Subdomain Folding of the Coiled Coil Leucine Zipper from the bZIP Transcriptional Activator GCN4[†]

Kevin J. Lumb, Chavela M. Carr, and Peter S. Kim[*]

Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142

ABSTRACT: One popular model for protein folding, the framework model, postulates initial formation of secondary structure elements, which then assemble into the native conformation. However, short peptides that correspond to secondary structure elements in proteins are often only marginally stable in isolation. A 33-residue peptide (GCN4-p1) corresponding to the GCN4 leucine zipper folds as a parallel, two-stranded coiled coil [O'Shea, E. K., Klemm, J. D., Kim, P. S., & Alber, T. A. (1991) *Science 254*, 539–544]. Deletion of the first residue (Arg 1) results in local, N-terminal unfolding of the coiled coil, suggesting that a stable subdomain of GCN4-p1 can form. N- and C-terminal deletion studies result in a 23-residue peptide, corresponding to residues 8–30 of GCN4-p1, that folds as a parallel, two-stranded coiled coil with substantial stability (the melting temperature of a 1 mM solution is 43 °C at pH 7). In contrast, a closely related 23-residue peptide (residues 11–33 of GCN4-p1) is predominantly unfolded, even at 0 °C, as observed previously for many isolated peptides of similar length. Thus, specific tertiary packing interactions between two short units of secondary structure can be energetically more important in stabilizing folded structure than secondary structure propensities. These results provide strong support for the notion that stable, cooperatively folded subdomains are the important determinants of protein folding.

One popular model for protein folding, the framework model, postulates initial formation of secondary structure elements, which then assemble into the native conformation [reviewed in Kim and Baldwin (1982)]. Isolated peptides that correspond to secondary structure elements, however, are often only marginally stable [e.g., Scholtz and Baldwin (1992) and Dyson and Wright (1993)]. This suggests that stabilizing interactions between transiently formed elements of secondary structure are required for proceeding to the native state. To address how elements of secondary structure associate and stabilize each other, we have investigated the folding of a two-stranded coiled coil, which is a simple motif consisting of just two α-helices (Figure 1A).

Coiled coils are formed from two, three, or four amphipathic α-helices that wrap around one another in a left-handed supercoil [Crick, 1953; Pauling & Corey, 1953; O'Shea et al., 1991; Harbury et al., 1993; see also Banner et al. (1987), Lovejoy et al. (1993), and Yan et al. (1993)]. This structural motif is found in many proteins (Cohen & Parry, 1990), including the leucine zipper, or bZIP,[1] class of transcription factors [Landschulz et al., 1988; O'Shea et al., 1989; for recent reviews, see Alber (1992), Hu and Sauer (1992), and Ellenberger (1994)]. The sequences of coiled coils are characterized by a heptad repeat of seven amino acid residues, denoted a to g, with a 4,3 repeat of predominantly hydrophobic

residues at positions a and d (Hodges et al., 1972; McLachlan & Stewart, 1975). The intercoil hydrophobic interface (Figure 1A) is formed by residues at positions a, d, e, and g (O'Shea et al., 1991; Harbury et al., 1993).

Studies of the folding of coiled coils have focused predominantly on tropomyosin, a 400-Å-long, two-stranded coiled coil. Tropomyosin unfolds thermally with a shallow pretransition, involving loss of about 20% helicity, followed by a major unfolding step (Lehrer, 1978). The nature of the pretransition is unclear, having been attributed to noncooperative helix fraying (Holtzer & Holtzer, 1992), unfolding in the vicinity of residues 130–190 (Ishii et al., 1992), and loss of coiled coil but not helical structure (Greenfield & Hitchcock-DeGregori, 1993).

Shorter coiled coils may exhibit simpler folding behavior than tropomyosin. A synthetic peptide, called GCN4-p1, corresponding to the leucine zipper of the yeast bZIP transcription factor GCN4, folds independently of the remainder of the GCN4 sequence as a parallel, two-stranded coiled coil (O'Shea et al., 1989, 1991; Oas et al., 1990). We define here peptides derived from the GCN4 coiled coil that form stable, folded structures. The results emphasize the importance of tertiary interactions in stabilizing marginally stable elements of structure during folding, and provide strong support for the notion that stable, cooperatively folded subdomains are the important determinants of protein folding.

EXPERIMENTAL PROCEDURES

The recombinant peptides GCN4-p3 and GCN4-p3C were expressed in *Escherichia coli* strain BL21 (DE3) pLysS using the T7 system (Studier et al., 1990). A synthetic gene encoding GCN4-p3 was prepared with optimal codon usage for *E. coli* (Grosjean & Fiers, 1982) and ligated into the *XbaI/EcoRI* site of pET3a (Studier et al., 1990) as a precursor to an

---

[*] To whom correspondence should be addressed.

[⊗] Abstract published in *Advance ACS Abstracts*, May 1, 1994.

[1] Abbreviations: 1D, one-dimensional; 2D, two-dimensional; $[\theta]_{222}$, molar ellipticity at 222 nm; Ac, acetyl; bZIP, basic-region leucine zipper; CD, circular dichroism; HPLC, high-performance liquid chromatography; HSMQC, ^{1}H detected heteronuclear single and multiple quantum correlation spectroscopy; LB, Luria–Bertani; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, 2D NOE spectroscopy; GCN4-p1$_{x-y}$, synthetic peptide corresponding to residues x to y of GCN4-p1; ppm, parts per million; T_m, midpoint of thermal denaturation.
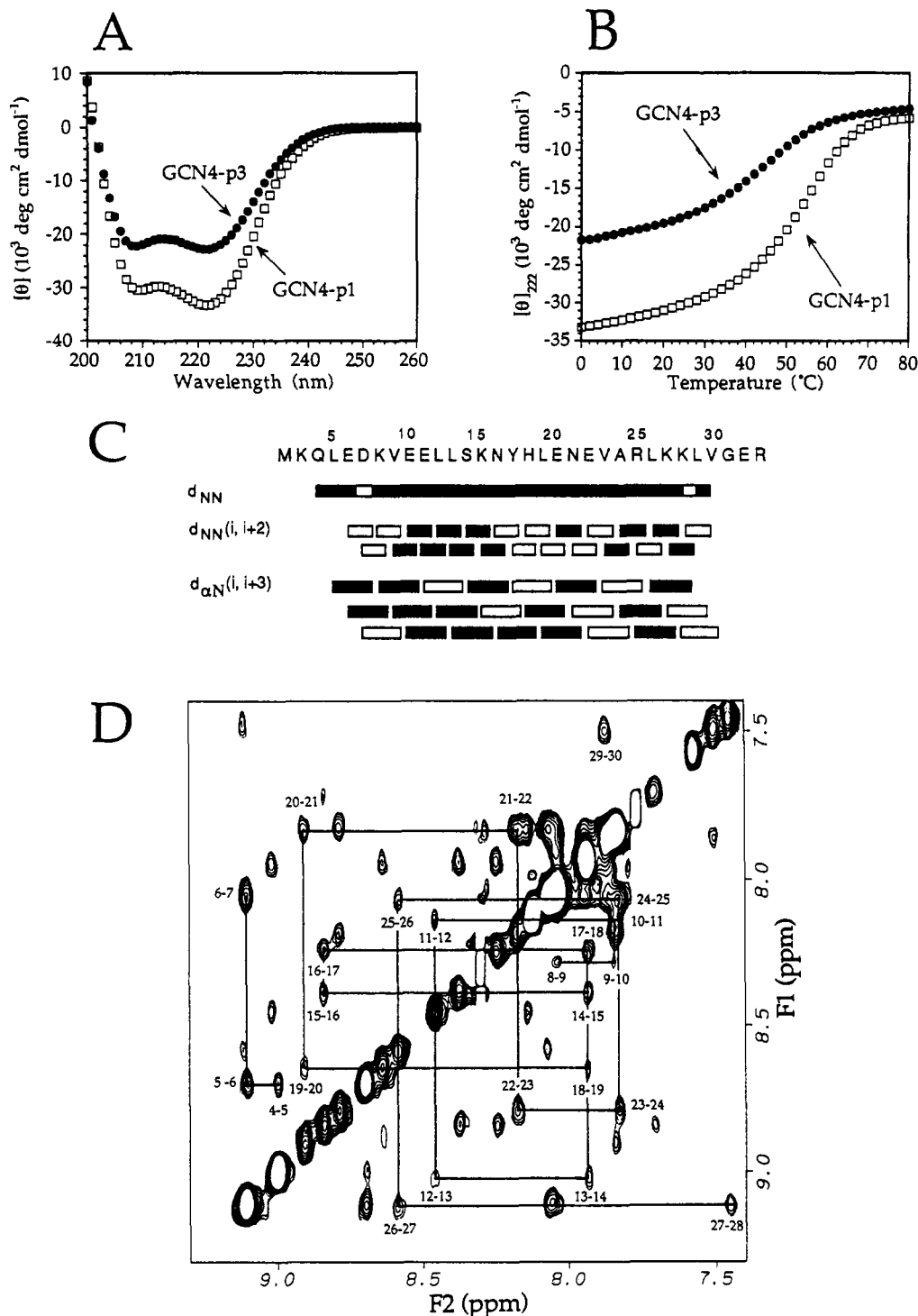
A

B

FIGURE 1: (A) Schematic representation of a parallel, two-stranded coiled coil (O'Shea et al., 1991). A side view is shown. For simplicity, the supercoiling of the helices is not depicted. The sequences of coiled coils are characterized by a heptad repeat of seven amino acid residues, denoted a to g, with a 4,3 repeat of predominantly hydrophobic residues at positions a and d (Hodges et al., 1972; McLachlan & Stewart, 1975). The hydrophobic interface between the two α-helices is formed by residues at positions a, d, e, and g. Prime (′) refers to positions from the other helix. One set of packing interactions consists of side chains from positions a, a′, g, and g′, whereas the second set consists of side chains from positions d, d′, e, and e′ (O'Shea et al., 1991). Residues at positions e and g pack against positions a and d, as well as participating in interhelical electrostatic interactions, which are indicated with bridges (position e to g′ of the preceding heptad). pK_a measurements in a designed heterodimeric coiled coil suggest, however, that interhelical salt bridges do not contribute significantly to the stability of a two-stranded coiled coil (O'Shea et al., 1993). (B) Helical wheel representation of GCN4-p3, which corresponds to residues Met 2–Arg 33 of GCN4-p1 (Table 1) and has an unblocked N-terminus. GCN4-p3 differs slightly from GCN4-p1 (O'Shea et al., 1989), which has a neutral (acetylated) N-terminus and begins at Arg 1, rather than Met 2. The view is from the N-terminus.

expression plasmid (pJH370) for a λ repressor–GCN4 leucine zipper fusion protein (Hu et al., 1990). The gene was then subcloned into the *XbaI/EcoRI* site of the high copy expression vector pAED4 (Doering, 1992) using standard procedures (Sambrook et al., 1989) and called p4LZ. The plasmid for GCN4-p3C (called p4LZGGC) was prepared by oligonucleotide-directed mutagenesis of p4LZ (Zoller & Smith, 1982). Sequences were confirmed by DNA sequencing (Sanger et al., 1977).

Cells were grown directly from a single colony in LB media containing 100 mg/L ampicillin and 25 mg/L chloramphenicol to an optical density of 0.6 at 600 nm and were induced with isopropyl β-D-thiogalactopyranoside (final concentration, 0.5 mM). After 3 h, cells were harvested by centrifugation and lysed by freezing followed by sonication in 50 mM Tris-base and 1 mM EDTA, pH 8.0. The pH of the lysate was reduced to 2 with HCl to precipitate impurities, and the resulting mixture was centrifuged. The soluble fraction was dialyzed against 10 mM sodium acetate, pH 4.0, and then loaded onto CM-Sepharose CL-6B. Contaminants were eluted in 10 mM sodium acetate, pH 4.0, and the GCN4 peptide was then eluted in the same buffer containing 1 M NaCl. Final purification was by reverse-phase HPLC on a Vydac C_{18} column using a linear water/acetonitrile gradient containing 0.1% trifluo-

roacetic acid. The yields of GCN4-p3 and GCN4-p3C from LB media were 60–90 and 40–60 mg/L, respectively. Uniformly ^{15}N-labeled GCN4-p3C was prepared in the same way except cells were grown in M9T minimal media (McIntosh et al., 1987) containing 1 g/L ($^{15}NH_4)_2SO_4$ to yield 25 mg/L HPLC-purified ^{15}N-labeled GCN4-p3C. All other peptides were synthesized using FMOC or t-Boc solid-phase synthesis and purified as described previously (O'Shea et al., 1989, 1993). The identity of each peptide was confirmed by laser desorption mass spectrometry on a Finnigan LASERMAT, and in all cases the expected and observed molecular masses agreed to within 2 Da. The numbering of residues of the peptides follows the GCN4-p1 sequence (O'Shea et al., 1989).

CD spectra were acquired on Aviv 60DS or 62DS spectrometers. Samples were prepared in 50 mM sodium phosphate and 150 mM NaCl. Peptide concentrations were determined by tyrosine absorbance in 6 M guanidine hydrochloride assuming an extinction coefficient at 276 nm of 1500 M^{-1} cm^{-1} (Edelhoch, 1967). Helix content was estimated from $[\theta]_{222}$ by assuming that a value of -33×10^3 deg cm^2 $dmol^{-1}$ corresponds to a helix content of 100% for a 33-residue coiled coil (O'Shea et al., 1989) and by correcting for the length dependence of $[\theta]_{222}$ (Chen et al., 1974). Thermal stability was determined at peptide concentrations (monomer) of 35 μM or 1 mM by monitoring the change in $[\theta]_{222}$ as a function of temperature. The temperature was increased in steps of 1 °C with an equilibration time of 120 s and a data collection time of 30 s. The T_m was determined from the minima of the first derivative of $[\theta]_{222}$ with T^{-1}, where T is in K (Cantor & Schimmel, 1980). All thermal melts were reversible (the folding and unfolding curves were superimposable, with ≥95% of the starting signal regained on cooling).

Apparent molecular masses were determined by sedimentation equilibrium on a Beckman XL–A ultracentrifuge at 0 °C. Samples were dialyzed against 50 mM sodium phosphate and 150 mM NaCl, pH 7.0, for at least 12 h. Typically, three samples of total peptide concentration of 166, 332, and 553 μM were run at rotor speeds of 35 000 or 40 000 rpm, and the absorbance at 276 nm was monitored. One peptide (GCN4-$p1_{11-33}$) was not significantly folded at these concentrations; in this case, a concentration of 1 mM was used and the absorbance at 244 nm was monitored. Data were fit to an ideal model plot of ln(absorbance) versus radial distance squared. Significant deviations of residuals, indicative of deviation from this ideal model, were not observed except for the partially folded peptide GCN4-$p1_{11-33}$. Partial molar volumes and solvent densities were calculated as described by Laue et al. (1992).

NMR spectroscopy was performed on a Bruker AMX spectrometer operating at 500.1 MHz for ^1H. Samples were typically 2 to 5 mM peptide in 50 mM sodium phosphate and 150 mM NaCl, pH 6.5, containing 10% D_2O and internally referenced to zero ppm with (trimethylsilyl)propionic acid. NOESY (Macura et al., 1981) and 2D ^1H–^{15}N HSMQC-NOESY (Gronenborn et al., 1989; Zuiderweg, 1990) data sets consisted of 512 t_1 increments defined by 1024 complex points and 96 transients collected using time-proportional phase incrementation (Marion & Wüthrich, 1983). Data were acquired at 5 °C using spectral widths of 6024.1 Hz in the ^1H dimension and 1250 Hz in the ^{15}N dimension and a recycle delay of 1.5 s. The NOESY mixing time was 150 ms. ^1Hs were decoupled in heteronuclear experiments using WALTZ-16 (Shaka et al., 1983). The water resonance was suppressed using a selective 1–1 echo sequence to avoid saturation transfer to the exchangeable amide NH protons

FIGURE 2: (A) CD spectra of GCN4-p1 (□) and GCN4-p3 (●) at 0 °C and pH 7. The minima at 208 and 222 nm indicate that both peptides are helical. However, GCN4-p3 is only approximately two-thirds helical by CD, in contrast to GCN4-p1, which is essentially fully helical. (B) Temperature dependence of $[\theta]_{222}$ for GCN4-p3 (●) and GCN4-p1 (□) at pH 7. Despite GCN4-p3 being only about two-thirds helical by CD, the temperature dependence of its CD signal exhibits a folded baseline (where $[\theta]_{222}$ shows little change with temperature) prior to the major unfolding transition region. (C) Continuous sets of $d_{NN}(i, i + 1)$, $d_{NN}(i, i + 2)$, and $d_{\alpha N}$ $(i, i + 3)$ NOEs. These NOEs indicate (Wüthrich, 1986) that GCN4-p3 is helical from approximately Leu 5 to at least Leu 29 (see text). A solid bar indicates that the NOE was observed unambiguously, whereas an open bar indicates that a NOESY cross peak was observed but could not be assigned unambiguously. (D) The amide NH region of the NOESY spectrum of GCN4-p3. $d_{NN}(i, i + 1)$ NOEs are observed from Gln 4 to Leu 29. For clarity, the contour plot has been made at a high level, so that the $d_{NN}(i, i + 2)$ cross peaks are not visible. The amide proton chemical shift degeneracy prevented unambiguous observation of the d_{NN} (28, 29) and the $d_{NN}(i, i + 1)$ NOEs for residues 30–33 (see also Figure 3).

(Sklenar & Bax, 1987). Data were resolution enhanced using a Gaussian function in t_2 and a 50° shifted squared sine bell in t_1 and zero-filled prior to Fourier transformation to give final digital resolutions of 2.9 and 1.2 Hz/point in the ^{1}H and ^{15}N dimensions, respectively. Spectra were assigned using a main chain directed approach (Englander & Wand, 1987)

using $d_{NN}(i, i + 1)$ NOEs and the Tyr 17 spin system as a unique reference point.

RESULTS AND DISCUSSION

GCN4-p3 (Figure 1B) corresponds to residues 2–33 of GCN4-p1 (Table 1) and has an unblocked N-terminus. It

Table 1: Circular Dichroism and Sedimentation Equilibrium Data

peptide	sequence	$[\theta]_{222}{}^a$ (×10³ deg cm² dmol⁻¹)		$T_m{}^b$ (°C)		molecular mass^c (Da)
		35 μM	1 mM	35 μM	1 mM	
GCN4-p1	AcRMKQLEDKVEELLSKNYHLENEVARLKKLVGER	−33.2	−33.3	56	68	7900 (8080)^d
GCN4-p3	MKQLEDKVEELLSKNYHLENEVARLKKLVGER	−22.0	−24.5	47	65	7790 (7681)
N-terminal modifications						
GCN4-p1ΔAc	RMKQLEDKVEELLSKNYHLENEVARLKKLVGER	−27.7	−30.0	49	67	7620 (7995)
GCN4-p3Ac	AcMKQLEDKVEELLSKNYHLENEVARLKKLVGER	−32.3	−33.3	51	69	7600 (7769)
N-terminal truncations						
GCN4-p1$_{4-33}$	AcQLEDKVEELLSKNYHLENEVARLKKLVGER	−31.1	−33.2	43	63	7120 (7246)
GCN4-p1$_{8-33}$	AcKVEELLSKNYHLENEVARLKKLVGER	−24.0	−33.2	35	50	5710 (6275)
GCN4-p1$_{11-33}$	AcEELLSKNYHLENEVARLKKLVGER	−8.8	−14.7^e	<0	<0	3600 (3900)^e
C-terminal truncations						
GCN4-p1$_{8-30}$	AcKVEELLSKNYHLENEVARLKKLV	−20.7	−27.8	22	43	5160 (5588)
GCN4-p1$_{8-27}$	AcKVEELLSKNYHLENEVARLK	−6.5	−7.0	<0	<0	

^{a} $[\theta]_{222}$ was measured at 0 °C at peptide concentrations (monomer) of 35 μM and 1 mM. GCN4-p1ΔAc, GCN4-p1$_{4-33}$, GCN4-p1$_{8-33}$, and GCN4-p1$_{8-30}$ are not fully helical at 35 μM, but are >90% helical at millimolar concentrations as judged by $[\theta]_{222}$ and continuous sets of $d_{NN}(i, i + 1)$ NOEs (data shown for GCN4-p1$_{8-30}$ in Figure 5C). ^{b} The T_m was determined at peptide concentrations (monomer) of 35 μM and 1 mM and is accurate to ±1 °C. ^{c} Apparent molecular masses were determined at 0 °C. The expected value for a dimer is enclosed in parentheses. ^{d} Data from O'Shea et al. (1989). ^{e} The concentration dependence of $[\theta]_{222}$ and the apparent molecular mass for a 1 mM sample of GCN4-p1$_{11-33}$ suggest that GCN4-p1$_{11-33}$ self-associates. The expected molecular mass indicated in parentheses is for a 1 mM solution of peptide, assuming an unfolded monomer–folded dimer equilibrium and a helix content of 45% estimated from the value of $[\theta]_{222}$.
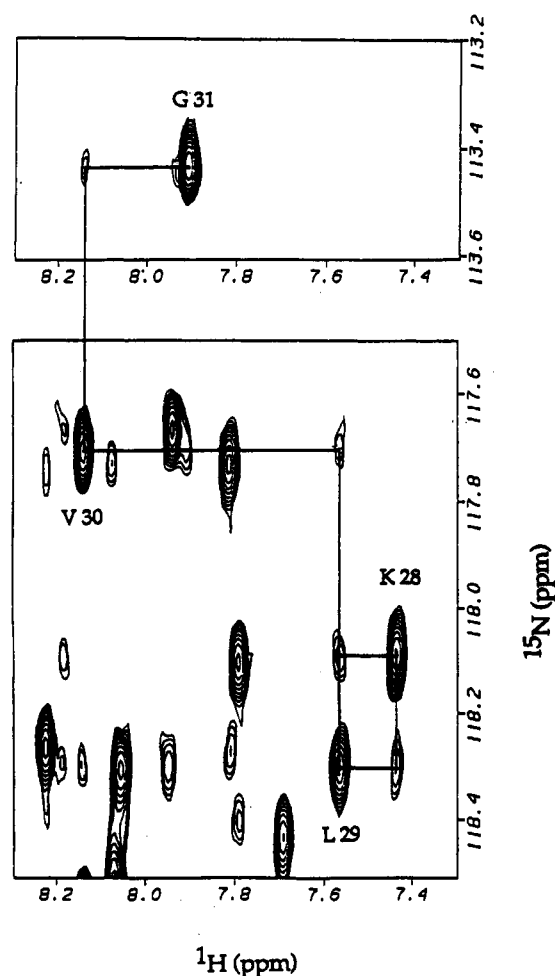
FIGURE 3: Region of the 2D ^{15}N–^{1}H HSMQC-NOESY spectrum of GCN4-p3C. In addition to the $d_{NN}(i, i + 1)$ NOEs observed in the NOESY spectrum of GCN4-p3 (Figure 2), $d_{NN}(i, i + 1)$ NOEs are observed for residues 29–31. NOEs are not observed for residues 31–33.

FIGURE 4: 1D NMR spectra of GCN4-p1$_{8-33}$, GCN4-p1$_{11-33}$, and GCN4-p1$_{8-27}$. The amide NH and aromatic region of the 1D NMR spectrum of GCN4-p1$_{8-33}$ is typical of a folded, two-stranded coiled coil (Oas et al., 1990). In contrast, the amide NH resonances in the 1D NMR spectrum of GCN4-p1$_{11-33}$ are broad, perhaps as a result of chemical exchange on the intermediate time scale between multiple conformations or between the unfolded and folded states. The spectrum of GCN4-p1$_{8-27}$ indicates that the peptide is predominantly unfolded, with the amide proton chemical shifts close to those expected for an unstructured peptide (Wüthrich, 1986).

differs slightly from GCN4-p1 (O'Shea et al., 1989), which has a neutral (acetylated) N-terminus and begins at Arg 1, rather than Met 2. The X-ray structure of GCN4-p1 is helical for residues Arg 1–Val 30 and is undefined for residues 31–33 (O'Shea et al., 1991). NMR studies indicate that GCN4-p1 in solution is helical from Arg 1 to at least Glu 32 (chemical shift degeneracy prevented the observation of NOEs to Arg 33; Oas et al., 1990). GCN4-p3 also differs slightly from GCN4-p2N (Goodman & Kim, 1991), which has the same sequence but with a neutral (amidated) C-terminus and a N-terminal Ac-Cys-Gly-Gly to allow interchain disulfide formation. N-terminally disulfide-bonded GCN4-p2N is helical as judged by NMR from Met 2 to Arg 33 and has a T_m of 83 °C (Goodman & Kim, 1991).

CD indicates that GCN4-p3 exists in a helical conformation (Figure 2A) which sedimentation equilibrium indicates is a dimer (Table 1). However, GCN4-p3 is only approximately two-thirds helical at 0 °C and pH 7.0 (Figure 2A and Table 1). In contrast, both GCN4-p1 (Figure 2A and Table 1; O'Shea et al., 1989) and the N-terminally disulfide-bonded GCN4-p2N (Goodman & Kim, 1991) are essentially 100% helical under identical conditions.
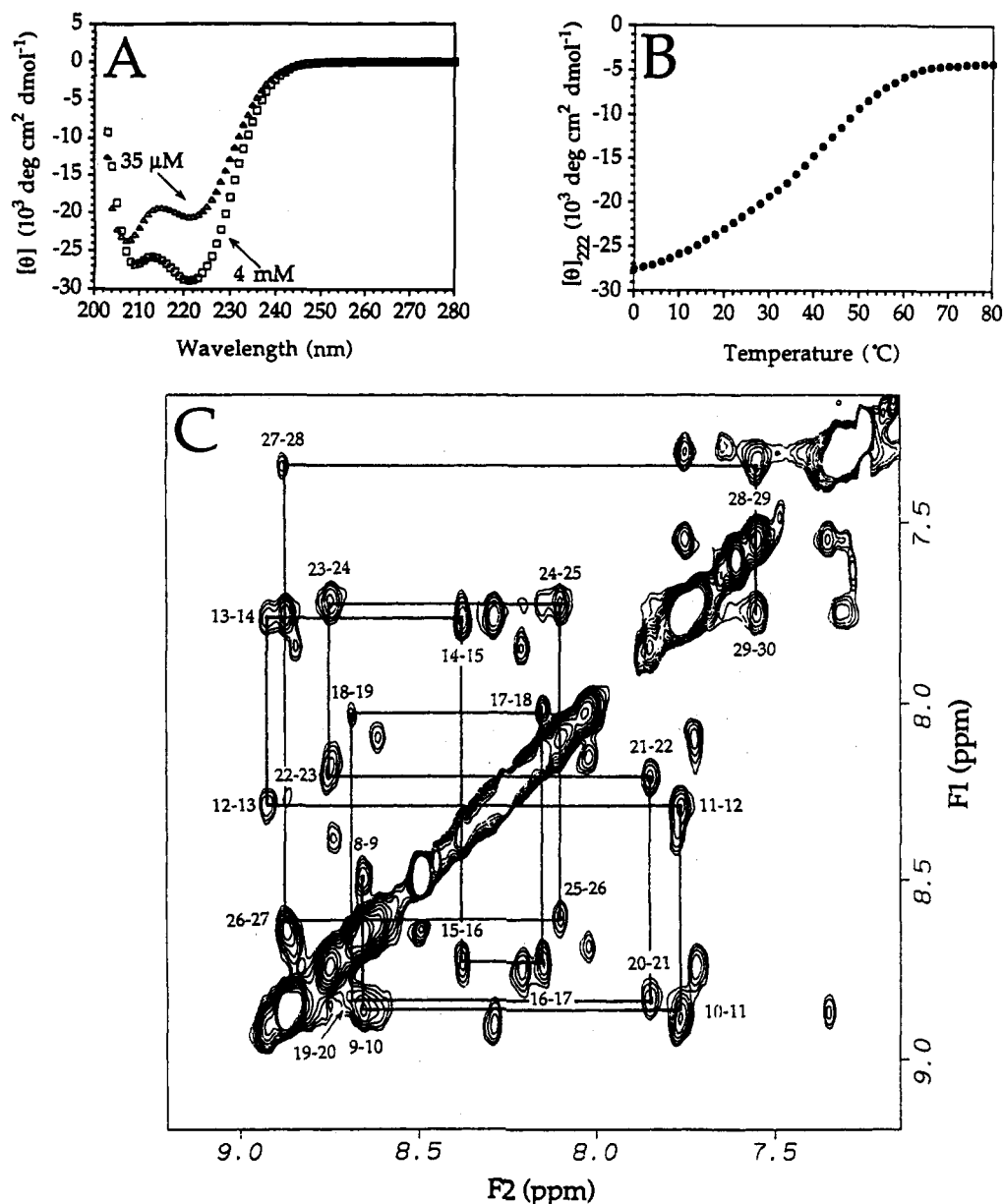
FIGURE 5: (A) CD spectra of GCN4-p1$_{8-30}$ at 0 °C and pH 7. These spectra indicate that the peptide is helical and that the helix content is concentration dependent, consistent with self-association. At millimolar concentrations the peptide is greater than 90% helical at low temperature. At 0 °C, GCN4-p1$_{8-30}$ is a dimer (Table 1). (B) Thermal melt of 1 mM GCN4-p1$_{8-30}$. The peptide exhibits a cooperative unfolding transition. (C) Amide NH region of the NOESY spectrum of GCN4-p1$_{8-30}$. $d_{NN}(i, i + 1)$ NOEs that are indicative of helical structure are observed from Lys 8 to Val 30. For clarity, the contour plot has been made at a high level, so that the $d_{NN}(i, i + 2)$ cross peaks are not visible.

Despite incomplete helix formation, GCN4-p3 exhibits a "folded" baseline during thermal unfolding; i.e., prior to the major unfolding step, there is a region where little helical signal is lost with increasing temperature (Figure 2B). This suggests that the CD spectrum of GCN4-p3 at low temperatures does not result from a weighted average of completely folded and completely unfolded peptide molecules. Instead, the data suggest that GCN4-p3 contains both an unfolded region and a separate folded region that is stable to thermal denaturation.

The helices of the GCN4 coiled coil are parallel (O'Shea et al., 1989, 1991). To confirm that the helices of GCN4-p3 are also parallel, the concentration dependence of the T_m of a disulfide-bonded variant of GCN4-p3 was studied. GCN4-p3C has the same sequence as GCN4-p3 except for a C-terminal Gly-Gly-Cys that allows formation of an interchain disulfide bond. If the two helices are disulfide bonded in the favorable orientation, the T_m will be independent of concentration. Conversely, it has been shown that if the GCN4

helices are disulfide bonded in an unfavorable orientation (i.e., antiparallel), then higher order oligomers form, reflected in a concentration dependence of T_m (O'Shea et al., 1989). The T_m of disulfide-bonded GCN4-p3C (77 ± 1 °C at pH 7) is independent of concentration over the range 50 μM to 2 mM, confirming that the helices of GCN4-p3 are parallel.

NMR indicates that GCN4-p3 is locally unfolded at the N-terminus, although it is not possible to define precisely the residue at which the helix begins. Medium-range $d_{NN}(i, i + 2)$ and $d_{\alpha N}(i, i + 3)$ NOEs that are indicative of α-helical structure are observed for residues 6-29 and 5-29, respectively (Figure 2C). Sequential $d_{NN}(i, i + 1)$ NOEs are observed for residues 4-30 (Figure 2D); however, these NOEs are observed in unfolded peptides as well as α-helices, and so, given the absence of medium-range NOEs to Gln 4, it cannot be certain that the helix begins at Gln 4. Chemical shift degeneracy precluded observation of helical NOEs for residues 30-33, although unambiguous $d_{NN}(i, i + 1)$ NOEs are observed for these residues of GCN4-p2N (Goodman & Kim, 1991)

and for residues 28–31 of GCN4-p3C (Figure 3). We conclude that GCN4-p3 is folded as a parallel, two-stranded coiled coil from approximately Leu 5 or Glu 6 to at least Leu 29 and that Met 2, Lys 3, and possibly Gln 4 are predominantly unfolded.

These results suggest that N-terminal deletions of residues from the GCN4 coiled coil sequence may be possible without loss of the coiled coil structure. Truncations were made at one residue before each of the hydrophobic interface residues at positions a and d (i.e., the first residue in the peptide was either at position g or c; Figure 1), since GCN4-p1ΔAc (which begins at Arg 1 and has a free N-terminus) is significantly more helical than GCN4-p3 (Table 1). It is likely that the loss of a favorable g–a packing interaction (Figure 1A; O'Shea et al., 1991; Greenfield et al., 1994) and a main-chain hydrogen bond (Arg 1-Leu 5) contributes to the lower helix content of GCN4-p3 compared to GCN4-p1ΔAc. The N-termini of the truncated peptides were synthetically neutralized (acetylated), since acetylation of the N-terminus of GCN4-p3 results in a fully folded, two-stranded coiled coil (GCN4-p3Ac in Table 1).[2] A significant increase in coiled coil stability upon N-terminal acetylation has also been observed for peptides corresponding to the N-terminus of tropomyosin (Greenfield et al., 1994).

Peptides beginning at residues Gln 4 (GCN4-p1$_{4-33}$) and Lys 8 (GCN4-p1$_{8-33}$) fold as stable, two-stranded coiled coils (Table 1 and Figure 4), although the helix content does not approach 100% by CD until the concentration of the peptide is increased to the millimolar level (Table 1). The next truncation was made at Glu 11. This 23-residue peptide (GCN4-p1$_{11-33}$) is largely unfolded by CD (Table 1) and NMR (Figure 4). Thus, N-terminal truncations of the GCN4 coiled coil beyond Glu 10 do not form stable structures.

C-terminal truncations were made from the sequence of GCN4-p1$_{8-33}$, which is the shortest of the N-terminal truncations that folds as a stable coiled coil. GCN4-p1$_{8-30}$ corresponds to a deletion of residues 31–33 and folds cooperatively as a two-stranded coiled coil (Figure 5 and Table 1). Disulfide-bonded GCN4-p1$_{8-30}$N (a variant of GCN4-p1$_{8-30}$ which has an N-terminal Ac-Cys-Gly-Gly to allow formation of an interchain disulfide bond) has a T_m (72 ± 1 °C at pH 7) that is independent of concentration over the range 35 μM to 2 mM, indicating that the helices are parallel (see above). A 20-residue peptide, GCN4-p1$_{8-27}$, corresponding to a further deletion of residues 28–30, is not folded by CD (Table 1) or NMR (Figure 4). Thus, the shortest peptide derived here from the GCN4 sequence that folds as a stable, two-stranded, parallel coiled coil with a cooperative thermal unfolding transition is the 23-residue peptide GCN4-p1$_{8-30}$.

This is the shortest isolated coiled coil to date [cf. Lau et al. (1984)], although shorter coiled coils can exist in the context of an entire domain (Marmorstein et al., 1992).

GCN4-p1$_{8-30}$, along with the longer peptides, folds as a stable, parallel, two-stranded coiled coil. This indicates that the GCN4 leucine zipper contains stable subdomains (i.e., cooperatively folded units of structure larger than an isolated helix, but smaller than an entire domain; Rose, 1979; Richardson, 1981; Oas & Kim, 1988; Staley & Kim, 1990; Wu et al., 1993). It is striking that GCN4-p1$_{11-33}$ is predominantly unfolded, even though it is the same length as GCN4-p1$_{8-30}$ and differs only at three terminal residues. In addition, deletion of the three C-terminal residues from GCN4-p1$_{8-30}$ results in an unfolded peptide. Isolated helices are often only marginally stable (Scholtz & Baldwin, 1992), including those of a peptide corresponding to the bZIP region of GCN4 (Thompson et al., 1993). Our results indicate, therefore, that the formation of specific tertiary interactions gives rise to the cooperative folding of GCN4-p1$_{8-30}$. Similar conclusions have been reached from studies of peptide fragments of BPTI, cytochrome *c*, and myoglobin (Oas & Kim, 1988; Staley & Kim, 1990; Wu et al., 1993; Shin et al., 1993).

The framework model for protein folding emphasizes the formation of secondary structure (Kim & Baldwin, 1982). Although the intrinsic propensities of the amino acids to form α-helices [for reviews, see Scholtz and Baldwin (1992) and Fersht and Serrano (1993)] or β-sheets (Kim & Berg, 1993; Minor & Kim, 1994; Smith et al., 1994) appear to be important determinants of protein folding, our results suggest that tertiary packing interactions (even between two short α-helices) can be energetically more important in establishing a cooperatively folded structure. This conclusion strongly reinforces the notion that protein folding can be understood, in large part, in terms of the formation of cooperatively folded subdomains, in which elements of secondary structure are stabilized by native-like tertiary interactions.

ACKNOWLEDGMENT

REFERENCES

Alber, T. (1992) *Curr. Opin. Genet. Dev. 2*, 205–210.

Banner, D. W., Kokkinidis, M., & Tsernoglou, D. (1987) *J. Mol. Biol. 196*, 657–675.

Cantor, C. R., & Schimmel, P. R. (1980) *Biophysical Chemistry*, Freeman, New York.

Chen, Y., Yang, J. T., & Chau, K. H. (1974) *Biochemistry 13*, 3350–3359.

Cohen, C., & Parry, D. A. D. (1990) *Proteins 7*, 1–15.

Creighton, T. E. (1993) *Proteins*, Freeman, New York.

Crick, F. H. C. (1953) *Acta Crystallogr. 6*, 689–697.

Doering, D. S. (1992) Ph.D. Thesis, Massachusetts Institute of Technology.

Dyson, H. J., & Wright, P. E. (1993) *Curr. Opin. Struct. Biol. 3*, 60–65.

Edelhoch, H. (1967) *Biochemistry 6*, 1948–1954.

Ellenberger, T. (1994) *Curr. Opin. Struct. Biol. 4*, 12–21.

Englander, S. W., & Wand, A. J. (1987) *Biochemistry 26*, 5953–5958.

Fersht, A. R., & Serrano, L. (1993) *Curr. Opin. Struct. Biol. 3*, 75–83.

---

[2] GCN4-p3 and GCN4-p1ΔAc are less stable at acidic than at neutral pH (the T_m of 35 μM GCN4-p3 decreases from 47 °C at pH 7 to 25 °C at pH 3, and the T_m of 35 μM GCN4-p1ΔAc decreases from 49 °C at pH 7 to 25 °C at pH 3). In contrast, the stabilities of the corresponding peptides with neutral (acetylated) N-termini are relatively insensitive to pH (the T_m of 35 μM GCN4-p1 is 56 °C at pH 7 and 52 °C at pH 3, and the T_m of 35 μM GCN4-p3Ac is 52 °C at pH 7 and 51 °C at pH 3). The pK_a of an α-amino group lies between 7 and 8 (Creighton, 1993), so the N-terminus will be more charged at acidic than at neutral pH. This might result in an unfavorable interaction either between the two like-charged termini or between the charged N-terminus and the helix dipole [cf. Shoemaker et al. (1987); see, however, Greenfield et al. (1994)]. Although at pH 10 the T_m of 35 μM GCN4-p3 (47 °C) is comparable to that of 35 μM GCN4-p1 (50 °C), GCN4-p3 is only 76% helical, whereas GCN4-p1 is essentially 100% helical. This suggests that, in addition to the potential for destabilizing charge effects at low pH, loss of other favorable interactions (e.g., the g–a packing interaction or the Arg 1·Leu 5 main-chain hydrogen bond) may contribute to the relatively low helix content of GCN4-p3.

Goodman, E. M., & Kim, P. S. (1991) *Biochemistry 30*, 11615–11620.

Greenfield, N. J., & Hitchcock-DeGregori, S. E. (1993) *Protein Sci. 2*, 1263–1273.

Greenfield, N. J., Stafford, W. F., & Hitchcock-DeGregori, S. E. (1994) *Protein Sci. 3*, 402–410.

Gronenborn, A. M., Bax, A., Wingfield, P. T., & Clore, G. M. (1989) *FEBS Lett. 243*, 93–98.

Grosjean, H., & Fiers, W. (1982) *Gene 18*, 199–209.

Harbury, P. B., Zhang, T., Kim, P. S., & Alber, T. (1993) *Science 262*, 1401–1407.

Hodges, R. S., Sodek, J., Smillie, L. B., & Jurasek, L. (1972) *Cold Spring Harbor Symp. Quant. Biol. 37*, 299–310.

Holtzer, M. E., & Holtzer, A. (1992) *Biopolymers 32*, 1589–1591.

Hu, J. C., & Sauer, R. T. (1992) *Nucleic Acids Mol. Biol. 6*, 82–101.

Hu, J. C., O'Shea, E. K., Kim, P. S., & Sauer, R. T. (1990) *Science 250*, 1400–1403.

Ishii, Y., Hitchcock-DeGregori, S., Mabuchi, K., & Lehrer, S. S. (1992) *Protein Sci. 1*, 1319–1325.

Kim, C. A., & Berg, J. M. (1993) *Nature 362*, 267–270.

Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459–489.

Landschulz, W. H., Johnson, P. F., & McKnight, S. L. (1988) *Science 240*, 1759–1764.

Lau, S. Y. M., Taneja, A. K., & Hodges, R. S. (1984) *J. Biol. Chem. 259*, 13253–13261.

Laue, T. M., Shah, B. D., Ridgeway, T. M., & Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., & Horton, J. C., Eds.) pp 90–125, The Royal Society of Chemistry, Cambridge.

Lehrer, S. S. (1978) *J. Mol. Biol. 118*, 209–226.

Lovejoy, B., Choe, S., Cascio, D., McRorie, D. K., DeGrado, W. F., & Eisenberg, D. (1993) *Science 259*, 1288–1293.

Macura, S., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259–281.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Marmorstein, R., Carey, M., Ptashne, M., & Harrison, S. C. (1992) *Nature 356*, 408–414.

McIntosh, L. P., Griffey, R. H., Muchmore, D. C., Nielson, C. P., Redfield, A. G., & Dahlquist, F. W. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1244–1248.

McLachlan, A. D., & Stewart, M. (1975) *J. Mol. Biol. 98*, 293–304.

Minor, D. L., & Kim, P. S. (1994) *Nature 367*, 660–663.

Oas, T. G., & Kim, P. S. (1988) *Nature 336*, 42–48.

Oas, T. G., McIntosh, L. P., O'Shea, E. K., Dahlquist, F. W., & Kim, P. S. (1990) *Biochemistry 29*, 2891–2894.

O'Shea, E. K., Rutkowski, R., & Kim, P. S. (1989) *Science 243*, 538–542.

O'Shea, E. K., Klemm, J. D., Kim, P. S., & Alber, T. A. (1991) *Science 254*, 539–544.

O'Shea, E. K., Lumb, K. J., & Kim, P. S. (1993) *Curr. Biol. 3*, 658–667.

Pauling, L. P., & Corey, R. B. (1953) *Nature 171*, 59–61.

Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.

Rose, G. D. (1979) *J. Mol. Biol. 134*, 447–470.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Scholtz, J. M., & Baldwin, R. L. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 95–118.

Shaka, A. J., Keeler, J., & Freeman, R. (1983) *J. Magn. Reson. 53*, 313–340.

Shin, H., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J., & Wright, P. E. (1993) *Biochemistry 32*, 6356–6364.

Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987) *Nature 326*, 563–567.

Sklenar, V., & Bax, A. (1987) *J. Magn. Reson. 74*, 469–479.

Smith, C. K., Withka, J. M., & Regan, L. (1994) *Biochemistry* (in press).

Staley, J. P., & Kim, P. S. (1990) *Nature 344*, 685–688.

Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60–89.

Thompson, K. S., Vinson, C. R., & Freire, E. (1993) *Biochemistry 32*, 5491–5496.

Wu, L. C., Laub, P. B., Elove, G. A., Carey, J., & Roder, H. (1993) *Biochemistry 32*, 10271–10276.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Yan, Y., Winograd, E., Viel, A., Cronin, T., Harrison, S. C., & Branton, D. (1993) *Science 262*, 2027–2030.

Zoller, M. J., & Smith, M. (1982) *Nucleic Acids Res. 10*, 6487–6500.

Zuiderweg, E. R. P. (1990) *J. Magn. Reson. 86*, 346–357.