Kinetic Role of Helix Caps in Protein Folding Is Context-Dependent†

Gregory T. Kapp, Jane S. Richardson, and Terrence G. Oas*

Department of Biochemistry, Box 3711, Duke University Medical Center, Durham, North Carolina 27710

Received September 17, 2003; Revised Manuscript Received December 12, 2003

ABSTRACT: Secondary structure punctuation through specific backbone and side chain interactions at the beginning and end of α-helices has been proposed to play a key role in hierarchical protein folding mechanisms [Baldwin, R. L., and Rose, G. D. (1999) *Trends Biochem. Sci. 24*, 26−33; Presta, L. G., and Rose, G. D. (1988) *Science 240*, 1632−1641]. We have made site-specific substitutions in the N- and C-cap motifs of the 5-helix protein monomeric λ repressor (λ_{6-85}) and have measured the rate constants for folding and unfolding of each variant. The consequences of C-cap changes are strongly context-dependent. When the C-cap was located at the chain terminus, changes had little energetic and no kinetic effect. However, substitutions in a C-cap at the boundary between helix 4 and the subsequent interhelical loop resulted in large changes to the stability and rate constants of the variant, showing a substantial kinetic role for this interior C-cap and suggesting a general kinetic role for interior helix C-caps. Statistical preferences tabulated separately for internal and terminal C-caps also show only weak residue preferences in terminal C-caps. This kinetic distinction between interior and terminal C-caps can explain the discrepancy between the near-absence of stability and kinetic effects seen for C-caps of isolated peptides versus the very strong C-cap effects seen for proteins in statistical sequence preferences and mutational energetics. Introduction of consensus, in-register N-capping motifs resulted in increased stability, accelerated folding, and slower unfolding. The kinetic measurements indicate that some of the new native-state capping interactions remain unformed in the transition state. The accelerated folding rates could result from helix stabilization without invoking a specific role for N-caps in the folding reaction.

The native and denatured conformational ensembles involved in the process of protein folding are important in proper cellular function. If the native state is the functional species, then that function is lost during any time spent in the unfolded state or in the folding process. The denatured or unfolded ensemble of structures may be more susceptible to protein turnover through cellular degradation pathways. Additionally, proper cellular localization and protein trafficking is often coupled to the folding process. A detailed understanding of the medically important process of protein misfolding and aggregation requires better molecular characterization of kinetic folding mechanisms.

Each residue in a protein does not have an equal role in the folding process. Several studies have shown that often only a small subset of residues attain nativelike structure in the transition state (*3*−*6*). In this paper, we consider the role of helix termini in protein folding by determining the kinetic consequences of site-specific substitutions in helix N- and C-cap structural motifs.

Helix N- and C-cap residues are found at the beginning and end of helices, ending the regular i to $i + 4$ hydrogen bond pattern of the helix and marking the junction to the adjacent loop region (*2*, *7*). Numerous types of N- and C-cap motifs, involving the cap residue and its neighbors, have been described and classified (*8*). In this study, the N- and C-cap residues are defined as the first and last residues of the helix with Cα atoms within the helical spiral (*7*) (see Materials and Methods for exact details). Extensive analyses by several groups have shown that helix caps at the N-terminus can stabilize and increase the helicity of model peptides (*9*−*12*), and the position-specific effects of N-cap amino acid changes in peptides correlate well with their occurrence statistics (*7*) and mutational stability effects in proteins (*13*−*21*). The thermodynamic consequences of substituting C-terminal helix cap residues in model peptides are also well-studied (*22*) but are disappointingly small and show no trace of the dramatic effect of C-cap glycine expected from its frequent occurrence in this structural motif (*7*, *23*). This difference could be related to whether the C-cap is at the chain terminus, but that relationship has not been previously investigated, and the peptide results have apparently discouraged experimental study of C-caps in proteins.

Although a kinetic role has long been hypothesized for helix caps (*1*, *2*), experiments have suggested different levels of involvement for helix capping residues in the folding process. Substitution of N-cap residues in either helix of barnase indicates that the side chains at these locations are primarily non-native in the folding transition state ensemble[1] (TS) (*24*). In the CI2 protein, however, residue Ala16 near the N-terminus of the helix is an important part of a nativelike core in TS and retains its importance even in a circularly permuted variant of CI2 (*25*, *26*). Pulsed hydrogen exchange experiments on lysozyme indicated that residues near the

† This work was partially supported by NIH Grant GM-45322 (T.G.O.) and NIH Grant GM-15000 (J.S.R.). G.T.K. is the recipient of an NSF Graduate Research Fellowship.

* Corrresponding author. E-mail: (T.G.O.) oas@duke.edu.

10.1021/bi035683k CCC: $27.50
Published on Web 03/12/2004

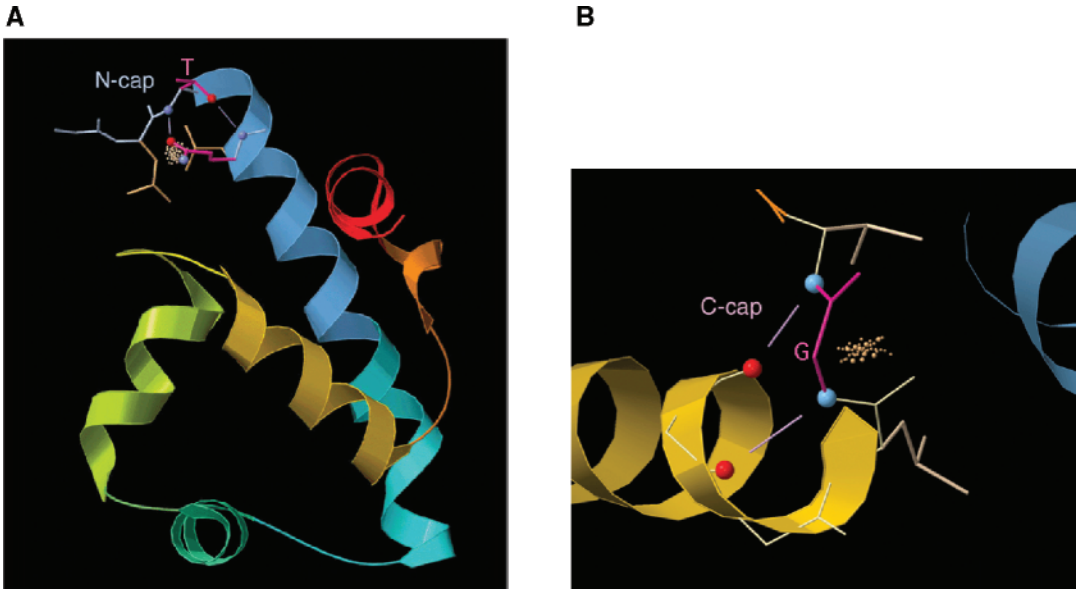

FIGURE 1: Classic N-cap and a classic C-cap illustrated on a ribbon diagram of the K70G* variant of λ_{6-85} (from chain 4 of the PDB file 1LMB). Panel A showing the N-cap of helix 1 (sky blue ribbon) shows the side chains of the Thr8 N-cap residue and the Gln11 N3 residue (in hot pink), the N-cap and cap-box side chain−backbone H-bonds (in purple, with color-coded balls on the H-bond O and N atoms), and the hydrophobic-staple side chains Leu7 and Leu12 with dots for their all-atom contacts (in peach). Panel B showing the improved C-cap of helix 4 (gold ribbon) includes the backbone (hot pink) of the Gly70 C-cap residue in Lα conformation, the two Schellman-motif backbone *i* to *i* − 3 and *i* to *i* − 5 H-bonds (in purple, with N and O balls), and the hydrophobic-staple side chains Ala and Val71 with dots for their all atom contacts (in peach). Figures are displayed in Mage (*50*), with all-atom contacts from Probe (*49*) and output rendered in Raster3D (*58*).

helix termini become protected from exchange along with the body of the helices in the first phase of folding (*27*). In the small domain of tumor suppressor p53, N-cap residues fold along with the body of the helix. Helical residues are not native in the p53 TS between the unfolded and the dimeric intermediate states, but almost all residues are highly nativelike in TS from the dimeric intermediate to the tetrameric native state (*28*). A range of kinetic effects has also been seen in single amino acid mutations at both helix ends in several other proteins (*3*, *5*, *6*).

The monomeric λ repressor protein (λ_{6-85}) is one of the fastest folding proteins known (*29*). It is small and helical, with short loops, high predicted helicity in helices 1 and 4 (*30*), and low contact order; however, none of these properties differs as much from average proteins as does its rapid folding. The extent to which almost all of the λ_{6-85} helix caps are complete and classic is, however, quite exceptional. A complete helix N-cap motif has three defining features (as shown in Figure 1A for the N-cap of λ_{6-85} helix 2): an H-bond from the N-cap Asn, Asp, Ser, or Thr (N/D/S/T) side chain to the backbone NH at N3 or N2 (*7*), a cap-box H-bond from an N3 (Gln, Glu, Ser, or Thr) side chain to the NH of the N-cap (*23*), and a hydrophobic-staple interaction between suitable side chains at N′ and N4 (*31*); see Materials and Methods for definitions. A survey of a 500-protein database showed that 10−20% of helices had all three N-capping features (depending on the stringency of criteria). λ_{6-85} is usually considered to have five helices (two with internal bends), and three of them, or 60%, have classic N-cap motifs (helices 1, 2, and 5).

A complete, classic helix C-cap motif also has three defining features, as shown in Figure 1B for the C-cap of λ_{6-85} helix 4: the C-cap residue (nearly always a Gly) in the Lα region of ϕ, ψ space near +40°, +35° with Schellman backbone H-bonds from C3 to C-cap and C4 to C′ and hydrophobic interactions between side chains at C′ and C4 (*7*, *32*, *33*). In our protein structure database, about 15−20% of helices have completely classic C-cap motifs; 33% as listed by Aurora et al. (*32*). λ_{6-85} has four classic C-caps (at the end of helices 1−4), or 80%, and reasons are given later as to why the absent C-cap motif at the chain C-terminus might not matter. The high content of classic helix caps (3−6 times more than normal) might partially explain the unusually rapid folding kinetics of λ_{6-85}. This paper will compare the folding kinetics, stability, and occurrence preferences of an N-cap and of helix C-caps at the chain terminus versus in the interior to explore their role in λ_{6-85} folding and to reconcile differences in C-cap preferences between helical peptides (*22*) and proteins.

In this study, we have introduced substitutions at several of the helix N- and C-caps in λ_{6-85}. It has previously been shown that λ_{6-85} folds in a two-state fashion (*34*); that the rate constants for the folding and unfolding can be monitored by NMR line shape analysis (*35*); and that these rate constants are sensitive to sequence (*29*, *36*, *37*). In the present study, we apply the same NMR line shape method to determine the kinetic consequences of capping perturbations. At the N-cap of helix 3, we have designed several substitutions to build a consensus capping box motif. Gly terminated C-caps have been introduced for helices 4 and 5. Each set of substitutions has also been further examined using the

[1] Abbreviations: N, native state ensemble; D, unfolded state ensemble; TS, transition state ensemble; λ_{6-85}, monomeric lambda repressor; WT, wild-type λ_{6-85}; WT*, λ_{6-85} with the G46A and G48A substitutions; NMR, nuclear magnetic resonance; CD, circular dichroism; HPLC, high performance liquid chromatography; MS, mass spectrometry; PDB, protein data bank; TMSP, sodium 3-(trimethylsilyl) propionate-2,2,3,3-d_4.

Φ_f analysis technique (*38*) to calculate the degree of native state interactions present in TS.

MATERIALS AND METHODS

Molecular Biology. Substitutions were introduced in two different sequences of monomeric λ repressor (λ_{6-85}). Sequence changes made in the WT background are denoted by the original amino acid and position followed by the new amino acid. Substitutions in the WT* sequence (already containing the stabilizing G46A and G48A substitutions) have similar nomenclature with an appended *. DNA substitutions were made using either the single-stranded mutagenesis protocol of Kunkel et al. (*39*) or the QuikChange kit (Stratagene). Each substitution was then verified by DNA sequencing of the entire gene.

λ repressor variants were expressed and purified using a previously published protocol (*29*). The final purity and homogeneity of each protein was checked by reverse-phase HPLC and electro-spray mass spectrometry. For each protein, the MS-determined mass was within 1 AMU of the expected mass, further confirming the proper amino acid substitutions. Both HPLC and mass spectrometry indicated no significant oxidation of the methionine residues. All protein concentrations were determined using the method of Edelhoch (*40*).

Equilibrium Measurements. Equilibrium stabilities of each variant were measured by automated urea titration using an Aviv 202 circular dichroism spectrometer interfaced to a Hamilton Microlab 500 titrator. CD samples consisted of 5−15 μM protein in a buffer of 20 mM sodium phosphate and 80 mM NaCl at pH 7.0. A titrant solution was made with urea (Nacalai Tesque, Kyoto, Japan) as an addition to the CD sample solution. The urea concentrations of the starting sample and titrant solution were determined from refractive index measurements (*41*). The CD sample was maintained at a constant 25 ± 0.1 °C throughout the titration. At each titration point, the urea concentration was increased, the sample was mixed for 30 s, and the signal at 222 nm was averaged for at least 30 s.

The raw CD signal was converted to mean residue ellipticity ($[\theta]_{222}$) and fit to a two-state transition using the equation of Nicholson and Scholtz (*42*), which yields the denaturant dependence of ΔG (m_{eq}) and the urea concentration at the midpoint of the denaturation curve ($C_{1/2}$). The product of m_{eq} and $C_{1/2}$ yields the equilibrium standard free energy of denaturation (ΔG_{N-D}) at 0 M urea, calculated using the linear extrapolation method (*43*).

Kinetic Measurements. NMR samples were prepared to span the entire range of the urea-induced folding transition. Samples contained approximately 200 μM protein in D_2O at pD 7.00 buffered with 20 mM deuterated sodium phosphate, 100 mM NaCl, 1 mM NaN_3, 20 μg/mL TMSP, and urea (Nacalai Tesque) previously exchanged with D_2O to isotopic purity of at least 99%. For each urea concentration, a Varian 500 MHz Innova instrument was used to collect a 1-D NMR spectrum. The aromatic region of each spectrum was analyzed using the ALASKA package (*44*) to determine rate constants for folding and unfolding. Rate constants derived from samples with native state populations of 3−97% were fit to eqs 1 and 2:

$$\ln k_f = \ln k_{f(4\text{ M urea})} + m_f([\text{urea}] - 4) \quad (1)$$

$$\ln k_u = \ln k_{u(4\text{ M urea})} + m_u([\text{urea}] - 4) \quad (2)$$

where $k_{f(4\text{ M urea})}$ and $k_{u(4\text{ M urea})}$ are the rate constants for folding and unfolding at a concentration of 4 M urea, and m_f and m_u reflect the urea dependence of each rate constant. Rate constants at 4 M urea are reported to minimize errors due to extrapolation of these fits.

Using the urea dependence of the rate constants, the solvent accessibility of TS as compared to the two end states was determined according to eq 3 (*45*):

$$\alpha = m_u/(m_u - m_f) \quad (3)$$

An α value of 1 indicates the solvent accessibility of TS is equivalent to the denatured state, while a value of 0 indicates nativelike solvent accessibility.

The effect of each substitution on the kinetics of the folding reaction can be quantified using eq 4 to calculate a Φ_f value[2] (*24*):

$$\Phi_f = \frac{-RT\ln\left(\frac{k_{f*}}{k_f}\right)}{-RT\ln\left(\frac{K_{variant}}{K_{wt}}\right)} = \frac{\ln\left(\frac{k_{f*}}{k_f}\right)}{\ln\left(\frac{k_{f*}k_u}{k_{u*}k_f}\right)} \quad (4)$$

where k_f and k_u are the rate constants before the substitutions, and k_{f*} and k_{u*} are the rate constants of the variants. All Φ_f values were calculated from rate constants at 4 M urea. A Φ_f of 1 indicates that the amino acid substitutions affect only the folding rate, while a Φ_f of 0 indicates kinetic effects on only the unfolding rate constants.

Definitions, Databases, and Modeling. Helix N- and C-caps are the interface residues that are half-in and half-out of a helix. In typical, unambiguous cases, the cap residue makes one helical backbone H-bond (the first or last one of the helix); it has nonhelical ϕ, ψ values, but its neighbor in the helix has helical ϕ, ψ; and its Cα lies near the next point on the spiral formed by the helical Cα positions, but the neighboring Cα outside the helix does not. When those criteria do not all agree, in this study we will use the Cα position as the dominant criterion because it is less sensitive to conformational distortions, gives the strongest correlations with sequence, and permits direct comparison to earlier statistics (*7*). In that system, the $+\phi$ glycines that end many

[2] Errors for Φ_f are propagated from errors in the individual k_f and k_u (wild type) and k_{f*} and k_{u*} (variant) measurements:

$$\sigma_\phi^2 = \left(\frac{\partial\phi}{\partial k_f}\right)^2\sigma_{k_f}^2 + \left(\frac{\partial\phi}{\partial k_u}\right)^2\sigma_{k_u}^2 + \left(\frac{\partial\phi}{\partial k_{f*}}\right)^2\sigma_{k_{f*}}^2 + \left(\frac{\partial\phi}{\partial k_{u*}}\right)^2\sigma_{k_{u*}}^2$$

$$\sigma_\phi = \left[\left[\frac{\ln\left(\frac{k_{f*}}{k_f}\right)}{k_f\ln\left(\frac{k_{f*}k_u}{k_fk_{u*}}\right)^2} - \frac{1}{k_f\ln\left(\frac{k_{f*}k_u}{k_fk_{u*}}\right)^2}\right]^2\sigma_{k_f}^2 + \left[-\frac{\ln\left(\frac{k_{f*}}{k_f}\right)}{k_f\ln\left(\frac{k_{f*}k_u}{k_fk_{u*}}\right)^2} + \frac{1}{k_f\ln\left(\frac{k_{f*}k_u}{k_fk_{u*}}\right)^2}\right]^2\sigma_{k_{f*}}^2 + \frac{\ln\left(\frac{k_{f*}}{k_f}\right)^2}{k_u{}^2\ln\left(\frac{k_{f*}k_u}{k_fk_{u*}}\right)^4}\sigma_{k_u}^2 + \frac{\ln\left(\frac{k_{f*}}{k_f}\right)^2}{k_{u*}{}^2\ln\left(\frac{k_{f*}k_u}{k_fk_{u*}}\right)^4}\sigma_{k_{u*}}^2\right]^{1/2}$$

helices are C-caps (*2*, *7*) rather than C′ residues (*31*, *32*). Positions at the helix ends are denoted relative to the caps as:

...N″ N′ N-cap N1 N2 N3 N4...
C4 C3 C2 C1 C-cap C′ C″...

where N1 through C1 are inside the helix, and primed positions are in the loops (or, at least, are outside this helix).

Statistical occurrence preferences for chain-terminal helix caps and for updated comparisons relative to λ_{6-85} were taken from the 500-protein, quality-filtered dataset of Lovell et al. (*46*). C-termini were defined as chain ends where the last residue included an oxt atom in the PDB file (*47*), and N-termini were defined as chains with the first residue number as 1 (or 2, in four cases) to avoid clipped or disordered ends. Helices were tabulated if their caps were within three residues of the end. There were 152 C-terminal helices, 76 that ended at the terminus and 76 with C-caps at Ct-1, Ct-2, or Ct-3 (1, 2, or 3 residues before the C-termini respectively). There were 69 N-terminal helices, 16 that started at the N-terminus and 53 with N-caps at residue 2, 3, or 4. A control set of 70 internal helices was sampled from the same files that provided the N-terminal examples.

Coordinates for residues 6−86 of chain 4 in the 1LMB crystal structure of the λ repressor dimer/DNA complex at 1.8 Å resolution (*48*) were used for modeling, with hydrogens added and optimized by Reduce (*49*). Empirical rules for improving capping motifs were taken from published summaries (*7*, *8*, *31*, *32*), with baseline positional occurrence frequencies (*7*). Negative design criteria were also applied by considering the effects of a sequence on possible alternative, offset capping positions. Individual variants were designed by adding the substituted side chains in ideal geometry and interactively evaluating the all-atom contacts in the MAGE/PROBE system (*50*) to find a rotameric conformation with good H-bonds and no steric clashes. If that was not possible without significant backbone motion (e.g., when trying to provide a cap-box H-bond for the N-cap of helix 4), the variant was not tried experimentally.

RESULTS

The sequence substitutions used in this study were made in one or both of two genetic backgrounds: wild type λ_{6-85} (WT) or the thermostable variant used in several previous studies, G46A/G48A. In the following, the G46A/G48A background is designated WT*, and all variants made in this background are designated with an asterisk following the sequence substitutions. The design of each λ_{6-85} variant began with careful inspection of the structure of the native state (*48*). As described in the introductory paragraphs, most of the helix ends in λ_{6-85} already show very classic N- or C-capping motifs. We have focused our experiments on those remaining helix caps that can be improved or changed without disruption of the overall native structure of the protein. Sequence changes at the helix caps were modeled to compare the residues with and without the features and interactions of consensus capping motifs while avoiding backbone motion or steric overlap between the altered side chains and the existing structure (see Materials and Methods).

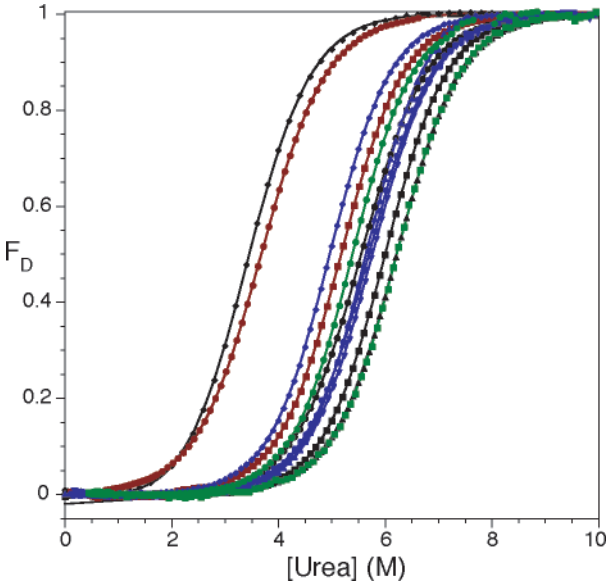

FIGURE 2: Plot of fraction denatured (F_D) at each urea concentration determined by circular dichroism. Colors are used to group the λ_{6-85} variants. N-cap variants in the WT* background: WT* (●), G43D/A46Q* (■), and G43D/S45A/A46Q* (▲). N-cap variants in the WT background: WT (red circle) and G43D/A46Q (red square). C-cap variants: K70G* (blue circle), K70A* (blue square), R85G+86K* (blue triangle), and R85A+86K* (blue down triangle). Curves represent the best fit to the fraction denatured data based on the equations of Nicholson and Scholtz (*42*).

Table 1: Equilibrium Parameters Determined by CD Urea Titration[a]

variant name	$C_{1/2}$ (M)	m_{eq} (cal/mol M)	ΔG_{N-D} (kcal/mol)
N-cap Variants			
WT*	5.24	980	5.13
G43D/A46Q*	6.00	1010	6.06
G43D/S45A/A46Q*	6.22	1010	6.28
WT	3.68	950	3.50
G43D/G46Q	5.18	1000	5.18
C-cap Variants			
R85G + 86K*	5.66	1040	5.89
R85A + 86K*	5.76	1000	5.76
K70G*	5.70	980	5.59
K70A*	4.82	1010	4.87

[a] All parameters were determined by two-state fits of the CD data. Errors in $C_{1/2}$ and m_{eq} are the standard deviations of multiple titrations of the same protein. Error in $C_{1/2}$ of ±0.05 M and error of ±30 cal/mol M in m_{eq} propagate to a ΔG_{N-D} error of approximately ±0.03 kcal/mol.

Each variant had a native CD spectrum consistent with the crystal structure of the WT protein. The equilibrium standard free energy of denaturation for each protein was determined by urea titration (Figure 2, Table 1). All titration curves were well-fit by a two-state model in which only fully native and fully denatured ensembles are significantly populated. The two-state model was also used to fit the rate constants for both folding and unfolding determined by NMR line shape analysis (Figure 3). The kinetic data show no evidence of any significantly populated partially folded state.

C-Cap Variants. At C-terminal helix caps, side chains are rarely involved in capping interactions, and the essential aspect is the C-cap residue's ability to adopt the positive backbone ϕ angles that allow favorable backbone-to-backbone interactions in a helix-breaking context. At the C-cap position, the prevalence of certain amino acids, notably

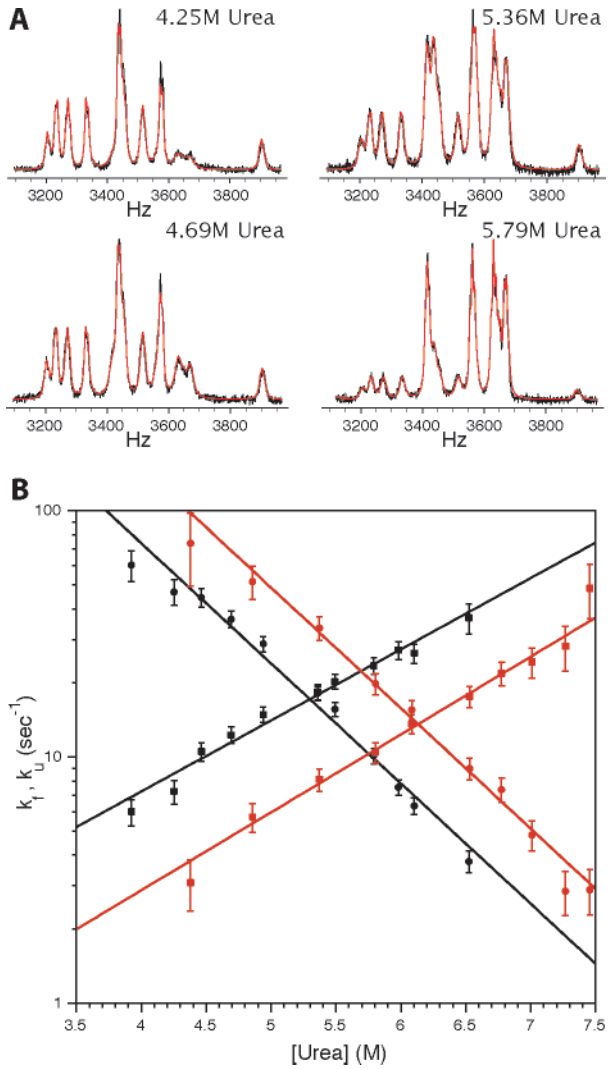

FIGURE 3: Representative data used to determine rate constants from NMR experiments. Panel A shows the aromatic region of the WT* protein at different urea concentrations around the midpoint in black. The red line is the fit using the ALASKA line shape analysis package. In panel B, the rate constants and errors determined from each NMR spectrum have been plotted. The k_f and k_u are plotted for both WT* (k_f ●, k_u ■) and G43D/A46Q* (k_f [red circle], k_u [red square]). The lines were fit using eqs 1 and 2 in the Materials and Methods.

Gly, is due to the side chain influence on the conformation of the backbone. In addition, a hydrophobic interaction is often found between residues surrounding the C-cap motif.

R85G+86K and R85G+86K**. The λ_{6-85} protein is derived from full-length N-terminal domain of λ repressor protein (102 amino acids) by a simple truncation at position 85 in the middle of helix 5. In this study, this nonnatural helix end was changed to place both Ala and Gly at the C-cap position, and Lys was added on the end to maintain a positive charge at the C-terminus. This new sequence allows for the formation of an αL C-cap motif (due to the charged C′) (*8*). The Gly variant should populate this structure, while the Ala variant should favor extension of helix 5 to the C-terminus. The location of the C-terminus of helix 5 (projected away from the protein into the solvent; see Figure 1) allows for the analysis of a chain-terminal C-cap in the absence of tertiary interactions from other helices.

Comparing the R85G+86K* protein to the R85G+86K* protein shows the Gly variant to be more stable by only 0.13 ± 0.04 kcal/mol. Studies of model peptides have shown that Ala in the C-cap position is 0.1 kcal/mol more stable than Gly (*22*). A helix ending at the end of the protein, without any tertiary interactions, is very similar to the peptides used as helical model systems. No difference was found in any of the rate constants or α values between R85G+86K* and R85A+86K*. When the stability difference is so small, the Φ_f values are unreliable. However, each variant can be compared to the WT* protein where there is a significant increase in stability. In each case, the Φ_f for the R85 residue is 0.33 ± 0.07, indicating that most of the energetic effects of extending helix 5 are not present in TS. This result is consistent with a TS that lacks a fully formed helix 5.

K70G and K70A**. The C-cap of helix 4 in λ_{6-85} contains the two backbone H-bonds characteristic of a Schellman motif with a hydrophobic interaction between the C4 (A66) and the C′ (V71) side chains and a C-cap residue (K70) in Lα conformation with positive ϕ and ψ angles (*33*). The C-cap Lys side chain could interact favorably with the helix macrodipole and protect the helix C-terminus from solvent during folding; it also helps bury Leu29 in the native tertiary structure. Working in the WT* background, we replaced the C-cap Lys with Gly, which should more readily adopt a positive backbone ϕ angle due to the lack of steric constraints imposed by a side chain Cβ. However, this K70G* substitution should lose the benefits of the various side chain interactions noted previously. As a control, we constructed the K70A* protein, which lacks the charge and most of the solvent protection ability but shares with Lys the disfavoring of Lα conformations. Comparison of K70G* and K70A* should indicate if the increased propensity of Gly for the positive ϕ angles has thermodynamic or kinetic effects on folding.

The K70A* protein is 0.26 ± 0.04 kcal/mol less stable than the WT* background due to the loss of the Lys side chain. Interestingly, the K70G* protein is 0.46 ± 0.04 kcal/mol more stable than the WT* background and thus 0.72 kcal/mol more stable than K70A*. The increased stability of K70G* is remarkable for two reasons: (1) any favorable interactions formed by the Lys side chain are absent and (2) Gly substitutions usually destabilize proteins by raising the chain entropy of the denatured state. Apparently, the standard free energy lost due to these factors is compensated by relaxation of backbone strain at or around residue 70. The same Ala to Gly substitution at the C-terminus of model helical peptides destabilized the helix by 0.1 kcal/mol (*22*). This difference between C-cap substitutions in the interior of the protein sequence and in single-helix model peptides should not be surprising. The backbone geometries that define C-caps and end the helix are important elements of protein tertiary structure. While these same geometries are accessible in model peptides, they are not determinants of simple helix formation. Secondary effects such as strain in the backbone around the C-cap would also be absent at the end of small peptides. In previous studies on other proteins, the energetic effect of most alanine to glycine substitutions at interior C-cap positions also show an increase in stability, as found for the helix 4 Gly C-cap in λ_{6-85}. In the two helices of Arc repressor, glycine C-caps stabilize the dimer by 1.0 and 1.3 ± 0.3 kcal/mol per residue over the alanine C-caps (*51*). Similar substitutions in barnase showed Gly C-caps to be more stable than Ala by 1.13 ± 0.04 kcal/mol at the C-cap

Table 2: Kinetic Parameters Determined by NMR Line Shape Analysis[a]

variant name	k_{f4M} (s^{-1})	m_f (M^{-1})	k_{u4M} (s^{-1})	m_u (M^{-1})	α
		N-cap Variants			
WT*	74 ± 4	−1.12 ± 0.03	7.2 ± 0.4	0.67 ± 0.04	0.37 ± 0.02
G43D/A46Q*	150 ± 20	−1.13 ± 0.05	2.9 ± 0.3	0.73 ± 0.05	0.39 ± 0.03
G43D/S45A/A46Q*	190 ± 30	−1.21 ± 0.07	1.7 ± 0.3	0.75 ± 0.08	0.38 ± 0.05
WT	29 ± 1	−1.36 ± 0.08	41 ± 2	0.56 ± 0.07	0.29 ± 0.04
G43D/G46Q	69 ± 4	−0.97 ± 0.03	10.0 ± 0.5	0.64 ± 0.03	0.40 ± 0.02
		C-cap Variants			
R85G + 86K*	110 ± 9	−1.17 ± 0.04	3.3 ± 0.3	0.78 ± 0.05	0.40 ± 0.03
R85A + 86K*	110 ± 10	−1.26 ± 0.06	3.3 ± 0.4	0.75 ± 0.07	0.37 ± 0.04
K70G*	240 ± 20	−1.10 ± 0.1	14 ± 2	0.68 ± 0.1	0.39 ± 0.09
K70A*	90 ± 10	−1.37 ± 0.05	10 ± 2	0.67 ± 0.07	0.33 ± 0.04

[a] The kinetic parameters for each protein were determined from a global fit of all NMR spectra. α was calculated according to eq 3 in the Materials and Methods.

of the first helix and 3.15 ± 0.04 kcal/mol for the C-cap of the second helix. The latter site of substitution is unusual in having a ϕ, ψ of +115°, +6° that is much more unfavorable for the Ala than the usual Lα values near +60°, +35° found for C-caps in λ_{6-85} and elsewhere (*52*). In the barstar protein, however, the Gly C-cap in the first helix is 1.3 kcal/mol less stable than the wild-type Ala (*5*); we do not see an explanation for that unusual behavior. All of these examples are Schellman motifs except for the αL motif (*32*) C-cap of barnase helix 1; within this very small sample, therefore, that distinction does not seem to matter.

Kinetically, the K70A* protein is very similar to the WT* background. The folding rate constant is the same within error, the K70A* unfolding rate constant is slightly greater, and TS has a similar solvent accessible surface area.

In contrast to the K70A substitution, the K70G change shows significant kinetic effects. As compared to the K70A* protein, the K70G* folding rate constant increases almost 3-fold, while unfolding is also slightly faster. A glycine C-cap at the boundary between the helix and the following loop residues is significantly more important to the folding reaction than the similar substitution at the C-cap of helix 5 (at the protein C-terminus). Alanine and glycine substitutions to interior C-caps (not at the chain terminus) of other proteins have also had large but rather variable kinetic effects. In both helices of the Arc repressor and in helix 1 of barnase, the Gly variant at the C-cap folds faster and unfolds more slowly than the Ala by factors of between about 1.5 and 25 (*52*, *53*). However, the Gly34 substitution in helix 2 of barnase folds slightly slower and unfolds about 300 times more slowly in 4 M urea (*52*). Collectively, these results suggest that formation of a helix C-cap with positive ϕ, ψ values may represent a high-energy rate-limiting step in the folding of helical proteins, particularly when the residue at that position is not glycine.

The faster folding K70G* protein has a α value similar to that of WT*. However, the Φ_f for the K70G substitution in the WT* background is 2.3 ± 0.7. This Φ_f outside the standard range of 0−1 suggests that the Lys residue is involved in higher energy interactions in TS than in the final native state. Φ_f values outside the standard range of 0−1 are not uncommon in protein mutagenesis studies, and they can result when substitutions cause changes in the populations of different pathways within the overall folding reaction (*54*), which would seem a plausible consequence of Gly versus Ala C-cap substitutions.

N-Cap Variants. Helix caps at the N-termini of helices are easily identified by the hydrogen bonding pattern between the amino acid side chains and the backbone. In one of the most common N-cap motifs, the helix capping box, a hydrogen bond is formed between the side chain of the N-cap residue and the backbone NH of the third helical residue (N3) plus a reciprocal hydrogen bond between the side chain of the N3 residue and the backbone NH of the N-cap residue. These reciprocal hydrogen bonds are often found along with a hydrophobic interaction, or staple, between the N′ and the N4 residues, then constituting a classic helix N-cap with all three defining features as seen in Figure 1A.

The amino acids at the N-terminus of helix 3 in WT λ_{6-85} form functionally important interactions at the DNA-binding site in the intact λ repressor, but they bear a poor resemblance in both sequence and structure to described N-cap motifs (*8*). The reciprocal hydrogen bonds characteristic of a N-terminal capping box are impossible at the N-terminus of helix 3 in either WT or WT* because the N-cap glycine (G43) and the N3 alanine (A46) are incapable of side chain hydrogen bonding. However, the hydrophobic interaction observed in most N-caps is present between the hydrophobic side chains of N′ (M42) and N4 (V47) (*31*).

We introduced a consensus capping box motif at the N-terminus of helix 3 by substituting the N-cap and N3 residues with side chains shown by modeling to be able to form the reciprocal hydrogen bonding pattern. The substitutions G43D and A46Q, along with the endogenous hydrophobic staple, yield a sequence capable of making the interactions that characterize a classic helix N-cap. The two substitutions were first introduced into the WT* background to generate the G43D/ A46Q* variant.

The G43D/A46Q* protein was stabilized, relative to the WT* background, by 0.93 ± 0.03 kcal/mol (Table 1). This degree of stabilization is consistent with the introduction of two hydrogen bonds and it is within the range of stabilization seen for similar N-capping box substitutions in other systems (*14*, *20*). The rate constants for the G43D/A46Q* variant indicate that the stabilization was the result of both a 2-fold increase in the folding rate (at 4 M urea) and approximately a 2-fold decrease in the rate constant for unfolding (Table 2). Analysis of the two substitutions together yields a Φ_f of 0.44 ± 0.06 (Table 3), indicating that only about half of the native state stability enhancement produced by the two new side chains is present in the transition state for folding. On the basis of the Φ_f analysis, the formation of full native state

Table 3: Φ_f Values Determined from Rate Constants

substitution(s)	original protein	new protein	Φ_f
G43/A46	WT*	G43D/A46Q*	0.44 ± 0.06
G43/S45/A46	WT*	G43D/S45A/A46Q*	0.40 ± 0.04
S45	G43D, A46Q*	G43D/S45A/A46Q*	0.31 ± 0.2
G43/A46	WT	G43D/A46Q	0.38 ± 0.02[a]
G48	G43D/A46Q	G43D/A46Q*	0.38 ± 0.05
K70	WT*	K70G*	2.3 ± 0.7
K70	WT*	K70A*	-0.53 ± 1[b]
K70	K70G*	K70A*	1.43 ± 0.5[c]
R85	WT*	R85G+86K*	0.33 ± 0.06
R85	WT*	R85G+86K*	0.33 ± 0.07

[a] α changed from 0.29 to 0.40 due to substitution. [b] α changed from 0.37 to 0.33 due to substitution; the very small $\Delta\Delta G_{N-D}$ due to the substitution makes the Φ_f value unreliable. [c] α changed from 0.39 to 0.33 due to substitution.

interactions is not an early obligatory step for these residues. This result implies that the newly designed N-cap motif for helix 3 is not fully present in TS.

To determine if the kinetic effects were due to a specific role of the helix cap motif or simply a result of the stabilization of helix 3, we made the G43D/S45A/A46Q* variant. The S45A substitution at the N2 position in helix 3 should have no role in any of the classic N-cap interactions; however, the S45A substitution should stabilize helix 3 since Ala has a greater helix propensity than Ser (*9*, *55*). Much like the earlier N-cap substitutions, the S45A substitution in the G43D/A46Q* background stabilizes the protein due to a faster folding rate constant and a slower unfolding rate constant. This suggests that the kinetic effects seen for the N-cap substitutions could result from simple stabilization of helix 3 and not from some specific role of the N-cap motif in the folding mechanism. The S45A substitution has a Φ_f of 0.31 ± 0.2; the three substitutions together (G43D, S45A, and A46Q) have a Φ_f of 0.40 ± 0.04.

To compare with G43D/A46Q*, the same substitutions were made in the WT sequence to produce the G43D/G46Q protein (in changing from the WT* to the WT background residue 46 and residue 48 change from Ala to Gly). In this background, the substitutions increased the stability by 1.68 $\pm$ 0.04 kcal/mol. The additional 0.75 kcal/mol of stabilization relative to the same substitutions in the WT* background may be due to the difference in the denatured state entropy of the two backgrounds. The kinetic effect of the G43D and G46Q substitutions in WT was similar to the effects in WT*. The rate constant for folding doubled, and the unfolding rate constant decreased 4-fold. The α-parameter (the solvent accessibility of TS relative to N and D) for G43D/G46Q increases from the lower WT value of 0.29 ± 0.04 to a value of 0.40 ± 0.02, similar to the WT* protein. Comparison of G43D/G46Q and G43D/A46Q* allows analysis of the effect of the G48A substitution in the 43D/46Q background. Since the α-value of both variants is the same, the G48A Φ_f of 0.38 ± 0.05 can be interpreted to mean that less than half of the energetic effects of the G48A substitution are present in TS. This Φ_f at position 48 is the same as that obtained for the N-cap substitutions at positions 43 and 46, suggesting at least the N-terminal half of helix 3 has a uniformly low Φ_f value.

Statistical Occurrence Preferences for Terminal versus Interior Helix Caps. To help interpret the dramatically different kinetic and energetic effects seen for chain-terminal versus interior helix C-caps, statistical amino acid preferences were investigated for C-caps of helices at the C-termini (Ct) of proteins. Position-specific preferences are indeed quite different and less distinct when the helix ends near the chain terminus, and almost all of that difference is concentrated at the very end (Table 4). Helices where the C-cap is 1, 2, or 3 residues before the chain end have amino acid preferences nearly indistinguishable from previously described overall C-cap preferences (*7*), with a relative Gly preference of 3.6 versus 3.9 and slightly greater preference for positive over negative charges (presumably because of the negatively charged C-terminus). In abrupt contrast, the preferences are very different (except for the charge bias) for helices that continue to the chain end: Ser and Gln are enriched 2-fold or more rather than disfavored, and Gly plummets to a preference of 0.5 (Table 4). Of the 17 Ser and Gln C-caps at the Ct position, 11 of those side chains H-bond to a backbone CO at either Ct-3 or Ct-4. The conformational properties of Gly, however, are not very useful at a Ct C-cap since there is no following chain to make other contacts if turned away from the helix. Even for C-caps at Ct-1 to Ct-3, the preference for Gly is only half as great when the end is free as when the end makes significant interactions with the rest of the molecule.

Table 4: Statistical Occurrence Preferences for Interior versus Chain-Terminal Helix C-caps

	overall	Ct-1 to Ct-3	chain term
Gly	3.9	3.6	0.5[a]
Gln	0.9	0.4	2.6[b]
Ser	0.8	0.4	1.9
Arg + Lys	1.2	1.6	1.6
Asp + Glu	0.5	0.3	0.2
Ala + Leu	0.8	1.3[b]	1.1
Pro	0.7	0.0	0.3
rest >3%	0.5	0.4	0.6

[a] Differ >3 SD. [b] Differ >2 SD from overall.

It is not known where the final helix terminates, either for the R85 variants of λ_{6-85} studied here or for model peptides studied in solution. Unsurprisingly, occurrence statistics in protein crystal structures show that if Ala is the penultimate chain residue (Ct-1), then 13/18 helices continue to the chain end, while if Gly is at Ct-1, then 14/17 helices end in a C-cap at that Gly: nine with a classic Schellman motif (*33*) and five with the α_L arrangement (*32*), which has a fully helical C1 conformation and an Lα C-cap with only one backbone H-bond. It is very likely, therefore, that an Ala to Gly change at Ct-1 moves the helix end one residue earlier. Since the chain-terminal carboxyl itself cannot H-bond to a backbone CO, an α-helical residue and a Schellman motif at Ct-1 would result in the same number of H-bonds, while an α_L motif would have one less. The helical H-bonds probably have somewhat better geometry, but an earlier C-cap has the advantage of placing the negatively charged terminus further away from the helix end. This apparently neutral balance is seen in the fact that there are 17 Gly and 18 Ala at the Ct-1 position in proteins where there is helix at least up to Ct-2. Overall, it should therefore not be surprising that Ala versus Gly makes no significant stability difference at the Ct or Ct-1 position of model helical peptides in solution (*22*, *30*). In λ_{6-85}, the C-terminal helix extends out somewhat from the rest of the protein, so the

possible difference in C-cap conformation between the R85G+86K* and the R85G+86K* variants is unlikely to provide significant interaction for Lys86. Thus, the statistical frequency data is consistent with the near-identical energetics and kinetics found for those two variants.

For helix N-caps, there is fairly close agreement between the helix-stabilization preferences measured for isolated peptides (*22*, *30*), where the cap is at the chain terminus, and the mutational stability effects (*13*−*20*), kinetic effects (these data) and occurrence frequencies (*7*) found for interior protein helices. That agreement was confirmed by statistical occurrence preferences tabulated for chain-terminal helix N-caps, where none of the individual amino acid preferences are found to differ significantly from the preference factors for interior helix N-caps. Like the C-terminal chain ends, there is a slightly stronger bias toward appropriately charged side chains (D + E over R + K at the N-terminus) than seen for interior helices. There are only 16 helices in our dataset with caps at residue 1, as compared with 76 helices that end at the Ct position; however, their amino acid distribution looks fairly typical for helix N-caps, with 50% Asn + Asp + Ser + Thr (N/D/S/T). Helix N-caps at residues 2−4 show only one strong difference from overall N-cap statistics: 70% of them have H-bonds from an N/D/S/T N-cap side chain to the backbone NH of N3 or N2, as opposed to only 38% in the reference data (*7*). Since the newer dataset has a higher average resolution and the expectations about N-caps have changed, we tabulated a control set of interior helices from the current dataset, which showed 49% with N-cap H-bonds, or 83% of the N/D/S/T residues. Eighty-two percent of the chain-terminal N-cap N/D/S/T have H-bonds (up from 71% in the 1988 dataset), so the observed H-bond difference comes entirely from the fraction of N/D/S/T residues for N-terminal versus interior N-caps: 85% versus 59% (or 54% in ref *7*), up by 3σ. Presumably, the N-cap H-bonds matter more where they can help prevent the chain end from fraying.

DISCUSSION

The high proportion of classic N- and C-capping motifs in the helices of the fast folding protein λ_{6-85} suggested a link between helix capping and fast folding kinetics. This notion is consistent with the view that helix caps may represent key determinants of folding mechanism, perhaps by providing the appropriate punctuation marks to prevent early-forming nascent helices from extending past the boundaries dictated by the final protein fold (*1*). Alternatively, helix caps may play a kinetically passive role and instead represent a commonly occurring thermodynamic solution to the problem of how to divert the backbone from a stable regular structure to nonregular structure (i.e., loops and turns (*56*)). If consensus helix cap interactions play an important role in the rate-determining steps in folding, we would expect sequence changes in these motifs to have dramatic effects on the rate constants. To further explore this hypothesis, we made sequence changes in λ_{6-85} to improve those helix caps that did not already conform to consensus capping motifs.

We made several sets of substitutions to the N-cap motif of helix 3, which does not conform to any of the consensus sequences, presumably because it is involved in DNA binding. Each sequence change designed to improve the N-cap motif did result in a significantly more stable protein with a concomitant decrease in unfolding rate and a less dramatic rise in folding rate constant (Tables 1 and 2). The relatively constant α among the N-cap variants suggests that all the variants share a TS solvent accessibility similar to the WT* variant. The intermediate Φ_f values of all the helix 3 N-cap substitutions (Table 3) indicate that not all of the capping interactions are formed in TS. If early formation of the helix 3 N-cap motif was an obligatory early step in λ_{6-85} folding, we would expect Φ_f values closer to unity. Although some of the N-cap interactions are formed in TS, this does not necessarily imply that they are part of the rate-determining step because nascent cap interactions could form much earlier than TS. The low Φ_f values of the N-cap substitutions and their moderate effects on folding rates indicate that the N-cap of helix 3 is not a necessary early step in the folding of λ_{6-85}, either because the cap is not necessary for the helix or because that helix is not a necessary part of TS.

These results for the helix 3 N-cap substitutions are consistent with our diffusion-collision model for the folding of λ_{6-85} (*57*). In the diffusion-collision mechanism, the fully native state is assembled from multiple collisions of partially folded native structure. In λ_{6-85}, these partially folded structures are the five helices, which form and decay rapidly on the time scale of the overall protein folding reaction. The diffusion-collision model predicts folding rates based on the intrinsic helicity of the five helices—more stable (more helical) helices yield more productive collisions and faster folding rates. Many groups have established that interactions in the N-cap motif increase the helicity of simple helical peptides (*9*−*12*), Thus, increased helicity of helix 3 due to N-cap substitutions alone can explain the increased folding rate constants. In each N-cap variant, the folding rate constant increases as the predicted helicity of helix 3 increases. It appears that N-cap substitutions affect the folding rate constant by stabilizing helix 3, thereby changing the productivity of collisions between helix 3 and other segments of structure. This indirect effect on folding rates suggests that helix N-caps play a passive role in folding mechanism and do not represent special kinetic punctuation marks to accelerate folding by preventing non-native helix extension.

The kinetic role of C-caps is not as easily described by the diffusion-collision model. Helical peptide studies of alanine and glycine in the C-cap position have shown that a glycine C-cap is only 0.1 kcal/mol less stable than an alanine C-cap (*22*). This small change in helix stability between Ala and Gly means that the diffusion-collision model would predict very similar rate constants for both alanine and glycine C-cap variants. Our results for the alanine and glycine substitutions at the terminal C-cap of helix 5 (Table 2) are consistent with this prediction. At the C-terminus of the protein, the alanine and glycine C-cap variants have the same rate constants. This result indicates that the C-cap at the C-terminus of the protein behaves like the C-cap of a free helical peptide.

At the internal C-cap of helix 4, however, the substitution of a nonglycine C-cap residue with a glycine results in a dramatic increase in the folding rate constant. Unlike the results at the terminal helix 5 C-cap, in the C-cap of helix 4, a glycine-based C-cap has a significantly greater folding rate

constant than either the alanine C-cap or the wild-type lysine C-cap. In striking contrast to the other N- and C-caps investigated in this study, the formation of an internal, nonglycine C-cap appears to be a rate-limiting step in λ_{6-85} folding. This strong preference for glycine indicates that the internal C-cap is not acting like the C-cap of a simple helical peptide. The statistical occurrence preferences also indicate that internal C-caps have significantly different residue compositions as compared to terminal C-caps. In a helical peptide, there is no structure C-terminal to the C-cap. At the internal C-cap of a protein, however, there is structure following the C-cap. At the C-cap of helix 4, the internal C-cap and the presence of structure C-terminal to the C-cap are important in determining the overall folding rate.

The C-cap motif has been described as having two different roles at the end of a helix (*32*). The first role is to adopt the positive ϕ and ψ dihedral angles and rigidify the final turn of the helix. Those positive ϕ and ψ angles also aid in the second role of the C-cap residue: to properly orient the trajectory of the backbone C-terminal to the helix. Comparison of the Ala-to-Gly substitutions at the internal C-cap of helix 4 and the terminal C-cap of helix 5 suggests that it is this second role that is responsible for the dramatic kinetic effects seen at the helix 4 C-cap. This distinction can solve the problem of what was formerly a rather troubling apparent contradiction between peptide and protein data on the importance of C-caps. In fact, both our results and the peptide data imply that the reversed backbone H-bonds of a Gly C-cap are not more stable than two helical H-bonds. C-caps do not stabilize helices; their function at the termini of interior helices seems to be their second role of ending the helix and orienting the following structure.

The internal C-cap at helix 4 must be forming the positive ϕ and ψ angles early in the folding reaction to correctly orient helix 5 relative to the end of helix 4. The glycine C-cap variant folds more rapidly because those positive ϕ and ψ angles are more accessible to the glycine residue. The nonglycine residues (Ala and Lys) sample the same positive ϕ and ψ angles less frequently due to the steric hindrance of the side chain Cβ. These steric effects make the formation of the native backbone angles a rate-limiting step for the nonglycine variants. The glycine substitution removes this limit and increases the rate constant for folding. The unusual folding Φ_f value of the K70G* variant, 2.3 ± 0.7, is evidence that the substitution has opened a new folding pathway in the overall folding mechanism (*54*). The K70G* variant, with a glycine at the internal C-cap position, folds rapidly because the glycine C-cap more frequently samples the positive ϕ and ψ angles needed to orient collisions of helix 4 and the C-terminal structure, helix 5. Thus, it appears that the kinetic role of sequence-optimized C-caps may be to promote the formation of native tertiary structure rather than to merely punctuate secondary structure.

CONCLUSION

The N-cap of helix 3 is only partly formed in TS and is unlikely to be rate limiting for folding. Instead, improvements to the N-cap motif affect the folding rate by increasing the intrinsic helicity of helix 3 thus increasing the productivity of collisions of helix 3 and other helices. Like the N-cap of helix 3, the C-cap of helix 5 appears to be only partially formed in TS, and the very small folding rate constant increases are consistent with very small increases in predicted helicity of helix 5. In contrast, the accelerated folding rate constant of the internal helix 4 glycine C-cap variant is not related to an increase in nascent helicity of helix 4. Instead, the internal C-cap affects the folding rate by adopting the required positive ϕ and ψ angles early in folding, thereby correctly orienting helix 5 relative to helix 4.

We predict that the introduction of nonglycine residues at the internal C-caps of helices 1−3 (all of which are glycines in both WT and WT*) would all result in significant decreases in folding rate constants modulated by the extent to which those particular helices are involved in TS. These substitutions would limit the accessibility of the positive ϕ and ψ angles, reducing the fraction of collisions in which the helices are properly oriented relative to each other. This should reduce the overall folding rate and possibly even change the overall mechanism of the folding process. Further studies are planned to test the general kinetic importance of internal C-caps both in λ_{6-85} and in other fast-folding proteins.

ACKNOWLEDGMENT

Bryan Arendall performed the MySQL database queries to generate files for statistical analysis of the chain-terminal helix caps. We also thank members of the Oas and Richardson labs for helpful discussions.

REFERENCES

1. Baldwin, R. L., and Rose, G. D. (1999) Is protein folding hierarchic? I. Local structure and peptide folding, *Trends Biochem. Sci. 24*, 26−33.
2. Presta, L. G., and Rose, G. D. (1988) Helix Signals in Proteins, *Science 240*, 1632−1641.
3. López-Hernández, E., and Serrano, L. (1996) Structure of the transition state for folding of the 129 aa protein Che Y resembles that of a smaller protein, CI-2, *Folding Des. 1*, 43−55.
4. Matouschek, A., Kellis, J. T., Jr., Serrano, L., and Fersht, A. R. (1989) Mapping the transition state and pathway of protein folding by protein engineering, *Nature 340*, 122−126.
5. Nölting, B., Golbik, R., Neira, J. L., Soler-Gonzalez, A. S., Schreiber, G., and Fersht, A. R. (1997) The folding pathway of a protein at high resolution from microseconds to seconds, *Proc. Natl. Acad. Sci. U.S.A. 94*, 826−830.
6. Srivastava, A. K., and Sauer, R. T. (2000) Evidence for Partial Secondary Structure Formation in the Transition State for Arc Repressor Refolding and Dimerization, *Biochemistry 39*, 8308−8314.
7. Richardson, J. S., and Richardson, D. C. (1988) Amino Acid Preferences for Specific Locations at the Ends of α Helices, *Science 240*, 1648−1652.
8. Aurora, R., and Rose, G. D. (1998) Temperature-sensitive mutations of bacteriophage T4 lysozyme occur at sites with low mobility and low solvent accessibility in the folded protein, *Protein Sci. 7*, 21−38.
9. Chakrabartty, A., Kortemme, T., and Baldwin, R. L. (1994) Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side chain interactions, *Protein Sci. 3*, 843−852.
10. Zhou, H. X., Lyu, P., Wemmer, D. E., and Kallenbach, N. R. (1994) _-Helix Capping in Synthetic Model Peptides by Reciprocal Side Chain−Main Chain Interactions: Evidence for an N Terminal "Capping Box", *Proteins 18*, 1−7.
11. Lyu, P. C., Zhou, H. X., Jelveh, N., Wemmer, D. E., and Kallenbach, N. R. (1992) Position-Dependent Stabilizing Effects in α-Helices: N-Terminal Capping in Synthetic Model Peptides, *J. Am. Chem. Soc. 114*, 6560−6562.
12. Lyu, P. C., Wemmer, D. E., Zhou, H. X., Pinker, R. J., and Kallenbach, N. R. (1993) Capping Interactions in Isolated

α-Helices: Position-Dependent Substitution Effects and Structure of a Serine-Capped Peptide Helix, *Biochemistry 32*, 421−425.
13. Serrano, L., and Fersht, A. R. (1989) Capping and α-helix stability, *Nature 342*, 296−299.
14. Lu, M., Shu, W., Ji, H., Spek, E., Wang, L., and Kallenbach, N. R. (1999) Helix Capping in the GCN4 Leucine Zipper, *J. Mol. Biol. 288*, 743−752.
15. Thapar, R., Nicholson, E. M., Rajagopal, P., Waygood, E. B., and Scholtz, J. M. (1996) Influence of N-Cap Mutations on the Structure and Stability of *Escherichia coli* HPr, *Biochemistry 35*, 11268−11277.
16. Serrano, L., Sancho, J., Hirshberg, M., and Fersht, A. R. (1992) α-Helix Stability in Proteins, *J. Mol. Biol. 227*, 544−559.
17. elMasry, N. F., and Fersht, A. R. (1994) Mutational analysis of the N-capping box of the α-helix of chymotrypsin inhibitor 2, *Protein Eng. 7*, 777−782.
18. Nicholson, H., Becktel, W. J., and Matthews, B. W. (1988) Enhanced protein thermostability from designed mutations that interact with α-helix dipoles, *Nature 336*, 651−656.
19. Nicholson, H., Anderson, D. E., Dao-pin, S., and Matthews, B. W. (1991) Analysis of the Interaction between Charged Side Chains and α-Helix Dipole Using Designed Thermostable Mutants of Phage T4 Lysozyme, *Biochemistry 30*, 9816−9828.
20. Zhukovsky, E. A., Mulkerrin, M. G., and Presta, L. G. (1994) Contribution to Global Protein Stabilization of the N-Capping Box in Human Growth Hormone, *Biochemistry 33*, 9856−9864.
21. Marshall, S. A., Morgan, C. S., and Mayo, S. L. (2002) Electrostatics significantly affect the stability of designed homeodomain variants, *J. Mol. Biol. 316*, 189−99.
22. Doig, A. J., and Baldwin, R. L. (1995) N- and C-capping preferences for all 20 amino acids in _-helical peptides, *Protein Sci. 4*, 1325−1336.
23. Harper, E. T., and Rose, G. D. (1993) Helix stop signals in proteins and peptides: the capping box, *Biochemistry 32*, 7605−7609.
24. Serrano, L., Matouschek, A., and Fersht, A. R. (1992) The Folding of an Enzyme. III. Structure of the Transition State for Unfolding of Barnase Analyzed by a Protein Engineering Procedure, *J. Mol. Biol. 224*, 805−818.
25. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) The Structure of the Transition State for Folding of Chymotrypsin Inhibitor 2 Analyzed by Protein Engineering Methods: Evidence for a Nucleation−Condensation Mechanism for Protein Folding, *J. Mol. Biol. 254*, 260−288.
26. Otzen, D. E., and Fersht, A. R. (1998) Folding of Circular and Permuted Chymotrypsin Inhibitor 2: Retention of the Folding Nucleus, *Biochemistry 37*, 8139−8146.
27. Radford, S. E., Dobson, C. M., and Evans, P. A. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways, *Nature 358*, 302−307.
28. Mateu, M. G., Sánchez Del Pino, M. M., and Fersht, A. R. (1999) Mechanism of folding and assembly of a small tetrameric protein domain from tumor suppressor p53, *Nat. Struct. Biol. 6*, 191−198.
29. Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W., and Oas, T. G. (1996) Microsecond Protein Folding through a Compact Transition State, *J. Mol. Biol. 263*, 311−322.
30. Muñoz, V., and Serrano, L. (1994) Elucidating the folding problem of helical peptides using empirical parameters, *Nat. Struct. Biol. 1*, 399−409.
31. Muñoz, V., Blanco, F. J., and Serrano, L. (1995) The hydrophobic-staple motif and a role for loop residues in α-helix stability and protein folding, *Nat. Struct. Biol. 2*, 380−385.
32. Aurora, R., Srinivasan, R., and Rose, G. D. (1994) Rules for α-Helix Termination by Glycine, *Science 264*, 1126−1130.
33. Schellman, C. (1980) The α-L conformation at the ends of helices, in *Protein Folding* (Rainer Jaenicke, Ed.) pp 53−61, Elsevier/North-Holland, New York.
34. Huang, G. S., and Oas, T. G. (1995) Structure and Stability of Monomeric λ Repressor: NMR Evidence for Two-State Folding, *Biochemistry 34*, 3884−3892.
35. Huang, G. S., and Oas, T. G. (1995) Submillisecond folding of monomeric λ repressor, *Proc. Natl. Acad. Sci. U.S.A. 92*, 6878−6882.
36. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997) The energy landscape of a fast-folding protein mapped by Ala → Gly Substitutions, *Nat. Struct. Biol. 4*, 305−310.
37. Myers, J. K., and Oas, T. G. (1999) Contribution of a buried hydrogen bond to λ repressor folding kinetics, *Biochemistry 38*, 6761−6768.
38. Matouschek, A., and Fersht, A. R. (1991) Protein engineering in analysis of protein folding pathways and stability, *Methods Enzymol. 202*, 82−112.
39. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection, *Methods Enzymol. 154*, 367−382.
40. Edelhoch, H. (1967) Spectroscopic Determination of Tryptophan and Tyrosine in Proteins, *Biochemistry 6*, 1948−1954.
41. Pace, C. N. (1986) Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves, *Methods Enzymol. 131*, 266−280.
42. Nicholson, E. M., and Scholtz, J. M. (1996) Conformational Stability of the *Escherichia coli* HPr Protein: Test of the Linear Extrapolation Method and a Thermodynamic Characterization of Cold Denaturation, *Biochemistry 35*, 11369−11378.
43. Pace, C. N., and Shaw, K. L. (2000) Linear extrapolation method of analyzing solvent denaturation curves, *Proteins Suppl.*, 1−7.
44. Burton, R. E., Busby, R. S., and Oas, T. G. (1998) ALASKA: A *Mathematica* package for two-state kinetic analysis of protein folding reactions, *J. Biomol. NMR 11*, 355−360.
45. Chen, B. L., Baase, W. A., Nicholson, H., and Schellman, J. A. (1992) Folding kinetics of T4 lysozyme and nine mutants at 12 °C, *Biochemistry 31*, 1464−1476.
46. Lovell, S. C., Davis, I. W., Arendall, W. B. I., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi, psi, and Cbeta deviation, *Proteins 50*, 437−450.
47. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235−242.
48. Beamer, L. J., and Pabo, C. O. (1992) Refined 1.8 Å crystal structure of the λ repressor−operator complex, *J. Mol. Biol. 227*, 177−196.
49. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side chain amide orientation, *J. Mol. Biol. 285*, 1735−1747.
50. Word, J. M., Bateman, R. C., Jr., Presley, B. K., Lovell, S. C., and Richardson, D. C. (2000) Exploring steric constraints on protein mutations using MAGE/PROBE, *Protein Sci. 9*, 2251−2259.
51. Milla, M. E., Brown, B. M., and Sauer, R. T. (1994) Protein stability effects of a complete set of alanine substitutions in Arc repressor, *Nat. Struct. Biol. 1*, 518−523.
52. Matthews, J. M., and Fersht, A. R. (1995) Exploring the energy surface of protein folding by structure−reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for structural elements for unfolding/ folding of barnase, *Biochemistry 34*, 6805−6814.
53. Milla, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995) P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding, *Biochemistry 34*, 13914−13919.
54. Ozkan, S. B., Bahar, I., and Dill, K. A. (2001) Transition states and the meaning of phi values in protein folding kinetics, *Nat. Struct. Biol. 8*, 765−769.
55. Cochran, D. A., and Doig, A. J. (2001) Effect of the N2 residue on the stability of the _-helix for all 20 amino acids, *Protein Sci. 10*, 1305−1311.
56. Martinez, J. C., Pisabarro, M. T., and Serrano, L. (1998) Obligatory steps in protein folding and the conformational diversity of the transition state, *Nat. Struct. Biol. 5*, 721−729.
57. Burton, R. E., Myers, J. K., and Oas, T. G. (1998) Protein Folding Dynamics: Quantitative Agreement of Theory and Experiment, *Biochemistry 37*, 5337−5343.
58. Merritt, E. A., and Bacon, D. J. (1997) Raster3D: Photorealistic Molecular Graphics, *Methods Enzymol. 277*, 505−524.

BI035683K