2013

THERMODYNAMIC AND KINETIC PROPERTIES OF CYTOCHROMES C AND C' IN THE DENATURED STATE: PROPENSITY FOR RESIDUAL STRUCTURE

Travis Arthur Danielson
The University of Montana

THERMODYNAMIC AND KINETIC PROPERTIES OF CYTOCHROMES C AND C' IN THE DENATURED STATE: PROPENSITY FOR RESIDUAL STRUCTURE

By

TRAVIS ARTHUR DANIELSON

Bachelor of Arts in Biology, The University of Montana, Missoula, MT, 2001
Bachelor of Arts in Psychology, The University of Montana, Missoula, MT, 2006

Thesis

presented in partial fulfillment of the requirements
for the degree of

Master of Science
in Biochemistry and Biophysics

The University of Montana
Missoula, MT

May 2013

Approved by:

Sandy Ross, Dean of The Graduate School
Graduate School

Bruce Bowler, Chair
Department of Chemistry and Biochemistry

J. Stephen Lodmell
Division of Biological Sciences

Klara Briknarova
Department of Chemistry and Biochemistry

Chapter 1: Loop Formation of Cytochrome *c'* in the Denatured State: Effect of Residual α-Helical Structure

Abstract

To expand our understanding of helical propensity in the denatured state, we have engineered serine in place of alanine near the center of the third helix (positions 83 and 87) in cytochrome *c'* (Cyt*c'*) and have measured histidine-heme loop formation in 3 M gdnHCl. A series of thirteen variants that include the A83S/A87S substitutions have been engineered with single surface histidine substitutions to provide a range of His-heme loop sizes from 10 to 111 residues and to provide direct comparison to previous studies with pWT Cyt*c'* . We observe decreased global stability for most variants and an average decrease of the midpoint of gdnHCl unfolding, C_m, by 0.197 ± 0.037 gdnHCl molar units for A83S/A87S variants compared to pWT. Loop stability versus loop size data yields a scaling exponent of 2.24 ± 0.24, similar to the pWT value of 2.5 ± 0.3, but with a crossover point and values that suggest that loop flexibility and stability increase in loops that contain the A83S/A87S substitutions. Kinetic data show nonrandom behavior in the denatured state similar to pWT, and also supports the suggestion of helical propensity only being important if the helical segment is contained within the loop, as k_f values are slightly larger for loops containing A83S/A87S. Molecular dynamic simulations of the thermal unfolding of Cyt*c'* show a perhaps surprising amount of structure retention in the third helix, even with the helical propensity lowering A83S/A87S substitutions. The data from these studies offer insight into residual structure provided by helical propensity in the denatured state.

Introduction

Early studies on protein folding reported the denatured state to be random coils without residual non-covalent structure, though did allow for the possibility of regions of ordered structure after chemical denaturation.[2] The random coil model persisted until the early 1990's when NMR studies showed chemical shifts that differed greatly from random coil values for a globular urea unfolded protein, the amino-terminal 63-residue domain of the 434-repressor.[3] This led to Nuclear Overhauser effect measurements of the protein in 7 M urea, and revealed a hydrophobic cluster corresponding to a region in the native protein.[4] During the same period of

time, thermodynamic studies of site-directed variants of staphylococcal nuclease also showed that the unfolded state was considerably more multifaceted than previously thought.[5] Substantiation of residual structure has been gathered through several studies using NMR and fluorescence resonance energy transfer (FRET).[6; 7; 8; 9] Levinthal's conjectured conformational search[10] will be greatly limited if pockets of residual structure provide nucleation sites for folds found in the native protein. These residual structures greatly contribute to the funneling of folding pathways to a single stable state by multiple routes in conformational space.[11] Currently, nonrandom structures of intermediate complexity in protein denatured states are widely accepted to occur.[12; 13]

Denatured state histidine-heme loop formation has been recently used to study a class II *c*-type cytochrome, cytochrome *c'* (Cyt*c'*) from *Rhodopseudomonas palustris* in 3 M guanidine-HCl (gdnHCl).[14; 15] These studies observed considerable scatter about the linear dependence of loop stability on loop size, divergent from data obtained with peptides of homogeneous sequence, revealing underlying residual structure. The observed scaling exponent, ν_3, for a freely joined random coil with excluded volume is 1.8 to 2.4.[16; 17] The scaling exponent for Cyt*c'* is very close to this range at 2.5 ± 0.3. The methodology was adapted from denatured state studies of iso-1-cytochrome *c*.[18; 19] This methodology takes advantage of the unique His-heme binding motif formed under denaturing conditions, as found in Figure 1. A loop is formed with the ionizable sidechain of histidine N-terminal to the site of heme attachment. Since the native sequence of Cyt*c'* has no other histidine residues besides His117, which is part of the heme binding motif, each histidine variant forms a unique loop. Thirteen single histidine mutants gave loop sizes between 10 and 111 residues as the size of the loop is determined by the position of the engineered histidine. The positions of the engineered histidines are found in Figure 2.

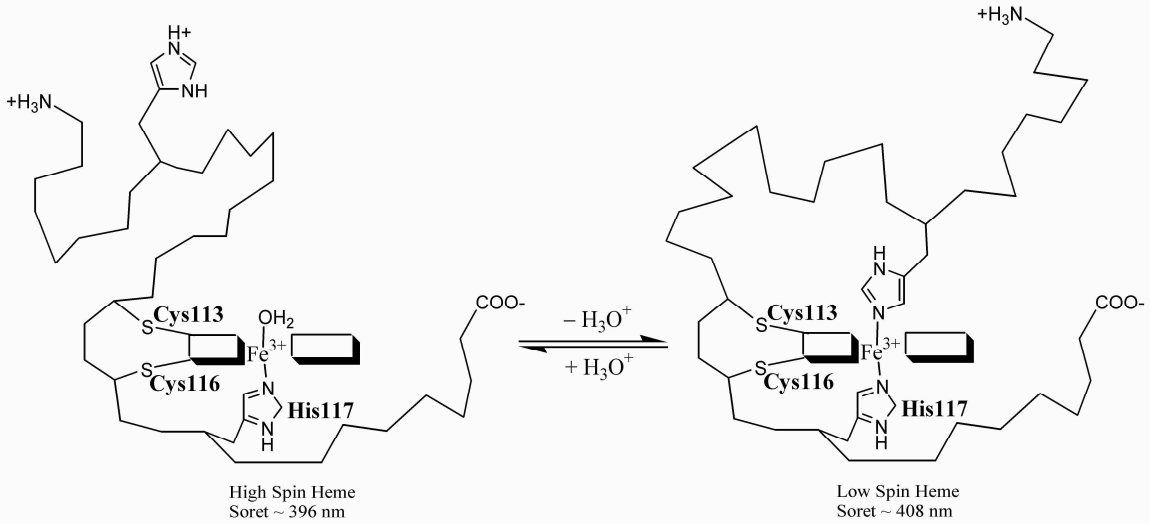


Figure 1: Schematic representation of His–heme loop formation in the denatured state of Cytc′.

Molecular dynamic simulations were also performed on Cyt*c'* to simulate the thermal unfolding pathway. The major residual structure throughout the simulations was within helix 3, approximately residues 75-95. This segment of residual structure was in agreement with the α-helix propensity predicted by the AGADIR algorithm. [20] To expand our understanding of helical propensity in the denatured state, we have engineered serine in place of alanine at positions 83 and 87. This placement and type of substitution is predicted to decrease the helical propensity of the third helix (Fig. 3 in Results).[21] By disrupting this helix and using the same set of engineered histidine loops, direct comparisons may be made, and insight into the effect of the propensity of helical structure on denatured state residual structure may be gained.

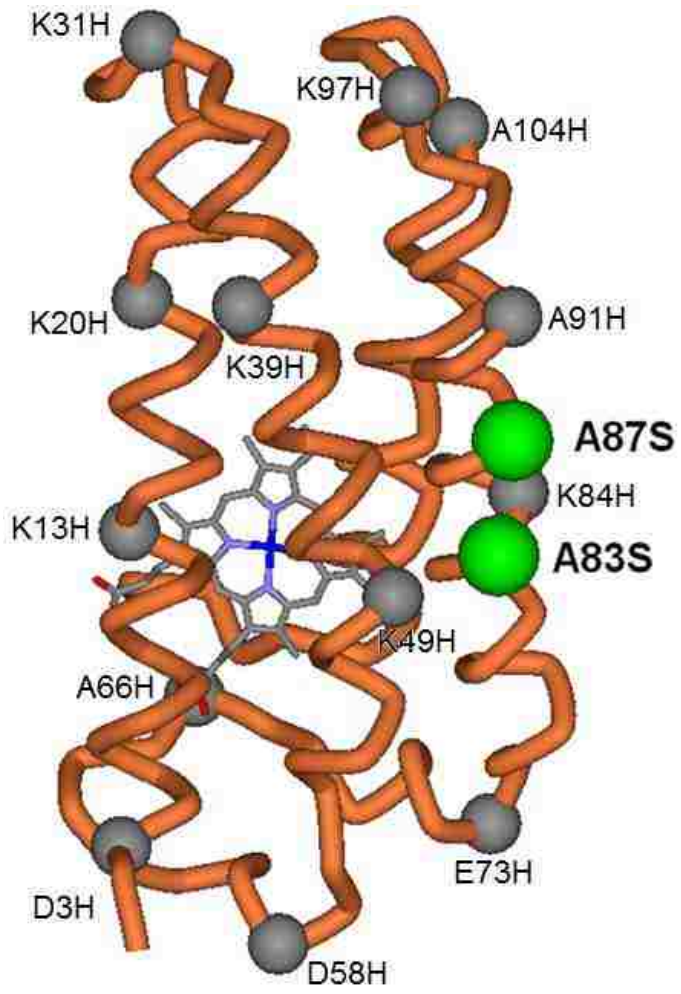


Figure 2: Structure of R. palustris Cytc′ (1A7V)[1] showing the positions of histidine and serine substitutions. The heme is shown in gray and blue (iron atom). The positions of the residues mutated in each histidine variant are shown with gray spheres. The positions of the residues mutated to serine are shown with green spheres. Mutations are indicated with standard notation next to each sphere.

Experimentals

Preparation of Cytochrome *c'* Variants by Mutagenesis

Using Stratagene's primer design program, primers were designed to introduce mutations to the pETcp vector[22] containing the pseudo wild type (pWT, Q1A mutation) gene for Cytochrome *c'*. The first mutagenesis performed was to mutate positions 83 and 87 from alanine to serine. The other single histidine mutations used this pWT A83S/A87S containing vector as a template. Site directed mutagenesis was performed using the Quikchange II PCR-based mutagenesis kit (Stratagene). The program used for PCR was 16 cycles of 30 seconds at 95 °C followed by 6 min at 68 °C. This program combines the annealing and extension steps of PCR to account for the high GC content of the Cyt*c'* gene. Typically, 2 µL of the PCR product was transformed into 100 µL of competent TG-1 *Escherichia coli* cells by heat shock as follows. A TG-1 cell aliquot was

removed from a -80° freezer and thawed on ice. PCR product was added and gently mixed by pipetting or by flicking the tube, which was then quickly centrifuged for ~2 seconds on a tabletop centrifuge equilibrated to 4°C to collect the solution at the bottom of the tube but without pelleting the cells. The cells were incubated on ice for 45 minutes, then transferred to a 42 °C water bath for 45 seconds before cooling on ice again for 5 minutes. The cells were transferred to a 15 mL tube containing 1 mL LB media and were placed in a 37 °C shaking water bath for 1 hour. 300 µL of the inoculated media was plated on LB-agar plates containing 100 µg/mL ampicillin and left to grow overnight at 37 °C. Five colonies from the plate were picked and used to inoculate individual solutions of 5 mL LB media containing 100 µg/mL ampicillin and left overnight to grow at 37 °C in a shaking water bath. DNA was extracted using the Promega Wizard Plus DNA miniprep kit and sent to the Murdock DNA Sequencing Facility at the University of Montana. T7 primers were used for sequencing.

Expression of Cytochrome *c'*

Cells were transformed with both pETcp (ampicillin resistance) and pEC86 (chloramphenicol resistance) vectors by heat shock. Typically 2 µL of pETcp of approximately 100 ng/µL concentration plus 1 µL of pEC86 of approximately 50 ng/µL concentration were added to a 5 µL aliquot of competent BL21-DE3 *E. coli* cells (Edge Bio) on ice and then the procedure described above for TG-1 cells was followed. The cells were plated on LB-agar plates containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol and left overnight to grow in a 37 °C incubator. The resulting full plate of colonies was resuspended in 3 mL sterile LB media and used to inoculate 1 L cultures of Terrific Broth (12 g/L tryptone, 24 g/L yeast extract, 4 mL glycerol, 2.3 g/L KH_2PO_4, 12.5g K_2HPO_4) also containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. These cultures were grown overnight at 30 °C or 37 °C with shaking at 165-200 rpm. Cell pellets were harvested using a Sorvall GS-3 rotor in a Sorvall RC-5C Plus floor model centrifuge running at 5000 rpm for 10 minutes. The supernatant was discarded and the resulting cell pellet was stored at -80 °C.

Lysis and Crude Purification of Cytochrome *c'* Variants

Three freeze (-80 °C)/thaw (4 °C) cycles were used to begin the lysis process. After the freeze/thaw cycles, the cell pellets were resuspended in room temperature lysis buffer (30% sucrose, 1 mM EDTA, 2mM PMSF, 30 mM Tris-HCl, pH 8.0) using a magnetic stir bar and stirplate or an orbital shaker at 200 rpm. This cell slurry was centrifuged using a GSA rotor in a Sorvall RC-5C+ floor model centrifuge at 10,000 rpm for 20 minutes. The cell pellet was discarded. The crude protein extract solution was adjusted to pH 5.0 using glacial acetic acid. This resulted in precipitation of many non-target proteins. The lysate was cleared by again centrifuging at 10,000 rpm for 20 minutes, with insoluble proteins formed into a pellet and discarded. A CM Sepharose column of approximately 100 mL was packed and equilibrated with CM Low buffer (5 mM NaOAc, pH 5.0). The protein extract solution was top loaded onto the Sepharose column with the assistance of a peristaltic pump, resulting in a dark orange/brown band on the top of the column. At least 1-2 column volumes of CM Low buffer was run through the column to wash non-Sepharose binding proteins away. Proteins that did bind to the Sepharose were eluted using a 200 mL linear salt gradient from 0 to 500 mM NaCl using CM High (5 mM NaOAc, 500 mM NaCl, pH 5.0).

Purification of Cytochrome *c'* Variants by HPLC

Protein solution eluted from the CM Sepharose column was exchanged into HPLC Buffer A (10 mM sodium phosphate, pH 6.0) using Millipore Centricon and Centriprep 10,000 MWCO (molecular weight cut-off) centrifugal filters running on a benchtop Jouan CR4i centrifuge. Purification continued by HPLC. Protein was loaded onto a Waters AP-1 ProteinPak SP-8HR cation exchange column and eluted with HPLC Buffer B (10 mM sodium phosphate, 500 mM NaCl, pH 6.0). The gradient used was as follows: 0-10 minutes 0% B, 10-11 minutes increase linearly to 2% B, 11-31 minutes linear increase to 12% B, 31-45.9 minutes increase linearly to 15%, 45.9-46 minutes increase linearly to 100% B, 46-58 minutes hold at 100% B, 58-58.1 minutes decrease to 0% B, 58.1-68 minutes hold at 0% B. The eluate was collected from the largest peak and exchanged back to HPLC Buffer A using the above mentioned Millipore centrifugal filters. A MALDI-TOF mass spectrum was collected to check purity, and this gradient consistently produced protein with one peak and at the expected mass. A Beckman Coulter DU

800 Spectrophotometer was used to check the protein concentration. As the spectrum is pH dependent, the concentration was measured in 100 mM sodium phosphate buffer at pH 7.0 using ε_{398} = 85,000 M^{-1} cm^{-1}.[23]

Global Thermodynamic Stability Measurements

To measure global stability, guanidine hydrochloride (gdnHCl) denaturation was employed and monitored by circular dichroism spectroscopy. Protein solution was exchanged back into HPLC buffer A via Centricon and Centriprep centrifugal filters and concentrated. Protein was then exchanged into CD Buffer (20 mM MES, 40 mM NaCl, pH 6.5) by loading the protein solution onto a G-25 column equilibrated in CD Buffer. Eluted protein concentration was checked by the method described above. A two syringe Hamilton Microlab 500 titrator was coupled with an Applied Photophysics Chirascan circular dichroism spectrometer. A volume of 850 µL of 4 µM native protein in CD Buffer was denatured by titrating in fully denatured protein of the same concentration in 6 M gdnHCl via the titrator unit. The action of the titrator unit was to remove an appropriate amount of solution from the sample cuvette and then substitute in denatured protein stock to achieve the next required overall concentration of gdnHCl. This process was controlled by the Chirascan software. The data were acquired in steps of 0.1 from 0 M to 0.5 M and 2.5 M to 3 M gdnHCl (baseline regions), but in steps of 0.05 from 0.5 M to 2.5 M gdnHCl (anticipated transition region). The steps were adjusted after the first run of CD to account for the actual transition region for each variant. A small stir bar was present in the cuvette and would stir at approximately 200 rpm for 100 seconds after addition of new denaturant stock to make certain mixing and equilibration was complete. The sample would settle without stirring for 10 seconds and data would be acquired. Ellipticity was monitored at 222 nm, diagnostic of α-helix, and 250 nm for 25 seconds. The 250 nm signal was subtracted as a baseline. All data were acquired at 25 °C.

Denatured State Loop Formation

His-heme loop stabilities of cytochrome *c'* variants were measured in the denatured state (3 M gdnHCl, 5 mM Na_2HPO_4 and 15 mM NaCl) by pH titration. Loop equilibria were monitored at the heme Soret band (398 nm). Titrations were carried out at room temperature (22 ± 1 °C).

The procedure for titration is as follows. A 3 × protein in 3 × buffer (9 µM protein concentration, 15 mM Na_2HPO_4, and 45 mM NaCl) stock solution was made. The refractive index of an approximately 6 M gdnHCl solution was measured using water as background to calculate the accurate concentration of gdnHCl using Equation 1.

Eq. 1: Molarity of gdnHCl = $57.147(\Delta N) + 38.68(\Delta N)^2 - 91.60(\Delta N)^3$ where ΔN is the difference in refractive index between the denaturant solution and water at the sodium D line.[24]

A 3 mL sample was created by the combination of appropriate amounts of 3 × 3 × protein buffer stock, 6 M gdnHCl, and water to create a 3 µM protein solution in 1 × buffer at 3 M gdnHCl. The pH was measured with a Denver Instruments UB-10 UltraBasic benchtop pH meter. To measure spectra of the sample, 60 µl of the sample was extracted and deposited into a clean Beckman Coulter MICRO CELL cuvette and spectra were acquired from 350 to 450 nm using a Beckman Coulter DU 800 spectrophotometer. To change the pH without changing the protein or gdnHCl concentration of the sample, a total of 60 µl of 3 × protein stock (20 µl), ~6 M gdnHCl (~30 µl), and HCl (~10 µl) in appropriate ratios was added back to the sample and thoroughly mixed by pipetting. The volume of HCl added was constant, but the concentration of the acid varied to adjust the pH of the sample by ~0.2 pH units with each addition. Between each pH change, the cuvette was rinsed with water, then ethanol, and dried. At least 3 runs were completed on each variant. The absorbance at 450 nm was used as a baseline and subtracted from the absorbance from 350 to 450 nm. The adjusted absorbance at 398 nm was plotted vs. pH and fit to a modified Henderson-Hasselbalch equation (Equation 4 found in Results) allowing for calculation of the apparent pK_a of loop formation, $pK_a(obs)$, and the number of protons, n_p, in the process. The spectra were also plotted across the pH range to identify the isosbestic point.

Kinetics of His-heme Loop Breakage

To monitor the rates of the His-heme loop breakage in the denatured state, stopped-flow mixing experiments were performed at 25 ± 0.1 °C in 3.0 M gdnHCl. These stopped-flow mixing experiments were carried out on an Applied Photophysics SX20 stopped-flow spectrophotometer with a 10 mm path length 20 µL flow cell. Dead time of the instrument was

0.7 ms as determined by the reduction of 2,6-dichlorophenolindophenol with ascorbate.[25] The experiments consisted of a 1:1 mixing of 6 µM protein stock solution in 10 mM MOPS, pH 6.80, at 3.0 M gdnHCl with a 100 mM sodium citrate buffer, pH 3.0 or 3.5, at 3.0 M gdnHCl to achieve the desired ending pH of 3.1 or 3.6 at 3 µM protein. The loop breakage was monitored at 398 nm to observe the Soret band shift. The final pH of each experiment was measured from the reaction products (stopped flow effluent) immediately following the reaction. Initial preparation of the protein was the above described HPLC methods, then exchange into 10 mM MOPS, pH 6.80, though a G25 Sephadex size exclusion chromatography column equilibrated in the MOPS buffer. Seven to fifteen stopped-flow runs were collected for each variant.

Helical Propensity Prediction

AGADIR (http://agadir.crg.es/) is a web-based tool for predicting the helical content of peptides.[20] The sequences of pWT and pWT/A83S/A87S were used as input. The default parameters (5 °C, pH 7, ionic strength of 0.1) of the tool were used in the calculation.

Molecular Dynamics Simulations

To study the unfolding pathway, molecular dynamic simulations were run on PDB 1A7V[1] as the starting structure for wild type, WT(1A7V),cytochrome *c'*, with and without the A83S/A87S mutations introduced into the 1A7V structure. Equipment and procedures from the Daggett lab were utilized to perform all simulations using their *in lucem* molecular mechanics (*il*mm) system. One simulation was performed at 298 K for each protein to serve as a native state control, but two simulations were performed at 498 K for each protein to model the unfolding pathway. Standard protocols for the Daggett lab were used for the simulations[26; 27; 28], and are outlined in Dar et al.[15] Briefly, the energy of the input structure was minimized, solvated in a rectangular water box, and then the whole system underwent another round of minimization. Each atom was assigned a starting velocity according to a Boltzman distribution, and the runs were performed.

Results

Design of Cytochrome *c'* Variants

To examine helical stability, we first used AGADIR to estimate the stability of the helices of Cytochrome *c'*. As seen in Figure 3, AGADIR calculations using pWT as the input sequence predicted a substantial rise in helical prediction for positions contained within helix 3 (positions 75-95), peaking at over 35% helix at position 83. The other three helices of Cyt*c'* showed far lower predicted helical propensity. Helix 1 (residues 3-26), peaked at approximately 9% predicted helicity, while helix 2 (residues 35-53) and helix 3 (residues 100-120) both peaked close to only 5% predicted helicity, giving us the clear choice to use helix 3 for this study. To examine the persistence of this helix, we chose serine to replace alanine at positions 83 and 87. These positions in the center of the helix should have a strong disrupting effect.[29] When the A83S/A87S mutations were added to pWT, the predicted helical content was less than 5% for the entire helix 3 region. Dar et al.[15] showed MD simulations where helix 3 of pWT Cyt *c'* presents good agreement with AGADIR algorithm predictions, and shows residual helical structure in the denatured state ensemble. The goal of mutating positions 83 and 87 to serine is to examine the linkage between the persistence of helix formation and properties of loop formation when the protein is subjected to denaturing conditions.

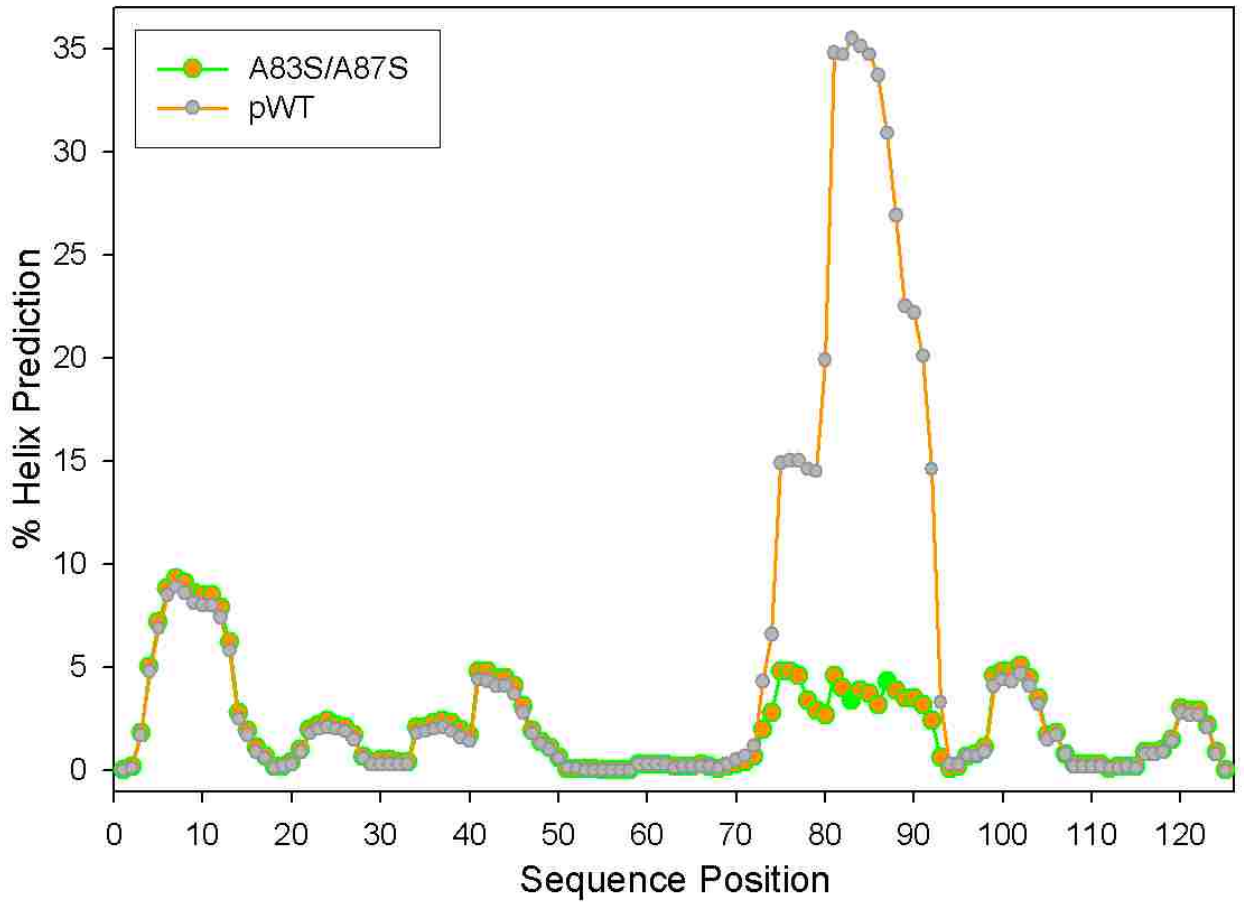


Figure 3: AGADIR calculations show predicted effects of mutations on the % α-helix of helix 3 of Cytochrome *c'* pWT vs. pWT/A83S/A87S.

Cyt*c'* can form loops between ionizable amino acid side chains, such as histidines that are good ligands, and the heme when the protein is in the denatured state. To take advantage of the 5-coordinate heme found in Cyt*c',* Rao et al.[14] designed thirteen single histidine variants to form loops with the heme in the denatured state. These loops, which form with an engineered histidine from the N-terminal side of the heme, span from 10 to 111 residues in length, a benefit of the attachment of the heme near the C-terminus though Cys 113, Cys 116 and His 117. The wild-type sequence of the protein has only one histidine, at position 117, which is the sole axial ligand to the heme. Since the wild-type sequence contains an N-terminal glutamine that can cyclize to pyroglutamate during expression from *E. coli*, our pseudo wild-type (pWT) contains a Q1A substitution to avoid a mixture of glutamine and pyroglutamate. To minimize

any effects on stability, positions selected for the single histidine mutants are solvent exposed.[14] The same thirteen single histidine variants were created with the A83S/A87S mutations so that direct comparisons may be made to pWT.

Global Stability of pWT A83S/A87S Variants

To evaluate the effect of destabilizing helix 3 with the A83S/A87S mutations on the stability of Cyt*c′*, the global stability of each variant was measured by gdnHCl denaturation monitored by circular dichroism spectroscopy (Figure 4). The ellipticity data at 222 nm versus gdnHCl concentration were fitted to a six parameter equation that evaluates slope and intercept of the native and denatured state baselines and uses a linear free energy relationship in the transition region.[30] Equations 2 and 3[18] provide the values for the free energy of unfolding in the absence of denaturant, $\Delta G_u^\circ(H_2O)$, midpoints of unfolding (C_m), and the *m*-value, which are collected in Table 1. Equation 3 allows for a six-parameter fitting of the ellipicity data at 222 nm (with ellipticity data at 250 nm subtracted as a baseline), which evaluates slope and intercept of the native (m_N, θ_N) and denatured state (m_D, θ_D) baselines and uses a linear free energy relationship (Eq. 2) and a two-state assumption to calculate ΔG_u in the transition region. [30]

Eq. 2 $$\Delta G_u = \Delta G^\circ_u(H_2O) - m[\text{denaturant}]$$

Eq. 3 $$\theta_{222} = \frac{(\theta_N + m_N[\text{gdnHCl}]) + (\theta_D + m_D[\text{gdnHCl}])e^{\left[\left(m[\text{gdnHCl}] - \Delta G^{\circ'}_u(H_2O)\right)/RT\right]}}{1 + e^{\left[\left(m[\text{gdnHCl}] - \Delta G^{\circ'}_u(H_2O)\right)/RT\right]}}$$

Figure 4 illustrates the changes in the curves for variants containing the A83S/A87S mutations. As compared to pWT, the A83S/A87S variant has a shifted curve to the left, indicating that less gdnHCl is required to denature the protein, and therefore the $\Delta G^\circ_u(H_2O)$ has been lowered. A rough relationship between denatured state loop size and global stability starts to emerge, as A83S/A87S variants with longer loop sizes are closer to the thermodynamic properties of A83S/A87S than are shorter loops. For example, K13H/A83S/A87S (loop size of 101 residues)

has a stability closer to A83S/A87S than E73H/A83S/A87S (loop size of 41 residues), which has a curve shifted towards lower gdnHCl concentration.

As seen in Figure 5, C_m values are lower for the A83S/A87S variants as compared to the pWT counterparts. As a general trend, C_m decreases as loop lengths get shorter. Figure 4 gives the amount of change that introducing A83S/A87S has on the C_m of each Cyt*c'* variant. On average, C_m decreases by 0.197 ± 0.037 gdnHCl molar units for A83S/A87S variants compared to pWT. All A83S/A87S variants, with the exception of K84H and A91H, achieved the average ΔC_m within error (Figure 6). The change between pWT and pWT/A83S/A87S for those variants were lower than the average, though the ΔC_m for the A91H variant was very close to the average ΔC_m if error is taken into account. The A83S/A87S/K84H variant has the least amount of change at roughly 0.08 gdnHCl molar units lower.

The overall free energy of unfolding follows the same broad trend of lower $\Delta G_u^\circ(H_2O)$ as loop size decreases (Figure 7). Most variants show a decrease in the free energy of unfolding, $\Delta G_u^\circ(H_2O)$, as compared to the pWT background. The amount decreased varied from about 0.5 to 2 kcal/mol. The variants D3H, K20H, A66H, and K84H showed no change, within error, from the pWT counterparts.

The denaturation midpoints indicate that every variant is fully unfolded at 3 M gdnHCl (see also Figure 2), the conditions we use in the next section to monitor denatured state loop formation.

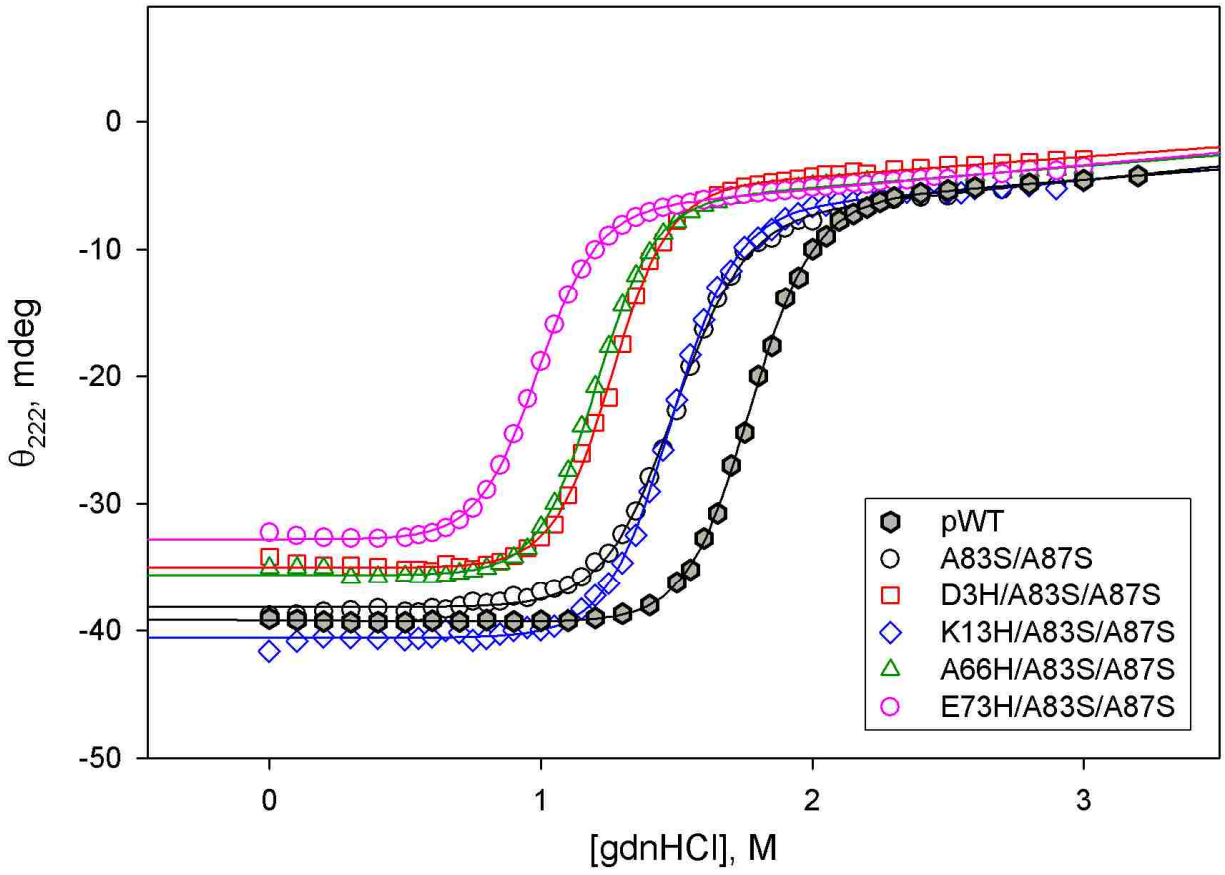


Figure 4: Representative curves of gdnHCl denaturation monitored by circular dichroism at 222 nm. Data were acquired at pH 6.5 and 25 °C.

Table 1: Thermodynamic properties for gdnHCl unfolding of cytochrome *c'* variants at pH 6.5 and 25 °C.

Variant	Loop Size	ΔG_u° (H_2O) (kcal/mol)	*m*-value (kcal/(mol*M))	C_m (M)
A83S/A87S/A104H	10	5.25 ± 0.18	5.81 ± 0.11	0.90 ± 0.01
A104H [a]	10	6.08 ± 0.26	5.59 ± 0.12	1.09 ± 0.05
A83S/A87S/K97H	17	5.88 ± 0.16	5.52 ± 0.25	1.07 ± 0.06
K97H [a]	17	6.80 ± 0.09	5.28 ± 0.07	1.29 ± 0.01
A83S/A87S/A91H	23	5.97 ± 0.04	5.59 ± 0.05	1.07 ± 0.01
A91H [a]	23	7.07 ±0.09	5.75 ± 0.20	1.23 ± 0.03

A83S/K84H/A83S	30	5.78 ± 0.24	5.80 ± 0.30	1.00 ± 0.01
K84H [a]	30	5.89 ± 0.34	5.32 ± 0.32	1.11 ± 0.01
E73H/A83S/A87S	41	5.27 ± 0.09	5.39 ± 0.07	0.98 ± 0.01
E73H [a]	41	5.76 ± 0.13	5.08 ± 0.14	1.13 ± 0.05
A66H/A83S/A87S	48	6.82 ± 0.11	5.46 ± 0.06	1.19 ± 0.01
A66H [a]	48	6.79 ± 0.09	4.88 ± 0.06	1.39 ± 0.04
D58H/A83S/A87S	56	5.91 ± 0.35	4.97 ± 0.33	1.18 ± 0.01
D58H [a]	56	6.72 ± 0.09	4.80 ± 0.11	1.40 ± 0.02
K49H/A83S/A87S	65	6.15 ± 0.08	5.05 ± 0.04	1.22 ± 0.01
K49H [a]	65	6.75 ±0.10	4.66 ± 0.07	1.45 ± 0.01
K39H/A83S/A87S	75	6.32 ± 0.09	5.18 ± 0.09	1.22 ± 0.02
K39H [a]	75	6.91 ± 0.15	4.70 ± 0.10	1.47 ± 0.01
K31H/A83S/A87S	83	6.60 ± 0.14	5.27 ± 0.08	1.25 ± 0.02
K31H [a]	83	7.41 ± 0.29	5.04 ± 0.20	1.47 ± 0.03
K20H/A83S/A87S	94	5.96 ± 0.34	4.91 ± 0.29	1.22 ± 0.01
K20H [a]	94	5.89 ± 0.13	4.17 ± 0.09	1.41 ± 0.01
K13H/A83S/A87S	101	7.41 ± 0.42	5.01 ± 0.29	1.472 ± 0.003
K13H [a]	101	8.17 ± 0.11	4.91 ± 0.07	1.66 ± 0.05
D3H/A83S/A87S	111	7.12 ± 0.30	5.69 ± 0.25	1.253 ± 0.003
D3H [a]	111	7.41 ± 0.17	5.08 ± 0.15	1.46 ± 0.05
A83S/A87S		6.78 ± 0.19	4.56 ± 0.20	1.49 ± 0.02
pWT [a]		8.81 ± 0.22	5.14 ± 0.13	1.71 ± 0.04

[a] Wild type data is from Rao et al.[14]

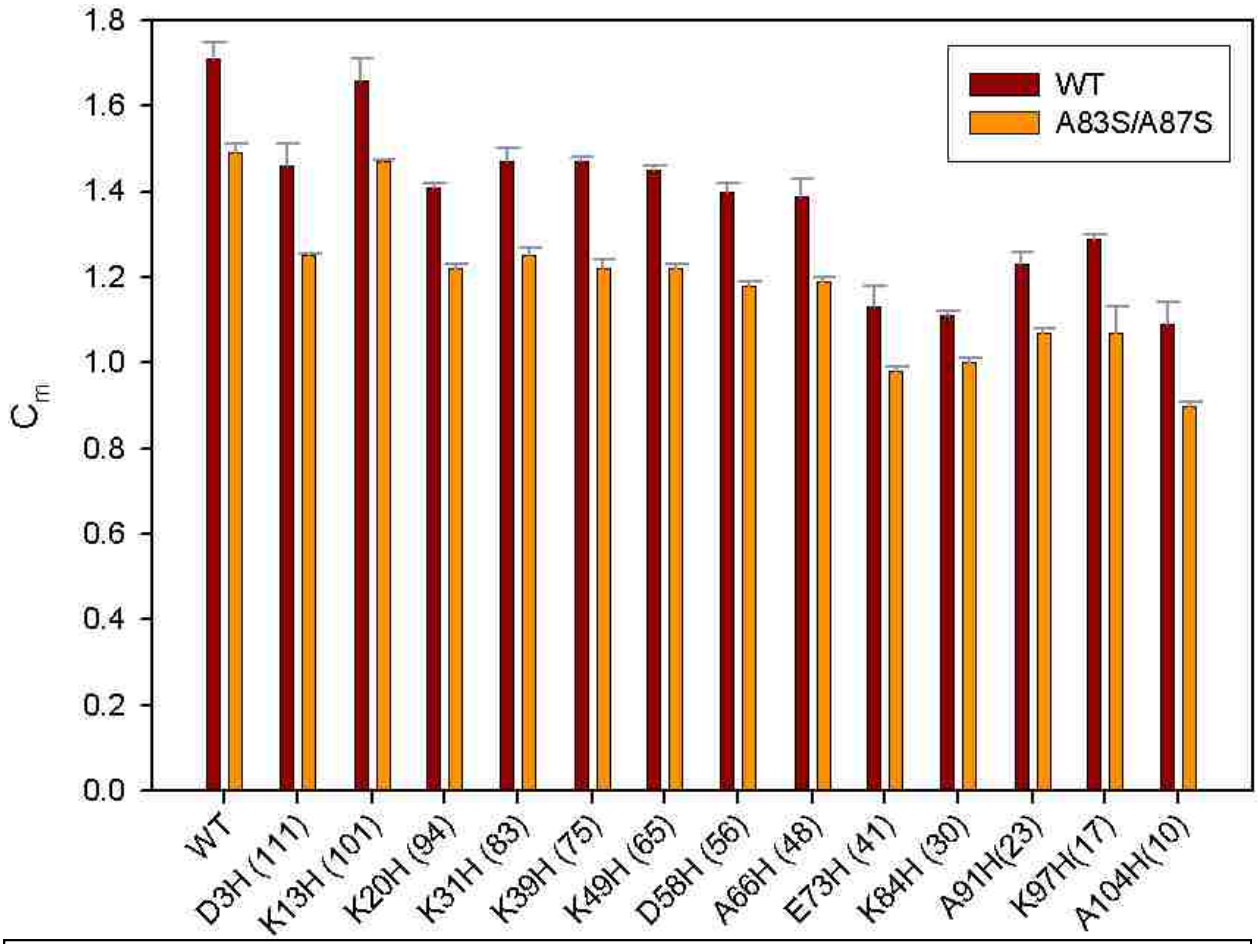


Figure 5: Midpoint of the unfolding transition (C_m) for Cyt *c'* variants. Histidine – heme loop sizes are found in parentheses. Variants with the A83S/A87S substitutions showed lowered midpoints. The shortest loops showed the smallest midpoints.

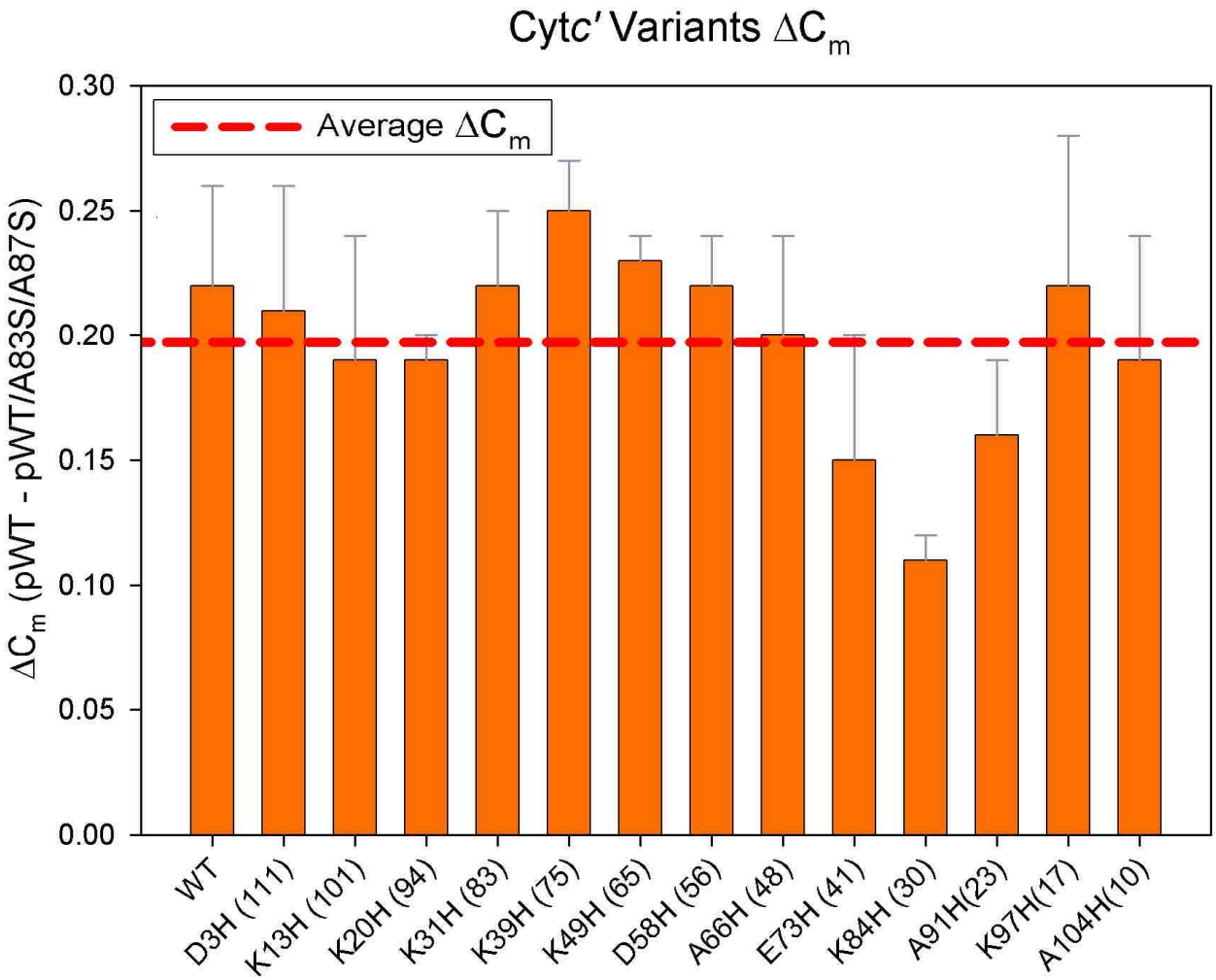


Figure 6: Changes in midpoint of the unfolding transition (ΔC_m) for Cytc′ variants. Histidine – heme loop sizes are found in parenthesis. On average, the ΔC_m of pWT/A83S/A87S is 0.197 ± 0.037 units lower (gdnHCl, M) than pWT.

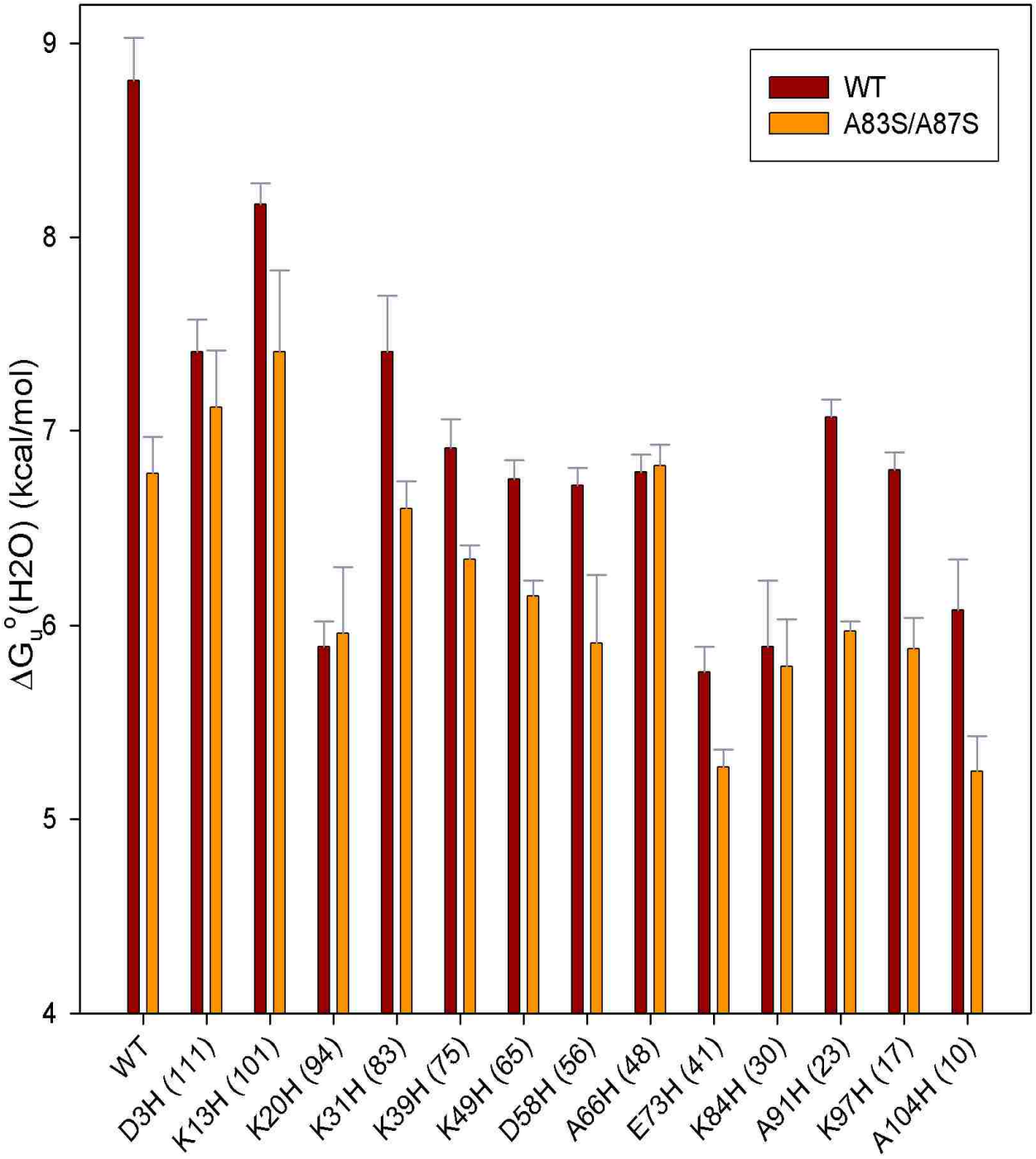


Figure 7: Free energy of unfolding ($\Delta G°_u(H_2O)$) of pWT[5] and A83S/A87S variants; histidine – heme loops sizes are in parentheses. Variants with A83S/A87S show a trend of lowered stability.

Denatured State Loop Formation Equilibria

Cyt*c′* has a covalently bound heme though a CXXCH heme attachement motif, and provides a useful approach to monitor loop formation in the denatured state. Since each variant can form a single His-heme loop, the ionizable side chain of histidine allows for loop stability measurements of each variant by simple pH titration. The engineered histidines provide a variety of loop lengths, and an observable pK_a, pK_a(obs), of loop formation yields information on relative stabilities of different sized loops. A pH titration allows detection of the His-heme loop breakage as the low spin Fe^{3+}-heme, bound to the strong field histidine side chain ligand, is replaced by the weak field water ligand as pH is lowered. This spin state change is detected by a shift in the Soret band near 400 nm (Figure 8). The Soret band shifts from 408 nm (the His-heme bound state) to 398 nm (loop broken state) as pH is lowered, generally from about pH 8 to about pH 3 depending on the variant. In Figure 6, the isosbestic point is initially at 404 nm.

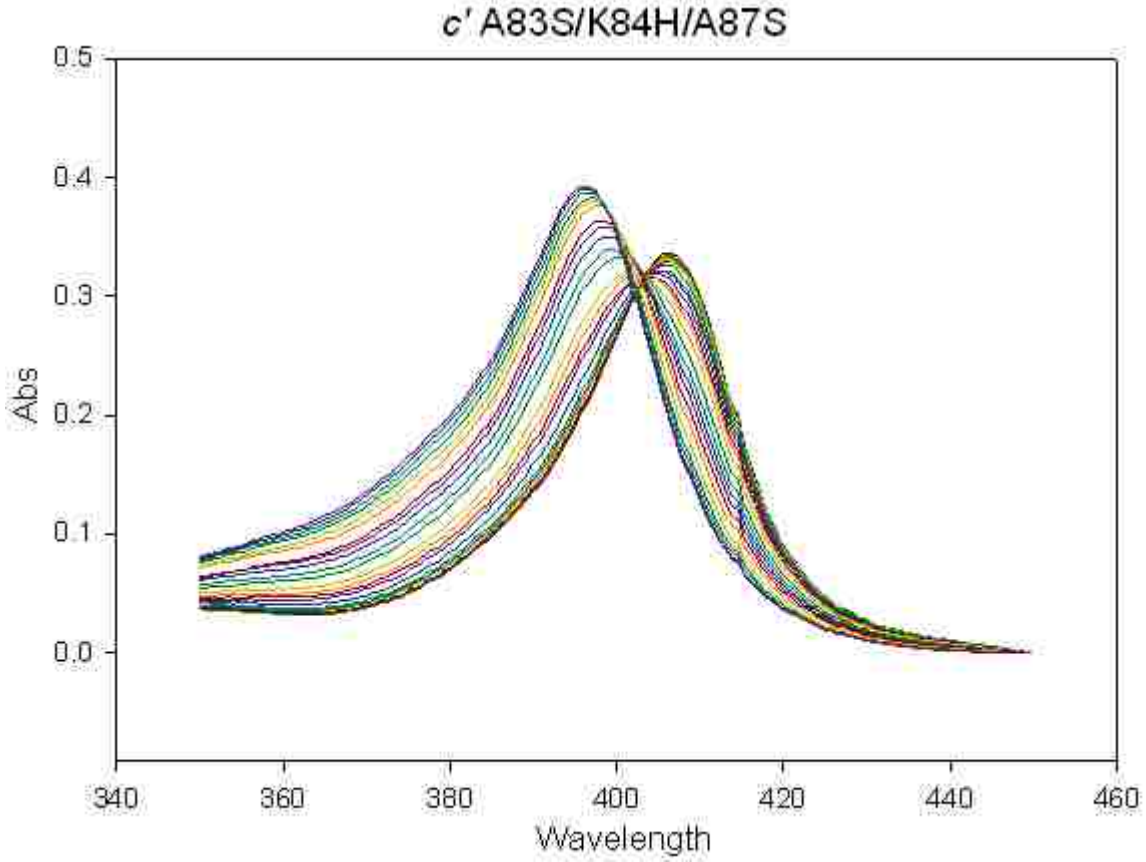


Figure 8: Representative absorbance spectra of Cyt*c′* from 350 nm to 450 nm over pH range pH 7.4 to 3.2 in increments of roughly 0.2 pH units in 3 M gdnHCl at 22 ± 1 °C.

Below pH 3.2, the isosbestic point shifts to 399 nm. The shift in isosbestic point likely results from protonation of His117 as the His117-heme bond breaks (see Figure 1). All of the Cyt*c*′ variants are fully unfolded in 3.0 M gdnHCl as evident in Table 1 and Figure 4. Also, because pH titrations are carried out in 3.0 M gdnHCl, electrostatic effects on denatured state loop formation should be minimized.

Denatured state loop formation equilibrium can be evaluated at 398 nm (Fig. 9), subtracting 450 nm as a baseline. Compared to pWT, A83S/A87S and the histidine variants show curves shifted towards a lower pK_a. As with global stability, the variants with the longest loops trend toward being the most similar compared to A83S/A87S and there is a distinct trend of decreasing loop stability with longer loop sizes. Data can be fit to the Henderson-Hasselbalch equation (Eq. 4), which produces pK_a(obs).

Eq. 4:

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

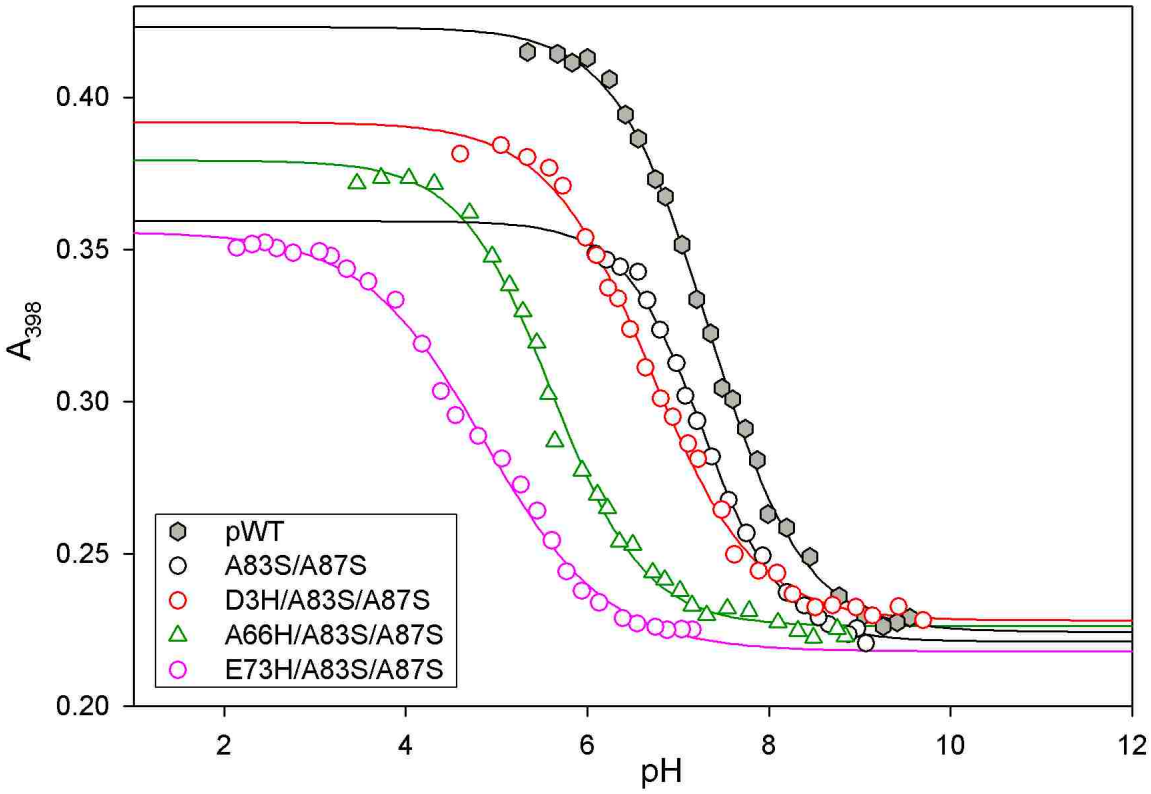


Figure 9: Representative titration curves of histidine-heme ligation as a function of pH at 22 ± 1 °C in 3.0 M gdnHCl.

More stable loops will have a lower pK_a(obs), as this indicates that a higher concentration of H^+ is required to break the His-heme loop. A two-step process describes the loop formation equilibrium as found in Equation 5; the deprotonated histidine is bound to the Fe^{3+} of the heme after ionization.

Eq. 5: heme-Fe^{3+}//$HisH^+$ → heme-Fe^{3+}//His + H^+ → heme-Fe^{3+}-His

In consideration of this two-step process, pK_a(obs) can be written as the summation of the intrinsic histidine ionization pK_a, $pK_a(HisH^+)$,and the pK for His-heme loop formation, pK_{loop}(His) with a fully deprotonated histidine (Eq. 6).

Eq. 6: $pK_a(obs) = pK_a(HisH^+) + pK_{loop}(His)$

Dar et al.[15] report an average $pK_a(HisH^+)$ of 6.78 ± 0.15 for all Cyt*c′* variants obtained in 3-6 M gdnHCl, which allows calculation of the pK_{loop}(His) value for each variant that displayed monophasic behavior in the pH titration data (Table 2).

Table 2: Cyt*c′* thermodynamic parameters in 3 M gdnHCl at 22 ± 1 °C for monophasic loop formation.

Variant	Loop Size	pK_a (obs)	n_p	pK_{loop}(His)
A83S/A87S/A104H	10	4.22 ± 0.03	1.13 ± 0.04	-2.56 ± 0.03
A104H [a]	10	4.12 ± 0.07	1.23 ± 0.21	-2.66 ± 0.07
A83S/A87S/K97H	17	4.40 ± 0.02	1.10 ± 0.02	-2.38 ± 0.02
K97H [a]	17	4.42 ± 0.07	0.97 ± 0.06	-2.36 ± 0.07
A83S/A87S/A91H	23	4.99 ± 0.04	1.02 ± 0.02	-1.79 ± 0.04
A91H [a]	23	5.08 ± 0.05	1.07 ± 0.03	-1.70 ± 0.05
A83S/K84H/A83S	30	4.94 ± 0.02	1.05 ± 0.02	-1.84 ± 0.02
K84H [a]	30	4.93 ± 0.04	0.97 ± 0.10	-1.85 ± 0.04
E73H/A83S/A87S	41	4.86 ± 0.02	0.68 ± 0.03	-1.92 ± 0.02

E73H [a]	41	5.07 ± 0.04	1.02 ± 0.10	-1.71 ± 0.04
A66H/A83S/A87S	48	5.66 ± 0.04	0.95 ± 0.12	-1.12 ± 0.04
A66H [a]	48	5.92 ± 0.09	1.04 ± 0.04	-0.86 ± 0.09
D58H/A83S/A87S	56	5.60 ± 0.03	1.00 ± 0.15	-1.18 ± 0.03
D58H [a]	56	5.68 ± 0.04	1.05 ± 0.03	-1.10 ± 0.04
K49H/A83S/A87S	65	5.99 ± 0.04	0.86 ± 0.04	-0.79 ± 0.04
K49H [a]	65	6.19 ± 0.02	1.01 ± 0.04	-0.59 ± 0.02
K31H/A83S/A87S	83	6.00 ± 0.03	1.02 ± 0.04	-0.78 ± 0.03
K31H [a]	83	6.18 ± 0.07	0.91 ± 0.08	-0.60 ± 0.07
A83S/A87S		7.21 ± 0.11	0.94 ± 0.15	-
pWT [a]		7.30 ± 0.05	0.97 ± 0.16	-

All given parameters are the average and standard deviation of at least three trials. The number of protons involved in the loop equilibrium is given by n_p. The values for pK_{loop}(His) are calculated by equation 6 using $pK_a(HisH^+)$ = 6.78, the average value for all variants.[15]

[a] Wild type data is from Rao et al.[14] and is converted to pK_{loop}(His) using $pK_a(HisH^+)$ = 6.78 from the more recent work of Dar et al.[15]

Loops displaying monophasic behavior appear to yield similar stabilities compared to their pWT counterparts, but upon closer examination show a trend of increased stability once a certain loop length is achieved. Figure 10 shows a comparison of monophasic pWT to A83S/A87S variants. A83S/A87S variants with loop lengths larger than 30 all show lower pK_a(obs), indicating increased loop stability. Figure 11 shows ΔpK_a(obs) for these variants. Two of these variants, A83S/K84H/A87S and A83S/A87S/K97H, show little change from their respective pWT counterparts, while A83S/A87S/A91H shows a ΔpK_a(obs) very close to the average ΔpK_a(obs). The variant A83S/A87S/A104H, the shortest loop in this study, has a negative ΔpK_a(obs), indicating a decrease in stability compared to pWT. Beyond loop size of 30, all variants are more stable than pWT counterparts, with three variants achieving slightly more than twice the average ΔpK_a(obs) of 0.102.

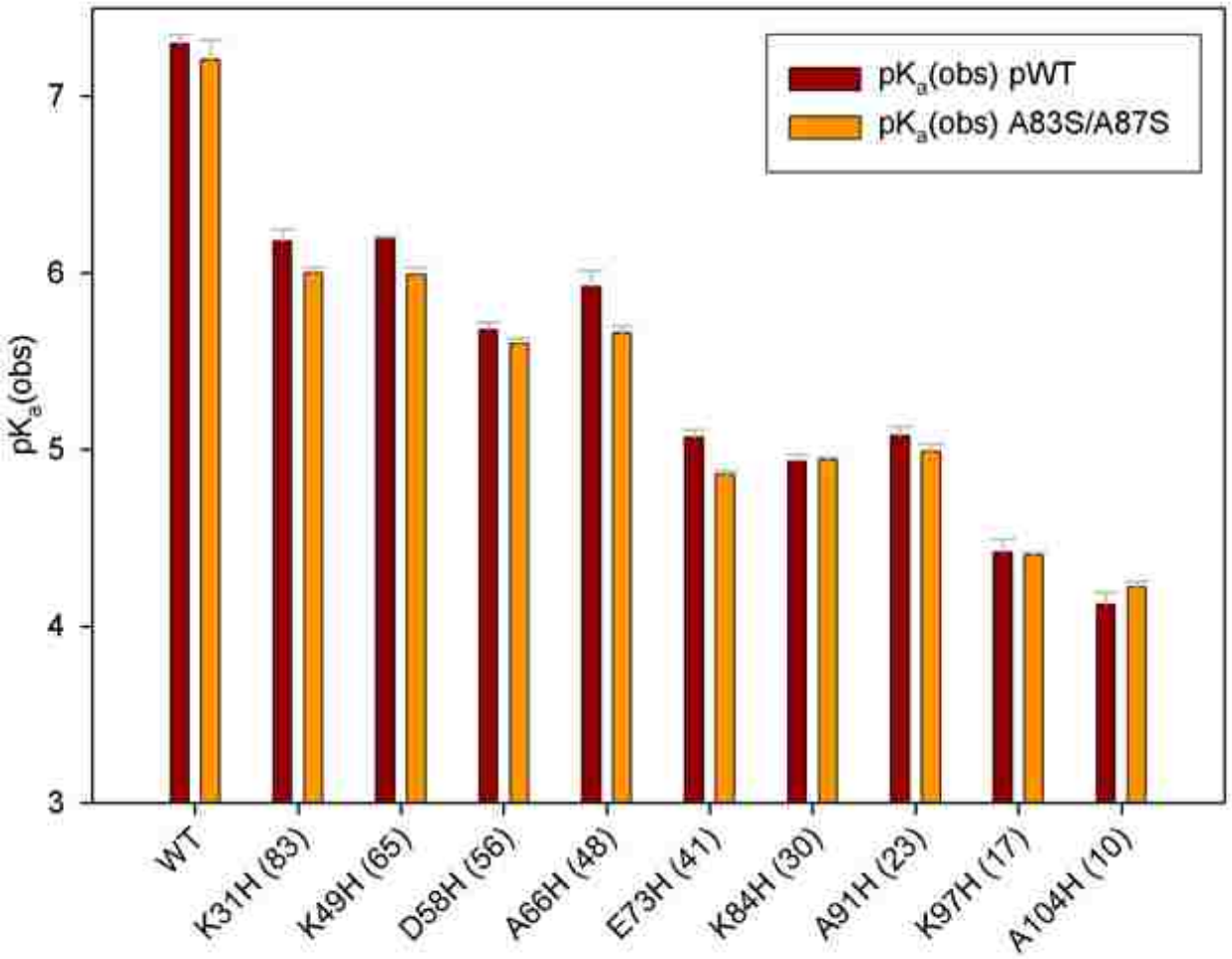


Figure 10: Observed pK$_a$ of loop equilibria at 22 ± 1 °C in 3 M gdnHCl. His-heme loop sizes are found in parentheses.

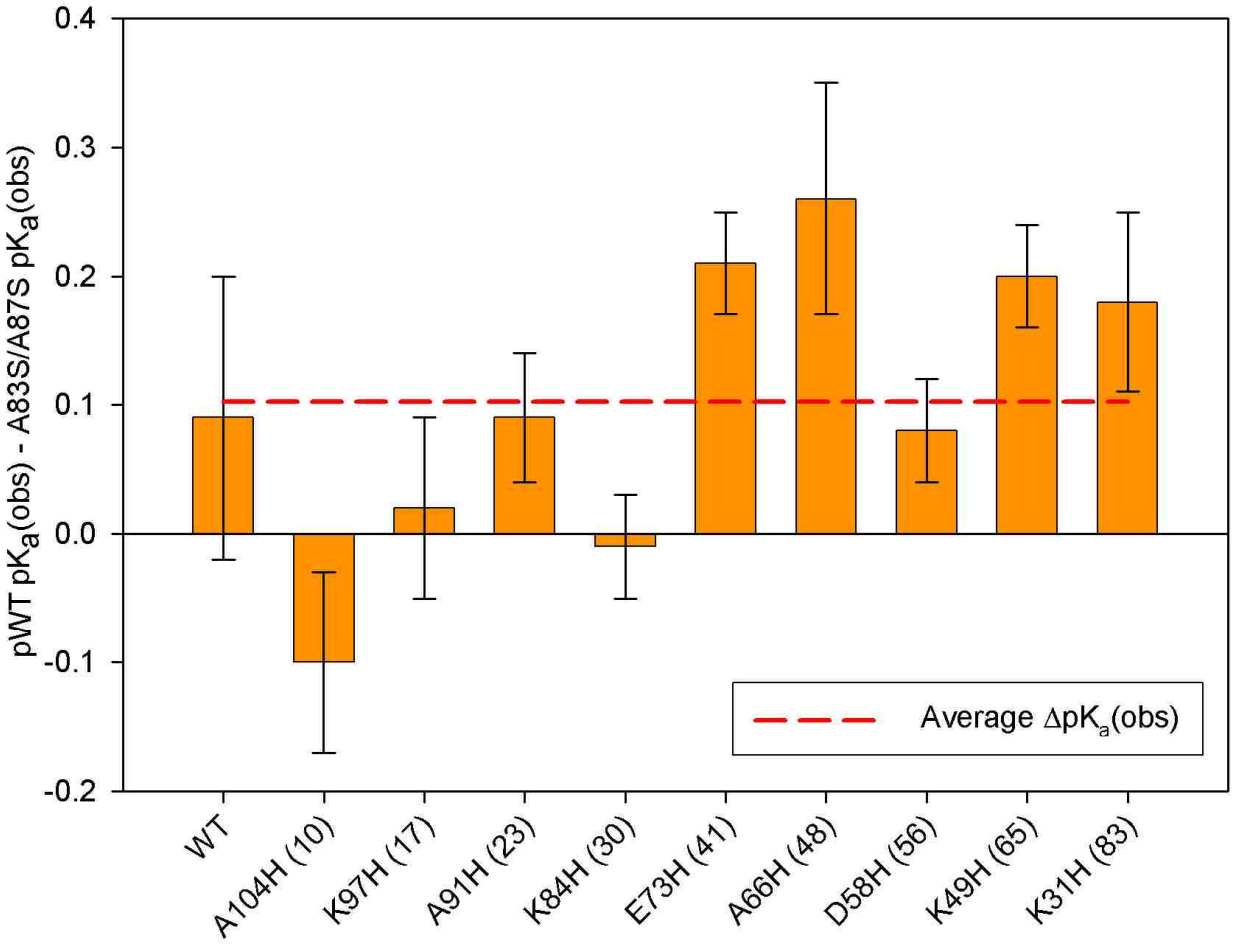


Figure 11: Changes in observed pK_a for variants of Cyt*c*′ variants which behaved in a monophasic manner. Loop sizes are found in parenthesis. The average $\Delta pK_a(obs) = 0.102$.

For variants with the longest loop sizes, biphasic loop formation was observed. For these variants, loop formation curves broadened and data did not fit to the Henderson-Hasselbalch equation well. For these variants, histidine-heme binding is insufficient to complete the spin state transition and so another ionizable ligand, lysine, binds to the heme and finishes the spin state transition. Data from these variants were fit to equation 7 and parameters from the fit are listed in Table 3. To fit data to this equation, the intrinsic pK_a of lysine, $pK_a(LysH^+)$ was set to 10.5. A_{HS} and A_{LS} correspond to absorbance at 398 nm for the high spin (loop broken) and low spin states (loop formed), and $pK_{loop}(Lys)$ is the stability of the Lys-heme loop.

Eq. 7:
$$A_{398} = \frac{A_{HS} + A_{LS}\left(\left(\frac{10^{-pK_{loop}(His)}}{1+10^{pK_a(HisH^+)-pH}}\right)+\left(\frac{10^{-pK_{loop}(Lys)}}{1+10^{pK_a(LysH^+)-pH}}\right)\right)}{1+\left(\frac{10^{-pK_{loop}(His)}}{1+10^{pK_a(HisH^+)-pH}}\right)+\left(\frac{10^{-pK_{loop}(Lys)}}{1+10^{pK_a(LysH^+)-pH}}\right)}$$

Table 3: Cytc′ thermodynamic parameters in 3.0 M gdnHCl at 22 ± 1 °C for biphasic loop formation. Wild type data is from Rao et al.[14]

Variant	Loop Size	pK_{loop}(His)[a]	pK_a(His$^+$)[a]	pK_{loop}(Lys)[a]
A83S/A87S/K39H	75	-0.93 ± 0.10	6.74 ± 0.09	-3.90 ± 0.36
K39H[b]	75	-0.81 ± 0.05	6.83 ± 0.09	-2.57 ± 0.14
K20H/A83S/A87S	94	-0.89 ± 0.02	6.78 ± 0.05	-3.42 ± 0.18
K20H[b]	94	-0.77 ± 0.04	6.81 ± 0.05	-3.25 ± 0.17
K13H/A83S/A87S	101	-0.32 ± 0.10	6.78 ± 0.14	-3.16 ± 0.01
K13H[b]	101	-0.22 ± 0.06	6.72 ± 0.09	-3.12 ± 0.07
D3H/A83S/K84H	111	-0.10 ± 0.12	6.67 ± 0.14	-3.27 ± 0.20
D3H[b]	111	0.53 ± 0.15	6.4 ± 0.3	-3.47 ± 0.06

[a] All given parameters are the average and standard deviation of at least three trials. The intrinsic pK_a for lysine was assumed to be 10.5.

[b] Data are from Rao et al.[14]

As seen in Table 3, the measure of His-heme loop stability, pK_{loop}(His), of each A83S/A87S variant is more negative than the pWT counterpart, though there is much overlap when error is taken into account. The most pronounced difference in pK_{loop}(His) is found between pWT/D3H and D3H/A83S/A87S, which is the longest loop length at 111 residues. As the loop size increases, pK_{loop}(His) becomes less negative, indicating that loop formation is less favorable as the loops become longer.

To be able to plot the monophasic and biphasic loop stability data on the same scale, Equation 2 allows calculation of the pK_{loop}(His) from the monophasic data (Table 2). The pK_{loop}(His), loop

stability, versus the logarithm of the loop size, Log(N), is found in Figure 12. This plot reveals that A83S/A87S variants more N-terminal to the disrupting A83S/A87S mutations (longer loop lengths) all have pK_{loop}(His) values more negative compared to pWT variants, indicating higher stability. For variants more C-terminal to the A83S/A87S mutations, A83S/A87S/A104H is less negative than its counterpart, while the K84H, K97H and A91H show little to no change, indicating lower or no change in stability as compared to pWT.

To compare the scatter about the best fit line, data from homopolymeric polyalanine inserts acquired at 3 M gdnHCl[31] are plotted alongside the pWT and A83S/A87S variants. While the homopolymeric polyalanine data have a tight correlation about the best fit line, the pWT and A83S/A87S variants show sizeable scatter about their respective best fit lines.

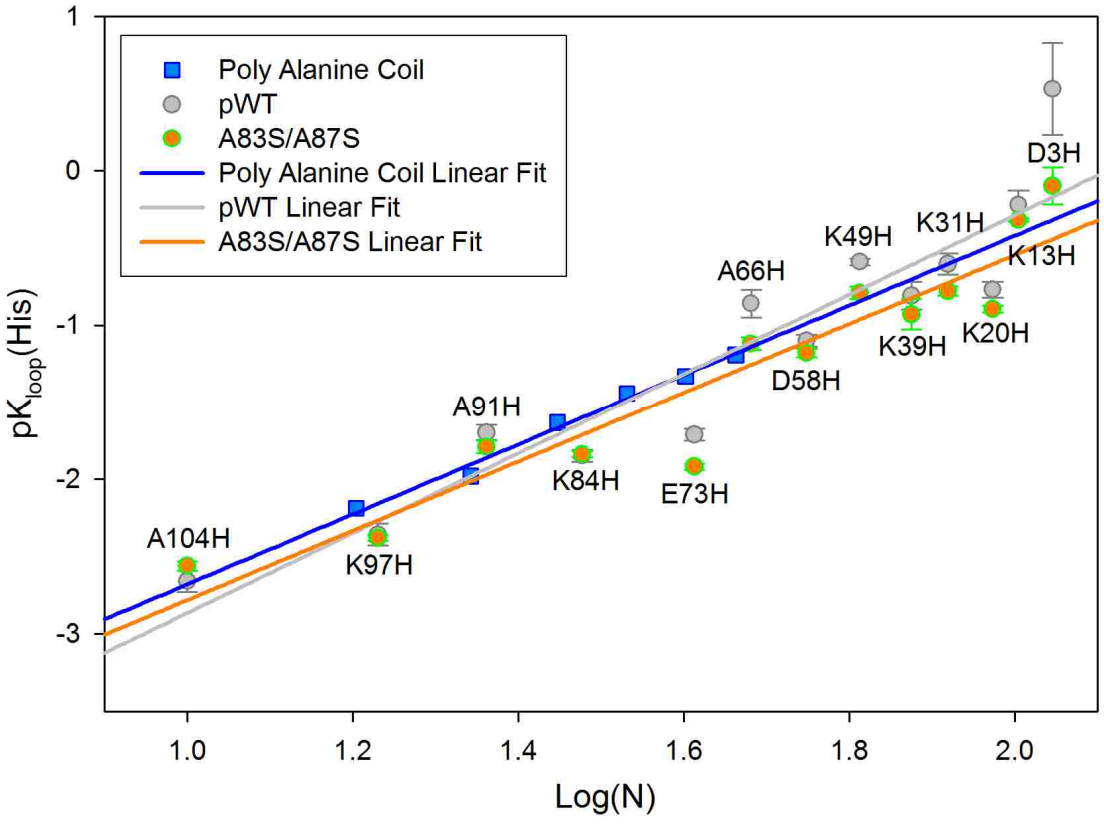


Figure 12: Plot of loop stability, pK_{loop}(His) vs. the logarithm of loop size (N).

Kinetics of Denatured State Loop Breakage and Formation

To gain further insight into the effects of loop length and residual structure in 3.0 M gdnHCl, the kinetics of histidine-heme loop formation and breakage of each variant were measured by downward pH-jump stopped-flow experiments in the denatured state. The Soret absorption band of the heme was employed to monitor the weak field water ligand replacing the strong field histidine side chain ligand. As demonstrated in previous studies in the Bowler lab, the model of rapid histidine deprotonation followed by binding of histidine to the heme is consistant with the kinetics of loop breakage and formation.[32] In this model, the observed rate constant has pH dependence as found in Equation 8, where k_b and k_f are the rate constants for loop breakage and formation, and $K_a(\mathrm{HisH^+})$ is the dissociation constant for histidine deprotonation. In this scheme, if pH << $pK_a(\mathrm{HisH^+})$, then k_{obs} is approximately equal to k_b.

Eq. 8:
$$k_{\mathrm{obs}} = k_{\mathrm{b}} + k_{\mathrm{f}}\left(\frac{K_{\mathrm{a}}\left(\mathrm{HisH^+}\right)}{K_{\mathrm{a}}\left(\mathrm{HisH^+}\right)+\left[\mathrm{H^+}\right]}\right)$$

Table 4 lists the k_b of A83S/A87S variants in two ways. The first way is as calculated from the full raw data from the stopped-flow experiments, while the second discards the first 4 ms of the data so as to be directly comparable with data from Dar et al.[15] The A83S/A87S k_b rates, calculated from raw data or using a 4 ms cutoff, are generally in good agreement with each other, and therefore the 4 ms cutoff rates will be used, allowing direct comparison with the data in Dar et al. [15] The k_b values of the A83S/A87S variants follow closely to the k_b values for the pWT variants, with the largest differences at the shortest and longest loop lengths. The data in Supplemental Table 1 (Chapter 1 Appendix) show k_b values of the A83S/A87S at pH 3.6 to be very similar to data found in Table 4, demonstrating that $k_b(\mathrm{obs}) \approx k_b$ at low pH as given in Eq. 8.

The k_b data of pWT A83S/A87S variants are plotted in Figure 13 as a function of loop size. The data fluctuates with loop size and has considerable range. Homopolymeric polyalanine inserts acquired in 3 M gdnHCl[31], also found in Figure 13, do not show this type of fluctuation or range, highlighting the nonrandom behavior of pWT and A83S/A87S variants. The polyalanine data

levels out as loop size increases as expected for a random coil, but pWT and A83S/A87S do not become uniform as loop size increases. A comparison of k_b between the pWT and A83S/A87S variants reveals very similar values for most variants, with the exception of the long loop of D3H. D3H/A38S/A87S has nearly doubled the pWT value.

Table 4: Kinetic parameters for loop breakage for Cytc' variants in 3 M gdnHCl and at 25 °C. WT data from Dar et al.[15]

Variant	Loop Size (n)	pWT/A83S/A87S k_b, s^{-1}(raw) pH 3.08 ± 0.05	pWT/A83S/A87S k_b, s^{-1} (4 ms cutoff) pH 3.08 ± 0.05	pWT k_b, s^{-1} (4 ms cutoff) pH 3.07 ± 0.02	4 ms cutoff difference
A104H	10	215.6 ± 4.4	199.0 ± 4.6	210 ± 4	11
K97H	17	77.5 ± 2.0	75.83 ± 2.3	86 ± 1	10.2
A91H	23	96.3 ± 1.8	94.42 ± 3.4	101 ± 1	6.6
K84H	30	65.7 ± 1.4	64.23 ± 1.9	62.4 ± 0.3	-1.8
E73H	41	35.0 ± 0.4	34.6 ± 0.4	32.7 ± 0.1	-1.9
A66H	48	141.3 ± 2.5	138.9 ± 3.6	137 ± 2	-1.9
D58H	56	72.5 ± 0.8	71.0 ± 0.9	66.5 ± 0.6	-4.5
K49H	65	171.7 ± 4.6	152.4 ± 4.3	156 ± 3	3.6
K39H	75	82.6 ± 1.0	80.6 ± 1.0	74 ± 1	-6.6
K31H	83	76.5 ± 0.7	74.6 ± 0.9	73 ± 1	-1.6
K20H	94	55.8 ± 0.6	53.5 ± 0.7	49 ± 1	-4.5
K13H	101	132.5 ± 3.7	120.1 ± 3.4	126.92 ± 0.01	6.9
D3H	111	121.2 ± 3.7	98.44 ± 4.7	59 ± 3	-39.4

His-heme loop formation rate constant, k_f, is calculated using pK_{loop}(His) and k_b ($k_f = k_b K_{loop}$(His)) and presented in Table 5. Figure 14 shows log of k_f as a function of loop size. The A83S/A87S variants followed fairly closely to pWT, but every A83S/A87S variant N-term to K84H shows an increase in k_f, though a few are within error of their respective counterparts. A83S/A87S/A104H is slightly slower to form, while A97H, A91H, and K84H are all very close to pWT. The longest loop at 111 is the only outlier as D3H/A83S/A87S is considerably faster to form than pWT/D3H, though pK_{loop}(His) is more difficult to measure at this loop length.

Table 5: Calculated kinetic parameters for loop formation for Cytc' variants in 3 M gdnHCl and at 25 °C. WT data from Dar, et al.[15]

Variant	Loop Size (n)	pWT/A83S/A87S k_f, s^{-1} pH 3.08 ± 0.05	pWT k_f, s^{-1} pH 3.07 ± 0.02	pWT – pWT/A83S/A87S
A104H	10	72269 ± 1286	88000 ± 14000	15731
K97H	17	18190 ± 325	19800 ± 3200	1610
A91H	23	5873 ± 266	5100 ± 600	-773
K84H	30	4443 ± 129	4500 ± 400	57
E73H	41	2876 ± 31	1700 ± 150	-1176
A66H	48	1830 ± 155	1000 ± 200	-830
D58H	56	1074 ± 31	850 ± 80	-224
K49H	65	946 ± 132	610 ± 30	-336
K39H	75	684 ± 32	480 ± 60	-204
K31H	83	446 ± 19	290 ± 50	-156
K20H	94	420 ± 15	290 ± 30	-130
K13H	101	250 ± 54	211 ± 3	-39
D3H	111	123 ± 52	18 ± 6	-105

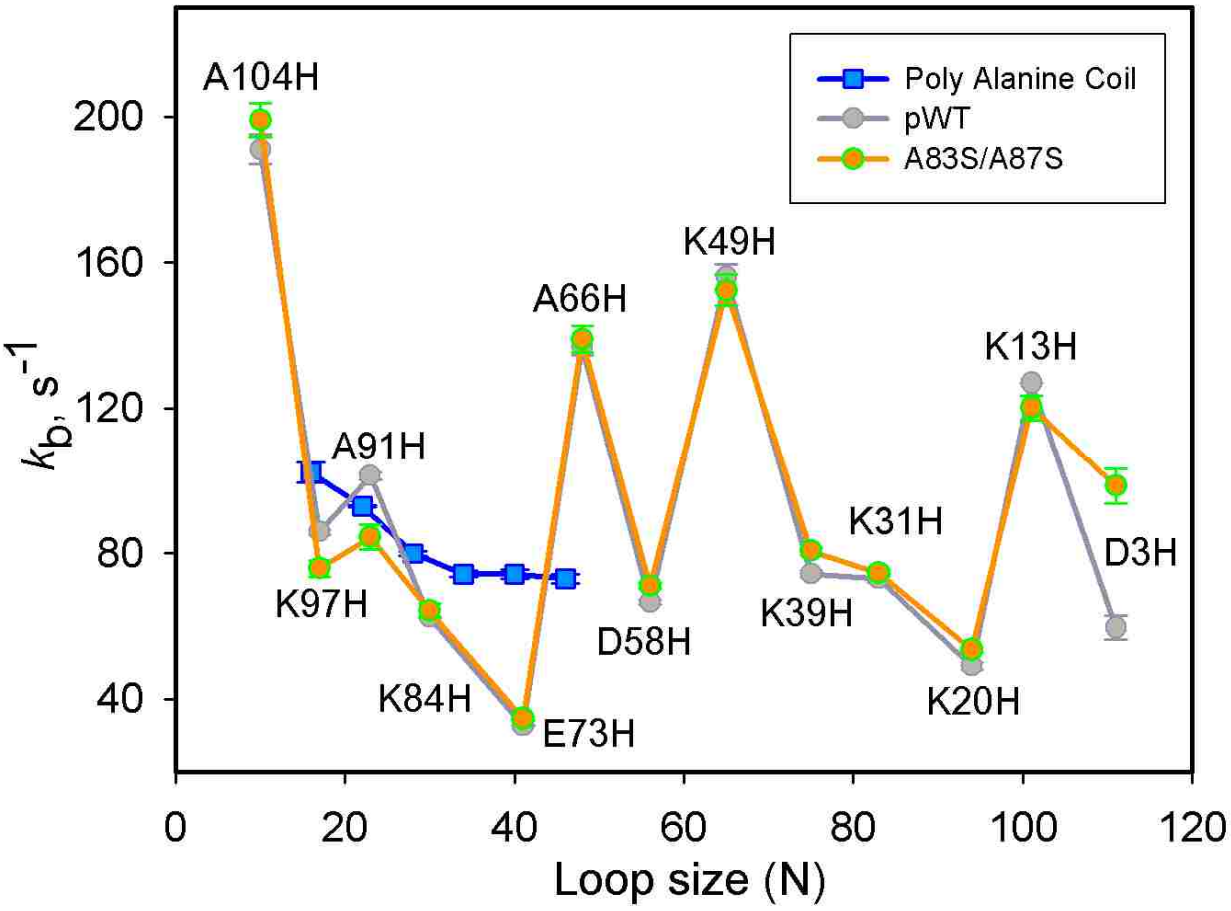


Figure 13: His-heme loop breakage rate constant, k_b, as a function of loop size, N, at pH 3.0, 22 ± 1 °C, in 3 M gdnHCl.

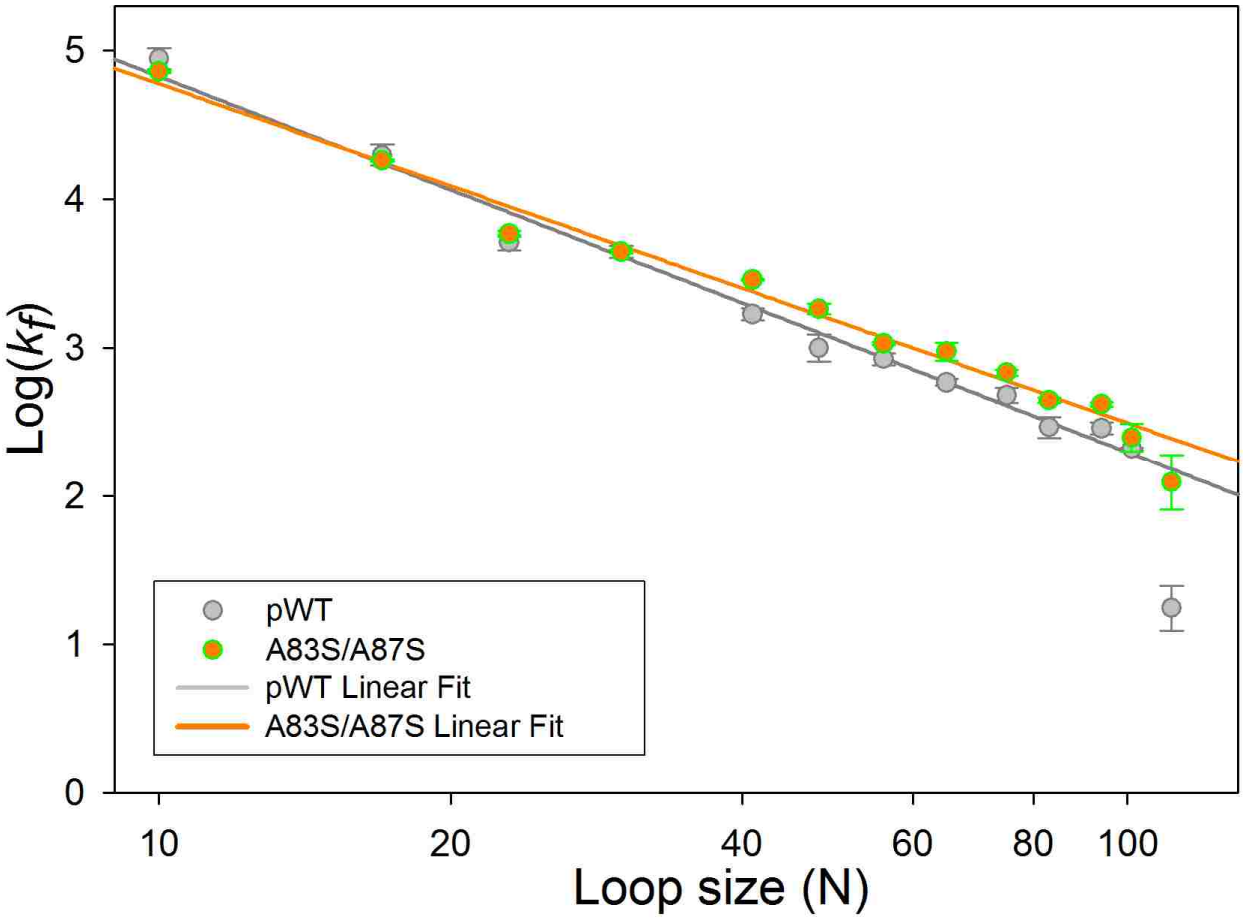


Figure 14: Log of His-heme loop formation rate constant, k_f, as a function of loop size, N, at pH 3.0, 22 ± 1 °C, in 3 M gdnHCl. The linear fits do not include the outlying D3H variant for either WT or A83S/A87S.

Molecular Dynamics Simulations

Molecular dynamics simulations were run on WT as well as WT/A83S/A87S Cyt*c* using the *il*mm system to simulate thermal unfolding of the protein. The Define Secondary Structure of Proteins (DSSP)[33] algorithm was used to characterize the results of the molecular dynamic simulations. DSSP employs a standardization of secondary structure assignment based on the coordinates of a protein. The 51 ns WT simulation at 298K DSSP results are found in Figure 15. The four native helical regions (residues 4-27, 34-54, 75-95, and 99-120) of the pWT protein are clearly visible, and rarely fluctuate from helical structure throughout the simulation. Figure 16 shows the WT/A83S/A87S structure overwhelmingly following the same pattern. There is a

slight opening in helix 2(residues 48-51) during the final 15 ns of the simulation, however these residues remained in flux between loop and helical structures and show very little difference when compared to pWT.

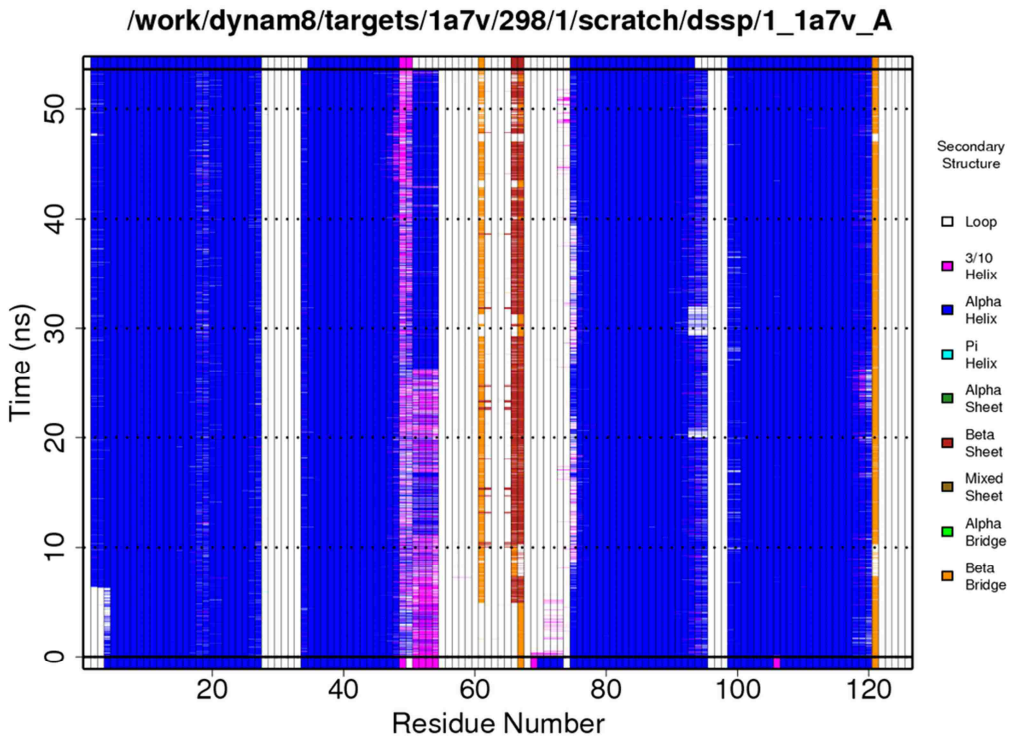


Figure 15: DSSP of WT Cyt*c'* simulation at 298 K up to 51 ns. Native helices are found in residues 4-27, 34-54, 75-95, and 99-120.

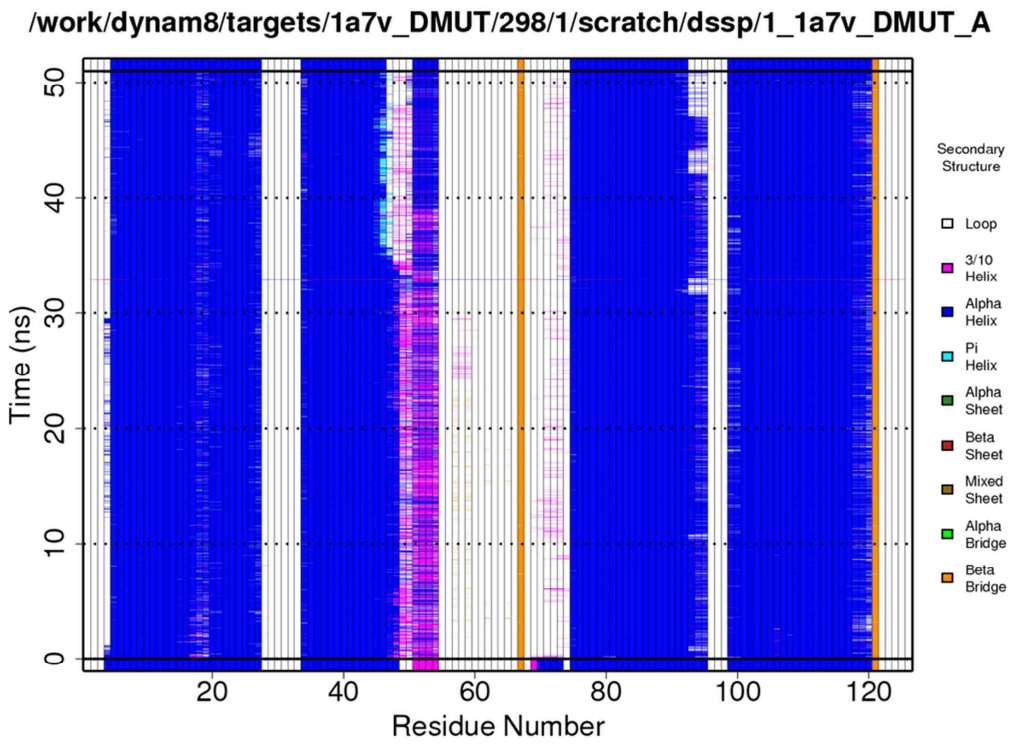


Figure 16: DSSP of WT/A83S/A87S Cytc′ simulation at 298 K up to 51 ns. Native WT helices are found in residues 4-27, 34-54, 75-95, and 99-120.

DSSP analyses of two simulations depicting the unfolding of WT at 498 K are found in Figure 17. The first simulation (top of the figure) shows that helix 1, 3, and 4 all retain some helical structure, while loop 2 melts into mostly loop structure with some flux back to helix through the 30 – 40 ns timepoints. The second simulation shows helix 1 and 2 melting into loop structure by 30 and 20 ns, respectively, with little fluctuation back to helical character. Helix 3 retains much of its helical character through the simulation, and the forth helix also retains more helical character than the first simulation.

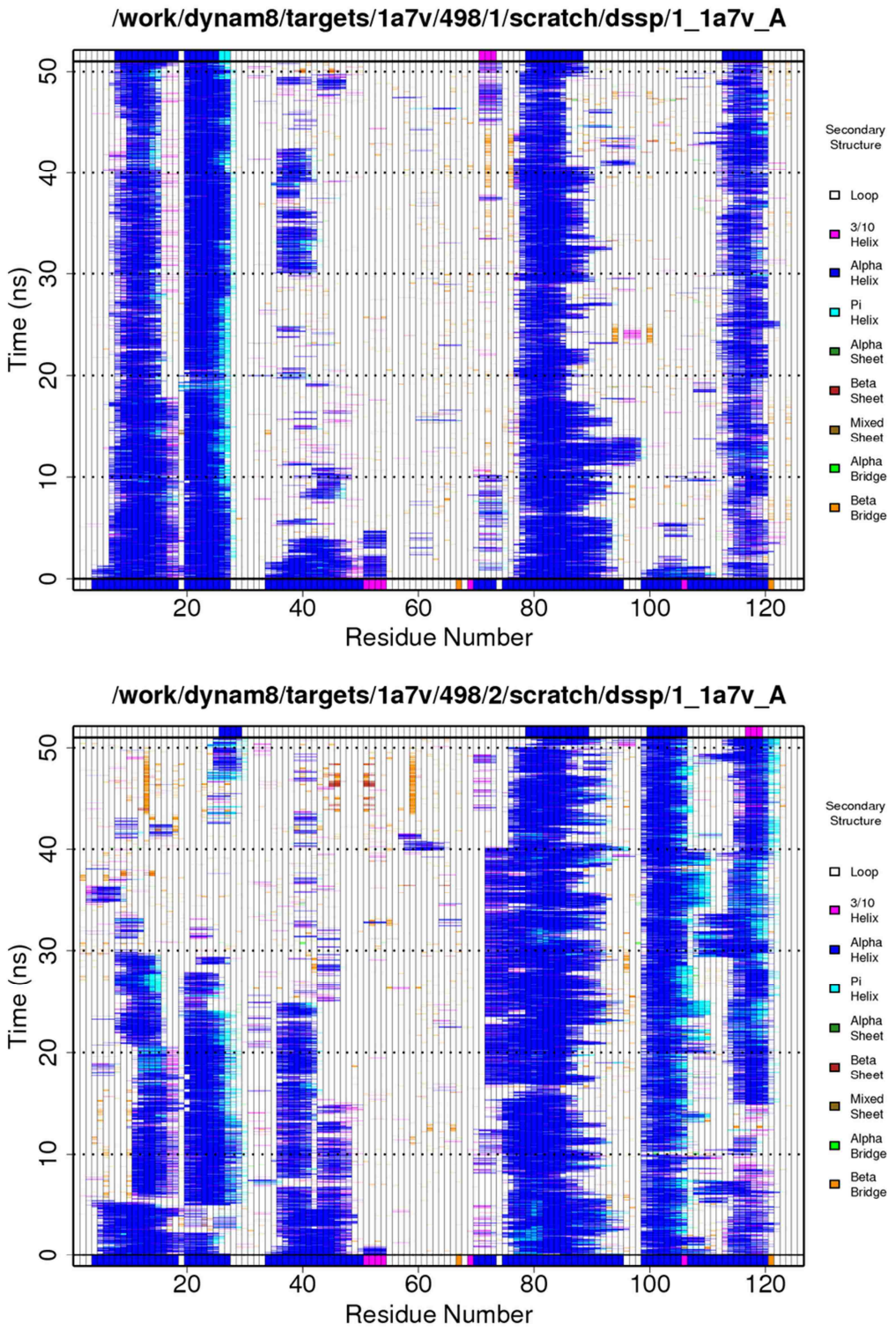


Figure 17: DSSP of WT Cyt*c'* (1A7V) simulations at 498 K up to 51 ns. Native helices are found in residues 4-27, 34-54, 75-95, and 99-120. The top figure is the first simulation and the bottom is the second.

Visually following the trajectories of WT as it unfolds at 498 K (Figure 18) shows the compactness of the protein even as helices lose secondary structure and portions of every helix remain at 51 ns except for Helix 1 of the second simulation.

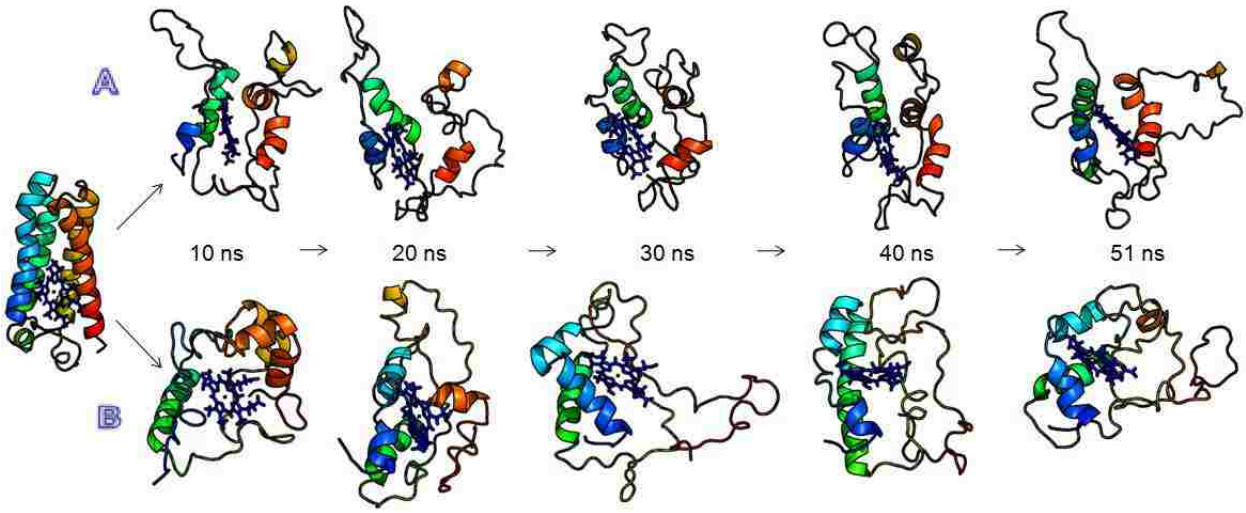


Figure 18: Snapshots of unfolding trajectories of pWT Cyt*c*′. The native structure and snapshots at 10, 20, 30, 40, and 51 ns are shown from two independent 498 K unfolding simulations. The native helical structures are shown by color: red, residues 4-27; yellow, residues 34-54; green, residues 75-95; blue, residues 99-120.

Two simulations of the thermal unfolding of pWT/A83S/A87S Cyt*c*′ were also performed out to 51 ns at 498 K. Secondary structure analysis is provided in Figure 19, and snapshots of the unfolding trajectories are presented in Figure 20. The first simulation, designated as simulation A, shows structure fluctuating over time, with all helices experiencing loss of structure to varying degrees. Helix 3 retains much of its structure for the duration of the simulation, while helix 4 also retains some structure, though not as much as helix 3. Helices 1 and 2 completely melt out by the end of the simulation. Helix 1 loses most of its structure at 9 ns and then regains some between 31-35 ns before finally melting out. Simulation B differs from A in several ways. While all helices experience fluctuation, helicies 1, 2, and 4 all retain some, but not all, structure throughout the simulation. Helix 3 loses structure from 15-30 ns, and then again after 45 ns. Helix 4 initially loses most of its structure, but regains a portion of it halfway through the simulation.

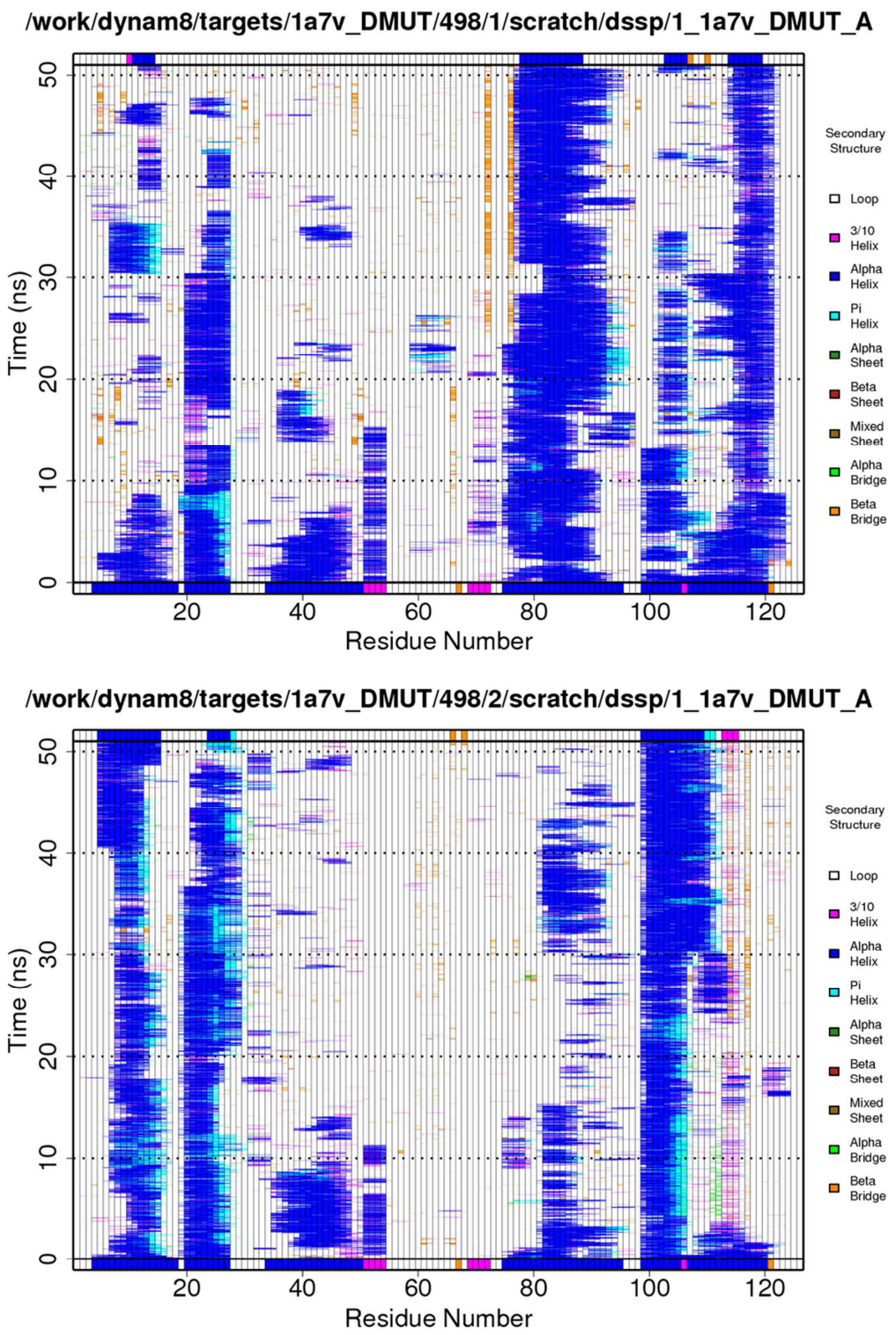


Figure 19: DSSP of WT/A83S/A87S Cyt*c'* simulations at 498 K up to 51 ns. Native helices are found in residues 4-27, 34-54, 75-95, and 99-120. The top figure is the first simulation and the bottom is the second.

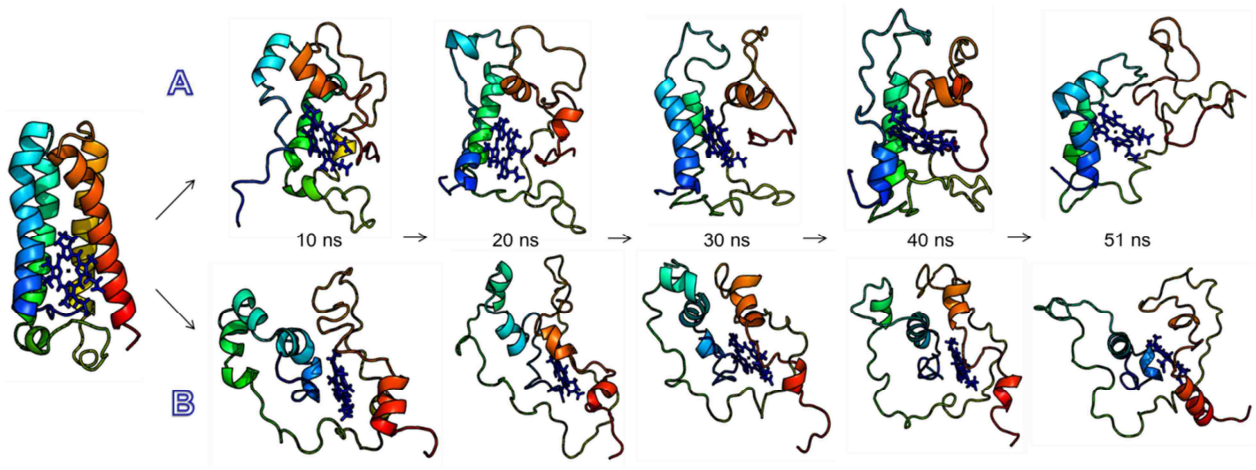


Figure 20: Snapshots of unfolding trajectories of WT/A83S/A87S Cyt*c*′. The native structure and snapshots at 10, 20, 30, 40, and 51 ns are shown from two independent 498 K unfolding simulations. The native helical structures are shown by color: red, residues 4-27; yellow, residues 34-54; green, residues 75-95; blue, residues 99-120.

Discussion

To further the study of Cyt*c*′ and the denatured state ensemble, we have expanded the application of the previously established thirteen single histidine variants[14] by introducing mutations in the center of the third helix in an attempt to establish the effects of disrupting the helix on the conformational biases of the denatured state ensemble. Here we have looked at thermodynamic and kinetic properties of Cyt*c*′, which include the A83S/A87S mutations, in the denatured state to be able to give a direct comparison to pWT.

Global Stability

Denaturation by gdnHCl was carried out on all variants, giving us $\Delta G_u°(H_2O)$, C_m, and *m*-values (Table 1). Of the thirteen variants studied, five have histidine substitutions in non-helical regions (D3H, K31H, D58H, A66H, K97H). The other eight variants have the single histidine mutation within a helix (K13H, K20H, K39H, K49H, E73H, K84H, A91H, and A104H). The mutations are destabilizing to the native protein, which can be explained by the lower helical propensity of histidine.[34; 35] Most of the variants, with the exception of E73H which is at the

end of the helix, have mutations that are approximately one turn into the helix. Two of the variants, K13H and K84H, are in the center of the helix. The anticipated decrease in helix propensity would be strongest in these center helix substitutions.[29; 34]

The comparison of $\Delta G_u°(H_2O)$ values of pWT to A83S/A87S variants did not have a consistent decrease for all variants (Fig. 7). Four variants (D3H, K20H, A66H, and K84H) showed no change, within error. The helix destabilizing histidine in the center of the third helix at position 84 explains the decreased change in $\Delta G_u°(H_2O)$ and C_m when comparing A83S/A87S to pWT. However, K20H is also in a helical region and D3H and A66H are in loop regions. This suggests that placement of the histidine in a loop region or helical region does not bring about a predictable effect based on secondary structure alone, thus factors beyond helix stability appear to affect global stability. Of the remaining variants that did show a relative change in $\Delta G_u°(H_2O)$ compared to pWT, all A83S/A87S variants had lower stability.

While the A83S/A87 $\Delta G_u°(H_2O)$ values had a general trend of lowered stability versus pWT with some variants as exceptions, the trend of C_m is much more plain (Figure 5). Here we see that each midpoint of the transition has lowered for every A83S/A87S variant, which reflects the destabilizing substitutions of alanine to serine in helical regions.[21] In addition, the plot of ΔC_m vs. sequence position appeared to be strikingly similar for each variant. Figure 6 reveals that the plot ΔC_m does seem to be featureless except for a valley starting with A66H and ending with K97H. The valley walls consist of the nearly identical values of E73H and A91H and dips to the lowest point at K84H, meaning that the least change is found in the variant surrounded by the destabilizing A83S/A87S mutations. This seemed counterintuitive until the effects of a histidine substitution in the very center of the helix are taken into account. AGADIR calculation of the K84H variant drops helix 3 from over 30% predicted helicity in pWT to under 10% in pWT/K84H. This diminishes the effects of serine substitutions and decreases the midpoint shift relative to other variants. Similar effects might have been expected at another variant near the center of a helix, K13H, but the predicted helical propensity of that region is small for pWT and moderates the histidine's destabilizing effect. This allows ΔC_m for K13H to reach the average ΔC_m.

Denatured State Loop Stability

Describing loop formation of a denatured protein can be a very useful approach to investigate the conformational properties and the degree to which a denatured protein fits or deviates from random coil behavior. Loop formation in denaturing conditions falls into well-developed polymer theory.[36] Scaling exponents can be calculated by determining the probability of closing loops of varying sizes for a particular polymer. The Jacobson-Stockmayer equation[37] (Eq. 9) is used to interpret loop formation as a function of loop size, where ν_3 is the scaling exponent, N is the number of

Eq. 9: $$\Delta S_{loop} = -\nu_3 R \ln(N) + R \ln\left((3/2\pi C_n \ell^2)^{\nu_3} V_i\right)$$

monomers in the loop, R is the gas constant, C_n is Flory's characteristic ratio, ℓ is the distance between monomers and V_i is the volume within which the ends of the chain must be constrained for the loop to form. This equation assumes random coil behavior, meaning that loop formation is assumed to be entirely entropic. The scaling exponent, ν_3, for a random coil with excluded volume is expected to be between 1.8 and 2.4 and can be easily determined from a plot of loop stability versus log of the loop size.

Rao et al. reported a scaling exponent of 2.5 ± 0.3 for pWT variants and considerable scatter about the best fit line.[14] To a large degree, A83S/A87S variants replicate this pattern of scatter about the best fit line (Fig. 12) and suggest a degree of robustness of the non-random behavior of the Cyt*c'* sequence. This is supported by Dar et al., who reported this same pattern of scatter for pWT data obtained in 6 M gdnHCl.[15] Data obtained for A83S/A87S variants are in some instances within error of pWT counterparts. However, the overall pattern of the small differences suggests a very real change in loop stability and yields a scaling exponent of 2.24 ± 0.24, similar to the value observed with the pWT variants.

The difference in scaling exponent creates a crossover between the best fit lines, and can be inferred to divide the protein into two sections; the region N-terminal to the disruption in helix 3, and the region C-terminal to the helix disruption. K84H can be thought to serve as an internal control, because of its position and since the pK_{loop}(His) values for pWT and A83S/A87S

are nearly identical. The engineered histidine at position 84 is a helix breaker, so helix 3 is disrupted in both pWT and pWT/A83S/A87S, and thus pK_{loop}(His) is expected to be similar for these two variants. With all the A83S/A87S variants after (more toward the N-terminus) showing more negative pK_{loop} values (higher stability) in relation to pWT variants, we can surmise that these loops are more flexible and this is conferring a small increase in loop stability. Positions more C-term to K84H (A91H, K97H, and A104H) all show little change, or perhaps lower stability for some positions, which is consistent with effects of helix propensity only being important if the helical segment is contained within the loop. These shortest loops may have less access to the heme (A104H in particular) because of possible excluded volume effects on the loop when the helix is disrupted.[38]

This pattern is also found in ΔpK_a(obs) (Fig. 11), where the data could be split into two sections; one N-term to the helix disrupting alanine to serine mutations, and one C-term. The N-term variants (E73H, A66H, D58H, K49H, and K31H) all achieve the average ΔpK_a(obs), but in the C-term group (A104H, K97H, A91H, K84H), only one variant (A91H) achieves ΔpK_a(obs). The low ΔpK_a(obs) values of the C-term group support the deduction that loop stability is unaffected in this region and that effects of helix propensity are only important if the helical segment is contained within the loop.

Interestingly, the scaling exponent of the A83S/A87S variants is nearly identical to the scaling exponent of poly-alanine coil inserts (2.26 ± 0.13), though the scatter of A83S/A87S reveals the effects of foldable native sequence and may be interpreted as areas of residual structure. This also strengthens the conclusions of Fitzkee and Rose[39], whose Monte Carlo simulations of rigid segments in protein denatured states showed the scaling exponent alone to be inadequate to determine random coil behavior.

Kinetics of Denatured State Loop Formation

Downward pH jump stopped-flow experiments were used to reveal how disrupting the third helix affects the loop kinetics of Cyt*c′* pWT/A83S/A87S variants in the denatured state. The results contained a strikingly similar k_b pattern to the data seen in Dar et al., displaying the same scatter up-and-down along the sequence (Fig. 13). This differs significantly from the poly-

alanine data, in which the k_b value levels out with increasing loop size. Since this is not the case with the A83S/A87S variants, this suggests nonrandom behavior in the denatured state.

Comparisons of the k_b pattern to denatured state pH titration data can also be made. When considering Figure 12, which is a plot of loop stability, pK_{loop}(His) vs. the logarithm of loop size, we see an arrangement of relative higher and lower stability variants. Except for the very longest and shortest loops, D3H and A104H respectively, the same relative pattern of higher and lower k_b values is seen for the plot of A83S/A87S variants. The lower k_b values comprise variants in positions of interest in the native structure. Positions 58 and 73, which are at the ends of a 20-residue Ω loop that brings helix 2 and 3 together, have low k_b values. Position 66, halfway between positions 58 and 73 in the loop between helix 2 and 3, has a high k_b value. This position also has a value above the best fit lines in Figure 12, indicating a less stable loop relative to positions 58 and 73 which are both below the best fit line. The k_b values for pWT/D58H/A83S/A87S and pWT/E73H/A83S/A87S are about 2 and 4 fold slower, respectively, than pWT/A66H/A83S/A87S. This translates into about a 0.40 kcal/mol and 0.82 kcal/mol [$RT \ln(k_{b66}/k_{b73})$] greater tendency for pWT/D58H/A83S/A87S and pWT/E73H/A83S/A87S than pWT/A66H/A83S/A87S to form persistent contacts in 3 M gdnHCl at 25 °C.

Another region of importance includes the positions 20, 31, and 39. As seen in Figure 13, these positions form a trough of low k_b values. The native structure of these positions establish the turn between helix 1 and 2. Relative to pWT/K13H/A83S/A87S, pWT/K20H/A83S/A87S has a 2.2 fold slower rate of loop breakage. On the other side of the trough, pWT/K39H/A83S/A87S has a rate of loop breakage 1.9 fold slower than pWT/K49H/A83S/A87S. This corresponds to an increased tendency of positions 20 and 39 to form a persistent contact by 0.48 and 0.38 kcal/mol over their neighbors at positions 13 and 49. The template of the natively folded protein may be molded by these energetic biases, as the loop between helix 2 and 3 and the turn between helix 1 and 2 would in all likelihood be favored even under denaturing conditions, allowing for efficient folding.

The amount of scatter about the best fit line found in k_f (Figure 14) is noticeably less than for pK_{loop}(His), which relates the importance of k_b in regards to deviation from random coil

behavior. Residual structure appears to be influenced by persistence of contacts, revealed by the scatter of k_b. This observation is supported by work done with iso-1-Cytc that also showed that deviations from random coil behavior can be attributed to loop persistence.[32; 40] While the overall pattern between pWT and pWT/A83S/A87S are similar, there are important observations to be made. There is a crossover of best fit lines approximately at position 97 (Figure 14), and while the changes brought by the mutations A83S/A87S are subtle, the k_f is larger for every variant N-term to position 84. This again divides the variants into portions N-terminal and C-terminal to the helix disrupting mutations, which has been seen in Figure 12, the plot of loop stability, pK_{loop}(His), vs. the logarithm of loop size. The k_f for positions 84, 91, 97, and 104 with the A83S/A87S mutations are either the same or are smaller than pWT. This supports the idea of helix propensity only being important if the helical segment is contained within the loop. The unchanged C-terminal portion of the A83S/A87S helix duplicates pWT, while the more flexible N-terminal loops have an increased k_f.

Molecular Dynamics Simulations of Cyt*c'* Unfolding

To serve as a baseline for comparison to WT/A83S/A87S, two unfolding simulations of WT at 498 K show preservation of the third helix throughout the duration. As this region had the highest predicted helical propensity, it was not surprising to see the structures endure. However, in the first simulation, helix 1 and 4 also retained more structure than expected, as parts of those helices remained at the 51 ns endpoint. The outcome of the second simulation was closer to the anticipated results, as the first and second helices lost structure near halfway through the simulation. However, the fourth helix also retained structure at the ends of the helix. All retained structure in both the pWT 498 K simulations corresponds well to the areas of increased predicted helix by the AGADIR algorithm.

Simulations of the unfolding of Cyt*c'* pWT/A83S/A87S show an increased response of the A83S/A87S variant to thermal denaturation, though not as much as anticipated. The first simulation shows retention of helical structure in the helix 3 region despite replacing helix favoring alanine residues with much less helix favorable serine residues. AGADIR predicted 5% or less helical propensity for that region of the protein, though the simulation suggests some

persistence of helix 3. The second simulation shows more disruption of the helix than the first simulation, but even though the helix melts into a loop structure at 15 ns, it regains some helical structure between 30 and 45 ns as well as continuously displaying flux between helix and loop when not in a stable helix structure. The accumulated results of the simulations speak to the robustness of this protein, as much structure was retained, or even regained after melting, at the high 498 K temperature (Figures 18 and 20).

Implications of Residual Structure

Studies of the conformational constraints on denatured state proteins show the importance of topology in contact formation and its effects on moderating early folding events. Smith et al. discuss the heme-edge vs. wrap-around topology of loops binding to the heme, and show that excluded volume plays a key role in limiting conformational space. The 2D square lattice model, using the asymmetry of the heme to study loop structure and excluded volume effects, clearly showed the favorable heme-edge topology favors loop formation. Topologies producing excluded volume reduced the conformations accessible to the loop. We can draw observations conducive to this model from the undisrupted chain stiffness of the third helix versus when the helix has serine substitutions. The third helix limits the loop flexibility in pWT, creating excluded volume effects. The increased loop flexibility, decreased excluded volume, and increased compactness produces more stable loops when the third helix is disrupted by the A83S/A87S substitutions. Residual structure also constrains conformational space and plays a role in directing the process of protein folding. In the protein used in this study, we have identified areas with more persistent contacts, which most likely impose topological constraints and reduction of conformational space. These structural biases in Cyt*c'* likely play a distinct role in efficient protein folding.

Conclusions

As there are many factors that influence protein folding, helical propensity remains an important part of understanding residual structure in the denatured state. The characterization of Cyt*c'* with A83S/A87S has shown the mutations to be destabilizing to global stability. The mutations also affect loop stability and kinetics. Loop stability and k_f values have increased,

albeit subtly, in the region N-terminal to the mutations, reflecting increased loop flexibility. k_b values show that persistent loops map to positions with important topological features in the native protein, particularly to positions of chain reversal. These persistent contacts form a template for folding events to form the native protein. MD simulations show particular robustness of the helices found in Cyt*c′*, with or without A83S/A87S, demonstrating the persistent residual structure in the compact denatured state of this protein. Further analysis of the MD simulations could provide more insight into these residual structures and the order of folding events for the native protein.

References

1. Shibata, N., Iba, S., Misaki, S., Meyer, T. E., Bartsch, R. G., Cusanovich, M. A., Morimoto, Y., Higuchi, Y. & Yasuoka, N. (1998). Basis for monomer stabilization in *Rhodopseudomonas palustris* cytochrome *c*' derived from the crystal structure. *Journal of Molecular Biology* **284**, 751-760.
2. Tanford, C. (1968). Protein denaturation. *Advances in Protein Chemistry* **23**, 121-282.
3. Neri, D., Wider, G. & Wuthrich, K. (1992). Complete 15N and 1H NMR assignments for the amino-terminal domain of the phage 434 repressor in the urea-unfolded form. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 4397-401.
4. Neri, D., Billeter, M., Wider, G. & Wuthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science (New York, N.Y.)* **257**, 1559-1563.
5. Shortle, D. (1995). Staphylococcal nuclease: a showcase of *m*-value effects. *Advances in Protein Chemistry* **46**, 217-245.
6. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M. & Schwalbe, H. (2002). Long-range interactions within a nonnative protein. *Science (Washington, DC, United States)* **295**, 1719-1722.
7. Shortle, D. (1996). Structural analysis of non-native states of proteins by NMR methods. *Current Opinion in Structural Biology* **6**, 24-30.
8. Lee, J. C., Engman, K. C., Tezcan, F. A., Gray, H. B. & Winkler, J. R. (2002). Structural features of cytochrome *c*' folding intermediates revealed by fluorescence energy-transfer kinetics. *Proceedings of the National Academy of Sciences, USA* **99**, 14778-14782.
9. Shan, B., Bhattacharya, S., Eliezer, D. & 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.
10. Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys. Physico-Chim. Biol.* **65**, 44-45.
11. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural Biology* **4**, 10-19.
12. Dill, K. A. & Shortle, D. (1991). Denatured states of proteins. *Annual Review of Biochemistry* **60**, 795-825.
13. 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.
14. Rao, K. S., Tzul, F. O., Christian, A. K., Gordon, T. N. & Bowler, B. E. (2009). Thermodynamics of loop formation in the denatured state of *Rhodopseudomonas palustris* cytochrome *c*': scaling exponents and the reconciliation problem. *Journal of Molecular Biology* **392**, 1315-1325.
15. Dar, T. A., Schaeffer, R. D., Daggett, V. & Bowler, B. E. (2011). Manifestations of native topology in the denatured state ensemble of *Rhodopseudomonas palustris* cytochrome *c*'. *Biochemistry* **50**, 1029-1041.
16. Chan, H. S. & Dill, K. A. (1990). The effect of internal constraints on the configurations of chain molecules. *Journal of Chemical Physics* **92**, 3118-3135.
17. Redner, S. (1980). Distribution functions in the interior of polymer chains. *Journal of Physics A: Mathematical and General* **13**, 3525-3541.
18. Hammack, B. N., Smith, C. R. & Bowler, B. E. (2001). Denatured state thermodynamics: residual structure, chain stiffness and scaling factors. *Journal of Molecular Biology* **311**, 1091-1104.

19. Wandschneider, E. & Bowler, B. E. (2004). Conformational properties of the iso-1-cytochrome *c* denatured state: dependence on guanidine hydrochloride concentration. *Journal of Molecular Biology* **339**, 185-197.
20. Muñoz, V. & Serrano, L. (1994). Elucidating the folding problem of α-helical peptides using empirical parameters III: temperature and pH dependence. *Journal of Molecular Biology* **245**, 297-308.
21. Pace, C. N. & Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical Journal* **75**, 422-7.
22. McGuirl, M. A., Lee, J. C., Lyubovitsky, J. G., Thanyakoop, C., Richards, J. H., Gray, H. B. & Winkler, J. R. (2003). Cloning, heterologous expression, and characterization of recombinant class II cytochromes c from Rhodopseudomonas palustris. *Biochimica et Biophysica Acta: General Subjects* **1619**, 23-28.
23. Lee, J. C., Gray, H. B. & Winkler, J. R. (2001). Cytochrome c ' folding triggered by electron transfer: Fast and slow formation of four-helix bundles. *Proceedings of the National Academy of Sciences, USA* **98**, 7760-7764.
24. Nozaki, Y. (1972). The preparation of guanidine hydrochloride. *Methods in Enzymology* **26**, 43-50.
25. Tonomura, B., Nakatani, H., Ohnishi, M., Yamaguchi-Ito, J. & Hiromi, K. (1978). Test reactions for a stopped-flow apparatus. Reduction of 2,6-dichlorophenolindophenol and potassium ferricyanide by L-ascorbic acid. *Analytical Biochemistry* **84**, 370-83.
26. Beck, D. A. C., Jonsson, A. L., Schaeffer, R. D., Scott, K. A., Day, R., Toofanny, R. D., Alonso, D. O. V. & Daggett, V. (2008). Dynameomics: mass annotation of protein dynamics and unfolding in water by high-throughput atomistic molecular dynamics simulations. *Protein Engineering Design and Selection* **21**, 353-368.
27. Beck, D. A. C. & Daggett, V. (2004). Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods* **34**, 112-120.
28. Beck, D. A. C., Armen, R. S. & Daggett, V. (2005). Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. *Biochemistry* **44**, 609-616.
29. Chakrabartty, A., Schellman, J.A., Baldwin, R.L. (1991). Large differences in the helix propensities of alanine and glycine. *Nature* **351**, 586-588.
30. Santoro, M. M. & Bolen, D. W. (1992). A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. *Biochemistry* **31**, 4901-4907.
31. Tzul, F. O. & Bowler, B. E. (2010). Denatured states of low complexity polypeptide sequences differ dramatically from those of foldable sequences. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11364-11369.
32. Kurchan, E., Roder, H. & Bowler, B. E. (2005). Kinetics of loop formation and breakage in the denatured state of iso-1-cytochrome *c*. *Journal of Molecular Biology* **353**, 730-743.
33. Kabsch, W., Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.
34. Baldwin, R. L. (1995). α-Helix formation by peptides of defined sequence. *Biophysical Chemistry* **55**, 127-135.
35. Rohl, C. A., Chakrabartty, A., Baldwin R.L. (1996). Helix propagations and N-cap propensities measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Science* **5**, 2623-2637.
36. Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*, W. H. Freeman and Co., San Francisco.
37. Jacobson, H. & Stockmayer, W. H. (1950). Intramolecular reaction in polycondensations. I. The theory of linear systems. *Journal of Chemical Physics* **18**, 1600-1606.

38. Smith, C. R., Mateljevic, N. & Bowler, B. E. (2002). Effects of topology and excluded volume on protein denatured state conformational properties. *Biochemistry* **41**, 10173-10181.
39. Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12497-12502.
40. Tzul, F. O. & Bowler, B. E. (2009). Importance of contact persistence in denatured state loop formation: kinetic insights into sequence effects on nucleation early in folding. *Journal of Molecular Biology* **390**, 124-134.

Chapter 1 Appendix

Table S1: Kinetic parameters for loop breakage for Cyt*c′* variants in 3 M gdnHCl and at 25 °C. WT data from Dar, et al.[15]

Variant	Loop Size (n)	pWT/A83S/A87S k_b, s^{-1}(raw) pH 3.62 ± 0.15	pWT/A83S/A87S k_b, s^{-1} (4 ms cutoff) pH 3.62 ± 0.15	pWT k_b, s^{-1} (4 ms cutoff) pH 3.64 ± 0.13	4 ms cutoff difference
A104H	10	252.4 ± 8.2	234.7 ± 1.4	228 ± 3	-6.7
K97H	17	98.5 ± 1.3	97.3 ± 1.6	94 ± 1	-3.3
A91H	23	101.5 ± 2.4	100.9 ± 2.8	105 ± 1	4.1
K84H	30	79.8 ± 0.7	77.9 ± 2.4	65.9 ± 0.6	-12.0
E73H	41	36.5 ± 0.8	36.1 ± 0.8	33.6 ± 0.2	-2.5
A66H	48	130.1 ± 3.1	127.4 ± 2.9	133 ± 2	5.6
D58H	56	75.2 ± 0.7	74.1 ± 0.8	73 ± 1	-1.1
K49H	65	151.0 ± 2.9	141.5 ± 3.9	164 ± 6	22.5
K39H	75	88.2 ± 1.7	87.5 ± 2.0	86 ± 1	-1.5
K31H	83	74.1 ± 0.8	73.21 ± 0.7	76 ± 1	2.8
K20H	94	55.2 ± 0.9	53.1 ± 1.3	54.2 ± 0.6	1.1
K13H	101	134.1 ± 3.7	124.3 ± 4.1	151 ± 4	26.7
D3H	111	130.9 ± 3.7	113.2 ± 4.9	82 ± 4	-31.2

Chapter 2: Studies of Residual Structure in the Denatured State of Iso-1-Cytochrome *c* by NMR

Abstract

To further investigate residual structure in denaturing conditions, we combine NMR methods with chemical shift secondary structure analysis on the iso-1 cytochrome *c* K54H variant in 3 and 6 M gdnHCl. We consider two pH conditions, pH 6.3 where the His54-heme loop is formed and pH 3.6 where the His54-heme loop is broken. A method designed to estimate the secondary structure propensities quantitatively was used to process chemical shift data obtained for each residue. Regions of residual structure were identified in 3 M gdnHCl, a condition at which the protein is fully denatured, as well as in the very harsh denaturant condition of 6 M gdnHCl. In 3 M gdnHCl at pH 3.6, a very defined helical region between residues 55-66 can be seen and may indicate a nucleation region for early folding events. The data presented here may contribute to the identification of residues and structural behavior involved in early folding which could lead to better understanding of protein folding pathways.

Introduction

The progression of events to a fully-folded protein from a disordered state is an area of interest held by many researchers.[1; 2] While early studies indicated the denatured state to be devoid of structure,[3] we now know that denatured proteins have much more complex behavior than random coils. This has been observed by NMR[4; 5; 6], fluorescence resonance energy transfer (FRET) studies[7], and small angle X-ray scattering (SAXS).[8]

Previous studies of yeast iso-1-cytochrome *c* (Cyt*c*) with single histidine substitutions to form His-heme loops showed one histidine variant with a surprisingly high affinity for the heme under denaturing conditions.[9; 10] This variant, K54H, displayed the lowest free energy of unfolding (ΔG_u), midpoint of unfolding (C_m), and most stable denatured state His-heme loop (lowest pK_a(obs)) of the seven variants investigated, which may point to a role for this region of the protein in guiding the collapse of the polypeptide chain during early folding events.[9; 11] This variant also had the largest denatured state loop formation *m*-value, indicating the highest burial of hydrophobic surface area upon loop formation.[12] Of the variants tested, this variant,

with a moderate-sized His-heme loop of 37 residues, had the highest propensity for forming a loop. It was an order of magnitude more favorable than most other His-heme loops, and twice as favorable as the longest His-heme loop, which was 83 residues long.[12] The presence of loops that connect local structural elements may be the most important feature in determining a folding pathway.[13] In addition, kinetic studies of the same set of variants showed an unexpectedly slow rate of breakage of the very stable His54-heme loop, spotlighting the strong stabilization of the closed loop.[14] Residual structure may be induced by the constraint imposed by loop formation. Therefore, this may be seen as a model system of what occurs as a protein collapses during folding and gives us an opportunity to characterize the nature of residual structure formation produced by His-heme loop formation.

To further investigate the residual structure produced by the His54-heme loop formed by the Cyt*c* K54H variant in denaturing conditions, we combine NMR methods with chemical shift secondary structure analysis. It is known that chemical shifts are responsive to secondary structure and can be used to characterize a residual structure.[15] By examining the protein with both the broken or formed His-heme loop (Figure 1), we identify regions of residual structure in harsh denaturing conditions, which may indicate nucleation regions for early folding events.

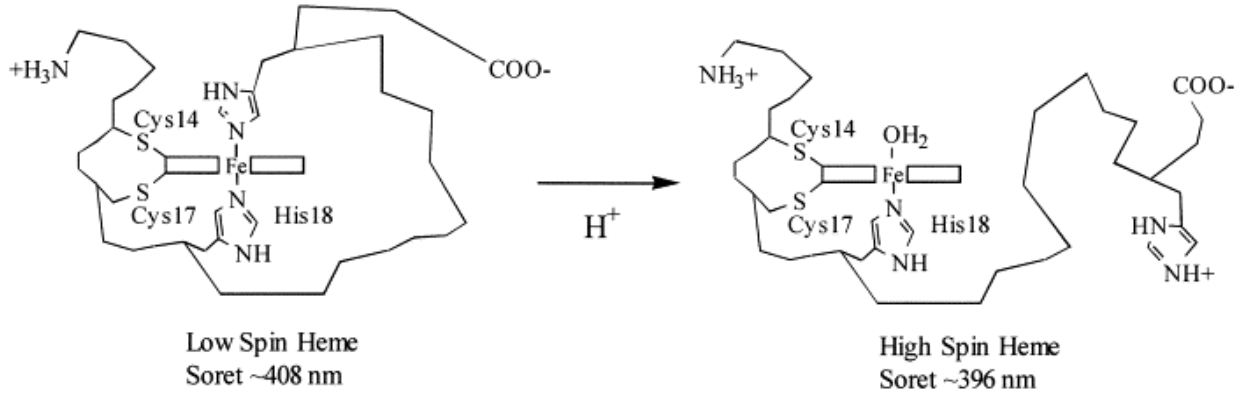


Figure 1: Schematic representation of His–heme loop formation in the denatured state of Cyt*c* .

Experimentals

Preparation of Iso-1-Cytochrome *c*

A RbsTM vector containing Cyt*c* with T(-5)P/N52I/K54H/K72A was prepared by former post-doctoral fellow Tanveer Dar. To verify that the sequence contained the desired mutations, a heat shock protocol was followed to introduce the plasmid into TG-1 cells for DNA amplification. 100 µL of TG-1 cells were thawed on ice and 2 µL of the approximately 100 ng/µL vector was added and mixed by gentle pipetting. This solution incubated on ice for 45 minutes and then incubated in a 42 °C water bath for 45 seconds before being placed back on ice for 5 minutes. Then, 1 mL LB-media was added. This mixture was transferred to a 15 mL tube and incubated at 37 °C for 1 hour in a shaking water bath. 300 µL of this inoculated media was plated on LB-agar plates containing 100 µg/mL ampicillin and left overnight in a 37 °C incubator. Five colonies from the plate were picked and used to inoculate individual solutions of 5 mL LB media containing 100 µg/mL ampicillin and left overnight to grow at 37 °C in a shaking water bath. DNA was extracted using the Promega Wizard Plus DNA miniprep kit and sent to the Murdock DNA Sequencing Facility at the University of Montana. The NRbs primer sequence d[CACAGGAAACAGCTATGACC] was used for sequencing.

Expression of Isotopically Labeled Iso-1-Cytochrome *c*

Cells were transformed with the RbsTM vector, which contained the Cyt*c* sequence with T(-5)P/N52I/K54H/K72A mutations as well as the ampicillin resistance gene. Normally, a 20 µL aliquot of competent BL21-DE3 *E. coli* cells (Edge Bio) was removed from the -80 °C freezer, thawed on ice and then 2 µL of the RbsTM vector was added. The cells were mixed by gentle pipetting and incubated on ice for 45 minutes before being transferred to a 42 °C water bath for 45 seconds. The cells were then incubated again on ice for 5 minutes. The cells were mixed into a 15 mL tube containing 1 mL LB media and incubated at 37 °C in a shaking water bath for 1 hour. 300 µL of this inoculated media was plated onto LB-agar plates containing 100 µg/mL ampicillin and left to incubate overnight at 37 °C. The resulting colonies were used in two different ways. In the first way, typically 5 individual colonies were picked and grown in 5 mL LB media containing 100 µg/mL ampicillin in a shaking water bath overnight at 37 °C. One mL of each of these starter cultures was pelleted by centrifugation to visually compare the color of

the cell pellet. The pellet with the deepest pink/red color was chosen, and the remaining 4 mL of that culture was used to inoculate 1 L of 100 µg/mL ampicillin M9 minimal media. Minimal media consisted of the following components: 10 mL of 0.1 g/mL $^{15}NH_4Cl$, 1 mL of 1 M $MgCl_2$, 100 µL of 1 M $CaCl_2$, 1 mg thiamine, 5 mg $FeSO_4 \bullet 7H_2O$, 1 mg biotin, 4 g ^{13}C glucose, 16.8 mg δ-aminolevulinic acid, and 100 mL of a 10 × M9 salts solution at pH 7.4 which contained 67.8 g/L Na_2HPO_4, 30 g/L KH_2PO_4, and 5 g/L NaCl. The water, M9 salts solution, and $MgCl_2$ solution was each autoclaved. The rest of the components were filter sterilized. The second way to inoculate the media was to resuspend the full plate of colonies into 3 mL sterile LB media and use this solution to inoculate a 1 L culture of minimal media containing 100 µg/mL ampicillin. These cultures were grown overnight at 30 °C or 37 °C with shaking at 165-200 rpm. Cell pellets were harvested using a Sorvall GS-3 rotor in a Sorvall RC-5C Plus floor model centrifuge running at 5000 rpm for 10 minutes. The supernatant was discarded and the resulting cell pellet was stored at -80 °C.

Expression of Iso-1-Cytochrome *c* in Rich Media

Cyt*c* was also expressed using autoclaved Terrific Broth (12 g/L tryptone, 24 g/L yeast extract, 4 mL/L glycerol, 2.3 g/L KH_2PO_4, 12.5 g/L K_2HPO_4) as the media. This media also contained 100 µg/mL ampicillin, which was filter sterilized and added after the media was autoclaved. The rest of the procedure was the same as in the isotopically labeled minimal media preparations.

Lysis and Crude Purification of Iso-1-Cyt*c*

The cell pellet was removed from the -80 °C freezer and allowed to thaw in the 4 °C cold box and then returned to the -80 °C freezer. Two more rounds of freeze/thaw were performed. On the final thaw, the cells were resuspended in Lysis Buffer (100 mM Tris, pH 8.0, 500 mM NaCl, 1 mM EDTA, 2 mM PMSF plus a few crystals of DNAse1 and RNAse). Lysis buffer was used at 2 ml/g of cells to be lysed. The cell slurry was run through the French press apparatus 8-10 times to break the cells. This cell slurry was centrifuged twice using a GSA rotor in a Sorvall RC-5C+ floor model centrifuge at 10,000 rpm for 30 minutes, keeping the supernatant. An ammonium sulfate cut was used to precipitate undesired proteins. The solution was brought up to 50% saturation of ammonium sulfate (31.4 g ammonium sulfate per 100 mL of lysate), and stirred 4-

5 hours or overnight at 4 °C. The solution was centrifuged at 10,000 rpm for 30 minutes, discarding the pellet. The supernatant was transferred to 3500 MWCO dialysis tubing and placed in 12.5 mM sodium phosphate dialysis buffer (2744 mL H_2O, 1.72 g sodium phosphate monobasic monohydrate, 3.11 g sodium phosphate dibasic anhydrous, 1.08 Na_2EDTA, 385 µL β-ME) and stirred gently overnight at 4 °C. The dialysis tubing was transferred to a second bottle of 12.5 mM sodium phosphate dialysis buffer and stirred gently for 24 hours. The Cyt*c* solution was removed from the dialysis bags and centrifuged at 10,000 rpm for 30 minutes to remove any precipitate.

Purification of Cyt*c* by CM-Sepharose

Approximately 100 mL of CM-Sepharose resin was equilibrated to a buffer containing 50 mM sodium phosphate, pH 7.2, 1 mM EDTA, and 2mM β-ME, and the Cyt*c* solution was added to the resin. After an hour of stirring at 4 °C, the resin was allowed to settle and the buffer was decanted off. The resin, now pink with Cyt*c* protein, was resuspended in a small amount of the 50 mM sodium phosphate, pH 7.2, 1 mM EDTA, and 2 mM β-ME buffer and was packed into a column. Approximately 300-400 mL of the buffer was run through the column as a wash. A 200 mL linear 0-0.8 M NaCl gradient was used to elute the protein from the CM-Sepharose with 50 mM sodium phosphate, pH 7.2, 1 mM EDTA, and 2 mM β-ME buffer and 50 mM sodium phosphate, pH 7.2, 1 mM EDTA, 2 mM β-ME, and 0.8 M NaCl buffer. The eluted protein was concentrated by ultrafiltration using Millipore Centricon and Centriprep 10,000 MWCO (molecular weight cut-off) centrifugal filters running on a benchtop Jouan CR4i centrifuge, and exchanged to HPLC A buffer (50 mM sodium phosphate, pH 7.0).

Purification of Cyt*c* by HPLC

Protein was purified by two methods. In one method, the protein was reduced by a small amount of sodium dithionite, and oxidized by potassium ferricyanide in the other. Protein was loaded onto a BioRad UnoS6 cation exchange column and eluted with HPLC Buffer B (50 mM sodium phosphate, 1 M NaCl, pH 7.0). The gradient used, with a flow rate of 3 mL/minute, was as follows: 0-7 minutes 0% B, 7-34 minutes increase linearly to 30% B, 34-40 minutes hold at 30% B, 40-43 minutes increase linearly to 100% B, 43-50 minutes hold at 100% B, 50-53 minutes

decrease linearly to 0% B, 53-63 minutes hold at 0% B. The eluent was collected from the largest peak and exchanged back to HPLC Buffer A (50 mM sodium phosphate, pH 7.0) using the above mentioned Millipore centrifugal filters and concentrated. The protein concentration was determined spectroscopically using a Beckman Coulter DU 800 Spectrophotometer at the isosbestic points 339, 526.5, and 541.75 nm of the Cyt*c* spectrum as well as the oxidized-state extinction coefficients at 360 and 550 nm.[16; 17]

Global Thermodynamic Stability Measurements

The same procedure found in Chapter 1 was used to measure global stability of Cyt*c* T(-5)P/N52I/K54H/K72A. The only differences were the buffer (20 mM Tris, 40 mM NaCl, 1.0 mM NaEDTA, pH 7.0) and the range and steps of the gdnHCl titration. The data were acquired in steps of 0.05 from 0 M to 2.2 M gdnHCl.

Denatured State Loop Formation

The same procedure as found in Chapter 1 was used to measure His-heme loop stability of Cyt*c* T(-5)P/N52I/K54H/K72A.

Preparation of Cyt*c* for NMR

For protein that was previously reduced by sodium dithionite potassium ferricyanide (5 mg per mg of protein) was used to oxidize it after HPLC. A MALDI-TOF mass spectrum was collected with an Applied Biosystems Voyager mass spectrometer to check purity, and this gradient consistently produced protein with one peak at 13252 Da, which is very near the expected mass of 13252.1 Da for uniformily ^{15}N-, ^{13}C-labeled protein. A 3 × protein in 3 × buffer (1200-1800 µM protein concentration, 15 mM Na_2HPO_4, and 45 mM NaCl) stock solution was made by loading the protein solution onto a G-25 column equilibrated in 3 × Pi Buffer. The ferricyanide was separated from the protein by this column, and the protein was exchanged into the buffer. A 600 µL sample was created by the combination of appropriate amounts of 3 × 3 × protein buffer stock, 6 M gdnHCl, D_2O and water to create a protein solution in 1 × buffer at 3 M gdnHCl and 10% D_20. The final protein concentration for experiments run in 3 M gdnHCl was ~700 µM. To prepare a sample in 6 M gdnHCl, a 6 × 6 × protein buffer stock, 8 M gdnHCl, D_2O and water

were used to create a protein solution in 1 × buffer at 6 M gdnHCl and 10% D_20. The final protein concentration for experiments run in 6 M gdnHCl was ~400 µM. The pH was measured with a Denver Instruments UB-10 UltraBasic benchtop pH meter. The procedure for changing the pH of the sample was very similar to the procedure found in the Cyt*c'* pH titration study. For example, to change the pH in 3 M gdnHCl without changing the protein or gdnHCl concentration of the sample, a total of 12 µl was added to the sample: 3 × protein stock (4 µl), ~6 M gdnHCl (~6 µl), D_20 (1.2 µl)and HCl (~.8 µl) in appropriate ratios and thoroughly mixed by pipetting. The volume of HCl added was constant, but the concentration of the acid varied to adjust the pH of the sample.

NMR Experiments

NMR experiments were run by the Briknarova lab. Data from the following experiments were collected: HSQC, HNCACB, CBCACONH, CCONH, HNCO, HNCACO [18] for 3 and 6 M gdnHCl at pH 6.3 and 3.6. HSQC data were also collected for 3 M gdnHCl at pH 5.0. The chemical shifts were recorded after processing the data with NMRpipe [19] and assignments were made using the Sparky NMR assignment and integration software (http://www.cgl.ucsf.edu/home/sparky/).[20]

Results

Design and Characterization of the His54 Variant

Specific mutations were introduced into the sequence of Cyt*c* WT for protein expression in *E. coli*. To avoid interference of His54-heme loop formation, a proline substitution at position -5 shifts the pK_a(obs) for loop formation of the N-terminal imino group with the heme to 6.67 ± 0.01 (T. Dar unpublished results) compared to 5.89 ± 0.06 without the substitution.[21] The N52I mutation serves as a stabilizing mutation, while the K72A mutation removes a potential binding ligand to the heme that affects the alkaline conformational transition.[22]

Denaturation by gdnHCl was carried out on Cyt*c* T(-5)P/N52I/K54H/K72A, giving $\Delta G_u°(H_2O)$, *m*-value, and C_m to respectively be 3.91 ± 0.15 kcal/mol, 3.93 ± 0.18 kcal/(mol*M), and 0.99 ± 0.01 M gdnHCl over 4 runs. This is very similar to the values for AcH54I52 iso-1-Cyt*c* which are 3.93

± 0.02 kcal/mol for ΔGu°(H_2O), 3.96 ± 0.02 kcal/(mol*M) for the *m*-value, and 0.99 ± 0.01 M gdnHCl as the midpoint, C_m.[23] A representative titration curve is found in Figure 2.

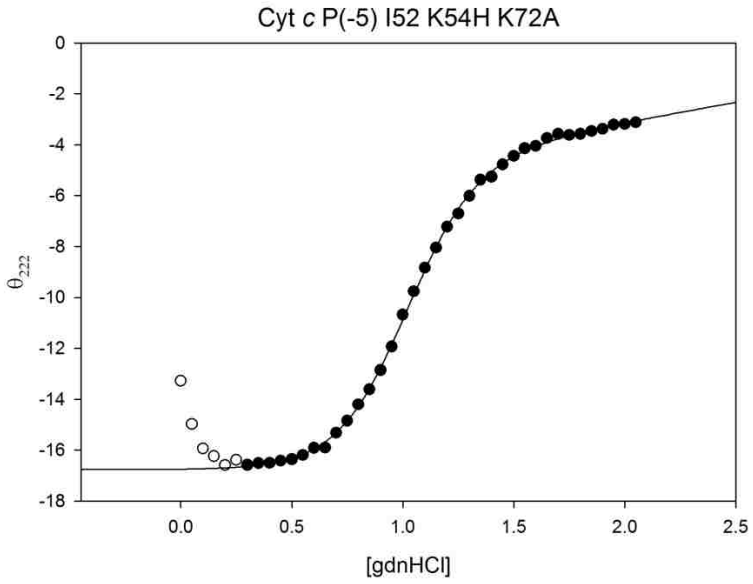


Figure 2: Representative curve of gdnHCl denaturation monitored by circular dichroism at 222 nm. Data were acquired at pH 7.0. White circles were not included in the fit.

The denatured state loop formation equilibrium was evaluated at 398 nm for Cyt*c* T(-5)P/N52I/K54H/K72A. Over three runs, the observed pK_a was 4.74 ± 0.04 and the average estimated number of protons involved in loop formation, n_p, was 0.78 ± 0.05. These values were also very similar to reported values of AcH54I52 which are 4.80 ± 0.02 for the observed pK_a and 0.83 ± 0.03 for n_p.[12] A representative titration curve is found in Figure 3. The data demonstrates that the His54-heme loop is fully formed near pH 6.3 and fully broken near pH 3.6.

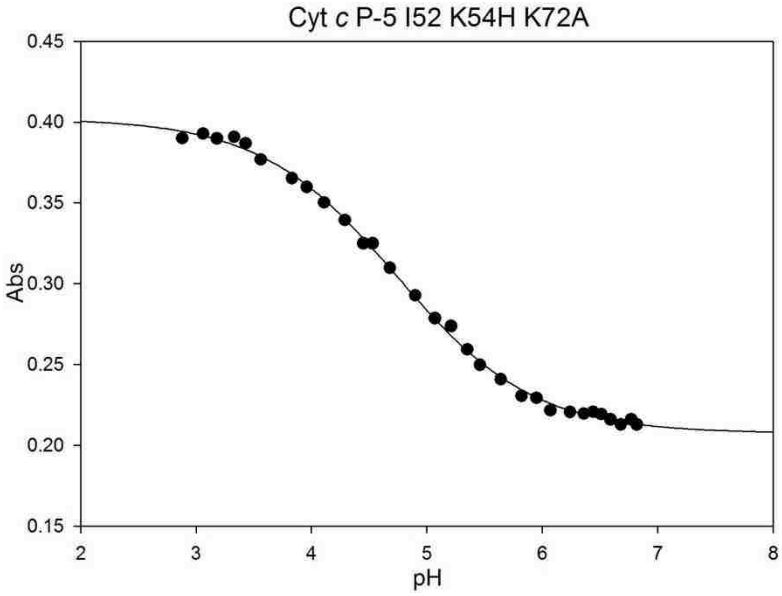
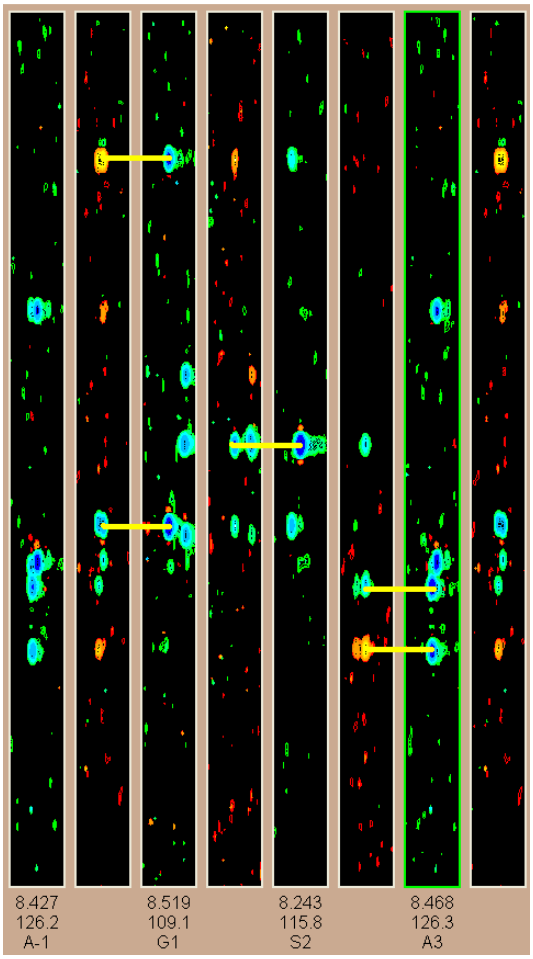


Figure 3: Representative titration curve of histidine-heme ligation as a function of pH at 22 ± 1 °C in 3 M gdnHCl.

NMR Assignments

NMR assignments were made in 3 M and 6 M gdnHCl at both pH 6.3 and pH 3.6. The protein is fully denatured in all conditions (Figure 2). At pH 6.3, the His-heme loop is formed while at pH 3.6, the His-heme loop is broken (Figure 3). To determine specific residues from chemical shifts, a process of identifying a residue with a unique chemical shift and then matching previous or next chemical shifts adjacent to the first residue was employed. For example, in Figure 4, strip plots show the connectivity from positions -1 through 3. Position 3 was first identified by its Cα and Cβ chemical shifts of 52.68 ppm and 19.27 ppm and the previous residue's Cα and Cβ chemical shifts of 58.4 ppm and 64.39 ppm. As alanine has distinctive Cα and Cβ chemical shifts of approximately 52 ppm and 20 ppm, and serine has the distinctive Cα and Cβ chemical shifts of approximately 58 ppm and 64 ppm, and the only place in the sequence where alanine was preceded by a serine occurs at position 3, the corresponding HSQC peak was assigned as A3. The connectivity is followed through position 2, as the serine residue NMR data shows a preceding residue having a chemical shift of approximately 45 ppm, another very distinctive shift, this time belonging to glycine. As this is the only occurrence of serine preceded by a glycine, the assignment was made. Some assignments were made with considerably more effort as several residues did have similar chemical shifts and sequence as other parts of the protein. In these instances, side-chain and terminal chemical shifts had to be

used to make the determination. Careful recording of chemical shifts and denatured state residue average chemical shift tables aided greatly. HSQC's with assigned residues are found in Figures 5 (3 M gdnHCl, pH 6.3 and pH 3.6) and 6 (6 M gdnHCl, pH 6.3 and pH 3.6). Chemical shifts are reported in the appendix to this chapter.



position		Cα	Cβ	prev Cα	prev Cβ
-1	A	52.95	19.65	56.13	33.75
1	G	45.61	-	53	19.69
2	S	58.41	64.24	45.62	-
3	A	52.68	19.67	58.4	64.39

Figure 4: Strip plot views of assigned residues in 3 M gdnHCl and at pH 6.36. Data were determined from HNCACB and CBCACONH experiments and chemical shifts are listed in the table above. Yellow lines indicate connectivity between residues.

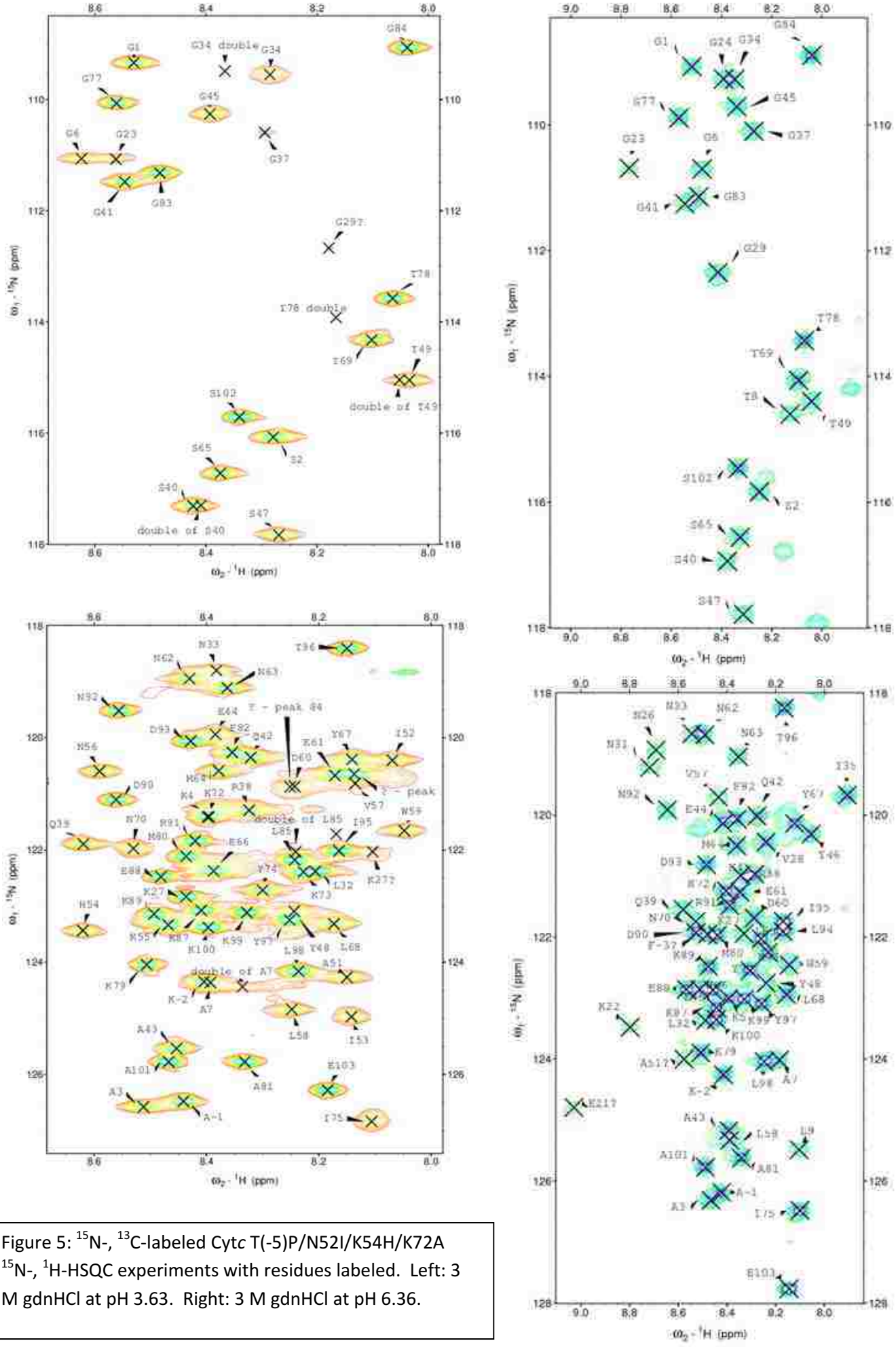


Figure 5: ^{15}N-, ^{13}C-labeled Cyt*c* T(-5)P/N52I/K54H/K72A ^{15}N-, ^{1}H-HSQC experiments with residues labeled. Left: 3 M gdnHCl at pH 3.63. Right: 3 M gdnHCl at pH 6.36.

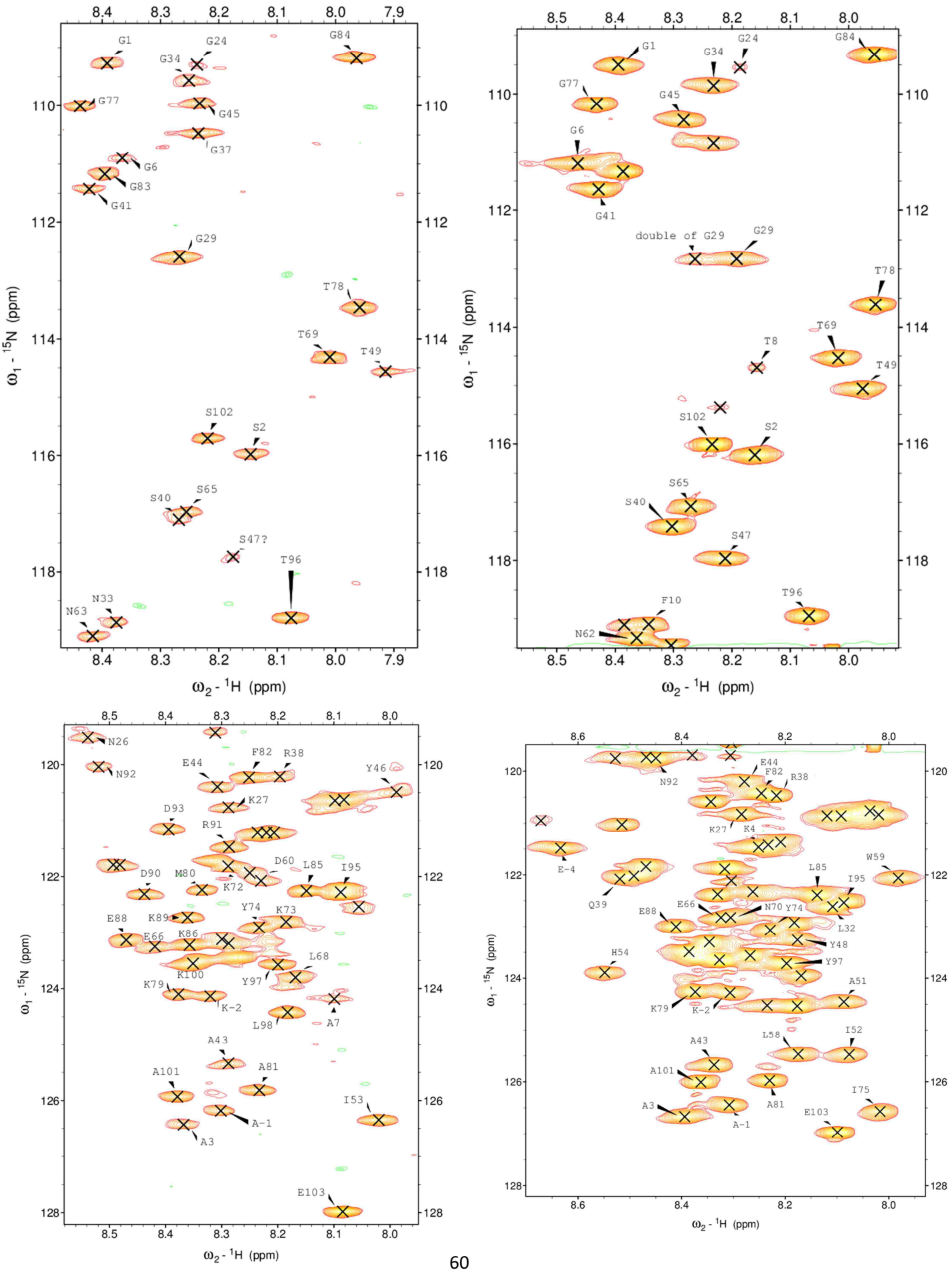


Figure 6: ^{15}N-, ^{13}C-labeled Cyt*c* T(-5)P/N52I/K54H/K72A ^{15}N-, ^{1}H-HSQC experiments with residues labeled. Left: 6 M gdnHCl at pH 6.36. Right: 6 M gdnHCl at pH 3.63.

Calculation of Secondary Structure Propensities

To examine the tendency of Cyt*c* protein to retain secondary structure in the denatured state, the chemical shifts obtained for each residue were analyzed with a method designed to estimate the secondary structure propensities quantitatively. The secondary structure propensity of Cyt*c* was calculated using Equation 1.[24]

Equation 1:

$$SSP_i = \frac{\sum_{j=i-2}^{i+2} \sum_X \begin{cases} \frac{|\Delta\delta X_j obs|}{\sigma_j \alpha} \text{ if } (\delta\Delta X_j obs)(\delta\Delta X_j \alpha) > 0 \\ -\frac{|\Delta\delta X_j obs|}{\sigma_j \beta} \text{ if } (\delta\Delta X_j obs)(\delta\Delta X_j \beta) > 0 \end{cases}}{\sum_{j=i-2}^{i+2} \sum_X \begin{cases} \frac{|\Delta\delta X_j \alpha|}{\sigma_j \alpha} \text{ if } (\delta\Delta X_j obs)(\delta\Delta X_j \alpha) > 0 \\ -\frac{|\Delta\delta X_j \beta|}{\sigma_j \beta} \text{ if } (\delta\Delta X_j obs)(\delta\Delta X_j \beta) > 0 \end{cases}}$$

Equation 1 shows the SSP score for a given residue i, where $\Delta\delta X_j$*obs* is the observed secondary chemical shift, $\Delta\delta X_j\alpha/\beta$ are the average secondary chemical shifts of fully formed secondary structure (α or β), and $\sigma_j\alpha/\beta$ are the standard deviations of $\Delta\delta X_j\alpha/\beta$. As the score reflects the fraction of secondary structure at a specific position, fully formed secondary structure would get an absolute value of 1. A positive score reflects α structure while a negative score reflects β structure. This method of using a weighted average over a five residue window has the advantage of minimizing chemical shifts that are poor measures of secondary structure, such as the chemical shifts for glycine. Since glycine has no C^β, it may give extreme values in the β-structure region.

SSP scores for Cyt*c* are found in Figure 7. The scores are given on a per residue basis for 3 and 6 M gdnHCl at pH 6.3 and 3.6. To get an SSP score for a specific residue, except for glycine, both C^α and C^β values must be supplied for that residue, and therefore scores for all residues in all conditions were not able to be calculated. In 3 M gdnHCl at pH 6.3, two regions were unable to be assigned. The region around the heme attachment (residues 10 through 21) and the region around K54H (residues 52 through 55, due to the binding of K54H to the heme,) were subject to paramagnetic effects and not able to be assigned and therefore have no SSP score.[25] The results show alternating regions of α and β structure propensity. Positions 7 to 9, 22 to 24, 31

to 37, 39, 47 to 51, 62 to 67, 78, 84, 92 to 97, and 102 to 103 all show α structure propensity in 3 M gdnHCl. Positions -5 to 6, 26 to 29, 38 to 46, 58 to 61, 68 to 77, 79 to 91, and 98 to 101 all show β structure propensity in 3 M gdnHCl at pH 6.3. A very similar pattern was found for assigned Cyt*c* residues in 6 M gdnHCl at pH 6.3, though with fewer residues assigned. The region around the heme and K54H are again missing.

The SSP scores for Cyt*c* in 3 M gdnHCl at pH 3.6 also show regions of α and β structure propensity, though differently than at pH 6.3. At the lower pH, the loop between K54H and the heme is broken and relieved of paramagnetic effects. This region, between residues 54 to 67, gives the highest score of α structure propensity of all the conditions and peaks at over 0.3, corresponding to a greater than 30% α structure propensity. There is another region of α structure propensity between residues 85 and 96, peaking at residue 90. These conditions also have the region of highest β structure propensity. Residues 70 through 77 form a region of β structure propensity peaking over 30 % at residue 75. The 6 M gdnHCl data at pH 3.6 also had fewer assignable residues, leaving gaps and skewing SSP scores as the weighted average was attenuated.

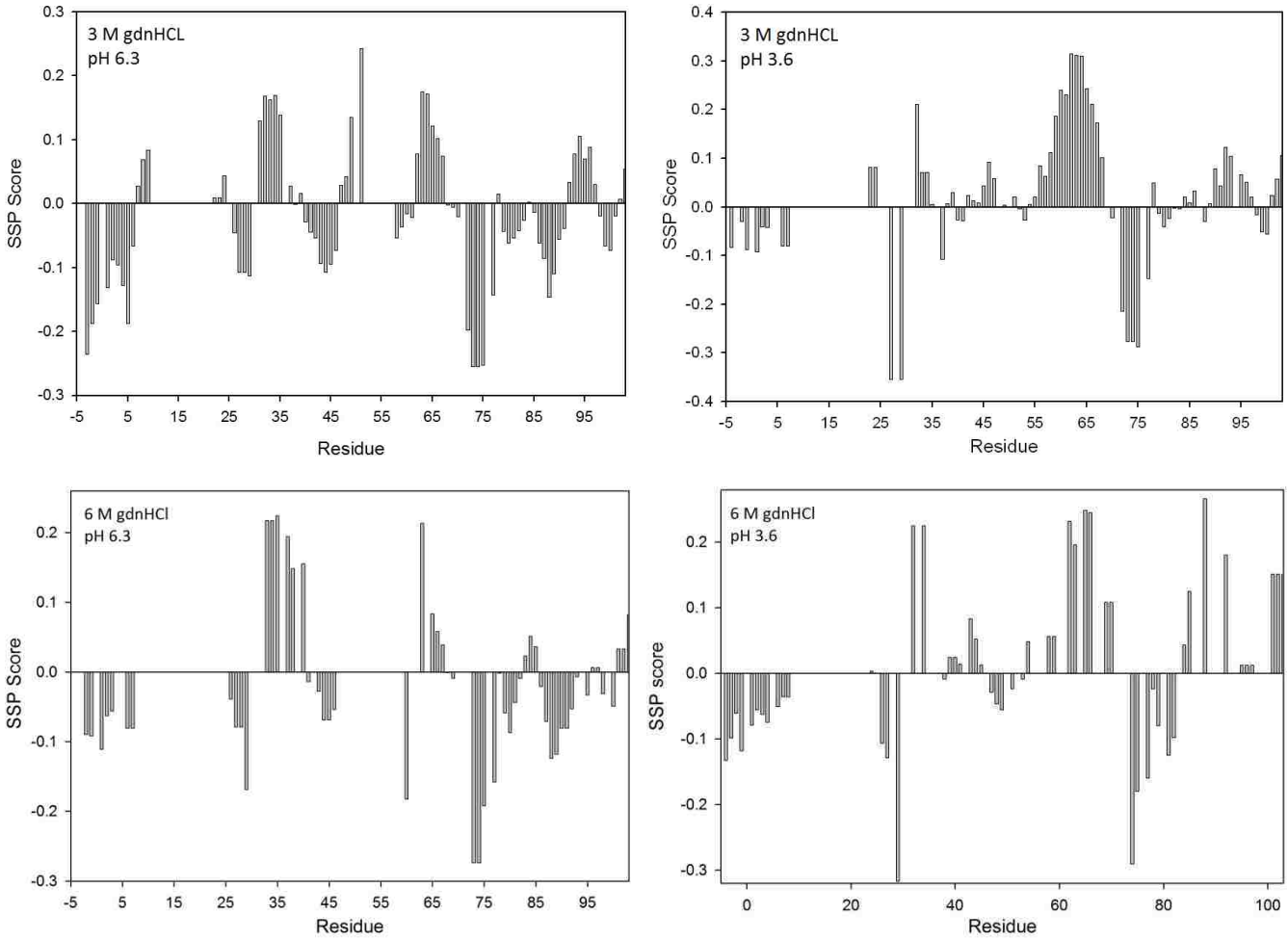


Figure 7: SSP scores on a per residue basis for Cyt*c* in 3 M (top) and 6 M (bottom) gdnHCl and at pH 6.3 (left) and pH 3.6 (right).

Discussion

While some denatured proteins can move toward random coil behavior in extreme denaturing conditions,[3; 26] we see Cyt*c* K54H having clear areas of residual structure based on chemical shift analysis even in the harsh denaturing conditions of 6 M gdnHCl. We see when the loop is not formed in 3 M gdnHCl at pH 3.6, a very defined helical region between residues 55-66, followed by a region displaying more beta sheet character in residues 69-77. While the data is less complete for the 6 M gdnHCl at pH 3.6 condition, the helical region between residues 55-66 is still apparent, as is the beta character region more towards the C-terminus. In the native protein, position 54 is in a helical region that precedes a short beta region (residues 57-60). A

helical region is again found beginning at residue 61 through 76 with a kink in the helix centered at position 71. These observations confer that while the native form has some common features as the secondary structure analysis in denaturing conditions, the denatured form has its own distinct character.

Figure 8 shows the change brought by increasing the gdnHCl concentration from 3 M to 6 M for both pH 6.3 and 3.6. The higher gdnHCl concentration was used to attempt to correct for paramagnetic effects as well as to observe the effect, if any, of harsher denaturing conditions on residual structure. Only residues with data in both gdnHCl concentrations can be considered for each pH. For most residues, the change between the two denaturing conditions is small and gives SSP scores with absolute values less than 0.1. The region between residues 25 and 46 mostly consists of a flexible loop in the native protein. When the loop is formed (pH 6.3), this region is primarily β-structure, but when the loop is broken (pH 3.6), this region starts taking on more α-helical character as the sequence progresses. The C-terminal region (residues 80-103) at pH 6.3 shows ΔSSP scores not present at pH 3.6. This may reflect the effect of having a constrained His-heme loop.

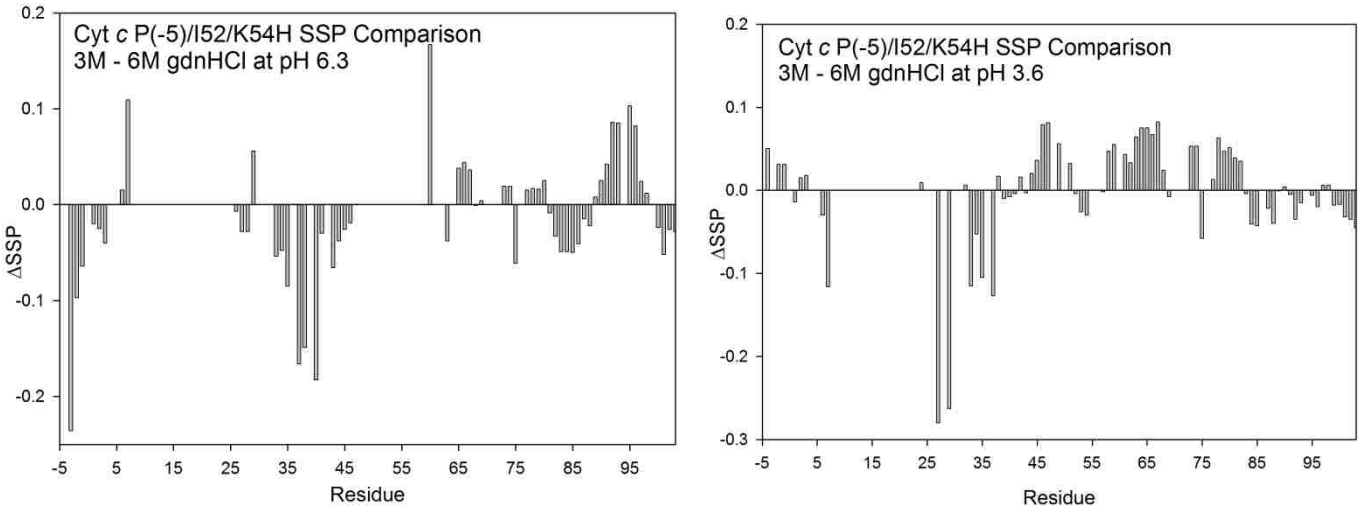


Figure 8: ΔSSP of Cyt*c* in gdnHCl for pH 6.3 (left) and pH 3.6 (right).

Jeng and Englander, using a combination of circular dichroism and hydrogen exchange, investigated Cyt*c* and reported that while many tertiary hydrogen bonds and sidechain interactions were lost under acidic conditions, nearly all the helical content was retained.[27] These regions possessed significant equilibrium stability and supported a more complex viewpoint than the two-state model. Krishna et al. built upon this work and described an ordered folding process to Cyt*c* that included specific foldons, cooperative native-like structural units. The work presented here shows residual structure that aligns with some of these foldon regions, while other regions do not share the same properties. In particular, the data from all the experiments show a beta sheet propensity in the N-terminal region while the foldon unit is helical. However, the data shows helical propensity in the 60's region which agrees well with the foldon unit described by Krishna et al. The data presented here also shows a consistent region around residue 75 that shows no helical propensity. This corresponds well to the described Ω-loop foldon unit, found at residues 71-85. Despite not having full agreement with the ordered foldon units, Cyt*c* K54H in the denatured state shows several areas of secondary structure propensity, which can be related to the tendency for some proteins to incorporate non-native energy minimizing partially folded forms during the conformational search to the native state.[28; 29; 30] The data presented here may contribute to the identification of residues and structural behavior leading to better understanding of early folding events.

Conclusions

As evident from the SSP score analysis, regions of residual structure exist in the conditions examined and support the current understanding of residual structure. In consideration of the progression of events to a fully-folded form of the protein, the clearly defined region of helical propensity from residues 54 to 67 in 3 M gdnHCl at pH 3.6 is likely a very important element as an SSP analysis of chemical shifts obtained for native Cyt*c* shows a strikingly similar pattern in this region. Further analysis of Cyt*c* K54H will include NMR relaxation experiments to more fully investigate any compact regions of the protein in the denatured state and to detect nonlocal residual structure.

References

1. Daggett, V. & Fersht, A. R. (2003). Is there a unifying mehcanism for protein fodling? *Trends in Biochemical Sciences* **25**, 18-25.
2. Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annual Review of Biophysics and Biomolecular Structure* **29**, 213-238.
3. Tanford, C. (1968). Protein denaturation. *Advances in Protein Chemistry* **23**, 121-282.
4. Neri, D., Billeter, M., Wider, G. & Wuthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science (New York, N.Y.)* **257**, 1559-1563.
5. Neri, D., Wider, G. & Wuthrich, K. (1992). Complete 15N and 1H NMR assignments for the amino-terminal domain of the phage 434 repressor in the urea-unfolded form. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 4397-401.
6. Zhang, O. & Forman-Kay, J. D. (1997). NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. *Biochemistry* **36**, 3959-3970.
7. Lee, J. C., Engman, K. C., Tezcan, F. A., Gray, H. B. & Winkler, J. R. (2002). Structural features of cytochrome *c*' folding intermediates revealed by fluorescence energy-transfer kinetics. *Proceedings of the National Academy of Sciences, USA* **99**, 14778-14782.
8. Konuma, T., Kimura, T., Matsumoto, S., Goto, Y., Fujisawa, T., Fersht, A. R. & Takahashi, S. (2011). Time-resolved small-angle X-ray scattering study of the folding dynamics of barnase. *Journal of Molecular Biology* **405**, 1284-1294.
9. Hammack, B. N., Smith, C. R. & Bowler, B. E. (2001). Denatured state thermodynamics: residual structure, chain stiffness and scaling factors. *Journal of Molecular Biology* **311**, 1091-1104.
10. 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. *Journal of Molecular Biology* **268**, 816-821.
11. Godbole, S., Hammack, B. & Bowler, B. E. (2000). Measuring denatured state energetics: deviations from random coil behavior and implications for the folding of iso-1-cytochrome *c*. *Journal of Molecular Biology* **296**, 217-228.
12. Wandschneider, E. & Bowler, B. E. (2004). Conformational properties of the iso-1-cytochrome *c* denatured state: dependence on guanidine hydrochloride concentration. *Journal of Molecular Biology* **339**, 185-197.
13. Klimov, D. K. & Thirumalai, D. (2002). Stiffness of the distal loop restricts the structural heterogeneity of the transition state ensemble in SH3 domains. *Journal of Molecular Biology* **315**, 721-737.
14. Kurchan, E., Roder, H. & Bowler, B. E. (2005). Kinetics of loop formation and breakage in the denatured state of iso-1-cytochrome *c*. *Journal of Molecular Biology* **353**, 730-743.
15. Cavalli, A., Salvatella, X., Dobson, C.M., & Vendruscolo, M. (2007). Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9615-9620.
16. Bowler, B. E., May, K., Zaragoza, T., York, P., Dong, A., & Caughey, W. S. (1993). Destabilizing Effects of Replacing a Surface Lysine of Cytochrome *c* with Aromatic Amino Acids: Implications for the Denatured State. *Biochemistry* **32**, 183-190.
17. Margoliash, E., & Frohwirt, N. (1959). Spectrum of horse-heart cytochrome c. *Biochemistry* **71**, 570-572.
18. Cavanagh, J., Fairbrother, W.J., Palmer, A.G., & Skelton, N.J. . (1995). *Protein NMR spectroscopy: Principles and practice.*, Academic Press, San Diego CA.

19. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277-293.
20. Goddard, T. D., Kneller, D.G. *SPARKY 3*, University of California, San Francisco, CA.
21. Hammack, B., Godbole, S., & Bowler, B.E. (1998). Cytochrome c folding traps are not due solely to histidine-heme ligation: direct demonstration of a role for N-terminal amino group-heme ligation. *Journal of Molecular Biology* **275**, 719-724.
22. Pollock, W. B., Rosell, F. I., Twitchett, M. B., Dumont, M. E., & Mauk, A. G. (1998). Bacterial expression of a mitochondrial cytochrome c. Trimethylation of lys72 in yeast iso-1-cytochrome c and the alkaline conformational transition. *Biochemistry* **37**, 6124-6134.
23. Wandschneider, E., Hammack, B. N., & Bolwer, B. E. . (2003). Evaluation of Cooperative Interactions between Substructures of Iso-1-Cytochrome c Using Double Mutant Cycles. *Biochemistry* **42**, 10659-10666.
24. Marsh, J. A., Singh, V.K., Jia, Z., & Forman-Kay, J.D. (2006). Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Science* **15**, 2795-2804.
25. Dyson, H. J., & Beattie, J. K. (1982). Spin state and unfolding equilibria of ferricytochrome c in acidic solutions. *Journal of Biological Chemistry* **257**, 2267-73.
26. Dill, K. A. (1999). Polymer principles and protein folding. *Protein Science* **8**, 1166-1180.
27. Jeng, M. F., & Englander, S. W. (1991). Stable Submolecular Folding Units in a Non-compact Form of Cytochrome c. *Journal of Molecular Biology* **221**, 1045-1061.
28. Feng, H., Zhou, Z. & Bai, Y. (2005). A protein folding pathway with multiple folding intermediates at atomic resolution. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5026-5031.
29. Nishimura, C., Dyson, H. J. & Wright, P. E. (2006). Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *Journal of Molecular Biology* **355**, 139-156.
30. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nature Structural Biology* **9**, 209-216.

Chapter 2 Appendix: Chemical Shifts

Parentheses denote chemical shifts seen by previous residue data only (from CBCACONH, HNCACO, and HNCO pulse sequences) and without direct connectivity data from the HNCACB pulse sequence.

3 M gdnHCl at pH 3.6

position		HSQC (^{15}N)	HSQC (^{1}H)	Cα	Cβ	Cγ	Cδ	Cε	COi
-5	P			(62.54)	(32.84)	32.77	26.7	49.78	(172.4)
-4	E	121.1	8.765	56.5	29.68				175.6
-3	F			(57.85)	(40.08)				(175.7)
-2	K	124.4	8.405	56.3	33.48	24.91	29.22	42.36	176.2
-1	A	126.5	8.441	53.07	19.74				178.4
1	G	109.4	8.530	45.71					174.4
2	S	116.1	8.278	58.51	64.35				174.5
3	A	126.6	8.512	52.89	19.82				177.9
4	K								
5	K			(56.95)	(33.45)	29.34	24.96		(177.4)
6	G	111.1	8.624	45.65					174.1
7	A	124.4	8.394	52.93	20.18				178.2
8	T								
9	L								
10	F								
11	K								
12	T								
13	R								
14	C								
15	L								
16	Q								
17	C								
18	H								
19	T								
20	V								
21	E								
22	K			(56.95)	(33.74)	29.29	24.97		(177.4)
23	G	111.1	8.564	45.65					174.1
24	G								
25	P								
26	N			(53.24)	(39.13)				(175.1)
27	K	122.8	8.437	56.45	33.36				176.3?
28	V			(62.09)	(33.2)				(176.3)
29	G	112.7	8.179	44.85					171.8

30	P								
31	N			(58.83)	(39.7)				(175.5)
32	L	122.4	8.206	55.75	42.55	25.25	23.19?		177.4?
33	N	118.8	8.382	53.65	39.43				175.9
34	G	109.6	8.286	45.85					174.3
35	I								174?
36	F			(57.61)	(39.77)				(176.6)
37	G	110.6	8.294	45.7					174.2
38	R			56.65	31.17	43.77	31.06	27.41	176.8
39	Q	121.9	8.62	56.36	29.71	34.2			176.2
40	S	117.3	8.424	58.74	64.18				175.2
41	G	111.5	8.546	45.78					174.3
42	Q	120.4	8.321	56.11	29.96	smear 25, 29			176
43	A	125.6	8.452	52.95	19.61				176
44	E	119.9	8.384	56.23	29.46	33.74			176.8
45	G	110.3	8.395	45.52					173.9
46	Y			58.34	(39.3)				174?
47	S	117.8	8.269	58.23	64.05				174.3
48	Y	123.1	8.244	(58.67)	38.94?				176.3
49	T	115.1	8.033	62.1	69.97				174.4
50	D			(53.89)	(39.84)				(175.4)
51	A	124.3	8.151	53.03	19.78				177.6
52	I	120.4	8.069	61.36	38.81	17.8?			176.2
53	I	125.0	8.142	61.18	38.89	27.48	17.65		176.2
54	H	123.4	8.620	55.35	29.28		25.03?		174.35
55	K	123.3	8.468	56.88?	33.51?	29.25	24.95	42.42	176.2
56	N	120.6	8.592	53.51	39.13				175.3
57	V	120.8	8.135	62.69	32.94	20.95			176.2
58	L	124.9	8.248	55.57	42.21	24.94	23.48		177.4
59	W	121.6	8.049	57.76	29.84				176.3
60	D	120.9	8.242	53.63	39.69				175.7?
61	E	120.7	8.171	56.63	29.2	33.6			176.2
62	N	119.0	8.430	53.74	39.25				175.2
63	N	119.1	8.363	53.7	39.16				175.4
64	M			56.22	32.96	38.64?	36.49?	34.67?	176.7
65	S	116.7	8.373	59.14	63.81				174.9
66	E	122.4	8.387	56.7?	29.25				175.9?
67	Y	120.4	8.141	58.1	39.01				175.9
68	L	123.3	8.173	55.41	42.67	25.23	23.53		177.4
69	T	114.3	8.103	61.94	70.14	21.83			174.2
70	N	122.0	8.531	51.6	39.23				173.2

71	P			(63.63)	(32.52)	27.45	50.91		(177.1)
72	K	121.4	8.395	56.75	33.11	29.17	42.33?		176.7
73	K	122.4	8.228	56.29	33.49	29.2	24.84	42.5	176.05
74	Y			57.92	39.26				175.2
75	I	126.8	8.106	58.28	39.27?				174.1
76	P			(63.61)	(32.61)	27.51	51.34		(177.4)
77	G	110.1	8.560	45.68					174.6
78	T	113.6	8.065	62.05	70.37	21.81			174.8
79	K	124.1	8.506	56.78	33.18	29.21?	24.89?	smear	176.7
80	M			56.3	33.3	38.56?			175.9
81	A	125.8	8.333	52.65	19.85				177.5
82	F	120.3	8.353	58.16	39.91				176.6
83	G	111.3	8.483	45.73					174.7
84	G	109.1	8.038	45.6					174.1
85	L	122.2	8.244	55.57	42.7	27.37	25.04	23.79	177.7
86	K			56.65	33.24		smear		176.8
87	K			56.73	33.33				176.7
88	E	122.5	8.482	56.32	29.51	33.78			176.3
89	K	123.1	8.493	56.82	33.44	29.21	24.88	42.5	176.4
90	D			53.83	39.9				175.8
91	R	121.9		56.83	31	27.25	43.66		176.35
92	N	119.5	8.557	53.72	39.18	41.41?			175.2
93	D	120.1	8.429	53.83	39.62				175.4
94	L			(55.75)	(42.52)	24.6	23.8	20.6?	(177.4)
95	I	122.0	8.164	61.54	38.77	27.67	17.68	12.92	176.4
96	T	118.4	8.151	61.92	70.01	21.73			174.3
97	Y	123.3	8.252	58.2	39.13	42.13?			175.7
98	L	124.2	8.236	55.29	42.82	27.21	25	23.7	177.2
99	K	123.1	8.329	56.71	33.33	29.23	24.94smear		176.8
100	K	123.4	8.398	56.84	33.42	smear 25-29			176.7
101	A	125.8	8.468	52.91	19.76				177.8
102	S	115.7	8.340	58.54	64.16				174.1
103	E	126.3	8.184	57.65	29.81				180.4

3 M gdnHCl pH 6.3

position		HSQC (^{15}N)	HSQC (^{1}H)	Cα	Cβ	Cγ	Cδ	Cε	COi
-5	P								
-4	E			(52.27)	(30.3)	36.39			(176.4)
-3	F	122.0	8.529	58	38.97				175.9

-2	K	124.3	8.416	56.2	33.53	29.23	24.72	42.53	176.1
-1	A	126.2	8.424	52.97	19.7				178.4
1	G	109.1	8.521	45.62	53.01				174.4
2	S	115.8	8.251	58.42	64.23				174.5
3	A	126.3	8.465	52.79	19.77				177.8
4	K	121.0	8.348	56.57	33.23	29.23	24.91	42.67	176.7?
5	K	123.0	8.394	56.7	33.23	24.89	29.11	42.5	177.1
6	G	110.7	8.477	45.4					173.8
7	A	124.0	8.188	52.64	19.94				177.9
8	T	114.6	8.127	62.02	69.59				174.1
9	L								
10	F								
11	K								
12	T								
13	R								
14	C								
15	L								
16	Q								
17	C								
18	H								
19	T								
20	V								
21	E								
22	K	123.5	8.797	(57.17)	(33.43)	25.05	29.4	42.98	(177.3)
23	G	110.7	8.767	45.57					174.3
24	G	109.3	8.391	45.62					172.2
25	P			(63.61)	(32.65)	27.36	50.09		(175.2)
26	N	118.9	8.693	53.58	39.18				175.3
27	K	122.0	8.33	56.47	33.44	29.23	24.91	42.67	176.4
28	V	120.5	8.241	62.36	33.33	21.04			176.3
29	G	112.4	8.414	44.79					172
30	P			(63.54)	(32.59)	27.3	50.12		(177.0)
31	N	119.2	8.716	53.52	39.01				175.6
32	L	123.4	8.488	55.46	42.3	23.27	25.27		177.3
33	N	118.7	8.548	(39.3)					175.9
34	G	109.3	8.343	45.58					174.3
35	I	119.7	7.909	61.77	38.71				176.1
36	F			(57.93)	(39.3)				(176.4)
37	G	110.1	8.269	45.57					174.2
38	R	121.0	8.287	(56.71)	(31.1)	27.39	43.71		(176.7)
39	Q	121.6	8.576	56.3	29.51	34.17			176.2

40	S	117.0	8.378	58.64	64.09				175.1
41	G	111.3	8.548	45.7					174.2
42	Q	120	8.279	56.55?	31.07				176.8?
43	A								
44	E	120.1	8.421	(53.84)	(39.3)				(176.9)
45	G	109.7	8.34	45.39					173.8
46	Y	120.3	8.061	58.14	39.3				175.9
47	S	117.8	8.31	58.1	64.09				174.4
48	Y			(58.87)	(38.8)				(176.4)
49	T	114.4	8.043	62.6	69.75				178.4
50	D								
51	A								
52	I								
53	I								
54	H								
55	K			(56.23)	(31.19)				(176.4)
56	N	117.9	8.016	53.68	39.3				175.2
57	V	119.7	8.437	62.58	33.27				175.9?
58	L								
59	W			(57.57)	(29.98)				(175.9)
60	D	121.7	8.289	54.16	41.6				176.3
61	E	121.3	8.348	(57.25)	(30.24)	36.53			(176.5)
62	N	118.7	8.493	53.78?	39.28				175.1
63	N	119.1	8.353	53.68	39.09				175.3
64	M	120.5	8.361	56	32.95				176.6
65	S	116.6	8.328	59.14	63.85				174.9
66	E	122.9	8.513	57ov	30.38				176.3
67	Y	120.1	8.124						
68	L	123.0	8.155	(55.33)	(42.58)	25.23	23.42		(177.4)
69	T	114.1	8.091	61.92	70.11	21.8			174.2
70	N	121.7	8.526	51.61	39.18				173.2
71	P			(63.49)	(32.59)	27.46	51.09		(177.1)
72	K	121.3	8.401	56.67	32.96	29.16	25.12	42.6?	176.7?
73	K	122.3	8.234	56.58	33.48				176.1
74	Y								
75	I								
76	P			(63.54)	(32.48)	27.54	51.32		(177.4)
77	G	109.9	8.569	45.61					174.5
78	T	113.4	8.071	62	70.32				174.8
79	K	123.9	8.505	56.74	33.14				176.7
80	M								

81	A								
82	F	120.1	8.358	(58.23)	(39.8)				(176.6)
83	G	111.1	8.493	45.69					174.6
84	G	108.9	8.041	45.52					174.1
85	L	122.1	8.262	55.47	42.7	25.2	23.74	27.45	177.6
86	K	122.9	8.46	56.69	33.25				176.7
87	K	123.2	8.444	(56.73)	(33.44)	29.19	24.91	42.49	(176.8)
88	E	122.9	8.566	56.76	30.6	36.5			176.7
89	K	122.5	8.477	56.77	33.44	29.21	2989	42.44	176.5
90	D	121.9	8.526	54.55	41.67				176.4
91	R	121.5	8.388						
92	N	119.9	8.645	(53.72)	(39.19)				(175.2)
93	D	120.8	8.485	54.46	41.32				176.2
94	L	121.9	8.17						
95	I	121.7	8.17	(61.62)	(38.68)	27.78	17.75	12.91	(176.5)
96	T	118.3	8.165	61.95	69.95	21.67			174.3
97	Y	123.1	8.259	58.12	39.13				175.7
98	L								
99	K								
100	K	123.4	8.442	(56.6)	(33.48)	29.27	24.94		(176.6)
101	A	125.8	8.488	52.89	19.79				177.8
102	S	115.5	8.335	58.43	64.18				173.9
103	E	127.8	8.145	58.37	31.16				181.5

6 M gdnHCl pH 3.6

position		HSQC (^{15}N)	HSQC (^{1}H)	Cα	Cβ	Cγ	Cδ	Cε	COi
-5	P			(62.73)	(32.85)	49.88	26.97		(172.2)
-4	E	121.5	8.364	56.7	29.82				175.6
-3	F			(57.63)	(40.18)				175.8
-2	K	124.3	8.305	56.6	33.49	29.55	24.99	42.7	176.2
-1	A	126.4	8.307	53.1	19.76				178.3
1	G	109.5	8.393	45.79					174.3
2	S	116.2	8.159	58.54	64.63				174.5
3	A	126.7	8.394	52.9	19.8				177.8
4	K	121.5	8.253	56.8?	33.33				174.2
5	K			(57.13)	(33.55)	29.54	25.02		(177.3)
6	G	111.2	8.466	45.67					173.9
7	A			(52.89)	(20.09)				178.1

8	T	114.7	8.156	62.23	70.37				174.5
9	L			(55.83?)	(42.35)	25	23.51		(177.2)
10	F	119.1	8.343	56.25?	?				175.9
11	K								
12	T								
13	R								
14	C								
15	L								
16	Q								
17	C								
18	H								
19	T								
20	V								
21	E								
22	K								
23	G			(45.47)					172.1
24	G	109.5	8.187	45.26					172.1
25	P			(63.41)	(32.85)				(176.8)
26	N	119.7	8.468	?	39.27				175?
27	K	120.8	8.284	56.34	33.06				176.6
28	V			(62.4)	(33.34)		21.19		(176.3)
29	G	112.8	8.192	45.02					171.8
30	P			(63.41)	(32.6)	27.27			(176.9)
31	N	119.8	8.530	53.54	39.16				175.3
32	L	122.6	8.108	55.58	42.42				177.2
33	N			(53.71)	(39.55)				(175.8)
34	G	109.9	8.231	45.96					174.1
35	I								
36	F			(58.01)	(39.91)				(176.5)
37	G	110.9	8.232	45.73					174.2
38	R	120.5	8.216	56.75	31.13	27.59	22.01	44	176.8
39	Q	122.1	8.519	56.46	29.72	34.44			176.2
40	S	117.4	8.303	58.81	64.45				175.1
41	G	111.6	8.428	45.83					174.2
42	Q			(56.18)	(30.08)	34			176.0
43	A	125.7	8.338	52.98	19.6				177.7
44	E	120.2	8.280	56.4	29.63	34.34			176.7
45	G	110.5	8.281	45.59					173.8
46	Y			(58.36)	(39.38)				176.0
47	S	118.0	8.211	58.35	64.32				174.3
48	Y	123.3	8.176	58?	39.13				176.2

49	T	115.1	7.975	62.07	70.39	21.89			174.3
50	D			(54.02)	(40.7)				175.5
51	A	124.5	8.087	53.01	19.84				177.5
52	I	125.5	8.077	(61.27)	(38.82)	27	17.8		(176.1)
53	I	125.5	8.081	61.22	38.8	17	27	12	176.1
54	H	123.9	8.546	55.46	29.25				174.3
55	K								
56	N								
57	V			(62.74)	(32.99)	21			176.0
58	L	125.5	8.174	55.55	42.42	25.03	24		177.3
59	W	122.1	7.982	57.65	29.95				176.2
60	D								
61	E			(56.63)	(29.40)	34.22			(176.0)
62	N	119.3	8.364	53.69	39.34				175.1
63	N	119.5	8.304	53.8	39.28				175.4
64	M			(56.2)	(33.1)				
65	S	117.1	8.270	59.04	64.14				174.8
66	E	122.8	8.324	56.54	29.45				176?
67	Y			(58.07)	(39.20)				(175.8)
68	L	124.0	8.170	55.51	42.71				177.4
69	T	114.5	8.018	61.96	70.57				174.2
70	N								
71	P								
72	K								
73	K			(56.69)	(33.67)	29	24.9		(176.1)
74	Y	123.1	8.229	58.11	39.41				175.2
75	I	126.6	8.018	58.38	39.16				174.2
76	P			(63.66)	(32.58)	27	51		(177.4)
77	G	110.2	8.432	45.75					174.5
78	T	113.6	7.953	62.06	70.63	22			174.8
79	K	124.3	8.373	56.84	33.33				176.6
80	M			(55.84)	(33.5)	32.6			(175.9)
81	A	126.0	8.231	52.69	19.84				177.4
82	F	120.4	8.247	58.21	40.04				176.5
83	G	111.3	8.386	45.73					174.6
84	G	109.3	7.958	45.65					174.6
85	L	122.4	8.137	55.64	42.78				177.6
86	K								
87	K			(56.75)	(33.40)	42.61	29	24	(176.7)
88	E	123.0	8.410	56.49	29.59				176.2
89	K			(56.77)	(33.50)	29.5	35.01		(176.4)

90	D	121.8	8.469	54.09	40.38				175.9
91	R			(56.83)	(31.13)	27.56			(176.2)
92	N	119.8	8.448	53.63	39.27				175.1
93	D	120.6	83.45	53.96	40.14				175.6
94	L	122.6	8.112	55.72	42.66	24	23		177.3
95	I	122.5	8.086	61.55	38.8	28	17	13.43	176.3
96	T	119.0	8.067	61.92	70.39	21.9			174.2
97	Y	123.7	8.197	58.15	39.27				175.7
98	L	124.5	8.175	55.39	42.49	24	27		177.2
99	K	123.6	8.267	56.79	33.33				176.7
100	K	123.7	8.327	56.8	33.35	42.71	29.5	25.08	176.6
101	A	126.0	8.364	52.97	19.79				177.7
102	S	116.0	8.235	58.61	64.44				174.1
103	E	127.0	8.099	58.16	29.79				180.8

6 M gdnHCl pH 6.3

position		HSQC (^{15}N)	HSQC (^{1}H)	Cα	Cβ	Cγ	Cδ	Cε	COi
-5	P								
-4	E								
-3	F			(57.89)	(40.27)				(175.8)
-2	K	124.1	8.322	56.44	33.65				176.2
-1	A	126.2	8.303	53.11	20.06				178.4
1	G	109.3	8.391	45.89					
2	S	116.0	8.147	58.51	64.49				174.5
3	A	126.4	8.369	52.88	20.13				177.8
4	K								
5	K			(57.12)	(33.48)				(177.3)
6	G	110.9	8.365	45.58					173.8
7	A	124.2	8.100	52.9	20?				178
8	T								
9	L								
10	F								
11	K								
12	T								
13	R								
14	C								
15	L								
16	Q								
17	C								

18	H								
19	T								
20	V								
21	E								
22	K								
23	G								
24	G	109.3	8.237	45.82					?
25	P			(63.59)	(32.72)	27.5	50.22?		(177.0)
26	N	119.5	8.539	53.78	39.29				175.5
27	K	120.8	8.288	55.98	33.14				176.6
28	V	120.8	8.099	62.45	33?	21.06			176.4
29	G	112.6	8.268	45.07					172
30	P								
31	N								
32	L			(55.81)	(42.39)				(177.3)
33	N	118.9	8.377	53.9	39.58				175.8
34	G	109.6	8.252	45.97					174.2
35	I	120.0	7.821	61.9	39.05				176.2
36	F			(58.17)	(40.02)				(176.5)
37	G	110.5	8.235	45.87					174.2
38	R	120.2	8.197	56.29	30.14				176
39	Q			(56.43)	(29.77)				(175.1?)
40	S	117.1	8.270	58.26	64.1?				176.3
41	G	111.4	8.422	45.9					174.3
42	Q			(56.24)	(30.00)	34.41			(176.0)
43	A	125.4	8.289	52.99	19.97				177.7
44	E	120.4	8.307	56.91	30.92				177
45	G	110.0	8.232	45.56					173.8
46	Y	120.5	7.988	58.35	39.52				176
47	S	117.7	8.176						174.3
48	Y			(58.76)	(39.03)				(176.4)
49	T	114.6	7.915		69.83?				174.5
50	D								
51	A								
52	I			(58.11)	(39.41)				(175.2)
53	I	126.4	8.02	58.31	39.34				174.3
54	H								
55	K								
56	N								
57	V								
58	L								

59	W			(57.55)	(30.34)				(176.1)
60	D	122.1	8.231	54.42	41.82				176.4
61	E								
62	N			(57.20)	(30.41				(176.5)
63	N	119.1	8.416	53.68	39.56				175.2
64	M			(56.20)	(33.15)				(176.6)
65	S	117.0	8.256	59. 27	64.1?				174.8
66	E	123.3	8.420	57.24	30.71				176.3
67	Y	120.8	8.060	58.11	39.28				175.9
68	L	123.8	8.169	55.42	42.88				177.4
69	T	114.3	8.009	61.93	70.5				174.3
70	N								173.3
71	P			(63.63)	(32.64)	27.82			(177.1)
72	K	121.8	8.288	(56.86)	33.16				(176.7)
73	K	122.8	8.186	56.76	33.75	42.89			176.2
74	Y	122.9	8.233	58.04	39.45				175.2
75	I	126.5	8.008	58.31	39.34				174.3
76	P			(63.80)	(32.78)	27.78			(174.6)
77	G	110.0	8.438	45.84					177.5
78	T	113.5	7.959	62.05	70.67				174.9
79	K	124.1	8.377	57.07	33.52	29.54	25.34	42.99	176.7
80	M	122.2	8.335	55.9	33.6				176
81	A	125.8	8.234	52.78	20.15				177.5
82	F	120.2	8.252	58.62	40.2				176.6
83	G	111.2	8.394	45.87					174.7
84	G	109.2	7.963	45.77					174.1
85	L	122.3	8.150	55.66	42.92				177.7
86	K	123.2	8.357	56.87	33.48	29.58	42.88		176.8
87	K	123.4	8.274	56.9	33.5	29.55	25.27	23?	176.9?
88	E	123.2	8.470	56.91	30.88				176.7
89	K	122.7	8.361	56.86	33.81	29.58	25.18	42.8	176.5
90	D	122.3	8.438	54.76	41.99				176.5
91	R	121.5	8.288	56.73	31.33	44.2	27.63		176.4
92	N	120.0	8.520	53.67	39.61				175.3
93	D	121.2	8.395	54.61	41.72				176.2
94	L			(55.65)	(41.47)	25.27	24.23		
95	I	122.3	8.089	61.55	39.1	29.03	17.94	13.24	(176.4)
96	T	118.8	8.077	61.94	70.33				174.3
97	Y	123.6	8.200	58.14	39.35				175.7
98	L	124.4	8.183	55.44	42.98				177.3
99	K			(56.84)	(33.46)				(176.9)

100	K	123.6	8.352	56.88	33.58	29.59	25.2		176.7
101	A	125.9	8.378	53.04	20				177.8
102	S	115.7	8.219	58.64	64.37				174.1
103	E	128.0	8.084	58.65	31.27				181.7