essentially similar mechanisms [and similarly constituted folding nuclei] with differences applying only to their individual rates of folding. Notably, the latter question assumes special importance in light of the emerging view that similar folded chain topologies generally owe to similar mechanisms of folding (*30–43*). If indeed β/α-barrels fold

Received 6 June 2002; accepted 22 August 2002.
Address correspondence to Purnananda Guptasarma. E-mail: pg@imtech.res.in

through similar mechanisms, it remains possible that even representatives such as TIM and aldolase actually fold through non–two-state mechanisms, with intermediate states being populated to such a poor extent, and possessing such short lives, that folding appears to occur without intermediates. A key question, therefore, is thus whether all β/α-barrels fold through half-barrel intermediates, or whether half-barrels themselves are assemblages of even smaller autonomously formed substructures. In a seminal paper, the view has recently been expressed by Zitzewitz et al. that folding of β/α-barrels could possibly involve the modular assembly of even smaller structural elements (*20*), based on experimental data suggestive of autonomous folding by chain fragments smaller than the half-barrel.

The possibility that structure formation could indeed occur through modular folding and assembly of substructures, is expanded in the sections below into a review of available literature supporting arguments favouring modular assembly, with a view to synthesize a phenomenological understanding of how such folding might occur in significantly detailed molecular terms. Naturally, the discussions here delve occasionally into the realms of speculation; however, we would like to emphasize that the attempt is predominantly to review and synthesize the implications of available experimental data that have a bearing on the issues at hand, and that nowhere in the review that follows are speculations made without basis.

THE TOPOLOGY AND ARCHITECTURE OF THE β/α-BARREL: A REFRESHER

The β/α-barrel is created through the folding of a polypeptide chain into a laminated structure consisting of two coaxial, cylindrical layers of segregated secondary structural elements: (a) an inner layer of eight "buried" beta strands that are hydrogen bonded into a beta sheet barrel, and (b) an outer layer of eight "solvent-contacting" alpha helices, with each alpha helix preferentially contacting only a single beta strand from the inner layer. Intriguingly, this radial (architectural) segregation of helices and sheets is not achieved through physical segregation of helix- and sheet-forming sequences into different regions of the polypeptide, as could have possibly been the case. Rather, helix- and sheet-forming sequences are placed alternately along the entire length of the chain, with loops separating them in such a manner that the trajectory of the peptide backbone ends up passing alternately through the inner and outer layers of the barrel, winding eight times around the core of an imaginary solenoid (Figs. 1, 2). Each such winding, referred to as a "β/α" or "α/β" unit, consists of a single β-strand (so-called because it participates in the formation of a beta sheet), a long loop, an α-helix and a short loop. The long loops, also known as β-α loops, come together at the mouth of the barrel to host the active site of the protein; however, it is now commonly accepted that residues important for catalysis are contributed mainly by loops from the β/α units located at the *C*-terminus of the chain. Helices from all β/α units make intimate contacts with partner strands from within the same unit, and occasionally also some interactions with a strand or a helix from another unit. Together the helices form an outer rind that covers the inner beta barrel, with the backbone of every helix running opposite to that of its partner strand. The last element in each β/α unit, the short loop (often ranging from only 1 to 4 residues in length) serves to

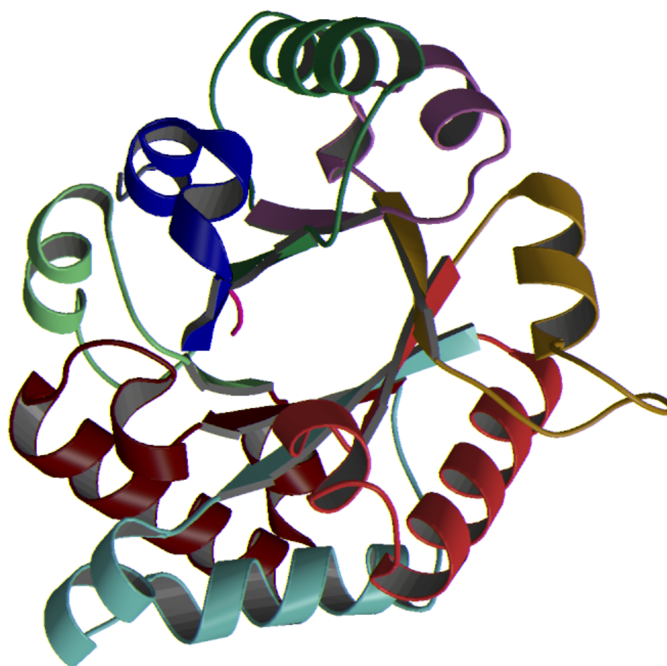

Figure 1. A ribbon diagram representation of the topology of the β/α-barrel drawn using the program MOLSCRIPT (*72*), showing the top view (or plan) of the structure as one looks down the barrel, clearly indicating the geometrical arrangements of the helices and strands (i.e., locations of all strands in the interior of the barrel, and all helices on the surface of the barrel).

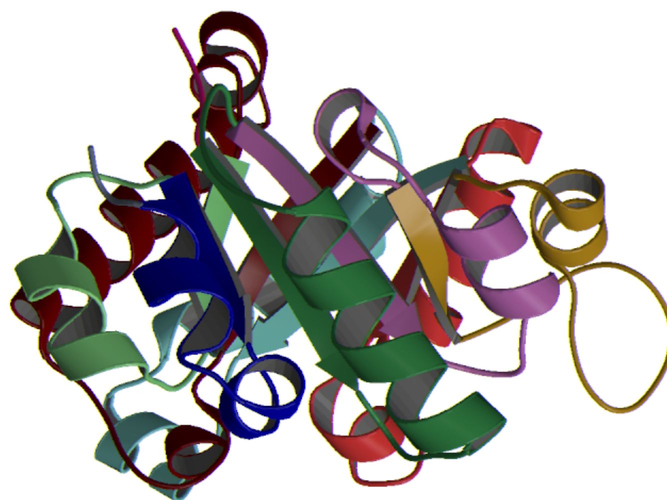

Figure 2. A ribbon diagram representation of the topology of the β/α-barrel, drawn using the program MOLSCRIPT (*72*), showing the side view (or elevation) of the structure.

separate the helix of one unit from the strand of the next unit. These tight loops can display local backbone hydrogen bonding interactions, reminiscent of β-turns or 3_{10} helices (*44*, *45*), although all short loops are not necessarily internally hydrogen-bonded.

Significantly, the strand in each β/α unit makes four classes of intimate contacts with helices and strands from both neighboring, as well as distal units: (1) backbone-backbone (hydrogen bond) contacts with strands from neighbouring units, to make beta sheets at the core of the barrel; (2) sidechain-sidechain contacts with partner helices, to facilitate the anchoring of helices onto themselves; (3) side chain-side chain contacts with strands from both neighbouring as well as distal units, to generate the nonpolar core of the beta barrel; and (4) in certain instances, side chain-side chain contacts with helices from neighbouring units, to strengthen interactions between units, and allow for multipoint anchoring of helices onto the beta barrel.

STRUCTURAL FEATURES THAT MAY INFLUENCE FOLDING MECHANISMS

Here, to prepare the ground for more detailed discussions to follow, we review the structural features of the β/α-barrel motif described specifically in relation to its average size, topology, and architecture, within the context of how each of these might be considered to support a greater likelihood of folding occurring through substructures, rather than through a molecule-wide, two-state transition.

Size. The β/α-barrel motif ranges in length from 250 to 350 residues (depending on the actual sizes of individual β-α loops in different proteins). It is probably the largest motif that has ever been assumed to constitute a single structural domain. If the entire motif were to have to form through a two-state transition, all elements of structure at every level of structural hierarchy would be required to form more or less simultaneously with the condensation of the unstructured chain around a single folding nucleus of key residues constituting a transition state for folding. Until such time, the entire chain would be required to remain in unstructured form, randomly exploring conformations through thermal motions. As is well known, the Levinthal paradox (*46*) points out how extremely time-consuming this can be for a polypeptide of even moderate length. According to the topomer sampling model (*47*, *48*), a recent model of folding (which attempts to find a way around the Levinthal paradox by proposing that random, domain-wide conformational search only applies to the early stages of chain collapse, with progressive restriction of conformational freedom occurring subsequently, on account of tentative interactions that occur in native-like chain topomers that have not yet undergone folding), the expected folding time for a chain of 100 residues turns out to be calculated to be of the order of 1 s. Extrapolation of the model's predictions to chain lengths of 250–350 residues (the range applicable to β/α-barrels) extends the range of expected folding times into several tens of minutes, and even hundreds of minutes. In contrast, rabbit TIM unfolds/folds on millisecond time scales (*11*). Prima facie it would appear, therefore, that randomness in conformational search must apply only to the very earliest moments of chain folding, if at all, with nonrandom conformational searches dominating at all later stages of folding. Below, we examine scenarios for nonrandom search from the viewpoints of the motif's topology and architecture.

Topology. As already discussed, if structure were to have to form through a two-state transition, the chain would be required to remain unstructured until the moment of adoption of a topomer capable of supporting the formation of a single folding nucleus. Such a topomer would need to support the mutual approach of all eight β-strands that are required to come together to form the inner beta barrel. While beta strands in the β/α-barrel are generally only 4–6 residues in length, they tend to be separated on an average by 21–30 residues (*49*), indicating that any mutual approach of strands limited by diffusional mechanisms would be likely to occur extremely slowly. The likelihood of all eight strands approaching each other simultaneously to participate in a two-state condensation to native structure would thus appear to be considerably lower that the likelihood of helices (specifically those not requiring stabilization through packing contacts with other structures) forming within the unfolded chain purely through local hydrogen bonding interactions. Such helices could conceivably associate with strand-designate sequences in the neighbouring regions of the chain; especially given the alternating occurrence of strand-forming and helix-forming sequences in the polypeptide, and the availability of long flexible loops separating strands from helices that are placed after them on the chain. Assuming limitations placed by requirements for diffusion-collision interactions, the probability of a strand colliding with a nearby nucleating-propagating helix with which it has side chain-side chain interactions in the native structure must clearly be much higher than the probability of a strand coming together with all seven other strands to support formation of a single folding nucleus and condensation of the entire structure around this nucleus. If strand-helix collisions could result in the stabilization of small structures, these could restrict the conformational freedom of the chain.

Architecture. If helical structures could be populated through nucleation-propagation mechanisms involving cooperatively occurring local hydrogen-bonding interactions, early forming helices resulting from such interactions could do one of two things. They could either interact with nearby regions of the chain through diffusion-collision mechanisms to form larger structures, using interactions similar to those applying to the native three-dimensional structure or, alternatively, wait for the inner beta barrel to form first. In this context, a rather peculiar feature of the β/α-barrel motif is that all secondary structural elements that can potentially form rapidly on account of being stabilized by predominantly local interactions (e.g., helices, short loops etc.) are actually located on the outside of the motif and in contact with the aqueous solvent. On the other hand,

all elements that are likely to form extremely slowly through diffusion-limited mechanisms because they are dependent on nonlocal interactions for their formation (e.g., beta strands forming sheets) turn out to be buried away within the motif's structure. The motif would thus appear to have a structure designed to frustrate attempts at rapid folding.

Thus, a key question is whether segments of the polypeptide remain unstructured before the chain is ready to adopt native structure through a single structural transition, or whether small structures form and undergo assembly. This question can be asked not merely of a whole barrel but also of a half-barrel. Next we summarize evidence from studies published in the past few years that indicate that helices that do form autonomously can form very much faster than the simplest beta sheets that occur in proteins (beta hairpins), and that peptides excised from helical regions of proteins too fractionally populate helical conformations even if they cannot fold autonomously, and quantitatively, into helices. Such arguments are used to reiterate a concept that has been around for a long time, namely that folding may begin with nucleation-propagation of helices; further we extend this possibility to propose that such nascent helices undergo induced-fit condensation (with strands making native sidechain contacts with them in the native structure) to create strand-helix assemblies.

Perspectives on (and a Possible Scheme for) Modular Folding

Rates of Formation Differ Among Secondary Structural Elements that Have the Potential to Form Autonomously. Helices in proteins are formed predominantly through local interactions between hydrogen-bonding partners (groups of atoms in the backbone) separated by no more than a few residues along a polypeptide chain. In contrast, beta sheets are formed through predominantly nonlocal interactions, with "along-chain" distances between hydrogen-bonding partners varying from a few tens of residues, for the farthest ends of a simple beta hairpin, to a few hundreds of residues for sheets that are formed entirely nonlocally. If diffusion were to limit the rate of formation of all secondary structural elements, one could argue that the average beta sheet ought to take much longer than the average alpha helix to form. Indeed, it is now known that helices are capable of forming over timescales as short as a few nanoseconds (*50*), or a few tens of nanoseconds (*51*), while even the simplest of beta sheet configurations (such as the beta hairpin) is found to take several microseconds to form (*52*). Furthermore, among proteins which fold rapidly (and quantitatively) to native state, those containing only alpha helical structures are found to fold on microsecond, or submicrosecond timescales, while those containing only beta structures take several milliseconds, or more, to fold (*48*). Thus, timescales of formation of helices and sheets appear to differ by two orders of magnitude, with the difference possibly increasing further in a manner dependent on the extent to which strands that are required to come together tend to be separated from each other in the primary structure.

It may be considered likely, therefore, that helices that are capable of forming autonomously (i.e., without a requirement for packing interactions with other structures) would actually form early-on during refolding. With regard to helices requiring packing interactions with other structures also, it has been shown, through NMR spectroscopic investigations of peptides derived from helices that do not form stably on their own in solution, that here too there is a tendency to populate helical conformations (*53*, *54*). Could early-forming helices, or even fractionally populated helices, hasten folding by interacting with neighboring designate β-strands that have not yet met their native hydrogen-bonding partners?

Diffusion-Collision Interactions of Nucleating-Propagating Helices with Neighbouring Elements. It is possible that fully formed helices capable of forming autonomously (i.e., without requiring additional stabilization through packing contacts with other structures), as also stretches of sequence that significantly populate helical structures without quantitatively adopting such structures, would have local diffusion-collision interactions with proximal regions of the chain.

Condensation of Strands Onto Helices to Form Strand-Helix Assemblies. In course of such collisions, the neighbouring 'strand-designate' sequences in the chain that normally make intimate non-covalent interactions with partner helices from within the same β/α unit in the native structure could potentially condense onto nascent helices through induced-fit interactions. Through such an interaction, a strand-designate sequence could stabilize a nucleating-propagating helix, and vice versa, to form a strand-helix assembly; however, without the characteristic beta sheet structure having yet been adopted by the strand. Because the component strands of such assemblies have not yet hydrogen bonded with strands from neighbouring assemblies to form beta sheets, we shall refer to them only as strand-helix assemblies to differentiate them from β/α units (in which strands have already qualified to be called β strands, on account of their having formed beta sheets).

Closure of Tight (α-β) Loops to Form β/α Unit Doublets Through Association of Strand-Helix Assemblies. The alpha helix of any β/α unit in the native structure is connected to the beta strand of the next β/α unit along the sequence through a short loop (the α-β loop already described) at the base of the barrel, generally between 1 and 4 residues long (*44*, *45*). In many cases, such short loops are conformationally constrained in the native structure through internal hydrogen bonding in the manner of a tight beta turn, or the turn of a 3_{10} helix. The closure of such loops at the C-termini of helices could be expected to force the coming together of adjacent strand-helix assemblies separated by tight loops that are internally hydrogen bonded in the native structure.

Ensuring the Proper Hydrogen-Bonding 'Register' for Beta Sheet Formation. When two strands approach each other to form a beta sheet, there are many registers in which hydrogen bonding can conceivably occur, since a C=O group from any residue in one strand can potentially hydrogen bond with

an N—H group from any other residue in the partner strand participating in beta-sheet formation. Even so, only one of these registers is actually adopted in a native structure. A possible perspective on a rapid identification of the appropriate register for hydrogen-bonding (that is not greatly rate-limited by nonlocal, diffusion-collision interactions among unstructured strands searching for a stable, equilibrium arrangement) is the occurrence of a native-like association of neighbouring β/α units. Such association could be facilitated by the closure of the tight loops zipping-up through a hydrogen bonding scheme, which would effectively ensure the occurrence of native-like side chain-side chain interactions between the associating strand-helix assemblies at their base. (The sheets and the helices in any β/α-barrel meet at a slight angle, ensuring that several, if not all, helices interact with at least two strands, and several strands with two helices. Further, in certain instances, helices from neighbouring strand-helix assemblies also make contact with each other in the native structure. Chances exist, therefore, that when two such assemblies are forced to come together through the closure of the tight α-β loop structures separating them in the sequence, the scope is effectively created for the occurrence of a number of native-like side chain-side chain interactions between two assemblies.) Such interactions could cause strands in adjacent assemblies to approach each other with no freedom in respect of the register in which hydrogen bonding would occur, thus leaving open to mutually approaching strand-helix assemblies only the option for formation of a beta sheet in the native register. A mechanism of this sort would reduce the time taken by every designate beta strand to actually form hydrogen bonds with its partner strands.

Assembly Into β/α Unit Quadruplets, Sextuplets, and Octuplets. Through local diffusion-collision interactions between adjacent β/α unit doublets in the chain, larger structures containing four, six, or eight β/α units could be gradually built up. At all stages, such structures would be curved, rather than planar, because of the natural handedness of twisting of beta strands and the need to accommodate the volumes of helices adsorbed onto strands. In proteins in which folding is required to proceed obligatorily through the formation and assembly of two β/α quadruplet (half-barrel) structures, curving could facilitate the rapid formation of nonpolar contacts between beta strands to form the core of the barrel. The proposed folding scheme is shown alongside (Fig. 3).

Only Near-Neighbour Interactions At All Levels. Most importantly, the formation of a β/α-barrel in the manner outlined here would be likely to occur at rates determined almost entirely by diffusion-collision interactions between adjacent structures in the polypeptide chain, through all the proposed stages of folding. Initially such interactions would occur between C═O and N—H groups separated by only a few residues, to generate helices. Following this, the helix would interact with one of its neighboring extended sequences (which it would be capable of stably binding to, through induced-fit sidechain contacts) to form strand-helix assemblies. Meanwhile, local hydrogen-bonding interactions at

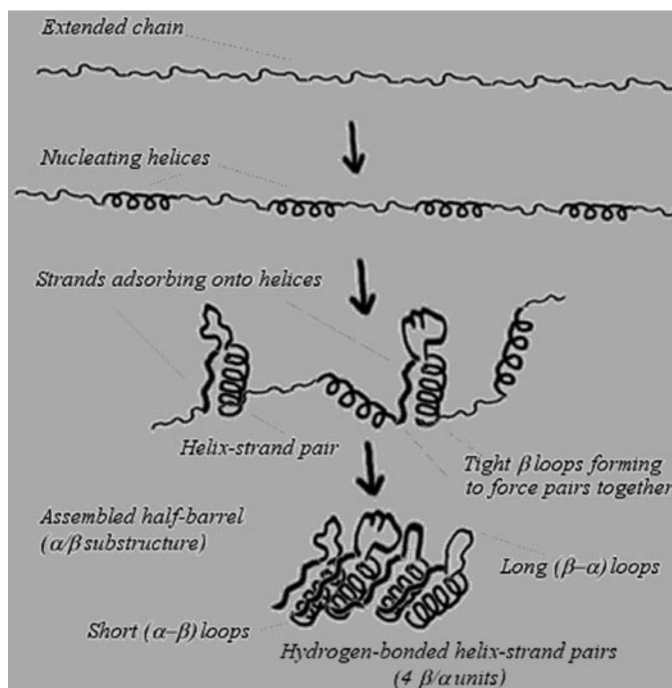

Figure 3. A schematic diagram illustrating the refolding of one half of the β/α-barrel motif. Note that whereas the figure shows helix-strand assemblies (or pairs) forming and assembling in a specific manner, it is meant to convey only that folding proceeds down multiple pathways involving the marked steps, with the precise details illustrating one of many possible pathways of formation and assembly in a refolding population. Note that during folding on the ribosome, the assembly could progress from the N- to the C-terminus. The figure has not been drawn to take into account either the relative dimensions of structural elements or the stereochemistry of their interactions.

the C-termini of fully extended helices would create tight α-β loops which, by virtue of their very formation, would bring together neighbouring strand-helix assemblies to create β/α doublets. Finally, β/α doublets would interact with neighbouring doublets to progressively generate quadruplets, sextuplets, and octuplets, concomitantly with the folding and assembly of the long β-α loops and nonpolar interactions at the core of the barrel. At no step in the scheme would nonlocal interactions (as between the eight beta strands of the barrel) ever be required to occur through random diffusional interactions in an unstructured chain undergoing thermal motions. Thus nonlocal interactions could never be rate-limiting for structure formation.

Rapid Multipathway Assembly Rather Than Cooperative Apparent Two-State Folding. The scheme could conceivably facilitate rapid folding by initiating at different helix-forming sequences in different molecules, and proceeding in a manner that would not really require any particular state to be populated for long lengths of time by any significantly large numbers of molecules, because various molecules could take various routes.

Folding could thus proceed in a nonrandom fashion, and yet appear to occur through a two-state reaction without detectable populated intermediates. Exceptions would occur only when the assembly of some β/α sub-structures would be prevented from occurring rapidly, e.g., on account of the existence of a kinetic barrier located close to the point in folding where various folding pathways converge. In other words, intermediates would only be seen in the folding of β/α-barrels characterized by high energy transition states of substructure assembly. In a β/α-barrel xylanase secreted by the *Bacillus* sp. NG-27, we see folding through an intermediate that has evidently folded around two nuclei and undergoes extremely slow assembly to native structure (manuscript submitted).

Commonality of Folding Mode, but not of Exact Folding Mechanism or Rate. A recent paper cites evidence for gene duplication within an α/β-barrel, suggesting that assembly of α/β units progresses to produce two substructures that constitute half-barrels (*29*). In another protein, such as α-TS, whereas a β/α quadruplet comprising the first four β/α units is clearly capable of folding autonomously (*20*), folding itself appears to proceed through the autonomous folding of a fragment that constitutes almost a β/α sextuplet (*13–23*). Similarly, unfolding appears to occur through the autonomous unfolding of two chain segments in two other enzymes (*24*, *25*) where the N-terminal segment approximates a β/α sextuplet (the first six strands and first five helices). Further complications could arise from cis-trans isomerization of proline residues in different β/α-barrels. It is conceivable, therefore, that the precise details of preferred pathways of folding differ sufficiently among the β/α-barrels to cause different proteins to preferentially populate different long-lived intermediates, while at the same time allowing some proteins like TIM and aldolase to form and assemble strand-helix assemblies and β/α unit doublets seamlessly, without detectable intermediates. Assuming that commonalities are restricted to commonalities in respect of only the mode of folding (i.e., multipathway folding, through modular assembly of supersecondary structural elements), it is possible that there is really no commonality of the actual folding mechanism among β/α-barrels despite the similarity of chain topology.

Lessons from Interrupted β/α-Barrels. It may be relevant to examine the few unusual β/α-barrel structures known in which large segments of polypeptide (much larger than the largest β-α loops) interrupt the β/α-barrel. In some such proteins, the interruption amounts to an entire protein 'domain,' occuring within the loop in the third β/α unit, whereas in others it occurs within the third, fourth, or other units (*55–59*). This suggests that the two parts of the β/α-barrel that are separated by an interrupting domain probably fold autonomously and remain folded but not assembled, to await the folding of the interrupting domain, so as to come together and assemble physically. Such interruptions suggest that there may be nothing sacrosanct about the need to fold through a half-barrel mechanism, and that what is actually important is the need to fold through the modular assembly of supersecondary structural elements as previously suggested (*20*), of course with differences in perspectives and detail applying to folding schemes as hypothesized here.

Possible Cotranslational Folding of β/α-Barrels. Assuming that folding occurs through the formation and modular assembly of supersecondary structural elements, it is not difficult to conceive of how modular assembly could occur extremely efficiently in course of folding on the ribosome. Following the formation of the very first strand-helix assembly in the chain, the newly synthesized strand of every successive β/α unit could be pulled towards the last strand of the preexisting structure through the closure of tight α-β loops as already suggested, such that the helix synthesized thereafter would have the opportunity of condensing not just against a single unstructured strand through induced-fit interactions, but rather against an entire superstructure of assembled β/α units ending with one unsatisfied strand available to make the right noncovalent contacts with the helix. Following the condensation of each helix onto the preexisting superstructure, the next strand could similarly be pulled onto the structure and hydrogen bonded through loop closure, and so on. Through such a mode of assembly the structure would finish forming virtually concomitantly with synthesis. If some aspects of such a folding mechanism were retained in course of refolding, one would predict that C-terminal segments of the barrel would need to be less stable on their own as compared to N-terminal segments, since they could have N-terminal segments already folded and available to pack against at the time of synthesis. Although such differences in stability could probably extend to the level of strand-helix assemblies, it is likely that larger structures also would show differences in stability. Indeed, as work with α-TS and with the HisF half barrel systems has shown, the genetically, or proteolytically excised N-terminal half-barrel is more stable than the C-terminal half-barrel (*19–22*, *30*).

The Significance of Entropic Contributions in Facilitating Early Enthalpic Contacts. The continuing debate about whether folding occurs through two-state (N $\Leftrightarrow$ U) transitions, or through the modular formation and assembly of supersecondary structural elements, focuses the spotlight on another very important issue—namely that of whether conformational entropy plays any role whatsoever in protein folding. Anfinsen's original hypothesis (*60*) proposing that folding takes a polypeptide chain to a state constituting the global minimum of Gibbs free energy has typically been interpreted to mean that the native conformation must correspond to a global minimum in internal energy (since a fully folded protein will have lost almost all of its conformational entropy). As pointed out by Srinivasan and Rose (*61*), this view is often extended to the erroneous conclusion that entropy plays no significant role in the thermodynamics of folding.

Today, the possible importance of entropic contributions (in particular, owing to steric repulsions) in the determination, in particular, of whether a segment of polypeptide chain will adopt the α, or β, type of secondary structure can no longer be ignored (*61*). The wealth of evidence showing that formation of

secondary structure can precede formation of tertiary structure (*62*), and recent evidence demonstrating that native-like topology can persist in nonnative and denatured states (*63–70*), emphasize the importance of explaining how secondary structures form, i.e., whether they form concomitantly with tertiary structure, or whether they form first and then preorganize the scheme of formation of tertiary structure—through modular assembly based on early enthalpic contacts that are facilitated by entropic contributions due to the steric repulsions of adjacent side chain groups, as well as adjacent precociously formed substructures in the chain (*61*, *68*).

It has been argued (*61*) that the only reasonable theory that can explain the existence of secondary structural elements in naturally occurring proteins involves: (a) early entropic contributions determining intrinsic chain biases for the adoption of α-, or β-, chain configurations, and further (b) such biases, in turn, determining the nature and probability of occurrence of enthalpic contacts between chain segments.

Because such early enthalpic contacts could conceivably compensate suitably for the loss of entropy caused by local steric repulsions between side chains and between side chains and backbone atoms, and further because the substructures (or secondary structural elements) happening to be stabilized by such enthalpic contacts could, in turn, conceivably facilitate further entropy-driven occurrences of new enthalpic contacts through further steric repulsion-driven entropic contributions (now owing to local collisions of secondary structural elements), such hand-in-hand control of folding by enthalpic and entropic contributions could potentially engender the modular formation and assembly of ever larger (supersecondary) structural elements, and finally the formation of the fully folded protein.

Thus, folding driven by the combined (and cooperating) effects of entropic and enthalpic contributions and occurring through the modular formation and assembly of secondary and supersecondary structural elements, is easily reconciled with the only satisfactory explanation that has yet been proposed for the occurrence of secondary structural elements in proteins (*61*, *68*)—an enigma that has never yet been either understood, or explained, purely on the basis of considerations of enthalpy-driven two-state models of folding alone (*71*).

ACKNOWLEDGEMENTS

We thank G. Sahni, R. Kishan, A. Ghosh, B. Kundu, A. Shukla, and S. Mukherji of the Institute of Microbial Technology for discussions, and the referee for the suggestion that a discussion on enthalpic and entropic contributions be included. SM is a doctoral research fellow of the Council of Scientific & Industrial Research in the laboratory of Amit Ghosh, who collaborated with PG on experimental studies of the folding of a β/α-barrel xylanase (to be published).

REFERENCES

1. Levitt, M., and Chotia, C. (1976) Structural patterns in globular proteins. *Nature* **261**, 552–558.
2. Richardson, J. S. (1981) The anatomy and taxonomy of protein structure. *Adv. Prot. Chem.* **34**, 67–339.
3. Farber, G. K., and Petsko, G. A. (1990) The evolution of α/β barrel enzymes. *Trends Biochem. Sci.* **15**, 228–234.
4. Branden, C., and Tooze, J. (1991) Alpha/beta structures. In *Introduction to Protein Structure.* pp. 43–57, Garland Publishing Inc., New York.
5. Wetlaufer, D. B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 697–701.
6. Branden, C., and Tooze, J. (1991) Motifs of protein structure. In *Introduction to Protein Structure*. pp. 11–31, Garland Publishing Inc., New York.
7. Privalov, P. L. (1979) Stability of proteins: small globular proteins. *Adv. Prot. Chem.* **33**, 167–241.
8. Chan, H. S., Bromberg, S., and Dill, K. A. (1996) Models of cooperativity in protein folding. In *Protein Folding* (Dobson, C. M., and Fersht, A. R. eds.). pp. 61–70, Cambridge University Press, Cambridge, England.
9. Jackson, S. E. (1998) How do small single domain proteins fold? *Fold. Des.* **3**, R81–91.
10. Waley, S. G. (1973) Refolding of triosephosphate isomerase. *Biochem. J.* **135**, 165–172.
11. Rietveld, A. W. M., and Ferreira, S. T. (1998) Kinetics and energetics of subunit dissociation/unfolding of TIM: the importance of oligomerization for conformational persistence and chemical stability of proteins. *Biochemistry* **37**, 933–937.
12. Rudolph, R., Siebendritt, R., and Kiefhaber, T. (1992) Reversible unfolding and refolding behavior of a monomeric aldolase from *Staphylococcus aureus*. *Protein Sci.* **1**, 654–666.
13. Yutani, K., Ogasahara, K., Suzuki, M., and Sugino, Y. (1979) Comparison of denaturation by guanidine hydrochloride of the wild type tryptophan synthase alpha-subunit of *Escherichia coli* and two mutant proteins (Glu 49 replaced by Met or Gln). *J. Biochem.* **9**, 915–921.
14. Matthews, C. R., and Crisanti, M. M. (1981) Urea-induced unfolding of the alpha subunit of tryptophan synthase: evidence for a multistate process. *Biochemistry* **20**, 784–792.
15. Matthews, C. R., and Crisanti, M. M. (1981) Characterization of the slow steps in the folding of the alpha subunit of tryptophan synthase. *Biochemistry* **20**, 2700–2706.
16. Miles, E. W., Yutani, K., and Ogasahara, K. (1982) Guanidine hydrochloride induced unfolding of the alpha subunit of tryptophan synthase and of the two alpha proteolytic fragments: evidence for stepwise unfolding of the two alpha domains. *Biochemistry* **21**, 2586–2592.
17. Beasty, A. M., and Matthews, C. R. (1985) Characterization of an early intermediate in the folding of the alpha subunit of tryptophan synthase by hydrogen exchange measurement. *Biochemistry* **24**, 3547–3553.
18. Chen, X., Rambo, R., and Matthews, C. R. (1992) Amino acid replacement can selectively affect the interaction energy of autonomous folding units in the alpha subunit of tryptophan synthase. *Biochemistry* **31**, 2219–2223.
19. Bilsel, O., Zitzewitz, J. A., Bowers, K. E., and Matthews, C. R. (1999) Folding mechanism of the alpha subunit of tryptophan synthase, an alpha/beta barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry* **38**, 1018–1029.
20. Zitzewitz, J. A., Gualfetti, P. J., Perkons, I. A., Wasta, S. A., and Matthews, C. R. (1999) Identifying the structural boundaries of independent folding domains in the alpha subunit of tryptophan synthase, a beta/alpha barrel protein. *Protein Sci.* **8**, 1200–1209.
21. Zitzewitz, J. A., and Matthews, C. R. (1999) Molecular dissection of the folding mechanism of the alpha subunit of tryptophan synthase: an amino-terminal autonomous folding unit controls several rate-limiting steps in the folding of a single domain protein. *Biochemistry* **38**, 10205–10214.
22. Gualfetti, P. J., Bilsel, O., and Matthews, C. R. (1999) The progressive development of structure and stability during the equilibrium folding of the alpha subunit of tryptophan synthase from *Escherichia coli. Protein Sci.* **8**, 1623–1635.

23. Eder, J., and Kirschner, K. (1992) Stable sub-structures of eightfold beta alpha-barrel proteins: fragment complementation of phosphoribosylanthranilate isomerase. *Biochemistry* **31**, 3617–3625.
24. Jasanoff, A., Davis, B., and Fersht, A. R. (1994) Detection of an intermediate in the folding of the (beta alpha)8-barrel N-(5′-phosphoribosyl)anthranilate isomerase from *Escherichia coli. Biochemistry* **33**, 6350–6355.
25. Sanchez del Pino, M. M., and Fersht, A. R. (1997) Nonsequential unfolding of the alpha/beta barrel protein indole-3-glycerol-phosphate synthase. *Biochemistry* **36**, 5560–5565.
26. Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., and David, D. R. (1988) Three-dimensional structure of the tryptophan synthase $\alpha_2\beta_2$ multienzyme complex from *Salmonella typhimurium. J. Biol. Chem.* **263**, 17857–17871.
27. Creighton, T. E. (1990) Protein folding. *Biochem. J.* **270**, 1–16.
28. Goldenberg, D. P. (1993) Mutational analysis of protein folding and stability. In *Protein Folding* (Creighton, T. E., ed.). pp. 353–403, W. H. Freeman, New York.
29. Lang, D., Thoma, R., Henn-sax, M., Sterner, R., and Wilmanns, M. (2000) Structural evidence for evolution of the beta-alpha barrel scaffold through gene duplication and fusion. *Science* **289**, 1546–1550.
30. Hocker, B., Beismann-Dreimeyer, S., Hettwer, S., Lustig, A., and Sterner, R. (2001) Dissection of a (beta/alpha)8 barrel enzyme into two folded halves. *Nat. Struct. Biol.* **8**, 32–36.
31. Shakhnovich, E. I., Abkevich, V. I., and Ptitsyn, O. (1996) Conserved residues and the mechanism of protein folding. *Nature* **379**, 96–98.
32. Michnick, S. W., and Shakhnovich, E. I. (1998) A strategy for detecting the conservation of folding-nucleus residues in protein superfamilies. *Fold. Des.* **3**, 239–251.
33. Mirny, L. A., Abkevich, V. I., and Shakhnovich, E. I. (1998) How evolution makes proteins fold quickly. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4976–4981.
34. Clarke, J., Cota, E., Fowler, S. B., and Hamill, S. J. (1999) Folding studies of immunoglobulin-like β-sandwich proteins suggest that they share a common folding pathway. *Structure* **7**, 1145–1153.
35. Alm, E., and Baker, D. (1999) Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11305–11310.
36. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M., and Dobson, C. M. (1999) Topology and contact order in protein folding. *Nat. Struct. Biol.* **6**, 1005–1009.
37. Martinez, J. C., and Serrano, L. (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.* **6**, 1010–1016.
38. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., and Baker, D. (1999) Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Biol.* **6**, 1016–1024.
39. Fersht, A. R. (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1525–1529.
40. Hamill, S. J., Steward, A., and Clarke, J. (2000) The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* **297**, 165–178.
41. Clementi, C., Nyemeyer, H., and Onuchic, J. N. (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and 'en-route' intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953.
42. Baker, D. (2000) A surprising simplicity to protein folding. *Nature* **405**, 39–42.
43. Clementi, C., Jennings, P. A., and Onuchic, J. N. (2000) How native-state topology affects the folding of dihydrofolate reductase and interleukin-1 beta. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5871–5876.
44. Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., and Waley, S. G. (1975) Structure of chicken muscle triosphosphate isomerase at 2. 5 Å resolution using amino acid sequence data. *Nature* **255**, 609–614.
45. Scheerlinck, J. P., Lasters, I., Claessens, M., De Maeyer, M., Pio, F., Delhaise, P., and Wodak, S. J. (1992) Recurrent alpha beta loop structures in TIM-barrel motifs show a distinct pattern of conserved structural features. *Proteins* **12**, 299–313.
46. Levinthal, C. (1968) Are there pathways for protein folding? *J. Chem. Phys.* **65**, 44–45.
47. Debe, D. A., Carlson, M. J., and Goddard, W. A. (1999) The topomer sampling model of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2596–2601.
48. Debe, D. A., and Goddard, W. A. (1999) First principles prediction of protein folding rates. *J. Mol. Biol.* **294**, 619–625.
49. Selvaraj, S., and Gromiha, M. M. (1998) Importance of long-range interactions in the $(\alpha/\beta)_8$ barrel fold. *J. Prot. Chem.* **17**, 691–697.
50. Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H., and Dyer, R. B. (1996) Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry* **35**, 691–697.
51. Eaton, W. A., Hofrichter, J., Monoz, V., and Thompson, P. A. (1998) Kinetics and dynamics of loops, alpha helices, beta hairpins and fast-folding proteins. *Acc. Chem. Res.* **31**, 745–753.
52. Munoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) Folding dynamics and mechanism of β-hairpin formation. *Nature* **390**, 196–199.
53. Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., and Wright, P. E. (1992) Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. I. Myohemerythrin. *J. Mol. Biol.* **226**, 795–817.
54. Luisi, D. L., Wu, W. J., and Raleigh, D. P. (1999) Conformational analysis of a set of peptides corresponding to the entire primary sequence of the N-terminal domain of the ribosomal protein L9: evidence for stable native-like secondary structure in the unfolded state. *J. Mol. Biol.* **287**, 395–407.
55. Stuart, D. I., Levine, M., Muirhead, H., and Stammers, D. K. (1979) Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. *J. Mol. Biol.* **134**, 109–142.
56. Hofmann, B. E., Bender, H., and Schulz, G. E. (1989) Three dimensional structure of cyclodextrin glycosyltransferase from *Bacillus circulans* at 3.4 Å resolution. *J. Mol. Biol.* **209**, 793–800.
57. Brayer, G. D., Luo, Y., and Withers, S. G. (1994) The structure of human pancreatic alpha-amylase at 1.8 Å resolution, and comparisons with related enzymes. *Protein Sci.* **4**, 1730–1742.
58. Essen, L. O., Perisik, O., Cheung, R., Katan, M., and Williams, R. L. (1996) Crystal structure of a mammalian phosphoinositide-specific phospholipase C-delta. *Nature* **380**, 596–602.
59. Shima, S., Warkentin, E., Grabarse, W., Sordel, M., Wicke, M., Thauer, R. K., and Ermler, U. (2000) Structure of co-enzyme F-420 dependent methylene tetrahydromethanopterene reductase from two methanogenic archae. *J. Mol. Biol.* **300**, 935–950.
60. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* **181**, 223–230.
61. Srinivasan, R., and Rose, G. D. (1999) A physical basis for protein secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14258–14263.
62. Arai, M., and Kuwajima, K. (2000) Role of the molten globule state in protein folding. *Adv. Protein Chem.* **53**, 209–282.
63. Wilson, G., Hecht, L., and Barron, L. D. (1996) Residual structure in unfolded proteins revealed by Raman optical activity. *Biochemistry* **35**, 12518–12525.
64. Bond, C. J., Wong, K. B., Clarke, J., Fersht, A. R., and Daggett, V. (1997) Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13409–13413.
65. Shortle, D., Simons, K. T., and Baker, D. (1998) Clustering of low-energy conformations near the native structures of small proteins. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11158–11162.
66. Neira, J. L., Sevilla, P., Menendez, M., Bruix, M., and Rico, M. (1999). Hydrogen exchange in ribonuclease A and ribonuclease S: evidence for residual structure in the unfolded state under native conditions. *J. Mol. Biol.* **285**, 627–643.

67. Denisov, V. P., Jonsson, B.-H., and Halle, B. (1999) Hydration of denatured and molten globule proteins. *Nat. Struct. Biol.* **6**, 253–260.
68. Shortle, D. (1999) Protein folding as seen from water's perspective. *Nat. Struct. Biol.* **6**, 203–205.
69. Shortle, D., and Ackerman, M. S. (2001) Persistence of native-like topology in a denatured protein in 8M urea. *Science* **293**, 487–489.
70. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein. *Science* **295**, 1719–1722.
71. Baldwin, R. L., and Rose, G. D. (1999) Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem. Sci.* **24**, 26–33.
72. Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.