Probing Backbone Hydrogen Bonds in the Hydrophobic Core of GCN4[†]

John W. Blankenship, Rema Balambika, and Philip E. Dawson*

Departments of Cell Biology and Chemistry and the Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, CVN-6, La Jolla, California 92037

Received September 16, 2002; Revised Manuscript Received October 29, 2002

ABSTRACT: Backbone amide hydrogen bonds play a central role in protein secondary and tertiary structure. Previous studies have shown that substitution of a backbone ester (−COO−) in place of a backbone amide (−CONH−) can selectively destabilize backbone hydrogen bonds in a protein while maintaining a similar conformation to the native backbone structure. The majority of these studies have focused on backbone substitutions that were accessible to solvent. The GCN4 coiled coil domain is an example of a stable α-helical dimer that possesses a well-packed hydrophobic core. Amino acids in the a and d positions of the GCN4 helix, which pack the hydrophobic core, were replaced with the corresponding α-hydroxy acids in the context of a chemoselectively ligated heterodimer. While the overall structure and oligomerization state of the heterodimer were maintained, the overall destabilization of the ester analogues was greater (average $\Delta\Delta G$ of 3+ kcal mol^{-1}) and more variable than previous studies. Since burial of the more hydrophobic ester should stabilize the backbone and reduce the $\Delta\Delta G$, the increased destabilization must come from another source. However, the observed destabilization is correlated with the protection factors for individual amide hydrogens from previous hydrogen exchange experiments. Therefore, our results suggest that backbone engineering through ester substitution is a useful approach for probing the relative strength of backbone hydrogen bonds.

Secondary structure has a key role in maintaining protein stability and tertiary structure, and mutations that disrupt secondary structural elements usually significantly destabilize proteins. The bulk of secondary structural interactions are mediated by intrabackbone amide hydrogen bonds (*1*), either between turns of an α-helix or between *β*-strands within a *β*-sheet. The contribution that backbone hydrogen bonds make toward stabilizing the overall tertiary structure—and concomitantly, their role in the protein folding process—has been under debate for some time (*2*−*8*).

Leucine zippers, typical coiled coil domains (*9*−*11*), are well-studied and highly α-helical small proteins. Leucine zippers have two (or more) α-helices wound around each other in a left-handed superhelix. They are known to contain several heptad repeats (designated by the positions abcdefg) (*12*, *13*), where the positions a and d are occupied by hydrophobic amino acids (such as Val or Leu) that interdigitate, forming a hydrophobic core between the helices. Positions e and g are typically charged residues, and positions b, c, and f are often polar or charged residues (*14*). The coiled coil domain from the bZIP repressor, GCN4, is the prototype of this fold and has been extensively characterized by protein engineering and biophysical analysis. The folding of the truncated GCN4 domain (which associates as a dimer) follows a two-state equilibrium, as followed by kinetic or equilibrium measurements (*9*, *15*, *16*). Engineered versions of GCN4 covalently linked by a flexible disulfide linker (either a N-terminal CGG− or a C-terminal −GGC extension) were found to retain the two-state folding equilibrium and fold as a monomer (*17*). There is some evidence by NMR spectroscopy that the hydrophobic core of GCN4 is highly stable, possessing slow amide proton exchange rates (*18*), as well as elements of helical structure present even in highly concentrated urea conditions (*19*). This stability has been partially attributed to the rigid and highly interdigitated "knobs into holes" packing of the hydrophobic core. As a result, numerous studies probing the stability, folding, and design of α-helices have focused on studying and engineering these interactions.

While the formation of backbone hydrogen bonds is both necessary for the folding of α-helices and for helping to maintain the overall fold of GCN4, no protein engineering studies have probed the contribution of specific hydrogen bonds to maintaining the stability and fold of GCN4. One method for doing so in a highly specific fashion is by site-specifically incorporating an unnatural amino acid analogue, such as an α-hydroxy acid. The resulting ester substitution has been found to effectively mimic the conformational and steric parameters of amide bonds in both model compounds and peptides (*20*−*22*). Introduction of an α-hydroxy acid into the sequence of a protein by either chemical synthesis (*23*−*27*) or recombinant expression (*28*) results in the replacement of an amide bond with an ester bond and the effective removal of at least one hydrogen bond donor, the amide NH. In addition to the loss of the amide NH, measurements by Arnett and co-workers on the enthalpy of ionization for model compounds in HSO_3F and the enthalpy of hydrogen bonding in *p*-fluorophenol established that the hydrogen bonding potential of any compound is tied to its

[†] This work was supported by NIH Grant GM59380-04, the Sloan Research Foundation, and the Louis R. Jabinson Fellowship (J.W.B.).

* Correspondence should be addressed to this author. Phone: (858) 784-7015; fax: (858) 784-7319; e-mail: dawson@scripps.edu.

10.1021/bi026862p CCC: $22.00
Published on Web 12/10/2002

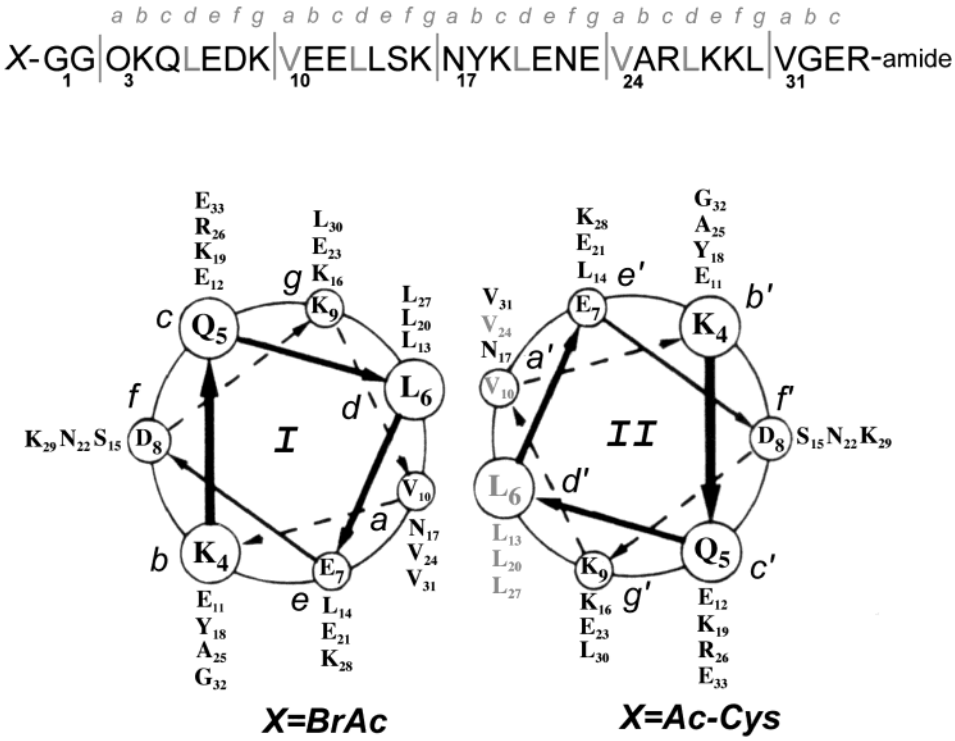

FIGURE 1: Amino acid sequence of the **gcn4-I** and the **gcn4-II** peptides placed over a helical wheel representation of the coiled coil. As described in the Experimental Procedures, O in the sequence represents norleucine; X refers to the sequence at the N-terminus of each a helix; I is bromoacetylated; and II has a *N*-acetylated cysteine residue. All ester mutations are in the second half of the heterodimer (II) and at either the a′ or d′ position.

basicity, and thus, its pK_a (*29*). From this analysis, the lowered carbonyl pK_a of ester model compounds such as ethyl acetate (−4.47) versus the carbonyl pK_a values of amide model compounds such as *N*,*N*-dimethylformamide (−1.5) or *N*,*N*-dimethylacetamide (−0.36) suggests that any hydrogen bond accepted by the ester carbonyl will be destabilized in relation to the same hydrogen bond formed by an amide carbonyl. Consequently, the amide to ester substitution destabilizes not one but two intrabackbone amide hydrogen bonds. This destabilization, $\Delta\Delta G_{amide\rightarrow ester}$, has been repeatedly confirmed by experiment (*23*−*25*, *27*, *28*). The reduction of overall hydrogen bonding potential, along with the change in the ΔG for water−octanol partitioning coefficients between *N*-methylacetamide and methyl acetate of −1.7 kcal mol^{-1} (*28*, *30*) implies that the amide to ester substitution is effectively increasing the hydrophobicity of the backbone. The resulting desolvation of the backbone upon folding has been attributed as a major stabilizing component in $\Delta\Delta G_{amide\rightarrow ester}$, although it is difficult to quantify (*28*). Previous backbone engineering studies on α-helical proteins and peptides have focused on partially solvent-exposed residues near the termini of secondary structure, many of which participate in only one intrabackbone hydrogen bond.

To get a better measure of the effects of an amide to ester substitution on the stability of protein structure, we have introduced a series of sequential single ester substitutions into the hydrophobic core of a GCN4-derived sequence (Figure 1) at the a and d positions. All positions should participate in two intrabackbone hydrogen bonds, and all esters should experience a uniformly desolvated environment within the hydrophobic core. While this highly desolvated environment should stabilize the more hydrophobic ester backbone, and thus reduce $\Delta\Delta G_{amide\rightarrow ester}$, the uniformity of the environment should give more consistent results between substitutions, and any destabilizing interaction should destabilize all residues equally. To follow the effect of a single ester introduction on the folding of the domain, we have used a chemoselective ligation to form a covalently linked heterodimer, in contrast to previous studies that focused on the effects of substitutions on symmetric homodimers.

EXPERIMENTAL PROCEDURES

Raw Materials. Boc-amino acids were obtained from Midwest Biotech (Fishers, IN). HBTU[1] was obtained from Qbiogene (Carlsbad, CA). MBHA resin was obtained from Peninsula Laboratories (Belmont, CA). All solvents (DMF, DCM, and acetonitrile) of high (spectroscopic grade) purity were purchased from Fisher. TFA was obtained from Halocarbon (River Edge, NJ). HF was purchased from Matheson Gas (Cucamonga, CA). DIEA, (*S*)-(+)-2-hydroxy-2-methylbutyric acid (the α-hydroxy acid corresponding to Valine, Hva hereafter) and (*S*)-(−)-2-hydroxyisocaproic acid (the α-hydroxy acid corresponding to Leucine, Hle hereafter) were obtained from Aldrich. Ultrapure guanidine hydrochlo-

[1] Abbreviations: HBTU, 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; DIEA, *N*,*N*,-diisopropylethylamine; TFA, trifluoroacetic acid; SPPS, solid-phase peptide synthesis; NEM, *N*-ethyl morpholine; DCM, dichloromethane; DMF, dimethylformamide; HOBT, 1-hydroxybenzotriazole hydrate; DIC, diisopropylcarbodiimide; DMAP, 4-(dimethylamino)pyridine; NMR, nuclear magnetic resonance; MBHA, *p*-methylbenzhydrylamine resin; HF, anhydrous hydrogen fluoride; HPLC, high-performance liquid chromatography; RP-HPLC, reversed-phase high-performance liquid chromatography; WT, wild type; CD, circular dichroism; Hva, *hV*, (*S*)-(+)-2-hydroxy-2-methylbutyric acid; Hle, *hL*, (*S*)-(−)-2-hydroxyisocaproic acid; bis-ANS, 4,4′-dianilino-1,1′-binaphthyl-5,5′-disulfonic acid; BrAc, bromoacetyl.

ride was purchased from ICN Biomedical (Aurora, OH). Bis-ANS and the NanoOrange assay kit were purchased from Molecular Probes (Eugene, OR).

Peptide Synthesis. All of the designed GCN4 peptides (**gcn4e**) were prepared by manual SPPS on a 0.4 mmol scale using the in situ neutralization/HBTU activation procedure for Boc chemistry as previously described (*31*). All peptides were synthesized on a MBHA resin to generate a C-terminal amide group ($-CONH_2$) upon cleavage from the resin, which neutralizes the C-terminus of the peptide. The peptide coupling was carried out with 5-fold excess (2.0 mmol) of activated amino acid for a minimum of 15 min. To serve as a starting sequence, the GCN4-p1 sequence was chosen. To minimize issues with oxidation upon cleavage and handling, the initial methionine residue (position 3 in our sequence) was replaced with the isosteric unnatural amino acid norleucine (designated as O or Nle hereafter). To also minimize possible stability and oligomerization dependence on pH (*32*), we substituted the central histidine residue (position 19 in our sequence) with a lysine (*33*). The two halves of the heterodimer were identical (Figure 1) save for the nature of the linker and the presence or absence of ester substitutions.

The first peptide (**gcn4-I** hereafter) had the WT sequence and had bromoacetic acid coupled to the amino terminus of the Gly−Gly linker, resulting in an N-terminal bromoacetyl (BrAc) group. The second peptide (**gcn4-II**) had an N_α-acetylated cysteine coupled to the amino terminus of the Gly−Gly linker and had either the WT sequence or an analogue sequence in its main chain. Any α-hydroxy acids were incorporated into the second chain (**gcn4-II**). The coupling of Hva and Hle was carried out using a DIC/HOBt activation method. Hva or Hle (2.2 mmol) in 4 mL of 50% DCM/DMF was activated with DIC (2.0 mmol) in the presence of HOBt (2.4 mmol) at 0 °C for 15 min. The mixture was added to the N-terminally deprotected resin, along with NEM (0.8 mmol), and coupled for 10 min at room temperature. A second coupling was carried out only if the coupling was less than 98%. The ester bond was formed using DIC/DMAP activation. The next amino acid (2.2 mmol) was taken in 4 mL of 50% DCM/DMF and activated with DIC (2.0 mmol) for 15 min at 0 °C. The mixture was added to resin with NEM (0.8 mmol) and less than 10 mg DMAP. This mixture was coupled for 1 h at room temperature. After the chain assembly was completed, the peptides were deprotected and cleaved from the resin by treatment with HF for 1 h at 0 °C with 4% anisole as a scavenger. The resultant peptides were precipitated with ice-cold anhydrous diethyl ether, dissolved in aqueous acetonitrile, and lyophilized.

HPLC. Analytical RP-HPLC was performed on a Hewlett-Packard HPLC 1050 system using Vydac C-18 columns (5 μm, 0.46 × 15 cm). Semipreparative RP-HPLC was performed on a Rainin HPLC system using a Vydac C-18 (10 μm, 1.0 × 15 cm). Linear gradients of A, H_2O/TFA (999:1), and B, acetonitrile/H_2O/TFA (900:99:1), were used to elute bound peptides.

Chemoselective Ligation. Chemoselective ligation was carried out by dissolving purified peptides in slightly basic buffer (100 mM NaP_i, pH 8.0) at moderate concentration—typically 0.2−0.3 mM (2−2.5 mg/mL) of each peptide. Each reaction was monitored by analytical RP-HPLC, and the formation of the thioether linkage (and concomitant disappearance of the reactants) was typically complete within 1 h. The resulting heterodimers (**gcn4e**) were purified by RP-HPLC.

Mass Spectrometry. Electrospray ionization mass spectrometry (ESI-MS) was performed on an API-III triple quadrupole mass spectrometer (PE-Sciex). Peptide masses were calculated from the experimental mass-to-charge (m/z) ratios from all of the observed protonation states of a peptide using MacSpec software (Sciex). Theoretical masses of peptides and proteins were calculated using MacProMass software (Beckman Research Institute, Duarte, CA).

Sedimentation Equilibrium. Sedimentation equilibrium experiments to analyze the oligomeric state of the **gcn4e** heterodimers were carried out in a Beckman XL-I analytical ultracentrifuge at 20 °C with a rotor speed of 37 000 rpm and detection at 276 nm. Samples were prepared at a concentration of 140 μM in 25 mM sodium phosphate and 150 mM sodium chloride at pH 7.0. The data were fit using XL-A software. All variants could only be fit to a single species model of the Lamm equation:

$$A_r = A_o \exp[\omega^2/2RT \cdot M(1 - \bar{v}\rho)](r^2 - r_o^2) \qquad (1)$$

where A_r is radial absorbance, A_o is the baseline absorbance, ω is the rotor speed (s^{-1}), R is the gas constant (J mol^{-1} K^{-1}), T is the temperature (K), υ is the partial specific volume (mL g^{-1}), ρ is the density of solvent (g mL^{-1}), r is the variable radius, and r_o is the meniscus radius. υ was set at 0.73, and ρ was set at 1.0.

Protein Concentration Measurements. Stock solutions of the purified **gcn4e** variants were initially prepared by dissolving the purified peptides (typically 2 mg) in buffer (25 mM NaP_i and 150 mM NaCl, pH 7.0). Concentrations were initially determined by the method of Gill and von Hippel (*34*), measuring the concentration of diluted fractions at 280 nm (ϵ_{280} = 2560 M^{-1} cm^{-1}, corresponding to two tyrosine residues). Owing to the low extinction coefficient of tyrosine in aqueous solution, concentration measurements for each variant were confirmed by amino acid analysis at the TSRI Core Facility according to their standard protocol (*35*). Finally, the NanoOrange (Molecular Probes) assay was used as a third standard to confirm protein concentration measurements. The manufacturer's standard protocol was used as described (*36*), with fluorescence detection carried out on an Aviv ATF 105 automated titrating spectrofluorometer. Samples were excited at 485 nm, and emission was monitored from 520 to 700 nm; emission maxima remained around 590 nm.

Concentration measurements derived from amino acid analysis were the most consistent and used to prepare final stock solutions for the equilibrium denaturation experiments. Stock solutions were diluted to final concentrations of 5 and 10 μM for the chemical denaturation experiments and 100 μM for the thermal denaturation experiments. The concentration of guanidine in the guanidine solutions was determined by diffractometry (*37*); typical concentrations were around 6.6−6.8 M guanidine.

Bis-ANS Binding. Bis-ANS binding studies were carried out in 25 mM sodium phosphate and 150 mM sodium chloride (pH 7.0) by titrating 1 μM of protein solution into

a prepared 1 μM bis-ANS solution. Fluorescence detection was monitored in an Aviv ATF 105 automated titrating spectrofluorometer by exciting the samples at 375 nm and monitoring the emission from 420 to 600 nm. Stocks of recombinant *Escherichia coli* thioredoxin and hen egg white lysozyme were used as controls.

CD Spectroscopy. The CD spectra were recorded on an Aviv 202 automated titrating spectropolarimeter. For wavelength scans, the samples were scanned every 1 nm from 195 to 260 nm, and 3 s was used to collect each data point. At least three scans were taken with each sample. The mean residue molar ellipticity was calculated using the following:

$$[\Theta] = (\Theta_{obs} \times MRM)/(10 \times l \times c) \quad (2)$$

where Θ_{obs} is the observed ellipticity expressed in millidegrees, *MRM* is the total molecular mass of the protein divided by the number of amino acid residues, l is the optical path length in cm, and c is the final peptide concentration in mg/mL. For all denaturation experiments, the change in the raw ellipticity was followed at 222 nm, and 30 s was used to collect each data point. For the chemical denaturation experiments, an equilibration time of 4 min was used, and guanidine was titrated in 0.2 M intervals between 0 and 6.0 M guanidine. For the thermal denaturation experiments, an equilibration time of 4 min was used, and the temperature was increased in 3 °C intervals between 25 and 97 °C. Increasing the equilibration time above 4 min for either the thermal or the chemical denaturation did not change the curve. The chemical denaturation experiments appeared to be fully reversible. The thermal denaturation experiments also appeared to be fully reversible for the WT version of **gcn4e**, although leaving the ester-substituted analogues at high temperature (97 °C) for extended periods of time (30+ min) produced variable amounts of cleavage at the ester bond.

Equilibrium Data Analysis. The folding of the **gcn4e** heterodimer can be described as a two-state, monomeric transition between a folded, native state N and a unfolded, denatured state U:

$$N \leftrightarrow U \quad (3)$$

Ellipticity measurements were normalized with respect to the fraction of folded versus unfolded at any point:

$$f_N = ([\theta] - [\theta]_U)/([\theta]_N - [\theta]_U) \quad (4)$$

where $[\theta]$ is the observed ellipticity at 222 nm, $[\theta]_U$ is the ellipticity of the unfolded state, and $[\theta]_N$ is the ellipticity of the folded state. The equilibrium constant of unfolding can be expressed as a ratio between the fraction of folded monomer f_N and the fraction of unfolded monomer f_U:

$$K_U = f_U/f_N = f_U/f_{1-U} \quad (5)$$

This allows the free energy of unfolding of the protein, ΔG_u, to be written as

$$\Delta G_u = -RT \ln(f_U/f_{1-U}) \quad (6)$$

where R is the gas rate constant, and T is the temperature. With the chemical denaturation experiments, the data was analyzed by fitting the whole denaturation curve with nonlinear regression (*38*) to

$$F = \frac{[(\alpha_F + \beta_F[D]) + ((\alpha_U + \beta_U[D]) \exp(m([D] - D_{50\%})))/RT]}{[1 + \exp(m([D] - D_{50\%}))/RT]} \quad (7)$$

$$F = \frac{[(\alpha_F + \beta_F[D]) + ((\alpha_U + \beta_U[D]) \exp(m([D] - DG_u^{H2O})))/RT]}{[1 + \exp(m([D] - \Delta G^{u^{H2O}}))/RT]} \quad (8)$$

wherein α_F and α_U are the baselines of the folding and unfolding states, respectively, β_F and β_U are the rate of change of those baselines, m is the slope of the unfolding transition in response to denaturant, $[D]$ is the concentration of the denaturant at any point, $D_{50\%}$ is the midpoint of denaturation, and ΔG_u^{H2O} is the free energy of unfolding of the protein in the absence of denaturant (*39*). $\Delta\Delta G_u^{H2O}$ was determined through the difference of values of ΔG_u^{H2O} since the values of m did not significantly change between variants. With the thermal denaturation experiments, the integrated unimolecular form of the van't Hoff equation (*40*) was used to derive the melting temperature by linear extrapolation:

$$K_U = \exp\left[\frac{\Delta H_{vH}}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right)\right] \quad (9)$$

where ΔH_{vH} is the apparent (van't Hoff) enthalpy, and T_m is the transition midpoint of the thermal melting curve. All data were fit using Kaleidagraph (Synergy Software, Reading, PA) and Microsoft Excel.

RESULTS

Peptide Synthesis. Owing to the length (~34 amino acids) of the GCN4 sequence and the relative ease with which α-hydroxy acids can be incorporated into peptide chains by solid-phase peptide synthesis, we chose chemical synthesis to construct the peptides for this study. Synthesis and purification of all sequences were straightforward, with high incorporation of the α-hydroxy acids into the backbone of the peptides. To isolate the effects of a single ester introduction into the backbone of the heterodimer, all mutations were made in the second peptide (**gcn4-II**), with the other peptide remaining as the WT gcn4-r sequence (**gcn4-I**).

Chemoselective Ligation. While tethered dimers have previously been constructed from coiled coil sequences, they utilized N-terminal disulfide bonds to cross-link the two monomers intermolecularly (*17*, *18*, *41*, *42*). While this is a convenient approach, it can be difficult to specifically construct and isolate heterodimers formed by this method. The resulting disulfide bond is also potentially reversible, which can complicate analysis. Therefore, we chose a convenient chemoselective ligation to form a thioether bond (*43*) (Figure 2) between one peptide with a WT sequence (**gcn4-I**) and a second peptide bearing the ester substitutions (**gcn4-II**). Thioether bonds are relatively unreactive and nonreversible and have been used as an alternative to disulfide bonds to cross-link coiled coil sequences (*43*). The thioether ligation was carried out under benign buffer

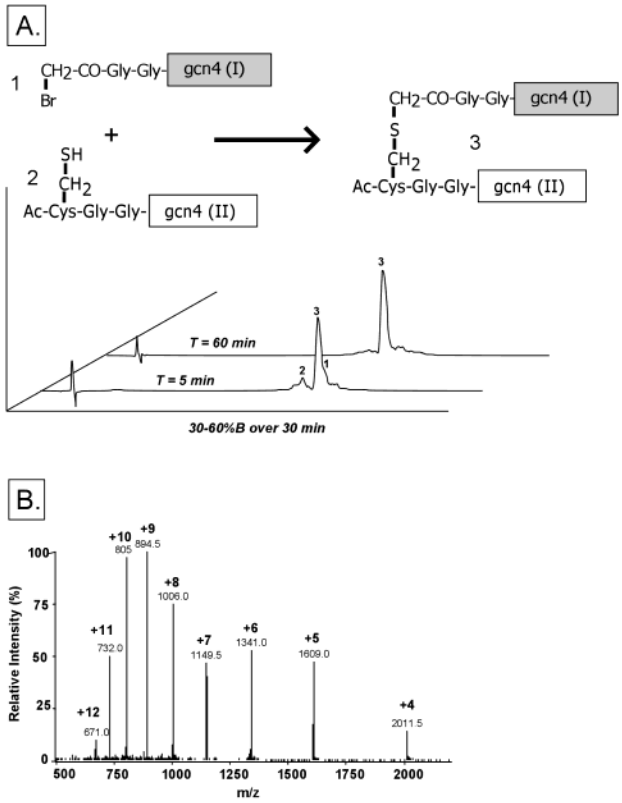

FIGURE 2: Formation of a GCN4 heterodimer by chemoselective ligation. (A) RP-HPLC traces following the course of the representative ligation (that of **L27*hL***) at 5 and 60 min. Formation of a thioether bond at basic pH proceeds rapidly between the two unprotected peptides; **gcn4-I** (1) and **gcn4-II** (2) react to form gcn4e (3), the linked heterodimer. (B) ES/MS spectra of ligated **L27*hL*** (3). The observed mass (8040.8) is in close agreement with the predicted mass (8041.0).

conditions (100 mM phosphate, pH 8). Similar to other ligations templated by protein folding (*44*, *45*), the process was extremely fast, coming to completion in less than 1 h (Figure 2). No differences were observed in the ligation rate of any of the sequences.

CD. All of the designed peptide sequences were examined by CD spectroscopy at 5, 10, and 100 μM. All exhibited strong elliptical minima at $\lambda = 208$ and 222 nm, characteristic of an α-helical structure. The WT version of **gcn4e** had a $[\Theta]$ at 222 nm of $-33\,055$ deg cm^2 mol^{-1}, and all of the other analogues had an equal or higher ellipticity. In addition, all had a $[\Theta]_{222:208}$ ratio between 0.98 and 1.02 (data not shown), characteristic of a system with interacting helices and typical of coiled coils (*46*−*48*).

Sedimentation Equilibrium Analysis. It has been found that many designed coiled coils will self-associate to form higher order oligomers, such as with some variants of the designed coiled coil **x19a** (*32*, *42*). In addition, changing the hydrophobic packing of residues in the a and d positions has been found to change the oligomerization state of GCN4-derived peptides (*49*, *50*). Thus, verification of the oligomerization state of our heterodimeric peptides was necessary. Analytical ultracentrifugation was used to analyze the sedimentation equilibrium of the heterodimers. All designed peptide sequences sedimented as monomers and could only be fit successfully to the single species form of the Lamm equation (Table 1).

Table 1: Results from Thermal Denaturation and Sedimentation Equilibrium Experiments

sequence	oligomerization	mass[a]	cross-linking	T_m (°C)
gcn4-p1[b]	dimer		no	54
gcn4-pVL[b]	dimer/trimer		no	95
ACID/BASE-pLL[c]	monomer/dimer[c]		yes[c]	>100
N19a[d]	monomer		yes	53.6
gcn4-p2N[e]	monomer		yes	83
gcn4e-WT	monomer	7434 (8042)	yes	86.5 ± 1.25
gcn4e-L6*hL*	monomer	6862 (8041)	yes	76.7 ± 0.58
gcn4e-V10*hV*	monomer	7311 (8041)	yes	72.0 ± 0.32
gcn4e-L13*hL*	monomer	7284 (8041)	yes	71.4 ± 0.66
gcn4e-L20*hL*	monomer	6958 (8041)	yes	72.1 ± 0.78
gcn4e-V24*hV*	monomer	6642 (8041)	yes	62.7 ± 0.26
gcn4e-L27*hL*	monomer	6409 (8041)	yes	68.8 ± 0.55

[a] Mass values are observed values from sedimentation equilibrim experiments; following values in paratheses are calculated. [b] From ref *63*. gcn4-pVL is the same as gcn4-p1 with a single point mutation (N16→V16). [c] From ref *41*. ACID/BASE-pLL is a disulfide-cross-linked heterodimer formed by linking ACID-pLL to BASE-pLL; the resulting monomer can dimerize with itself. [d] From ref *32*. [e] From ref *18*. gcn4-p2N is the same as gcn4-p1 with an N-terminal CGG extension.

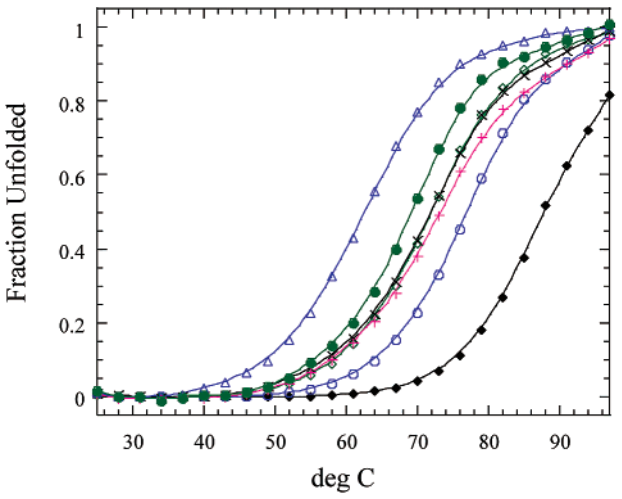

FIGURE 3: Thermal denaturation of **gcn4e** variants. The fraction unfolded as monitored at 222 nm is plotted against the temperature. The WT heterodimer is plotted as black filled diamonds (◆), **L6*hL*** is plotted as blue open circles (○), **V10*hV*** is plotted as green open diamonds (◇), **L13*hL*** is plotted as black cross marks (X), **L20*hL*** is plotted as pink plus marks (+), **V24*hV*** is plotted as blue open triangles(Δ), and **L27*hL*** is plotted as green filled circles (●).

Bis-ANS Binding. To examine the possibility that the introduction of the ester bonds would disrupt and expand the hydrophobic core, resulting in a molten globule state, the constructs were tested for their ability to bind the hydrophobic dye bis-ANS (*51*). No significant increase in fluorescence emission was detected for either the constructs or well-folded protein controls (data not shown), suggesting that all of the **gcn4e** heterodimers are well-folded and retain a relatively compact hydrophobic core.

Thermal and Chemical Denaturation. To examine the thermal stability of the heterodimers, all constructs were thermally denatured and monitored by CD spectroscopy (Figure 3). The thermal unfolding of all proteins studied could be fit to a two-state equilibrium (Table 1) using the van't Hoff relationship (eq 9). Fitting the melting curve with other methods assuming a constant ΔC_p did not improve the

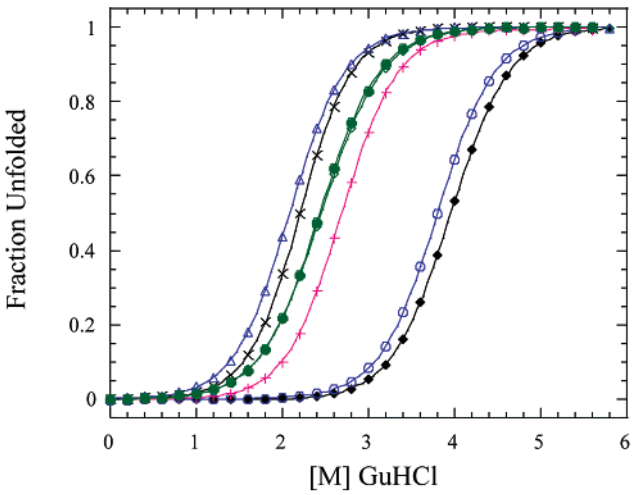

FIGURE 4: Guanidine denaturation of **gcn4e** variants. The fraction unfolded as monitored at 222 nm is plotted against the molarity of titrated guanidine. Plotting legend is as per Figure 3.

Table 2: Stability Measurements from Guanidine Denaturation

sequence	m[a]	$D_{50\%}$[b]	ΔG_u^{H2O} [c]	$\Delta\Delta G_u^{H2O}$ [c]
gcn4-p2′ [d]	1.6 ± 0.1	n/a	8.9 ± 0.4	n/a
gcn4e-WT	1.88 ± 0.10	3.96	7.41 ± 0.06	n/a
gcn4e-L6*hL*	1.76 ± 0.03	3.80	6.70 ± 0.21	0.70
gcn4e-V10*hV*	1.72 ± 0.11	2.45	4.21 ± 0.04	3.20
gcn4e-L13*hL*	1.86 ± 0.08	2.33	4.32 ± 0.03	3.09
gcn4e-L20*hL*	1.88 ± 0.10	2.69	5.04 ± 0.01	2.37
gcn4e-V24*hV*	1.85 ± 0.13	2.08	3.85 ± 0.21	3.56
gcn4e-L27*hL*	1.74 ± 0.15	2.44	4.24 ± 0.19	3.17

[a] In kcal mol^{-1} M^{-1}. [b] In M GuHCl. [c] In kcal mol^{-1}. [d] From ref *17*. This homodimeric sequence is cross-linked by a disulfide bond.

fit. As was found with cross-linking the original gcn4-p1 domain with a N-terminal Cys−Gly−Gly linker (gcn4-p2) (*18*), there was an observable increase in the thermal stability of our designed cross-linked WT heterodimer **gcn4e** (T_m = 86.5 °C) above that of the original unlinked, dimeric gcn4-p1 at a similar concentration (T_m = 54 °C at 5.5 μM). A slight increase in stability was observed over that of the previously reported gcn4-p2 (T_m = 83 °C) (Table 1). All of the ester analogues (**L6*hL*→L27*hL***) showed a noticeable decrease in thermal stability (average ΔT_m = −15.9 °C), although all had a higher T_m than the original gcn4-p1 dimer. The least destabilizing of the mutations was the Leu→Hle 6 analogue, **L6*hL*** (ΔT_m = −9.8 °C); the most destabilizing of the mutations was the Val→Hva 24 analogue, **V24*hV*** (ΔT_m = −23.8 °C). No general trend was observed between the a position substitutions and the d position substitutions within the helix.

Similarly, the stability of all of the **gcn4e** heterodimers against guanidine denaturation was also monitored by CD spectroscopy (Figure 4). The chemical unfolding of all proteins studied could be similarly fit to a two-state equilibrium (Table 2). All of the ester analogues showed a sizable decrease in the free energy of unfolding from chemical denaturation as opposed to the WT **gcn4e** sequence, averaging a loss around 3 kcal mol^{-1} (Table 2). As with the thermal unfolding data, the least destabilizing to the free energy was **L6*hL*** ($\Delta\Delta G_u^{H2O}$ = −0.70 kcal mol^{-1}), and the most destabilizing was **V24*hV*** ($\Delta\Delta G_u^{H2O}$ = −3.56 kcal mol^{-1}). However, the slope of the unfolding transition, m, remained relatively constant at around 1.8 kcal mol^{-1} M^{-1}, although slight fluctuations (∼0.1 kcal mol^{-1} M^{-1}) were observed. Also, similar to the thermal unfolding data, no general trend was observed between the a position substitutions and the d position substitutions.

DISCUSSION

The designed heterodimeric coiled coil sequences were readily amenable to synthesis and purification and folded and behaved similar to other model coiled coils. The degree of destabilization observed in these ester-substituted proteins, however, is somewhat surprising. The introduced esters should be uniformly protected from solvent within the hydrophobic core. The increased hydrophobicity of the ester bond relative to the amide should stabilize the folded state when the ester bond is buried. This desolvation of the ester backbone has been estimated to destabilize the ester in the unfolded state by up to 1.7 kcal/mol as compared to the amide (*28*). This was approximated from the increased partition coefficient of methyl acetate in water-saturated octanol (log P = 0.18) over that of *N*-methyl acetamide (log P = −1.05) (*30*). These partition coefficients reflect the slight preference for an ester to be in octanol over water and the strong preference for an amide to be in water over octanol. If the desolvation of our ester backbone is increased relative to previous studies, then the free energy of desolvation should come close to the theoretical maximum (−1.7 kcal mol^{-1}). This should decrease the effects of solvation on the free energy of any ester substitution since destabilization of the unfolded state results in a net stabilization of the folded state. As a result, our variants should have a similar or lower $\Delta\Delta G_{amide\rightarrow ester}$ than previous substitutions if the strength of the hydrogen bonds remain constant.

Previous studies of ester substitutions in proteins such as T4 lysozyme (*28*) and CI-2 (*25*) observed different changes in free energy at solvent-exposed positions. The most destabilizing single ester incorporation in the T4 lysozyme, Leu 44, destabilized the protein by 1.73 kcal mol^{-1}. Also, the quadruple ester analogue of CI-2, which disrupted three hydrogen bonds, destabilized the protein by 2.93 kcal mol^{-1}. These previous studies from esters in variably solvent-exposed positions in helices suggest that the cost of destabilizing an individual backbone hydrogen bond by an amide to ester substitution is equal or less than 1 kcal mol^{-1}. In contrast, the average cost of ester introduction into the core of GCN4, which affects two hydrogen bonds, is 3 kcal mol^{-1}; in one case (**V24*hV***) the change in free energy is almost 3.6 kcal mol^{-1} for two bonds, nearly twice what was previously observed. Since solvent accessibility is relatively constant in our samples, the additional—and variable—destabilization must come from another source.

An alternative source of destabilization that has been suggested for the amide to ester substitution is an altered interaction with the helix dipole since the charge distributions are different between the two groups (*28*). Again, there is no clear correlation between position along the helix and destabilization. The only potential exception would be the most N-terminal of the ester substitutions, **L6*hL***, which affects the residue (Leu/Hle 6) that caps the N-terminus of the helix and occupies the first d position (Figure 1) (*14*). Some evidence exists from NMR studies that Leu6 is the initial residue to receive a helical H-bond (*18*). As a result,

Table 3: Table of Ester Analogues Introduced into the GCN4 Heterodimer (gcn4e)

number	sequence (I)	sequence (II)	disrupted H bonds[a]	length[b]	$\log(k_{ex}/k_{int})$[c]	ΔG_{op}[d]
WT	WT	WT	n/a			
L6*hL*	WT	Leu→Hle 6	Gly 2 CO→Leu 6 NH	3.08	−2.9	3.94
			Gln 5 CO→Lys 9 NH	3.01	−4.1	5.57
V10*hV*	WT	Val→Hva 10	Leu 6 CO→Val 10 NH	3.04	−5.2	7.06
			Lys 9 CO→Leu 13 NH	2.90	−5.4	7.34
L13*hL*	WT	Leu→Hle 13	Lys 9 CO→Leu 13 NH	2.90	−5.4	7.34
			Glu 12 CO→Lys 16 NH	3.11	−4.6	6.25
L20*hL*	WT	Leu→Hle 20	Lys 16 CO→Leu 20 NH	2.96	−5.3	7.20
			Lys 19 CO→Glu 23 NH	3.06	−5.3	7.20
V24*hV*	WT	Val→Hva 24	Leu 20 CO→Val 24 NH	3.01	−6.0	8.15
			Glu 23 CO→Leu 27 NH	2.94	−5.7	7.74
L27*hL*	WT	Leu→Hle 27	Glu 23 CO→Leu 27 NH	2.94	−5.7	7.74
			Arg 26 CO→Leu 30 NH	3.19	−4.0	5.43

[a] Backbone amide hydrogen bonds. [b] Length in Å (averaged between the two subunits) of the disrupted hydrogen bond in the crystal structure of GCN4-p1 (PDB: 2ZTA). [c] Log of the observed (k_{ex}) versus predicted (k_{int}) proton exchange rates of the analogous amino acid residues from ref *18*. [d] In kcal mol^{-1}, calculated from eq 10. All substitutions are in the second half of the heterodimer (see Figure 1). Backbone amide to ester substitutions effectively destabilize two hydrogen bonds in the context of an α helix; the effects of all substitutions are listed above.

the ester in **L6*hL*** may not destabilize two hydrogen bonds of equal occupancy. One interesting observation is that while **L6*hL*** is the least destabilizing variant of all of ester substitutions, it has a more pronounced effect on the thermal stability of the protein ($\Delta T_m = -9.8$ °C) than to the stability of the protein against chemical denaturation ($\Delta\Delta G_u^{H2O} = -0.70$ kcal mol^{-1}). It was previously observed that modification or truncation of the N-terminal cap of gcn4-p1 significantly destabilizes the protein against thermal denaturation (*52*), which is comparable with our current observations.

In the absence of a correlation with the position along the helix and destabilization, another possibility for the increased destabilization of our analogues is a repulsive electrostatic interaction between the main-chain ester oxygen and the $i + 4$ amide carbonyl oxygen. While in previous studies there was the potential for a solvent-accessible amide carbonyl to make a hydrogen bond to solvent, the amide carbonyls in our study are inaccessible to solvent. This buried partial negative charge could be a potential source of additional destabilization, although a growing body of evidence suggests that burial of polar groups within a hydrophobic core is actually stabilizing, owing to favorable polar−polar van der Waals interactions (*53*). In addition, crystal structures of ester-containing helices have shown that the peptide backbone geometry remains essentially unchanged, with the distance between the mainchain ester oxygen and the $i + 4$ amide carbonyl oxygen staying between 3.0 and 3.3 Å (*54*, *55*). Finally, the increased destabilization we observe appears to vary considerably between residues; **V24*hV*** is 1.2 kcal mol^{-1} more destabilizing than **L20*hL***, and any effect that was nonspecific to an individual residue should destabilize all residues equally.

Another potential explanation for the increased destabilization of these analogues is that the additional flexibility within the backbone granted by the disruption of the hydrogen bonds could weaken the packing of the hydrophobic core. Previous studies have shown that single mutations that disrupt the hydrophobic core can have a heavy cost (*56*, *57*), and this is equally true for coiled coils. In the studies carried out by Hodges and Tripet on the designed coiled coil **x19a** (*58*), changing the packing within the hydrophobic core at the d position resulted in a heavy penalty. For example, altering the packing of the core by mutating Leu to Phe in both monomers destabilized the protein by −2.73 kcal mol^{-1} ($\Delta\Delta G_u$), and changing the electrostatic environment by mutating Leu to Gln in the same position destabilized the protein by −3.24 kcal mol^{-1} ($\Delta\Delta G_u$). While these mutations affected both monomers, the destabilization is on a similar scale to what we have observed. However, if the packing of the hydrophobic core was significantly disrupted by the introduction of these mutations, we would expect the protein to shift toward a more molten globule-like state with lesser association between the helices and a more solvent-accessible core. Formation of such a molten globule has been observed on the designed coiled coil RLP (*59*−*61*). One consistent indicator of the presence of extensive molten globules is increased fluorescence emission from ANS or bis-ANS binding, which was not observed in our variants. Similarly, the ester analogues lacked any significant shift in the slope of the unfolding transition in relation to denaturant, *m*, which was common to the other mutations in previous studies that did disrupt the hydrophobic core (*58*). *m* is proportional to the amount of solvent-accessible surface area exposed in the transition to the unfolded state for proteins that undergo a two-state unfolding transition (*62*). In combination with the lack of a shift in oligomerization state, shift in the ratio of $[\Theta]_{222:208}$, or increased bis-ANS fluorescence emission, the evidence suggests that the structure of the proteins was not significantly disrupted by the substitutions, although structural studies would be needed to come to a definite conclusion.

A final possibility to consider is that the destabilization resulting from the ester substitutions is more dependent on the specific strength of the individual hydrogen bonds themselves than the solvation environment of the protein backbone. While the length of the hydrogen bonds from the crystal structure are not characteristic of strong hydrogen bonds (Table 3), the amide exchange protection factors as determined from H/D exchange (*18*) are reasonably high (Table 3). This degree of amide protection reflects the shift in equilibrium from the unstructured random coil state (k_{int}) to the folded state (k_{ex}) and since the protein appears to follow EX2 behavior, local opening and closing of individual hydrogen bonds and the structure around them is necessary for exchange. As a result, the amide exchange can be related

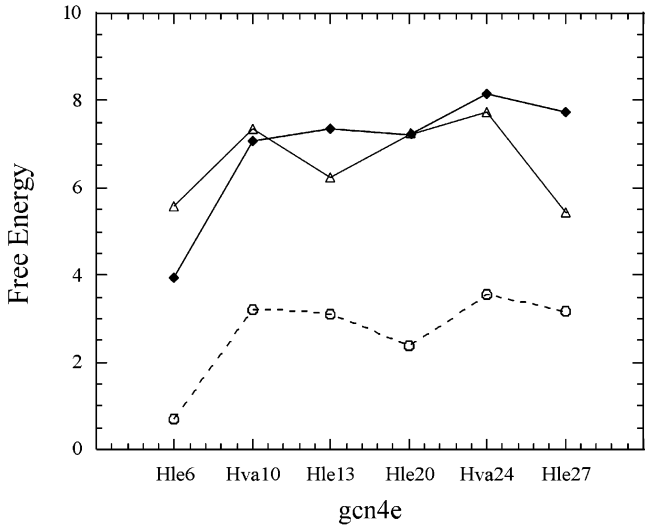

FIGURE 5: Correlation between the measured change in free energy $\Delta\Delta G_u^{H2O}$ (○) and the free energy calculated from hydrogen bond exchange, ΔG_{op}. Hydrogen bond exchange data is from ref *18*. Plotted are the ΔG_{op} values for the hydrogen bonds destabilized by the replacement of the donating amide NH (◆) and the hydrogen bond accepted by the preceding carbonyl (Δ). Free energy is in kcal mol^{-1}.

to the free energy for the structural opening reactions that give way to hydrogen exchange, ΔG_{op}:

$$\Delta G_{op} = -RT \ln(k_{ex}/k_{int}) \quad (10)$$

Plotting ΔG_{op} for the individual NH hydrogen bonds along with $\Delta\Delta G_u^{H2O}$ for the corresponding ester substitutions reveals a similar relative trend (Figure 5) with an average difference of 4 kcal mol^{-1}. This should not be overinterpreted; the value of ΔG_{op} is dependent on the overall local stability rather than a specific strength of a given hydrogen bond, and the $\Delta\Delta G$ values reflect the specific destabilization of the amide hydrogen bonds from the ester substitution. Still, the general trends we observe in the ester-induced destabilization are reflected in the amide protection factors—the lowest belong to the substitution that is least destabilizing to the protein (**L6*hL***), and the highest belong to the substitution that is the most destabilizing to the protein (**V24*hV***).

In conclusion, while the exact thermodynamic and structural effects of an amide to ester substitution within a protein are still being determined, our results suggest that the observed $\Delta\Delta G_{amide\rightarrow ester}$ values are dominated by the free energy cost of destabilizing the individual hydrogen bonds. As a result, ester substitution may be applicable to the general analysis of protein structure and function of the backbone in a manner analogous to side-chain alanine scans and reflect sequence-specific information about the local strength of hydrogen bonds. Since ester substitutions are complementary to other techniques for determining the relative strength of hydrogen bonds, the backbone specific nature of the modification means that they can be combined in parallel with a wide range of other techniques and modifications.

ACKNOWLEDGMENT

The authors graciously acknowledge the assistance of Dr. Peter Wright, Dr. Jane Dyson, and Linda Tennant with CD spectrometer access; Songpon Deechongkit with analytical ultracentrifugation; and the TSRI Core Facility with amino acid analysis.

REFERENCES

1. Baker, E. L., and Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol. 44*, 97−179.
2. Pauling, L., Corey, R. B., and Branson, H. R. (1951) *Proc. Natl. Acad. Sci. U.S.A. 37*, 205−11.
3. Schnellman, J. A. (1955) *C. R. Trav. Lab. Carls. C. 29*, 230−59.
4. Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1.
5. Klotz, I. M., and Franzen, J. S. (1962) *J. Am. Chem. Soc. 84*, 3461−6.
6. Kim, P. S., and Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631.
7. Creighton, T. E. (1990) *Biochem. J. 270*, 1.
8. Dill, K. A. (1990) *Biochemistry 29*, 2733.
9. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989) *Science 243*, 538−42.
10. Rasmussen, R., Benvegnu, D., O'Shea, E. K., Kim, P. S., and Alber, T. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 561−4.
11. Oas, T. G., McIntosh, L. P., O'Shea, E. K., Dahlquist, F. W., and Kim, P. S. (1990) *Biochemistry 29*, 2891−4.
12. McLachlan, A. D., and Steward, M. (1975) *J. Mol. Biol. 98*, 293.
13. Cohen, C., and Parry, D. A. D. (1986) *Trends Biochem. Sci. 11*, 245.
14. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) *Science 254*, 539−44.
15. Krylov, D., Mikhailenko, I., and Vinson, C. (1994) *EMBO J. 13*, 2849−61.
16. Kenar, K. T., Garcia-Moreno, B., and Freire, E. (1995) *Prot. Sci. 4*, 1934−8.
17. Moran, L. B., Schneider, J. P., Kentsis, A., Reddy, G. A., and Sosnick, T. R. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 10699−704.
18. Goodman, E. M., and Kim, P. S. (1991) *Biochemistry 30*, 11615−20.
19. Holtzer, M. E., Lovett, E. G., d'Avignon, D. A., and Holtzer, A. (1997) *Biophys. J. 73*, 1031−41.
20. Blom, C. E., and Gunthard, H. H. (1981) *Chem. Phys. Lett. 84*, 267.
21. Ingwall, R. T., and Goodman, M. (1974) *Macromolecules 7*, 598−606.
22. Wiberg, K. B., and Laiding, K. (1987) *J. Am. Chem. Soc. 109*, 5935−43.
23. Lu, W., Qasim, M. A., Laskowski, M., and Kent, S. B. (1997) *Biochemistry 36*, 673−9.
24. Zhou, J., Case, M. A., Wishart, J. F., and McLendon, G. L. (1998) *J. Phys. Chem. B 102*, 9975−80.
25. Beligere, G. S., and Dawson, P. E. (2000) *J. Am. Chem. Soc. 122*, 12079−82.
26. Wales, T. E., and Fitzgerald, M. C. (2001) *J. Am. Chem. Soc. 123*, 7709−10.
27. Low, D. W., and Hill, M. G. (2002) *J. Am. Chem. Soc. 122*, 11039−40.
28. Koh, J. T., Cornish, V. W., and Schultz, P. G. (1997) *Biochemistry 36*, 11314−22.
29. Arnett, E. M., Mitchell, E. J., and Murty, T. S. S. R. (1974) *J. Am. Chem. Soc. 96*, 3875−91.
30. Leo, A., Hansch, C., and Elkins, D. (1971) *Chem. Rev. 71*, 525−55.
31. Schnolzer, M., Alewood, P., Jones, A., Alewood, D., and Kent, S. B. H. (1992) *Int. J. Pept. Protein Res. 40*, 180−93.
32. Wagschal, K., Tripet, B., Lavigne, P., Mant, C., and Hodges, R. S. (1999) *Protein Sci. 8*, 2312−29.
33. Lovett, E. G., Davignon, D. A., Holtzer, M. E., Braswell, E. H., Zhu, D., and Holtzer, A. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 1781−5.
34. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−26.
35. http://corefac.scripps.edu/aaamm.html, October, 2002.
36. http://www.probes.com/media/pis/mp06666.pdf, October, 2002.
37. Nozaki, Y. (1972) *Methods Enzymol. 26*, 43.
38. Jackson, S., Moracci, M., Elmasry, N., Johnson, C. M., and Fersht, A. R. (1993) *Biochemistry 32*, 11259−69.
39. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−79.
40. John, D. M., and Weeks, K. M. (2000) *Protein Sci. 9*, 1416−9.

41. Lumb, K. J., and Kim, P. S. (1995) *Biochemistry 34*, 8642−8.
42. Wagschal, K., Tripet, B., and Hodges, R. S. (1999) *J. Mol. Biol. 285*, 785−803.
43. Muir, T. W., Williams, M. J., Ginsberg, M. H., and Kent, S. B. H. (1994) *Biochemistry 33*, 7701−8.
44. Camarero, J. A., Pavel, J., and Muir, T. W. (1998) *Angew. Chem., Int. Ed. 37*, 347−9.
45. Beligere, G. S., and Dawson, P. E. (1999) *J. Am. Chem. Soc. 121*, 6332−3.
46. Cooper, T. M., and Woody, R. W. (1990) *Biopolymers 30*, 657−76.
47. Lau, S. Y., Taneja, A. K., and Hodges, R. S. (1984) *J. Biol. Chem. 259*, 13253−61.
48. Sonnichsen, F. D., Van Eyk, J. E., Hodges, R. S., and Sykes, B. D. (1992) *Biochemistry 31*, 8790−8.
49. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) *Science 262*, 1401−7.
50. Harbury, P. B., Kim, P. S., and Alber, T. (1994) *Nature 371*, 80−3.
51. York, S. S., Lawson, R. C., and Worah, D. M. (1978) *Biochemistry 21*, 4480−6.
52. Lumb, K. J., Carr, C. M., and Kim, P. S. (1994) *Biochemistry 33*, 7361−7.
53. Pace, C. N. (2001) *Biochemistry 40*, 310−3.
54. Karle, I. L., Das, C., and Balaram, P. (2001) *Biopolymers 59*, 276−89.
55. Aravinda, S., Shamala, N., Das, C., and Balaram, P. (2002) *Biopolymers 64*, 255−67.
56. Kellis, J. T. J., Nyberg, K., Sali, D., and Fersht, A. R. (1988) *Nature 333*, 784−6.
57. Parsell, D. A., and Sauer, R. T. (1989) *J. Biol. Chem. 264*, 7590−5.
58. Tripet, B., and Hodges, R. S. (2000) *J. Mol. Biol. 285*, 785−803.
59. DeGrado, W. F., Raleigh, D. P., and Handel, T. (1991) *Curr. Opin. Struct. Biol. 1*, 984−93.
60. Betz, S. F., Raleigh, D. P., and DeGrado, W. F. (1993) *Curr. Opin. Struct. Biol. 3*, 601−10.
61. Betz, S. F., Liebman, P. A., and DeGrado, W. F. (1997) *Biochemistry 36*, 2450−8.
62. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) *Protein Sci. 4*, 2138−48.
63. Akey, D. L., Malashkevich, V. N., and Kim, P. S. (2001) *Biochemistry 40*, 6352−60.

BI026862P