JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

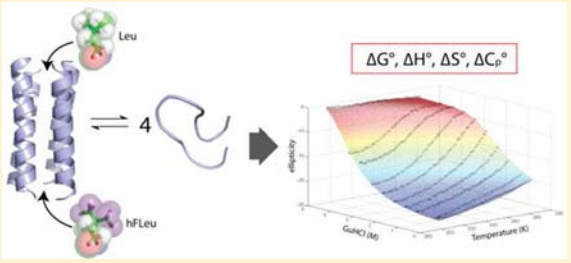
Influence of Fluorination on the Thermodynamics of Protein Folding

Benjamin C. Buer,[†] Benjamin J. Levin,[†] and E. Neil G. Marsh*[,†,‡]

Departments of [†]Chemistry and [‡]Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States

Supporting Information

ABSTRACT: The introduction of highly fluorinated analogues of hydrophobic amino acid residues into proteins has proved an effective and general strategy for increasing protein stability toward both chemical denaturants and heat. However, the thermodynamic basis for this stabilizing effect, whether enthalpic or entropic in nature, has not been extensively investigated. Here we describe studies in which the values of $\Delta H°$, $\Delta S°$, and $\Delta C_p°$ have been determined for the unfolding of a series of 12 small, *de novo*-designed proteins in which the hydrophobic core is packed with various combinations of fluorinated and non-fluorinated amino acid residues. The increase in the free energy of unfolding with increasing fluorine content is associated with increasingly unfavorable entropies of unfolding and correlates well with calculated changes in apolar solvent-accessible surface area. $\Delta C_p°$ for unfolding is positive for all the proteins and, similarly, correlates with changes in apolar solvent-accessible surface area. $\Delta H°$ for unfolding shows no correlation with either fluorine content or changes in apolar solvent-accessible surface area. We conclude that conventional hydrophobic effects adequately explain the enhanced stabilities of most highly fluorinated proteins. The extremely high thermal stability of these proteins results, in part, from their very low per-residue $\Delta C_p°$, as has been observed for natural thermostable proteins.



INTRODUCTION

Non-canonical (or non-proteogenic) amino acids provide a promising avenue for the design of proteins with novel properties. Fluorinated amino acids have attracted much attention in this respect, and over the past decade numerous studies have examined the effect of introducing highly fluorinated analogues of hydrophobic amino acids on the stability and biological activity of small proteins and peptides.[1−3] These include studies on *de novo*-designed proteins that form regular coiled-coil structures,[4−14] which can be synthesized chemically, and on natural proteins that adopt more complex structures, in which the fluorinated residues are generally introduced biosynthetically.[15−25] There have also been several studies investigating the effects of fluorination on the activity of biologically active peptides such as antimicrobial peptides and glucagon.[26−29] In almost all cases, replacing a hydrophobic residue with its fluorinated analogue stabilizes the protein structure against thermal unfolding and denaturation by reagents such as urea, guanidinium chloride (GuHCl), and organic solvents.[4,8,13,14,17,19−23] At the same time, the biological activity of the protein or peptide is usually little altered by fluorination.

The structural basis for the enhanced stability of highly fluorinated proteins has remained obscure. To address this, our laboratory recently determined high-resolution structures for a pair of *de novo*-designed 4-helix bundle proteins to examine the influence of fluorination on protein structure.[30] In one protein (α_4H) the hydrophobic core comprised all leucine residues at the canonical 'a' and 'd' positions, whereas in the second protein (α_4F_3a) the residues at the 'a' positions were all replaced by hexafluoroleucine (hFLeu) so that half of the core residues were fluorinated.[8,14,30] Although α_4F_3a contained 72 fluorine atoms in the hydrophobic core, its structure was remarkably little perturbed: the C_α atoms of α_4F_3a were displaced by an rmsd of only 0.95 Å from the coordinates of α_4H and, except at one position, all the core residues adopted the same packing arrangement in both proteins.

From analysis of these structures we concluded that, although larger, the fluorinated residues are minimally perturbing because they closely match the shape of the hydrocarbon side chains they replace. The increased stability of the fluorinated protein could be adequately explained by the increase in buried hydrophobic surface area that accompanied fluorination without the need to invoke specific "fluorous" interactions between fluorinated side chains.

The focus of this study is on how fluorination affects the thermodynamics of protein folding. Whereas there have been many studies demonstrating, through different measurements, that fluorination stabilizes proteins against unfolding, there is very little data on how the enthalpic and entropic contributions to protein folding change as a consequence of fluorination. Most studies have either focused on increases in T_m as a measure of protein stability or have compared changes to the overall free energy of unfolding, $\Delta\Delta G_u°$, resulting from fluorination. Here we present thermodynamic data for 12 proteins that are variants of the *de novo*-designed anti-parallel 4-helix bundle protein α_4H, which we have described previously.[8,10,13,14,30] This protein has served as a model system in our laboratory with which to investigate the effects of

Received: April 12, 2012
Published: July 16, 2012

   dx.doi.org/10.1021/ja303521h | J. Am. Chem. Soc. 2012, 134, 13027−13034

fluorination on protein structure and stability. The 12 proteins differ only in the makeup of the hydrophobic residues at 'a' and 'd' positions that form the hydrophobic core of the 4-helix bundle.

EXPERIMENTAL PROCEDURES

Peptide Synthesis. L-5,5,5,5′,5′,5′-Hexafluoroleucine was synthesized as described previously[31] and converted to its Boc derivative using standard methods. 4,4,4-Trifluoroethylglycine was purchased from SynQuest Laboratory and enzymatically resolved as described previously.[32] Boc- and Fmoc-protected *β-tert*-butyl-L-alanine were purchased from AnaSpec Inc. Peptides were synthesized by manual Fmoc procedures (α_4H and $\alpha_4$tbA$_6$) or manual Boc procedures [α_4F_2(6,24), α_4F_2(10,20), α_4F_2(13,17), α_4F_3a, α_4F_3d, α_4F_3(6−13), α_4F_3(17−24), $\alpha_4F_3af_3d$, α_4F_3a-tbA$_3$d, and α_4F_6] according to established protocols.[14,33]

Circular Dichroism. CD spectra of protein unfolding were recorded with an Aviv 62DS spectropolarimeter at 222 nm with a 1 mm path length cuvette. To examine the thermal unfolding of the peptide, stock solutions were prepared containing 40 μM peptide (concentration of monomer) in 10 mM potassium phosphate buffer, pH 7.0, with 9−12 different concentrations of GuHCl. The temperature was increased from 4 to 90 °C in increments of 2 °C; ellipticity measurements were recorded with a 10 s averaging time. For all the proteins thermal unfolding was fully reversible (see Supporting Information).

Data Modeling. Thermal unfolding of the proteins was modeled assuming a two-state equilibrium, shown by eq 1, between the folded tetrameric protein (F) and the unfolded monomeric peptide (U), which is characterized by an equilibrium constant $K(T,[\mathrm{GuHCl}])$ that is dependent on temperature and denaturant concentration.

$$\mathrm{F} \Leftrightarrow 4\mathrm{U} \tag{1}$$

Equation 2 relates $K(T,[\mathrm{GuHCl}])$ to [F], [U] and [P], which are the concentrations of folded tetramer, unfolded monomer, and total protein, respectively, so that [P] = 4[F] + [U].

$$K(T,[\mathrm{GuHCl}]) = \frac{[\mathrm{U}]^4}{[\mathrm{F}]} = \frac{4[\mathrm{U}]^4}{[\mathrm{P}]-[\mathrm{U}]} \tag{2}$$

Rearrangement of eq 2 results in the polynomial expression in eq 3.

$$[\mathrm{U}]^4 + \frac{K(T,[\mathrm{GuHCl}])}{4}[\mathrm{U}] - \frac{K(T,[\mathrm{GuHCl}])}{4}[\mathrm{P}] = 0 \tag{3}$$

For fixed [P], given any non-negative value of $K(T,[\mathrm{GuHCl}])$, eq 3 has a unique solution for [U] between 0 and [P]. Equation 3 can be solved numerically, which allows $K(T,[\mathrm{GuHCl}])$ to be calculated at each condition of temperature and GuHCl concentration.

To calculate the values $\Delta H°$, $\Delta S°$, and $\Delta C_p°$ associated with protein unfolding, $K(T,[\mathrm{GuHCl}])$ was fitted to the Gibbs−Helmholtz equation (eq 4), modified by assuming that the Gibbs free energy, $\Delta G°$, varies linearly with GuHCl concentration as described by eq 5, to give eq 6.

$$\Delta G°(T) = \Delta H° - T\Delta S° + \Delta C_p°\left(T - T_0 + T\ln\frac{T_0}{T}\right) \tag{4}$$

$$\Delta G°([\mathrm{GuHCl}]) = \Delta G°(0\ \mathrm{M\ GuHCl}) - m[\mathrm{GuHCl}] \tag{5}$$

$$\Delta G°(T,[\mathrm{GuHCl}]) = \Delta H° - T\Delta S° + \Delta C_p°\left(T - T_0 + T\ln\frac{T_0}{T}\right) - m[\mathrm{GuHCl}] \tag{6}$$

In these equations T is temperature, T_0 is the reference temperature of 25 °C, $\Delta H°$ is the change in enthalpy, $\Delta S°$ is the change in entropy, and $\Delta C_p°$ is the change in heat capacity, each at the reference temperature T_0. It has been observed that $\Delta C_p°$ and m change little over the measured range of denaturant concentration and temperature and are assumed to be constant.[34,35] $K(T,[\mathrm{GuHCl}])$ is then given by eq 7 and global fitting of $K(T,[\mathrm{GuHCl}])$ as a function of T and [GuHCl] allows the values of $\Delta H°$, $\Delta S°$, $\Delta C_p°$, $\Delta G°$, and m to be calculated.[34]

$$K(T,[\mathrm{GuHCl}]) = \exp\left(\frac{-\left(\Delta H° - T\Delta S° + \Delta C_p°\left(T - T_0 + T\ln\frac{T_0}{T}\right) - m[\mathrm{GuHCl}]\right)}{RT}\right) \tag{7}$$

Treatment of Baselines. Circular dichroism at 222 nm was used to monitor protein unfolding. Plotting the ellipticity of α_4 proteins as a function of GuHCl concentration and temperature results in a two-dimensional surface with the pre- and post-transition base planes corresponding to the ellipticity of folded protein (θ_f) and unfolded protein (θ_u). The ellipicity of the unfolded and folded proteins is assumed to vary linearly with T and [GuHCl] and were modeled using eqs 8 and 9, where the parameters a, b, c, d, e, and f describe the ellipticity of the folded and unfolded states at various temperatures and GuHCl concentrations.

$$\theta_u(T,[\mathrm{GuHCl}]) = a + bT + c[\mathrm{GuHCl}] \tag{8}$$

$$\theta_f(T,[\mathrm{GuHCl}]) = d + eT + f[\mathrm{GuHCl}] \tag{9}$$

The observed ellipticity is the sum of the contributions from the unfolded and folded fractions of protein and is described by eq 10.

$$\theta_{obsd} = \theta_u(T,[\mathrm{GuHCl}])\frac{[\mathrm{U}]}{[\mathrm{P}]} + \theta_f(T,[\mathrm{GuHCl}])\frac{[\mathrm{P}]-[\mathrm{U}]}{[\mathrm{P}]} \tag{10}$$

Equations 3, 7, 8, and 9 were substituted implicitly into eq 10, which was fitted to the data using the program MATLAB (MathWorks Inc.), see Supporting Information, to calculate values for a, b, c, d, e, f, $\Delta H°$, $\Delta S°$, $\Delta C_p°$, and m. Data sets comprised 430−512 data points and robust fits were obtained for each data set. As discussed below, reliable values for e and f could not be obtained for the less stable proteins as there were insufficient data points to define the folded base plane. In such cases e and f were set to zero.

Surface Area Calculations. Protein surface areas were analyzed using MSMS in the program Chimera with a probe radius of 1.4 Å corresponding to a water molecule and a vertex density of 10. To calculate apolar core surface area, the surface area of tripeptides of Ala-X-Ala, where X is Leu, tFeG, hFLeu or tBAla, were measured and the difference from Ala-Gly-Ala used. Apolar solvent-accessible surface area (ΔASA_{ap}) was considered as the side chain surface area of the 24 hydrophobic core residues, which would be exposed to solvent upon protein unfolding.

RESULTS AND DISCUSSION

In this investigation we sought to better understand how fluorination changes the thermodynamics of protein folding. In particular, we wished to determine whether the generally observed increases in the thermodynamic stability of fluorinated proteins arise primarily from changes in enthalpy or entropy and the origin of the increased thermal stability that fluorination generally confers on proteins. The model protein α_4H, that is designed to form a tetrameric, anti-parallel 4-α-helix bundle,[14] and its fluorinated variants provide a good system with which to examine this question.

α_4H adopts a well-folded structure, consistent with its design, which has recently been experimentally verified by X-ray crystallography.[30] The Leu residues at 'a' and 'd' positions pack the hydrophobic core in six layers in a regular, repeating manner. In a series of prior investigations[8,10,13,14] we have synthesized and characterized various fluorinated versions of α_4H in which hFLeu replaces Leu in various combinations at the 'a' and 'd' positions (summarized in Figure 1). All of these

	abcdefg	abcdefg	abcdefg	a
α_4H	Ac–GN	ADE**L**YKE	**L**ED**L**QER	**L**RK**L**RKK	**L**RSG–NH_2
$\alpha_4$tbA$_6$	Ac–GN	ADE**X**YKE	**X**ED**X**QER	**X**RK**X**RKK	**X**RSG–NH_2
α_4F$_2$(6,24)	Ac–GN	ADE**X**YKE	LED**L**QER	LRK**L**RKK	**X**RSG–NH_2
α_4F$_2$(10,20)	Ac–GN	ADE**L**YKE	**X**ED**L**QER	LRK**X**RKK	LRSG–NH_2
α_4F$_2$(13,17)	Ac–GN	ADE**L**YKE	LED**X**QER	**X**RK**L**RKK	LRSG–NH_2
α_4F$_3$(6-13)	Ac–GN	ADE**X**YKE	**X**ED**X**QER	LRK**L**RKK	LRSG–NH_2
α_4F$_3$(17-24)	Ac–GN	ADE**L**YKE	LED**L**QER	**X**RK**X**RKK	**X**RSG–NH_2
α_4F$_3$a	Ac–GN	ADE**L**YKE	**X**ED**L**QER	**X**RK**L**RKK	**X**RSG–NH_2
α_4F$_3$d	Ac–GN	ADE**X**YKE	LED**X**QER	LRK**X**RKK	LRSG–NH_2
α_4F$_3$a-tbA$_3$d	Ac–GN	ADE**X**YKE	**X**ED**X**QER	**X**RK**X**RKK	**X**RSG–NH_2
α_4F$_3$af$_3$d	Ac–GN	ADE**X**YKE	**X**ED**X**QER	**X**RK**X**RKK	**X**RSG–NH_2
α_4F$_6$	Ac–GN	ADE**X**YKE	**X**ED**X**QER	**X**RK**X**RKK	**X**RSG–NH_2

L = Leu
X = tBAla
X = hFLeu
X = tFeG

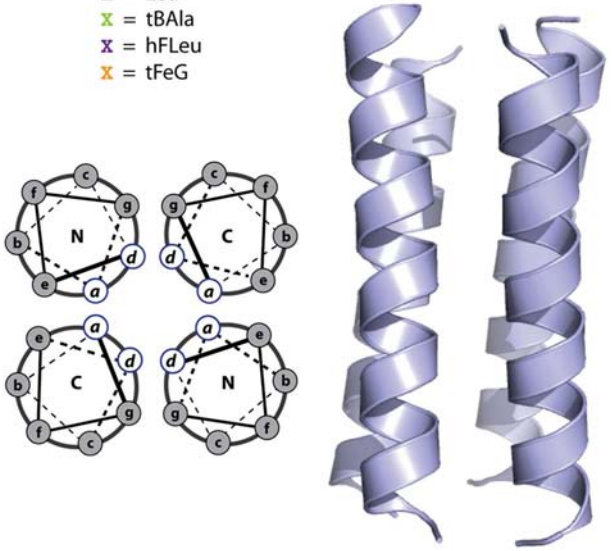


Figure 1. (Top) Sequences of α_4 proteins with hydrocarbon and fluorocarbon analogues of Leu that are substituted at 'a' and 'd' positions. (Bottom) Helical wheel diagram of anti-parallel 4-helix bundle with hydrophobic residues at 'a' and 'd' positions and coiled-coil structure of an anti-parallel 4-helix bundle.

proteins appear to adopt well-folded 4-helix bundles very similar to α_4H as judged by CD, analytical ultracentrifugation, exclusion of hydrophobic dyes such as ANS, and in the case of α_4F$_3$a, as confirmed by X-ray crystallography.[30]

For the current study we synthesized two additional fluorinated variants of α_4H: α_4F$_3$(6–13) and α_4F$_3$(17–24) in which, respectively, the first three and last three hydrophobic core positions contain hFLeu. This results in an arrangement in which two Leu and two hFLeu residues pack each layer of the hydrophobic core, rather similarly to that of the α_4F$_3$a and α_4F$_3$d proteins. We also synthesized two peptides containing *β-tert*-butylalanine (tbAla) at 'a' and 'd' positions. We chose this amino acid as a non-fluorinated mimic of hFLeu, because although its shape is slightly different its side-chain volume is very similar. In one peptide, α_4F$_3$a-tbA$_3$d, the 'a' positions contain hFLeu and the 'd' positions tbAla; in the other, $\alpha_4$tbA$_6$, all the 'a' and 'd' positions contain tBAla. Each of these peptides appears to adopt a well-folded 4-helix bundle structure as judged by the criteria mentioned above. The sequences and structures of all 12 peptides used in this study are summarized in Figure 1.

Van't Hoff Analysis. We used a Van't Hoff analysis to determine the thermodynamic parameters $\Delta H°$, $\Delta S°$, and $\Delta C_p°$ associated with the unfolding of each protein. Thermal denaturation curves were measured for each protein in 9–12 different concentrations of GuHCl by following changes in ellipticity at 222 nm. The addition of GuHCl perturbs the unfolding temperature, allowing measurements to be made over a wider temperature range so that $\Delta C_p°$ can be reliably determined.[34] The family of unfolding curves was then globally fitted to the Gibbs–Helmholtz equation using the program MATLAB to obtain values for $\Delta H°$, $\Delta S°$, $\Delta C_p°$, and m for each protein at the reference temperature, 298 K, in the absence of denaturant (Figure 2 and Supporting Information). The data analysis assumes a two state model for protein unfolding, and that $\Delta C_p°$ and m, the change in the free energy of folding with GuHCl concentration, are independent of temperature. These assumptions are generally considered to be valid for most proteins.[34,35]

Robust fits were obtained for each of the proteins studied; however, it was not possible to reliably determine the slope of the lower base-plane for some of the proteins because they begin to unfold at very low concentrations of GuHCl. Therefore in calculating the thermodynamic parameters for α_4H, α_4F$_3$af$_3$d, α_4F$_2$(6,24), α_4F$_2$(10,20), and α_4F$_2$(13,17) a flat lower base-plane was assumed. This approximation may introduce some error into the analysis of these less stable proteins; however, when the same approximation was applied to some of the more stable proteins, for which both the upper and lower base-planes could be reliably determined, the calculated values for $\Delta H°$, $\Delta S°$, and $\Delta C_p°$ did not change by more than ~10%. The values for $\Delta H°$, $\Delta S°$ and $\Delta C_p°$ for each protein are presented in Table 1, together with the value of $\Delta G_u°$, calculated from $\Delta H°$ and $\Delta S°$. It is evident that this method gives slightly larger values, by 1–2 kcal/mol, for $\Delta G_u°$ than those we have previously measured by GuHCl denaturation at single temperatures; the reason for this small discrepancy is unclear. As illustrated in Figure 3, the majority of the increase in stability across the series of peptides can be ascribed to entropic effects

Correlations with Change in Solvent-Accessible Surface Area. Changes in the thermodynamics of natural proteins are well known to correlate with changes in solvent-accessible surface area (ΔASA) that accompany the transition from folded to unfolded protein.[36–39] In many studies the distinction is made between polar and apolar ΔASA as the sign $\Delta C_p°$ for unfolding is expected to be positive for an increase in apolar ASA and negative for an increase in polar ASA.[36] Correlation of $\Delta C_p°$ with ΔASA of natural proteins is complicated by the fact that changes in polar ASA and apolar ASA are themselves highly correlated.[38,39] In the α_4 proteins, only the hydrophobic core residues vary so that only apolar ASA is changed (ΔASA_{ap}).

For six of the proteins, α_4H, α_4F$_3$af$_3$d, α_4F$_3$a, α_4F$_3$d, α_4F$_3$(6–13) and $\alpha_4$tbA$_6$, crystal structures are available (ref 30 and B.C.B. and E.N.G.M., unpublished results) and were used to calculate ΔASA_{ap} of the core 'a' and 'd' residues, assuming an extended conformation in the unfolded form. For the other proteins ΔASA_{ap} was calculated from the surface areas of the individual 'a' and 'd' residues comprising the hydrophobic core; i.e. it was assumed that the structures are not significantly different for those for which crystal structures are available.

Plots of $\Delta G_u°$, $\Delta S°$, and $\Delta C_p°$ against apolar surface area for each peptide are shown in Figure 4. The trend of increasing $\Delta G_u°$ with increasing apolar surface area is similar to what we have observed in previous studies of these peptides. $\Delta G_u°$ correlates well with increasing apolar surface area (R = 0.925). Fits of the data yield values of $\Delta\Delta G_u°$ = 28.3 cal mol^{-1} Å^{-2} for

dx.doi.org/10.1021/ja303521h | J. Am. Chem. Soc. 2012, 134, 13027–13034

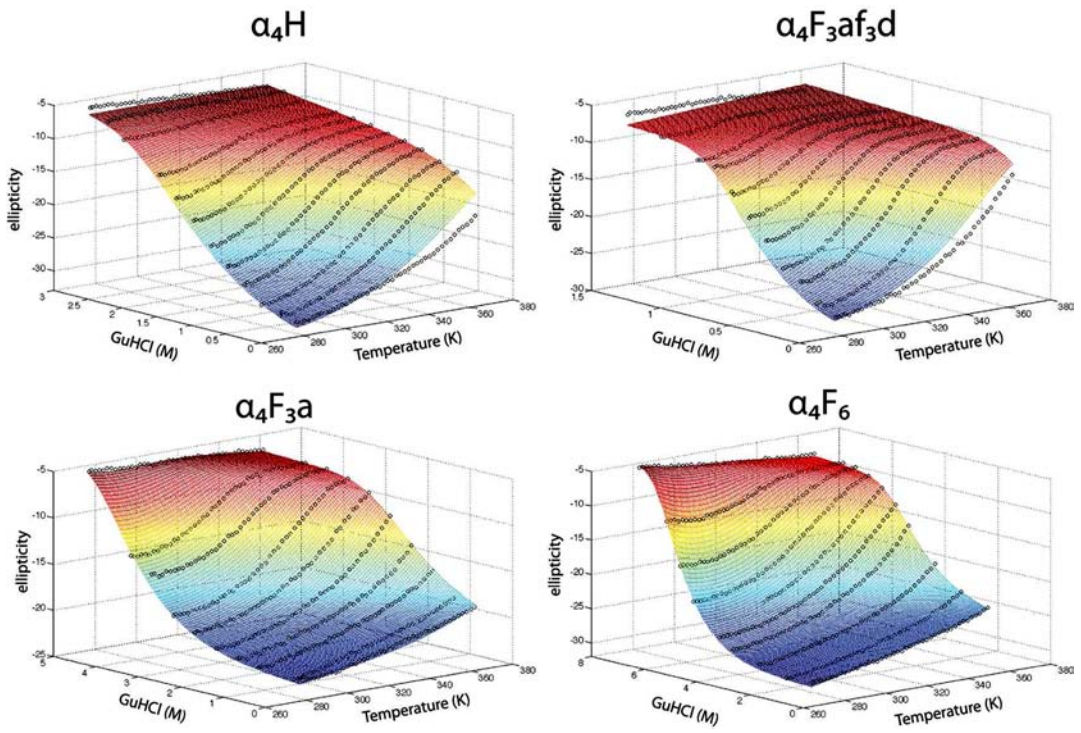


Figure 2. Unfolding of representative α_4 proteins as a function of temperature and GuHCl. Fits to the data are represented by a colored surface with blue as the folded base plane and red as the unfolded base plane.

Table 1. Summary of the Thermodynamic Parameters Derived from Temperature and GuHCl-Induced Unfolding of Proteins at T = 298 K

protein	ΔH° (kcal/mol)	$T\Delta S^\circ$ (kcal/mol)	ΔC_p° (kcal mol^{-1} K^{-1})	m (kcal mol^{-1} M^{-1})	ΔG°_{unfold} (kcal/mol)
α_4H	15.87 ± 0.57	−4.69 ± 0.48	0.22 ± 0.02	2.19 ± 0.07	20.56 ± 0.74
$\alpha_4F_3af_3d$	20.70 ± 1.05	−0.13 ± 0.90	0.19 ± 0.03	6.32 ± 0.28	20.83 ± 1.38
α_4F_2(6,24)	16.13 ± 0.62	−5.86 ± 0.52	0.29 ± 0.02	2.51 ± 0.18	21.99 ± 0.81
α_4F_2(10,20)	14.55 ± 0.48	−8.26 ± 0.41	0.37 ± 0.02	2.15 ± 0.05	22.81 ± 0.63
α_4F_2(13,17)	12.54 ± 0.38	−9.44 ± 0.32	0.34 ± 0.02	2.19 ± 0.05	21.98 ± 0.50
α_4tbA_6	13.17 ± 0.28	−12.56 ± 0.21	0.56 ± 0.02	2.35 ± 0.05	25.72 ± 0.35
α_4F_3(6−13)	19.56 ± 0.61	−6.43 ± 0.38	0.53 ± 0.03	2.71 ± 0.09	25.99 ± 0.72
α_4F_3(17−24)	12.95 ± 0.39	−12.02 ± 0.28	0.60 ± 0.03	2.38 ± 0.08	24.97 ± 0.47
α_4F_3a	19.07 ± 0.60	−7.99 ± 0.38	0.54 ± 0.04	2.29 ± 0.09	27.05 ± 0.71
α_4F_3d	21.37 ± 0.30	−6.99 ± 0.19	0.69 ± 0.02	2.68 ± 0.04	28.36 ± 0.36
$\alpha_4F_3atbA_3d$	12.95 ± 0.32	−17.46 ± 0.27	0.64 ± 0.02	2.45 ± 0.05	30.41 ± 0.41
α_4F_6	18.30 ± 0.28	−13.42 ± 0.17	0.63 ± 0.04	2.21 ± 0.04	31.72 ± 0.33

the change in free energy of unfolding per unit hydrophobic surface area. These values lie close to the generally accepted energetic contribution of the hydrophobic effect to protein folding of ΔG_u° = 25−30 cal mol^{-1} Å^{-2} that has been determined from numerous studies on hydrophobic small molecules and proteins.[37,40,41]

Plots of $T\Delta S^\circ$ against apolar surface area (Figure 4B) also show a good correlation, R = 0.834. Notably, the correlation is not quite as strong as that found with ΔG_u°, indicating that a small degree of enthalpy−entropy compensation is present. At 298 K the area coefficient for the entropy change, $T\Delta S^\circ$ = −31.6 cal mol^{-1} Å^{-2}, implying that most of the increase in the free energy of unfolding is due to entropic effects. This is in accord with the increased stability associated with fluorination arising primarily from the hydrophobic effect in which the increase in entropy of folding is ascribed to the release of water molecules that form an ordered clatherate around the hydrophobic side chains in the unfolded state.

It is known that natural proteins show a strong correlation of ΔASA with m, the coefficient for the change in ΔG_u° with denaturant concentration.[39] However, for the α_4 proteins no correlation between ΔASA and m appears to be present in the data. This may be because m is better correlated with polar ΔASA, including the peptide backbone, which should hydrogen-bond with GuHCl. In this study, polar residues were kept invariant; however, for natural proteins polar ΔASA and apolar ΔASA are, themselves, highly correlated,[38,39] making it difficult to distinguish the effect of polar and apolar surface area changes on m.

Significantly, there appears to be no correlation between ΔH° and apolar surface area (Figure 4D) or between ΔH° and the number of trifluoromethyl groups in the protein. An increase in ΔH° would be expected if electrostatic interactions arising from the permanent dipole moments of the trifluoromethyl groups and dipole moments of hydrogen-bonding moieties of the protein were contributing to the

dx.doi.org/10.1021/ja303521h | J. Am. Chem. Soc. 2012, 134, 13027−13034

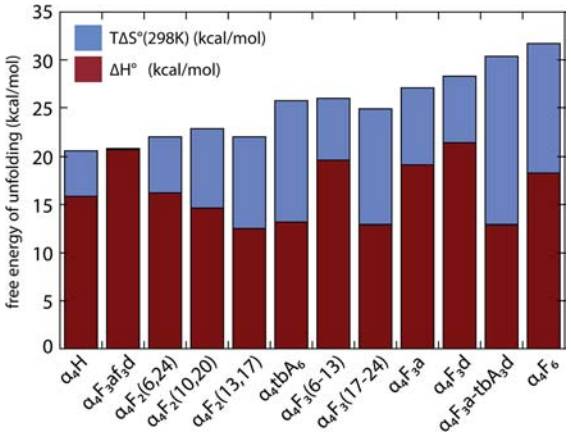


Figure 3. Comparison of the entropic ($T\Delta S°$) (T = 298 K) and enthalpic ($\Delta H°$) contributions to the free energy of unfolding ($\Delta G_u°$) of the α_4 proteins.

greater thermodynamic stability. Such interactions, sometimes described as the "polar hydrophobic effect", have been observed when certain fluorinated small molecules bind to proteins.[42,43] The absence of any correlation indicates that if any specific fluorine−fluorine non-covalent interactions do occur, they do not make a substantive contribution to $\Delta G_u°$.

Changes in Heat Capacity of Unfolding, $\Delta C_p°$. The parameter $\Delta C_p°$ is particularly informative because a positive value is expected if unfolding is dominated by solvation of hydrophobic side chains, whereas a negative $\Delta C_p°$ indicates that solvation of polar residues dominates unfolding.[36] For the α_4 series of proteins (which all have the same number of residues) the per-residue $\Delta C_p°$ values are positive and range from 2 to 6 cal mol^{-1} K^{-1} residue^{-1}. These values are much lower than those typically measured for natural, well-folded proteins for which per-residue $\Delta C_p°$ generally lies within the range of 10−15 cal mol^{-1} K^{-1} residue^{-1}.[39] The small value of $\Delta C_p°$ is in part responsible for the very high thermal stability of these proteins; i.e., $\Delta H°$ and $\Delta S°$ change only slowly as a function of temperature. We note that the α_4 proteins are extremely thermostable, with T_m of the most stable proteins estimated to be above 220 °C in the absence of denaturant.

There are several explanations for the low per-residue $\Delta C_p°$ values. One is that this is the result of a molten globule state.[44] However, this seems highly unlikely because the α_4 proteins

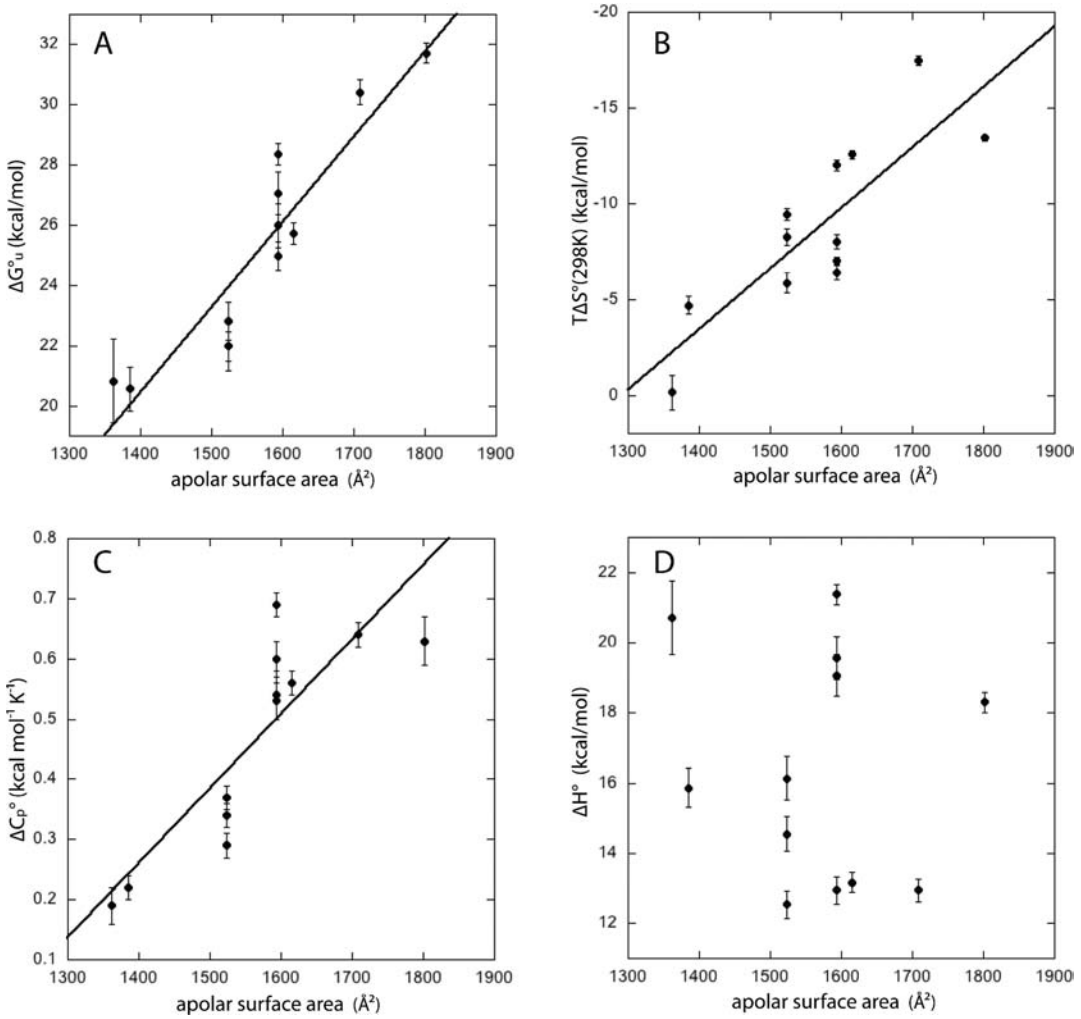


Figure 4. Plots of thermodynamic parameters as a function of protein apolar surface area. (A) Plot of $\Delta G_u°$ (T = 298 K) against apolar surface area; slope =28.3 cal mol^{-1} Å^{-2} (R = 0.925). (B) Plot of $T\Delta S°$ against apolar surface area; slope = −31.6 cal mol^{-1} Å^{-2} (T = 298 K; R = 0.834). (C) Plot of $\Delta C_p°$ against apolar surface area; slope =1.2 cal mol^{-1} K^{-1} Å^{-2} (R = 0.856). (D) Plot of $\Delta H°$ (T = 298 K) against apolar surface area, showing no correlation to apolar surface area.

dx.doi.org/10.1021/ja303521h | J. Am. Chem. Soc. 2012, 134, 13027−13034

such as α_4F_6 exhibit well-dispersed amide resonances in $^{15}N-^{1}H$ HSQC NMR spectra and exclude hydrophobic dyes such as ANS from their cores[13]—both traits of well-folded proteins. Furthermore, the crystal structures of even the least stable proteins show a well-packed hydrophobic core, although the first and last one or two residues are typically not resolved in the structure implying some fraying of the ends of the 4-helix bundle.[30] Another explanation is that the α_4 proteins have a much lower proportion of apolar residues (7 out of 27 or 26%) than most natural proteins (42%).[45] Changes in polar surface area solvent accessibility are associated with negative ΔC_p° values, which would offset the positive ΔC_p° associated with the change in apolar surface area and decrease the net per-residue ΔC_p°.

A third explanation for the low ΔC_p° values is that they are the result of residual structure in the unfolded state.[46,47] Support for this idea derives from studies on natural proteins from thermophilic organisms, which also tend to exhibit low per-residues ΔC_p° values, possibly as an evolutionary strategy to maintain the folded state at high temperature. Various lines of evidence point to such proteins retaining a compact structure in the unfolded state, possibly through non-specific hydrophobic contacts, and thereby reducing the amount of solvent exposed area that is buried upon folding. It is unclear how structured the α_4 proteins are in the unfolded state. No residual secondary structure was evident in the CD spectra of the unfolded proteins, although this technique would not detect compact structures lacking regular secondary structure.

ΔC_p° increases fairly linearly (R = 0.856) with increasing apolar surface area (Figure 4C), which is consistent with the hypothesis that ΔC_p° is proportional to the change in apolar surface area upon unfolding.[34,36,44] The area coefficient calculated from these data is 1.2 cal mol^{-1} K^{-1} Å^{-2}. The literature values for area coefficients calculated from other sets of proteins vary quite widely from 0.16 to 0.5 cal mol^{-1} K^{-1} Å^{-2}.[36,38] In part, this may be due to differences in the way ΔASA_{ap} is calculated and uncertainties in the surface area of the unfolded state. In particular, the various natural globular proteins that comprised the basis sets of previous analyses were diverse in both sequence and fold,[38,39] making the comparison of ΔASA_{ap} much less straightforward. Nevertheless, it appears that for the α_4 proteins ΔC_p° is unusually sensitive to changes in ΔASA_{ap}; the reasons for this are currently unclear.

It is also unclear whether the atypical properties of ΔC_p° associated with the α_4 proteins arise from fluorination, *per se*, or the de novo-designed 4-helix coiled-coil scaffold. We suspect the latter because α_4tbA_6 and $\alpha_4F_3atbA_3d$ which both contain the larger unnatural leucine analogue *β-tert*-butylalanine follow the same trend even though this side chain contains no fluorine.

Evaluation of Thermodynamic Convergence. A consequence of ΔC_p° being non-zero is that entropy and enthalpy of protein folding is temperature dependent. Various studies have observed that for most proteins ΔS and ΔH converge toward common values at a particular temperature, designated T_S* or T_H*, respectively, when normalized for molecular weight or number of residues.[38,48] It has been argued, based on the temperature at which the entropy of dissolution of small hydrophobic compounds extrapolates to zero, that the entropy of unfolding at T_S*, ΔS*, represents the value at which the hydrophobic contribution to entropy is minimal.[49] By analogy, the enthalpy unfolding at T_H*, ΔH*, represents the value at which the hydrophobic contribution to enthalpy is minimal; i.e., ΔS* comprises mainly the configurational entropy change in unfolding, and ΔH* comprises mainly electrostatic and van der Waals interactions. For a set of structurally related proteins that exhibit convergent behavior ΔS*, ΔH*, T_S*, and T_H* can be calculated by plotting ΔC_p° vs ΔS° or ΔH° using the standard thermodynamic eqs 11 and 12, relating temperature, heat capacity, and entropy and enthalpy, to obtain ΔS*, ΔH*, T_S*, and T_H*.

$$\Delta S_0 = \Delta S^* + \Delta C_p \ln\left(\frac{T_0}{T_S^*}\right) \tag{11}$$

$$\Delta H_0 = \Delta H^* + \Delta C_p(T_0 - T_H^*) \tag{12}$$

The α_4 proteins represent a highly structurally related series and might be expected to show convergent behavior. The plot of ΔS° against ΔC_p° for the α_4 proteins is shown in Figure 5A and exhibits a moderate correlation of these two thermodynamic quantities with R = 0.709. From these data ΔS^* = −0.2 ± 1.0 cal mol^{-1} K^{-1} and T_S^* = 44.4 ± 6 °C.

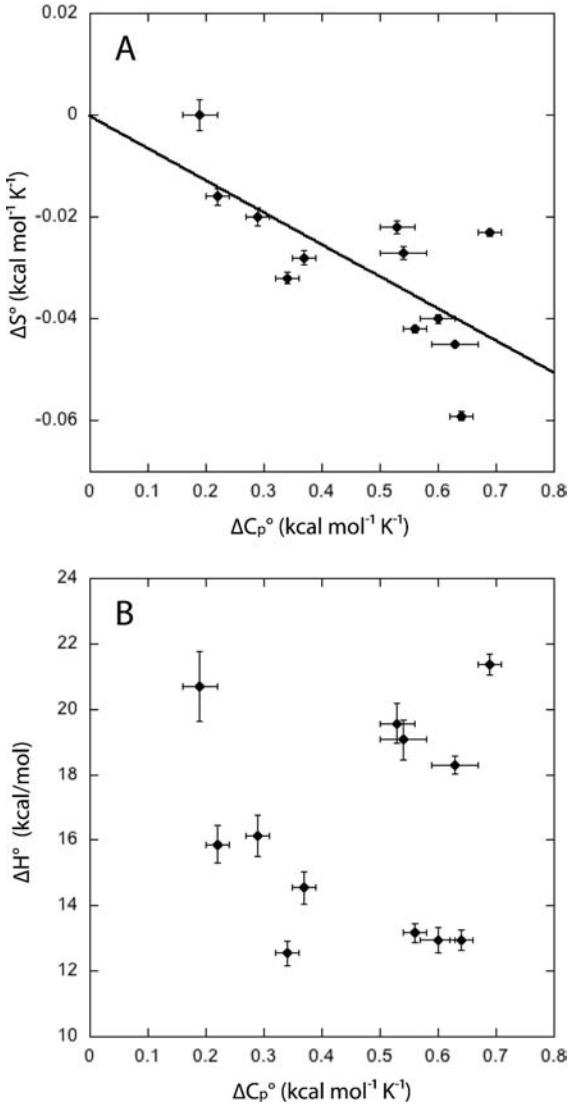


Figure 5. Plots of thermodynamic parameters as a function of protein heat capacity. (A) Plot of ΔS° against ΔC_p°; intercept = −0.2 cal mol^{-1} K^{-1} (T = 298 K; R = 0.709). (B) Plot of ΔH° (T = 298 K) against ΔC_p° shows no correlation.

dx.doi.org/10.1021/ja303521h | J. Am. Chem. Soc. 2012, 134, 13027−13034

This suggests that at the convergence temperature the entropy due to non-hydrophobic effects, i.e., configurational entropy, is very small. There is no correlation evident between ΔH° and ΔC_p° (Figure 5B), an observation that is consistent with the lack of correlation of ΔH° and ΔASA_{ap} (Figure 4D). This further suggests that any hydrophobic contribution to the enthalpy of unfolding is small.

In their original study Murphy et al.[48] used a set of 11 proteins that exhibited a well-correlated relationship between per-residue ΔS° and ΔC_p° that yielded T_S^* = 112 °C and which coincided with T_S^* for the dissolution of hydrophobic small molecules. This was cited as evidence that ΔC_p° reflects the contribution of the hydrophobic effect to the entropy of unfolding. Subsequent studies, using a more extensive set of globular proteins, have found the correlation to be much weaker and T_S^* significantly lower; for example, Robertson and Murphy[38] analyzed a set of 49 proteins and found T_S^* = 65 °C. Nevertheless, the interpretation of T_S^* and T_H^* as representing the temperature at which the hydrophobic contribution to ΔS and ΔH reaches a minimum appears to be generally accepted. It is evident that T_S^* for the α_4 proteins is much lower than the convergence temperature for globular proteins in general. This is primarily a consequence of the high thermal stability and small ΔC_p° [because T_S^* is determined by slope of the plot in Figure 5A, such that $\Delta\Delta S^\circ/\Delta\Delta C_p^\circ = \ln(T_0/T_S^*)$]. This appears to be a general phenomenon, as recent analysis of the thermodynamics of a large group of 116 proteins,[50] divided into thermostable and mesostable proteins based on their T_m values, observed that, on average, thermostable proteins have smaller per-residue ΔC_p° and larger ΔS° and thus will exhibit lower T_S^* than mesostable proteins.

CONCLUSIONS

In conclusion, Van't Hoff analysis of this series of model proteins, in which the fluorine content is progressively increased, show no evidence for any unusual mechanism of stabilization, for example, fluorous interactions or polar hydrophobic effects, that could be specifically attributed to fluorine. Rather, the correlations between increasing buried apolar surface area and increasing entropy and heat capacity indicate that conventional hydrophobic effects are responsible for the stabilizing effects of highly fluorinated amino acids such as hFLeu. The high thermal stability of these proteins arises from the relatively large enthalpic contribution to ΔG_u° and the relatively small ΔC_p° exhibited by these proteins.

As we have discussed previously,[30] the most likely reason that fluorination has proved such a generally effective strategy for stabilizing proteins while maintaining structure and activity is that fluorination closely preserves the shape of the side chain while at the same time increasing surface area. This allows the larger fluorinated residue to be accommodated within the existing protein fold with minimal perturbation to the structure and activity of the protein. Whether fluorination contributes to stabilizing residual structure in the unfolded state, as suggested by the small ΔC_p° of these proteins remains to be determined.

ASSOCIATED CONTENT

Supporting Information

MATLAB script for the routine used to fit the unfolding data sets and figures showing the unfolding surfaces for all the proteins analyzed in this study. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

nmarsh@umich.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the U.S. Department of Defense Multidisciplinary University Research Initiative (DoD 59743-CH-MUR to E.N.G.M.).

REFERENCES

(1) Jäckel, C.; Koksch, B. *Eur. J. Org. Chem.* **2005**, *2005*, 4483.
(2) Marsh, E. N. G. *Chem. Biol.* **2000**, *7*, R153.
(3) Yoder, N. C.; Yüksel, D.; Dafik, L.; Kumar, K. *Curr. Opin. Chem. Biol.* **2006**, *10*, 576.
(4) Bilgiçer, B.; Fichera, A.; Kumar, K. *J. Am. Chem. Soc.* **2001**, *123*, 4393.
(5) Bilgiçer, B.; Kumar, K. *Tetrahedron* **2002**, *58*, 4105.
(6) Bilgiçer, B.; Kumar, K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15324.
(7) Bilgiçer, B.; Xing, X.; Kumar, K. *J. Am. Chem. Soc.* **2001**, *123*, 11815.
(8) Buer, B. C.; de la Salud-Bea, R.; Al Hashimi, H. M.; Marsh, E. N. G. *Biochemistry* **2009**, *48*, 10810.
(9) Chiu, H.-P.; Suzuki, Y.; Gullickson, D.; Ahmad, R.; Kokona, B.; Fairman, R.; Cheng, R. P. *J. Am. Chem. Soc.* **2006**, *128*, 15556.
(10) Gottler, L. M.; de la Salud-Bea, R.; Marsh, E. N. G. *Biochemistry* **2008**, *47*, 4484.
(11) Jäckel, C.; Salwiczek, M.; Koksch, B. *Angew. Chem., Int. Ed.* **2006**, *45*, 4198.
(12) Jäckel, C.; Seufert, W.; Thust, S.; Koksch, B. *ChemBioChem* **2004**, *5*, 717.
(13) Lee, H.-Y.; Lee, K.-H.; Al-Hashimi, H. M.; Marsh, E. N. G. *J. Am. Chem. Soc.* **2006**, *128*, 337.
(14) Lee, K.-H.; Lee, H.-Y.; Slutsky, M. M.; Anderson, J. T.; Marsh, E. N. G. *Biochemistry* **2004**, *43*, 16277.
(15) Chiu, H.-P.; Kokona, B.; Fairman, R.; Cheng, R. P. *J. Am. Chem. Soc.* **2009**, *131*, 13192.
(16) Horng, J.-C.; Raleigh, D. P. *J. Am. Chem. Soc.* **2003**, *125*, 9286.
(17) Montclare, J. K.; Son, S.; Clark, G. A.; Kumar, K.; Tirrell, D. A. *ChemBioChem* **2009**, *10*, 84.
(18) Niemz, A.; Tirrell, D. A. *J. Am. Chem. Soc.* **2001**, *123*, 7407.
(19) Son, S.; Tanrikulu, I. C.; Tirrell, D. A. *ChemBioChem* **2006**, *7*, 1251.
(20) Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A. *Angew. Chem., Int. Ed.* **2001**, *40*, 1494.
(21) Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790.
(22) Tang, Y.; Tirrell, D. A. *J. Am. Chem. Soc.* **2001**, *123*, 11089.
(23) Wang, P.; Fichera, A.; Kumar, K.; Tirrell, D. A. *Angew. Chem., Int. Ed.* **2004**, *43*, 3664.
(24) Wang, P.; Tang, Y.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 6900.
(25) Woll, M. G.; Hadley, E. B.; Mecozzi, S.; Gellman, S. H. *J. Am. Chem. Soc.* **2006**, *128*, 15932.
(26) Gottler, L. M.; de la Salud Bea, R.; Shelburne, C. E.; Ramamoorthy, A.; Marsh, E. N. G. *Biochemistry* **2008**, *47*, 9243.
(27) Gottler, L. M.; Lee, H.-Y.; Shelburne, C. E.; Ramamoorthy, A.; Marsh, E. N. G. *ChemBioChem* **2008**, *9*, 370.
(28) Meng, H.; Krishnaji, S. T.; Beinborn, M.; Kumar, K. *J. Med. Chem.* **2008**, *51*, 7303.
(29) Meng, H.; Kumar, K. *J. Am. Chem. Soc.* **2007**, *129*, 15615.
(30) Buer, B. C.; Meagher, J. L.; Stuckey, J. A.; Marsh, E. N. G. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 4810.

(31) Anderson, J. T.; Toogood, P. L.; Marsh, E. N. G. *Org. Lett.* **2002**, *4*, 4281.
(32) Buer, B. C.; Chugh, J.; Al-Hashimi, H. M.; Marsh, E. N. G. *Biochemistry* **2010**, *49*, 5760.
(33) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Intl. J. Pept. Protein Res.* **1992**, *40*, 180.
(34) Kuhlman, B.; Raleigh, D. P. *Protein Sci.* **1998**, *7*, 2405.
(35) Yi, Q.; Scalley, M. L.; Simons, K. T.; Gladwin, S. T.; Baker, D. *Folding Design* **1997**, *2*, 271.
(36) Prabhu, N. V.; Sharp, K. A. *Annu. Rev. Phys. Chem.* **2005**, *56*, 521.
(37) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
(38) Robertson, A. D.; Murphy, K. P. *Chem. Rev.* **1997**, *97*, 1251.
(39) Myers, J. K.; Pace, N. C.; Scholtz, M. J. *Protein Sci.* **1995**, *4*, 2138.
(40) Chothia, C. *Nature* **1974**, *248*, 338.
(41) Eriksson, A.; Baase, W. A.; Zhang, X.; Heinz, D.; Blaber, M.; Baldwin, E. P.; Matthews, B. *Science* **1992**, *255*, 178.
(42) Biffinger, J. C.; Kim, H. W.; DiMagno, S. G. *ChemBioChem* **2004**, *5*, 622.
(43) Müller, K.; Faeh, C.; Diederich, F. *Science* **2007**, *317*, 1881.
(44) Betz, S. F.; DeGrado, W. F. *Biochemistry* **1996**, *35*, 6955.
(45) Gerstein, M. *Folding Design* **1998**, *3*, 497.
(46) Robic, S.; Guzman-Casado, M.; Sanchez-Ruiz, J. M.; Marqusee, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11345.
(47) Wallgren, M.; Aden, J.; Pylypenko, O.; Mikaelsson, T.; Johansson, L. B. A.; Rak, A.; Wolf-Watz, M. *J. Mol. Biol.* **2008**, *379*, 845.
(48) Murphy, K. P.; Privalov, P. L.; Gill, S. J. *Science* **1990**, *247*, 559.
(49) Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8069.
(50) Sawle, L.; Ghosh, K. *Biophys. J.* **2011**, *101*, 217.