BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2011,** *50,* 1029–1041 1029
DOI: 10.1021/bi101551h

Manifestations of Native Topology in the Denatured State Ensemble of *Rhodopseudomonas palustris* Cytochrome *c′*†

Tanveer A. Dar,‡,§ R. Dustin Schaeffer,∥ Valerie Daggett,∥,⊥ and Bruce E. Bowler*,‡

‡*Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Dynamics, University of Montana, Missoula, Montana 59812, United States,* ∥*Biomolecular Structure and Design Program, and* ⊥*Department of Bioengineering, University of Washington, Seattle, Washington 98195-5013, United States.* §*Current address: School of Biosciences and Biotechnology, Baba Ghulam Shah Badshah University, Rajouri-185131 Jammu and Kashmir, India*

Received September 24, 2010; Revised Manuscript Received December 22, 2010

ABSTRACT: To provide insight into the role of local sequence in the nonrandom coil behavior of the denatured state, we have extended our measurements of histidine−heme loop formation equilibria for cytochrome *c′* to 6 M guanidine hydrochloride. We observe that there is some reduction in the scatter about the best fit line of loop stability versus loop size data in 6 M versus 3 M guanidine hydrochloride, but the scatter is not eliminated. The scaling exponent, ν_3, of 2.5 ± 0.2 is also similar to that found previously in 3 M guanidine hydrochloride (2.6 ± 0.3). Rates of histidine−heme loop breakage in the denatured state of cytochrome *c′* show that some histidine−heme loops are significantly more persistent than others at both 3 and 6 M guanidine hydrochloride. Rates of histidine−heme loop formation more closely approximate random coil behavior. This observation indicates that heterogeneity in the denatured state ensemble results mainly from contact persistence. When mapped onto the structure of cytochrome *c′*, the histidine−heme loops with slow breakage rates coincide with chain reversals between helices 1 and 2 and between helices 2 and 3. Molecular dynamics simulations of the unfolding of cytochrome *c′* at 498 K show that these reverse turns persist in the unfolded state. Thus, these portions of the primary structure of cytochrome *c′* set up the topology of cytochrome *c′* in the denatured state, predisposing the protein to fold efficiently to its native structure.

The essential and nontrivial role of the denatured state of proteins in their folding reactions has been increasingly scrutinized in recent years (*1*−*5*). Nowadays, it is widely accepted that protein denatured states contain nonrandom or residual structure (*4*, *5*). Much of the evidence for residual structure comes from spectroscopic techniques like NMR and fluorescence resonance energy transfer (FRET)[1] (*6*−*16*). Thermodynamic data indicate that residual structure can have a significant impact on the stability of denatured states and thereby affect the conformational distribution that comprises the denatured state ensemble (DSE) (*1*, *2*, *4*). Thus, residual structure can limit the conformational search toward a thermodynamic (or kinetic) minimum to increase protein folding efficiency (*17*). In contrast, SAXS and viscometric measurements have shown that denatured proteins scale with length, $R_g \propto N^{\nu}$, as expected for a random coil ($\nu \sim 0.6$) (*18*−*20*). This "reconciliation problem" (*5*, *19*) has been a subject of considerable interest.

Unlike SAXS and viscometric measurements, which interrogate the ensemble properties of proteins of different lengths, thermodynamic and kinetic studies on loop formation with DNA hairpin loops (*21*) and denatured proteins (*22*, *23*) are able to probe loop formation at different points along the length of a biopolymer. Loop formation data are usually interpreted using the Jacobson−Stockmayer equation (eq 1 (*24*, *25*))

$$\Delta S_{loop} = -\nu_3 R \ln n + R \ln\left(\left(\frac{3}{2\pi C_n l^2}\right)^{\nu_3} V_i\right) \qquad (1)$$

where n is the number of monomers in the loop, R is the gas constant, C_n is Flory's characteristic ratio, l is the distance between monomers, V_i is the approach volume of the atoms involved in loop formation, and ν_3 is the scaling exponent for loop formation. The Jacobson−Stockmayer equation assumes random coil behavior. Thus, loop formation is entirely entropic. For a freely jointed random coil, $\nu_3 = 1.5$ (*2*, *4*, *26*), and when excluded volume is taken into account, $\nu_3 = 1.8$−2.4 (*2*, *4*, *26*−*28*).

†This work was supported by the National Institutes of Health (R01GM074750 to B.E.B. and GMS R01GM050789 to V.D.). The simulations for Dynameomics were performed using computer time through the DOE Office of Biological Research as provided by the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

*To whom correspondence should be addressed. Phone: (406) 243-6114. Fax: (406) 243-4227. E-mail: bruce.bowler@umontana.edu.

[1]Abbreviations: FRET, fluorescence resonance energy transfer; Cyt*c′*, cytochrome *c′*; iso-1-Cyt*c*, yeast iso-1-cytochrome *c*; DSE, denatured state ensemble; MD, molecular dynamics; pWT, pseudo-wild-type, contains the mutation Q1A relative to the naturally occurring cytochrome *c′*; GdnHCl, guanidine hydrochloride; pK_{loop}(His), stability of a histidine−heme loop in the denatured state; m_{eq}, derivative of $2.3RTpK_{loop}$(His) with respect to GdnHCl concentration; ν_3, scaling exponent for loop formation; k_b, rate constant for His−heme loop breakage; $m_b^{\ddagger}$, derivative of $RT \ln k_b$ with respect to GdnHCl concentration, k_f, rate constant for His−heme loop formation; $m_f^{\ddagger}$, derivative of $RT \ln k_f$ with respect to GdnHCl concentration; $k_b^{XM}{}_{corr}$, rate constant for His−heme loop breakage at a molar guanidine hydrochloride concentration X, corrected for viscosity; $m_b^{\ddagger}{}_{corr}$, derivative of $RT \ln k_b^{XM}{}_{corr}$ with respect to GdnHCl concentration; TS, transition state. Format for indicating mutations: single letter abbreviation for original amino acid followed by position number followed by one letter abbreviation for new amino acid; for example, D3H indicates that the aspartate at position 3 has been replaced with histidine.

Published on Web 12/29/2010

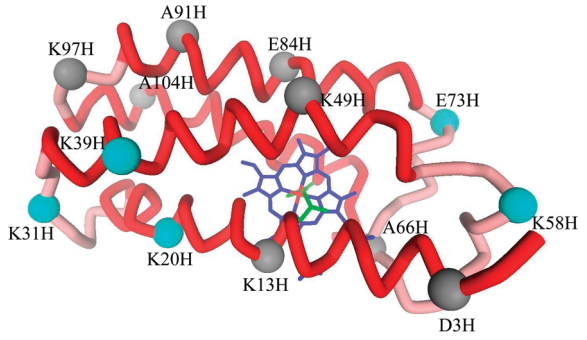

FIGURE 1: Structure of *R. palustris* Cyt*c*′ (1A7V) (*34*) showing the positions of histidine substitutions. The heme is shown in blue with the iron atom in red. The His117 axial ligand and the residue blocking the sixth coordination site, Leu12, are shown in green. The four helices are shown in red, and the loops connecting the helices are shown in pink. The positions of the residues mutated in each variant are shown with gray spheres, except for the residues with the slowest loop breakage rates which are shown in cyan. Mutations are indicated with standard notation (see Abbreviations footnote) next to each sphere.

The data on loop formation with DNA hairpins and denatured proteins have yielded ν_3 as high as 4−7 (*21*−*23*), much larger than expected for a random coil or a random coil with excluded volume. The large scaling exponents were attributed to residual structure, which was supported by the observation that ν_3 decreased with denaturant concentration (*23*). Recent molecular dynamics simulations confirm that large scaling exponents can result from residual structure (*29*).

For cases where the scaling exponent is at or near that expected for a random coil, simulations have shown that local nonrandom structure can coexist with random coil scaling exponents (*30*, *31*). Since nonrandom structure observed by NMR is often (*11*), but not always (*6*), local, these simulations provide a reasonable resolution of many aspects of the "reconciliation problem". In an extreme example, Monte Carlo simulations by Fitzkee and Rose showed that the scaling properties of a denatured protein can be insensitive to large amounts of structure (*32*).

Recently, we studied denatured state histidine−heme loop formation at 3 M guanidine hydrochloride (GdnHCl) with *Rhodopseudomonas palustris* cytochrome *c*′ (Cyt*c*′ (*33*, *34*)). Using a set of 13 single-histidine variants (Figure 1), we used His−heme loop equilibria to investigate the conformational properties of the denatured state every 7−11 residues along the sequence. In contrast to our earlier work on His−heme loop formation with yeast iso-1-cytochrome *c* (iso-1-Cyt*c*) where ν_3 was ~4 in 3 M GdnHCl, we observed $\nu_3 \sim 2.5$ in 3 M GdnHCl, which is consistent with random coil behavior (*33*). However, there was considerable scatter of the loop stability versus loop size data about the best fit line used to obtain ν_3. Thus, consistent with recent simulations (*30*−*32*), the denatured state of Cyt*c*′ has random coil scaling properties despite displaying nonrandom behavior along the polypeptide chain.

The origin of the nonrandom denatured state behavior was unclear in our initial studies. Local sequence hydrophobicity and local ϕ,ψ preferences (*30*, *31*) are both likely to be involved in producing nonrandom behavior in denatured states. FRET studies on the denatured state of Cyt*c*′ show that ~50% of the chains adopts compact structures under denaturing conditions (*8*, *9*). Given the high hydrophobic content of the primary sequence of Cyt*c*′ (*8*), residual structure due to hydrophobic clusters seems to be a likely cause for nonrandom behavior. To probe this possibility, we have measured the effect of stronger denaturing conditions on denatured state loop formation for the same set of Cyt*c*′ variants. We have previously shown that loop breakage rates play a key role in deviations from random coil behavior in the denatured state of iso-1-Cyt*c* (*35*, *36*). Thus, we also have measured rates of His−heme loop breakage in the denatured state of all 13 variants of Cyt*c*′ in both 3 and 6 M GdnHCl. To provide further insight into the results from the thermodynamics and kinetics of His−heme loop formation and breakage, we have carried out molecular dynamics (MD) simulations of the unfolding of Cyt*c*′.

MATERIALS AND METHODS

Preparation of Cytochrome c′ Variants. The Cyt*c*′ variants were prepared and purified from BL21(DE) cells that were cotransformed with the pEC86 vector (*37*) as described previously (*33*). Briefly, the plasmid pETcp (*38*), carrying the gene for Cyt*c*′, as well as plasmid pEC86, obtained from Thöny-Meyer, was transformed into BL21(DE3) cells under selection for ampicillin (pETcp) and chloramphenicol (pEC86). An osmotic shock method was used for the extraction of all Cyt*c*′ variants, as described earlier (*33*). For the final HPLC purification step, all variants were exchanged into low salt buffer A (5 mM sodium phosphate, pH 6.0), and high salt buffer B (5 mM sodium phosphate and 0.25 M NaCl, pH 6.0) was used to elute pure protein using the same gradient as described earlier (*33*).

Protein concentrations were determined using 100 mM sodium phosphate buffer at pH 7.05 as the spectrum is pH dependent. Scans were taken from 250 to 750 nm, and absorbance at 398 nm was used with a molar extinction coefficient of 85000 M^{-1} cm^{-1} to determine protein concentration (*39*). All variants were prepared in the pseudo-wild-type (pWT) background, which carries the Q1A mutation to avoid partial formation of pyroglutamate at the N-terminus (*38*).

Denatured State pH Titrations. UV−vis-monitored pH titrations for all Cyt*c*′ variants at 3 μM concentration were carried out in 6, 5, or 4 M GdnHCl in the presence of 5 mM Na_2HPO_4 and 15 mM NaCl, as described previously (*23*). Absorbance at 398 nm, A_{398}, was plotted against pH, and data were fit to a modified form of the Henderson−Hasselbalch equation (eq 2)

$$A_{398} = \frac{A_{LS} + A_{HS}10^{n_p(pK_a(obs) - pH)}}{1 + 10^{n_p(pK_a(obs) - pH)}} \quad (2)$$

to obtain the apparent pK_a, pK_a(obs), and the number of protons, n_p, involved in the process (A_{LS} is A_{398} of the low-spin form at high pH and A_{HS} is A_{398} of the high-spin form at low pH). The values reported are the average of three separate experiments, and the error given is the standard deviation.

For some variants, a biphasic titration was observed, with His−heme binding dominating below pH 7 and Lys−heme binding above pH 7. Fitting of these titration curves was done using eq 3 as described previously (*23*, *33*). In eq 3, pK_a($HisH^+$) is the pK_a of the histidine and pK_a($LysH^+$) is the pK_a of the lysine(s) that bind to the heme in denaturing GdnHCl solutions and pK_{loop}(His) (= −log[K_{loop}(His)]) and pK_{loop}(Lys) (= −log[K_{loop}(Lys)]) are the pK's for binding of a fully deprotonated histidine or lysine, respectively, to the heme in denaturing solutions of GdnHCl. pK_a($LysH^+$) is set to 10.5 for all fits to biphasic

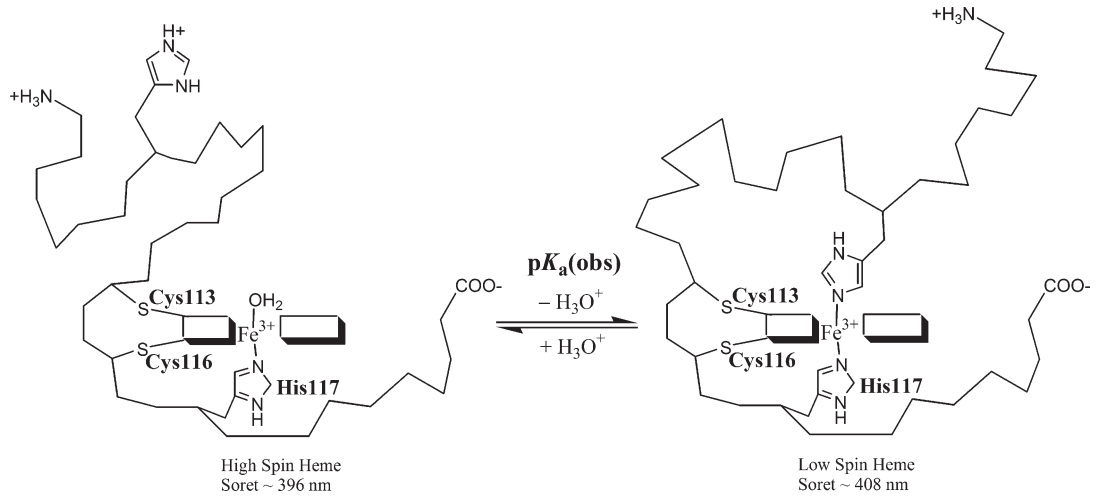

FIGURE 2: Schematic representation of the proton-mediated His−heme loop formation equilibrium for cytochrome *c*′ under denaturing conditions.

pH titration data (*23*, *33*). All other parameters in eq 3 are as in eq 2.

$$A_{398} = \left[A_{HS} + A_{LS}\left(\frac{10^{-pK_{loop}(His)}}{1+10^{pK_a(HisH^+)-pH}} + \frac{10^{-pK_{loop}(Lys)}}{1+10^{pK_a(LysH^+)-pH}}\right)\right] / \left[1+\frac{10^{-pK_{loop}(His)}}{1+10^{pK_a(HisH^+)-pH}} + \frac{10^{-pK_{loop}(Lys)}}{1+10^{pK_a(LysH^+)-pH}}\right] \quad (3)$$

Kinetic Measurements. Stopped-flow mixing was used to monitor the breakage of the His−heme bond in the denatured state. Loop breakage kinetic experiments were carried out at 3−6 M GdnHCl. Stock solutions of each variant at 6 μM concentration were prepared at the desired GdnHCl concentration in 10 mM MOPS, pH 6.80. The breakage reaction was initiated by 1:1 (v/v) mixing of protein stock solution with a solution containing 100 mM buffer, at the desired GdnHCl concentration, to reach a final reaction mixture of 3 μM protein, at the desired GdnHCl concentration and the desired ending pH of 3.1 or 3.6. For these experiments, an Applied Photophysics SX20 stopped-flow spectrophotometer with a 10 mm path length 20 μL flow cell was used. The dead time of the instruments was 0.7 ms for our mixing conditions as determined by the reduction of 2,6-dichlorophenolindophenol with ascorbate (*40*). Reactions were monitored by absorbance spectroscopy at 398 nm to observe the Soret band shift resulting from His−heme bond breakage. All reactions were carried out at 25 °C. Final reaction pH was determined by collecting the product of the mixing reaction and immediately measuring the pH. All data were fit to a single exponential rise to maximum equation. Fitting of data to a double exponential equation does not significantly improve the fits.

Molecular Dynamics Simulations of Cytc′. Simulations of Cyt*c*′ were run at high temperature to sample the unfolding pathway. Chain A from PDB 1MQV (pWT variant carrying the F32W/W72F mutations) was used as the starting structure (*8*). Cyt*c*′ is a member of the four-helical up-and-down bundle metafold, rank 19, of our consensus domain dictionary (Schaeffer et al., submitted for publication) used for our Dynameomics project; this project contains native and unfolding simulations of representatives of all protein folds (www.dynameomics.org) (*41*, *42*). The heme cofactor was based on our standard parameters (*43*) (also Amanda Jonsson, unpublished results, in preparation) and covalently bonded to the protein through thioether linkages at C113 and C116. In addition, a pseudobond between the Nε2 atom of H117 and the heme iron was imposed. A single 298 K simulation was run for 40 ns as a native state control. Two 498 K simulations were run for 60 ns to model the unfolding pathway.

Unfolding simulations were run using our standard protocols (*41*, *44*, *45*). The input structure was minimized for 1000 steps of steepest descent (SD) minimization. Following minimization, the structure was solvated in a rectangular water box extending at least 10 Å from any protein atom. The water density of the box was adjusted to match the experimental density (0.829 g/mL at 498 K) by altering the box dimensions slightly (*46*, *47*). The waters of the combined system were then minimized for an additional 1000 steps, followed by 500 steps of dynamics/heating of the entire system, 500 further steps of water minimization, and a final 500 steps of minimization of the entire system. The flexible three-center (F3C) water model was used (*48*) along with periodic boundary conditions to minimize edge effects. Simulations were performed using the Levitt et al. force field (*43*) and the microcanonical ensemble (NVE; number of particles, volume, and total energy held fixed). All simulations were performed using *in lucem* molecular mechanics (*il*mm) (*49*). In all 498 K simulations, an 8 Å force-shifted nonbonded cutoff was used, and the nonbonded pair interaction list was updated every 2 steps. Nonbonded interactions between charged groups separated by three bonds were scaled to 0.4 of their full force (*50*). Finally, a 2 fs time step was used, and structures were saved every 500 steps, resulting in a trajectory of 1 ps snapshots.

RESULTS

Denatured State His−Heme Loop Formation Methods. We use His−heme loop formation in the denatured state to measure both the equilibria and kinetics of formation of simple polypeptide loops (*2*, *4*, *22*, *23*, *33*, *35*, *36*, *51*). The method requires a *c*-type cytochrome where the heme is attached covalently to the polypeptide chain through a CXXCH heme binding motif. The site of attachment can be near the N-terminus of the protein as with mitochondrial cytochromes *c*, such as iso-1-Cyt*c* (*2*, *4*, *22*, *23*, *35*, *36*, *51*), or near the C-terminus as with the type 2 cytochrome *c*, Cyt*c*′ (*33*), used in the current work.

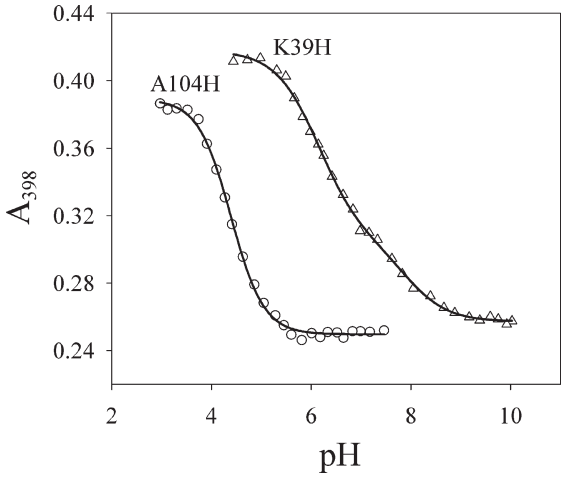

FIGURE 3: Representative titration curves for His−heme loop formation as a function of pH at 22 ± 1 °C in 6 M GdnHCl, 5 mM sodium phosphate, and 15 mM NaCl. Absorbance at 398 nm, A_{398}, versus pH is shown for the A104H (○) and K39H (△) variants of Cyt*c*′. The data for the A104H variant are monophasic and are fit (solid line) to a modified form of the Henderson−Hasselbalch equation (eq 2 in Materials and Methods) yielding pK_a(obs) and n_p, the number of protons coupled to loop formation (parameters are collected in Table 1). The data for the K39H variant are fit (solid curve) to a model involving heme binding by His at lower pH and Lys at higher pH (eq 3 in Materials and Methods) yielding pK_a(HisH$^+$), pK_{loop}(His), and pK_{loop}(Lys) (parameters are collected in Table 2).

With Cyt*c*′, the loop is formed with a histidine N-terminal to the site of heme attachment (Figure 2). The native sequence of Cyt*c*′ contains no other histidines besides His117, which is part of the heme binding motif. Thus, each histidine variant (see Figure 1) forms a unique His−heme loop under denaturing conditions, and the size of the loop is determined by the sequence position of the engineered histidine. Since histidine has an ionizable side chain, the stability of a His−heme loop can be measured by a simple pH titration (Figure 2), providing an apparent pK_a, pK_a(obs). The pK_a(obs) for loop formation provides information on the relative stability of loops of different sizes. The lower the pK_a(obs), the more stable the loop (i.e., higher [H$^+$] needed to break the loop). His−heme loop formation is monitored at 398 nm, a wavelength which is particularly sensitive to the change in Fe^{3+} spin state when the weak field ligand, water, replaces the strong field histidine ligand upon loop breakage. We note that electrostatic contributions to denatured state loop formation measurements are expected to be suppressed in concentrated GdnHCl solutions (*52*−*54*).

The binding of histidine to the heme can be broken down into two steps, ionization of the histidine, followed by binding of the deprotonated histidine to the Fe^{3+} of the heme. Thus, pK_a(obs) can be written as the sum of the two p*K*'s (eq 4), pK_a(HisH$^+$):

$$pK_a(\text{obs}) = pK_a(\text{HisH}^+) + pK_{loop}(\text{His}) \quad (4)$$

the intrinsic pK_a for ionization of histidine in denaturing concentrations of GdnHCl, and pK_{loop}(His) (= −log[K_{loop}(His)]), the p*K* for His−heme loop formation with a fully deprotonated histidine. For histidines at positions 3, 13, 20, and 39 of Cyt*c*′, pK_a(HisH$^+$) = 6.7 ± 0.2 in 3 M GdnHCl (*33*). Similarly, for histidines at positions 73, 89, and 100 of iso-1-Cyt*c*, we found pK_a(HisH$^+$) = 6.6 ± 0.1 at GdnHCl concentrations ranging from 3 to 6 M (*23*). Thus, we typically subtract a constant value for pK_a(HisH$^+$) from pK_a(obs) to obtain pK_{loop}(His).

Table 1: Thermodynamic Parameters from Monophasic pH Titration Data for Denatured State Loop Formation of Cytochrome *c*′ Variants in Concentrated GdnHCl Solutions and at 22 ± 1 °C[a]

variant	loop size (*n*)	pK_a(obs)	n_p[b]	pK_{loop}(His)[c]
		6 M GdnHCl		
A104H	10	4.40 ± 0.02	1.33 ± 0.08	−2.38 ± 0.02
K97H	17	4.79 ± 0.04	1.31 ± 0.10	−1.99 ± 0.04
A91H	23	5.38 ± 0.02	1.10 ± 0.02	−1.40 ± 0.02
K84H	30	5.21 ± 0.03	1.02 ± 0.04	−1.57 ± 0.03
E73H	41	5.42 ± 0.01	1.07 ± 0.03	−1.36 ± 0.01
pWT		7.39 ± 0.02	1.16 ± 0.03	
		5 M GdnHCl		
K84H	30	5.10 ± 0.02	1.06 ± 0.03	−1.69 ± 0.02
E73H	41	5.35 ± 0.02	1.04 ± 0.02	−1.43 ± 0.02
		4 M GdnHCl		
K84H	30	4.82 ± 0.05	1.01 ± 0.04	−1.96 ± 0.05
E73H	41	5.16 ± 0.03	0.99 ± 0.05	−1.63 ± 0.03

[a]All parameters are the average and standard deviation of at least three trials. [b]Number of protons involved in loop formation equilibrium. [c]Calculated using eq 4 with pK_a(HisH$^+$) = 6.78, the average value for all variants from 3 to 6 M GdnHCl.

For a random coil, the stability of a loop is expected to be governed by the entropy of loop formation as given by the Jacobson−Stockmayer equation (eq 1 (*24*, *25*)). Assuming random coil behavior and that the Fe^{3+}−imidazole bond energy is identical for all His−heme loops, we can write ΔG_{loop}(His) = $2.3RT$pK_{loop}(His) = $-T\Delta S_{loop}$, where ΔS_{loop} is given by eq 1. Thus, for a random coil, the dependence of pK_{loop}(His) on the log of loop size, log *n*, should be linear with a slope equal to the scaling exponent, ν_3, as given by eq 5 (*22*, *23*, *33*), where pK_{loop}(His)$_{ref}$ is the pK_{loop}(His) for *n* = 1.

$$pK_{loop}(\text{His}) = pK_{loop}(\text{His})_{ref} + \nu_3 \log n \quad (5)$$

*Denatured State Loop Formation for Cyt*c*′ in GdnHCl Solution*. To extend our previous studies on denatured state His−heme loop formation for Cyt*c*′ in 3 M GdnHCl (*33*), we have carried out denatured state titrations in 6 M GdnHCl for all 13 single-histidine variants of this protein (we note that global unfolding of all variants is completed well before 3 M GdnHCl (*33*)). Variants with *n* = 10−41 yield monophasic titration behavior for denatured state His−heme loop formation, and the data can readily be fit to a modified form of Henderson−Hasselbalch equation (Figure 3). Table 1 summarizes the pK_a(obs) parameters derived from pH titration data for this group of variants. The pK_a(obs) values increase with loop size, as expected, if entropy plays a significant role in the loop equilibria. However, clearly there are irregularities in this dependence. It is also evident that the number of protons, n_p, released is close to 1, consistent with the reaction scheme in Figure 2. Table 1 contains data for pseudo-wild-type Cyt*c*′ (pWT (*38*)). The pK_a(obs) seen for pWT is slightly higher than the value seen in 3 M GdnHCl and is consistent with Lys−heme binding, as demonstrated for iso-1-Cyt*c* in this pH regime (*55*).

For the variants with denatured state loop sizes of 48−111, biphasic titration behavior was observed (Figure 3). This observation indicates that loop stability has decreased such that the high- to low-spin transition cannot be completed by the binding

Table 2: Thermodynamic Parameters from Biphasic pH Titration Data for Denatured State Loop Formation of Cytochrome c' Variants in Concentrated GdnHCl Solutions and at 22 ± 1 °C[a]

variant	loop size (n)	pK_a(HisH$^+$)	pK_{loop}(His)	pK_{loop}(Lys)[b]
		6 M GdnHCl		
A66H	48	6.88 ± 0.03	−0.76 ± 0.07	−3.04 ± 0.07
D58H	56	6.86 ± 0.08	−0.93 ± 0.09	−3.29 ± 0.26
K49H	65	7.01 ± 0.03	−0.25 ± 0.07	−3.19 ± 0.14
K39H	75	6.71 ± 0.15	−0.35 ± 0.09	−3.08 ± 0.11
K31H	83	6.82 ± 0.04	−0.14 ± 0.03	−3.40 ± 0.20
K20H	94	6.63 ± 0.01	−0.31 ± 0.01	−3.08 ± 0.02
K13H	101	6.88 ± 0.02	0.03 ± 0.01	−3.00 ± 0.03
D3H	111	7.02 ± 0.13	0.31 ± 0.14	−3.12 ± 0.05
		5 M GdnHCl		
K39H	75	6.69 ± 0.05	−0.48 ± 0.01	−3.34 ± 0.14
K20H	94	6.81 ± 0.05	−0.42 ± 0.04	−3.21 ± 0.05
		4 M GdnHCl		
K39H	75	6.73 ± 0.02	−0.67 ± 0.04	−3.33 ± 0.04
K20H	94	6.73 ± 0.02	−0.55 ± 0.03	−3.48 ± 0.09

[a]All parameters are the average and standard deviation of at least three trials. [b]The ionization constant of lysine, pK_a(LysH$^+$), is assumed to be 10.5.

of histidine to the heme. Instead, lysine−heme binding completes the transition above pH 7. The data were fitted to an equilibrium equation involving both His−heme and Lys−heme loop formation (eq 3 in Materials and Methods). The analysis permits the direct extraction of pK_{loop}(His), pK_a(HisH$^+$) for the histidine involved in loop formation, and pK_{loop}(Lys) (see Table 2). As with the pK_a(obs) for variants in Table 1, pK_{loop}(His) becomes less negative (less favorable) as loop size increases. As for the pK_a(obs) data, the progression of pK_{loop}(His) toward less negative values as n increases is irregular. The values for pK_a(HisH$^+$) for each histidine in 6 M GdnHCl obtained from fits to eq 3 are relatively insensitive to sequence position. For eight variants, pK_a(HisH$^+$) = 6.85 ± 0.14. This value is within error of pK_a(HisH$^+$) = 6.7 ± 0.2 observed for four variants of Cytc' in 3 M GdnHCl.

The K84H, E73H, K39H, and K20H variants yielded the largest changes in pK_a(obs) or pK_{loop}(His) at 6 M versus 3 M GdnHCl. Thus, pH titrations also were carried out at 4 and 5 M GdnHCl for these variants to better determine the denaturant dependence of His−heme loop stability. The resulting pK_a(obs) or pK_{loop}(His) and pK_a(HisH$^+$) parameters are collected in Tables 1 and 2. Inspection of the data shows that His−heme loop formation becomes less favorable as GdnHCl concentration increases. The data in Table 2 also show that the magnitude of pK_a(HisH$^+$) is insensitive to GdnHCl concentration for the K39H and K20H variants. Thus, we average the pK_a(HisH$^+$) data for all Cytc' variants obtained in 3−6 M GdnHCl yielding pK_a(HisH$^+$) = 6.78 ± 0.15.

Trends in pK_{loop}(His) in 3 M versus 6 M GdnHCl. To allow comparison of loop stability data for the variants having monophasic versus biphasic titration behavior, we use the average value of pK_a(HisH$^+$) to extract pK_{loop}(His) from monophasic titration data (eq 2, Table 1). The pK_{loop}(His) values for all the variants are plotted against n in Figure 4. pK_{loop}(His) becomes less negative (less favorable) both at higher GdnHCl concentration and with increasing loop size. The scatter of the data about the fit to eq 5 is decreased in 6 M GdnHCl compared to the poorer solvent conditions of 3 M GdnHCl (R^2 increases from 0.87 in 3 M GdnHCl to 0.94 in 6 M GdnHCl). But, the scatter is not eliminated in 6 M GdnHCl, and interestingly the pattern of scatter is the same at both GdnHCl concentrations. For comparison, Figure 4 also shows data for His−heme loop formation in 6 M GdnHCl with homopolymeric polyalanine sequences inserted into iso-1-Cyt$c$ (*56*). The scatter about the fit to eq 5 for this low complexity sequence is minimal compared to that observed for loop formation with the Cytc' sequence. A scaling exponent, ν_3, of 2.5 ± 0.2 is obtained for His−heme loop formation with Cytc' in 6 M GdnHCl, which is consistent with the scaling exponent range of 1.8−2.4 for a random coil with excluded volume (*26*−*28*). The ν_3 in 6 M GdnHCl is also within error of ν_3 = 2.6 ± 0.3 observed for the fit to the 3 M GdnHCl Cytc' data in Figure 4.

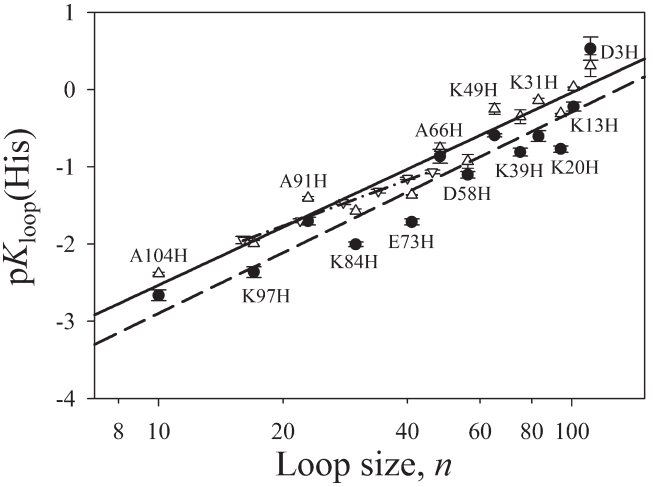

FIGURE 4: Plot of the denatured state loop stability, pK_{loop}(His), versus loop size, n (on a logarithmic scale), for all variants of Cytc' at 6 M (△) and 3 M (●) GdnHCl. Data for homopolymeric polyalanine inserts acquired in 6 M GdnHCl (▽, taken from ref *56*) are shown for comparison. Fits to eq 5 (6 M GdnHCl, solid line; 3 M GdnHCl, dashed line; polyalanine, dash-dotted line) yield ν_3 reported in the text. All pK_{loop}(His) values were determined at 22 ± 1 °C. For consistency, we have used the value of pK_a(HisH$^+$) reported here, which is derived from a much larger number of measurements, to reevaluate pK_{loop}(His) in 3 M GdnHCl. This results in pK_{loop}(His) being 0.08 more negative than previously reported for 3 M GdnHCl (Table 3 in ref *33*).The His−heme loop equilibrium for the K84H variant was repeated, here, yielding pK_a(obs) = 4.78 ± 0.03 [pK_{loop}(His) = −2.00 ± 0.03]. Data for each mutation site are labeled using the standard notation for single site mutations (see Abbreviations footnote).

For the four variants, K84H, E73H, K39H and K20H, where we have measured pK_{loop}(His) at 3, 4, 5, and 6 M GdnHCl, pK_{loop}(His) becomes progressively less negative as GdnHCl concentration increases (Tables 1 and 2, Figure S1 of the Supporting Information). Thus, His−heme loop formation in the denatured state is disfavored incrementally as GdnHCl concentration increases.

Kinetics of Denatured State Loop Breakage and Formation in 3−6 M GdnHCl. We have demonstrated in previous work that the kinetics of loop breakage and formation is consistent with a model which involves rapid histidine deprotonation followed by binding of the histidine to the heme (*35*). Thus, the observed rate constant has the pH dependence given

by eq 6

$$k_{obs} = k_b + k_f\left(\frac{K_a(HisH^+)}{K_a(HisH^+) + [H^+]}\right) \quad (6)$$

where k_b and k_f are the rate constants for loop breakage and formation, respectively, and $K_a(HisH^+)$ is the dissociation constant for histidine deprotonation. Therefore, if pH $\ll$ pK_a-($HisH^+$), then k_{obs} is equal to k_b.

The kinetics of His−heme loop breakage were measured at 25 °C in both 3 and 6 M GdnHCl by downward pH-jump mixing experiments monitored via the heme Soret absorption band. Figure S2 of the Supporting Information shows representative kinetic traces for the A91H variant of Cyt*c*′ for loop breakage in 3 and 6 M GdnHCl. The data fit well to a single exponential equation, as expected for the breakage of an intramolecular His−heme loop. For all variants, k_{obs} for downward pH jumps to pH 3.1 is similar to or slightly smaller than k_{obs} for downward jumps to pH 3.6 (Table S1 of the Supporting Information). This observation indicates that k_{obs} has reached its lower limit of k_b in this pH range (eq 6). Thus, we use k_{obs} obtained at pH 3.1 for k_b.

The k_b data at 3 M (●) and 6 M (△) GdnHCl are plotted as a function of loop size in Figure 5 (top panel). The magnitude of k_b varies up and down in an irregular manner with loop size. For comparison, k_b data in 3 M GdnHCl as a function of loop size for homopolymeric polyalanine inserts in iso-1-Cyt*c* (*56*) is also shown in Figure 5 (▽, top panel). For short polyalanine loops, k_b is higher due to chain stiffness. For longer loops, k_b levels out to a constant value as would be expected for a random coil. Thus, the irregular variation in k_b with loop size for the Cyt*c*′ variants is indicative of nonrandom behavior in the denatured state.

In the majority of cases, k_b is larger or within error the same at 3 M GdnHCl compared to 6 M GdnHCl. Only at loop sizes of 41 (E73H), 75 (K39H), and 94 (K20H) does k_b increase from 3 to 6 M GdnHCl. We measured k_b at 4 and 5 M GdnHCl (Table S2 of the Supporting Information) for the A104H (largest decrease in k_b as GdnHCl concentration increases from 3 to 6 M) and K39H (largest increase in k_b as GdnHCl increases from 3 to 6 M) variants. The results (Figure 5, inset, top panel) show a monotonic decrease in k_b as GdnHCl concentration increases for the A104H variant. For the K39H variant, k_b increases from 3 to 4 M GdnHCl and then appears to level off. The progressive decrease in k_b as GdnHCl concentration increases for the A104H variant suggests that the increased viscosity of more concentrated GdnHCl solutions slows loop breakage. Correction of k_b for the A104H variant for the increase in viscosity as GdnHCl concentration increases shows that $RT \ln k_b$ is nearly denaturant independent (Figure S3 of the Supporting Information). Thus, for the A104H variant the effect of GdnHCl on k_b is almost purely due to viscosity. Studies on denatured state dynamics have focused on viscosity effects on loop formation kinetics (*57*, *58*), and thus little is known about the viscosity dependence of loop breakage kinetics in the denatured state. Studies on the viscosity dependence of protein folding show that both folding and unfolding are viscosity dependent (*59*−*61*). A plot of τ_b (k_b^{-1}) versus the viscosity of GdnHCl for the A104H variant is linear and within error goes through the origin when corrected for nonviscosity effects (Figure S4 of the Supporting Information). The observation that the viscosity dependence of τ_b goes through the origin is consistent with the breakage reaction not being limited by any intrinsic constraints (internal friction) of the polypeptide chain consistent with some (*59*, *62*), but not all (*60*, *61*, *63*), protein folding reactions.

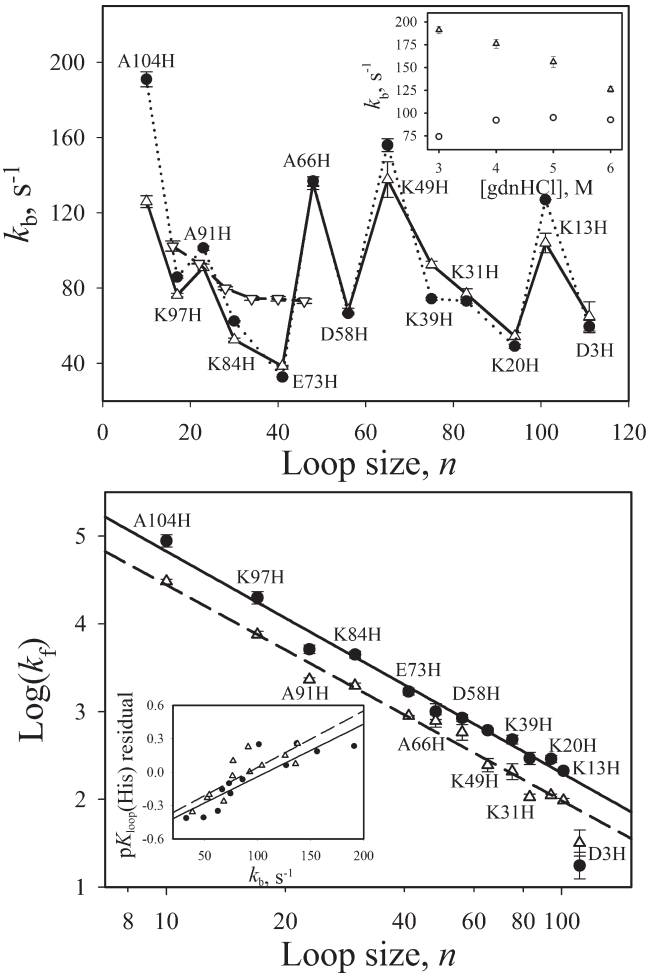

FIGURE 5: Loop kinetics as a function of loop size, n, at 25 °C for all variants of Cyt*c*′ in 3 M (●) and 6 M (△) GdnHCl. Rate constants for loop breakage, k_b, as a function of n are shown in the top panel. The k_{obs} values at pH ~3.1 (Supporting Information Table S1) are used in this plot. Data for homopolymeric polyalanine inserts acquired in 3 M GdnHCl (▽, taken from ref *56*) are shown for comparison. The inset to the top panel shows plots of k_b versus GdnHCl concentrations for the A104H (△) and K39H (○) variants of Cyt*c*′. The values of k_b used in these plots for the A104H variant are those corrected for the contribution from the forward rate constant at pH 3.1 (Tables S1 and S3 in the Supporting Information). The logarithm of the rate constant for loop formation, log k_f, as a function of n (axis is logarithmic) is shown in the bottom panel. Values of k_f are calculated using a two-state assumption [$k_f = k_b K_{loop}$(His)] and the k_b data at 25 °C. The k_f value for the A104H variant is calculated using the corrected k_b for the A104H variant in Table S1 (see Supporting Information Table S3). The solid (3 M GdnHCl) and dashed (6 M GdnHCl) straight lines are best fits of the data to a linear equation. R^2 for the fit to the 3 M GdnHCl data is 0.988. R^2 for the fit to the 6 M GdnHCl data is 0.984. The scaling exponents, ν_3, obtained from these fits are reported in the text. The inset to the bottom panel shows a plot of the pK_{loop}(His) residual from the fits in Figure 4 plotted against k_b. Data are shown for both 3 M (●) and 6 M (△) GdnHCl. The solid (3 M GdnHCl) and dashed (6 M GdnHCl) straight lines are best fits of the data to a linear equation. R^2 for the fit to the 3 M GdnHCl data is 0.764. R^2 for the fit to the 6 M GdnHCl data is 0.694. Data for the D3H variant are not included in the inset due to the larger error in pK_{loop}(His) for this variant. Data for points are identified using the standard notation for single site mutations (see Abbreviations footnote).

Loop formation rate constants, k_f, were calculated from K_{loop}(His) and k_b, assuming two-state behavior for denatured

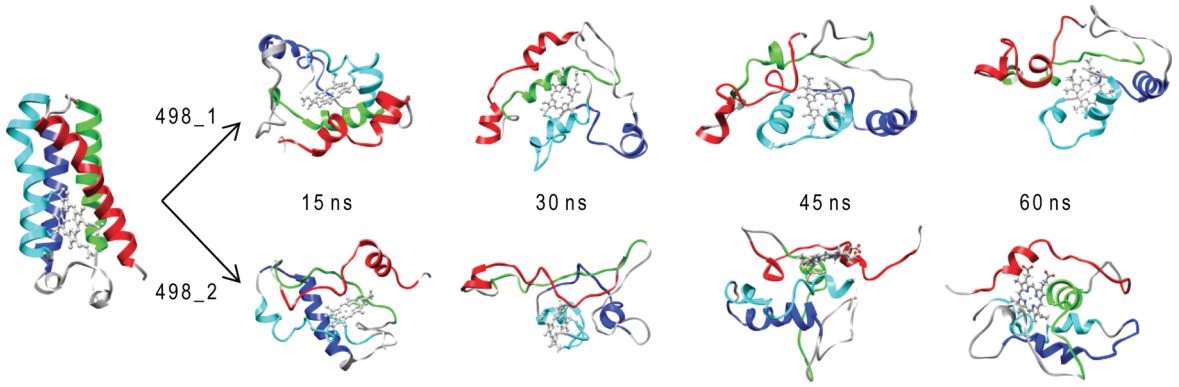

FIGURE 6: Snapshots from unfolding trajectories of Cyt*c*′. The starting structure and the 15, 30, 45, and 60 ns snapshots from the two independent 498 K unfolding simulations are shown. Structures are colored by native helical structure: red, residues 4−27; green, 34−54; blue, 75−95; cyan, 99−120.

state His−heme loop formation (Table S3 of the Supporting Information). Linear fits to plots of log k_f versus n (plotted logarithmically) yield $\nu_3 = 2.5 \pm 0.1$ at both 3 and 6 M GdnHCl (Figure 5, bottom), consistent with the thermodynamic data in Figure 4. The scatter about the best fit lines is lower for the kinetic data than for the thermodynamic data. Thus, as with our previous work on the kinetics and thermodynamics of denatured state loop formation with yeast iso-1-Cyt*c* (*23*, *35*, *36*, *64*), much of the deviation from random coil behavior in the denatured state results from k_b, not k_f.

MD Simulations of the Unfolding of Cytc′. We have performed two fairly long MD simulations of the thermal unfolding of Cyt*c*′ at 498 K. Various representative snapshots from the simulations are presented in Figure 6. As can be seen, there is some residual helical structure in the denatured state ensembles, structure that fluctuates over time. The position of the major residual helix (approximately residues 75−95, helix 3, colored blue in Figure 6) is in good agreement with the predictions of the Agadir algorithm (ref *65*, Figure S5 of the Supporting Information). Agadir also predicts that residues 2−18 should have some residual helical structure but that region is mostly unstructured in the MD-generated structures (red helix in Figure 6). In addition, the denatured ensemble is compact with distances from residues 32 and 72 to the heme iron of < 25 Å (21.7 ± 3.3 Å; see Figure S6 of the Supporting Information), which is in agreement with the compact portions of the distributions seen in denaturing concentrations of GdnHCl with FRET experiments for both Trp32−heme (F32W/W72F variant) and Trp72−heme (pWT variant) (*8*, *9*).

To interrogate the nature and extent of intrinsic residual structure in the denatured state of Cyt*c*′, the covalent His−heme loops probed experimentally are not present in the simulations. In this way we can examine the nature of any residual structure in the absence of potential perturbations imposed upon loop formation. In particular, we are interested in investigating whether there is residual structure around residues 20, 31, 39, 58, and 73 since k_b is lower for these positions (Figure 5, top panel).

We first consider the region around residues 20, 31, and 39, which contains the C-terminal portion of helix 1, the turn between helix 1 and helix 2, and the N-terminus of helix 2 (Figure 7A). This turn, or loop, is maintained, but altered, in the denatured ensembles (see the segment between the red and green helices in Figure 6). The turn slides toward the C-terminus of helix 1 due to stabilizing interactions caused by formation of dynamic hydrophobic clusters involving primarily the following residues: Leu12, Met15, Ala18, Trp32, Val36, and Val37. These residues each formed well-populated contacts ($> 10\%$ of the time) in the denatured state (20−60 ns). Leu12 primarily contacted Met15 (78%), Val37 (16%), and Leu41 (40%). In addition to Leu12, Met15 contacted Ala18 (29%), Trp32 (10%), and Val37 (11%). Trp32 also contacted Val36 (42%) and Val37 (25%). Five of these interactions are nonnative, and the precise partners fluctuated over time (consequently not all of the interacting residues are observed in any one MD snapshot).

The large Ω loop between helices 2 and 3 was also largely maintained in the denatured state (the gray regions between the green and blue helices in Figure 6, and see detail in Figure 7B). Persistent interactions were observed between Leu51 and Leu54 (55%), Leu51 and Phe55 (49%), and Leu54 and Phe 78 (24%). While the loop is large and variable, it served to keep Leu54 and Ala76 in proximity, which in turn kept helices 2 and 3 near one another. It is surprising that such a large loop was maintained. Residues 50−58 are very polar, and they remained extended to interact with water (Figure 7B). After ~30 ns in the 498_1 simulation, Phe72 and Phe55 came together, and other hydrophobic residues (Leu51, Leu68, and Phe78) then joined in to form a dynamic hydrophobic cluster that effectively pinched off the loop and persisted throughout the remainder of the simulations; effectively, the configuration of that loop and contacts are the same in many other structures in the ensemble, but particular interactions may differ from structure to structure. (Figure 7B).

While these two loops were dynamic in the simulations with interchanging residues aiding in their maintenance such that chain reversal occurred, they effectively kept helices 1 and 2 and helices 2 and 3 in proximity. Note how the red and green regions in Figure 8 stay near one another throughout the folding/unfolding process even though the helical structure in the region varies and the interhelical interactions vary. The same is true for interactions between the segments of the chain corresponding to helices 2 and 3 (green and blue in Figure 8). Precise interactions were not maintained, but interactions between the regions were maintained such that the topology of the native state was maintained. The same cannot be said for helix 4; instead, it was more mobile and interacted with different regions of the structure over time.

DISCUSSION

In the present work, we have measured the denaturant dependence of the thermodynamics and kinetics of His−heme

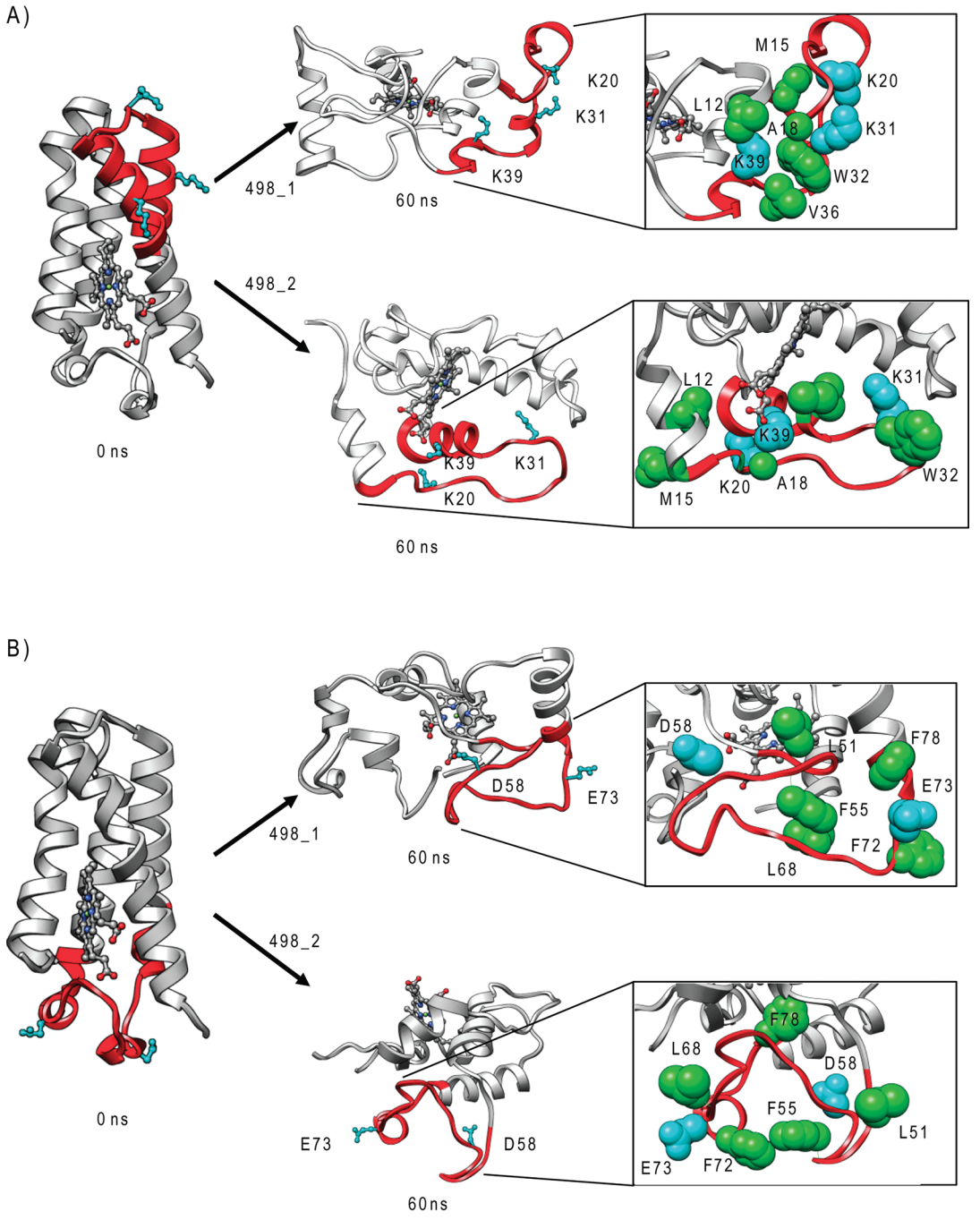

FIGURE 7: (A) Residual structure in the region including the C-terminus of helix 1, the turn connecting helix 1 and helix 2, and the N-terminus of helix 2. Residues 15−44 are colored red and shown in both the starting structure (left) and the final structure of 498_1 (top center) and 498_2 (bottom center). Mutation sites are shown in cyan. Side chain positions of hydrophobic cluster participants are shown in detail for both 498_1 (right top) and 498_2 (right bottom). (B) Residual structure in the region including the Ω loop between helices 2 and 3. Residues 50−78 are colored red and shown in both the starting structure (left) and the final structure of 498_1 (top center) and 498_2 (bottom center). Mutation sites are shown in cyan. Side chain positions of hydrophobic cluster participants are shown in detail for both 498_1 (top right) and 498_2 (bottom right).

loop formation in the denatured state of Cyt*c*′. FRET methods have shown that a significant fraction (~50%) of the denatured state of this protein remains compact even in 4 M GdnHCl in contrast to iso-1-Cyt*c* where the fraction of the denatured state occupying compact structures is only about 10% in high concentrations of GdnHCl (*8*, *9*). Thus, the present results provide an apt comparison with our previous results on His−heme loop formation in the denatured state of iso-1-Cyt*c* (*23*). In the following sections, we will compare and contrast the effects of denaturant concentration on the equilibria and kinetics of loop formation for these two proteins. Finally, we will consider the physical basis for the observed results and the broader implications for protein folding in the context of the MD simulation results.

*Denatured State Equilibrium Loop Formation for Cyt*c′ *Is Weakly Dependent on GdnHCl Concentration*. Theoretical (*66*) and empirical (*67*) studies show that the *m*-value for GdnHCl denaturation of proteins is proportional to the change in solvent-accessible surface area (ΔSASA) upon unfolding. By analogy, *m*-values derived from equilibrium loop formation in the denatured state as a function of GdnHCl concentration, m_{eq}, should reflect the ΔSASA when the His−heme loop breaks. Our

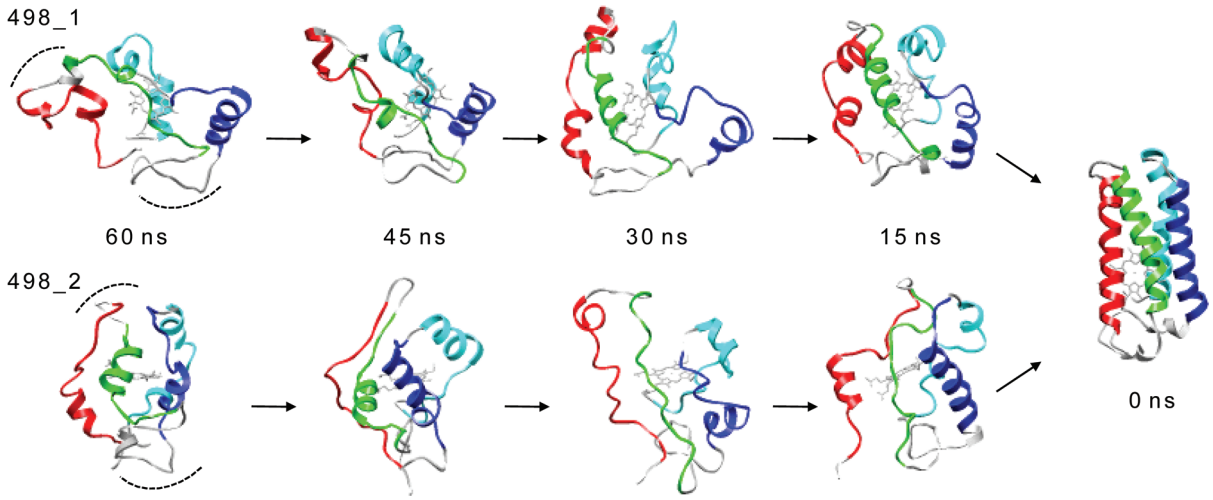

FIGURE 8: Snapshots of the folding (reverse unfolding) of Cyt*c*′ in the context of the two reverse turns. Snapshots from 60, 45, 30, and 15 ns are shown in reverse along with the starting structure. Coloring is the same as in Figure 6. The turn between helix 1 and helix 2 is oriented at the top of the protein in all structures and indicated with a dashed arc in the 60 ns snapshots. The turn between helices 2 and 3 is at the bottom in each case and marked with a dashed arc in the 60 ns structures.

previous results with iso-1-Cyt*c* were consistent with hydrophobic residual structure melting out in a loop size dependent manner as GdnHCl concentration increased from 2 to 6 M (*23*). Denaturant *m*-values as large as 0.45 kcal mol^{-1} M^{-1} were observed for $n = 37$ decreasing to 0.15 kcal mol^{-1} M^{-1} for $n = 83$. The scaling exponent, ν_3, also decreased from ~4.4 in 2 M GdnHCl to ~2.1 in 6 M GdnHCl. Thus, for Cyt*c*′ we expected that the scatter about the best fit line of pK_{loop}(His) versus log *n* would decrease dramatically in 6 M GdnHCl relative to our previous results for 3 M GdnHCl (*33*) as hydrophobic residual structure melted out. Similarly, a small decrease in ν_3 toward the center of the usual range for a random coil with excluded volume was anticipated.

Neither of these predictions is borne out (Figure 4). The pattern of deviation about the best fit line of pK_{loop}(His) versus log *n* is essentially identical; albeit the magnitude of the deviations is somewhat compressed in 6 M GdnHCl. Similarly, the scaling exponent is essentially unchanged; it does not drop from ~2.6 to near 2. We find that m_{eq} is modest for Cyt*c*′ compared to iso-1-Cyt*c*, ranging from 0.05 to 0.21 (Table S4 of the Supporting Information). Since the denatured state of Cyt*c*′ is more compact than that of iso-1-Cyt*c* (*8*), the enhancement of structure caused when loop formation constrains hydrophobic residues into a smaller volume (mass action effect (*22*, *23*)) may be attenuated. Consistent with this view, the MD simulations of the unfolding of Cyt*c*′ carried out without loop constraints yield unfolded conformations that are quite compact. Long MD simulations done at 498 K for chymotrypsin inhibitor 2 (CI2) (*68*, *69*) and engrailed homeodomain (*70*−*72*) yield considerably more expanded denatured states.

Deviations from Random Coil Behavior Are Attributable to k_b. A key observation in the kinetic data at both 3 and 6 M GdnHCl is that k_b varies over about 4−6-fold, with substantial differences between k_b for His−heme loops of similar size (compare E73H and A66H in Figure 5, top panel). Our work on iso-1-Cyt*c* also showed that deviations from random coil behavior in equilibrium His−heme loop formation data are due to loop persistence (*35*, *36*). The considerably decreased scatter about the best fit line of a plot of log k_f versus log *n*, relative to the best fit lines for the plots of pK_{loop}(His) versus log *n*, at both 3 and 6 M GdnHCl further substantiates this observation (Figure 4 versus Figure 5, bottom panel). Furthermore, a plot of the residuals from the fit of eq 5 to the pK_{loop}(His) versus *n* data in Figure 4 against k_b is significantly correlated for data acquired at both 3 and 6 M GdnHCl (Figure 5, inset, bottom panel). Thus, our loop breakage data for both Cyt*c*′ and iso-1-Cyt*c* indicate that the rim of a folding funnel for a foldable protein sequence is not smooth. This behavior contrasts starkly with the convergence toward a single value of k_b for $n > 30$ for His−heme loop breakage with low complexity sequences such as polyalanine shown in Figure 5, top panel (*56*).

It is interesting to note that k_b for Cyt*c*′ is both larger and smaller than the k_b observed for long polyalanine sequences. The physical properties of alanine are often viewed as neutral relative to other side chains. From the k_b data in Figure 5, it is evident that sequence properties can both slow and speed loop breakage relative to alanine. The variability in k_b for Cyt*c*′ could relate to differences in sequence hydrophobicity or residual secondary structure. We expect that both of these factors will contribute most to the closed loop form (and thus to k_b) due to the reduced conformational space (i.e., lower entropy) of this state. On this basis, we expect deviations from random coil behavior to be expressed more prominently in k_b than in k_f.

As discussed in the Results section, viscosity affects the rates of loop breakage (see Figure 5 inset, top panel). Thus, to estimate the relative ΔSASA for loop breakage, we calculated the denaturant *m*-value for k_b after correction for the visocity of GdnHCl, $m_{b\,corr}^{\ddagger}$, for all variants (Table S5 and Figure S3 of the Supporting Information). The values range from 0 to 0.1 kcal mol^{-1} M^{-1}, indicating that a modest increase in solvent exposure occurs en route to the transition state (TS) for loop breakage.

Since His−heme bond formation is reaction-limited not diffusion-limited (*35*, *73*, *74*), k_f should be independent of viscosity from the perspective of intrachain diffusion. Thus, we use uncorrected kinetic *m*-values to evaluate ΔSASA as the TS for loop formation is approached. The kinetic *m*-values for loop formation, $m_f^{\ddagger}$, range from 0.05 to 0.21 kcal mol^{-1} M^{-1} (Table S4 of the Supporting Information). The $m_f^{\ddagger}$ is generally similar in magnitude to m_{eq} for loop formation, indicating that most of the modest increase in burial of surface area upon loop formation occurs before the TS for loop formation. Since the *m*-value for equilibrium unfolding of Cyt*c*′ by GdnHCl is near 5 kcal mol^{-1} M^{-1} (*33*), the maximal m_{eq} and $m_f^{\ddagger}$ values correspond

to complete burial of the surface area of about 5 residues of this 125-residue protein upon loop formation.

Variation in k_b with Sequence Position and Its Relationship to Efficient Folding. If the variation in k_b with loop size seen in Figure 5 was simply due to differences in the tendency of different loops to promote residual structure, one would predict that the magnitude of k_b would tend toward a common value (as for His−heme loop formation with polyalanine sequences) at high denaturant concentration; i.e., all of the residual structure would melt out. Clearly, the pattern of variation in k_b is the same in 3 and 6 M GdnHCl (Figure 5 and Figure S3 of the Supporting Information). Thus, the sequence-dependent variation in k_b is insensitive to solvent. One possible explanation is that the groups that contribute to the variation in k_b are not well-solvated. Due to the steric constraints that likely exist around the His−heme bond, poor solvation is likely for the residues adjacent to the histidine in the primary sequence. Thus, the GdnHCl-insensitve variations in k_b can be viewed as providing a readout of the portions of the protein sequence with a high versus a low tendency to form noncovalent interactions in the DSE or early in folding.

In analogy to the zipper model of Dill and co-workers (*75*−*77*), interactions between sequentially adjacent "sticky points" in the protein sequence could provide low entropy pathways for folding by stabilizing small loops in the DSE that template the correct topology. Two low points in the k_b versus sequence position plot are at residues 58 and 73. These residues are at the ends of a 20-residue Ω loop that brings helices 2 and 3 together (cyan spheres in Figure 1). At the same time, residue 66 in the middle of this loop has a relatively fast breakage rate. Loop breakage for residue 73 is about 4-fold slower than for residue 66, which corresponds to about a 0.85 kcal/mol [$RT \ln(k_{b66}/k_{b73})$] greater tendency for residue 73 than for residue 66 to form a persistent contact in 3 M GdnHCl at 25 °C. Similarly, residue 58 would have about a 0.4 kcal/mol greater tendency to form a persistent contact in 3 M GdnHCl at 25 °C. Thus, the tendency for residues around positions 58 and 73 to form a persistent loop early in folding should be significant. A second low point in the k_b versus sequence position plot encompasses positions 20, 31, and 39. These residues demarcate the turn between helices 1 and 2 (cyan spheres in Figure 1). Relative to the surrounding residues 49 and 13, the rates of loop breakage are 1.7−3.2-fold slower for residues 20, 31, and 39. Thus, residues 20, 31, and 39 will have a 0.3−0.7 kcal/mol preference to form persistent interactions in the DSE than residues 13 and 49. These energetic biases in the DSE would likely be adequate to transiently establish the Ω loop that brings helices 2 and 3 together and the turn between helices 1 and 2 allowing for efficient folding. Thus, our His−heme loop breakage data suggest that such "sticky points" in protein sequences may be important in templating two critical chain reversals in the topology of this four helix bundle protein.

The MD simulations provide strong support for this interpretation of our loop breakage data. Dynamic hydrophobic clusters help to stabilize both the turn between helices 1 and 2 and pinch off the ends of the Ω-loop between helices 2 and 3 (Figure 7). Thus, the MD simulations show that the loops punctuate the structural ensemble of the unfolded state of Cyt*c*′, leading to a bias toward chain reversals that set up the topology of the native structure (Figure 8). These chain reversals in the DSE keep portions of the primary sequence that form neighboring helices in the native state in close proximity in the DSE even when they are not helical. Thus, helix formation is facilitated as docking can occur productively because the turns that are part of the residual structure in the DSE have set up the necessary topology.

Microsecond mixing folding experiments show changes in the fluorescence of Trp32 on a 150 μs time scale that are consistent with the helix 1−turn−helix 2 structure forming early in folding (*9*). Similar changes were not observed for Trp72 in these experiments, which was interpreted as lack of formation of the Ω loop at the base of helices 2 and 3 early in folding. Our results suggest that this large loop is present but does not assemble against the heme until later in folding. The persistence of the Ω loop at the base of helices 2 and 3 indicates that moderate-sized loops (*78*), in addition to small loops (*75*−*77*, *79*), can be important in guiding the early stages of folding. Interestingly, the high residual helical structure observed in the DSE for helix 3, in agreement with Agadir (*65*), does little to guide folding, underscoring the predominant role of the chain reversals in the DSE (Figure 8).

Physical Properties Leading to Deviations from Random Coil Behavior. If the relative stability of a loop has a significant component from the local interaction with the heme surface, then it is reasonable to expect that simple properties like local sequence hydrophobicity might correlate with loop stability. Figure S6 in the Supporting Information shows a plot of the Eisenberg hydrophobicity (*80*) of Cyt*c*′ as a function of sequence position. For the D58H through the K13H variants, local hydrophobicity correlates moderately with the scatter about the fit of eq 5 to the pK_{loop}(His) versus log *n* data in Figure 4 (Figure S7 of the Supporting Information). The correlation coefficient for this subset of the data is $R^2 = 0.49$ in 3 M GdnHCl. However, R^2 is near 0 for the entire data set. Thus, loop stability appears to be in part affected by local hydrophobicity, but other factors clearly are important.

We also considered the effects of local ϕ,ψ preferences using the Agadir algorithm (*65*) to predict helical propensity. As can be seen in Figure S5 of the Supporting Information, the variants from A91H to A66H occur in or immediately adjacent to a region of high helical propensity. While it is unlikely that long stretches of helical structure would persist in 6 M GdnHCl, the high preference for this region of ϕ,ψ space may influence the chain properties leading to the similar values of pK_{loop}(His) for the A91H, K84H, and E73H variants.

Effects of Denaturant Concentration on the DSE. The role of the denatured state in protein folding has been broadly debated (*1*−*5*). There is evidence that residual structure melts out in the presence of denaturants (*10*, *81*) and thus one can question whether the behavior of a denatured protein under strong denaturing conditions can provide insight into features of the denatured state important to folding (*82*). However, recent work indicates that osmolytes and denaturants work primarily by modulating the thermodynamics of the protein backbone with only modest effects on side chain thermodynamics (*83*, *84*). In other words, denaturants and osmolytes affect the stability but not the specificity of protein folding, in line with the thinking of Lattman and Rose regarding the effects of mutations on protein stability versus specificity (*85*).

Our His−heme loop formation data on the denatured state of Cyt*c*′ are consistent with these ideas about denaturants, the denatured state, and protein folding. In particular, local sequence variability in the denatured state is not eliminated in the presence of high concentrations of denaturants. This behavior is evident in the persistence in the pattern of scatter of pK_{loop}(His) versus *n* at 3 M compared to 6 M GdnHCl in Figure 4. It is also evident in

the lack of elimination of sequence-dependent variability in His−heme loop breakage rates, k_b, in 6 M compared to 3 M GdnHCl in Figure 5. Interestingly, the "sticky parts" of the sequence of Cyt*c*′ define important chain reversals that must be established during the folding of Cyt*c*′. The interactions expected on the basis of the His−heme loop breakage rates are also observed in the unfolded state using MD simulations. Both sets of data show that the key contacts necessary to establish the topology of a protein are prominent contributors to the DSE of this protein, clearly biasing the topology of the DSE of Cyt*c*′ toward that of the native state (Figure 8) so as to promote efficient folding.

CONCLUSIONS

The GdnHCl concentration dependence of the kinetics and thermodynamics of loop formation in the denatured state of Cyt*c*′ reveals sequence-dependent variability in the tendency of different parts of the chain to form persistent contacts. Mapping of the positions of the histidines that form the most persistent loops onto the sequence of Cyt*c*′ shows that the portions of the Cyt*c*′ sequence most likely to form persistent loops define important chain reversals in the tertiary structure of Cyt*c*′. MD simulations show that chain reversals, involving loops as large as ~20 residues, occur in the DSE in regions of the protein that form persistent His−heme loops in the presence of GdnHCl. These observations suggest that the primary sequence of a protein encodes the information necessary for a protein chain to efficiently establish its topology either as part of the DSE, as observed here, or early in folding.

SUPPORTING INFORMATION AVAILABLE

Figures S1−S7 and Tables S1−S5 as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Cho, J.-H., Saito, S., and Raleigh, D. P. (2008) Non-native electrostatic interactions in the denatured state ensemble: effects on protein stability and folding, in Unfolded Proteins: From Denatured to Intrinsically Disordered (Creamer, T. P., Ed.) pp 51−69, Nova Science Publishers, Hauppauge, NY.
2. Bowler, B. E. (2008) Thermodynamic approaches to understanding protein denatured states, in Unfolded Proteins: From Denatured to Intrinsically Disordered (Creamer, T. P., Ed.) pp 23−50, Nova Science Publishers, Hauppage, NY.
3. McCarney, E. R., Kohn, J. E., and Plaxco, K. W. (2005) Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. *Crit. Rev. Biochem. Mol. Biol. 40*, 181–189.
4. Bowler, B. E. (2007) Thermodynamics of protein denatured states. *Mol. BioSyst. 3*, 88–99.
5. Fleming, P. J., and Rose, G. D. (2005) Conformational properties of unfolded proteins, in Protein Folding Handbook, Part I (Buchner, J., and Kiefhaber, T., Eds.) pp 710−736, Wiley-VCH, Weinheim, Germany.
6. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein. *Science 295*, 1719–1722.
7. Shortle, D. (1996) Structural analysis of non-native states of proteins by NMR methods. *Curr. Opin. Struct. Biol. 6*, 24–30.
8. Lee, J. C., Engman, K. C., Tezcan, F. A., Gray, H. B., and Winkler, J. R. (2002) Structural features of cytochrome *c*′ folding intermediates revealed by fluorescence energy-transfer kinetics. *Proc. Natl. Acad. Sci. U.S.A. 99*, 14778–14782.
9. Kimura, T., Lee, J. C., Gray, H. B., and Winkler, J. R. (2007) Site-specific collapse dynamics guide the formation of the cytochrome *c*′ four-helix bundle. *Proc. Natl. Acad. Sci. U.S.A. 104*, 117–122.
10. Shan, B., Bhattacharya, S., Eliezer, D., and Raleigh, D. P. (2008) The low-pH unfolded state of the C-terminal domain of the ribosomal protein L9 contains significant secondary structure in the absence of denaturant but is no more compact than the low-pH urea unfolded state. *Biochemistry 47*, 9565–9573.
11. Mittag, T., and Forman-Kay, J. D. (2007) Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol. 17*, 3–14.
12. Saxena, A. M., Udgoankar, J. B., and Krishnamoorthy, G. (2006) Characterization of intramolecular distances and site-specific dynamics in chemically unfolded barstar: evidence for denaturant-dependent non-random structure. *J. Mol. Biol. 359*, 174–189.
13. Lyubovitsky, J. G., Gray, H. B., and Winkler, J. R. (2002) Mapping the cytochrome *c* folding landscape. *J. Am. Chem. Soc. 124*, 5481–5485.
14. Marsh, J. A., and Forman-Kay, J. D. (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J. Mol. Biol. 391*, 359–374.
15. Ensign, A. A., Jo, I., Yildirim, I., Krauss, T. D., and Bren, K. L. (2008) Zinc porphyrin: a fluorescent acceptor in studies of Zn-cytochrome *c* unfolding by fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U.S.A. 105*, 10779–10784.
16. Pletneva, E. V., Gray, H. B., and Winkler, J. R. (2005) Many faces of the unfolded state: conformational heterogeneity in denatured yeast cytochrome *c*. *J. Mol. Biol. 345*, 855–867.
17. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol. 4*, 10–19.
18. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. J., Ruczinski, I., Doniach, S., and Plaxco, K. W. (2004) Random coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12491–12496.
19. Millett, I. S., Doniach, S., and Plaxco, K. W. (2002) Toward a taxonomy of the denatured state: small angle X-ray scattering studies of unfolded proteins. *Adv. Protein Chem. 62*, 241–262.
20. Tanford, C. (1968) Protein denaturation. *Adv. Protein Chem. 23*, 121–282.
21. Kuznetsov, S. V., Shen, Y., Benight, A. S., and Ansari, A. (2001) A semiflexible polymer model applied to loop formation in DNA hairpins. *Biophys. J. 81*, 2864–2875.
22. Hammack, B. N., Smith, C. R., and Bowler, B. E. (2001) Denatured state thermodynamics: residual structure, chain stiffness and scaling factors. *J. Mol. Biol. 311*, 1091–1104.
23. Wandschneider, E., and Bowler, B. E. (2004) Conformational properties of the iso-1-cytochrome *c* denatured state: dependence on guanidine hydrochloride concentration. *J. Mol. Biol. 339*, 185–197.
24. Cantor, C. R., and Schimmel, P. R. (1980) Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules, W. H. Freeman, San Francisco.
25. Jacobson, H., and Stockmayer, W. H. (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Chem. Phys. 18*, 1600–1606.
26. Chan, H. S., and Dill, K. A. (1990) The effects of internal constraints on the configurations of chain molecules. *J. Chem. Phys. 92*, 3118–3135.
27. de Gennes, P.-G. (1979) Scaling Concepts in Polymer Physics, Cornell University Press, Ithaca, NY.
28. Redner, S. (1980) Distribution functions in the interior of polymer chains. *J. Phys. A: Math. Gen. 13*, 3525–3541.
29. Feige, M. J., and Paci, E. (2008) Rate of loop formation in peptides: a simulation study. *J. Mol. Biol. 382*, 556–565.
30. Jha, A. K., Colubri, A., Freed, K. F., and Sosnick, T. R. (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl. Acad. Sci. U.S.A. 102*, 13099–13104.
31. Tran, H. T., Wang, X., and Pappu, R. V. (2005) Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. *Biochemistry 44*, 11369–11380.
32. Fitzkee, N. C., and Rose, G. D. (2004) Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12497–12502.
33. Rao, K. S., Tzul, F. O., Christian, A. K., Gordon, T. N., and Bowler, B. E. (2009) Thermodynamics of loop formation in the denatured state of *Rhodopseudomonas palustris* cytochrome *c*′: scaling exponents and the reconciliation problem. *J. Mol. Biol. 392*, 1315–1325.
34. Shibata, N., Iba, S., Misaki, S., Meyer, T. E., Bartsch, R. G., Cusanovich, M. A., Morimoto, Y., Higuchi, Y., and Yasuoka, N. (1998) Basis for monomer stabilization in *Rhodopseudomonas palustris* cytochrome *c*′ derived from the crystal structure. *J. Mol. Biol. 284*, 751–760.

35. Kurchan, E., Roder, H., and Bowler, B. E. (2005) Kinetics of loop formation and breakage in the denatured state of iso-1-cytochrome *c*. *J. Mol. Biol. 353*, 730–743.
36. Tzul, F. O., and Bowler, B. E. (2009) Importance of contact persistence in denatured state loop formation: kinetic insights into sequence effects on nucleation early in folding. *J. Mol. Biol. 390*, 124–134.
37. Arslan, E., Schulz, H., Zufferey, R., Künzler, P., and Thöny-Meyer, L. (1998) Overproduction of the *Bradyrhizobium japonicum c*-type cytochrome subunits of the cbb_3 oxidase in *Escherichia coli*. *Biochem. Biophys. Res. Commun. 251*, 744–747.
38. McGuirl, M. A., Lee, J. C., Lyubovitsky, J. G., Thanyakoop, C., Richards, J. H., Gray, H. B., and Winkler, J. R. (2003) Cloning, heterologous expression, and characterization of recombinant class II cytochromes *c* from *Rhodopseudomonas palustris*. *Biochim. Biophys. Acta 1619*, 23–28.
39. Lee, J. C., Gray, H. B., and Winkler, J. R. (2001) Cytochrome *c′* folding triggered by electron transfer: fast and slow formation of four-helix bundles. *Proc. Natl. Acad. Sci. U.S.A. 98*, 7760–7764.
40. Tonomura, B., Nakatani, H., Ohnishi, M., Yamaguchi-Ito, J., and Hiromi, K. (1978) Test reactions for a stopped-flow apparatus. Reduction of 2,6-dicholorophenolindophenol and potassium ferricyanide by L-ascorbic acid. *Anal. Biochem. 84*, 370–383.
41. Beck, D. A. C., Jonsson, A. L., Schaeffer, R. D., Scott, K. A., Day, R., Toofanny, R. D., Alonso, D. O. V., and Daggett, V. (2008) Dynameomics: mass annotation of protein dynamics and unfolding in water by high-throughput atomistic molecular dynamics simulations. *Protein Eng. Des. Sel. 21*, 353–368.
42. Van der Kamp, M. W., Anderson, P. C., Beck, D. A. C., Benson, N. C., Jonsson, A. L., Merkley, E. D., Schaeffer, R. D., Scouras, A. D., Simms, A., Toofanny, R. D., and Daggett, V. (2010) Dynameomics: a comprehensive database of protein dynamics. *Structure 18*, 423–435.
43. Levitt, M., Hirshberg, M., Sharon, R., and Daggett, V. (1995) Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Comput. Phys. Commun. 91*, 215–231.
44. Beck, D. A. C., and Daggett, V. (2004) Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods 34*, 112–120.
45. Beck, D. A. C., Armen, R. S., and Daggett, V. (2005) Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. *Biochemistry 44*, 609–616.
46. Day, R., Bennion, B. J., Ham, S., and Daggett, V. (2002) Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol. 322*, 189–203.
47. Haar, L., Gallagher, J. S., and Kell, G. S. (1984) NBS/NRC Steam Tables: Thermodynamics and Transport Properties and Computer Programs for Vapor and Liquid States of Water in SI Units, Hemisphere Publishing Corp., Washington, DC.
48. Levitt, M., Hirshberg, M., Sharon, R., Laidig, K., and Daggett, V. (1997) Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J. Phys. Chem. B 101*, 5051–5056.
49. Beck, D. A. C., Alonso, D. V. O., and Daggett, V. (2000−2010) *In lucem* molecular mechanics (ilmm), University of Washington, Seattle, WA.
50. Armen, R. S., Bernard, B. M., Day, R., Alonso, D. O. V., and Daggett, V. (2005) Characterization of a possible amyloidogenic precursor in glutamine-repeat neurodegenerative diseases. *Proc. Natl. Acad. Sci. U.S.A. 102*, 12433–12438.
51. Godbole, S., and Bowler, B. E. (1997) A histidine variant of yeast iso-1-cytochrome *c* that strongly affects the energetics of the denatured state. *J. Mol. Biol. 268*, 816–821.
52. Cho, J.-H., Sato, S., and Raleigh, D. P. (2004) Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state. *J. Mol. Biol. 338*, 827–837.
53. Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. *Biochemistry 38*, 4896–4903.
54. Monera, O. D., Kay, C. M., and Hodges, R. S. (1994) Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci. 3*, 1984–1991.
55. Smith, C. R., Wandschneider, E., and Bowler, B. E. (2003) Effect of pH on the iso-1-cytochrome *c* denatured state: changing constraints due to heme ligation. *Biochemistry 42*, 2174–2184.
56. Tzul, F. O., and Bowler, B. E. (2010) Denatured states of low complexity polypeptide sequences differ dramatically from those of foldable sequences. *Proc. Natl. Acad. Sci. U.S.A. 107*, 11364–11369.
57. Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M., and Kiefhaber, T. (1999) The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci. U.S.A. 96*, 9597–9601.
58. Möglich, A., Krieger, F., and Kiefhaber, T. (2005) Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. *J. Mol. Biol. 345*, 153–162.
59. Jacob, M., Geeves, M., Holtermann, G., and Schmid, F. X. (1999) Diffusional barrier crossing in a two-state protein folding reaction. *Nat. Struct. Mol. Biol. 6*, 923–926.
60. Qiu, L., and Hagen, S. J. (2004) A limiting speed for protein folding at low solvent viscosity. *J. Am. Chem. Soc. 126*, 3398–3399.
61. Qiu, L., and Hagen, S. J. (2005) Internal friction in the ultrafast folding of the tryptophan cage. *Chem. Phys. 312*, 327–333.
62. Plaxco, K. W., and Baker, D. (1998) Limited internal friction in the rate-limited step of a two-state protein folding reaction. *Proc. Natl. Acad. Sci. U.S.A. 95*, 13591–13596.
63. Hagen, S. J., Qui, L., and Pabit, S. A. (2005) Diffusional limits to the speed limit of protein folding: fact or friction? *J. Phys.: Condens. Matter 17*, S1503–S1514.
64. Tzul, F. O., Kurchan, E., and Bowler, B. E. (2007) Sequence composition effects on denatured state loop formation in iso-1-cytochrome *c* variants: polyalanine versus polyglycine inserts. *J. Mol. Biol. 371*, 577–584.
65. Muñoz, V., and Serrano, L. (1994) Elucidating the folding problem of α-helical peptides using empirical parameters III: temperature and pH dependence. *J. Mol. Biol. 245*, 297–308.
66. Schellman, J. A. (1978) Solvent denaturation. *Biopolymers 17*, 1305–1322.
67. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: relation to changes in accessible surface area of protein unfolding. *Protein Sci. 4*, 2138–2148.
68. Kazmirski, S. L., Wong, K.-B., Freund, S. M. V., Tan, Y.-J., Fersht, A. R., and Daggett, V. (2001) Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A. 98*, 4349–4354.
69. Day, R., and Daggett, V. (2005) Ensemble versus single-molecule protein unfolding. *Proc. Natl. Acad. Sci. U.S.A. 102*, 13445–13450.
70. Mayor, U., Johnson, C. M., Daggett, V., and Fersht, A. R. (2000) Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. U.S.A. 97*, 13518–13522.
71. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M. V., Alonso, D. O. V., Daggett, V., and Fersht, A. R. (2003) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature 421*, 863–867.
72. DeMarco, M. L., Alonso, D. O. V., and Daggett, V. (2004) Diffusing and colliding: the atomic level folding/unfolding pathway of a small helical protein. *J. Mol. Biol. 341*, 1109–1124.
73. Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) Diffusion-limited contact formation in unfolded cytochrome *c*: estimating the maximum rate of protein folding. *Proc. Natl. Acad. Sci. U.S.A. 93*, 11615–11617.
74. Hagen, S. J., Hofrichter, J., and Eaton, W. A. (1997) Rate of intrachain diffusion of unfolded cytochrome *c*. *J. Phys. Chem. B 101*, 2352–2365.
75. Dill, K. A., Fiebig, K. M., and Chan, H. S. (1993) Cooperativity in protein-folding kinetics. *Proc. Natl. Acad. Sci. U.S.A. 90*, 1942–1946.
76. Weikl, T. R., and Dill, K. A. (2003) Folding kinetics of two-state proteins: effect of circularization, permutation, and crosslinks. *J. Mol. Biol. 332*, 953–963.
77. Weikl, T. R., and Dill, K. A. (2003) Folding rates and low-entropy-loss routes of two-state proteins. *J. Mol. Biol. 329*, 585–598.
78. Orevi, T., Ishay, E. B., Pirchi, M., Jacob, M. H., Amir, D., and Haas, E. (2009) Early closure of a long loop in the refolding of adenylate kinase: a possible key role of non-local interactions in the initial folding steps. *J. Mol. Biol. 385*, 1230–1242.
79. Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008) The protein folding problem. *Annu. Rev. Biophys. 37*, 289–316.
80. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol. 179*, 125–142.
81. Zhang, O., and Forman-Kay, J. D. (1997) NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. *Biochemistry 36*, 3959–3970.

82. Cho, J.-H., and Raleigh, D. P. (2009) Experimental characterization of the denatured state ensemble of proteins. *Methods Mol. Biol.* (*Totowa, NJ*) *490*, 339–351.
83. Auton, M., Holthauzen, L. M. F., and Bolen, D. W. (2007) Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Natl. Acad. Sci. U.S.A. 104*, 15317–15322.
84. Bolen, D. W., and Rose, G. D. (2008) Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem. 77*, 339–362.
85. Lattman, E. E., and Rose, G. D. (1993) Protein folding—what's the question? *Proc. Natl. Acad. Sci. U.S.A. 90*, 439–441.