JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

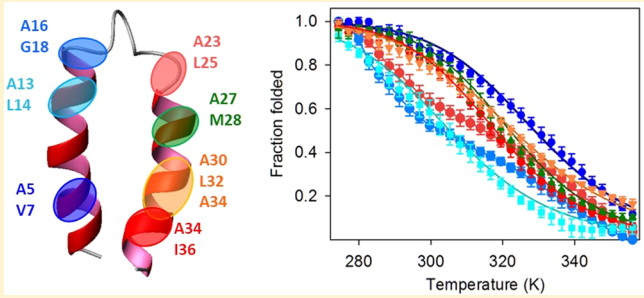
Site-Specific Thermodynamic Stability and Unfolding of a *de Novo* Designed Protein Structural Motif Mapped by ^{13}C Isotopically Edited IR Spectroscopy

Ginka S. Kubelka and Jan Kubelka*

Department of Chemistry, University of Wyoming, Laramie, Wyoming 82071, United States

Supporting Information

ABSTRACT: The mechanism of protein folding remains poorly understood, in part due to limited experimental information available about partially folded states. Isotopically edited infrared (IR) spectroscopy has emerged as a promising method for studying protein structural changes with site-specific resolution, but its full potential to systematically probe folding at multiple protein sites has not yet been realized. We have used ^{13}C isotopically edited IR spectroscopy to investigate the site-specific thermal unfolding at seven different locations in the *de novo* designed helix-turn-helix protein *αtα*. As one of the few stable helix-turn-helix motifs, *αtα* is an excellent model for studying the roles of secondary and tertiary interactions in folding. Circular dichroism (CD) experiments on the full *αtα* motif and its two peptide fragments show that interhelical tertiary contacts are critical for stabilization of the secondary structure. The site-specific thermal unfolding probed by ^{13}C isotopically edited IR is likewise consistent with primarily tertiary stabilization of the local structure. The least thermally stable part of the *αtα* motif is near the turn where the interhelical contacts are rather loose, while the motif's center with best established core packing has the highest stability. Similar correlation between the local thermal stability and tertiary contacts was found previously for a naturally occurring helix-turn-helix motif. These results underline the importance of native-like tertiary stabilizing interactions in folding, in agreement with recent state-of-the art folding simulations as well as simplified, native-centric models.



INTRODUCTION

Much of our current understanding of protein folding has been gathered from studies on small proteins, subdomains, and *de novo* designed mini-proteins. Small proteins are especially attractive for theoretical simulations of folding, particularly if they fold very fast. Recent successes of many computer models in simulating folding to the correct native states[1−6] may seem to suggest that for these small proteins the folding problem has been solved. However, reliable convergence to the folded state does not guarantee correct prediction of the folding mechanism. In fact, many successful folding simulations provide contradictory answers about folding pathways.[7−9] Therefore, while the problem of the native structure may appear solved, the folding process itself remains very much an open question.

The main reason why our knowledge of the folding mechanism remains sketchy is that the information available from experiments about other than fully folded proteins has been severely limited. For example, the majority of experiments on the folding of small proteins focus on fast kinetics, but typically rely on nonspecific spectroscopic methods to monitor the progress of folding.[10] While such experiments can provide valuable insights into the folding process, e.g., through studying various effects on the kinetic rates, they typically offer very little structural detail. As a consequence, it is difficult to accurately reconstruct the folding pathways from the experimental data and to validate the folding mechanism predicted by simulations.

Detailed characterization of the folding mechanism requires high-resolution probes that can report on local structural changes at specific locations within the protein. Isotopically edited infrared (IR) spectroscopy has emerged as one of the experimental methods capable of providing such high-resolution structural information. It takes advantage of the frequency shift of the amide I′ band (mostly C═O stretch, prime denotes N-deuterated amides in D_2O) upon the isotopic substitution by ^{13}C or $^{13}C═^{18}O$ on selected amide groups and has been extensively used to study peptide aggregation and fibril formation[11−16] as well as solvation.[17−19] Applications to folding have been mostly limited to *α*-helical oligopeptides[20−23] or short tryptophan zipper *β*-hairpins.[24−27] Notable exceptions are the villin headpiece subdomain (HP-35), whose folding was investigated with a single $^{13}C═^{18}O$ labeled residue,[28] and the N-terminal subdomain of protein L9 (NTL9) where a $^{13}C═^{18}O$ and an engineered azide label where used to probe, respectively, the backbone and side-chain environment of a single methionine.[29]

Received: January 27, 2014
Published: March 31, 2014

   dx.doi.org/10.1021/ja500918k | *J. Am. Chem. Soc.* 2014, 136, 6037−6048

While $^{13}C{=}{=}^{18}O$ labeling provides a greater isotopic shift and therefore better separation between the labeled and unlabeled signals, it is significantly more costly and typically limited to a single or very few labeled amino acids. On the other hand, with ^{13}C editing it is possible to introduce a much greater number of labeled amino acids, not only at multiple different positions along the sequence but also positioned in pairs or triplets in close proximity. The latter allows for the vibrational coupling of the ^{13}C-labeled amide oscillators,[30] which is key to the sensitivity of the amide I′ IR to the polypeptide backbone conformation (secondary structure). Therefore, in addition to providing probes of local structure in a number of multiple sites, the observed spectral changes are also straightforward to interpret: they arise due to changes of the local secondary structure. However, other than the preliminary report from this laboratory,[31] a systematic investigation of folding incorporating pairs and triplets of ^{13}C labels at multiple positions throughout the protein has not been reported.

This preliminary study[31] focused on a 40-residue helix-turn-helix subdomain of a viral coat protein P22.[32] We showed that each of the four isotopically labeled segments, probed by the ^{13}C amide I′, unfolded via a different thermodynamic transition, from the least stable N-terminus to the helical structure near the turn, all of which were distinct from the overall unfolding followed by circular dichroism (CD) and (unlabeled) IR amide I′. This signifies structurally heterogeneous unfolding, which is being reported increasingly often for small proteins studied by multiple, site-specific experimental probes.[28,33−38] The order of the local stabilities of the α-helices, consistent with the dominant role of interhelical tertiary interactions in their stabilization as judged by the absence of any helical structure in the fragments[31] provides important insights into the folding pathways. Recent ultralong folding simulations on small, fast folding proteins[4] suggest that folding generally follows a defined route of sequential stabilization of secondary structure by tertiary contacts, starting from the most stable element that tends to persist in the unfolded state. This is consistent with a body of experimental data from hydrogen/deuterium exchange (HDX),[39] one of the few site-specific methods, on larger proteins. The principle of sequential stabilization has also recently been implemented as a general scheme for simulation of folding pathways.[5] Therefore, probing thermodynamic stabilities using multiple structural probes,[35,40] in particular site-specific ones that are used to sense the local structure throughout the whole protein,[36,38] is key to characterizing the folding mechanism.

Here we apply isotopically edited IR to investigate the site-specific thermal unfolding of another helix-turn-helix motif, the *de novo* designed 38 residue $\alpha t\alpha$.[41] Helix-turn-helix motifs were chosen as models for studying protein folding for several reasons. Also termed two-helix bundles or helical hairpins, they are the simplest α-helical structures that combine both secondary and tertiary elements and contain all the fundamental interactions responsible for stabilizing protein structures. They represent the next step up in the protein structural hierarchy from α-helices, which still remain intensely studied models for elementary processes in protein folding. An equivalent motif in β-sheet proteins, the β-hairpin, has been a prototype of protein folding studies for a decade and a half.[42] However, unlike α-helices, the β-strands are not autonomous secondary structural elements: they require neighboring β-strands for hydrogen bonding. Investigation of the folding mechanism of helix-turn-helix motifs should therefore provide interesting insights into the roles of secondary and tertiary interactions and their interplay[4,5] in stabilization of protein structures and folding pathways. As the basic building units of larger α-helical domains, helix-turn-helix motifs may even constitute the critical early intermediates facilitating the folding of helical domains.[43] In contrast to β-hairpins, only a handful of folding studies have been reported for these elementary helical motifs.[31,43−45]

The close similarity to the naturally occurring P22 subdomain helix-turn-helix motif is another reason for studying the *de novo* designed $\alpha t\alpha$. The importance of the overall folded structure topology versus specific sequence in determining the folding mechanism remains the subject of ongoing discussion,[46,47] as does the role of natural sequence evolution in contrast to the *de novo* protein design, particularly in relation to the folding cooperativity.[48−50] Comparison of site-specific unfolding between these two motifs can therefore offer new important insights into these fundamental questions. Finally, the $\alpha t\alpha$ motif in particular and autonomously stable helix-turn-helix motifs in general should be of interest for theoretical folding simulations.

Since our preliminary study on the P22 subdomain,[31] new methods for the analysis of protein thermal unfolding monitored by amide I′ IR spectra have been developed. In particular, it was necessary to take into account the temperature-dependent frequency shifts,[51−53] which frustrates the standard multivariate analyses. For this reason new methods that explicitly consider spectral frequency shifts named shifted multivariate spectra analysis (SMSA) and its variant parametric SMSA (pSMSA), which fits the data directly to a thermodynamic model, were developed and extensively tested.[54] Here, this method is applied for the first time to the analysis of thermal unfolding of a protein with multiple site-specific ^{13}C isotopic editing.

EXPERIMENTAL METHODS

Protein Synthesis and Purification. The protein $\alpha t\alpha$, DWLKARVEQE^{10}LQALEARGTD^{20}SNAELRAMEA^{30}KLKAEIQK, was synthesized by standard FMOC-based solid-phase peptide synthesis on a Tribute automated peptide synthesizer (Protein Technologies, Inc., Tucson, AZ). The termini were left uncapped, in contrast to the originally designed $\alpha t\alpha$ sequence[41] to more closely mimic natural proteins. Seven isotopically labeled variants (Table S1) were also synthesized using amino acids ^{13}C labeled on C=O. All ^{13}C labeled amino acids were purchased FMOC protected for peptide synthesis from Cambridge Isotopes Laboratories (Tewksbury, MA) except for methionine and isoleucine, which were only available unprotected. For these two the FMOC protecting group was attached using a standard protocol.[55] Purification of the synthetic peptides was performed by reverse-phase HPLC, and the purity of the final product was verified by MALDI-TOF mass spectrometry. Peptide fragments corresponding to the N-terminal helix (helix 1) (DWLKARVEQE^{10}LQALEAR) and the C-terminal helix (helix 2) (AELRAMEAKL^{10}KAEIQK) of $\alpha t\alpha$ were also synthesized and purified using the same methods.

Sample Preparation. To allow for direct comparison between the CD and IR spectral data, both types of experiments were conducted under the same solution conditions in a D_2O based buffer. Prior to sample preparation, amide protons in the $\alpha t\alpha$ and peptide samples were exchanged for deuterium by dissolving in D_2O, followed by lyophilization. HCl (~0.1 M) was added to the D_2O to simultaneously remove the residual trifluoroacetic acid (TFA) remaining after HPLC purification. For spectral measurements the samples were dissolved in a 100 mM D_2O phosphate buffer at pH 2.3 (uncorrected). Peptide concentrations for the CD measurements were 20 μM, determined spectrophotometrically, except for the helix 2 fragment, where it was

obtained by weight. Samples for IR experiments were prepared at ~3 mM concentrations by weight. Although the concentrations of CD and IR samples differ considerably, concentration effects due to intermolecular effects can be safely excluded, since $\alpha t\alpha$ is monomeric up to 5 mM,[56] and all thermal denaturation data were fully reversible with no signs of aggregation. CD (ellipticity at 222 nm) was used to verify that the thermal unfolding is concentration independent up to ~330 μM $\alpha t\alpha$.

CD Measurements. All CD measurements were performed on a Jasco J-815 spectropolarimeter using a 1 mm path length quartz cuvette. Spectra were collected every 5 °C from (nominally) 0 to 85 °C. The sample temperature was controlled by a Peltier device, and the exact sample temperature was calibrated with a thermocouple. At least two complete spectral sets were collected for each sample to ensure reproducibility. The peptide spectra were corrected by subtracting the CD traces of just the buffer measured in the same sample cell under identical conditions.

IR Measurements. IR spectra were collected on a Bruker Tensor 27 FTIR equipped with a RT-DLaTGS detector as an average of 256 scans at a nominal resolution of 4 cm^{-1}. Three independently prepared samples of the unlabeled as well as each of the ^{13}C isotopically labeled $\alpha t\alpha$ variants were measured to ensure reproducibility and allow statistical analysis of the results. The samples were placed in a custom-built temperature controlled cell with CaF_2 windows separated by a 50 μm Teflon spacer. The temperature was controlled by a liquid bath driven by the spectra acquisition software. All samples were measured every 3 °C from (nominally) 0 to 87 °C. The exact sample temperature was again calibrated using a thermocouple. Buffer spectra and those of dilute TFA were collected under identical conditions as those of the peptide samples for the subsequent subtraction. All spectra were corrected for the temperature variation of the optical path length.[57]

DATA ANALYSIS AND MODELING

All experimental data processing and modeling was performed in MATLAB (The MathWorks, Inc., Natick, MA) using programs written in-house.

CD Spectra. The CD data on $\alpha t\alpha$ unfolding were analyzed in two ways, both using a two-state thermodynamic model with temperature-dependent intensity baselines. First, the mean residue molar ellipticity at 222 nm was modeled, as it has been most frequently used for quantifying helical content, and its dependence on the helical length and temperature-dependent baselines are best established.[58] The expression for the observed ellipticity as a function of temperature (T) is

$$[\Theta]_{222}(T) = [\Theta]^{F}_{222}(T)X_F + [\Theta]^{U}_{222}(T)X_U \quad (1)$$

where X_F and X_U are molar fractions, respectively, of the folded and unfolded states:

$$X_F = 1 - X_U = \frac{1}{1+K} \quad (2)$$

$$K = \exp\left(-\frac{\Delta G}{RT}\right) = \exp\left[-\frac{\Delta H}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right)\right] \quad (3)$$

where K is the unfolding equilibrium constant, ΔG, ΔH, and T_m are the unfolding Gibbs free energy, enthalpy and midpoint transition temperature, respectively, and R is the universal gas constant. Since $[\Theta]_{222}$ is most sensitive to the α-helical content, folded $\alpha t\alpha$ can be approximated as equivalent of two helices of 14−15 and 10−11 amides in length (of 37 amides total).[41] For helical peptides the standard assumption is that the ellipticities are linearly temperature dependent:

$$[\Theta]^{F}_{222}(T) = [\Theta]^{F}_{222}(T_0) + \alpha_F(T - T_0)$$
$$[\Theta]^{U}_{222}(T) = [\Theta]^{U}_{222}(T_0) + \alpha_U(T - T_0) \quad (4)$$

where T is the temperature and T_0 is taken at 0 °C. Feasible limits for the parameters in eq 4 were estimated from the generally accepted dependence on helix length:[58]

$$[\Theta]^{F}_{222}(T) = [\Theta_\infty(T_0) + (\partial\Theta_\infty/\partial T)(T - T_0)](n - k_n)/n \quad (5)$$

where Θ_∞ is the mean residue molar ellipticity for an infinitely long α-helix, n is the number of helical amide groups, and k_n is a constant. Values for the parameters Θ_∞, $(\partial\Theta_\infty/\partial T)$ and k_n proposed by several groups[58−61] yield for the two helices in $\alpha t\alpha$ $(-22\,000 \le [\Theta]^{F}_{222}(0) \le -15\,000)$ deg·cm^2·dmol^{-1} and $(60 \le \alpha_F \le 120)$ deg·cm^2·dmol^{-1}·K^{-1}. For the "random coil" peptides the ellipticity is typically assumed to be positive at low temperature and decrease as the temperature is increased.[60,62] However, an increase with temperature is feasible in proteins due to residual structure; the unfolded ellipticity parameters were therefore left unconstrained. The model was fitted to experimental data by optimizing the parameters ΔH, T_m, $[\Theta]^{F}_{222}(0)$, $[\Theta]^{U}_{222}(0)$, α_F, and α_U by the standard nonlinear least-squares procedure.

The second approach for the analysis of the CD data is global fitting of the temperature-dependent spectra into a bilinear model:

$$\mathbf{D}_{\lambda,T} = \sum_{d=F,U} \mathbf{S}_{\lambda,d}\mathbf{C}_{d,T} + \mathbf{E}_{\lambda,T} \quad (6)$$

where $\mathbf{D}$ is the matrix of the experimental spectra, $\mathbf{S}$ is the matrix of the folded and unfolded state spectra, $\mathbf{C}$ of their respective temperature-dependent contributions, and $\mathbf{E}$ the errors, λ denotes wavelength, and T temperature. The $\mathbf{C}$ matrix contains temperature-dependent baselines:

$$C_{d,T} = [1 - b_d(T - T_0)]X_d \quad (7)$$

where b_d is the linear slope of the intensity (assumed to decrease with temperature) and X_d with $d = \{F, U\}$ are molar fractions of the folded and unfolded states given in eqs 2 and 3. The parameters of the model, ΔH, T_m, b_F, and b_U, were again optimized by the nonlinear least-squares minimization with the spectra $\mathbf{S}$ obtained by the linear least-squares fit at each iteration.

The thermodynamic model (eq 3) assumes that the enthalpy (ΔH) and entropy ($\Delta S = \Delta H/T_m$) are independent of temperature, which is equivalent to neglecting the unfolding heat capacity (ΔC_p). This assumption is common for small proteins where the ΔC_p is generally small and when it cannot be determined independently, as the unfolding curves obtained from spectroscopic methods are not sufficient to reliably estimate the ΔC_p.[24,28,31,36,54,63] We have estimated the ΔC_p for $\alpha t\alpha$ using several empirical methods[64−67] based on the change in solvent accessible surface area (SASA), calculated from the protein structure using ProtSA.[68,69] The estimated ΔC_p values (Table S2) are indeed small, even compared to other similarly sized helical proteins.[70] Nevertheless, to ensure that our modeling of unfolding thermodynamics is not significantly affected by the neglect of the heat capacity, we have also explicitly considered ΔC_p (assumed temperature independent) in the model, which changes the expression for the equilibrium constant (eq 3) to

dx.doi.org/10.1021/ja500918k | J. Am. Chem. Soc. 2014, 136, 6037−6048

$$K = \exp\left(-\frac{\Delta G}{RT}\right) = \exp\left\{-\frac{\Delta H_{T_m}}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right) + \Delta C_p\left[1 - \frac{T_m}{T} - \ln\left(\frac{T}{T_m}\right)\right]\right\} \quad (8)$$

where ΔH_{T_m} is the unfolding enthalpy at the transition midpoint. The ΔC_p was treated either as a free parameter or constrained to the mean value from the three empirical estimation methods (MF, SLF, and MPS) that we previously found to agree best with experiment.[70]

Amide I′ IR Spectra. The unfolding thermodynamics as probed by the amide I′ IR spectra was modeled using the recently developed SMSA/pSMSA methodology,[54] which allows global fitting of the amide I′ bands while explicitly accounting for the frequency (as well as intensity) shifts with temperature.

Since the SMSA/pSMSA model is applicable only to amide I′, it is necessary first to correct for the overlapping side-chain and C-terminal group IR bands.[71] Deconvolution utilizing the maximum entropy method (MEM) was used for this purpose since, among all other possible approaches, it is the method least affected by arbitrary choices of parameters as discussed in detail previously.[54] The MEM deconvolution was implemented based on Burg-related entropy without sign restriction, essentially as described by Lorenz-Fonfria and Padros,[72] but using the MATLAB Optimization Toolbox fmincon utility to minimize the objective function. The $\alpha t\alpha$ spectra were deconvolved with Pearson VII pseudo-Voigt functions[54] with the exponents M = 3.0 and width (fwhm) linearly increasing with temperature from 22 to 24 cm^{-1}. A nonlinearity entropy factor (NEF) of 3.0 was used. The noise in the experimental spectra was estimated as RMSD of repeated FTIR scans of D_2O to be ~1.2 × 10^{-4} au in the region of interest (1800−1500 cm^{-1}) for 256 scans. Assuming the same average noise value for the sample, after the D_2O subtraction the standard deviation of the absorbance signal was taken as $\sigma = \sqrt{2} \times 1.2 \times 10^{-4}$ au ≈ 1.7 × 10^{-4} au. The deconvolution of the experimental spectrum by MEM method is illustrated in Figure 1.

The side-chain correction was performed by truncating the deconvolved spectrum at points of minimum absorbance between the amide I′ and side-chain/terminus peaks and subsequent convolution with the same line shape (Figure 1). Simple truncation is possible due to the great degree of narrowing provided by the MEM deconvolution; other possible approaches, such as approximating the "tails" of the deconvolved peaks by fitting or interpolation result in only negligible changes (on the order of σ) in the final reconstructed amide I′. The truncation also naturally corrects for a baseline originating from overlapping "tails" of the absorption bands to the lower frequency, which are also deconvolved by the MEM procedure (Figure 1). On the other hand, this procedure only eliminates contributions of the side-chain groups that are sufficiently separated from the amide modes to be resolved by the MEM deconvolution, i.e., carboxylic groups of acidic residues (D, E) as well as the C-terminus and guanidinium group of R, but not of amide groups of N and Q. Since subtraction of these would require estimation based on empirical amino acid data[71] or band fitting, again in the interest of minimizing arbitrary manipulations of the data, these were left uncorrected. The contribution of these groups (1xN, 3xQ) is, however, relatively small compared to 37 backbone

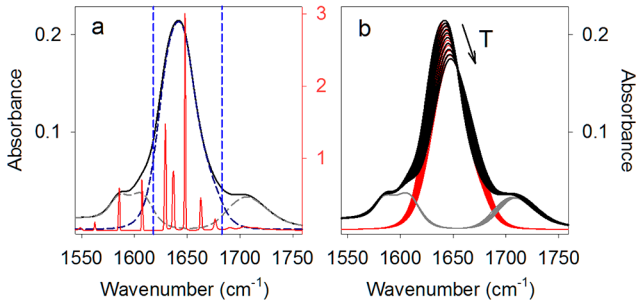


Figure 1. Correction of the experimental amide I′ spectrum of the unlabeled $\alpha t\alpha$ for side-chain and C-terminal carboxylic group bands by MEM deconvolution. (a) Experimental spectrum (solid black) was deconvolved using MEM method (red). The side-chain and terminal peaks (neutral and charged carboxylic groups at ~1715 cm^{-1} and ~1580 cm^{-1}, respectively, and guanidinium group of arginine at ~1605 cm^{-1}) were separated by truncation at points of minimum absorption (vertical dashed blue lines). The side-chain and terminal peaks along with the residual baseline were obtained by reconvolution (dashed gray) and subtracted from the original spectrum to obtain the amide I′ (dashed black). (b) Temperature-dependent experimental spectra (black), side-chain, terminus, and baseline (gray) and amide I′ after correction (red).

amides, and most importantly, the side-chains do not affect the analyses of ^{13}C isotopically edited IR data (see below).

The corrected amide I′ was modeled using the parametric shifted multivariate spectra analysis (pSMSA) method, described in detail elsewhere.[54] In brief, the bilinear scheme (eq 6) is generalized to allow for frequency shifts of the spectra:

$$\mathbf{D}_{\nu,T} = \sum_d \mathbf{S}_{\nu-\delta\nu_{d,T},d}\mathbf{C}_{d,T} + \mathbf{E}_{\nu,T} \quad (9)$$

where $\delta\nu_{d,T}$ are the frequency shifts, effected by multiplication by a factor $e^{i2\pi t\delta\nu_{d,t}}$ in the time domain. The frequency shifts of amide I′ are linear:[51]

$$\delta\nu_{d,T} = a_d(T - T_0) \quad (10)$$

where a_d is the slope of the frequency shift with temperature for the spectrum d, $d = \{F, U\}$. The matrix $\mathbf{C}$ is again given by eq 7, and the model is optimized by finding the parameters of the two-state thermodynamic model (eq 3) along with the intensity slopes b (eq 7) and frequency slopes in vector a (eq 10) that minimize the squared error between model and data. The frequency slopes a are constrained to be nonnegative (amide I′ frequencies shift higher with temperature) and to not exceed the maximum value measured for NMA (0.08 cm^{-1}·K^{-1}).[52] The intensity slopes b are likewise kept positive and less than 5 × 10^{-3} K^{-1}; this value is greater than the maximum measured intensity decrease with temperature measured for amide I′ of NMA in D_2O (~2 × 10^{-3} K^{-1}) since the intensity baselines may also need to account for additional effects.[54]

Isotopically Edited Amide I′ Spectra. Site-specific unfolding is probed by the changes in the amide I′ of the ^{13}C labeled amide groups (Figure 2), which was again modeled using SMSA/pSMSA methods.[54] The ^{13}C amide I′ bands were isolated as the positive part of the difference spectra of ^{13}C $\alpha t\alpha$ isotopologues minus the spectra of the unlabeled protein.[31,73] Due to the relatively small difference signals, significant errors may arise due to small intensity variations from, e.g., incompletely corrected optical path length changes,[57] imperfect buffer subtraction, and other effects. Since side-chain IR bands are independent of isotopic labeling, they can serve as an

dx.doi.org/10.1021/ja500918k | J. Am. Chem. Soc. 2014, 136, 6037−6048

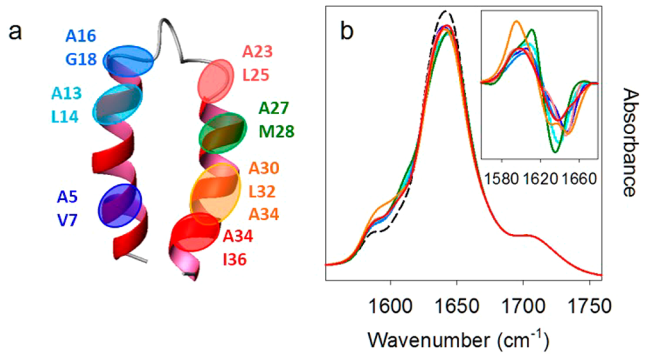


Figure 2. Studied isotopologues of *ατα* and their amide I′ spectra. (a) The *ατα* motif with highlighted ^{13}C isotopically labeled amino acids in seven different locations, color-coded to match the spectra in b). (b) Experimental amide I′ spectra of the seven isotopically labeled variants and the unlabeled *ατα* spectrum (dashed black) at the lowest measured temperature (0.5 °C). Inset: difference spectra of each labeled *ατα* with the unlabeled spectrum subtracted. For calculation of the difference spectra the intensity was normalized to the ∼1715 cm^{-1} band (due to neutral carboxylic groups).

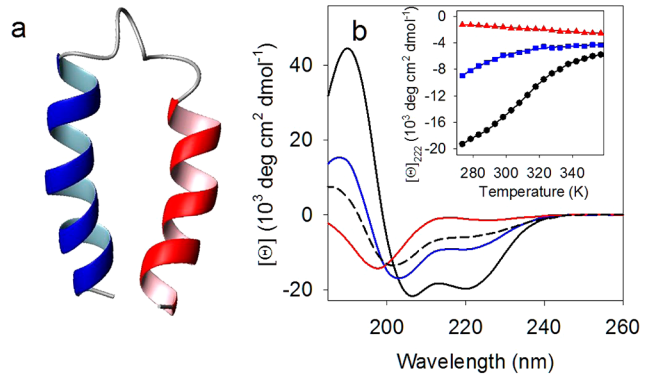


Figure 3. Stability of the *ατα* motif and its two fragments corresponding to the individual helices. (a) Structure of the *ατα* motif with the N-terminal (blue) and C-terminal (red) helices, whose corresponding sequences were synthesized and their stability studied by CD. (b) CD spectra of the full *ατα* motif (solid black), the N-terminal fragment (solid blue), and the C-terminal fragment (solid red) at the lowest measured temperature (0.5 °C). Dashed black spectrum is the *ατα* motif at the highest temperature (84 °C). Inset: mean residue molar ellipticity at 222 nm as a function of temperature for the full motif (black), N-terminal fragment (blue), and C-terminal fragment (red).

"internal standard" to correct for these artifacts. The spectra were normalized to the (neutral) carboxylic group band at ∼1715 cm^{-1} (Figures 1 and 2) simultaneously with subtraction of the residual TFA absorption and a cubic polynomial correction for small residual baseline. Although simply discarding the negative difference may truncate the ^{13}C spectrum due to overlapping signal cancellation, it was adopted as the way of approximating the ^{13}C band that is least affected by arbitrary assumptions. Other approaches, e.g., band fitting or even deconvolution, as described above for side-chain signal correction, were considered but abandoned as they necessarily yield variable results depending on the choice of parameters.

Separated ^{13}C amide I′ spectra as a function of temperature were then fitted using pSMSA method as described in the previous section, again assuming a two-state model. An example of the analysis is shown in Figure S1. For each of the seven isotopically edited variants, three sets of data, obtained from independently prepared and measured samples, were analyzed. The pSMSA was able to fit the data with a high level of confidence and reproducibility for all ^{13}C labeled *ατα* variants, except two, both labeled in the turn region (^{13}C A16/G18 and A23/L25). For these two isotopologues, the SMSA was therefore used assuming two states, which in all cases produced excellent fits. The fractions of the two states determined by SMSA were arbitrarily normalized to range from 0 to 1.

RESULTS

Stability of the *ατα* Motif and Its Fragments. First we compared the stability of the complete *ατα* motif with that of the peptide fragments corresponding to the individual helices using CD. The CD spectra of the *ατα* and the two fragments at the lowest temperature (0.5 °C) are shown in Figure 3. The mean residue molar ellipticity at 222 nm as a function of temperature for all three peptides is shown in the inset. It is evident that the fragment corresponding to the C-terminal helix is essentially structure-less, and the CD also shows no evidence of melting of a helix with temperature; in fact the $[\Theta]_{222}$ decreases slightly with increasing temperature. By contrast, a small amount of residual helical structure is present at low temperature in the peptide corresponding to the N-terminal helix. However, the residual helix amount is comparable to that in the unfolded (high temperature) complete motif: although the latter has less negative $[\Theta]_{222}$ (−5800 deg·cm^2·$dmol^{-1}$) than the N-terminal fragment at 0.5 °C (−8900 deg·cm^2·$dmol^{-1}$), the *ατα* ellipticity is normalized to the total number of residues, not all of which are helical when folded. If the per residue normalization considered only 23−25 helical residues, based on the NMR structure,[41] the ellipicities for the unfolded *ατα* and for the N-terminal fragment at low temperature would be nearly the same.

The original study on the capped *ατα* motif[56] reported greater residual helical structure in both capped fragments, but particularly in the C-terminal peptide, which was found to have comparable or even slightly greater amount of helix than the N-terminal one. This, along with the systematically higher CD ellipticities for the capped sequences[56] might be explained by the stabilizing effects of the N- and C-terminal capping groups. On the other hand, the thermal melting curves for the *ατα* and the N-terminal fragment[56] are very similar to the present data.

These results demonstrate that the stability of the *ατα* motif cannot be attributed to the independent folding of the constituent *α*-helices. Rather, the helical structure exists primarily due to stabilization by tertiary, interhelical interactions. The dependence of the helical structure on the tertiary contacts is even more pronounced in this uncapped sequence than in the original *ατα* with N- and C-terminal protecting groups,[56] which had significant amounts of *α*-helix in both separate fragments. On the other hand, since the N-terminal helix appears to retain some residual helical structure, it is a likely candidate for the initiation site for folding.[4]

Global Unfolding Thermodynamics. The overall thermal unfolding of the *ατα* motif was measured by CD and amide I′ IR spectroscopies, for consistency both carried out in D_2O based buffers. The results are summarized in Figure 4 and the resulting thermodynamic parameters from a two-state model fit are listed in Table 1.

The two-state model yields an excellent fit to the mean residue molar ellipticity (Figure 4b) with the folded baseline consistent with the estimated limits considering the two *α*-

dx.doi.org/10.1021/ja500918k | J. Am. Chem. Soc. 2014, 136, 6037−6048

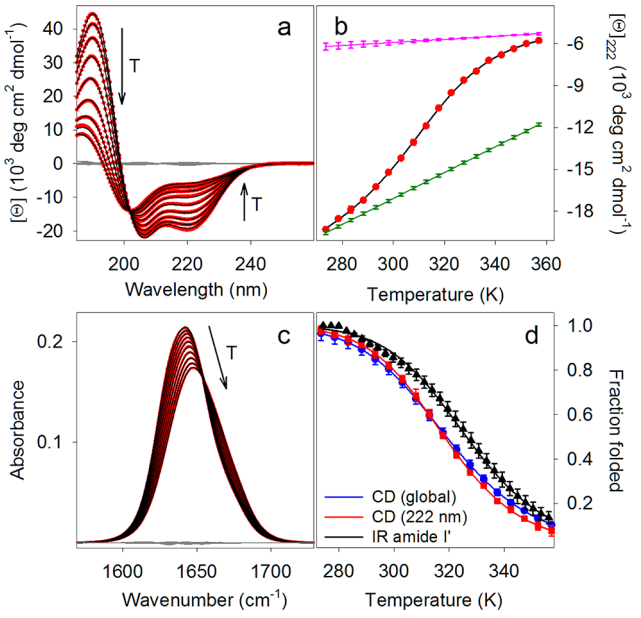


Figure 4. Global thermal unfolding of the *αtα* motif probed by CD and amide I′ IR. (a) CD spectra (red dots) and global fit (black lines) to a two state model with intensity baselines. Gray lines are the residuals of the fit. (b) Mean residue molar ellipticity at 222 nm (red circles) and fit to a two-state model (black line) with folded (green) and unfolded (pink) baselines. (c) Amide I′ IR after correction for side-chain and terminal absorption (red) and pSMSA model fit (black). Residuals are shown in gray. (d) Comparison of the unfolding transitions obtained from the global fit to the CD data (blue), CD ellipticity at 222 nm (red), and IR amide I′ (black).

Table 1. Parameters of Global Thermal Unfolding of the *αtα* Motif[a]

parameter	$[\Theta]_{222}$[b]	CD spectra global[c]	amide I′ IR pSMSA[d]
ΔH (kJ·mol^{-1})	61 ± 3	53 ± 4	62 ± 3
T_m (K)	318 ± 1	318 ± 1	326 ± 2
$[\Theta^F_{222}(T_0)]$ (deg·cm^2·dmol^{-1})	−19500 ± 100	–	–
$[\Theta^U_{222}(T_0)]$ (deg·cm^2·dmol^{-1})	−6200 ± 400	–	–
α_F (deg·cm^2·dmol^{-1}·K^{-1})	92 ± 2	–	–
α_U (deg·cm^2·dmol^{-1}·K^{-1})	12 ± 6	–	–
b_F (10^{-3} K^{-1})	–	2.2 ± 0.2	1.52 ± 0.02
b_U (10^{-3} K^{-1})	–	0.8 ± 0.3	1.5 ± 0.1
a_F(10^{-2} cm^{-1}·K^{-1})	–	–	6.5 ± 0.5
a_U(10^{-2} cm^{-1}·K^{-1})	–	–	7.8 ± 0.2

[a]Assuming temperature independent enthalpy and entropy of unfolding (ΔC_p of unfolding = 0). Unfolding parameters with explicit consideration of ΔC_p are shown in Table S3. [b]CD mean residue molar ellipticity at 222 nm with linear temperature-dependent baselines, eqs 1−4. [c]CD global fit with linear intensity baselines (slopes b_F, b_U), eqs 6 and 7. [d]Eq 8, linear temperature dependence of intensity (b_F, b_U) and frequency (a_F, a_U) given by eqs 7 and 10, respectively.

helices (Data Analysis and Modeling section). The unfolded ellipticity is negative, and the baseline has positive slope, which is opposite to what is generally accepted for the "random coil" peptides, but consistent with residual helical structure in the unfolded state evident from the shape of the CD (Figure 4a) as well as from the ellipticity (Figures 3 inset and 4b). By contrast, the global fit to the CD data (Figure 4a) is statistically not very good (reduced sum of squares $\chi_\nu^2 \cong 1.2$ and associated $P_\chi(\chi^2, \nu) \cong 10^{-13}$), but the estimated thermodynamic parameters are very consistent between the two sets of CD data (Table 1) and, within error, essentially the same as those found from the CD ellipticity at 222 nm (Figure 4d). The low confidence of the fit may be in part due to intrinsic spectra changes or shifts with temperature, similar to the IR (see below).

The pSMSA modeling (eq 9) of the amide I′ IR is shown in Figure 4c. The resulting folded and unfolded state spectra and their contributions, along with the temperature-dependent intensity baselines as given by the model, are shown in Figure S2. The resulting unfolding transition is shown in Figure 4d, and the parameters are summarized in Table 1. Obviously, the unfolding curves obtained from CD and IR are significantly different, with the temperature midpoints ∼8 K apart. As detailed in the Experimental Methods section, despite very different protein concentrations required by these two techniques, any intermolecular effects that could cause concentration-dependent behavior are extremely unlikely. The "probe-dependent" unfolding thermodynamics, which has been previously observed in several small proteins,[28,33−36,38,74] must therefore come from complex, structurally heterogeneous processes. This is further supported by the statistically low quality of the two-state pSMSA fit to the amide I′ ($\chi_\nu^2 \cong 1.5$ and associated $P_\chi(\chi^2, \nu) \approx 0$).

Explicit consideration of the unfolding heat capacity (ΔC_p) in the thermodynamic model does not significantly affect the results, as shown in Figure S3 and Table S3. If left as an adjustable parameter, modeling of the CD spectra yields negligible ΔC_p (identically zero for the global CD analysis) and, consequently, the same thermodynamic parameters as presented above. Constraining the ΔC_p to the mean of the three empirical estimates ∼560 J·mol^{-1}·K^{-1} (Table S2) shifts the unfolding to ∼5K lower temperatures (Figure S3 and Table S3), in addition to giving a statistically worse fit. Interestingly, the optimum ΔC_p obtained from the amide I′ IR data happens to agree with the empirical estimates based on the ΔSASA (Table S2), but the unfolding enthalpy (at the transition midpoint) and the transition midpoint temperature are the same, within error, as with zero heat capacity (Table 1).

Site-Specific Unfolding Thermodynamics. Site-specific thermal unfolding of the *αtα* motif was investigated with isotopically edited IR using seven ^{13}C labeled variants (Figure 2 and Table S1). For five isotopologues labeled on the two *α*-helices, the local unfolding as probed by the changes in the ^{13}C amide I′ was successfully modeled by pSMSA using a two-state model. The analysis is illustrated in Figure S1 for one of the isotopologues. The results are summarized in Figure 5.

Figure 5b−f shows the ^{13}C amide I′ spectra for each labeled *αtα* variant, obtained from the differences with the unlabeled spectra (Figure S1), along with the fits of the pSMSA model. In all cases, statistically the quality of fits (Figure 5b−f) was excellent within the estimated uncertainties ($P_\chi(\chi^2, \nu) \gtrsim 0.5$). The positions of the ^{13}C amide I′ bands are consistent with expected *α*-helical signatures for the particular arrangement of labels,[30] arising from the labeled amide I′ normal mode distributions and their intensities as a consequence of the vibrational coupling. Likewise, the shifts with temperature correspond to what is expected for the transition to an unordered structure. Specifically, the bands arising from ^{13}C amino acids in neighboring positions (sequential) are higher in frequency and more intense than the next-nearest neighbor

dx.doi.org/10.1021/ja500918k | J. Am. Chem. Soc. 2014, 136, 6037−6048

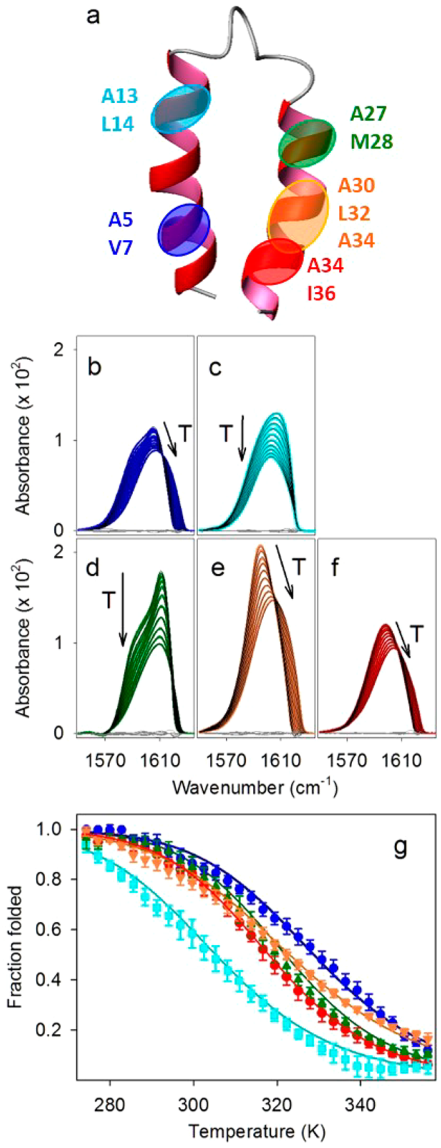


Figure 5. Site-specific thermal unfolding of the *αtα* helices from ^{13}C isotopically edited amide I′ IR spectroscopy. (a) Structure of *αtα* motif with highlighted positions of the ^{13}C isotopic labels. The colors correspond to the ^{13}C amide I′ spectra in (b–f) and the unfolding curves in (g). (b–f) Amide I′ spectra (positive parts of the differences labeled–unlabeled) for the individual *αtα* isotopologues (colored lines matching the positions of the isotopic labels in (a), the two-state pSMSA model fits (black) and residuals (gray). (g) Unfolding transitions (fraction of the folded state) for the individual isotopically labeled regions.

(alternate) label ones. Helix unfolding causes the ^{13}C amide I′ for sequential labels to lose more intensity, but no significant overall shift of the band to higher frequency, while alternately labeled helix bands always shift higher. This is clearly seen in Figure 5 for both sequential (A13/L14 and A27/M28; Figure 5c,d, respectively) and three alternate ^{13}C substitutions (A5/V7, A30/L32/A34 and A34/I36; Figure 5b,e,f respectively) as well as from the quantitative comparison of the ^{13}C amide I′ frequencies at the lowest and highest measured temperatures for all *αtα* isotopologues (Table S4). In fact, if peak positions are considered, the sequentially ^{13}C labeled amide I′ appears to shift down upon unfolding (Table S4), although comparison of peak frequencies is somewhat misleading due to varying band shapes. These general patterns illustrate the usefulness of pairs (or triplets) of isotopic labels, as the changes due to their coupling allow for tracing of changes in the local helical structure.

The frequency shifts with temperature obtained from the pSMSA model are likewise consistent with expectations. In analogy to the unlabeled amide I′ (previous section) generally the folded state spectrum shifts less with temperature than that of the unfolded state; the latter is expected to have more solvent exposed amides.[51] The exception is the A27/M28 labeled protein, where the shifts are approximately the same. In all alternately labeled positions the slopes of the unfolded state frequency shifts with temperature correspond to those found for NMA,[51] while in both sequentially labeled cases (A13/L14 and A27/M28) the unfolded frequency shifts are smaller. This may suggest some correlation between the frequency shifts and the label arrangements, possibly due to residual vibrational coupling between sequential amides in the unfolded states. However, since a variety of effects influence the frequency slopes, particularly the solvent exposure of the amide groups,[17,18,53] along with its possible change during unfolding, it is difficult to justify any such patterns with certainty.

The unfolding curves obtained from the pSMSA model are overlaid in Figure 5g, and the resulting thermodynamic parameters are summarized in Table 2. The differences in the thermodynamic stability in the various labeled segments are apparent. The most stable part of the structure is the center of the N-terminal helix (labeled at A5/V7), next is the middle segment of the C-terminal helix (A30/L32/A34), followed by its N-terminus and the C-terminus, respectively. While the three labeled parts of the C-terminal helix show very similar unfolding, with the midpoint transition temperatures within ~4 K of each other, the differences are significant, in particular between the least stable C-terminus and the other two regions. The global unfolding transition, as obtained from the unlabeled amide I′ (Figure 4), falls in between those of the A5/V7 and

Table 2. Parameters of Site-Specific Thermal Unfolding of the *αtα* Motif[a,b]

parameter	A5/V7	A13/L14	A27/M28	A30/L32/A34	A34/I36	A16/G18	A23/L25
ΔH (kJ·mol^{-1})	64 ± 5	54 ± 4	64 ± 4	53 ± 4	65 ± 2	–	–
T_m (K)	329 ± 1	305 ± 1	322 ± 1	323 ± 1	318 ± 1	–	–
b_F (10^{-3} K^{-1})	3.6 ± 0.3	2.5 ± 0.2	3 ± 1	1.9 ± 0.4	1.4 ± 0.2	–	–
b_U (10^{-3} K^{-1})	2.7 ± 0.2	2.5 ± 0.5	4.6 ± 0.2	2.1 ± 0.3	2.3 ± 0.1	–	–
a_F (10^{-2} cm^{-1}·K^{-1})	4.9 ± 0.4	1.9 ± 0.5	3.6 ± 0.4	3.7 ± 0.2	3.1 ± 0.1	2.5 ± 0.6	2.3 ± 0.5
a_U (10^{-2} cm^{-1}·K^{-1})	7.7 ± 0.2	5 ± 1	3 ± 1	7.8 ± 0.1	7.8 ± 0.1	7.7 ± 0.2	7.7 ± 0.1

[a]Assuming temperature-independent enthalpy and entropy of unfolding (ΔC_p of unfolding = 0). Unfolding parameters with explicit consideration of ΔC_p are shown in Table S5. [b]Parametetric SMSA model, eq 9. Linear temperature dependence of intensity (b_F, b_U) and frequency (a_F, a_U) as determined by eqs 7 and 10, respectively.

dx.doi.org/10.1021/ja500918k | J. Am. Chem. Soc. 2014, 136, 6037–6048

A30/L32/A34 ones. The least stable of the measured segments is near the C-terminus of the N-terminal helix (A13/L14). This order of relative thermal stabilities is consistent with the higher residual structure found in the corresponding fragment (Figure 3), and the order of unfolding transitions has the "symmetry" expected from the stabilization by tertiary, interhelical contacts.

The site-specific unfolding was also modeled with explicit consideration of unfolding heat capacity (ΔC_p), and the results are presented in Figure S4, Table S5. The ΔC_p was treated as an adjustable parameter in all cases: while the global unfolding ΔC_p can be estimated (Table S2), it is not at all clear what to anticipate for the individual local transitions. Intuitively, since the global unfolding parameters are averages, some of the site-specific ΔC_p should be lower, and others higher than the overall unfolding heat capacity. As shown in Table S5, this appears to be the case: the values range from identically zero to somewhat higher than the empirical estimate (Table S2) or global ΔC_p found from the amide I′ IR (Table S3). Considerable error bars reflect the difficulty of consistently determining unfolding heat capacities from the unfolding curves, as expected, particularly for small signals. Like for the global unfolding, the heat capacity has little effect on the site-specific unfolding thermodynamics, giving the same order of stabilities (Figure S4) and very similar thermodynamic parameters (Table S5).

The unfolding of the last two isotopically labeled $\alpha t\alpha$ motifs, with ^{13}C at A16/G18 and A23/L25, is shown in Figure 6. For these two variants, it was not possible to fit the ^{13}C amide I′ spectra to the two-state thermodynamic model in pSMSA with any physically reasonable and consistent parameters. The most likely reason is that these parts of the motif exhibit a considerable degree of disorder, as evident from the NMR structure[41] as well as the HDX experiments.[75]

However, the changes observed in the amide I′ spectra (Figure 6b,c) do suggest some structural transitions. Indeed, without the constraints of the thermodynamic model (eqs 2 and 3), the two-state SMSA can consistently fit the data with very high confidence ($P_\chi(\chi^2, \nu) \approx 1$) and with frequency shift parameters consistent with expectations (Table 2).

The populations are not uniquely determined by SMSA model (as signified by the star in Figure 6d), as they involve arbitrary scaling (in Figure 6d they are normalized between 1 and 0) and also "rotational ambiguity,"[54] which means that the solution of eq 9 is specified only up to an arbitrary rotation matrix (pSMSA by virtue of forcing a thermodynamic model removes this ambiguity). Nevertheless, they do provide a qualitative picture of the transition probed by the ^{13}C amide I′ spectra: gradual with the most pronounced changes at low temperatures. The initial state of the N-terminal segment (A16/G18) appears to decay more rapidly than the C-terminal one (A23/L25), but overall the curves roughly parallel each other. Such correlation would again be consistent with the tertiary interactions between the segments stabilizing their local structure.

While without the model it is not possible to quantify the unfolding thermodynamics of these two segments, based on the known structural disorder it can be safely concluded that they are less stable than the other parts of the motif probed by ^{13}C isotopically edited IR. Furthermore, from the SMSA results in Figure 6, it appears likely that the A16/G18 labeled region is less stable than its opposite on the A23/L25 side of the turn.

Combined, the above data provide a site-specific map of the thermal stabilities within the studied motif, which is qualitatively illustrated in Figure 7. The unfolding commences

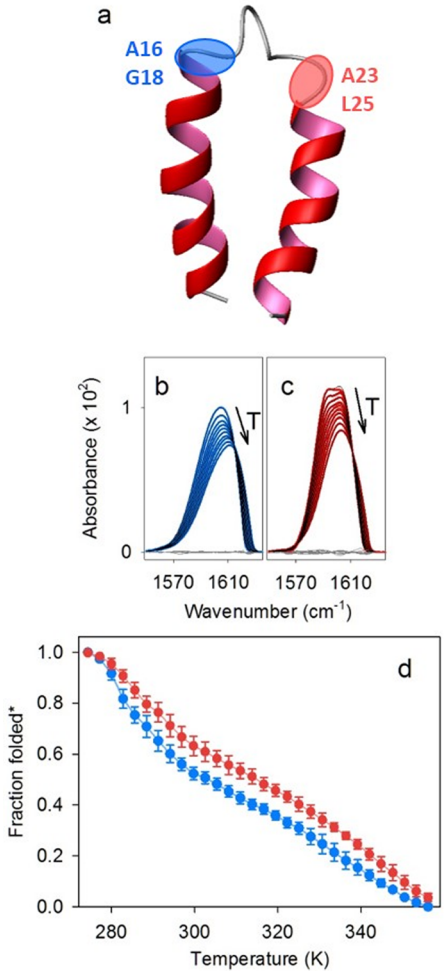


Figure 6. Site-specific thermal unfolding of the two $\alpha t\alpha$ segments near the turn from ^{13}C isotopically edited amide I′ IR spectroscopy. (a) Structure of $\alpha t\alpha$ motif with highlighted positions of the ^{13}C isotopic labels. The colors correspond to the ^{13}C amide I′ spectra in (b, c) and the unfolding curves in (d). (b, c) Amide I′ spectra (positive parts of the differences labeled−unlabeled) for the individual $\alpha t\alpha$ isotopologues (colored lines matching the positions of the isotopic labels in a), the two-state SMSA model fits (black), and residuals (gray). (d) Unfolding transitions for the individual isotopically labeled regions. The curves were arbitrarily normalized to unity, the fraction folded may therefore not correspond to the true fraction of the folded state (highlighted by the star).

from the turn region and its immediate vicinity, which are considerably disordered even in the maximally folded structure. The next least stable segment is the C-terminus of the N-terminal helix, which is partly opposite to the mostly disordered terminus of the second helix. Its unfolding is followed by the very C-terminus of the motif; the N-terminus, though not specifically probed, is likely to be more stable given the higher stability of the neighboring segment of the N-terminal helix. The most stable part of the motif is at its center, where the helices come to the closest contact, between residues V7 to L11 and A30 to A34 on the two helices. The first helix is more stable than the second one, consistent with the fragment results (Figure 3), but within the motif, its stability depends in large part on tertiary contacts with the partially unfolded second helix.

dx.doi.org/10.1021/ja500918k | J. Am. Chem. Soc. 2014, 136, 6037−6048

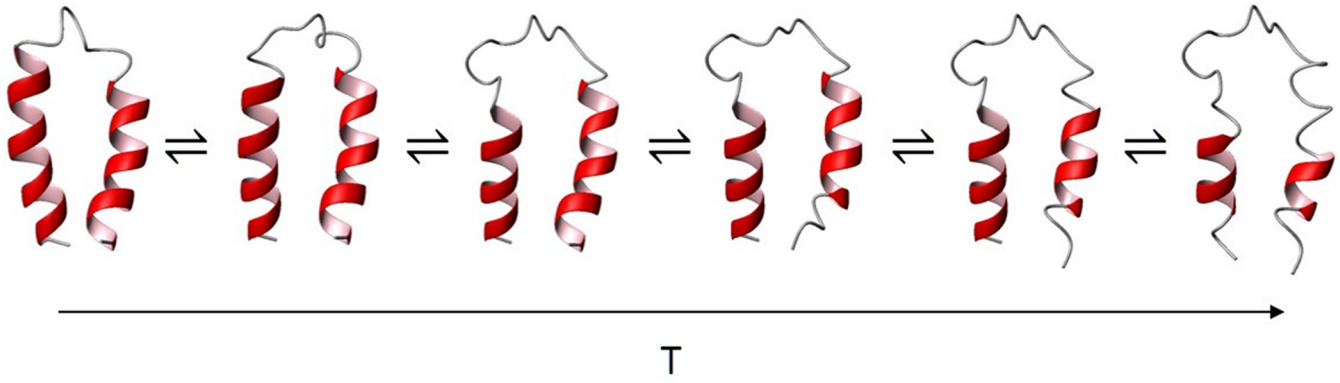


Figure 7. Cartoon representation of the progress of thermal unfolding of *αtα*.

DISCUSSION

The majority of small proteins are believed to undergo a cooperative, two-state folding transition.[76,77] Two-state folding by definition means that no intermediates are ever populated to be experimentally detected. An increasing number of experiments that employ multiple, independent structural probes is inconsistent with this simple picture.[28,33−39,74] Rather, folding appears to proceed through a sequence of intermediate steps[39] or even gradually through a continuum of effective intermediate states.[35−38,74] In some cases, additional experiments revealed more complex folding for proteins originally believed to fold in two-state like fashion,[38] and in others an apparent two-state folding can be turned to a heterogeneous process by varying experimental conditions.[78] Folding of the *αtα* motif, presented above, is clearly more complex than a simple two-state mechanism.

Thermal Unfolding of *αtα*. One of the experimental signatures of deviations from the two-state folding is the experimental probe dependence of the unfolding thermodynamics.[35,40] For the *αtα* motif, this becomes apparent already from the differences in the global unfolding (Figure 4) determined from the CD and IR. Furthermore, the statistical tests suggest that neither IR nor the CD (if analyzed globally) can be described by two thermodynamic states. However, although the two-state model must fail for the whole motif, it may very well be applicable to individual short segments. In the limit of a single peptide bond or residue, it is feasible that the configuration can be described either as native or unfolded, as, e.g., statistical mechanical Ising-like models of folding assume with considerable success.[63,79−81] It is therefore expected that the two-state description might be appropriate for very short segments but might differ between them, as is indeed observed here. Even for the two labeled parts neighboring the turn (Figure 6) two states were able to account for the observed changes, even though they could not be reconciled with the simple mass action model. This is largely because the thermodynamic model relies on the curvature and sigmoidal shape of the transition; for a gradual change it is essentially impossible to obtain a reliable fit. Unfortunately, this is somewhat of a problem, since without the constraints of the thermodynamic model the solutions are not unique due to the rotational ambiguity, although in the two-state case this often reduces to simple scaling.[54] Nevertheless, even the most fundamental parameters such as midpoint transition temperatures cannot be uniquely determined without such a model.

While the thermal unfolding of *αtα* has not been investigated with site-specific resolution, Xian et al. reported the HDX protection factors for the amide protons of a number of residues in the original capped motif at 18 °C.[75] The protection factors are directly related to the stability of the "closed" (i.e., hydrogen bonded or helical) state of the amide. These data suggest a somewhat different order of site-specific thermodynamic stabilities than do our data. The highest protection factors are found for helix 2 (A30), whereas the most protected residues in helix 1 are A13 and L14, significantly more than A5 on the same helix and also A34 on helix 2. By contrast, our thermal unfolding experiments show that the A5/V7 part of the helix 1 is the most stable, followed by the three labeled regions on helix 2 (A30/L32/A34, A27/M28 and A34/I36) and the A13/L14 segment of the helix 1. No protection was found for G18 and A23 close to the turn, which agrees with our results, supporting disorder and fraying of A16/G18 and A23/L25 labeled parts (Figure 6).

The HDX results are somewhat counterintuitive. First, they seem to be at odds with the NMR structural data: the RMSD from the average structure and the differences of the chemical shift from "random coil" values, both essentially indicating the degree of structural ordering.[41] These appear quite symmetric between the two helices, with the best defined structure in the middle and perhaps slightly less disorder in the N-terminal helix than the C-terminal one. On helix 1, positions A5, V7, A13, and L14 are all similar in RMSD, but the former two appear to have slightly more helical chemical shifts. Residues A34 and I36 have higher RMSD than A13 and L14, but, again, more helical chemical shifts. Second, the stabilities indicated by HDX do not reflect the symmetry between sites in interhelical contacts, expected from the dominant stabilizing role of the tertiary interactions.[75] For example, L14 should be opposite to and in contact with L25, yet their protection factors are very different. Similarly, V7 would be in contact with L32 and, although not measured, from the value for the neighboring A5 and the overall trend, the V7 protection factor would be significantly smaller. This asymmetry was explained by a register shift of the helices and supported by the statistical mechanical model[82] based on helix−coil transition theory augmented with tertiary interactions. Our data, on the other hand, are consistent with both the NMR (RMSD and chemical shift index) as well as the intuitively expected correspondence between the regions related through interhelical contacts.

One possible reason for the discrepancies between the HDX and our results may be that Xian et al.[75] effectively studied a different molecule: the capped version of *αtα*. As already discussed, in the capped sequence both helices, but especially the C-terminal one, retained significant residual structure in

dx.doi.org/10.1021/ja500918k | J. Am. Chem. Soc. 2014, 136, 6037−6048

isolation. In such case, the tertiary stabilization may be less important and perhaps allow different arrangements of the helices to account for the distribution of the protection factors. Another reason may be the obvious differences in the methodology: while HDX measures the stability at a particular temperature, our determination relies on measuring—and fitting—the entire transition. For the latter we necessarily rely on a thermodynamic model. As already pointed out above, the model depends on the curvature of the transition to determine the fractions of the folded and unfolded states and assumes closure, i.e., that both states must add to 100%. However, there is no guarantee that the initial state is fully folded, as best seen for the two turn flanking parts (Figure 6), where the two-state model fails completely. A close examination of the unfolding curves in Figure 5 (also the IR unfolding curve in Figure 4d) also suggests deviations from the two-state unfolding transitions, especially at low temperatures.

Implications for the Folding Mechanism. Even a simple helix-turn-helix motif, depending on the relative importance of specific stabilizing forces, can fold through a variety of possible scenarios. An obvious possibility is docking of two independently formed helices. At the other extreme is the nonspecific hydrophobic collapse, where the hydrophobic (tertiary) interactions form before any secondary structure. Another possible mechanism is "zipping" the structure down from the turn, in analogy to the proposed folding of a GB1 β-hairpin.[42] Finally, the motif can fold completely cooperatively, in a two-state, all-or-none fashion with the entire structure forming simultaneously.

Our data indicate that none of the above is actually realized. Although we have not followed the actual folding process, stabilities of the individual structural elements are critical determining factors for the folding mechanism. Recent state-of-the-art, ultralong molecular dynamics simulations of folding of several small, fast folding proteins[4] show that folding pathways follow the stabilization of native secondary structure elements by key tertiary contacts, which also establish a native-like topology. Initiation sites for folding are generally the most stable segments of native structure that persist in unfolded states; for helical proteins these are the most stable helices.[4] Based on these general principles, the above data paint a detailed picture of folding of the $\alpha t\alpha$ motif. The obvious initiation site is the center of the N-terminal helix, which has the highest propensity for retaining structure in the unfolded state. The opposite part of the second helix forms upon contact with this first native segment and both short helical regions are stabilized by the hydrophobic interactions. This also establishes the overall native topology of the motif (Figure 7). Then the helical structure propagates in both directions toward the turn and the termini simultaneously with zipping-up of the hydrophobic core.

It is possible that the folding process involves other steps, such as non-native helix contacts and subsequent creeping motion to align the register, as observed in a folding simulation study of a simplified helix-turn-helix model.[83] However, while these are feasible in folding from completely random or extended structures, real unfolded states are far from such. It has been shown that unfolded states contain a substantial amount of residual structure,[84] and non-negligible amount of residual helix is also apparent from the CD data (Figures 3, 4). It is also likely that tertiary interactions must exist to stabilize the helices, and these are also native-like. Therefore, consistent with the already discussed folding simulations, as well as other evidence including HDX experiments[39] and simplified, explicitly native-centric models of protein folding,[63,79−81] it is likely that folding starts with native contacts and proceeds through increasingly native-like structures rather than traps or off-pathway intermediates.

Comparison to other Helix-Turn-Helix Motifs. One of the reasons for studying the folding of $\alpha t\alpha$ is that it is one of the few known autonomously stable and monomeric helix-turn-helix proteins. Helix-turn-helix motifs are interesting as the simplest structures where long-range, tertiary contacts stabilize the arrangement of secondary elements (α-helices). Compared to the overwhelming popularity of the β-hairpin motif, however, only a handful of folding studies on helix-turn-helix motifs have been reported. These include the *de novo* designed 34-residue Z34C,[44] which requires a disulfide bridge for stability,[85] a 44-residue subdomain of an engrailed homeodomain (En-HD),[43] and a 51 residue SAP domain.[45] The latter two differ somewhat from the classical α-helical hairpin fold with antiparallel oriented α-helices. The Z34C and En-HD subdomains were found to fold ultrafast (in microseconds) and through an apparent two-state mechanism, based on the same unfolding transition monitored by CD and IR amide I′ in Z34C[44] and the same relaxation kinetics rate measured with IR and fluorescence,[43] although the multiphasic kinetics might suggest a more complex process. The ultrafast folding subdomain of En-HD may represent an important early intermediate in folding of the homeodomain, which further highlights the potential importance of these simple motifs in folding of larger helical proteins. The SAP domain folds slower, with simple (exponential) kinetics, also implying a two-state process.[45] However, a detailed investigation using multiple, site-specific structural probes has not been carried out for any of these motifs.

In our laboratory,[31] site-specific unfolding was previously studied for a 40-residue subdomain of the P22 coat protein from virus *Salmonella typhimurium bacteriophage*,[32] a helix-turn-helix motif very similar to $\alpha t\alpha$. It is therefore possible to directly compare the local thermodynamic stabilities and, by the arguments of the previous section, the folding mechanism of the two very similar structures but with completely different and unrelated sequences. The site-specific unfolding of the P22 subdomain was analyzed differently,[31,73] based on a three-state model from the global unfolding, since the SMSA/pSMSA methodology was not available. However, the preliminary analysis of the $\alpha t\alpha$ data with this method[86] yields qualitatively identical results with the same order of site-specific thermal stabilities.

Unfolding of the two motifs shows some similarities but also some important differences. In analogy to $\alpha t\alpha$, the P22 is stabilized mainly by tertiary interactions between the two helices: the corresponding peptide fragments again have essentially no residual helical structure in isolation.[31] The progress of unfolding is, however, nearly the opposite to that of the $\alpha t\alpha$. The P22 starts unfolding from the N-terminus, where the longer N-terminal helix does not seem to have interhelical helical contacts with the shorter, second α-helix. The unfolding then proceeds through the center of the motif toward the turn. In contrast to $\alpha t\alpha$, the turn in P22 is well ordered[32] and shows a distinct unfolding transition.[31] The most stable parts of the structure are the α-helical termini next to the turn segment. This implies that the folding of the P22 subdomain is initiated close to the turn, and its mechanism is "zipping" of the structure from the turn toward the termini. While completely

dx.doi.org/10.1021/ja500918k | J. Am. Chem. Soc. 2014, 136, 6037−6048

different from the αtα, which unfolds from the turn part (Figure 7), there seems to be the same superficial correlation between the thermal stability and the best defined structure based on the NMR coordinates.

What makes the comparison even more interesting is that αtα is a *de novo* designed sequence, while the P22 subdomain is natural. Artificially designed peptides and proteins have been very important models for studying folding, starting with the alanine-rich helical oligopeptides,[87] β-hairpins,[88] and several mini-proteins.[89,90] It is, however, believed that designed and natural sequences exhibit very different folding behavior, namely that the natural proteins fold cooperatively while the artificial ones do not.[48,49] In this case, folding of both motifs can be considered cooperative in the sense that both secondary and tertiary interactions must act in concert to stabilize their structures. On the other hand, neither unfolds through an all-or-none process; both motifs unfold gradually with different parts exhibiting distinct thermodynamic stabilities. It is interesting to note that if judged by the CD alone, the αtα would appear to exhibit a more cooperative unfolding than the P22: the ellipticity at 222 nm of the αtα is distinctly sigmoidal, (Figure 4) while for the P22 subdomain it is more gradual, perhaps suggesting a pretransition.[31] Therefore, the differences in cooperativity for natural versus *de novo* designed proteins are not borne out in this study.

CONCLUSION

We have investigated thermal unfolding of the αtα motif, its two peptide fragments, and seven ^{13}C isotopically edited variants labeled to sense the site-specific changes in the secondary structure. The recently developed methodology—shifted multivariate spectra analysis—has been utilized to ensure the intrinsic spectral changes are properly taken into account. The data show that αtα unfolds through a continuum of intermediates with varying degrees of native structure and provide the local thermodynamic stabilities for the labeled sites. Although the unfolding is heterogeneous, it is far from random but reflects the importance of the interactions responsible for stabilizing the native-like structure in the particular parts of the motif. As demonstrated by the low residual helical structure in the peptide fragments, the tertiary interactions play a decisive role in stabilizing the helical structure. Likewise, the order of unfolding of individual ^{13}C labeled segments correlates with the apparent stabilization by the tertiary, interhelical contacts. Differences between the site-specific thermal unfolding of αtα and another very similar helix-turn-helix motif highlight the importance of the particular sequence and associated specific inter-residue interactions in the protein stability and folding mechanism. Our results underscore the importance of high-resolution, site-specific experimental studies of protein folding but also highlight the limitations of the traditional, simple mass action models of data analysis.

ASSOCIATED CONTENT

Supporting Information

Amino acid sequence of αtα and its isotopologues; estimated unfolding heat capacities; example site-specific IR data analysis; results of two-state global amide I′ IR unfolding analysis; global and site-specific unfolding thermodynamics considering heat capacity changes; ^{13}C amide I′ band frequencies. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

jkubelka@uwyo.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Milan Balaz for the use of CD and HPLC instruments and Benjamin Anderson for proofreading of the manuscript. This work was supported by the National Science Foundation CAREER 0846140 grant.

REFERENCES

(1) Ding, F.; Tsao, D.; Nie, H. F.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010.
(2) Freddolino, P. L.; Harrison, C. B.; Liu, Y. X.; Schulten, K. *Nat. Phys.* **2010**, *6*, 751.
(3) Bowman, G. R.; Voelz, V. A.; Pande, V. S. *J. Am. Chem. Soc.* **2011**, *133*, 664.
(4) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517.
(5) Adhikari, A. N.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17442.
(6) He, Y.; Mozolewska, M. A.; Krupa, P.; Sieradzan, A. K.; Wirecki, T. K.; Liwo, A.; Kachlishvili, K.; Rackovsky, S.; Jagiela, D.; Slusarz, R.; Czaplewski, C. R.; Oldziej, S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 14936.
(7) Sato, S.; Religa, T. L.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6952.
(8) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Biophys. J.* **2011**, *100*, L47.
(9) Kuhrova, P.; De Simone, A.; Otyepka, M.; Best, R. B. *Biophys. J.* **2012**, *102*, 1897.
(10) Buchner, G. S.; Murphy, R. D.; Buchete, N.-V.; Kubelka, J. *Biochim. Biophys. Acta, Proteins Proteomics* **2011**, *1814*, 1001.
(11) Petty, S. A.; Decatur, S. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14272.
(12) Paul, C.; Axelsen, P. H. *J. Am. Chem. Soc.* **2005**, *127*, 5754.
(13) Mehta, A. K.; Lu, K.; Childers, W. S.; Liang, Y.; Dublin, S. N.; Dong, J. J.; Snyder, J. P.; Pingali, S. V.; Thiyagarajan, P.; Lynn, D. G. *J. Am. Chem. Soc.* **2008**, *130*, 9829.
(14) Moran, S. D.; Woys, A. M.; Buchanan, L. E.; Bixby, E.; Decatur, S. M.; Zanni, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 3329.
(15) Chi, H.; Welch, W. R. W.; Kubelka, J.; Keiderling, T. A. *Biomacromolecules* **2013**, *14*, 3880.
(16) Welch, W. R. W.; Keiderling, T. A.; Kubelka, J. *J. Phys. Chem. B* **2013**, *117*, 10359.
(17) Manas, E. S.; Getahun, Z.; Wright, W. W.; Degrado, W. F.; Vanderkooi, J. M. *J. Am. Chem. Soc.* **2000**, *122*, 9883.
(18) Walsh, S. T. R.; Cheng, R. P.; Wright, W. W.; Alonso, D. O. V.; Daggett, V.; Vanderkooi, J. M.; DeGrado, W. F. *Protein Sci.* **2003**, *12*, 520.
(19) Fesinmeyer, R. M.; Peterson, E. S.; Dyer, R. B.; Andersen, N. H. *Protein Sci.* **2005**, *14*, 2324.
(20) Silva, R.; Kubelka, J.; Bouř, P.; Decatur, S. M.; Keiderling, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8318.
(21) Huang, C. Y.; Getahun, Z.; Zhu, Y. J.; Klemke, J. W.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2788.
(22) Barber-Armstrong, W.; Donaldson, T.; Wijesooriya, H.; Silva, R. A.; Decatur, S. M. *J. Am. Chem. Soc.* **2004**, *126*, 2339.
(23) Ihalainen, J. A.; Paoli, B.; Muff, S.; Backus, E. H. G.; Bredenbeck, J.; Woolley, G. A.; Caflisch, A.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9588.
(24) Hauser, K.; Krejtschi, C.; Huang, R.; Wu, L.; Keiderling, T. A. *J. Am. Chem. Soc.* **2008**, *130*, 2984.
(25) Huang, R.; Wu, L.; McElheny, D.; Bouř, P.; Roy, A.; Keiderling, T. A. *J. Phys. Chem. B* **2009**, *113*, 5661.

(26) Smith, A. W.; Lessing, J.; Ganim, Z.; Peng, C. S.; Tokmakoff, A.; Roy, S.; Jansen, T. L.; Knoester, J. *J. Phys. Chem. B* **2010**, *114*, 10913.
(27) Jones, K. C.; Peng, C. S.; Tokmakoff, A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 2828.
(28) Brewer, S. H.; Song, B. B.; Raleigh, D. P.; Dyer, R. B. *Biochemistry* **2007**, *46*, 3279.
(29) Nagarajan, S.; Taskent-Sezgin, H.; Parul, D.; Carrico, I.; Raleigh, D. P.; Dyer, R. B. *J. Am. Chem. Soc.* **2011**, *133*, 20335.
(30) Huang, R.; Kubelka, J.; Barber-Armstrong, W.; Silva, R.; Decatur, S. M.; Keiderling, T. A. *J. Am. Chem. Soc.* **2004**, *126*, 2346.
(31) Amunson, K. E.; Ackels, L.; Kubelka, J. *J. Am. Chem. Soc.* **2008**, *130*, 8146.
(32) Sun, Y.; Parker, M. H.; Weigele, P.; Casjens, S.; Prevelige, P. E.; Krishna, N. R. *J. Mol. Biol.* **2000**, *297*, 1195.
(33) Yang, W. Y.; Pitera, J. W.; Swope, W. C.; Gruebele, M. *J. Mol. Biol.* **2004**, *336*, 241.
(34) Ahmed, Z.; Beta, I. A.; Mikhonin, A. V.; Asher, S. A. *J. Am. Chem. Soc.* **2005**, *127*, 10943.
(35) Garcia-Mira, M. M.; Sadqi, M.; Fischer, N.; Sanchez-Ruiz, J. M.; Muñoz, V. *Science* **2002**, *298*, 2191.
(36) Sadqi, M.; Fushman, D.; Muñoz, V. *Nature* **2006**, *442*, 317.
(37) Udgaonkar, J. B. *Annu. Rev. Biophys.* **2008**, *37*, 489.
(38) Kishore, M.; Krishnamoorthy, G.; Udgaonkar, J. B. *Biochemistry* **2013**, *52*, 9482.
(39) Englander, S. W.; Mayne, L.; Krishna, M. M. G. *Q. Rev. Biophys.* **2007**, *40*, 287.
(40) Muñoz, V. *Int. J. Quantum Chem.* **2002**, *90*, 1522.
(41) Fezoui, Y.; Connolly, P. J.; Osterhout, J. J. *Protein Sci.* **1997**, *6*, 1869.
(42) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
(43) Religa, T. L.; Johnson, C. M.; Vu, D. M.; Brewer, S. H.; Dyer, R. B.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9272.
(44) Du, D. G.; Gai, F. *Biochemistry* **2006**, *45*, 13131.
(45) Dodson, C. A.; Ferguson, N.; Rutherford, T. J.; Johnson, C. M.; Fersht, A. R. *Protein Eng. Des. Sel.* **2010**, *23*, 357.
(46) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298*, 937.
(47) Zarrine-Afsar, A.; Larson, S. M.; Davidson, A. R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 42.
(48) Scalley-Kim, M.; Baker, D. *J. Mol. Biol.* **2004**, *338*, 573.
(49) Watters, A. L.; Deka, P.; Corrent, C.; Callender, D.; Varani, G.; Sosnick, T.; Baker, D. *Cell* **2007**, *128*, 613.
(50) Zhang, Z.; Chan, H. S. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 2920.
(51) Amunson, K. E.; Kubelka, J. *J. Phys. Chem. B* **2007**, *111*, 9993.
(52) Ackels, L.; Stawski, P.; Amunson, K. E.; Kubelka, J. *Vib. Spectrosc.* **2009**, *50*, 2.
(53) Kaminský, J.; Bouř, P.; Kubelka, J. *J. Phys. Chem. A* **2011**, *115*, 30.
(54) Kubelka, J. *Anal. Chem.* **2013**, *85*, 9588.
(55) Sureshbabu, V. V.; Narendra, N. In *Amino Acids, Peptides and Proteins in Organic Chemistry*; Hughes, A. B., Ed.; Wiley-VCH: Weinheim, 2011; Vol. *4*, p 1.
(56) Fezoui, Y.; Braswell, E. H.; Xian, W.; Osterhout, J. J. *Biochemistry* **1999**, *38*, 2796.
(57) Amunson, K. E.; Anderson, B. A.; Kubelka, J. *Appl. Spectrosc.* **2011**, *65*, 1307.
(58) Toniolo, C.; Formaggio, F.; Woody, R. W. In *Comprehensive Chiroptical Spectroscopy*; Berova, N., Polavarapu, K. N., Woody, R. W., Eds.; John Wiley & Sons, Ltd: Hoboken, NJ, 2012; Vol. *2*, p 499.
(59) Gans, P. J.; Lyu, P. C.; Manning, M. C.; Woody, R. W.; Kallenbach, N. R. *Biopolymers* **1991**, *31*, 1605.
(60) Luo, P. Z.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8413.
(61) Chin, D. H.; Woody, R. W.; Rohl, C. A.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15416.
(62) Park, S.-H.; Shalongo, W.; Stellwagen, E. *Protein Sci.* **1997**, *6*, 1694.
(63) Kubelka, J.; Henry, E. R.; Cellmer, T.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18655.
(64) Makhatadze, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307.
(65) Murphy, K. P.; Freire, E. *Adv. Protein Chem.* **1992**, *43*, 313.
(66) Spolar, R. S.; Livingstone, J. R.; Record, M. T. *Biochemistry* **1992**, *31*, 3947.
(67) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138.
(68) Bernado, P.; Blackledge, M.; Sancho, J. *Biophys. J.* **2006**, *91*, 4536.
(69) Estrada, J.; Bernado, P.; Blackledge, M.; Sancho, J. *BMC Bioinf.* **2009**, *10*, 104.
(70) Buchner, G. S.; Shih, N.; Reece, A. E.; Niebling, S.; Kubelka, J. *Biochemistry* **2012**, *51*, 6496.
(71) Anderson, B. A.; Literati, A.; Ball, B.; Kubelka, J. *Biopolymers* **2014**, *101*, 536.
(72) Lorenz-Fonfria, V. A.; Padros, E. *Appl. Spectrosc.* **2005**, *59*, 474.
(73) Buchner, G. S.; Kubelka, J. In *Intrinsically Disordered Protein Analysis*; Uversky, V. N., Dunker, A. K., Eds.; Springer Science +Business Media: New York, 2012; Vol. *1*.
(74) Ma, H. R.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2283.
(75) Xian, W. J.; Connolly, P. J.; Oslin, M.; Hausrath, A. C.; Osterhout, J. J. *Protein Sci.* **2006**, *15*, 2062.
(76) Barrick, D. *Phys. Biol.* **2009**, *6*, 015001.
(77) Lane, T. J.; Schwantes, C. R.; Beauchamp, K. A.; Pande, V. S. *J. Chem. Phys.* **2013**, *139*, 145104.
(78) Liu, F.; Du, D.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2369.
(79) Muñoz, V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11311.
(80) Henry, E. R.; Eaton, W. A. *Chem. Phys.* **2004**, *307*, 163.
(81) Henry, E. R.; Best, R. B.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 17880.
(82) Hausrath, A. C. *Protein Sci.* **2006**, *15*, 2051.
(83) Hoffmann, D.; Knapp, E. W. *J. Phys. Chem. B* **1997**, *101*, 6734.
(84) Bowler, B. E. *Curr. Opin. Struct. Biol.* **2012**, *22*, 4.
(85) Starovasnik, M. A.; Braisted, A. C.; Wells, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10080.
(86) Buchner, G. S., Diploma Thesis, University of Wuerzburg, 2010.
(87) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286.
(88) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5578.
(89) Struthers, M.; Ottesen, J. J.; Imperiali, B. *Folding Des.* **1998**, *3*, 95.
(90) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425.