Short-Lived α-Helical Intermediates in the Folding of β-Sheet Proteins

ARTICLE *in* BIOCHEMISTRY · JULY 2010

Impact Factor: 3.02 · DOI: 10.1021/bi100288q · Source: PubMed

7 AUTHORS, INCLUDING:

Eefei Chen
University of California, Santa Cruz
35 PUBLICATIONS **676** CITATIONS
SEE PROFILE

Dawn E Bowles
Duke University Medical Center
38 PUBLICATIONS **1,047** CITATIONS
SEE PROFILE

William Parker
Duke University Medical Center
141 PUBLICATIONS **3,328** CITATIONS
SEE PROFILE

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 5609–5619 5609
DOI: 10.1021/bi100288q

Short-Lived α-Helical Intermediates in the Folding of β-Sheet Proteins

Eefei Chen,[‡] Mary Lou Everett,[§] Zoie E. Holzknecht,[§] Robert A. Holzknecht,[§] Shu S. Lin,[§,∥] Dawn E. Bowles,[§] and William Parker*[,§]

[‡]*Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95060,* [§]*Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, and* [∥]*Department of Immunology, Duke University, Durham, North Carolina 27710*

Received February 26, 2010; Revised Manuscript Received May 26, 2010

ABSTRACT: Several lines of evidence point strongly toward the importance of highly α-helical intermediates in the folding of all globular proteins, regardless of their native structure. However, experimental refolding studies demonstrate no observable α-helical intermediate during refolding of some β-sheet proteins and have dampened enthusiasm for this model of protein folding. In this study, β-sheet proteins were hypothesized to have potential to form amphiphilic helices at a period of <3.6 residues/turn that matches or exceeds the potential at 3.6 residues/turn. Hypothetically, such potential is the basis for an effective and unidirectional mechanism by which highly α-helical intermediates might be rapidly disassembled during folding and potentially accounts for the difficulty in detecting highly α-helical intermediates during the folding of some proteins. The presence of this potential was confirmed, indicating that a model entailing ubiquitous formation of α-helical intermediates during the folding of globular proteins predicts previously unrecognized features of primary structure. Further, the folding of fatty acid binding protein, a predominantly β-sheet protein that exhibits no apparent highly α-helical intermediate during folding, was dramatically accelerated by 2,2,2-trifluoroethanol, a solvent that stabilizes α-helical structure. This observation suggests that formation of an α-helix can be a rate-limiting step during folding of a predominantly β-sheet protein and further supports the role of highly α-helical intermediates in the folding of all globular proteins.

The folding of all globular proteins via highly helical intermediate structures was hypothesized by V. Lim in the late 1970s (*1*, *2*). At that time, several reasons why such a hypothesis merited attention were enumerated. First, all globular proteins, even those whose native structure is characterized by predominantly β-sheet and the total absence of α-helix, have substantial potential to form α-helices with both a hydrophobic and a hydrophilic surface, termed amphiphilic α-helices. Second, since α-helices are rigid structures, the amphiphilic α-helix is well-suited to participate in specific interhelical interactions. Third, an α-helix with a typical length of approximately 10−15 residues possesses an overall shape that is approximately spherical, thus comprising a small, densely packed globule that will diffuse in solution much faster than any other conformation of the polypeptide chain.

A final argument made by Lim in support of highly helical intermediates in the folding of all proteins was that the formation of α-helix, being dependent on local interactions, was much faster than the formation of structure such as β-sheet, which depends on interactions between sections of the polypeptide chain that are distant from each other. This argument has found support from a number of studies evaluating the formation of α-helix in peptides. For example, Wright and colleagues (*3*) as well as Baldwin and colleagues (*4*−*6*) have demonstrated that helix formation can occur in short peptides with naturally occurring sequences as a result of stabilization by $i + 3$ and particularly $i + 4$ side chain pairs involved in charge−charge interactions, hydrogen bonding, or possibly van der Waals interactions. Further, work with a variety of alanine-rich model peptides (*7*−*13*) has shown that helices can form rapidly, in the range of a few hundred nanoseconds. Lednev and colleagues (*11*), for example, determined a time constant of ∼800 ns for the folding of a 21-residue alanine-based peptide into an α-helix. These data support the argument that rapid formation of α-helices during the early stages of protein folding can be achieved on the basis of local interactions.

Subsequent work has provided additional support for the idea that highly helical intermediates are involved in the folding of all proteins, regardless of their native structure. First, a quantitative evaluation of a variety of predominantly β-sheet proteins showed that the potential to form amphiphilic α-helix in those proteins was not found in randomly generated amino acid sequences, indicating that the potential to form amphiphilic α-helix was due to specific features in the naturally occurring sequences rather than happenstance (*14*). Second, amino acid segments that have amphiphilic α-helical potential were observed to form ring-shaped clusters, 20−30 Å in diameter, within the native structure of predominantly β-sheet proteins (*14*). The segments with amphiphilic α-helical potential that form these clusters are spaced at average intervals of 13 Å between segments, approximately the same as the spacing of helices in a helical bundle (*15*). Further, formation of such clusters does not occur when the location of segments with amphiphilic α-helical potential was randomly assigned in the amino acid sequence (*15*). These "dormant α-helical domains" within the structure of β-sheet proteins may be the footprints of helical bundles that once existed but are no longer present in the native structures (*14*). A third factor suggesting that highly α-helical intermediates are important for the folding of β-sheet proteins is the observation that, although

*To whom correspondence should be addressed: Department of Surgery, Duke University Medical Center, Box 2605, Durham, NC 27710. Phone: (919) 681-3886. Fax: (919) 681-7263. E-mail: bparker@duke.edu.

Published on Web 06/01/2010

the presence of amphiphilic α-helical potential in β-sheet proteins is ubiquitous, the location of such potential is not conserved within a given type of fold (*15*). This observation indicates that the amphiphilic α-helical potential in β-sheet proteins is apparently not involved in the stabilization of any particular native structure but rather is a candidate for the highly specific information that directs protein folding. This conclusion is further supported by the observation that segments with amphiphilic α-helical potential are not consistently associated with any particular type of secondary structure (*15*). A fourth factor that supports Lim's proposal is the collection of experimental observations that point toward a distinction or separation between amino acid sequence information that determines the native state stability and sequence information that governs the folding pathway(s) (*16*). The presence of dormant α-helical domains in a variety of proteins is an ideal candidate for primary structure components which determine protein folding pathway(s) in a manner independent of information necessary for stabilization of the native structure.

Although this wide range of theoretical considerations and indirect observations points strongly toward highly α-helical intermediates in the folding of all proteins, experimental evidence has been less convincing. Some experimental support has been established, as the early folding steps of globular α-helical proteins such as cytochrome *c* (*17*) and apomyoglobin (*18*) are known to involve the condensation of α-helices that have been identified as amphiphilic (*19*). Further, observations made when evaluating the folding of some β-sheet proteins provide strong support for Lim's hypothesis. For example, the folding of β-lactoglobulin is characterized by a long-lived and highly α-helical intermediate that is readily detected by stopped-flow circular dichroism (SFCD)[1] spectroscopy (*20*, *21*). However, experiments have failed to detect any apparently highly α-helical intermediate in the folding of some other β-sheet proteins. For example, SFCD studies reveal that folding of rat intestinal fatty acid binding protein (RIFABP) is characterized by a transition from aperiodic structure to predominantly β-strand without any apparent highly α-helical intermediate (*22*).

Because α-helices associated with the native state of a wide range of proteins are relatively stable because of optimized hydrogen bonding interactions, it might be supposed that any intermediate α-helical structure would also be relatively stable. With this in mind, the observation that some β-sheet proteins fold without detectable α-helical intermediates could suggest that α-helical intermediates are not involved in the folding of all proteins. However, this conclusion is impossible to reconcile with the theoretical considerations of Lim and a variety of quantitative analyses of amphiphilic α-helical potential in β-sheet proteins as discussed above. To reconcile these apparently conflicting indicators, we postulate an α-helical intermediate that is crucial to the folding process but is either relatively unstable or less stable than subsequent intermediates. A central question concerning this proposal asks about the nature of the forces that make subsequent intermediates even more stable than the inherently stable α-helical intermediate. A possible solution to this question lies in Lim's original proposal, which postulated that intermediate α-helices were essentially "stretched" into a linear conformation during later stages of the folding process. This proposal is supported by studies (*15*, *23*) demonstrating that segments with amphiphilic α-helical potential in β-sheet proteins generally form relatively extended structures "which may be obtained by stretching out the normal α-helix" (*23*).

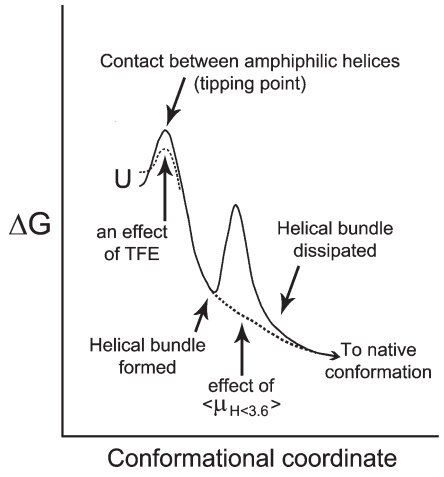

FIGURE 1: Qualitative free energy diagram corresponding to some of the early folding steps of predominantly β-sheet proteins. The ensemble of unfolded proteins ("U", top left), via an ensemble of diffusion−collision pathways, forms an ensemble of loosely formed helical bundles, which are then reorganized nonhelical, more nativelike structures. The transition state (the "tipping point" of the folding process) is that point just prior to stabilization of amphiphilic helices by coalescence. The potential reduction by TFE [≤ ∼10% (v/v)] of the energy barrier preceding the transition state is illustrated by the dashed line at the top left of the diagram. This effect is anticipated on the basis of the observation that TFE supports helix formation in a wide range of peptides, regardless of amino acid composition. TFE may increase the free energy of the unfolded ensemble and/or may decrease the free energy of the transition state at the point of contact between two amphiphilic helices (*29*, *38*, *46*). Both possibilities are illustrated by the dashed line in the top left of the diagram. Although TFE apparently stabilizes intermediate helical bundles in at least one protein (*20*, *21*), the effect of TFE on intermediate helical bundles in general has not been characterized and is therefore not illustrated in the diagram. The dashed line in the middle of the diagram illustrates the possible effect of the amphiphilic helical potential at 3.6 residues/turn being exceeded by the potential at periods shorter than 3.6 residues/turn. This potential is expected to reduce or perhaps, as indicated in the diagram, eliminate altogether the putative barrier to dissolution of the intermediate α-helical bundles.

The potential for a given sequence to form amphiphilic α-helices is calculated (*24*) on the basis of a periodicity of hydrophobicity of 3.6 residues. However, if, in addition to having substantial potential to form an amphiphilic α-helix, a given sequence had even greater amphiphilic potential at shorter periods, such as 3.0 residues, then the driving force of hydrophobic collapse would tend to overtwist any putative α-helix. Such overtwisting of the α-helical intermediate would simultaneously result in further coalescence of the hydrophobic core and an elongation of the local protein conformation. By this means, ensembles of intermediate helical bundles present during the folding process might be less stable than subsequent less helical, more nativelike structures (Figure 1). Thus, a greater amphiphilic helical potential at a period less than 3.6 residues/turn compared to the potential at 3.6 residues/turn is a potential indicator of ensembles of helical intermediates that are present transiently within ensembles of folding pathways but form no populated intermediate states.

To probe the validity of the hypothesis mentioned above, we evaluated the periodicity of hydrophobicity in a number of predominantly β-sheet proteins and in predominantly α-helical

[1]Abbreviations: TFE, 2,2,2-trifluoroethanol; SFCD, stopped-flow circular dichroism; RIFABP, rat intestinal fatty acid binding protein; GdnHCl, guanidine hydrochloride.

proteins. In addition, we evaluated the effects of 2,2,2-trifluoroethanol (TFE), a solvent that stabilizes α-helical structure, on the folding of RIFABP. The folding of RIFABP is well-characterized in the literature (*22*, *25*−*28*) and, as pointed out above, lacks a highly α-helical intermediate that can be detected by SFCD (*22*). Since TFE stabilizes α-helical conformations in a variety of peptides, regardless of amino acid composition (*29*−*31*), it is expected that TFE will drive the formation of transient α-helices in an ensemble of unfolded structures. An increased level of formation of transient α-helix in the unfolded ensembles will, in turn, increase the likelihood that collisions between different segments of the protein chain will lead to productive folding of those segments (*32*), if indeed α-helices are the critical structural elements involved in the initial steps of structure formation. Thus, it is hypothesized that TFE will accelerate the folding of RIFABP (Figure 1) if a rate-limiting step in folding is the diffusion and collision of amphiphilic α-helices, as proposed by Lim. Indeed, low concentrations of TFE [< ∼10% (v/v)] have been shown to profoundly accelerate the folding of a wide range of proteins, including some predominately β-sheet proteins (*33*). This effect of TFE is thought to be associated with some influence of TFE on the transition state and is observed in the absence of any increased stability of the native state (*29*). With this in mind, it is expected that the effects of α-helix-stabilizing solvents on the folding of RIFABP will provide insight into the potential role of α-helical intermediates in the folding of this protein, even if those intermediates are very short-lived.

MATERIALS AND METHODS

Cloning and Purification of Rat Intestinal Fatty Acid Binding Protein (*RIFABP*). RIFABP cDNA was amplified from reverse-transcribed RNA (purified from small intestine epithelial cells of adult male rats) using the following primers: forward (5′-TTAGGATCCATGGCATTTGATGGCACTTGG-3′) and reverse (5′-TTAGGATCCTCATTCCTTCTTAAAGATCCG-3′). Ligation, transformation, and protein induction were performed using the Affinity Protein Expression System (Stratagene, La Jolla, CA) for the pCAL-n vector according to the manufacturer's instructions, the one exception being that subcloning was performed in DH10B *Escherichia coli* (Invitrogen, Carlsbad, CA) rather than in XL1-Blue *E. coli*. The procedure resulted in expression of a fusion product with a thrombin cleavage site (TCS) linking a calmodulin binding peptide (CBP) with RIFABP. Positive clones were confirmed for the RIFABP insert by DNA sequencing using a CBP primer (5′-GAATTTCATAGCCGTCTCAGC-3′).

The CBP−TCS−RIFABP fusion product was expressed in BL21-Gold(DE3) *E. coli*. The identity of the RIFABP component was confirmed in the extract of the *E. coli* by molecular weight and by reactivity with a goat anti-fatty acid binding protein antibody [I-FABP (C-20), sc16063 (Santa Cruz Biotechnology, Inc., Santa Cruz, CA)] as assessed by immunoblotting. The CBP−TCS−RIFABP fusion product was purified using the Affinity Protein Purification System (Stratagene) as instructed by the manufacturer, with the addition of a step to remove any lipid from the fusion product. The CBP−TCS−RIFABP fusion product was delipidated using hydroxyalkoxypropal dextran [Lipidex 1000 (PerkinElmer Inc., Waltham, MA)] in 10 mM potassium phosphate (pH 7.4) at 37 °C for 2−3 h. Following delipidation, the CBP−TCS−RIFABP fusion product was cleaved using Thrombin CleanCleave (Sigma-Aldrich, St. Louis, MO), and the free CBP was removed from solution by binding to a calmodulin affinity column. The resulting RIFABP construct showed a single band by Coomassie staining of a sodium dodecyl sulfate−polyacrylamide gel. The construct contained, in addition to the naturally occurring amino acids in RIFAPB, a glycine-serine-methionine tag at the N-terminus. This tag corresponds to the C-terminal portion of the thrombin cleavage site (glycine-serine) with the N-terminal methionine, which is normally post-translationally removed from RIFABP. This GSM−RIFABP construct was used in all folding experiments described in this study.

Stopped-Flow Circular Dichroism. GSM−RIFABP refolding experiments were performed on a stopped-flow circular dichroism spectrophotometer (model 202-SF, AVIV, Lakewood, NJ) with a 1 mm path length cell. The CD data were monitored at 222 nm for 20 s with a mixing dead time of 30−130 ms. Although the dead time of the instrument was ∼12 ms, an apparent mixing artifact lasted for up to 130 ms depending on the solutions mixed in each run. Protein refolding was initiated by addition of 1 part of unfolded protein in 2.5 M GdnHCl, 20 mM KaP, and 0.1 mM EDTA (pH 7.2) to 4 parts of a renaturing buffer containing 20 mM KaP, 0.1 mM EDTA, either 0.63 M GdnHCl or 0.31 M GdnHCl, and various concentrations of TFE (pH 7.2). This resulted in a final solution containing 0.5 mg/mL RIFABP in 20 mM KaP, 0.1 mM EDTA, and either 1.0 or 0.75 M GdnHCl with various concentrations of TFE. The kinetic traces were smoothed using a 15-point Savitzky−Golay algorithm (Pro-Matlab, version R2006a) before the data were fit using the equation $f = A_o + A[\exp(-k_{obs}t)]$, where t is time, A_o is the final signal, and A is the amplitude of the observable phase (SigmaPlot, version 2.01). The relative amplitude corresponding to k_{obs} was calculated by normalization of the amplitude of the observable phase (A) with respect to the final signal (A_o), and the burst phase amplitude, reported as a percentage of the native CD signal, was calculated on the basis of normalization of the difference between the burst phase and unfolded protein CD signals with respect to the difference between the native and unfolded protein CD signals.

Evaluation of the Amphiphilic Helical Potential and the "Periodicity Ratio". The helical hydrophobic moment ($\langle\mu_H\rangle$), a quantitative measure of helical amphiphilicity, was calculated according to Eisenberg et al. (*24*) using the following equation

$$\langle\mu_{H,resX}\rangle = \left\{\left[\sum_{n=resX-5}^{resX+5} H_n \sin(\delta n)\right]^2 + \left[\sum_{n=resX-5}^{n=resX+5} H_n \cos(\delta n)\right]^2\right\}^{1/2} /11$$

where $\langle\mu_{H,resX}\rangle$ is the helical hydrophobic moment for a specific residue (resX), n is a specific residue in an 11-residue segment centered on resX, δ is the distance between residues as viewed down a helical axis (e.g., 100° for an α-helix), and H_n is the hydrophobic value assigned to residue n. Hydrophobic values were assigned to each amino acid using the method of Eisenberg (*34*).

As previously described, regions with high amphiphilic helical potential were considered to be regions with at least three consecutive $\langle\mu_H\rangle$ values of >0.35. This cutoff value was used since it was previously found to be useful in predicting the occurrence of native α-helix in α-helical proteins when a period of 3.6 residues/turn is used for the calculations (*19*).

When comparing the relative ratios of amphiphilic helical potential at different periodicities, P_A and P_B, we define the

Table 1: Basis Set of β-Sheet Protein Structures[a]

	structure	fold/superfamily/family	no. of residues	no. of segments
1	2cpl	cyclophilin-like	165	6
2	1hxn	four-blade β-propeller	219	7
3	1ifb	lipocalins/lipocalins/fatty acid binding protein-like	131	6
4	1rbp	lipocalins/lipocalins/retinol binding protein-like	182	6
5	1tgn	trypsin-like serine proteases/trypsin-like serine proteases/eukaryotic proteases	229	6
6	1lmw	trypsin-like serine proteases/trypsin-like serine proteases/eukaryotic proteases	253	6
7	1pgs	nucleoplasmin-like/VP (viral coat and capsid proteins)	314	10
8	1ton	trypsin-like serine proteases/trypsin-like serine proteases/eukaryotic proteases	235	6
9	1whi	ribosomal protein L14	122	7
10	1gof residues 151−537	seven-blade β-propeller/galactose oxidase, central domain/galactose oxidase, central domain	387	7
11	1gof residues 538−639	immunoglobulin-like β-sandwich	102	5
12	1mai	PH domain-like barrel	131	7
13	2alp	trypsin-like serine proteases/trypsin-like serine proteases/prokaryotic proteases	198	8
14	2cab	carbonic anhydrase	260	8
15	1thw	osmotin, thaumatin-like protein	207	8
16	1ivc	six-blade β-propeller/sialidases/sialidases (neuraminidases)	388	10
17	2pec	single-stranded right-handed β-helix/pectin lyase-like/pectate lyase-like	353	11
18	1snc	OB fold	149	10
19	1ppf	trypsin-like serine proteases/trypsin-like serine proteases/eukaryotic proteases	218	8
20	1ihs	trypsin-like serine proteases/trypsin-like serine proteases/eukaryotic proteases	295	11
21	2bbk	seven-blade β-propeller/YVTN repeat-like/quinoprotein amine dehydrogenase/methylamine dehydrogenase, H-chain	355	11
22	1kap	single-stranded right-handed β-helix/β-roll, serralysin-like metalloprotease, C-terminal domain	224	7
23	1kit	six-blade β-propeller/sialidases/sialidases (neuraminidases)	368	10
24	1bgl	supersandwich	293	12
25	1lla	immunoglobulin-like β-sandwich	249	6

[a]The Protein Data Bank file designation is listed (in the structure column), along with the fold designation according the SCOP, the number of amino acids in the structure, and the number of regions with high amphiphilic α-helical potential determined as described in Materials and Methods.

periodicity ratio (P_A/P_B) as

$$P_A/P_B = \frac{\text{for all RHP}_A \text{ with } \langle\mu_A\rangle > \langle\mu_B\rangle,\ \sum(\langle\mu_A\rangle - \langle\mu_B\rangle)}{\text{for all RHP}_B \text{ with } \langle\mu_B\rangle > \langle\mu_A\rangle,\ \sum(\langle\mu_B\rangle - \langle\mu_A\rangle)} \quad (1)$$

where RHP_A and RHP_B are residues that fall within regions with high amphiphilic helical potential at period A and period B, respectively, and $\langle\mu_A\rangle$ and $\langle\mu_B\rangle$ are the magnitudes of the helical hydrophobic moments at period A and period B, respectively.

Selection of Basis Sets of α-Helical and β-Sheet Folds. Fifty amino acid sequences defined by SCOP as a protein folding domain were selected as test cases for the study, half being classified as β-sheet folds (Table 1) and half being classified as α-helical folds (Table 2). The following selection criteria were used. (a) Nonglobular folds were avoided. (b) Smaller folds with fewer than five segments having high amphiphilic α-helical potential were avoided, since the parameters described in the text are averages over an entire fold, and thus, small folds might tend to add variance to the data. Finally, (c) a wide range of folds were sampled. Although five eukaryotic proteases belonging to the trypsin-like serine protease fold were included in the basis set of β-sheet proteins, it has been previously shown (*15*) that the locations of segments with high amphiphilic α-helical potential are not conserved within this family of proteins, and thus, each protein adds additional information to the basis set.

The basis set of β-sheet folds had an average length of 241 ± 86 (mean ± standard deviation) amino acids, with 7.96 ± 2.05 segments having high amphiphilic α-helical potential per fold. The basis set of α-helical folds had an average length of 197 ± 70 amino acids, with 7.56 ± 1.97 regions having high amphiphilic α-helical potential per fold. Because the distinction between folding domains within a single protein was not important for the calculations, and because the α-helical sequences tended to be shorter than the β-sheet sequences, both α-helical folds in hemocyanin as defined by SCOP were treated as a single fold for the purpose of deriving the basis sets.

RESULTS

Amphiphilic Helical Potential at 3.0 Residues per Turn in RIFABP and Equine β-Lactoglobulin. To determine whether a potential to form amphiphilic helices at a period of < 3.6 residues/turn that matches or exceeds the potential at 3.6 residues/turn might be involved in the folding of β-sheet proteins, the amphiphilic helical potential at 3.6 and 3.0 residues/turn was evaluated in selected proteins. Two predominantly β-sheet proteins, which have widely different folding behavior despite having the β-barrel, lipocalin fold (*35*), were selected for the initial study. As shown in Figure 2A, RIFABP, which does not form an observable α-helical intermediate during refolding (*22*), has significantly more amphiphilic helical potential at 3.0 residues/turn than at 3.6 residues/turn. On the other hand, equine β-lactoglobulin, which forms a detectable, highly α-helical intermediate during folding (*20*, *21*), has substantially more amphiphilic helical potential at 3.6 residues/turn than at 3.0 residues/turn (Figure 2B). These findings are consistent with the idea that RIFABP has "escape routes" leading away from an α-helical intermediate or intermediates, whereas β-lactoglobulin lacks such routes and provides a reasonable explanation for the differences in the observed folding kinetics between the two proteins despite their similar structures.

A schematic diagram showing the putative "folding code" of RIFABP is shown in Figure 3. Six regions with high amphiphilic

Table 2: Basis Set of α-Helical Protein Structures[a]

	structure	fold	no. of residues	no. of segments
1	1gln	an anticodon-binding domain of class I aminoacyl-tRNA synthetases	163	9
2	1agr	regulator of G-protein signaling (RGS)	353	10
3	1agr (α-subunit)	transducin (α-subunit), insertion domain	205	6
4	2abk	DNA glycosylase	211	7
5	1rgp	GTPase activation domain (GAP)	242	10
6	1bvp	a virus capsid protein α-helical domain	215	8
7	1ird	globin-like	141	5
8	1s0p	N-terminal domain of adenylylcyclase associated protein (CAP)	176	6
9	2j0o	IpaD-like	318	10
10	1hw1	GntR ligand-binding domain-like	152	8
11	1yoz	AF0941-like	116	6
12	1eyv	NusB-like	156	8
13	1rfz	YutG-like	168	8
14	1dbh	DBL homology domain (DH domain)	207	11
15	1hbn	methyl-coenzyme M reductase α- and β-chain C-terminal domain	280	6
16	2fef	PA2201 C-terminal domain-like	159	6
17	1jr3	post-AAA+ oligomerization domain-like	126	5
18	1hus	ribosomal protein S7	155	7
19	1aep	apolipophorin III	161	8
20	1o5h	methenyltetrahydrofolate cyclohydrolase-like	214	7
21	1g7d	ERP29 C domain-like	106	6
22	1dvk	functional domain of splicing factor Prp18	173	6
23	1lla	hemocyanin (both domains)	379	11
24	1m9c	retrovirus capsid protein, N-terminal core domai	146	5
25	2icw	superantigen MAM	213	10

[a]The Protein Data Bank file designation is listed (in the structure column), along with the fold designation according the SCOP, the number of amino acids in the structure, and the number of regions with high amphiphilic α-helical potential determined as described in Materials and Methods.

α-helical potential are identified (Figure 3, HPA, HPB, ..., HPF). All of those regions but one (Figure 3, HPB), which is associated with the two native α-helices found in the native structure (Figure 3, H1 and H2), apparently have an accompanying escape route or "release" in the form of amphiphilic helical potential at 3.0 residues/turn.

Transition in the Amphiphilic Helical Potential between 3.6 and 3.0 Residues per Turn. The amphiphilic helical potential at 3.0 residues/turn was used as an indicator of amphiphilic helical potential at a period of <3.6 residues/turn for several reasons. A helical period of 3.0 residues/turn (a) describes a structure that cannot maintain hydrogen bonding consistent with α-helices, (b) is sufficiently different from a period of 3.6 residues/turn such that the amphiphilic helical potentials at the two periods do not strictly correlate with one another, and (c) is not an apparent indicator of common secondary structure components. However, if indeed amphiphilic helical potential at periods shorter than 3.6 residues/turn serves as an effective escape route or release for intermediate helical bundles, then a smooth transition between amphiphilic helical potential at 3.6 and 3.0 residues/turn is expected. As shown in Figure 4, the amphiphilic helical potential at periodicities between 3.6 and 3.0 residues/turn generally fell between that seen at 3.6 and 3.0 residues/turn. Thus, optimization of hydrophobic interactions is expected to contribute to an energetically smooth transition for segments making the transition from intermediate α-helical bundles to a more extended conformation.

Amphiphilic Helical Potential at 3.0, 3.6, and 4.2 Residues per Turn in Globular Proteins Having Predominantly β-Sheet Structure and in Those Having Predominantly α-Helical Structure. To further evaluate the possibility that a potential to form amphiphilic helices at a period of <3.6 residues/turn might be involved in the folding of β-sheet proteins, the amphiphilic helical potential at 3.0, 3.6, and 4.2 residues/turn was evaluated in 25 predominantly β-sheet proteins (Table 1) and, as a control, in 25 α-helical proteins (Table 2). The 50 proteins were selected without prior knowledge of their helical potential at 3.0 and 4.2 residues/turn as described in Materials and Methods. Evaluation of the amphiphilic helical potential in these proteins revealed that proteins having β-sheet folds, in general, have a higher ratio of amphiphilic helical potential at a period of 3.0 compared to 3.6 residues/turn than do proteins with an α-helical fold ($p = 0.0005$; top diagram in Figure 5). A useful "control" for this calculation is to evaluate the amphiphilic helical potential at a period of 4.2 residues/turn in the same proteins. An amphiphilic helical period of 4.2 residues/turn is not expected to effectively provide a means by which intermediate α-helices may unfold, since hypothetical helical structures with a period greater than 3.6 residues/turn (a) are not expected to be stable due to an inability to maintain van der Waals interactions in the core of the helix as the diameter of the helix increases and (b) would lead to relatively compact conformations rather than the elongated conformations that are characteristic of non-α-helical regions with high amphiphilic α-helical potential in the native protein. In contrast to the preference of β-sheet proteins for amphiphilic helical potential at a period of 3.0 residues/turn, an amphiphilic helical potential at a period of 4.2 residues/turn was not favored particularly by β-sheet proteins over α-helical proteins ($p = 0.87$; bottom diagram in Figure 5). Thus, the fact that β-sheet proteins favor an amphiphilic helical period of 3.0 residues/turn and not 4.2 residues/turn indicates that the tendency for β-sheet proteins to favor a period of 3.0 residues/turn is a specific organizational feature within the amino acid sequence. These observations are consistent with the idea that, after initial coalescence of amphiphilic α-helices during the folding of β-sheet proteins, amphiphilic helical potentials with a period of <3.6 residues/turn provide a driving force for breaking the hydrogen bonds involved in α-helix formation. This driving force potentially reduces substantially

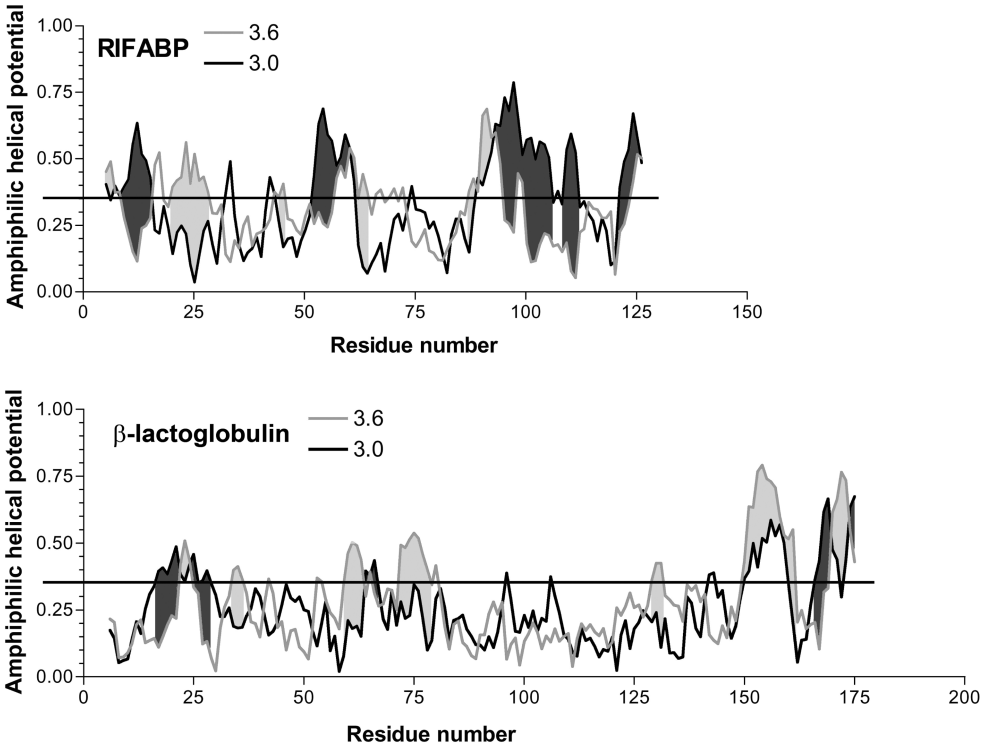

FIGURE 2: Amphiphilic helical potential at periods of 3.6 and 3.0 residues/turn in RIFABP and in equine β-lactoglobulin. The amphiphilic α-helical potential with a period of 3.6 residues/turn (gray line) and the amphiphilic helical potential with a period of 3.0 residues/turn (black line) were calculated according to the method of Eisenberg (*24*) as described in Materials and Methods. The horizontal line at $y = 0.35$ reflects a "cutoff" value set roughly at a point where three or more consecutive values that exceed that potential are indicative of native amphiphilic α-helices in α-helical proteins (*19*). Areas above this cutoff in which the amphiphilic α-helical potential (period of 3.6 residues/turn) exceeds the amphiphilic helical potential at 3.0 residues/turn are shaded light gray. Areas above this cutoff in which the amphiphilic helical potential at a period of 3.0 residues/turn exceeds the amphiphilic α-helical potential are shaded dark gray. The N-terminal methionine (post-translationally removed and thus not assigned a residue number in the native protein) is designated as residue zero in the calculations.

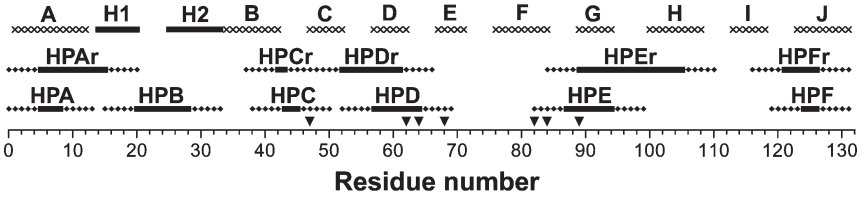

FIGURE 3: Putative folding code for RIFABP. The seven hydrophobic residues that are involved in the formation of a hydrophobic core early in the folding process are indicated by the triangles. The six regions with amphiphilic α-helical potential (as defined in Materials and Methods) are labeled HPA−HPF. The central portion of each region, which is associated with high amphiphilic α-helical potential, is shown as a solid block. Five residues on each side of that central portion, which are potentially involved in amphiphilic α-helix formation, are indicated by the stippled line. (A high potential at a given residue involves five residues on each side of that residue, since a window of 11 residues was used in the calculations as described in Materials and Methods.) In the middle row, the areas with high amphiphilic potential at a period of 3.0 residues/turn ($\langle\mu_{3.0}\rangle > 0.35$), which correspond to particular regions with high amphiphilic α-helical potential, are shown. Again solid and stippled lines represent the central portion and the five outlying residues of each region, respectively. These putative escape routes, or release routes, are labeled according to their associated region with high amphiphilic α-helical potential, along with the suffix r, for release. In the top row of the diagram, the native structure is indicated, with the 10 β-strands labeled A−J, and the two α-helices labeled H1 and H2. No release sequence was identified for HPB, which is associated with the two α-helices in the native structure. The N-terminal methionine (post-translationally removed and thus not assigned a residue number in the native protein) is designated as residue zero in the calculations.

or eliminates the energy trough that might otherwise be associated with those intermediate helical clusters (illustrated in Figure 1). This mechanism, which essentially entails a one-way, nonreversible path from amphiphilic α-helix formation to coalescence of those helices to their disruption during further coalescence of hydrophobic residues, might play a major role in the folding of many β-sheet proteins.

Effect of 2,2,2-Trifluoroethanol (TFE) on the Refolding of the GSM−RIFABP Construct. TFE, which strongly drives formation of α-helix (*36*, *37*), is expected to accelerate the folding of a β-sheet protein for which formation of a strongly helical intermediate is a rate-limiting step. RIFABP was selected for study because the kinetics and early folding of that protein are well-characterized (*22*, *26*−*28*) and because the protein exhibits no apparent highly α-helical intermediate during folding. As shown in Figure 6A, the CD signal at 222 nm for the GSM−RIFABP construct [RIFABP containing a glycine-serine-methionine tag at the N-terminus of the protein (see Materials and Methods)] refolding in 1.0 M GdnHCl was substantial in the burst phase (~48% of the native state CD signal), and recovery of the native state occurred within a few seconds. The presence of TFE accelerated the rate of refolding as indicated by an increase

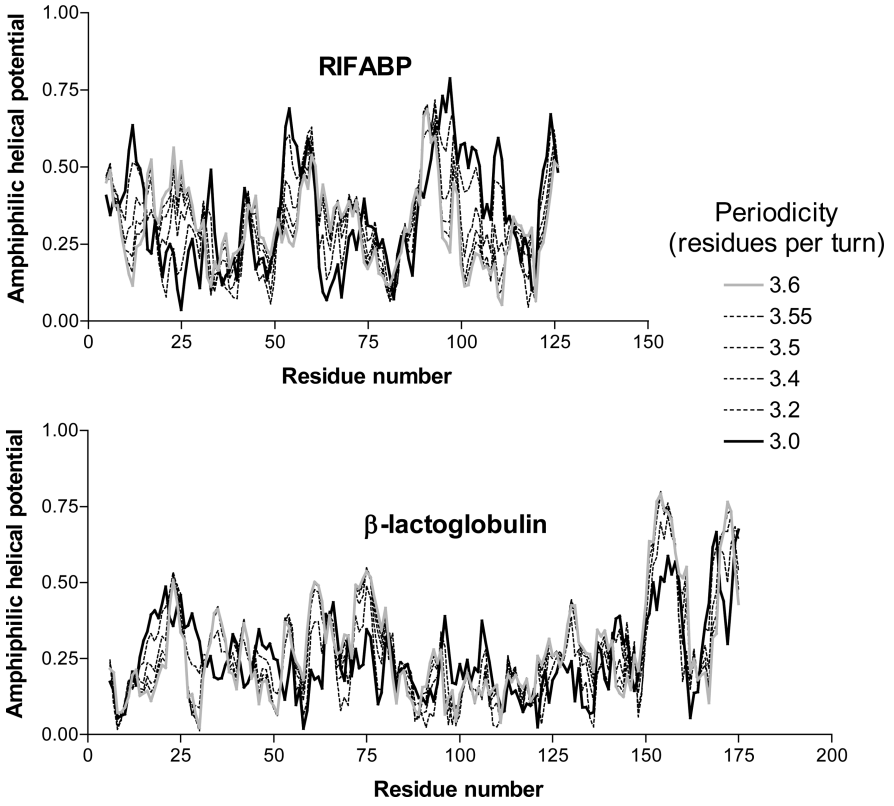

FIGURE 4: Amphiphilic helical potential at selected periods ranging from 3.6 to 3.0 residues/turn in RIFABP and in equine β-lactoglobulin. The amphiphilic helical potential at periods between 3.6 and 3.0 residues/turn (dashed lines) generally lies between the potential at 3.6 and 3.0 residues/turn.

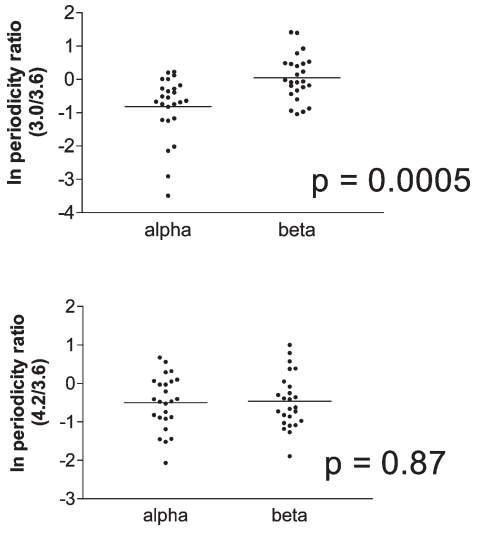

FIGURE 5: Relative periodicity of amphiphilic helical potential in predominantly α-helical proteins compared to that in predominantly β-sheet proteins. Fifty amino acid sequences comprising globular protein folding domains classified by SCOP as either α-helical folds ($n = 25$) or β-sheet folds ($n = 25$) (*35*) were selected such that the sequences chosen (a) represented a wide range of globular folds and (b) were of sufficient size to contain five or more regions of amphiphilic α-helical potential using a cutoff of three consecutive values of > 0.35. The natural log of the periodicity ratios for the proteins is shown, and p values were determined using an unpaired, two-tailed t test. The periodicity ratio of amphiphilic helical potential in general (P_A/P_B) is defined in eq 1 (see Materials and Methods for additional details). For example, $P_{3.0}/P_{3.6}$ for RIFABP is calculated as the area shaded dark gray in Figure 2A divided by the area shaded light gray in Figure 2A. The mean of each set is indicted by the horizontal line.

in the size of the burst phase and a decrease in the time to recovery of the native state (see, for example, Figure 6B,C). Although TFE can denature globular proteins at relatively high concentrations (*36*), the native states of globular proteins are generally not disrupted by concentrations of TFE below 10% (*29*, *33*, *38*). Consistent with this view, the native state CD signal of the GSM−RIFABP construct was observed in the presence of 1.0 M GdnHCl at TFE concentrations at or below 10% (v/v), indicating that the cosolvent did not destabilize the native state extensively at those TFE concentrations. However, as expected on the basis of the effects of TFE on a wide range of proteins (*33*, *38*), higher concentrations of TFE induced a CD signal that was indicative of a non-native, highly helical conformation [e.g., 20% (v/v) TFE but not 10% TFE induced a $> 20\%$ increase in negative ellipticity at 222 nm in the presence of 0.5 M GdnHCl]. The changes in burst phase amplitude as a function of TFE concentration in 1.0 M GdnHCl and in 0.75 M GdnHCl are shown in Table 3. These TFE-induced increases in burst phase amplitude can be attributed to protein conformational changes and indicate that TFE facilitates accelerated refolding of the GSM−RIFABP construct.

Acceleration of refolding was apparent when the CD signals measured in 1.0 and 0.75 M GdnHCl with and without TFE were fit to single-exponential processes as described in Materials and Methods. The GSM−RIFABP construct refolded in 1.0 and 0.75 M GdnHCl with rate constants, k_{obs}, of 1.9 and 2.7 s^{-1}, respectively, and relative amplitudes of 0.41 and 0.53, respectively. The refolding rate constant increased with increasing TFE concentrations in both 1.0 and 0.75 M GdnHCl (Figure 6D and Table 3). The relative amplitude of this folding phase tended to decrease as the TFE concentration increased for both 1.0 and 0.75 M GdnHCl (Table 3). In 10% TFE, the data for refolding in both 0.75 and 1.0 M GdnHCl could not be fit reliably because the protein had either completely or almost completely folded within the dead time of the SFCD instrument. However, given that the mixing dead time was approximately 130 ms for GSM−RIFABP refolding in 1.0 M

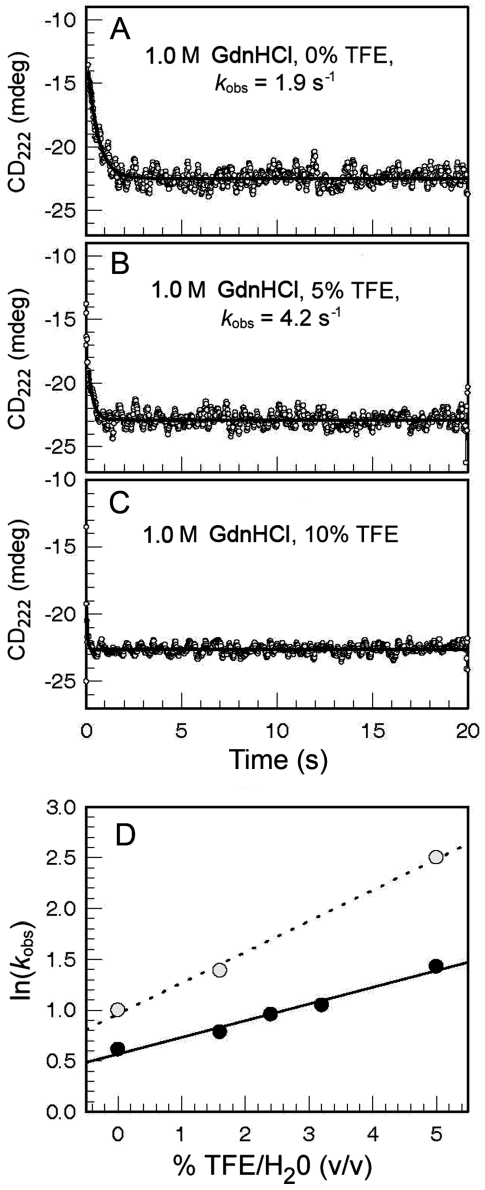

FIGURE 6: Acceleration of refolding of the GSM−RIFABP construct by TFE. The GSM−RIFABP construct was cloned, expressed, and purified as described in Materials and Methods. Stopped-flow CD measurements were obtained for the GSM−RIFABF construct refolding in 0.75 M GdnHCl and 1.0 M GdnHCl and in the presence of various concentrations of TFE. Eight to twelve sets of data were collected under each condition and averaged. The stopped-flow data for the refolding of the GSM−RIFABP construct in 1.0 M GdnHCl with (A) 0, (B) 5, and (C) 10% TFE are shown. The stopped-flow data are represented by the open circles, and the corresponding single-exponential fits are shown by the black lines. (D) The refolding rate constant, k_{obs}, measured in 0.75 M GdnHCl (○) and 1.0 M GdnHCl (●) is shown as a function of TFE concentration. The uncertainties of each k_{obs} value are ±10−15%. Refolding data measured with 10% TFE in 0.75 M GdnHCl or in 1.0 M GdnHCl could not be fit reliably because the protein folded or nearly folded within the dead time of the instrument. The CD signals for the native and unfolded (in 2.5 M GdnHCl) states of the GSM−RIFABP construct are −22.5 and −6.2 mdeg, respectively.

GdnHCl with 10% TFE, the rate constant is expected to be greater than 7.7 s^{-1} under those conditions. Similarly, for refolding in 0.75 M GdnHCl and 10% TFE, where the mixing dead time was ~30 ms, the rate constant is expected to be greater than 33 s^{-1}. These rate constants suggest that TFE accelerates GSM−RIFABP refolding >12- and >4-fold in 0.75 and 1.0 M GdnHCl, respectively.

Table 3: Effect of TFE on the Refolding of the GSM−RIFABP Construct

[GdnHCl] (M)	[TFE] (%, v/v)	burst phase amplitude (% of native state CD signal)	k_{obs} (s^{-1})	relative amplitude
1.0	0.0	48	1.9	0.41
1.0	1.6	60	2.2	0.42
1.0	2.4	60	2.6	0.4
1.0	3.2	66	2.9	0.35
1.0	5.0	79	4.2	0.35
1.0	10.0	94	>7.7[a]	ND[b]
0.75	0.0	53	2.7	0.53
0.75	1.6	72	6.8	0.55
0.75	5.0	91	12.2	0.39
0.75	10.0	100	>33[a]	ND[b]

[a]The k_{obs} (s^{-1}) values at 10% TFE were estimated on the basis of refolding completely or almost completely within the dead time of the instrument in 1.0 M GdnHCl (~130 ms) and in 0.75 M GdnHCl (~30 ms). [b]Not determined.

DISCUSSION

The observation that β-lactoglobulin rapidly forms an α-helical intermediate that is stable for several milliseconds during refolding (*20*, *21*) indicates that an intermediate containing substantial amounts of non-native α-helical structure can be stable. However, several factors point toward the inherent instability of the isolated α-helix and to the importance of a well-formed hydrophobic core as a stabilizing agent for amphiphilic α-helices. For example, an amino acid sequence that forms α-helix when located in one protein may form a different structure when in another protein (*39*−*41*), indicating that long-range interactions are important for stabilizing native structure. Consistent with this conclusion is the observation that peptide preparations derived by proteolytic digestion of animal tissue-derived proteins fail to form stable, periodic structures such as α-helices (*42*). With this in mind, it is reasonable to hypothesize that intermediate α-helical bundles formed during folding may be less stable than subsequent intermediates if (a) the hydrophobic core of the intermediate is not well formed and (b) subsequent intermediates have a better formed hydrophobic core.

The refolding of RIFABP has been extensively described in the literature, and the results described in this study must be interpreted in light of that work. The protein is an excellent candidate for folding studies because it contains no proline or cysteine residues that might lead to misfolded structures. Of primary importance is the observation that early folding events of RIFABP involve formation of a hydrophobic core consisting of residues 47, 62, 64, 68, 82, 84, and 89 (*27*). All of these residues are contained within regions having high amphiphilic α-helical potential and escape routes of amphiphilic helical potential with a period of <3.6 residues/turn (Figure 3). However, a substantial fraction of the RIFABP can be characterized as having high amphiphilic α-helical potential and accompanying escape routes (Figure 3), and thus, the association between amphiphilic helical potential and those residues forming a hydrophobic core during the early stages of folding may be coincidental. Nevertheless, the observed association is consistent with the idea that intermediate α-helices form but are rapidly lost during the folding of RIFABP. Further, experimental studies have led to the conclusion that β-barrels utilize multiple nucleation sites for folding and acquisition of native long-range interactions (*28*). For IFABP, that process includes a very rapid burst phase (rate of >10000 s^{-1}) in

which the hydrophobic core described above is formed (*27*). The involvement of multiple sites in a rapid process is consistent with the condensation of transient amphiphilic α-helices into short-lived α-helical bundles as proposed herein.

We hypothesized that an α-helical intermediate has not been detected by SFCD measurements of RIFABP refolding because such an intermediate is very short-lived, with perhaps only a small fraction of the molecules adopting a helical conformation at any given time. The observation of predicted escape routes for transient amphiphilic α-helices in the primary structure of predominantly β-sheet proteins certainly supports this hypothesis. If correct, this scenario presents a substantial challenge to detection by SFCD of transient α-helical intermediates during the refolding of RIFABP. In our study, TFE, which promotes formation of α-helix (*36*, *37*), was included in the refolding experiments to probe for the putative involvement of α-helical intermediates: TFE could potentially stabilize transient α-helical bundles, making those putative bundles observable by the SFCD approach, or TFE might accelerate the folding of RIFABP if formation of the transient α-helix is a rate-limiting step. The observation that 10% TFE accelerates the refolding of the GSM−RIFABP construct by more than 10-fold in the presence of 0.75 M GdnHCl is consistent with the idea that formation of intermediate α-helices is a rate-limiting step in the refolding of that protein and thus provides at least indirect experimental evidence supporting the formation of transient α-helices in the refolding of RIFABP. Further, the observation that the refolding of a wide range of proteins is accelerated by TFE (*33*) is consistent with the idea that a universal paradigm associated with protein folding is the rate-limiting coalescence of amphiphilic α-helices.

A potential problem affecting time-resolved folding studies is the question of whether proteins tend to fold via populated intermediates. One side of this argument suggests that proteins do not generally fold through an intermediate and that populated intermediates represent misfolded ensembles trapped in nonproductive pathways or that they arise from aggregation of denatured protein. Certainly, the detection of stable "intermediates" during the folding of some proteins can be due to trapping of nonproductive states. For example, whether the human spliceosomal protein U1A refolds rapidly from the unfolded monomer or slowly from an aggregate depends on the solvent conditions and protein concentration (*43*). However, in their studies of RIFABP folding, Ropson and colleagues (*22*) demonstrated that the amplitudes and relaxation times for both folding and unfolding were independent of the protein concentration in the range of 5−20 μM. Because the experimental conditions in the SFCD studies presented here were similar to those used by Ropson, with a GSM−RIFABP construct concentration of ∼5 μM, the observed kinetic traces are not apparently due to folding from protein aggregates, and the effects of TFE on the folding kinetics are not likely due to any potential influence of TFE on intermolecular interactions. Thus, the results strongly support the view that TFE promoted formation of structure that was productive in the folding process.

Of note is that the refolding rate of the GSM−RIFABP construct (final phase; k_{obs} = 1.9 s^{-1} in 1.0 M GdnHCl) observed here was somewhat faster than the previously reported refolding rate constant for RIFABP (0.33 s^{-1} in 1.0 M GdnHCl) (*22*). The source of this difference in rate constants is unknown, although the respective N-termini of the GSM−RIFABP construct and RIFABP may be responsible. One straightforward but speculative explanation is that the positively charged N-terminus of RIFABP interacts unfavorably with the helical dipole of the putative, transient N-terminal α-helix, inhibiting formation of that helix during folding. The GSM tag may provide sufficient separation between the helical dipole and the positive charge associated with the N-terminus, favoring formation of that helix and thereby accelerating the folding process. Interestingly, modification of Leu-64 also results in an increase in the refolding rate of RIFABP (*26*). Leu-64 is a component of the hydrophobic core that is formed most rapidly during folding (*27*) and is positioned very close to the N-terminus in the native structure. Further, Leu-64, like the N-terminus, is associated with high amphiphilic α-helical potential that is not α-helical in the native structure. Thus, it might be hypothesized that transient α-helices in the far N-terminus and in the region near Leu-64 both strongly influence the kinetics of refolding.

Although the acceleration of GSM−RIFABP folding by TFE supports the idea that the GSM−RIFABP construct may fold via the rate-limiting formation of a very short-lived α-helical intermediate(s), it does not prove that such an intermediate exists. For example, one potential explanation for the observed acceleration of folding of the GSM−RIFABP construct is that TFE accelerated formation of one or both of the two α-helices that occur in the native structure of RIFABP (*44*), and that these α-helices serve as a nucleus for the hierarchical folding of the rest of the protein. However, an alternative explanation consistent with Lim's hypothesis, that TFE accelerates folding by inducing formation of intermediate, substantially non-native, amphiphilic α-helical structures throughout the sequence, seems more likely for several reasons. First, the portion of the hydrophobic core that forms earliest in the folding process (*27*) is remote from the α-helices in the native structure of RIFABP. Second, the observation that RIFABP folds into a compact all-β-sheet fold even in the absence of the native α-helical segments (*45*) suggests that the folding of RIFABP does not depend on those native helices. Third, TFE is expected to induce formation of α-helix throughout the sequence, not only in the regions that form native α-helix. Nevertheless, the acceleration of GSM−RIFABP refolding by TFE does not constitute proof that non-native α-helices form early in the refolding of that protein but rather is consistent with that proposed folding mechanism.

One of the primary features of the Lim model described here is a flexibility in the folding code that is reflected in substantial variation in this code between proteins, even proteins that have similar folds. Although amphiphilic α-helical potential has been observed in all naturally occurring amino acid sequences evaluated to date, the location of that potential is not conserved within a given fold (*15*). In this study, additional flexibility and variation is evident. Although the potential to form amphiphilic α-helical bundles during folding is apparently ubiquitous in globular proteins, an escape route utilizing an amphiphilic helical potential with a period of < 3.6 residues is evidently not always present, as is seen for example in some parts of equine β-lactoglobulin (Figure 2). While predominantly β-sheet proteins tended to have a greater ratio of amphiphilic helical potential at a period of 3.0 residues/turn to that of 3.6 residues/turn ($P_{3.0}/P_{3.6}$) than did predominantly α-helical proteins, this observation applied to the average protein, not to all proteins. For example, four of the predominantly β-sheet proteins in the basis set had a lower $P_{3.0}/P_{3.6}$ than did the average predominantly α-helical protein. The observation probably accounts in part for variations in observed kinetics of folding even among proteins with similar folds and provides a basis for understanding at least in part why

some predominantly β-sheet proteins fold with relatively stable (and thus observable) α-helical intermediates, while others do not. This understanding of the folding process suggests a fundamental basis for folding that is compatible with a wide range of kinetics in the presence or absence of populated intermediate states.

In this work, we demonstrate that a correlate of Lim's model is predictive of features of the amino acid sequence. Specifically, a potential for amphiphilic helical potential at a period of < 3.6 residues/turn, which is expected to provide a unidirectional escape route for intermediate α-helical bundles, is indeed found to a greater extent in β-sheet proteins than in α-helical proteins. This result is predicted on the basis of the proposed model since such escape routes are presumably needed to a lesser extent in predominantly α-helical proteins than in predominantly β-sheet proteins. The ability of this model of protein folding to predict previously unidentified features of sequence information in a wide range of proteins provides exceptionally strong support for the model.

Lim's proposal provides a general mechanism by which all proteins might rapidly find a relatively compact conformation with a hydrophobic core, thus answering the Levinthal paradox. The amphiphilic α-helical potential present in a wide range of naturally occurring amino acid sequences, including those from viruses, bacteria, yeast, and humans, provided the basis for the initial proposal and strongly suggests that Lim's proposal does indeed describe a universal mechanism of protein folding. The focus of this study has been on predominantly β-sheet proteins, in which case most intermediate α-helices are necessarily non-native. A simpler paradigm by which amphiphilic α-helices facilitate protein folding involves proteins like myoglobin and cytochrome *c*, in which the helices formed early during folding are maintained in the native state. However, it seems likely that formation of intermediate, non-native, amphiphilic α-helical regions during folding also occurs in some proteins that contain native α-helical structure. Indeed, the presence of amphiphilic helical potential at a period of < 3.6 residues/turn is found in most α-helical proteins, although to a lesser extent on average than in β-sheet proteins. Thus, a wide range of proteins, including some α-helical proteins, probably utilize the rapid formation of amphiphilic α-helices and subsequent overtwisting of those helices during folding in a process driven by optimization of amphiphilic helical potential at a period of < 3.6 residues/turn.

ACKNOWLEDGMENT

We thank Susanne Meza-Keuthen for help with the data analysis, Roberto A. Bogomolni and David S. Kliger for helpful suggestions during preparation of the manuscript, Katelyn B. Connell for invaluable assistance with the CD studies, and Susan Marqusee for the generous use of her laboratory facilities.

REFERENCES

1. Lim, V. I., Mazanov, A. L., and Efimov, A. V. (1978) Stereochemical theory of the 3 dimensional structure of globular proteins. Part 1. Highly helical intermediate structures. *Mol. Biol.* (*Moscow*) *12*, 206–213.
2. Lim, V. I. (1978) Polypeptide chain folding through a highly helical intermediate as a general principle of globular protein structure formation. *FEBS Lett. 89*, 10–14.
3. Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988) Conformation of peptide fragments of proteins in aqueous solution: Implications for initiation of protein folding. *Biochemistry 27*, 7167–7175.
4. Huyghues-Despointes, B. M., Scholtz, J. M., and Baldwin, R. L. (1993) Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Protein Sci. 2*, 80–85.
5. Marqusee, S., and Baldwin, R. L. (1987) Helix stabilization by $Glu^{-}...Lys^{+}$ salt bridges in short peptides of de novo design. *Proc. Natl. Acad. Sci. U.S.A. 84*, 8898–8902.
6. Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M., and Baldwin, R. L. (1985) Nature of the charged-group effect on the stability of the C-peptide helix. *Proc. Natl. Acad. Sci. U.S.A. 82*, 2349–2353.
7. Huang, C. Y., Getahun, Z., Zhu, Y., Klemke, J. W., DeGrado, W. F., and Gai, F. (2002) Helix formation via conformation diffusion search. *Proc. Natl. Acad. Sci. U.S.A. 99*, 2788–2793.
8. Huang, C. Y., Getahun, Z., Wang, T., DeGrado, W. F., and Gai, F. (2001) Time-resolved infrared study of the helix-coil transition using ^{13}C-labeled helical peptides. *J. Am. Chem. Soc. 123*, 12111–12112.
9. Huang, C. Y., Klemke, J. W., Getahun, Z., DeGrado, W. F., and Gai, F. (2001) Temperature-dependent helix-coil transition of an alanine based peptide. *J. Am. Chem. Soc. 123*, 9235–9238.
10. 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.
11. Lednev, I. K., Karnoup, A. S., Sparrow, M. C., and Asher, S. A. (1999) α-Helix Peptide Folding and Unfolding Activation Barriers: A Nanosecond UV Resonance Raman Study. *J. Am. Chem. Soc. 121*, 8074–8086.
12. Thompson, P. A., Eaton, W. A., and Hofrichter, J. (1997) Laser temperature jump study of the helix⇌coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry 36*, 9200–9210.
13. Chen, E., Kumita, J. R., Woolley, G. A., and Kliger, D. S. (2003) The kinetics of helix unfolding of an azobenzene cross-linked peptide probed by nanosecond time-resolved optical rotatory dispersion. *J. Am. Chem. Soc. 125*, 12443–12449.
14. Parker, W., and Stezowski, J. J. (1996) The surface of β-sheet proteins contains amphiphilic regions which may provide clues about protein folding. *Proteins 25*, 253–260.
15. Parker, W., Sood, A., and Song, A. (2001) Organization of regions with amphiphilic α-helical potential within the three-dimensional structure of β-sheet proteins. *Protein Eng. 14*, 315–319.
16. Lattman, E. E., and Rose, G. D. (1993) Protein folding: What's the question? *Proc. Natl. Acad. Sci. U.S.A. 90*, 439–441.
17. Roder, H., Elove, G. A., and Englander, S. W. (1988) Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature 335*, 700–704.
18. Ballew, R. M., Sabelko, J., and Gruebele, M. (1996) Direct observation of fast protein folding: The initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. U.S.A. 93*, 5759–5764.
19. Parker, W., and Song, P. S. (1990) Location of helical regions in tetrapyrrole-containing proteins by a helical hydrophobic moment analysis. Application to phytochrome. *J. Biol. Chem. 265*, 17568–17575.
20. Hamada, D., Segawa, S.-I., and Goto, Y. (1996) Non-native α-helical intermediate in the refolding of β-lactoglobulin, a predominantly β-sheet protein. *Nat. Struct. Biol. 3*, 868–873.
21. Hamada, D., and Goto, Y. (1997) The equilibrium intermediate of β-lactoglobulin with non-native α-helical structure. *J. Mol. Biol. 269*, 479–487.
22. Ropson, I. J., Gordon, J. I., and Frieden, C. (1990) Folding of a predominantly β-structure protein: Rat intestinal fatty acid binding protein. *Biochemistry 29*, 9591–9599.
23. Lim, V. I. (1978) Stereochemical theory of the 3 dimensional structure of globular proteins. Part 2. Transition of the highly helical intermediate structure into the native one. *Mol. Biol.* (*Moscow*) *12*, 214–218.
24. Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1982) The helical hydrophobic moment: A measure of the amphiphilicity of a helix. *Nature 299*, 371–374.
25. Burns, L. L., Dalessio, P. M., and Ropson, I. J. (1998) Folding mechanism of three structurally similar β-sheet proteins. *Proteins 33*, 107–118.
26. Kim, K., Ramanathan, R., and Frieden, C. (1997) Intestinal fatty acid binding protein: A specific residue in one turn appears to stabilize the native structure and be responsible for slow refolding. *Protein Sci. 6*, 364–372.
27. Yeh, S. R., Ropson, I. J., and Rousseau, D. L. (2001) Hierarchical folding of intestinal fatty acid binding protein. *Biochemistry 40*, 4205–4210.

28. Bagby, S., Go, S., Inouye, S., Ikura, M., and Chakrabartty, A. (1998) Equilibrium folding intermediates of a Greek key β-barrel protein. *J. Mol. Biol. 276*, 669–681.
29. Buck, M. (1998) Trifluoroethanol and colleagues: Cosolvents come of age. Recent studies with peptides and proteins. *Q. Rev. Biophys. 31*, 297–355.
30. Luo, P., and Baldwin, R. L. (1997) Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry 36*, 8413–8421.
31. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1998) Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease T1. *Protein Sci. 7*, 383–388.
32. Myers, J. K., and Oas, T. G. (2002) Mechanism of fast protein folding. *Annu. Rev. Biochem. 71*, 783–815.
33. Hamada, D., Chiti, F., Guijarro, J. I., Kataoka, M., Taddei, N., and Dobson, C. M. (2000) Evidence concerning rate-limiting steps in protein folding from the effects of trifluoroethanol. *Nat. Struct. Biol. 7*, 58–61.
34. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol. 179*, 125–142.
35. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol. 247*, 536–540.
36. Shiraki, K., Nishikawa, K., and Goto, Y. (1995) Trifluoroethanol-induced stabilization of the α-helical structure of β-lactoglobulin: Implication for non-hierarchical protein folding. *J. Mol. Biol. 245*, 180–194.
37. Hirota, N., Mizuno, K., and Goto, Y. (1997) Cooperative α-helix formation of β-lactoglobulin and melittin induced by hexafluoroisopropanol. *Protein Sci. 6*, 416–421.
38. Reiersen, H., and Rees, A. R. (2000) Trifluoroethanol may form a solvent matrix for assisted hydrophobic interactions between peptide side chains. *Protein Eng. 13*, 739–743.
39. Guo, J.-T., Jaromczyk, J. W., and Xu, Y. (2007) Analysis of chameleon sequences and their implications in biological processes. *Proteins 67*, 548–558.
40. Mezei, M. (1998) Chameleon sequences in the PDB. *Protein Eng. 11*, 411–414.
41. Takano, K., Katagiri, Y., Mukaiyama, A., Chon, H., Matsumura, H., Koga, Y., and Kanaya, S. (2007) Conformational contagion in a protein: Structural properties of a chameleon sequence. *Proteins 68*, 617–625.
42. Parker, W., Partis, M., and Song, P. S. (1992) N-terminal domain of Avena phytochrome: Interactions with sodium dodecyl sulfate micelles and N-terminal chain truncated phytochrome. *Biochemistry 31*, 9413–9420.
43. Silow, M., and Oliveberg, M. (1997) Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl. Acad. Sci. U.S.A. 94*, 6084–6086.
44. Scapin, G., Gordon, J. I., and Sacchettini, J. C. (1992) Refinement of the structure of recombinant rat intestinal fatty acid-binding apoprotein at 1.2-Angstrom resolution. *J. Biol. Chem. 267*, 4253–4269.
45. Ogbay, B., Dekoster, G. T., and Cistola, D. P. (2004) The NMR structure of a stable and compact all-β-sheet variant of intestinal fatty acid-binding protein. *Protein Sci. 13*, 1227–1237.
46. Kentsis, A., and Sosnick, T. R. (1998) Trifluoroethanol promotes helix formation by destabilizing backbone exposure: Desolvation rather than native hydrogen bonding defines the kinetic pathway of dimeric coiled coil folding. *Biochemistry 37*, 14613–14622.