

University of Montana

ScholarWorks at University of Montana

Graduate Student Theses, Dissertations, &
Professional Papers

Graduate School

2009

Thermodynamics and Kinetics of Iso-1-cytochrome c Denatured State

Franco Ollan Tzul
The University of Montana

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

Let us know how access to this document benefits you.

Recommended Citation

Tzul, Franco Ollan, "Thermodynamics and Kinetics of Iso-1-cytochrome c Denatured State" (2009).
Graduate Student Theses, Dissertations, & Professional Papers. 1112.
<https://scholarworks.umt.edu/etd/1112>

This Dissertation is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.

THERMODYNAMICS AND KINETICS OF ISO-1-CYTOCHROME C DENATURED

STATE

By

FRANCO OLLAN TZUL

B. Sc. Chemistry & Biology, Regis University, Denver, CO, USA, 1997

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Chemistry & Biochemistry

The University of Montana
Missoula, MT

March 2009

Approved by:

Perry Brown, Associate Provost for Graduate Education
Graduate School

Dr. J.B. Alexander Ross, Committee Chairperson
Department of Chemistry & Biochemistry

Dr. Bruce E. Bowler, Committee Member
Department of Chemistry & Biochemistry

Dr. Klára Briknarová, Committee Member
Department of Chemistry & Biochemistry

Dr. Michael DeGrandpre, Committee Member
Department of Chemistry & Biochemistry

Dr. Michele A. McGuirl, Committee Member
Division of Biological Sciences

Thermodynamics and Kinetics of Iso-1-cytochrome *c* Denatured State

Chairperson: J. B. Alexander Ross

Various diseases result from protein misfolding. Curing these conditions requires understanding the principles governing folding. Efforts toward understanding how proteins fold have focused on the transition state rather than the earliest folding events. We study these initial events using the assumption that protein folding must involve the formation of the most primitive structure possible – a simple loop. Our laboratory has developed a system of studying simple loops in the denatured state using *c*-type cytochromes. New insights into how the properties of these loops impact the denatured state are outlined in this thesis.

First, studies on a 22-residue loop revealed a previously unreported finding that equilibrium loop formation was not strongly affected by sequence composition. While loop formation rates depended only on sequence composition, loop breakage rates also depended on sequence order. Second, thermodynamic and kinetic studies on homopolymeric inserts in “poor” and “good” solvents revealed that homopolymeric non-foldable protein sequences behave like a random coil. However, heteropolymeric foldable sequences have scaling factors higher than those of a random coil, suggesting the presence of residual structure in denatured proteins. Thus, peptide models with homopolymeric sequences do not adequately describe the nature of foldable sequences. Third, we investigated the kinetics of reversible oligomerization in the denatured state using a P25A yeast iso-1-cytochrome *c* variant. The findings indicated that intermolecular aggregation in a denatured protein is extremely fast – 10^7 - 10^8 $M^{-1}s^{-1}$ and that the P25A mutation strongly affects intermolecular aggregation. This work suggests that equilibrium control of folding *versus* aggregation is advantageous for productive protein folding *in vivo*. Fourth, we use time-resolved FRET to follow compact and extended distributions of a protein under denaturing conditions. Our findings revealed three major populations in the unfolded state when no loop is present whereas only two populations remain when the loop forms. The most extended population is lost upon loop formation showing that simple loop formation dramatically constrains the denatured state.

Thus, thermodynamic and kinetic studies on simple loops using a variety of spectroscopic techniques have enhanced understanding of the initial events of protein folding and the role of the denatured state in modulating protein aggregation.

Acknowledgements

I would first like to thank my advisor Dr. Bruce Bowler for all his patience and guidance throughout my graduate educational development. He has been an inspiration of the type of scientist I would like to mature into and has shown me the dedication and hard work required to be successful in academia. I will forever cherish the kindness he has shown me, both in and out of the lab environment.

I would also like to thank past and present members of the Bowler group, especially Jasmina, Saritha and Eydiejo for their mentoring; as well as Swati and Michael for their added friendship – making it a joy to work in the Bowler group. I would also like to thank Natasa Mateljevic and Nicole Branan for assistance in the preparation of the AcA25H26I52 variant; Sudhindra, Tanveer and Melisa for valuable discussions and insights into life beyond graduate school; Ayesha Sharmin, Reuben Darlington and Dr. Ross for data analysis and guidance with FRET experiments and software training.

I must thank my dear friends, Edmir, Nelda, Michael, Behrang, Joe, Tom, Ricardas, Ryan, Jessica, Kurtis, Becky and Jeff for making my educational experience a wonderful one even during star-crossed situations.

I especially would like to thank my parents, Martha and Franco Sr., for all they have sacrificed of themselves so that I may be able to pursue a tertiary education; my sisters Marsha and Jasmine for their continued support and encouragement; my wife, Gianna, for her selflessness, support, encouragement, sacrifice and patience during my studies.

Finally, to my committee members – Drs. Michele McGuirl, Klára Briknarová, Michael DeGrandpre and J. B. Alexander Ross – thank you for your guidance and time.

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	vii
List of Tables	x
Chapter 1 <i>The Protein Denatured State</i>	
1.1 Proteins and the Polypeptide Chain	1
1.2 Protein Folding Models	3
1.3 Difficulty Studying the Denatured State	8
1.4 Random or Non-Random Denatured State?	10
1.5 Experimental Evidence Supporting Non-random Coil	11
1.6 Denatured State Loop Formation	17
1.7 Inspiration for Dissertation	21
1.8 Goals of Dissertation	22

Chapter 2 *Local Sequence and Composition Effects on a 22-residue Loop with a Single Gly₂Ala Insert at the N-terminus of Iso-1-cytochrome c*

2.1 Introduction	26
2.2 Materials and Methods	33
2.3 Results	46
2.4 Discussion	62
2.5 Conclusion	75

Chapter 3 *Thermodynamics and Kinetics of Loop Formation with Homopolymeric Polyalanine Sequences from the N-terminus of Iso-1-cytochrome c in Poor and Good Solvent Conditions*

3.1 Introduction	76
3.2 Materials and Methods	82
3.3 Results	90
3.4 Discussion	106
3.5 Conclusion	115

Chapter 4 *Competition between Reversible Aggregation and Loop Formation in Denatured Iso-1-cytochrome c*

4.1 Introduction	117
4.2 Materials and Methods	119
4.3 Results	126
4.4 Discussion	151
4.5 Conclusion	162

Chapter 5 Probing Denatured State Distributions Relevant to the Initial Events of Protein Folding using Time-Resolved Förster Resonance Energy Transfer (TR-FRET)

5.1 Introduction	164
5.2 Materials and Methods	168
5.3 Results	176
5.4 Discussion	188
5.5 Conclusion	196
References	199
Appendices	216
Appendix A	216
Appendix B	217
Appendix C	219

List of Figures

<u>FIGURE</u>		<u>PAGE</u>
1.1	Predetermined Pathway-Optional Error Folding Model	5
1.2	Folding Funnel Energy Landscape	6
1.3	Folding Funnel with Intermolecular Aggregation	8
1.4	Schematic Representation of m -values and Denatured State Compaction	12
1.5	Hydrogen Exchange Experimental Scheme	14
1.6	Schematic Representation of Loop Formation in the Denatured State – Heme Edge <i>versus</i> Wrap Around	18
2.1	Schematic Representation of Denatured State His-heme Ligation	29
2.2	B-factor Putty Representation of Disordered Region of Insertions	34
2.3	MALDI-TOF Spectrum of Fragmented Protein	40
2.4	MALDI-TOF Spectrum of EDTA Minimized Protein Fragmentation	41
2.5	MALDI-TOF Spectrum of Optimized Purification with EDTA	42
2.6	Typical CD Denaturation Curve Obtained with AP PiStar 180	46
2.7	Typical CD Denaturation Curve Obtained with AP Chirascan	47
2.8	Representative Equilibrium Loop Formation Titration with DU- 640	51
2.9	pH Titration Reproducibility	54
2.10	pK_a Plotted as a Function of Glycine Percentage Segregated by Local Sequence	55
2.11	Loop Breakage and Formation Rate Constants as a Function of Glycine Percentage Segregated by Local Sequence	61
3.1	Typical CD Denaturation Curve of Poly(Ala) Variants Obtained with AP Chirascan	90
3.2	Typical CD Denaturation Curve of Cytochrome c' Variants Obtained with AP Chirascan	91
3.3	Stability Plots of Cytochrome c' Global Stability and m -values as a Function of Loop Size	93
3.4	Representative Equilibrium Loop Formation Titration of Poly(Ala) Variants with Isoplot	94
3.5	Equilibrium Loop Formation Titration with DU-800 in 3 M and 6 M gdnHCl	95
3.6	Equilibrium obtained Scaling Factors of Poly(Ala) Variants in “Poor” and “Good” Solvents	97
3.7	Representative Equilibrium Loop Formation Titration of Biphasic Cytochrome c' Variants	98

3.8	Representative Equilibrium Loop Formation Titration of Monophasic Cytochrome <i>c'</i> Variants	99
3.9	Comparison of Scaling Factors for Non-Foldable Poly(Ala) and Foldable Cytochrome <i>c'</i> Variants in 3 M gdnHCl	101
3.10	Representative Loop Breakage Trace of Poly(Ala) Variants using Stopped-Flow Kinetics	102
3.11	Plot of Loop Formation Propensities for C-terminal Loops	109
3.12	Plot of Loop Size <i>versus</i> 3 M and 6 M Viscosity Corrected k_b Rate Constants for Poly(Ala) Variants	113
3.13	Plot of Loop Size <i>versus</i> 3 M and 6 M Viscosity Corrected k_f Rate Constants for Poly(Ala) Variants	114
3.14	Scaling Factors obtained from Kinetics Data for Poly(Ala) Variants in “Poor” and “Good” Solvents	115
4.1	CD Denaturation Curve of AcA25H26I52 Variant Obtained with AP Chirascan	126
4.2	Reaction Scheme for Competition of Intramolecular Loop Formation with Intermolecular Aggregation	127
4.3	Plot of $pK_a(\text{obs})$ <i>versus</i> [Cyt c] for AcA25H26I52 and AcH26I52 Variants	128
4.4	k_{obs} <i>versus</i> pH for the Fast Phase of His26-heme Bond Formation and Breakage for the AcH26I52 Variant	132
4.5	Sequential Mixing Data for His-heme Bond Formation of AcH26I52	133
4.6	Plot of Amplitude <i>versus</i> Aging Time for Slow Phase His-heme Bond Formation and Slow and Fast Phase Bond Breakage for the AcH26I52 Variant	135
4.7	Sequential Mixing Data for His-heme Bond Breakage of AcH26I52	137
4.8	A_{398} <i>versus</i> Time for Upward pH Jumps at 2 and 45 μM Final Concentrations for the AcH26I52 Variant	143
4.9	A_{398} <i>versus</i> Time for Downward pH Jumps at 3.75 and 60 μM Final Concentrations for the AcH26I52 Variant	144
4.10	A_{398} <i>versus</i> Time for Upward pH Jumps at 7.5 and 15 μM Final Concentrations for the AcA25H26I52 Variant	145
4.11	A_{398} <i>versus</i> Time for Downward pH Jumps at 7.5 and 45 μM Final Concentrations for the AcA25H26I52 Variant	146
4.12	Concentration Dependence of Fractional Amplitudes for Loop Formation of the AcH26I52 Variant	147
4.13	Concentration Dependence of Fractional Amplitudes for Loop Formation of the AcA25H26I52 Variant	148
4.14	Concentration Dependence of the Slow Phase Rates for AcH26I52 and AcA25H26I52 His-heme Bond Formation	148
4.15	Concentration Dependence of the Fractional Amplitudes for the	150

	Kinetic Phases of AcH26I52 and AcA25H26I52 His-heme Bond Breakage	
4.16	Comparison of Equilibrium versus Kinetic Control for Production of Intramolecular Loops <i>versus</i> Intermolecular His-heme Dimers for the AcH26I52 Variant	160
5.1	Position of Engineered Cysteine Residues for I-AEDANS Attachment in Iso-1-cytochrome <i>c</i>	168
5.2	Cysteine-AEDANS Conjugation Reaction	173
5.3	CD Denaturation Curve of the AcH54I52C37_AEDANS Variant Obtained with AP Chirascan	178
5.4	Equilibrium Loop Formation Titration of the AcH54I52C37_AEDANS Variant in 3 M gdnHCl	178
5.5	TR-FRET trace for the AcH100I52 Variant under Native Conditions in the Presence and Absence of Imidazole	180
5.6	TR-FRET Trace of β -ME-AEDANS Control	182
5.7	A Typical TR-FRET trace of the AcH54I52C37_AEDANS Variant at a given pH – Individual Fits	184
5.8	Lifetime and Corresponding Amplitudes for TR-FRET Data of AcH54I52C37_AEDANS Fitted Individually and Globally	185
5.9	A Typical TR-FRET trace of the AcH54I52C37_AEDANS Variant at a given pH – Global Fit	186
5.10	Plot of FRET Efficiency <i>versus</i> Distance for AEDANS-Heme Pair	190
5.11	Eisenberg Hydrophobicity Plot of AcTMI52 using ProtScale	193
5.12	Schematic Representation of the Effects of Loop Formation in the DSE	196

List of Tables

TABLE		PAGE
2.1	Variant Nomenclature and Sequence Composition of Insert	36
2.2	Stability Parameters for First Generation Gly X Variants	47
2.3	Stability Parameters for Second Generation Gly X Variants	48
2.4	Equilibrium Loop Formation Parameters for First Generation Gly X Variants	52
2.5	Equilibrium Loop Formation Parameters for Second Generation Gly X Variants	56
2.6	Loop Breakage and Formation Rate Constants for Gly X Variants	59
3.1	Stability Parameters for NH5A-X Variants	91
3.2	Stability Parameters for Cytochrome <i>c</i> ' Variants	92
3.3	Equilibrium Loop Formation Parameters for NH5A-X Variants in 3 M and 6 M gdnHCl	96
3.4	Equilibrium Loop Formation Parameters for Cytochrome <i>c</i> ' Variants in 3 M gdnHCl	100
3.5	Loop Breakage Parameters for NH5A-X Variants in 3 M and 6 M gdnHCl	104
3.6	Calculated Loop Formation Parameters for NH5A-X Variants in 3 M and 6 M gdnHCl	105
3.7	Viscosity Corrected Loop Breakage and Formation Parameters for NH5A-X Variants in 6 M gdnHCl	105
4.1	Equilibrium His-heme Bond Formation Parameters for the AcA25H26I52 Variant	128
4.2	Apparent His-heme Intramolecular and Intermolecular Binding Constants for Iso-1-cytochrome <i>c</i> variants	129
4.3	Rate Constants and Amplitudes for pH Jumps for the AcH26I52 Variant	131
4.4	His-heme Bond Formation Sequential Mixing Kinetic Parameters for the AcH26I52 Variant	134
4.5	Concentration Dependent His-heme Bond Formation Sequential Mixing Kinetic Parameters for the AcH26I52 Variant	136
4.6	His-heme Bond Breakage Sequential Mixing Kinetic Parameters for the AcH26I52 Variant	137
4.7	Concentration Dependent Kinetic Parameters for His-heme Bond Formation for the AcH26I52 Variant	139
4.8	Concentration Dependent Kinetic Parameters for His-heme Bond Breakage for the AcH26I52 Variant	140
4.9	Concentration Dependent Kinetic Parameters for His-heme Bond Formation for the AcA25H26I52 Variant	141
4.10	Concentration Dependent Kinetic Parameters for His-heme Bond Breakage for the AcA25H26I52 Variant	142
4.11	Equilibrium Parameters for Intramolecular and Intermolecular	155

	His26-heme Bond Formation	
5.1	TR-FRET Parameters for the AcH100I52 Variant in the Presence and Absence of Imidazole	181
5.2	Individually Fitted TR-FRET Parameters for the AcH54I52C37_AEDANS Variant	183
5.3	Globally Fitted TR-FRET Parameters for the AcH54I52C37_AEDANS Variant	187

CHAPTER 1

THE PROTEIN DENATURED STATE

1.1 Proteins and the polypeptide chain

Proteins are one of the most important molecular constituents of all living things, and hence of life. Their ubiquitous presence in all biochemical and structural aspects of everyday existence make obvious the magnitude of their importance. At an introductory level, a protein is a polymer chain of various amino acids that are linked through condensation polymerization *in vivo*. There are three, sometimes four, different structural conformations that may constitute a protein. Namely, in order of formation, the primary structure, the secondary structure, the tertiary structure and sometimes the quaternary structure. The primary structure simply encompasses the linear sequence (order) of the linked amino acids that make up the polypeptide polymer chain. The secondary structure is a spatial arrangement of the amino acids residues that are *nearby* in the sequence.¹ It entails structural motifs both periodic (repeating pattern of dihedral angles), such as α -helices, 3_{10} helices, β -strands and polyproline II helices, as well as aperiodic (regular but non-repeating pattern of dihedral angles), such as reverse turns/loops, Ω -loops, β -hairpins, and random coils.² After these secondary structural motifs are formed, they then enable contact formation from amino acid residues that are normally *far apart* in the primary sequence. This leads to the formation of disulfide bridges and other enthalpic interactions (electrostatic, van der Waals, hydrogen bonds) of the polypeptide chain creating a tertiary structure thereby culminating in a fully folded, functional protein.

When two polypeptide chains are involved in achieving the fully folded protein, it is said to have a quaternary structure.

It has long been realized, through the works of Christian Anfinsen in the late 1950's and early 1960's, that the protein's primary structure holds the key to the protein folding process,^{3; 4} i. e. it contains the code for folding to the functional native state. Anfinsen's work on ribonuclease A clearly demonstrated that the compact, three-dimensional structure (native state) could be reached from the primary amino acid sequence (denatured state) through purely physicochemical processes without the need for molecular machinery such as molecular chaperones.⁵ The folding process is an intricate one and the exact mechanism has eluded the general scientific community for the past several decades. In the late 1960's Levinthal suggested this folding must occur according to specified pathways and he sought those using computational means for cytochrome *c*, lysozyme and myoglobin.^{6; 7} However, he was unable to determine "the uniqueness of the proposed folding process"⁶ and suggested that computational means, and hence proteins, could not be expected to search randomly through all possible structural conformations for the lowest energy structure of the native fold.⁶ Currently, it is believed that the folded native state of a protein is the main determinant of the folding process^{1; 8} and ultimately provides the stability needed for the native protein to be functional. However, the timescale for this folding process can be very fast - microseconds to milliseconds.⁹⁻¹¹ Thus, the paradox of fast folding from an astronomical number of possible conformations is known as the protein folding problem. Such folding efficiency can only be rationalized by subscribing to a "cooperative" folding mechanism

in which the polypeptide residues use “cumulative selection” sampling to reach the native state.^{1; 7; 12; 13}

1.2 Folding models

To accommodate the cooperative folding notion, several folding models have been proposed to describe the possible pathways a protein’s primary sequence goes through before achieving the native fold. First, the “framework model” describes folding as a stepwise mechanism involving a hierarchical assembly of local elements of secondary structure from the primary sequence but independent from tertiary structure.¹⁴⁻¹⁶ This greatly reduces the conformational search and tertiary structure is attained by diffusion and collision of the local elements of secondary structure whereby favorable amalgamation occurs.¹⁴⁻¹⁶ This model does not place emphasis on the formation of native tertiary contacts directly from the primary structure; thus, secondary structure formation is independent of how the final folded tertiary protein should look. Second, the “nucleation-condensation” model is a modified form of the “nucleation” model and describes folding from more of a helix-coil perspective. The model describes folding *via* the formation of a loosely packed (extended) nucleus, derived from initial helix or sheet “seeding” of **native** secondary structural elements, which becomes more compact in the transition state and is directly responsible for the formation (condensation) of higher order tertiary structure.¹⁷⁻¹⁹ Third, the “hydrophobic collapse” model hypothesizes that the native protein fold is formed from a “molten globule” as a result of the polypeptide chain having a concentrated region of hydrophobic side chains.²⁰⁻²² The molten globule

then quickly rearranges due to the narrowed conformational search leading up to the native fold.

Currently, experimental evidence has led to a more modern and general description of protein folding. Unfortunately, there are still two main competing models. The “predetermined pathway – optional error (PPOE)” model claims that all of a protein population folds by essentially the same stepwise pathway.²³ This single pathway is defined by predetermined cooperative native-like foldon units (intermediates) and how those foldon units interconnect in the final native fold.²³ This model predicts the transition state as a single obligatory step having a few well-defined structures that all protein molecules need to pass through. The intermediates are all downhill from the transition state, thus are “hidden” and are only seen when there is an error in folding (misfold); thus, proteins behave as two-state or multi-state folders, depending on the spectroscopic probe being used.²³⁻²⁵ This model is mainly derived from hydrogen exchange data. Figure 1.1 depicts an energy diagram to convey the main idea of this model.

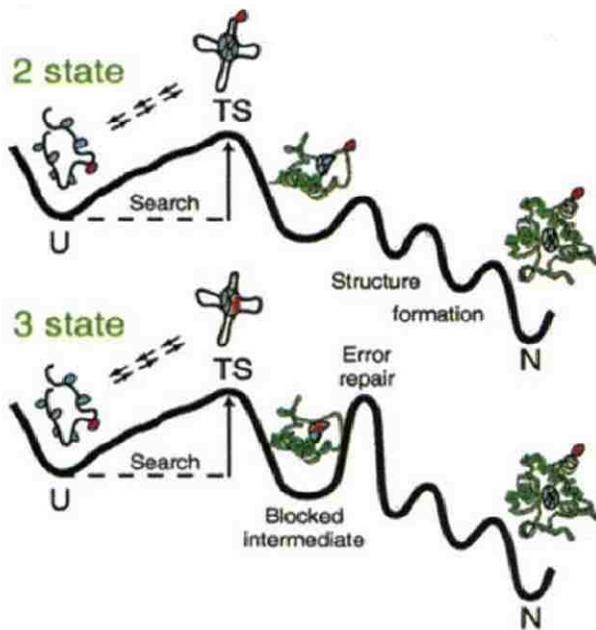


Figure 1.1: A schematic energy profile representation of a two-state and three-state folder having hidden intermediates as described by the “Predetermined Pathway – Optional Error” method.^{23;25} Figure modified from Rumbley *et. al.*, 2001.²⁵ U represents the unfolded state, TS is the transition state and N is the native state.

Secondly, the more popular “folding funnel” model, is derived from statistical mechanics and concepts of polymer physics, rather than those of classical chemical dynamics – hence, is called the ‘new view’.²⁶ This model represents the energy landscape of the protein folding pathway as an energy funnel (see Figures 1.2 and 1.3).

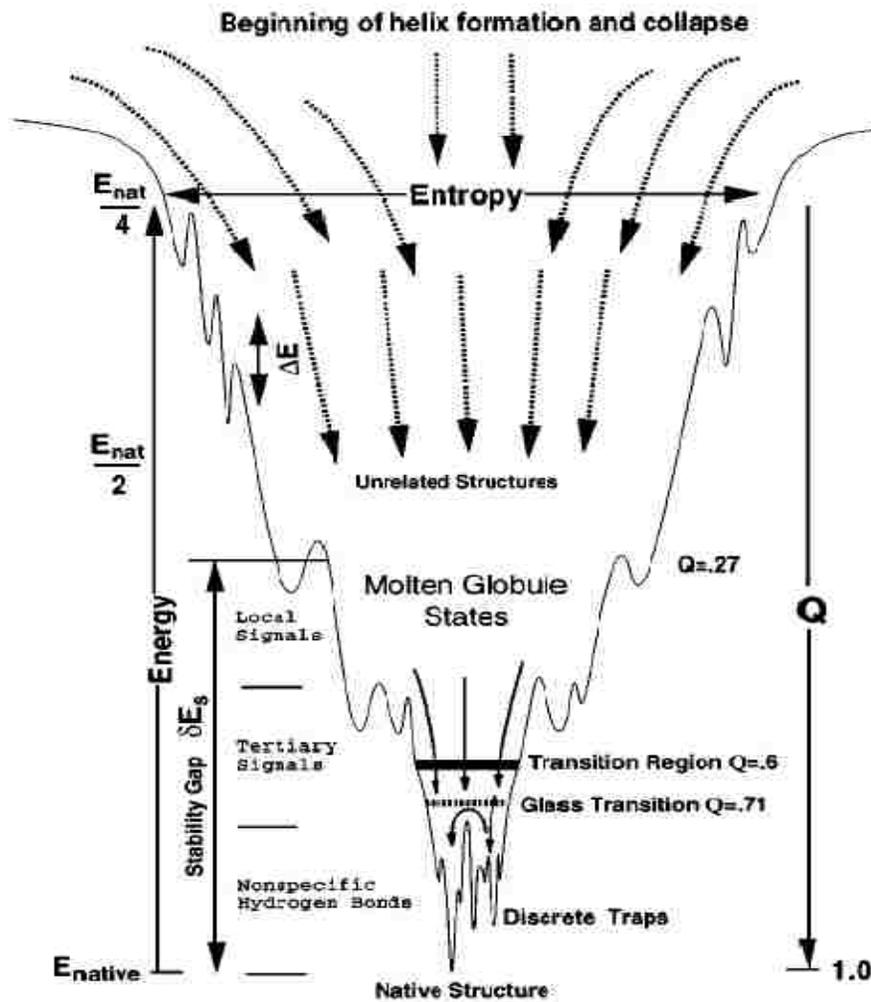


Figure 1.2: Folding Funnel 2-D representation of the rugged energy landscape of protein folding. Q represents the fraction of structures formed with native-like contacts. Diagram taken from Onuchic *et. al.*, 1997.⁵

At the top (rim) of the funnel is a heterogeneous mixture of rapidly-exchanging, high enthalpy, high entropy, polypeptide conformations in the unfolded denatured state ensemble (DSE). There are multiple pathways an unfolded protein can take down a rugged landscape, guided only by the increased enthalpic stabilization of transiently forming native-like structures, toward the “minimally frustrated” global minimum.⁵ In this view the transition state is not viewed as a single obligatory step through which all

molecules have to pass. Rather, the transition state is an ensemble of many different chain conformations.⁵ However, some native contacts will be more probable than others due to the polymeric nature of the chain and the topology of the native state.⁵ This model has been well received because it consolidated many of the ideas from previously mentioned models, as well as provided plausible explanations for protein behavior both *in vitro* and *in vivo*.

Significant advancement in understanding the major steps and criteria used by nature to achieve efficient protein folding has been made. Unfortunately, such advancement has been in the wake of a concerted and multi-dimensional effort to understand some of the many diseases believed to be due to misfolding of proteins. Some of the more prominent protein misfolding diseases are cystic fibrosis²⁷, spongiform encephalopathies (prion diseases such as kuru, Cruetzfeldt-Jakob, Mad Cow, Chronic Wasting and Scrapie) and possibly Parkinson's, Lou-Gehrig's and Alzheimer's diseases.^{1; 8; 26; 28-31} The rugged or bumpy nature of the "folding funnel" model has provided plausible explanations of these protein diseased-states *via* intermolecular interactions as seen in Figure 1.3. Hence, protein folding has and will continue to be a critical research area of past and future science.

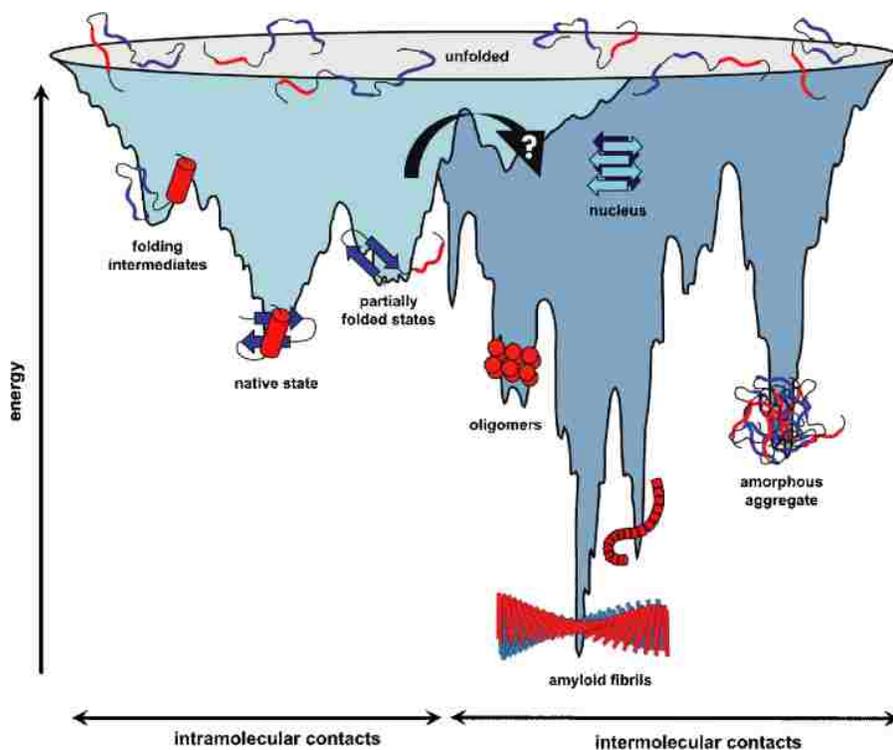


Figure 1.3: 2-D diagram of rugged “folding funnel” with plausible intermolecular interactions leading to various protein diseased-states. Diagram taken from Jahn *et. al.*, 2005.³²

1.3 Difficulty studying the denatured state

Considerable focus has been exerted on understanding the previously mentioned protein misfolding diseases. Much of that focus has been directed toward the native state of the protein. This “top down” approach has been readily embraced due to the well ordered nature of the compact native state. Abundant meaningful structural data can be obtained from X-Ray and NMR studies since the native state directly reports on the various enthalpic networks formed. This detailed information is readily available, thus making it easier to analyze the thermodynamic contributions of weak non-covalent interactions on protein stability. While useful, such native state studies do not contribute to elucidating the initial stages that may underlie a protein misfolding event which leads

to a disease-causing protein. It would seem logical to focus on the sequential steps a protein's polypeptide backbone must navigate to reach the native state.

Unfortunately, this is easier said than done. First, the data that did arise from viscosity and radius of gyration measurements by some research groups, demonstrated that the denatured state was nothing more than a random coil.³³ Under this classification, interactions in a protein's polypeptide chain would be limited to residues that are near each other in the primary sequence.³⁴ This view dominated for many years because the denatured state loses many of the probes that are used to monitor structural features in the native state. In NMR spectroscopy, for example, there is considerable loss of chemical shift dispersion when a protein is denatured.³⁵ This decrease in dispersion of the ^1H and ^{13}C resonances leads to extensive overlap of peaks.³⁵ Also, the increased motions between the residues of the denatured state weaken or eliminate the nuclear Overhauser enhancements (NOE) between protons. Therefore, it is very difficult to define structure under these conditions.³⁵ An added detriment of the denatured state with respect to NMR and other techniques, is that a single parameter cannot be interpreted for a single protein conformation, rather, it would be a value for the dynamically averaged ensemble.³⁵ More importantly, one of the fundamental requirements for characterizing the thermodynamic nature of a structure or "state" is to quantitatively compare it to some reference in order to determine the magnitude of non-covalent interactions and their contributions to the "state" being studied.³⁶ Sadly, in the denatured state these non-covalent interactions are not as well-formed as in the compact well-ordered native state.³⁷ Structural loss in the denatured state is not cooperative i.e. not consistent with a first order thermodynamic transition, is of small magnitude, and thus, is difficult to measure.

Fortunately, sophisticated instrumentation and methodologies have matured considerably allowing thermodynamic^{37; 38} and structural³⁹⁻⁴¹ studies to be successful. Site-directed mutagenesis has opened up new avenues of denatured state research, enabling new questions to be asked and much more data to be generated. Hence, the denatured state has seen a much needed renewed surge of interest in recent years.

1.4 Random or non-random denatured state?

Early work by Tanford in the late 1960's, in moderate to highly denaturing guanidine hydrochloride (gdnHCl) and urea solutions, enabled measurements of the radius of gyration, R_g , of the polypeptide chain.^{33; 42-44} R_g shows a power law dependence on the number of monomers, n , in a polymer chain, $R_g \propto n^\nu$. For a random coil $\nu = 0.6$ in a "good" solvent.⁴⁵ Tanford obtained values of his denatured proteins with $\nu \sim 0.67$.^{33; 42; 43} These works were crucial in the perception of the protein denatured state being viewed as a random coil. However, Tanford did not rule out the possibility that natural polymers such as proteins, could still have certain regions of their backbone incompletely solvated by the denaturant, even in harsh denaturing conditions such as 6 M gdnHCl conditions.^{37; 46} Additionally, other data from approximately the same era as Tanford's work, supported the notion that the denatured state still contained residual structural entities.⁴⁷⁻⁵⁰ Thus, for the most part, it is unclear that the initial wide acceptance of the denatured state being a random coil was justified.

1.5 Experimental evidence of non-random nature of the denatured state

Over the past four decades, since describing the denatured state of proteins as a random coil, tremendous advances have been made in elucidating this elusive state. Structural data from numerous NMR^{35; 39; 51-53} studies have shown that the denatured state is not a random coil of aperiodic amino acid arrangement. Structural data from a few small angle X-ray scattering (SAXS) and Fourier transform infrared spectroscopy (FT-IR) studies on thermally-unfolded ribonuclease A have shown that residual secondary structure still exists with some degree of compaction, even when subjected to increased unfolding conditions.⁵⁴ SAXS studies on Che Y in 5 M gdnHCl have shown that even though the overall protein can be described as a worm-like chain, there are certain regions (~ 25% of the residues) that have residual structure.⁵⁵ Thermodynamic evidence supporting residual structure presence³⁸ and a host of hydrogen exchange data on various proteins demonstrate that the denatured state contains residual structure with stabilities in the range of 0.5 to 2 kcal/mol.⁵⁶⁻⁵⁹

As mentioned previously, the introduction of site-directed mutagenesis (SDM) has opened up a world of possibilities to understanding the denatured state. This technique enables one to probe various regions of the protein at will. One of the pioneering works resulting from exploitation of SDM were studies of *m*-value ($d\Delta G/d[\text{denaturant}]$; proportional to change in solvent exposed surface area upon unfolding) effects on Staphylococcal Nuclease (SNase) variants by Shortle's laboratory.^{60; 61} These denaturant *m*-values studies were the first to demonstrate the complex and intricate behavior of the denatured state. Shortle discovered that the *m*-values differed widely from mutant to mutant. He categorized variants as m^+ and m^-

relative to the m -value of the wild type. Shortle proposed that the changes seen in the m -values for different mutants were a result of changes in degree of compactness (residual structure) present in the denatured state.^{38; 61} A more compact denatured state would have less exposed surface area upon unfolding (m^-), while a less compact denatured state would have more exposed surface area upon unfolding (m^+), relative to the wild type as seen in Figure 1.4.

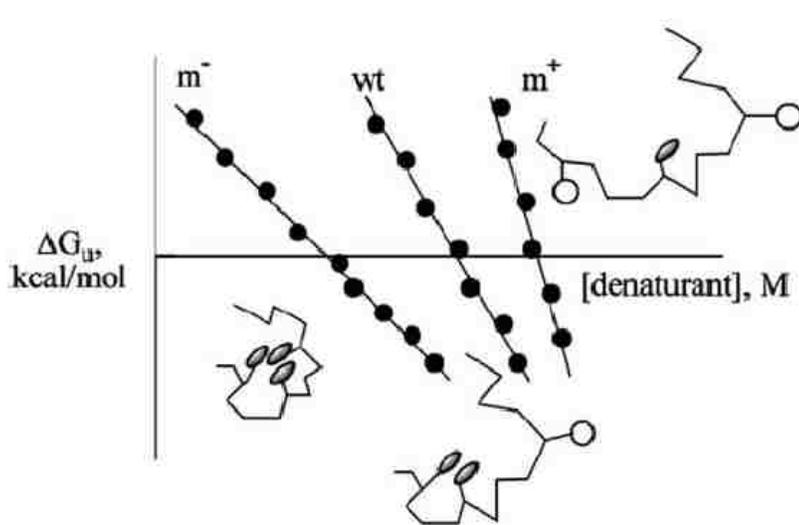


Figure 1.4: $\Delta G_u^{0'}$ versus [denaturant] data for SNase wild type (wt), m^- and m^+ variants. Cartoon representations show how the degree of compactness of the denatured states, from specific mutations, affected the m -values. Diagram taken from Bowler, 2007.³⁸

SDM has also contributed tremendously to probing the electrostatic interactions and their contributions to overall protein stability.^{36; 62-67} An interesting observation from SDM studies on λ Cro, iso-1-cytochrome c and RNase T1, is a phenomena known as the “reverse hydrophobic effect”.^{66; 68-70} It describes altering the stabilization of the denatured state by mutating protein residues natively found on the surface from polar to

hydrophobic residues. Such mutations cause alterations to the compactness of the denatured state.³⁷ Resulting stabilization of the denatured state provides strong evidence that the unfolded state does not contain free energy neutral residues and does, in fact, contain residual structure. However, the inference of residual structure from studies of *m*-values or even changes in heat capacity would benefit from a more direct thermodynamic measurement of residual structure. One technique that can accomplish that is Hydrogen Exchange (HX). How well the amide (NH) groups on the polypeptide backbone are protected from exchange with deuterated solvent can be used to directly evaluate whether hydrophobic clusters exist and in exactly what regions.^{71;72} Application of this technique to the denatured state has not been met with much success largely due to the elimination of the very property (residual structure) that you are trying to monitor *via* the use of denaturants.³⁸ However, the laws of thermodynamics and the Boltzmann distribution dictates that intermediate states as well as unfolded states should exist with the native state conformation, albeit at much lower populations under native conditions.⁷¹ Thus, the presence of residual structure can also be monitored under native conditions. Figure 1.5 below illustrates a typical HX experiment used in such evaluations.

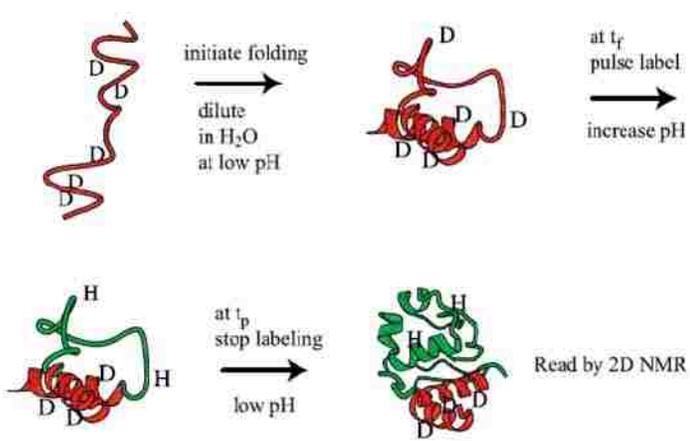
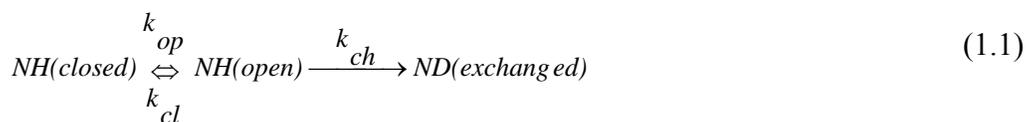


Figure 1.5: Typical native state HX experiment depicting protected regions of a protein that can be identified via 2D NMR. Diagram taken from Krishna *et. al.*, 2004.⁷²

The mechanism for the above hydrogen exchange experiment is given as:



where k_{op} and k_{cl} are the rate constants for unfolding and folding the structure that protects the amide NH, respectively, and k_{ch} is the rate constant for $^1\text{H} \rightarrow ^2\text{H}$ amide exchange. For EX1 hydrogen exchange conditions (favored at high pH), the observed rate constant, k_{ex} , will be independent of pH since it depends only on k_{op} . While for EX2 hydrogen exchange conditions (favored at low pH), k_{ex} will change by a factor of 10 for each unit change in pH since k_{ch} is dominated by base catalysis in the region where measurements are usually made (pH 5 to 8). Therefore, in a protein's native state, $k_{cl} \gg k_{ch}$ (EX2 conditions, low pH & temperature) and the observed $k_{ex}^{EX2} = K_{op}k_{ch}$, where K_{op} is k_{op}/k_{cl} , which is the equilibrium constant for unfolding the structure that protects the

amide NH from exchange.⁷² Therefore, the change in free energy can be evaluated for the amide residue that was exchanged using Eq 1.2 below:

$$\Delta G_{HX} = -RT \ln K_{op} = -RT \ln \left(\frac{k_{ex}^{EX2}}{k_{ch}} \right) \quad (1.2)$$

Using this technique, interesting results in recent years have shown that certain protein residues have ΔG_{HX} that is about 0.5 to 2.0 kcal/mol higher than $\Delta G_u^{0'}$ measured by conventional CD and fluorescence probes.³⁸ This higher free energy has been termed superprotection and these values are in line with expected stabilities conferred by residual structure.³⁸ Thus, native state HX has gained popularity in certain laboratories and has provided direct evidence of residual structure by means of superprotected protein regions, notably those that correspond to β -sheets and to a lesser extent α -helices in the native fold.^{38; 71; 72} Besides HX, another field that has greatly benefitted from SDM is fluorescence spectroscopy. Exploitation of SDM in this field has also provided a wealth of information on the reverse hydrophobic effect⁷³ as well as a picture of the compact and extended distributions that exist in the denatured state.⁷⁴ This information is mainly from time-resolved Förster resonance energy transfer (TR-FRET) and has been widely exploited in the past three to four years. Various distributions of extended and compact conformations exist in the denatured state ensemble (DSE) and the steady-state interconversion of these distributions can be directly monitored using TR-FRET under variable denaturant concentrations.⁷⁴⁻⁷⁶ Actual dynamic interconversions of the various conformations can be monitored in real time with single molecule FRET studies. These powerful techniques are demonstrating that some proteins have hydrophobic regions that

influence the distribution of the DSE toward highly compact conformations,⁷⁵ surprisingly akin to native-like distance distributions. These “supercompacted” conformations have been shown to be robust and persist even in denaturing conditions.^{74;}
⁷⁶ Other less hydrophobic regions of a protein do not seem to possess these “supercompacted” conformations or are minimally present at best and are quickly converted to compact and extended degenerate⁷⁷ denatured state conformations when subjected to denaturing conditions.^{74; 76} There are questions as to the effect denaturant has on the apparent extended and compact distributions seen with increasing [denaturant].⁷⁸ But it is undeniable that the DSE has “supercompacted” conformations as well as less compact conformations and more extended conformations.

These recent findings provide strong support that protein folding is directed by the most stable cluster able to act as a seeding template for the rest of the protein⁷⁹ and not necessarily dictated by the overall global stability of the protein.⁷⁵ Last year some astounding thermodynamic work of phi-value analysis on the Notch Ankyrin repeat protein, using the powerful tool of SDM yet again, tracked the transition state as the protein folds.⁷⁹ This work demonstrated that the pathway a protein takes down the folding funnel can be re-routed by manipulating the most stable repeat unit in the energy landscape.⁷⁹ Though this work is not directly related to residual structure in the denatured state, it serves as a conceptual proof that protein folding is guided by the most stable cluster that is capable of acting as a nucleation point. The wealth of evidence on residual structure presented in this chapter so far has unequivocally shown that residual structure exists in the denatured state and it provides high stability to the denatured state. Thus, residual structure is expected to behave similarly to the individual repeats of Notch

Ankyrin, which serve to limit the conformational search, thereby creating biases in the energy landscape which direct folding pathways.⁷⁹ Besides the growing support for its importance in directing protein folding, it has also been implied that the varying compactness (residual structure) and relative degeneracy of the DSE may have a crucial physiological importance in preventing protein misfolding, hence misfolding diseases.⁷⁷ Therefore, a focused effort on studying residual structure and the denatured state with its heterogenous conformations is required and is expected to be rewarding.

1.6 Denatured state loop formation

In order to pursue denatured state and residual structure studies, our laboratory has developed methods to measure the equilibria and kinetics of formation of simple polypeptide loops under denaturing conditions.^{37; 38} We use variants of the *c*-type cytochrome, yeast iso-1-cytochrome *c*, that have been engineered to contain only a single histidine besides the native heme ligand, His18. Histidines on the N-terminal side of the heme attachment site (Cys14, Cys17 and His18) will form a loop (heme-edge) which includes residues between the engineered histidine and Cys14 (Figure 1.6). Histidines on the C-terminal side of the heme attachment site will form a loop between His18 and the engineered histidine (wrap-around). A simple pH titration allows evaluation of the stability of a given histidine-heme loop, providing an apparent pK_a , $pK_a(\text{obs})$, which is influenced by the local ligand environment. Since a higher proton concentration is required to break a more stable loop, a lower $pK_a(\text{obs})$ indicates a more stable loop. Loop breakage is detected at 398 nm which is the wavelength to which the low spin Fe^{3+} -heme Soret band shifts when a high spin Fe^{3+} -heme forms due to the weak field water

ligand being bound to the Fe^{3+} instead of the previously bound strong field histidine ligand (see Figure 1.6).

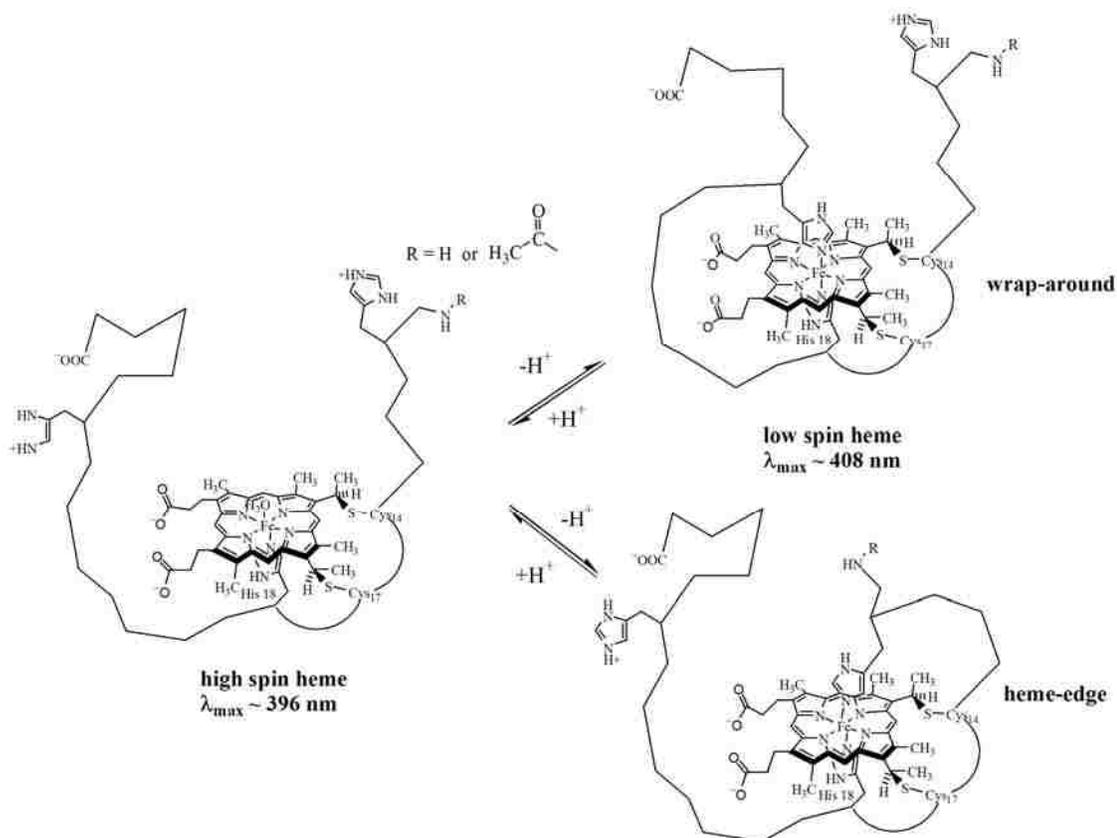


Figure 1.6: Schematic representation of denatured state His-heme ligation from the N-terminal side of the heme of iso-1-cytochrome *c* (heme-edge) as well as from the C-terminal side (wrap-around). Diagram taken from Bowler, 2008.³⁷

Thus, the relative stabilities of different loops can be evaluated easily by a fit of absorbance data at 398 nm plotted against pH, to a rearranged form of the Henderson-Hasselbalch equation (Eq 1.3) to obtain the apparent pK_a , $\text{pK}_a(\text{obs})$, and the number of protons, n , involved in the process,

$$A_{398} = \frac{(A_{LS} + A_{HS} 10^{n(pK_a(obs)-pH)})}{(1 + 10^{n(pK_a(obs)-pH)})} \quad (1.3)$$

where A_{LS} is the absorbance at 398 nm in the low spin form of the heme with histidine bound and A_{HS} is the absorbance of the high spin form of the heme with histidine displaced by water. Typically, the apparent pK_a is lower than the pK_a of free histidine because deprotonation of histidine is coupled to favorable formation of a bond between Fe^{3+} in the heme and the histidine imidazole side-chain. The stability of the His-heme bond is dependent on three main factors. First, chain stiffness can decrease the stability if the polypeptide chain forming the loop is too short or too sterically restrictive, thereby hindering effective His-heme molecular orbital overlap.^{80; 81} These hindrances can be relieved to increase the bond strength by either increasing loop size or decreasing residue bulk. Second, loop entropy decreases the His-heme bond strength as the loop becomes larger. Too long a loop increases the conformational space for the histidine to search, thereby decreasing the probability of making an actual His-heme bond. Third, residual structure formation increases the His-heme bond strength due to the presence of additional stabilizing contributions in the loop, if the structure is induced by loop formation. The summed stability of these contributors is conferred onto the His-heme bond.

We have used this method to evaluate the conformational properties of a denatured protein^{37; 38} including the scaling properties of a polypeptide chain,⁸⁰ the effect of varying denaturing conditions on loop formation,⁸² the effects of local excluded

volume on loop formation,⁸¹ and most recently the effects of sequence composition on loop formation.⁸³

The Jacobson–Stockmayer equation^{84; 85} (Eq 1.4) is typically used to predict loop formation equilibria:

$$\Delta S_{loop} = -v_3 R \ln(n) + R \ln[(3 / 2\pi C_n l^2)^{v_3} V_i] \quad (1.4)$$

where n is the number of monomers in the loop, R is the gas constant, C_n is Flory's characteristic ratio, which is sensitive to chain flexibility,⁸⁶ l is the distance between monomers, V_i is the approach volume of the atoms involved in loop formation and v_3 is the scaling exponent for loop formation. C_n is a ratio of the dimension of the polymer in question to those of a “random flight” chain. Realistic chains are always more expanded than a “random-flight” chain, therefore, C_n will be greater than 1. The degree of stiffness (flexibility) of a polymer chain can therefore be evaluated from the value of C_n . Furthermore, as described above, the stiffness of the polymer chain can be modulated by sequence composition having compact versus bulky residues or by changing the length of the chain.

Another informative component of the Jacobson-Stockmayer equation is the scaling factor, v_3 , which indicates how well the polymer chain of interest behaves relative to a random coil. For a freely jointed random coil, the scaling factor is 1.5.^{37; 84; 85; 87} When excluded volume is taken into account, the value of v_3 increases to a range of 1.8 to 2.4.³⁷

From Eq 1.4, since loop formation is governed mainly by entropy, it can be shown that $\Delta pK_a(\text{obs})$ for a variant X *versus* a variant Y with the same loop size is given by Eq (1.5):

$$\Delta pK_a(\text{obs}) = pK_a(\text{obs})_X - pK_a(\text{obs})_Y = \left(\frac{-v_3}{\ln(10)}\right) \left[\ln\left(\frac{C_{n,Y}}{C_{n,X}}\right)\right] \quad (1.5)$$

If we use $v_3 = 2.1$ (average of theoretical values for a random coil with excluded volume),^{38; 88} we are able to predict the effect of C_n on the expected $\Delta pK_a(\text{obs})$ for the various loop variants we evaluate.

1.7 Inspiration for dissertation work

Much of my research has been based on previous experimental results from our laboratory that were inconclusive or warranted further investigation. Previous work on denatured state loop breakage kinetics demonstrated that an optimally stable loop of 37 residues (AcH54I52 variant) had an extremely slow breakage rate.⁸⁹ No previous data were available to validate the importance of loop breakage in modulating loop stability.⁸⁹ Thus, we investigated loop breakage as a tool to analyze the basis of loop stability (Chapter 2). The same 37 residue loop (AcH54I52) demonstrated unique loop stabilization under various [gdnHCl] as well as a relatively high m -value compared to other loops.^{82; 90} Thus, investigations into the nature of the loop stability and possible residual structure stabilization lead to TR-FRET studies of various loops (Chapter 5). Furthermore, to account for excluded volume effects of heme on equilibrium loop

formation⁸² and kinetic⁸⁹ loop breakage and formation data of loops formed from the C-terminus of iso-1-cytochrome *c*, we engaged in a comparative study of loops formed from the N-terminal side of the heme instead (Chapter 3). This project also served to fill the huge gap of experimental data on the behavior of foldable *versus* non-foldable protein sequences. Finally, a loop forming variant (AcH26I52) demonstrated a significantly concentration dependent $pK_a(\text{obs})$ and had biphasic kinetics of loop formation.⁸² Thus, the nature of these kinetic phases was investigated to determine whether it was due to aggregation or proline isomerization (Chapter 4).

1.8 Goals of dissertation

At the start of my research into the denatured state, the presence of residual structure was well accepted and there was considerable evidence that it is thermodynamically important to the folding process as seen from Section 1.5. Modern computational and experimental data had re-evaluated the nature of the disordered polypeptide chains against a random coil dimension. These evaluations indicated that disordered chains found in the DSE were more expanded than previously predicted from the allowed ϕ and ψ angles of amino acids using the Ramachandran Plot. This discrepancy was attributed computationally to the need for excluded volume effects in describing a polypeptide and experimentally to the existence of expanded structures such as polyproline II structures in the DSE. Conversely, there has been little data on how disordered foldable sequences compared to disordered non-foldable homopolymeric protein sequences. We hypothesized that foldable and non-foldable polymer types would behave differently and sought to demonstrate this possibility. Also, much of the data in

the literature was targeted at defining the transition state for protein folding but not much focus was on the initial events, such as simple loops and how the properties of those loops affected the properties of the denatured state. Thus, the following chapters describe our efforts to contribute to and further elucidate the properties of denatured proteins. We do so using a model that dictates protein folding must go through the formation of the most primitive structure possible – a simple loop. To do so my research focuses on what occurs at the top (rim) of the free energy landscape in the initial stages leading to the polypeptide chain folding down the “folding funnel”.

The first project (Chapter 2) looks at sequence composition and sequence order of an engineered 22-residue loop at the N-terminal side of yeast iso-1-cytochrome *c*. We asked, what are the conformational constraints that enable formation and persistence of a simple loop as the fraction of the flexible amino acid glycine is varied. We observed a previously unreported finding that the equilibrium for loop formation was not strongly affected by sequence composition, although significant compensating changes in rates of loop formation and breakage occur. Furthermore, through kinetic loop breakage studies, we found that for simple loops, sequence composition is less important than local sequence order in maintaining formed contacts; whereas, loop formation depended only on sequence composition.

The second project (Chapter 3) examines the thermodynamic and kinetic behaviour of homopolymeric inserts of yeast iso-1-cytochrome *c* in “poor” and “good” solvents. In particular, we ask the question of how a disordered non-foldable

homopolymer differs from a disordered foldable heteropolymer protein sequence. We discovered that homopolymeric sequences have thermodynamic properties in line with a random coil. However, heteropolymeric sequences, while having scaling exponents similar to a random coil, scatter much more widely about the line that describes random coil scaling. This is highly suggestive of residual structure being present in denatured proteins. It also demonstrates that many of the peptide models for heteropolymeric sequences do not adequately describe the nature of foldable sequences.

The third project (Chapter 4) investigates the kinetics of reversible oligomerization in the denatured state using a P25A yeast iso-1-cytochrome *c* variant. The findings indicate that intermolecular aggregation in a denatured protein is extremely fast – on the order of 10^7 - 10^8 $M^{-1}s^{-1}$. The P25A mutation was found to control the rate of that intermolecular aggregation. This work is suggestive that equilibrium control of the partitioning between folding and aggregation is advantageous for productive protein folding *in vivo*.

The fourth project (Chapter 5) uses fluorescence spectroscopy and time-resolved Förster Resonance Energy Transfer (TR-FRET) to follow compact and extended distributions of simple loops under denaturing conditions. In particular, we hypothesized that formation of a simple loop in the denatured state will lead to compaction of the part of the denatured protein within the loop. Our main fluorescence donor was the extrinsic fluorophore, 1,5 I-AEDANS, which was attached to the protein *via* an engineered cysteine. This work provided insights into the conformations that exist as a primitive

loop closes with increasing pH. The findings show that there are three major distributions in the fully unfolded denatured state ensemble when no loop is present. That distribution changes to two distributions as the loop closes at the same denaturant concentration. The surprising finding is that “supercompacted” structures still exist when there is no stabilizing loop present. This data provides evidence of denatured state heterogeneity in a protein variant under conditions that remove the influences of competing heme ligands which would otherwise result in misligation. This misligation-free environment, at constant denaturant concentration, allows for data interpretation of populations without the influence of changing denaturant concentration producing changes in populations. More investigations with probes at various other positions and farther distances from the FRET acceptor need to be pursued in the future to get a wider view, however.

CHAPTER 2

LOCAL SEQUENCE AND COMPOSITION EFFECTS ON A 22-RESIDUE LOOP WITH A SINGLE GLY_ALA INSERT AT THE N-TERMINUS OF ISO-1-CYTOCHROME C

2.1 Introduction

The denatured state of a protein is the starting point for folding a protein to its native structure, which confers its function. Early studies on denatured proteins indicated that disordered polypeptides had properties consistent with a random coil.³³ This suggests that the denatured state of a protein is devoid of any persistent structure or has minimal structural periodicity.³³ Nowadays, it is widely accepted that protein denatured states can contain structures of intermediate complexity or non-random structures known as residual structure.^{45; 91} Non-random structure results in part from the intrinsic bias of the polypeptide chain toward extended structures such as the polyproline II helix.^{45; 92; 93} Structural data, particularly from NMR studies of denatured proteins, and computational studies demonstrate the presence of persistent secondary structure and hydrophobic clusters under denaturing conditions^{35; 51; 55; 91; 94-98} which are likely important factors in protein folding.^{35; 91} Thermodynamic approaches to characterizing protein denatured states also show strong deviations from the properties expected for a random coil.^{37; 38; 64} Such residual structure can significantly impact the stability of the denatured state.³⁸ In particular, studies on the pH dependence of the stability of the N-terminal domain of ribosomal protein L9 (NTL9)^{63; 65; 99} and staphylococcal nuclease,³⁶ as well as investigations into the effects of surface electrostatics on the stability of RNase T1, RNase Sa and T4 lysozyme^{66; 67; 100} show that ionic interactions within residual structures

can stabilize protein denatured states by 1 to 4 kcal/mol.^{38; 63-67; 99; 100} Recently, these electrostatic interactions were suggested to be the stabilizing forces of initial short range contacts made in the denatured state polypeptide, which guide the chain down the folding funnel.¹⁰¹

These lingering structures in a protein's denatured state can limit the conformational search or "walk across conformational space" towards a thermodynamic minimum (bottom of folding funnel). Recently, work on Notch Ankyrin domains has shown unequivocally that the protein folding pathway is determined by nucleation of the most thermodynamically stable domain(s).⁷⁹ By analogy, stable residual structures present in the denatured state ensemble should dictate early nucleation events, which in turn guide the denatured polypeptide down the 'folding funnel'. The Notch Ankyrin work highlights the importance of studying residual structure in the denatured state and strongly supports its suspected role of limiting the "walk across conformational space" towards a thermodynamic minimum. The presence of residual structure can translate into a direct influence on the kinetics of protein folding,^{63; 100} and in so doing, greatly increases protein folding efficiency,^{7; 67} thereby satisfying Levinthal's Paradox. Unfortunately, the question of how this increased efficiency is accomplished is still not fully understood.

In order to determine the mechanism by which residual structure increases protein folding efficiency, the formation and persistence of such structure must first be assessed. Statistically it has been shown that simple loops are much easier to form as compared to more complex β turns, β sheets and α -helices.¹⁰² Loop formation is necessary in nucleation of more complex structures such as α -helical secondary structure^{103; 104} and

relatively large loops have been implicated in the folding mechanism of some proteins.¹⁰⁵⁻¹⁰⁷ Hence, loop formation is of great interest because contact between two monomers along a polypeptide chain is the most primitive type of structure which forms during protein folding^{102; 108} and can be considered a basic step necessary for protein folding which may set a “speed limit” for folding.^{102; 109; 110}

Consequently, it is necessary to understand the fundamental conformational constraints that act on a disordered polypeptide chain. To probe such constraints, we have developed methods to measure the equilibria and kinetics of formation of simple polypeptide loops under denaturing conditions.^{37; 38} We use variants of the *c*-type cytochrome, yeast iso-1-cytochrome *c*, that have been engineered to contain only a single histidine besides the native heme ligand, His18. Histidines on the N-terminal side of the heme attachment site (Cys14, Cys17 and His18) will form a loop which includes residues between the engineered histidine and Cys14 (Figure 2.1). Histidines on the C-terminal side of the heme attachment site will form a loop between His18 and the engineered histidine. A simple pH titration allows evaluation of the stability of a given histidine-heme loop, providing an apparent pK_a , $pK_a(\text{obs})$. Since a higher proton concentration is required to break a more stable loop, a lower $pK_a(\text{obs})$ indicates a more stable loop. Thus the relative stabilities of different loops can be evaluated easily. We have used this method to evaluate the conformational properties of a denatured protein^{37; 38} including the scaling properties of a polypeptide chain,⁸⁰ the effect of varying denaturing conditions on loop formation,⁸² the effects of local excluded volume on loop formation,⁸¹ and most recently the effects of sequence composition on loop formation.⁸³

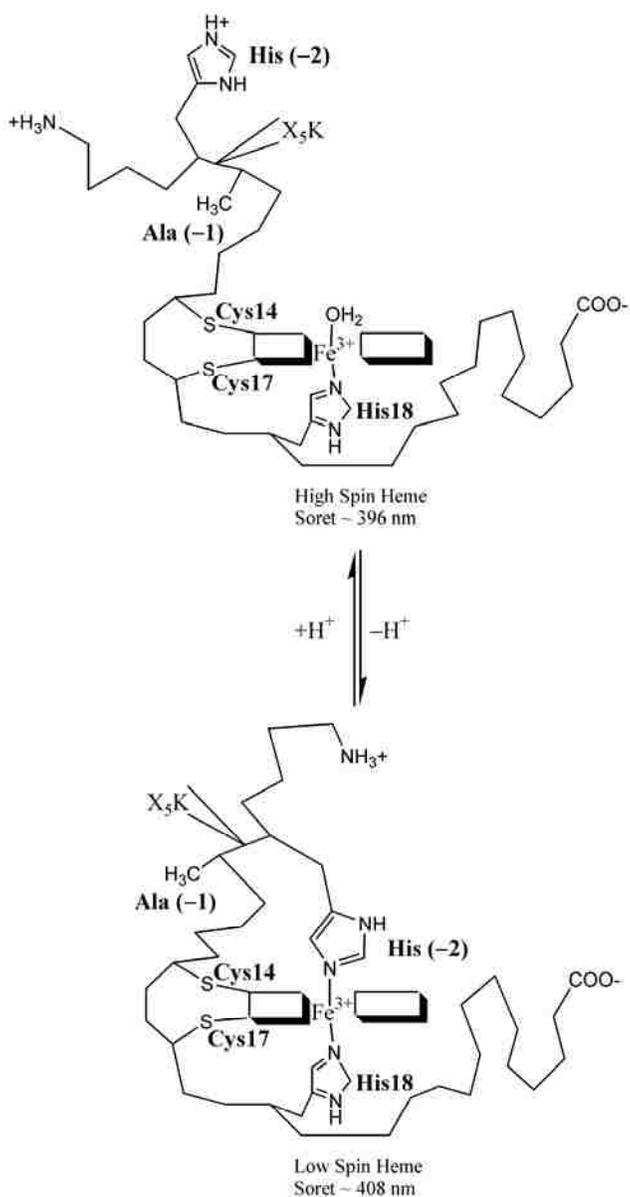


Figure 2.1: Schematic representation of denatured state His-heme ligation from the N-terminal side of the heme of iso-1-cytochrome *c*. Ala/Gly inserts were made following a histidine substituted in place of Lys(-2) for the NH5A and Gly X variants. Ala(-1) terminates each insert sequence.

Studies on the kinetics of loop formation using synthetic peptides have provided key insights into the effects of sequence composition on the conformational constraints that limit the rate of loop formation in a disordered polypeptide.^{102; 108} Except for glycine

and proline, the effects of sequence composition are minimal.^{108; 110; 111} However, intramolecular hydrogen bonds from Thr, Ser, and Gln side chains may stiffen the polypeptide backbone.¹¹¹⁻¹¹⁴

The work presented in this chapter uses a single 22-residue loop formed from the N-terminal side of yeast iso-1-cytochrome *c* to examine the effects of sequence composition and sequence order on the kinetics and thermodynamics of loop formation and breakage in the denatured state. We inserted the sequence AAAAAK in between histidine at position -2 (K(-2)H) and Ala at position -1 of yeast iso-1-cytochrome *c* (Figure 2.1; we use horse cytochrome *c* numbering, thus the 5 amino acids preceding Gly1 of yeast iso-1-cytochrome *c* are designated -5 to -1).⁸³ Under denaturing conditions, His(-2)-heme binding forms a 22 residue loop which acts as a simple test of important factors in polymer theory, enabling insight into the behavior of natural polymers.

First, we investigate flexibility of the main chain due to steric bulk of the R-group and the effect that less conformationally restrictive amino acid residues would have on polypeptide compactness and thus the equilibrium stability of loop formation. A more flexible chain is more compact³³ and loop formation might be expected to be more favorable in a more compact polymer. However, Robinson and Sauer¹¹⁵ suggested that equilibrium stability of loops or linkers between two subunits of a protein is primarily dependent on composition and residue ratios rather than sequence. In particular, a balance between adequate *versus* too much flexibility was essential for optimal stability. To accomplish this first investigation, we progressively changed each of the alanines in the insert to a glycine, producing a set of 22 residue loops where the percent glycine

within the loop varied from 9% to 32%. Polymer theories predict that the average end-to-end distance of a chain will decrease as glycine content increases.⁸⁵ Therefore, the probability of closing a loop should increase with increased glycine content. In fact, we observed an identical $pK_a(\text{obs})$ for equilibrium loop formation in 3 M gdNHCl for the all alanine insert (NH5A variant) and the all glycine insert (Gly 5 variant). The decrease in Flory's characteristic ratio, C_n , for the 22-residues contained in the loop for the Gly 5 variant relative to the NH5A variant predicted a decrease in the $pK_a(\text{obs})$ of ~ 0.4 . Therefore, we subsequently tested three hypotheses that could explain the lack of a decrease in the equilibrium $pK_a(\text{obs})$ with increased glycine content for the Gly 5 variant relative to the NH5A variant. The first hypothesis is that increased glycine content leads to faster rates of loop breakage, counterbalancing the faster rates of loop formation expected for the more flexible glycine sequence. We have measured rates of loop breakage to test this hypothesis.

The second hypothesis is based on NMR studies on short poly-glycine peptides which indicate that the persistence length of glycine is greater than previously thought.¹¹⁶ In this study, it was suggested that the greater than expected persistence length of poly-glycine stretches was either due to local sequence effects or to peptide backbone hydration. The latter is not accounted for in early polymer models. In the early part of this investigation,⁸³ the $pK_a(\text{obs})$ decreased as the first two glycines were inserted and then increased as the remaining glycines were added. The glycines were inserted contiguous to each other in this study (see Table 2.1). This observation raised the possibility that a critical threshold of contiguous glycines is necessary to extend the backbone through local sequence or hydration effects. Similarly, in the NMR studies on

poly-Gly peptides, the residual dipolar coupling for a peptide with the sequence Ac-YES-G₆-ATD was very different from that of a peptide with the sequence Ac-YGEGSGAGTGDG, where the stretch of contiguous glycines is broken up. Thus, to test the effect of local sequence/hydration on the conformational properties of glycine rich polypeptide segments, we have prepared variants with the same glycine content as in the early part of this investigation,⁸³ but with a decrease in the number of contiguous glycine residues.

The third hypothesis is that local sequence dominates early in folding when the overall structure of a protein is largely disordered. As discussed above, the work of Robinson et al.¹¹⁵ suggested that equilibrium stability of loops or linkers between two subunits of a protein is primarily dependent on sequence composition rather than specific sequence. This investigation tests whether those findings are applicable to nucleating structures such as simple loops in the denatured state or if they are only applicable to the large contact interfaces of subunits or subdomains. In particular, we test whether the amino acid – glycine *versus* alanine – next to the histidine involved in loop formation has a dominant impact on loop dynamics. In the early part of this investigation, variants with 1 to 3 glycines in the 5 amino acid insert in front of the histidine had an alanine next to the histidine, whereas the inserts with 4 and 5 glycines had a glycine next to the histidine (see Table 2.1).

2.2 Materials and Methods

2.2.1 Design of first generation poly-Ala/Gly variants

Mutants were designed with the triple mutant (TM) variant of yeast iso-1-cytochrome *c* as the template.^{117; 118} This TM variant has the mutations H26N, H33N and H39Q, which remove all histidines that can participate in denatured state His-heme ligation. Since the TM variant does not form His-heme loops in the denatured state, it also acts as a control to assess the effects of such loops on global stability. We have found in previous work that the N-terminal amino group competes with His-heme loop formation in the denatured state of iso-1-cytochrome *c*,^{80; 119} and must be blocked through *in vivo* N-terminal acetylation in yeast to allow His-heme affinity in the denatured state to be measured accurately.⁸⁰ We have also studied variants having the mutation K(-2)H that can make a 16 residue His-heme loop in the denatured state.^{81; 89} The histidine at position -2 in these variants has a high affinity for the heme due to its location on the N-terminal side of the site of heme attachment to iso-1-cytochrome *c*.⁸¹ Thus, homopolymeric sequence insertions have been made on the N-terminal side of the site of heme attachment to make N-terminal amino group competition with His-heme binding in the denatured state insignificant. This is important for the current work since the N-terminus of cytochrome *c* is not acetylated in *Escherichia coli* (*E. coli*), which is being used to express the variant proteins described here.

We initially inserted the sequence KAAAAA in between F(-3) and K(-2) of the TM variant to produce the variant NK5A. Alanines were used in this insert because they have one of the simplest side chains and on the Ramachandran plot most amino acids are found to be similar to alanine. Lysine was added to this insert to maintain solubility and

prevent aggregation. The insertion is in a disordered region (see Figure 2.2) of the protein, as judged from the high thermal factors for residues -5 to -1 in the crystal structure of yeast iso-1-cytochrome *c*.¹²⁰ Therefore, mutations in this region are not expected to have significant effects on the overall structure or stability of the native fold.

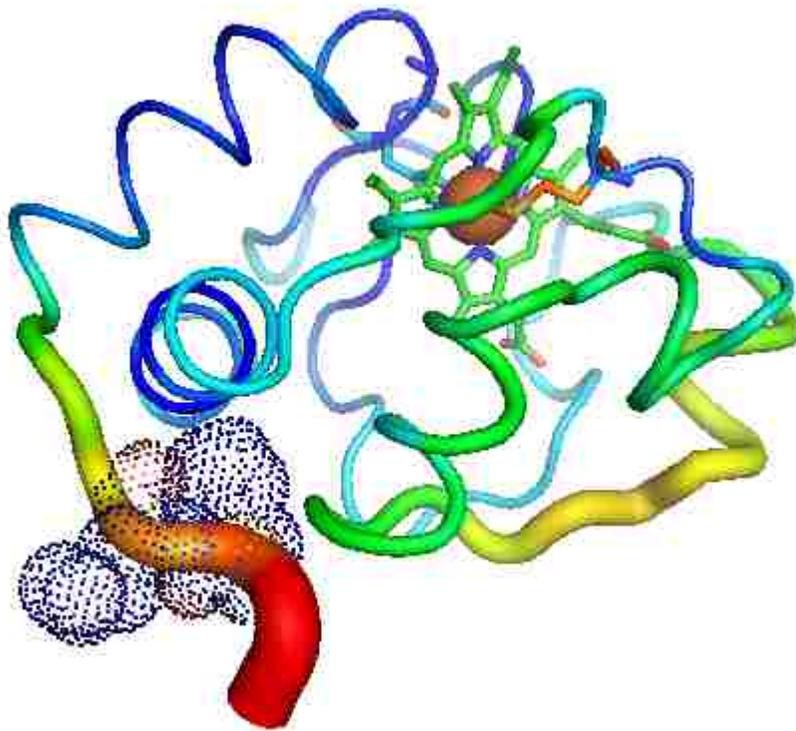


Figure 2.2: B-factor putty representation of iso-1-cytochrome *c* showing disordered region (orange to red region). Position of insertions [between F(-3) and K(-2)] is indicated with dots. Pymol was used in conjunction with the 2YCC PDB file to generate the structure.

The inserted lysine in this cassette was then mutated to a histidine residue to produce the NH5A variant which is capable of forming a 22-residue His-heme loop (the loop begins with Cys14, which forms a thioether bond to a heme vinyl substituent and ends at the

engineered histidine which binds to the Fe³⁺ in the heme) in the denatured state. Figure 2.1 is a schematic representation of how this loop would form from the N-terminal side of the heme.

To test the effect of chain flexibility, the alanines in the NH5A insert were progressively replaced with glycines to produce variants Gly 1 to Gly 5 (Table 2.1). Sequential replacement of Ala with Gly could have been accomplished in many ways. We arbitrarily chose to replace Ala with Gly outward from the center of the five Ala insert. We note that there are no prolines in any of the His-heme loops which could modulate loop formation through *cis/trans* proline isomerization.^{110; 121} The loop also contains two additional glycines (Gly1 and Gly6) and two β -substituted amino acids (Thr8 and Thr12). Our previous work indicates that changes in β -substituted amino acid content have modest effects on loop equilibria.⁸¹ Including glycines 1 and 6, the overall glycine content in the 22-residue loop goes from 9% to 32% as the NH5A variant is progressively converted to the Gly 5 variant, substantially increasing the flexibility of the polypeptide segment involved in loop formation.

2.2.3 Design of second generation poly-Ala/Gly variants

The first generation of poly-Ala/Gly inserts was made by progressively replacing the five alanines in the NH5A variant with glycine, working from the center of the insert outwards (Table 2.1).⁸³ In this set of variants, all of the glycines were contiguous and the amino acid next to His (-2) was Ala in some cases and Gly in others.

% Gly in Insert	First Generation Variants	Second Generation Variants
0	NH5A (HAAAAAK)	--
20	Gly 1 (HAAGAAK)	Gly1v2 (HGAAAAK)
40	Gly 2 (HAGGAAK)	--
60	Gly 3 (HAGGGAK)	Gly3v2 (HGAGAGK)
80	Gly 4 (HGGGGAK)	Gly4v2 (HGGAGGK) Gly4v3 (HAGGGGK)
100	Gly 5 (HGGGGGK)	--

To test the role of local sequence or backbone hydration in extending the main chain of poly-Gly sequences, we have prepared two new variants where the number of contiguous glycines is decreased, Gly3v2 and Gly4v2 (Table 2.1). In Gly3v2, the three glycines are now separated by alanines in the 5 amino acid insert. In Gly4v2, the sole alanine is now placed in the center of the insert, so each run of glycines is only two amino acids in length. We note that the kinetic experiments on the His-heme loops will also test the degree of extension of the poly-Gly inserts. If hydration or local sequence sterics extend the main chain of the poly-Gly inserts to the level of the poly-Ala insert (or greater,¹²² then the forward rate constant for loop formation, k_f , should not be significantly increased when alanines are replaced by contiguous glycines and k_f should be sensitive to whether or not the glycines are contiguous.

We also made several variants that test whether the amino acid next to His(-2) on the C-terminal side (i.e., within the loop) – Ala *versus* Gly – affects His-heme loop formation (Table 2.1). The Gly1v2 variant places the sole Gly in the insert next to

His(-2) for comparison with the Gly 1 variant which has Ala next to His(-2). The Gly4v3 variant places the sole Ala in the insert next to His(-2), for comparison with the Gly 4 variant which has Gly next to His(-2).

2.2.4 Site-directed mutagenesis

All variants were made using the unique restriction site elimination method¹²³ as previously described.¹²⁴ Single stranded DNA (ssDNA), carrying the appropriate template DNA for site-directed mutagenesis, was obtained using helper phage R408¹²⁵ via infection of transformed *E. coli* TG-1 cells. The NK5A variant was initially made using single-stranded pBTR1 vector (provided by Grant Mauk at the University of British Columbia)^{126; 127} DNA carrying the TM variant of iso-1-cytochrome *c* as the template and the following primer: 5'-GA ACC GGC CTT AGC AGC AGC AGC AGC TTT GAA TTC AGT C-3'.⁹⁰ Underlined sequence indicates region of mutation. The *EcoRV*-selection primer, 5'-d(GTGCCACCTGACGTCTAAGAAACC)-3' was used together with the NK5A mutagenic primer in the mutagenesis reaction. It converts a unique *EcoRV* restriction site in the pBTR1 vector carrying the TM variant, into an *Aat2* restriction site. The NH5A variant was subsequently made using single-stranded pBTR1 vector DNA carrying the NK5A variant of iso-1-cytochrome *c* as the template by mutating the lysine to a histidine using the following mutagenic primer: 5'-GA ACC GGC CTT AGC AGC AGC AGC AGC GTG GAA TTC AGT C)-3'.⁸³ The *Aat2*-selection primer, 5'-d(GTGCCACCTGATATCTAAGAAACC)-3' was used together with the NK5A mutagenic primer in the mutagenesis reaction. It converts a unique *Aat2* restriction site in the pBTR1 vector carrying the NK5A variant, into an *EcoRV* restriction

site. The Gly 1 and Gly 2 variants were made using single-stranded pBTR1 vector DNA carrying the NH5A variant of iso-1-cytochrome *c* as the template and the following primers: 5'-d(CTTAGCAGCACCAGCAGCGTG)-3', and 5'-d(CTTAGCAGCACCGCCAGCGTGGAA)-3', respectively. The Gly 3 variant was made using single-stranded pBTR1 vector DNA carrying the Gly 2 variant of iso-1-cytochrome *c* as the template and the following primer: 5'-d(GGCCTTAGCGCCACCGCCAGC)-3'. The Gly 4 variant was made using single-stranded pBTR1 vector DNA carrying the Gly 3 variant of iso-1-cytochrome *c* as the template and the following primer: 5'-d(GCCACCGCCACCGTGGAATTC)-3'. The Gly 5 variant was made using single-stranded pBTR1 vector DNA carrying the Gly 4 variant of iso-1-cytochrome *c* as the template and the following primer: 5'-d(ACCGGCCTTACCGCCACCGC)-3'. The EcoRV- selection primer was used together with the Gly 1, Gly 2 and Gly 4 mutagenic primers in the mutagenesis reactions, while the Aat2- selection primer was used together with the Gly 3 and Gly 5 mutagenic primers in the mutagenesis reactions.

The Gly1v2 variant was made using single-stranded pBTR1 vector DNA carrying the NH5A variant of iso-1-cytochrome *c* as the template and the following primer: 5'-d(CAGCAGCAGCACCGTGGAATTCAG)-3'. The Gly3v2, Gly4v2 and Gly4v3 variants were made using single-stranded pBTR1 vector DNA carrying the Gly 5 variant of iso-1-cytochrome *c* as the template and the following primers: 5'-d(GCCTTACCTTGCACCAGCACCGTGGA)-3' for the Gly3v2 variant, 5'-d(CTTACCGCCAGCGCCACCGTG)-3' for the Gly4v2 variant and 5'-d(GCCACCGCCAGCGTGGAATTC)-3' for the Gly4v3 variant. The EcoRV-

selection primer was used together with the above mutagenic primers in the mutagenesis reactions.

2.2.5 Expression and purification of protein variants

Iso-1-cytochrome *c* mutant genes were expressed from BL21-DE3 *Escherichia coli* cells (Novagen) using the pBTR1 vector as previously described using standard protocols from the Bowler laboratory.^{83; 89} This vector co-expresses heme lyase to allow covalent attachment of heme to iso-1-cytochrome *c* in the cytoplasm of *E. coli*. Rich growth media (2xYT) was used for expression with incubation periods of 16-18 hours. Typical yields ranged from 30 – 45 mg per liter of culture, depending on the variant. Initial pH titrations with NH₅A and Gly 1 variants gave apparent pK_a (pK_a(obs)) values that were much higher than expected and not in accord with His-heme ligation based on our previous work with histidine at position -2.⁸¹ Subsequent MALDI-TOF analyses of the proteins showed mass fragmentation around a specific region of mass/charge ratios (see Figure 2.3). Further analyses of the fragmentation indicated that the N-terminal region of the protein, H₃N⁺...F(-3)HAAAAAK(-2)...COO⁻, was being cleaved after the second alanine residue in the insert, thereby forming the H₃N⁺...F(-3)HAA//AAAK(-2)...COO⁻ fragments. On suspicion of possible metal-catalyzed cleavage,^{128; 129} all buffers and 18Ω deionized water were tested using Buettner's ascorbate test.¹³⁰ The results indicated that the high salt HPLC buffer (1 M NaCl, 50 mM sodium phosphate) likely had contamination with redox active metals. To neutralize the effects of metal contamination, 1.0 mM EDTA was added to all buffers. Mass spectral analysis using a Bruker Daltronics Reflex IV MALDI-TOF mass spectrometer demonstrated that this step

successfully minimized the cleavage problem (see Figure 2.4). Similar problems were not encountered with the NK5A variant, suggesting that the histidine was necessary for fragmentation and a likely site for metal binding. The possibility of a metalloprotease co-eluting with the HPLC purified protein could not initially be ruled out since EDTA would be expected to inhibit a metalloprotease as well. Nonetheless, the protein isolation protocol was then further optimized with more rapid processing by ultrafiltration after HPLC purification. These improvements allowed for acceptable experimentation on the initial set of sensitive variants (NH5A, Gly 1 to Gly 5) with minimal to no fragmentation (see Figure 2.5).

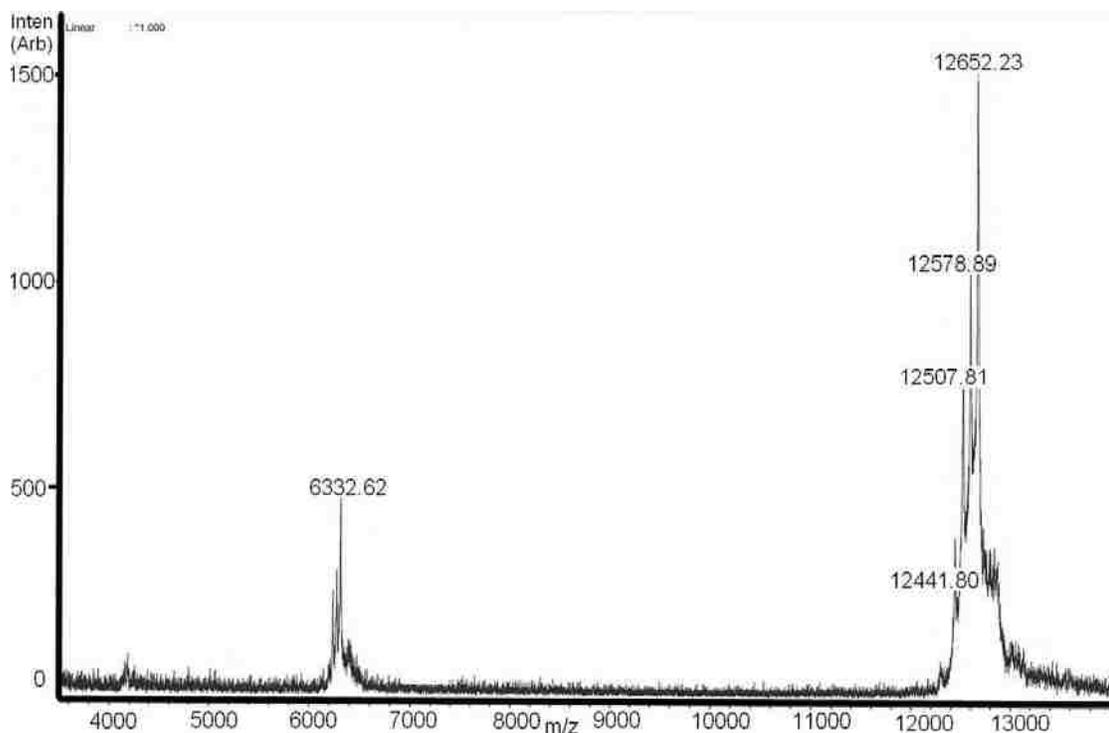


Figure 2.3: MALDI-TOF spectrum of NH5A variant showing drastic fragmentation after HPLC purification. Protein samples were prepared in 1 mL of saturated sinapic acid matrix (Fluka) containing 30% acetonitrile and 0.1% trifluoroacetic acid to a final protein concentration of 6 μ M. Horse cytochrome *c* was used as an external standard. Samples were analyzed on a Bruker Daltronics Reflex IV MALDI-TOF mass spectrometer. Expected m/z: 13, 089; Expected m/z for NH5A cleavage between 2nd and 3rd alanines of the insert: 12, 432. Higher m/z peaks correspond to one less amino acid cleaved from the N-terminus. Instrumental resolution: At 10, 000 \pm 10 m/z.

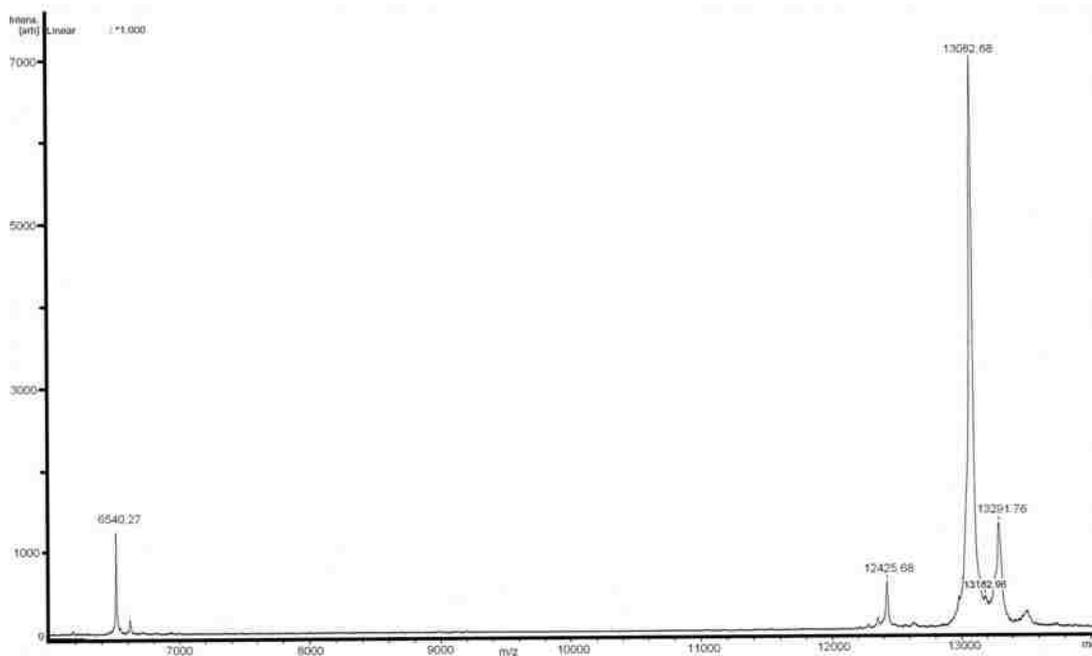


Figure 2.4: MALDI-TOF Spectrum of NH5A variant showing reduced fragmentation after HPLC purification with 1.0 mM EDTA in buffers. Expected m/z: 13, 089; Expected m/z for NH5A cleavage between 2nd and 3rd alanines of the insert: 12, 432. Instrumental resolution: At 10, 000 \pm 10 m/z.

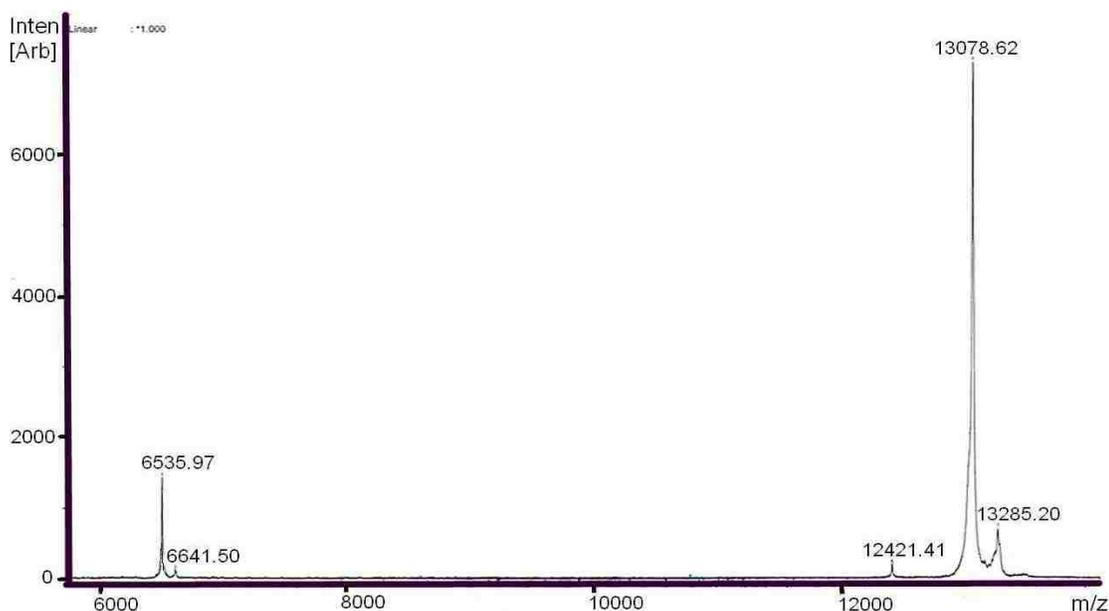


Figure 2.5: MALDI-TOF Spectrum of NH5A variant showing minimal fragmentation after HPLC purification with 1.0 mM EDTA in buffers and added optimization techniques. Expected m/z: 13, 089; Expected m/z for NH5A cleavage between 2nd and 3rd alanines of the insert: 12, 432. Instrumental resolution: At 10, 000 \pm 10 m/z.

Expression and isolation of the second generation variants (Gly1v2, Gly3v2, Gly4v2 and Gly4v3) suffered from a similar fragmentation problem as for NH5A and Gly 1 to Gly 5.⁸³ However, cleavage of the protein near the N-terminus was drastically minimized through diligent and rapid protein workups with sufficient protease inhibitor (3 mM PMSF) present in cell lysates. Cleavage is more than likely due to trace amounts of redox active metals,¹³⁰ as suspected earlier, since the possibility of a metalloprotease co-eluting with the HPLC purified protein can likely be ruled out because rapid processing by ultrafiltration down to \sim 200 μ L after HPLC purification also seems to arrest the cleavage process relative to processing to a larger final volume. Furthermore, buffer solutions above physiological pH enable longevity of intact proteins presumably due to deactivation of redox active metals as metal hydroxide precipitating out of

solution. An Applied Biosystems Voyager – DE PRO Biospectrometry Workstation MALDI-TOF instrument was used to verify the integrity of all the isolated as well as purified second generation iso-1-cytochrome *c* variants.

Proteins were oxidized with potassium ferricyanide just prior to each experiment and separated from the oxidizing agent by Sephadex G-25 chromatography. The Sephadex G-25 resin was equilibrated to a buffer appropriate to the experiment. All buffers used for experiments contained 1.0 mM EDTA. Fragmentation is effectively neutralized by adding 1.0 mM EDTA to all buffers used in protein isolation and purification coupled with rapid ultracentrifugation.⁸³ Ultrafiltration devices were also pre-washed with EDTA-containing buffers before use. All proteins were analyzed by MALDI-TOF mass spectrometry, as described above, prior to and after each experiment to verify the integrity of the protein. If more than 10% degradation of the protein was observed, the experimental data were discarded and the experiment repeated with freshly purified protein.

2.2.6 Protein stability measurements

The stability of all variants was monitored at 25 °C as a function of gdnHCl concentration using either an Applied Photophysics PiStar 180 Spectrometer or an Applied Photophysics Chirascan circular dichroism spectrometer coupled to a Hamilton MICROLAB 500 Titrator using methods described previously.¹³¹ Data were acquired at pH 7.0 in the presence of 20 mM Tris, 40 mM NaCl as buffer with 1.0 mM EDTA to minimize protein cleavage. Ellipticity data at 250 nm, acquired simultaneously, were subtracted as background at each [gdnHCl]. The data were fit to a linear free energy

relationship, as described previously,^{80; 117} to extract the free energy of unfolding in the absence of denaturant, $\Delta G_u^{o'}(\text{H}_2\text{O})$, and the m -value (rate of change of free energy as a function of denaturant concentration: $\Delta G = \Delta G_u^{o'}(\text{H}_2\text{O}) - m[\text{gdnHCl}]$).^{132; 133} Two-state folding is assumed in these fits. At low gdnHCl concentrations, slight increases in ellipticity at 222 nm have been observed possibly due to specific Cl^- binding.^{48; 134} The ellipticity subsequently levels off before the unfolding transition. Thus, a constant native state baseline has been used here as previously described.^{48; 80} Reported parameters are the average and standard deviation of three independent trials. Fragmentation greater than 10% of the major peak required a restart of the entire experimental attempt (i.e. starting from re- isolation and re-purification of the protein).

2.2.7 Equilibrium loop formation in the denatured state

Equilibrium loop formation in the denatured state (3 M gdnHCl, 5 mM Na_2HPO_4 , 15 mM NaCl with 1.0 mM EDTA) was monitored through pH titrations using either a Beckman DU-640 UV-Vis spectrophotometer or a Beckman DU-800 UV-Vis spectrophotometer. All titrations were done at 3 μM protein concentration and at room temperature, 22 ± 1 °C. Titrations were done in sets of three starting from about pH 7 down to about pH 3 in intervals of approximately 0.2 pH units. Spectra from 350 to 450 nm were acquired at each pH. Titration procedures have been described previously.⁸² Data at 398 nm *versus* pH were fit to a modified form of the Henderson-Hasselbalch equation allowing extraction of the apparent pK_a for loop formation, $\text{pK}_a(\text{obs})$, and the number of protons, n , involved in the process. Reported parameters are the average and

standard deviation of three independent trials.

2.2.8 pH dependent stopped-flow kinetic measurements

In order to monitor the breakage of the His-heme bond in the denatured state for all variants, stopped flow mixing methods were used (Applied Photophysics SX-20 spectrometer). Reaction progress was monitored by absorbance spectroscopy at 398 nm to observe the Soret band shift resulting from His-heme bond breakage.⁸² All data were collected at 25 °C. For pH dependent His-heme bond breakage reactions, the 10 mm pathlength of the 20 μ L flow cell was used. The final reaction mixture was obtained from 1:1 mixing of 6 μ M protein, 3 M gdnHCl, 10 mM MES pH 6.2, 1.0 mM EDTA with 3 M gdnHCl, 100 mM glycine buffer containing 1.0 mM EDTA to achieve the desired ending pH of 3.0 or 3.5. All kinetics experiments were done the same day each variant was HPLC purified. This minimized the N-terminal cleavage of sensitive variants in the lower pH starting buffer (pH 6.2). Final reaction pH was determined by collecting the product of the mixing reaction and immediately measuring pH. Using the method of reduction of 2,6-dichlorophenolindophenol,¹³⁵ a 0.7 ms dead time was determined for the 20 μ L flow cell under our mixing conditions. Loop breakage rate constants were obtained by adjusting the stop time to the dead time of the instrument and then removing data up to the first 3 ms to deal with an instrumental glitch. All data were fit to a single exponential rise to maximum equation. Double exponential fits were attempted but did not significantly improve the fit to the data.

2.3 Results

2.3.1 Stability of variants

The global stability of all variants was measured by gdnHCl denaturation monitored by circular dichroism spectroscopy. Figures 2.6 and 2.7 are representative of data collected on an Applied Photophysics PiStar 180 and an Applied Photophysics Chirascan circular dichroism spectrometer, respectively. The former instrument was used for data collection of the first generation variants (NH5A, Gly 1, Gly 2, Gly 3, Gly 4 and Gly 5); while the latter instrument was used for data collection of the second generation variants (Gly1v2, Gly3v2, Gly4v2 and Gly4v3). A summary of the thermodynamic parameters obtained from the data is presented in Tables 2.2 and 2.3.

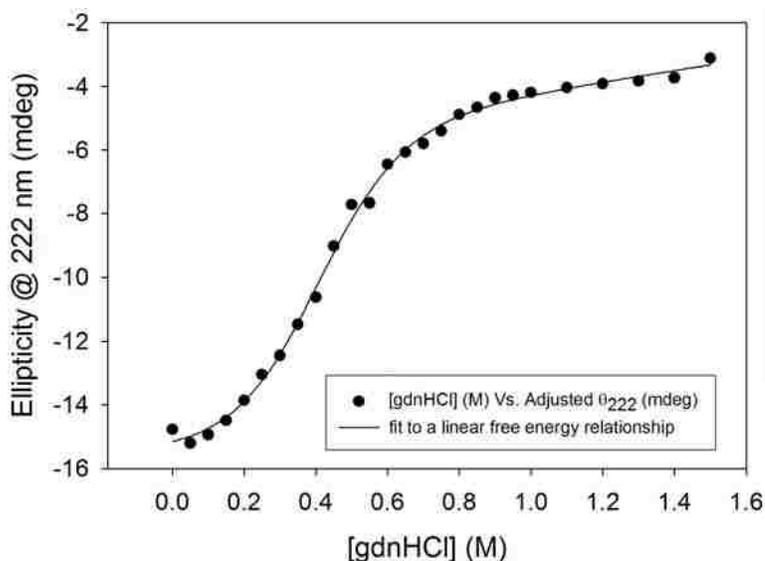


Figure 2.6 A typical gdnHCl protein denaturation titration monitored at 222 nm using an Applied Photophysics PiStar 180 circular dichroism spectrometer (Gly 4 variant Run #2). Data were collected at 25 °C.

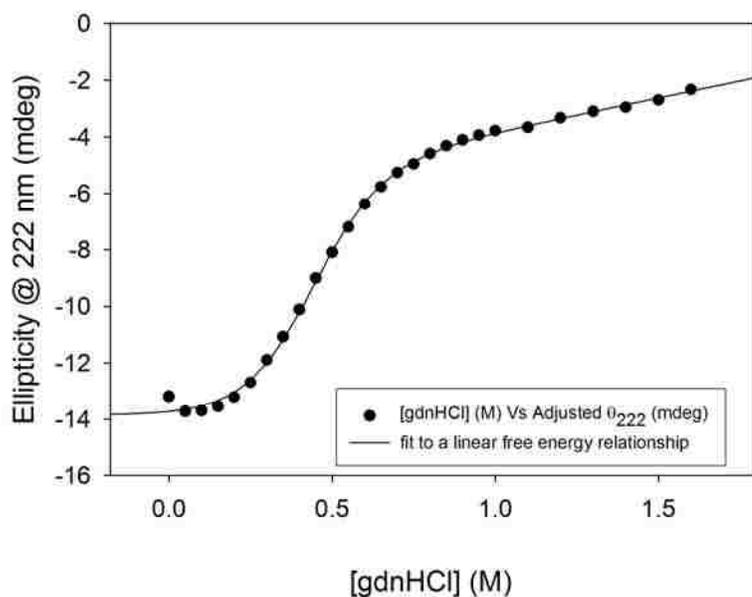


Figure 2.7: A typical gdnHCl protein denaturation titration monitored at 222 nm using an Applied Photophysics Chirascan circular dichroism spectrometer (Gly 4v2 variant Run #2). Data were collected at 25 °C.

Table 2.2: Thermodynamic parameters for gdnHCl induced unfolding of iso-1-cytochrome *c* variants at pH 7 and 25 °C for first generation variants

Iso-1-cytochrome <i>c</i> variant	$\Delta G_u^{\circ}(\text{H}_2\text{O})$ (kcal/mol) ^a	<i>m</i> -value (kcal/(mol* <i>M</i>)) ^a	<i>C_m</i> (<i>M</i>) ^a
TM ^b	4.02±0.17	3.99±0.17	1.008±0.0003
NK5A	4.21±0.16	4.11±0.01	1.03±0.04
NH5A	2.00±0.08	4.52±0.18	0.44±0.03
Gly 1	1.87± 0.28	4.36±0.63	0.43±0.02
Gly 2	2.05±0.16	4.42±0.19	0.46±0.03
Gly 3	1.97±0.19	4.42±0.67	0.45±0.04
Gly 4	2.04±0.17	4.72±0.29	0.43±0.03
Gly 5	1.92±0.21	5.06±0.84	0.38±0.03

^a Average values reported are based on three separate experiments and MALDI-TOF analysis was done before each run to assess the degree of fragmentation.

^b Data values obtained from Godbole and Bowler¹¹⁸ for comparison purposes.

Immediately obvious from Table 2.2 is the fact that all the mutants capable of forming the 22 residue His–heme loop in the denatured state have approximately half the free energy of unfolding, $\Delta G_u^{\circ}(\text{H}_2\text{O})$, of the TM and NK5A variants. Coupled with the denatured state loop formation results discussed later, this difference in $\Delta G_u^{\circ}(\text{H}_2\text{O})$ strongly suggests that the His–heme loop formed under denaturing conditions stabilizes the denatured state relative to the fully folded form (see Discussion). No significant differences in the stabilities of the NH5A and Gly 1 to Gly 5 variants are observed. The denaturation midpoints, C_m , are all less than 0.5, which indicates that these proteins are fully unfolded in 3 M gdnHCl, the conditions for the pH titrations used to determine denatured state loop stability.

2.3.2 Global stability of new Gly X variants

Table 2.3: Thermodynamic parameters for gdnHCl induced unfolding of iso-1-cytochrome *c* variants at pH 7.0 and 25 °C for second generation variants

Iso-1-cytochrome <i>c</i> variant	$\Delta G_u^{\circ}(\text{H}_2\text{O})$ (kcal/mol) ^a	<i>m</i> -value (kcal/(mol*M)) ^a	C_m (M) ^a
Gly1v2	2.70 ± 0.07	5.52 ± 0.25	0.49 ± 0.03
Gly3v2	2.33 ± 0.24	5.19 ± 0.64	0.45 ± 0.07
Gly4v2	2.34 ± 0.16	5.42 ± 0.46	0.43 ± 0.05
Gly4v3	2.25 ± 0.09	5.23 ± 0.08	0.43 ± 0.02

^aValues reported are based on an average of 3 separate experiments. The reported error is the standard deviation.

Data were collected with an Applied Photophysics Chirascan circular dichroism spectrometer coupled to a Hamilton Microlab 500 titrator as previously described.¹³¹

All second generation variants have higher global stability ranging from about 0.3 to 0.8 kcal/mol more than their first generation counterparts (compare Tables 2.2 and 2.3).⁸³ Specifically, the largest change is observed for Gly1v2 (2.70 kcal/mol) compared to Gly 1 (1.87 kcal/mol) while the smallest is Gly4v3 (2.25 kcal/mol) compared to Gly 4 (2.02 kcal/mol). The denaturation midpoints, C_m , of the second generation variants are all less than 0.5, confirming that these variants are fully denatured in 3 M gdnHCl, the experimental conditions used for denatured state loop formation.

2.3.3 Equilibrium loop formation in the denatured state for initial variants

One of the advantages of using His-heme ligation as a probe is that a direct correlation can be made between the apparent pK_a , $pK_a(\text{obs})$, of titrating the histidine off the heme and the strength of that ligand interaction. Typically, the apparent pK_a is lower than the pK_a of free histidine because deprotonation of histidine is coupled to favorable formation of a bond between Fe^{3+} in the heme and the histidine imidazole side-chain. The stability of the His-heme bond is dependent on three main factors as previously described. First, chain stiffness can decrease the stability if the polypeptide chain forming the loop is too short or too sterically restrictive, thereby hindering effective His-heme molecular orbital overlap.^{80; 81} These hindrances can be relieved to increase the bond strength by either increasing loop size or decreasing residue bulk. Second, loop entropy decreases the His-heme bond strength as the loop becomes larger. Too long a loop increases the conformational space for the histidine to search, thereby decreasing the probability of making an actual His-heme bond. Third, residual structure formation increases the His-heme bond strength due to the presence of additional stabilizing

contributions in the loop, if the structure is induced by loop formation. The summed stability of these contributors is conferred onto the His–heme bond. Lower pK_a values for the His–heme ligation are a direct indication that the interaction is stronger and thus requires more acid to break the bond by protonating the imidazole ring of the histidine. Using this probe of relative denatured state loop stability for the NH5A and Gly X (where X = 1 to 5) variants, we were able to obtain direct information on the effect that sequence composition and thus main-chain sterics has on loop formation in the denatured state.

Loop breakage is detected at 398 nm which is the wavelength to which the low spin Fe^{3+} –heme Soret band shifts when a high spin Fe^{3+} –heme forms due to the weak field water ligand being bound to the Fe^{3+} instead of the previously bound strong field histidine ligand (see Figure 2.1). A typical pH titration in 3 M gdnHCl is shown in Figure 2.8 along with a plot of spectra as a function of pH as an inset. For the variants in this series of experiments, the pH range of pH 7 to 3 was more than adequate to define the apparent pK_a . In particular, upper and lower baselines are well-formed allowing for straightforward extraction of $pK_a(\text{obs})$ from fits of the data to the Henderson–Hasselbalch equation. The inset in Figure 2.8 shows spectra as a function of pH for the titration in Figure 2.9 and is typical for titrations with these variants. There are two isosbestic points at 403 nm (above pH 3.5) and 398 nm (below pH 3.5). This second isosbestic point is possibly due to His18 titrating off the heme. Since all experiments monitor loop breakage at 398 nm, the apparent pK_a values for loop breakage obtained at 398 nm should be insensitive to breakage of the His18–heme bond.

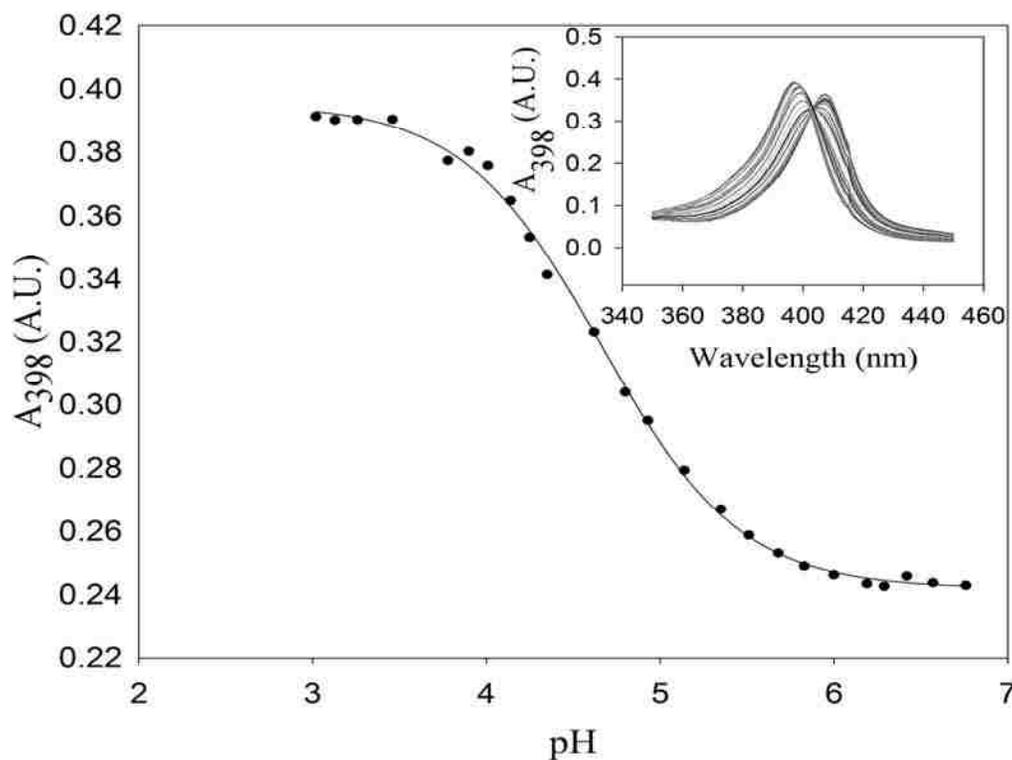


Figure 2.8: A representative pH titration curve monitoring His-heme ligation in 3 M gdnHCl, 5 mM sodium phosphate, 15 mM NaCl, 1.0 mM EDTA. Data were acquired at room temperature, $22 \pm 1^\circ\text{C}$, using a Beckman DU-640 spectrophotometer following previously discussed titration procedures.^{80; 117-119} Data were fitted to a modified form of the Henderson-Hasselbalch equation, allowing extraction of both the $\text{pK}_{\text{a}}(\text{obs})$ and the number of protons coupled to loop formation.^{80; 117-119} Titrations were done in sets of three starting from pH 7 down to about pH 3 in intervals of approximately 0.2 pH units. The inset shows absorbance spectra from 350 to 450 nm as a function of pH over the pH range 7 to 3 (bottom to top at 398 nm) in approximate increments of 0.2 pH units.

Table 2.4 summarizes the $\text{pK}_{\text{a}}(\text{obs})$ results for the six different His-heme loop-forming variants (NH5A and Gly 1 to Gly 5), as well as for the NK5A and TM variants. This Table also shows the experimental values of n , the number of protons involved in loop breakage. All variants yield $n \sim 1$, which is in agreement with the scheme for His-heme loop equilibria in Figure 2.1.

Table 2.4: Thermodynamic parameters for denatured state loop formation at 22 °C in 3 M gdnHCl for variants of iso-1- cytochrome *c*

Iso-1-cytochrome <i>c</i> variant	pK _a (obs) ^a		<i>n</i> -value ^{a,b}	
	Data Set 1	Data Set 2	Data Set 1	Data Set 2
TM ^c	5.94 ± 0.08		0.9 ± 0.1	
H(-2)I52 ^d	4.22 ± 0.09		1.11 ± 0.07	
NK5A ^d	6.22 ± 0.06		0.88 ± 0.12	
NH5A	4.64 ± 0.02	4.59 ± 0.04	1.06 ± 0.03	0.99 ± 0.05
Gly 1	4.54 ± 0.02	4.56 ± 0.02	0.97 ± 0.02	1.02 ± 0.06
Gly 2	4.45 ± 0.03	4.53 ± 0.02	1.03 ± 0.02	0.98 ± 0.04
Gly 3	4.51 ± 0.02	4.53 ± 0.02	1.11 ± 0.12	0.99 ± 0.02
Gly 4	4.58 ± 0.03	4.58 ± 0.01	1.08 ± 0.04	1.05 ± 0.01
Gly 5	4.59 ± 0.02	4.60 ± 0.02	1.04 ± 0.03	1.13 ± 0.10

^a Average and standard deviation reported are based on three titration experiments carried out in sequence. MALDI-TOF analysis was done before each run to ensure minimal fragmentation. Parameters were obtained as described in the legend to Figure 2.8.

^b Proton transfer numbers (*n*-values) for each variant are expected to be 1 (see Figure 2.1).

^c Data values obtained from Godbole and Bowler.¹¹⁸

^d Data obtained from the Ph. D. thesis of Eydiejo Kurchan.⁹⁰ This variant, like the previously reported TM and H(-2)I52 variants, does not show any fragmentation.^{83; 90}

Previous experiments done in our laboratory with the TM variant gave pK_a(obs) = 5.94 in 3 M gdnHCl for N-terminal amino group ligation to the heme.^{118; 119} The NK5A variant yields a pK_a(obs) value of 6.22 in 3 M gdnHCl, an increase of ~ 0.3 unit relative to the TM variant (Table 2.4). This increase is consistent with the N-terminal amino group being six amino acid residues further from the heme.

Replacing the lysine in NK5A with a histidine (NH5A) causes a drastic drop in the pK_a(obs) value to ~ 4.6, consistent with the high affinity of histidines for the heme observed for histidines on the N-terminal side of the site of heme attachment in iso-1-cytochrome *c*.⁸¹ Thus, interference from the N-terminal amino group in assessing pK_a(obs) for histidine–heme ligation with the variants studied here should be minor. Another comparison in this regard is instructive. Previous studies gave pK_a(obs) = 4.20 ± 0.04 for the variant AcH(-2).⁸¹ This variant has an acetylated N terminus and a histidine

at position -2. The H(-2)I52 variant also has a histidine at position -2 (I52 indicates a stabilizing N52I mutation).⁸⁹ The H(-2)I52 variant is expressed in bacteria and thus its N terminus is not acetylated. Its $pK_a(\text{obs})$ reported here is 4.22 ± 0.09 .⁹⁰ The fact that these variants' $pK_a(\text{obs})$ values are identical within error demonstrates that the N-terminal amino group does not perturb $pK_a(\text{obs})$ values for His-heme ligation from the N-terminal side of the heme.

The $pK_a(\text{obs})$ values in Table 2.4 also show that His-heme ligation is stronger for the 16 residue loop formed by the H(-2)I52 variant than for the longer 22 residue loop formed by the NH5A variant. The increase in the $pK_a(\text{obs})$ of ~ 0.4 unit for His-heme ligation in 3 M gdnHCl caused by the AAAA₄K insert is similar to the ~ 0.3 unit increase for N-terminal amino group-heme ligation for the NK5A *versus* the TM variant. Thus, loop affinity scales similarly with loop size for the two types of denatured state loop formation.

The $pK_a(\text{obs})$ results for the Gly X variants in Table 2.4 are surprising. The expected trend of progressively decreasing $pK_a(\text{obs})$ values with increasing chain flexibility was not observed. The initial data set collected showed a decrease in $pK_a(\text{obs})$ values, though subtle, for the series of variants, NH5A, Gly 1 and Gly 2. However, instead of continuing this trend, the $pK_a(\text{obs})$ values increased for the series of variants, Gly 3, Gly 4 and Gly 5. The differences in the $pK_a(\text{obs})$ values were very small. Therefore, to evaluate the error in the data due to calibrating the pH meter, a second independent data set of $pK_a(\text{obs})$ values was collected (Table 2.4). Figure 2.9 shows the plot of both data sets and even though the differences are small, the trend is still the same. There is a subtle decrease in the $pK_a(\text{obs})$ reaching a minimum for the Gly 2 or Gly 3

variant. This result is in accord with the chain being more flexible and thus more compact upon addition of less sterically restrictive glycine *versus* alanine. At higher glycine content, there is a slight increase in the $pK_a(\text{obs})$ values. The greater flexibility presumably leads to a greater loss in conformational space when the loop forms for inserts with three or more of the five alanine residues replaced with glycine. These observations are in accord with the results on linker composition in Arc repressor by Robinson and Sauer.¹¹⁵ In that work it was shown that an adequate percentage of glycine residues in the linker was very useful in bringing the two subunits together. However, having too many glycine residues in the linker was counterproductive.

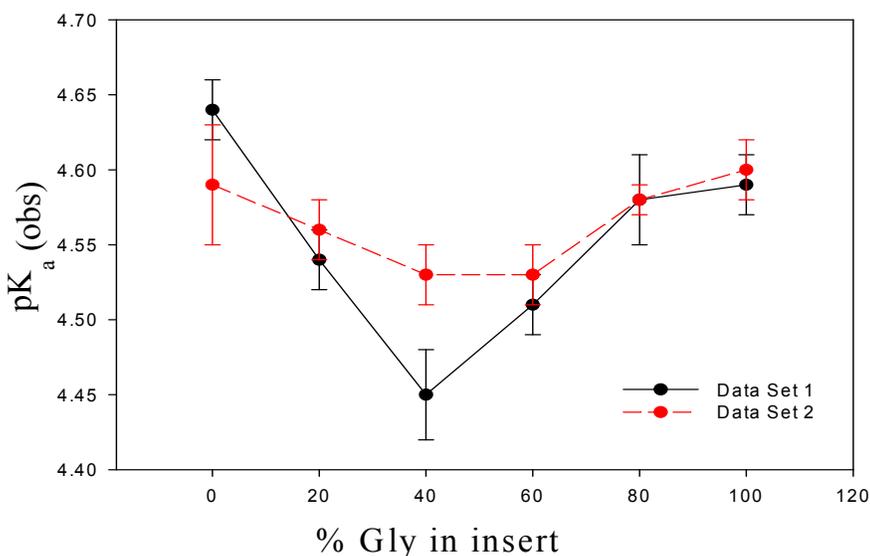


Figure 2.9: Graph of the two independent data sets showing pH titration reproducibility due to error in calibrating the pH meter. The errors for each $pK_a(\text{obs})$ are the standard deviation for three pH titrations done sequentially for each variant in each data set. The % Gly in insert axis indicates the percentage of the five alanine residues in the insert in the NH5A variant which has been replaced with glycine. Thus, the NH5A variant has 0% Gly in insert and Gly 5 variant has 100% Gly in insert.

2.3.4 Equilibrium loop formation in the denatured state for second generation variants

The stabilities of the His-heme loops under denaturing conditions (3 M gdnHCl) were measured for all second generation variants. Figure 2.10A represents a typical titration curve for equilibrium loop formation.

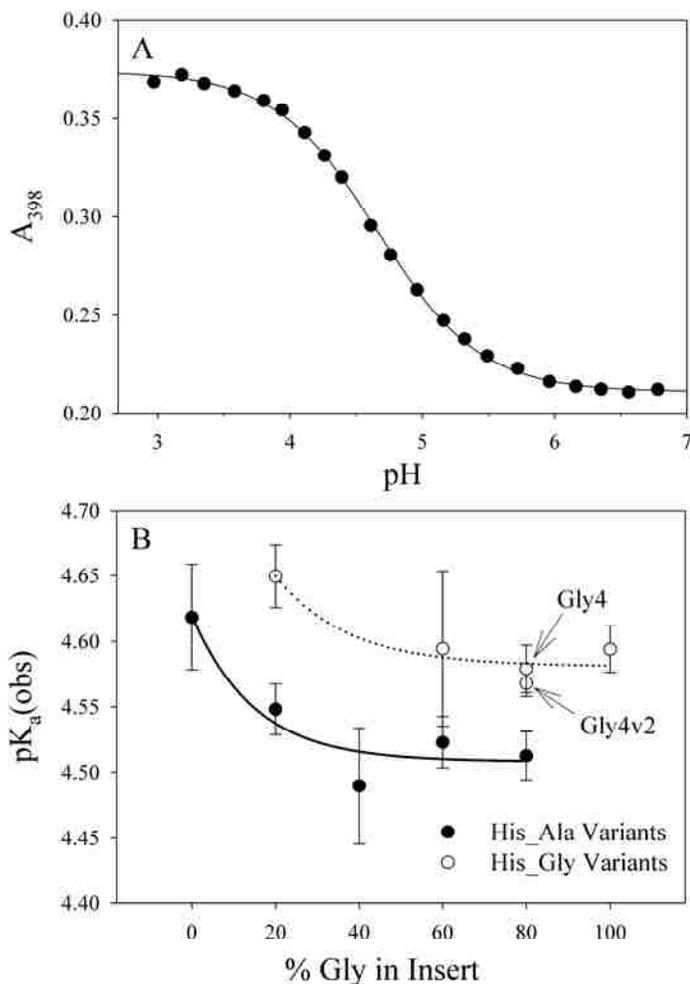


Figure 2.10: (A) A representative pH titration curve monitoring His-heme ligation in 3 M gdnHCl, 5 mM sodium phosphate, 15 mM NaCl, 1.0 mM EDTA. Data are for the Gly3v2 variant and were acquired at room temperature, $22 \pm 1^\circ\text{C}$. The solid curve is a fit of the data to a modified form of the Henderson-Hasselbalch equation. The parameters from the fit, $pK_a(\text{obs})$ and the number of protons, n , coupled to loop formation are reported in Table 2.5. (B) Graph of $pK_a(\text{obs})$ versus % Gly in Insert. Data are segregated based on whether Gly follows His, His_Gly variants (\circ), or Ala follows His, His_Ala variants (\bullet). The solid and dotted curves in panel B are meant to guide the eye and have no physical significance.

The thermodynamic parameters from the equilibrium loop formation experiments are presented in Table 2.5. Proton transfer numbers (n-values) are at or near the expected value of 1 based on a one proton process as depicted in Figure 2.1. In Figure 2.10B, the data are segregated based on whether Ala (His_Ala variants) or Gly (His_Gly variants) is next to the histidine involved in loop formation. In all cases, the pK_a(obs) is higher when Gly is the residue next to the histidine (and within the loop formed by histidine binding to the heme). This result indicates that chain flexibility immediately next to the histidine disfavors loop formation. Thus, the equilibrium data indicate that sequence immediately adjacent to the histidine that forms the loop (and within the loop) is important for loop stability.

Table 2.5: Thermodynamic parameters for denatured state loop formation at 22 ± 1 °C in 3 M gdnHCl for second generation Gly X variants of iso-1-cytochrome *c*

Iso-1-cytochrome <i>c</i> variant	pK_a(obs)^a	n^{a,b}
Gly1v2	4.65 ± 0.02	1.02 ± 0.10
Gly3v2	4.59 ± 0.06	1.09 ± 0.05
Gly4v2	4.57 ± 0.01	1.01 ± 0.01
Gly4v3	4.51 ± 0.02	1.03 ± 0.01

^aValues reported are based on an average of 3 separate experiments. The reported error is the standard deviation.
^bn is the apparent number of protons linked to breakage of the His-heme loop.

Furthermore, the apparent pK_a(obs) of the Gly 4 variant with 4 contiguous glycines is the same as that for the Gly4v2 variant with two sets of two contiguous

glycines (see arrows in Figure 2.10B). As seen in Table 2.1, these variants both have Gly next to the His involved in His-heme loop formation, controlling for the local sequence effect of the residue next to the histidine. Thus, the data for the Gly 4 and Gly4v2 variants do not support unusual extension of the backbone due to contiguous glycines. Variants Gly 3 and Gly3v2 were also designed to probe backbone hydration due to contiguous glycines. However, given the clear segregation of the $pK_a(\text{obs})$ data for the His_Ala *versus* the His_Gly variants, the difference in the $pK_a(\text{obs})$ for these two variants is more likely due to the Gly 3 variant having Ala next to His(-2) and Gly3v2 having Gly next to His(-2) (Figure 2.10B).

Interestingly, for both the His_Ala and His_Gly series of variants, increased glycine content in the insert initially confers greater loop stability (Figure 2.10B). However, above approximately 40% glycine in the insert, loop stability levels out. This trend is more pronounced for the His_Ala series than the His_Gly series.

2.3.5 Measurement of loop breakage kinetics

In previous studies,⁸⁹ we have shown that loop formation and breakage kinetics are consistent with a model involving a rapid protonation equilibrium (of histidine) followed by His-heme loop formation. This model predicts that k_{obs} has the pH dependence given by Eq 2.1,

$$k_{\text{obs}} = k_b + k_f \left(\frac{K_a(\text{HisH}^+)}{[\text{H}^+] + K_a(\text{HisH}^+)} \right) \quad (2.1)$$

where $K_a(\text{HisH}^+)$ is the ionization constant of the histidine involved in loop formation and k_f and k_b are the rate constants for loop formation and loop breakage respectively. Thus, if $\text{pH} \ll \text{p}K_a(\text{HisH}^+)$, k_b can be obtained. Table 2.6 summarizes the kinetic parameters from loop breakage for both new, as well as previously designed variants. We measured k_{obs} using downward pH jumps to both pH 3.5 and 3.0. The k_{obs} values are uniformly about 10% lower at pH 3.0 *versus* pH 3.5, consistent with the smaller contribution expected from k_f at lower pH (Eq 2.1). Thus, we use the k_{obs} values at pH 3.0 for k_b .

Table 2.6: Kinetic parameters for denatured state loop breakage and loop formation at 25 °C in 3 M gdnHCl for variants of iso-1-cytochrome *c*

Variant	Loop Breakage ^a		Loop Formation ^{a, b}
	$k_{\text{obs}}, \text{s}^{-1}$ (pH 3.50)	$k_{\text{obs}}, \text{s}^{-1}$ (pH 3.00)	$k_{\text{f}}, \text{s}^{-1}$
His Ala Variants^c			
NH5A	100.5 ± 1.0	93.0 ± 0.2	8910 ± 830
Gly 1	91.6 ± 1.8	83.1 ± 1.9	9360 ± 470
Gly 2	87.4 ± 0.9	76.8 ± 0.7	9900 ± 1000
Gly 3	92.1 ± 1.3	81.6 ± 1.0	9740 ± 460
Gly4v3	94.9 ± 3.0	87.2 ± 1.0	10670 ± 470
His Gly Variants^d			
Gly1v2	110.3 ± 3.6	101.8 ± 2.3	9080 ± 540
Gly3v2	109.5 ± 3.9	96.7 ± 3.1	9790 ± 1370
Gly4v2	114.0 ± 2.5	100.8 ± 3.8	10860 ± 480
Gly 4	116.2 ± 2.2	107.0 ± 1.8	11230 ± 500
Gly 5	122.0 ± 2.3	111.4 ± 1.3	11300 ± 480

^aThe errors reported for k_{obs} are the standard deviations of the mean and those for k_{f} are from standard propagation of errors.
^bLoop closure rate constants are beyond the limitation of our SX-20 stopped-flow apparatus. Thus, these values were extracted using pK_{loop} obtained as described in the text. k_{f} was calculated as $k_{\text{f}} = k_{\text{b}} \times K_{\text{loop}} = k_{\text{b}} \times 10^{-\text{pK}_{\text{loop}}}$, using k_{obs} for loop breakage at pH 3.00 for k_{b} .
^cVariants where the amino acid on the C-terminal side of the His which forms the loop with the heme is an alanine.
^dVariants where the amino acid on the C-terminal side of the His which forms the loop with the heme is a glycine.

Loop formation rates are much faster than the deadtime of stopped-flow instrumentation.⁸⁹ Thus, these rate constants were calculated by extracting the pK_{loop} (pK for loop formation with a fully deprotonated His) from the apparent $\text{pK}_{\text{a}}(\text{obs})$, using Eq 2.2

$$pK_a(obs) = pK_a(HisH^+) + pK_{loop} \quad (2.2)$$

This equation is a reasonable approximation if $pK_a(obs)$ is at least one unit less than $pK_a(HisH^+)$. Since $pK_a(HisH^+)$ equals 6.6 ± 0.1 in 3 M gdnHCl,⁸² the approximation is reasonable for the data presented here (Table 2.5, Fig. 2.10). pK_{loop} was then used in conjunction with the loop breakage rate constants, k_b , to extract loop formation rate constants (Table 2.6). Figures 2.11A and 2.11B show loop breakage and loop formation rate constants, respectively segregated based on the amino acid immediately adjacent to the histidine involved in loop closure. It is evident that loop breakage is faster if glycine is next to histidine. We note that with $pK_a(His) = 6.6 \pm 0.1$, Eq 2.1 indicates that the contribution of k_f to k_{obs} at pH 3 is about 2 to 3 s^{-1} . Thus, equating k_{obs} to k_b at this pH is a reasonable assumption.

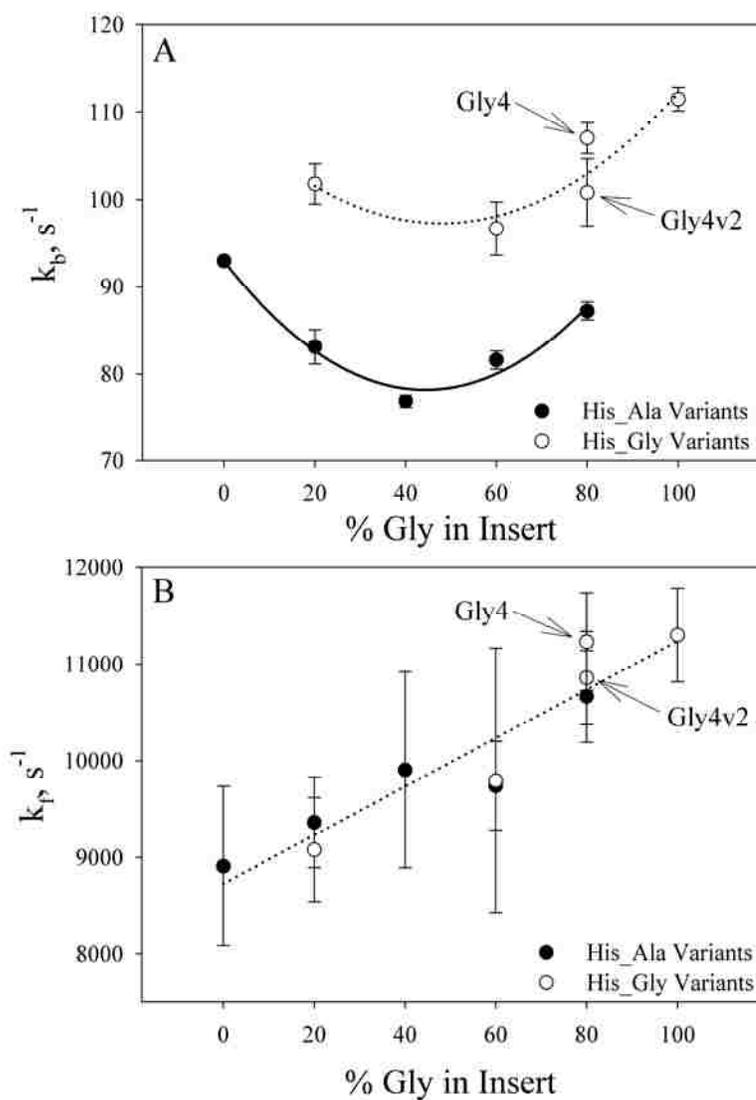


Figure 2.11: (A) Loop breakage rate constants, k_b , at 25 °C in 3 M gdnHCl plotted as a function of % Gly in Insert and grouped based on whether Gly follows His, His_Gly variants (○) or Ala follows His, His_Ala variants (●). The dotted and solid lines are meant to guide the eye and have no physical significance. (B) Loop formation rate constants, k_f , at 25 °C in 3 M gdnHCl plotted as a function of % Gly in Insert and grouped based on local sequence, His_Gly variants (○) or His_Ala variants (●), as in part A. The dotted line is meant to guide the eye and has no physical significance.

Figure 2.11A shows a decrease in loop breakage rate constants for inserts with low glycine content. For the His_Ala variants, the decreasing trend is more pronounced, reaching a minimum at 40% glycine content in the insert. Above this glycine content

there is a significant increase in loop breakage rate constants. The trend is similar for the His_Gly variants, except the decrease in k_b is less pronounced at low glycine percentage with a minimum at 40% glycine content, as well. Furthermore, k_b does not increase significantly until >60% glycine content in the insert, for the His_Gly variants. A comparison of the kinetics data for the Gly 4 and Gly4v2 variants (Table 2.6, Fig. 2.11A) suggests that the number of contiguous glycine residues adjacent to the histidine has a small but significant effect on k_b , as well. Figure 2.11B shows that k_f increases with increasing glycine content in the insert for both His_Ala and His_Gly combinations. In contrast to k_b , k_f depends within error, only on glycine content and not on the specific sequence.

2.4 Discussion

2.4.1 Effects of denatured state loop formation on overall protein stability

It is becoming increasingly evident that the denatured state of a protein can have an important impact on protein stability.^{35; 36; 38; 63; 65-67; 91; 99; 100; 136} In previous work, we have noted correlations between denatured state loop stability and global protein stability.^{80; 81; 124} All variants capable of forming a His-heme loop in the denatured state are strongly destabilized relative to the NK5A variant (Table 2.2) which forms the much less favorable N-terminal amino group-heme loop in the denatured state. How much of this destabilization can be accounted for by differences in denatured state loop stability? The stabilities of the NH5A and Gly X variants are all within error the same, with an average stability of 1.98 kcal/mol, which is 2.23 kcal/mol less stable than the NK5A

variant (Table 2.2). At pH 7, where global stability is measured, both types of loops are fully or almost fully formed. Thus, the difference in denatured state stability due to loop formation can be evaluated as $\ln(10)RT\Delta pK_a(\text{obs})$. The average $pK_a(\text{obs})$ value for the NH5A and Gly X variants is 4.56 (Table 2.4). Thus, relative to the less stable loop formed by the NK5A variant, the stabilization of the denatured state for the NH5A and Gly X variants is 2.24 kcal/mol. Thus, denatured state loop stability completely accounts for the decrease in $\Delta G_u^{\circ}(\text{H}_2\text{O})$ for the NH5A and Gly X variants relative to the NK5A variant. While denatured state His-heme loop formation is specific to heme proteins, denatured state metal ligation may be a general modulator of metalloprotein stability that must be considered.

As discussed above, the decrease in stability of all variants relative to a variant with the five alanine insert but Lys(-2) instead of His(-2) (NK5A, $\Delta G_u^{\circ}(\text{H}_2\text{O}) \sim 4$ kcal/mol)⁸³ could be attributed completely to stabilization of the denatured state by His-heme loop formation. For the second generation variants, the dominant effect on stability relative to the NK5A variant is still the very stable His-heme denatured state loop. The differences in the stability of variants with the same number of glycines in some cases are consistent with changes in denatured state loop stability and in other cases are not. Thus, some of the changes in protein stability may reflect effects on the native state.

2.4.2 Effects of sequence composition on loop formation in the denatured state for first generation variants

Perhaps what is most surprising is the observation that the $pK_a(\text{obs})$ is identical for the NH5A and the Gly 5 variants. The increase in overall glycine content from 9%

(NH5A) to 32% (Gly 5) in the 22 residues extending from the histidine, which forms the loop to the heme, should significantly decrease the average end-to-end separation of the histidine and the heme. The Jacobson–Stockmayer equation^{84; 85} (Eq (2.3)) is typically used to predict loop formation equilibria:

$$\Delta S_{loop} = -v_3 R \ln(n) + R \ln[(3 / 2\pi C_n l^2)^{v_3} V_i] \quad (2.3)$$

where n is the number of monomers in the loop, R is the gas constant, C_n is Flory's characteristic ratio, which is sensitive to chain flexibility,⁸⁶ l is the distance between monomers, V_i is the approach volume of the atoms involved in loop formation and v_3 is the scaling exponent for loop formation. Using the Jacobson–Stockmayer equation it can be shown that $\Delta pK_a(obs)$ for the Gly 5 variant *versus* the NH5A variant is given by Eq (2.4):

$$\Delta pK_a(obs) = pK_a(obs)_{Gly5} - pK_a(obs)_{NH5A} = \left(\frac{-v_3}{\ln(10)}\right) \left[\ln\left(\frac{C_{n,NH5A}}{C_{n,Gly5}}\right)\right] \quad (2.4)$$

If we use $v_3 = 2.1$ (average of theoretical values for a random coil with excluded volume),^{38; 88} then set the Flory characteristic ratios for the NH5A variant (9% Gly, 22 residue loop) to 5.5 and for the Gly 5 variant (32% Gly, 22 residue loop) to 3.5,⁸⁶ the expected ΔpK_a would be -0.41. Our experimental results clearly are counter to this prediction, the question then is why?

The peptide systems used to monitor the kinetics of loop formation allow much more freedom in the stereochemistry of loop formation than is true for our histidine–heme loop formation system. Histidine–heme binding requires axial docking of the histidine ligand to the iron in the heme and this constraint would be expected to seriously affect formation of short loops. Since our set of variants forms loops from the N-terminal side of the heme (see Figure 2.1),⁸¹ the heme excluded volume effect, which causes a decrease in the favorability of loop formation for loops <35 residues formed from the C-terminal side of the heme,⁸⁰ is not a factor. Also, the favorability of loop formation is higher for the 16 residue *versus* the 22 residue N-terminal histidine–heme loop (Table 2.4), indicating that chain stiffness is not a dominant factor at this loop length. Thus, for the set of 22 residue loops studied here, heme excluded volume and chain stiffness are unlikely to influence the observed result. The clustering of glycine residues near the histidine which forms the loop could also influence loop formation. However, the observations of Robinson and Sauer¹¹⁵ argue against this possibility. Their work shows that the overall percentage Gly composition of the linker in single chain Arc repressors, and not its specific sequence, is the dominant factor controlling stability and folding kinetics.

The assumption implicit in the Jacobson–Stockmayer treatment of loop formation is that the entropy decrease can be approximated by the required volume reduction to form a loop given a Gaussian distribution of conformations (see Eq (2.3)). Thus, a more compact chain should form loops more readily. However, this approximation may not account well for the larger density of conformations in a more flexible polypeptide chain with high glycine content. Comparison of rates of loop formation for polyserine *versus*

poly(GlySer) peptides show that the glycine-containing sequences have rates of loop formation that are faster by a factor of 2 for loop sizes from 3 to 12.¹¹¹ The $\log(k)$ versus $\log(n)$ curves for both are decreasing linearly by $n = 12$, and while data are only available for poly (GlySer) for larger loop sizes the factor of 2 difference looks to persist to larger loop sizes based on extrapolation of the polyserine data. Krieger *et al.*¹¹¹ noted that the enhanced rate of loop formation for poly(GlySer) sequences is smaller than might be expected given the greater flexibility and shorter end-to-end distance expected for poly(GlySer) versus polyserine peptides, likely as a result of the greater conformational space available to glycine-containing sequences. In a more recent study¹¹⁰ a single glycine was embedded into polyserine sequences capable of forming loops of 4 to 12 residues progressively dropping glycine content from ~ 25% to ~ 8%. Rates of loop formation for this set of peptides move smoothly from the faster poly(GlySer) to the slower polyserine rate versus loop length curve. Thus, we expect that rates of loop formation should increase smoothly going from the NH5A (9% Gly) to the Gly 5 (32% Gly) variant. The invariance in the equilibrium $pK_a(\text{obs})$ values from Table 2.4 suggests that the rate of loop breakage increases in nearly equal measure with the rate of loop formation. Thus, the greater flexibility speeds loop breakage, as well, which presumably reflects the greater number of ways the histidine can move away from the heme after the Fe^{3+} -histidine bond breaks.

Recent studies on polyglycine sequences using residual dipolar NMR and small-angle X-ray scattering methods have provided evidence that polyglycine segments prefer elongated conformations that are in fact more extended than polyalanine.¹²² Stiffening of the backbone by hydration provides a possible explanation. The low sensitivity of

$pK_a(\text{obs})$ to replacement of alanine with glycine could also result from polyglycine sequences being more elongated than previously thought. Thus, we carried out kinetic experiments on loop formation and breakage with the NH5A and the first and second generation Gly X variants to determine whether the insensitivity of $pK_a(\text{obs})$ to the percentage of Gly in the insert is due to compensating effects, of both faster loop formation and breakage, as opposed to polyglycine sequences being more elongated in aqueous solution than previously thought.

2.4.3 Does hydration or local sequence sterics make polyGly segments unexpectedly extended?

If local sequence sterics or backbone hydration, due to contiguous glycines, results in polypeptide chain extension in the denatured state, then the denatured state of our Gly X variants should be more expanded than expected.¹²² If this is true, one might expect glycine percent in our insert to have a minimal effect on k_f . However, we observe an approximately 25% increase in k_f going from 0% to 100% glycine in our insert. Also, if hydration or local sequence sterics cause poly-Gly segments to be extended, one might expect slower loop formation for contiguous *versus* non-contiguous glycines in the insert. Our observation that Gly 3 ($9740 \pm 460 \text{ s}^{-1}$) and Gly3v2 ($9790 \pm 1370 \text{ s}^{-1}$) variants have similar values for k_f suggests otherwise. Similarly, the values of k_f are within error for the Gly 4 variant ($11230 \pm 500 \text{ s}^{-1}$) which has 4 contiguous glycines and Gly4v2 ($10860 \pm 480 \text{ s}^{-1}$), which has the four glycines separated into two groups of two contiguous glycines by an alanine.

Thus, we see no clear evidence that hydration or local sequence sterics of poly-Gly segments causes such sequences to be as extended or more extended than poly-Ala segments.¹²² Rather, as with previously published experimental and computational studies, the polyglycine stretches in our variants appear to be flexible enough to cause compaction of the denatured state permitting more rapid loop formation.^{92; 111}

2.4.4 Effect of the sequence immediately adjacent to histidine on loop dynamics

Our data indicate that the amino acid adjacent to the histidine involved in His-heme loop formation has a significant effect on loop formation. In our equilibrium data (Figure 2.10B) the stability is uniformly lower (i.e., higher $pK_a(\text{obs})$) for loops with a glycine next to the His which forms the loop than for those loops with an Ala next to the His which forms the loop. Comparison of loop formation rate constants, k_f , and loop breakage rate constants, k_b , (Figure 2.11) shows that the effect on equilibrium loop stability of the amino acid within the loop that is next to the histidine can be attributed primarily to k_b . The loop breakage rate constants, k_b , for the His_Gly variants are clearly larger than those of the His_Ala variants. The rate constants for loop formation, k_f , on the other hand are not sensitive to the amino acid next to the histidine (Figure 2.11B). Since glycine is more flexible than alanine, it is likely that the faster breakage rates for the His_Gly variants *versus* the His_Ala variants are due to lower main chain rotational barriers for glycine compared to alanine.¹¹⁰ The greater main chain flexibility of glycine residues likely allows the histidine ligand to swing away from the heme iron faster after the His-heme bond breaks than when alanine is next to the histidine. Although the effect is subtle, the enhanced flexibility effect seems to extend beyond the residue next to the

His involved in loop formation. k_b for Gly 4 ($107.0 \pm 1.8 \text{ s}^{-1}$, 3rd residue from the His is Gly), is slightly faster than that of Gly4v2 ($100.8 \pm 3.8 \text{ s}^{-1}$, 3rd residue from the His is Ala).

2.4.5 Sequence composition effects on the loop formation rate constant, k_f

As glycine content in a polypeptide chain increases, Flory's characteristic ratio for an infinite chain, $C_n(\infty)$, decreases from a value of ~ 9 for alanine (and most other amino acids except proline) to ~ 2 for pure glycine.⁸⁶ $C_n(\infty)$ initially decreases rapidly as % Gly increases, dropping from 9 to 4 as % Gly increases from 0 to 30. The change is much more gradual above 30% Gly. As $C_n(\infty)$ decreases, the random coil is expected to be more compact and the rate of contact between two monomers separated by a given number of residues should increase. Thus, the increase in the rate of contact formation should be most pronounced as % Gly content increases from 0% to 30%.

Kiefhaber and coworkers,^{108; 110; 111} observed that poly(Ser) peptides (% Gly = 0, $C_n(\infty) = 9$) have about 2- to 3-fold slower contact rates compared to poly(Gly-Ser) peptides (% Gly = 50, $C_n(\infty) = 3$) for loop sizes ranging from 3 to 12 residues. For a set of peptides with a single Gly in the middle of a poly(Ser) chain ranging from 4 (% Gly = 25) to 12 residues (% Gly = 8.3), the rate constant for loop formation moved progressively from near coincidence with the rate *versus* loop size curve observed for poly(Gly-Ser) to coincidence with that for poly(Ser). While a % Gly series at a single loop size was not done in this study, the results suggest that a progressive decrease in the rate constant for loop formation should be observed as % Gly decreases at a single loop size.

There are several key differences between our system and that of Kiefhaber and co-workers. First, His-heme loop formation is reaction limited not diffusion limited,^{89; 137} Thus, our values of k_f do not provide a measure of the rate of first contact. The magnitude of k_f does still scale with the end-to-end distance distribution of the chain and thus will reflect compaction of the chain due to increased chain flexibility. Second, the characteristic ratio, C_n , varies considerably with chain length for short chain lengths and the chain length at which it approaches $C_n(\infty)$ varies considerably with % Gly.⁸⁶ For poly(Ala), C_n reaches 90% of $C_n(\infty)$ at a chain length of 64 whereas C_n reaches 90% of $C_n(\infty)$ at a chain length of ~ 16 for poly(Gly-Ala) (% Gly = 50). Thus, C_n is not the same for the equivalent % Gly for our 22-residue loop and for Kiefhaber's data for the chain lengths where poly(Gly-Ser) and poly(Ser) can be compared (chains of 3 to 12 residues). Finally, our insert is placed near the N-terminus of iso-1-cytochrome *c* in a segment of the protein with irregular and dynamic structure that can readily accept inserts.¹²⁰ Thus, unlike Kiefhaber's system the glycines are not evenly distributed throughout the sequence. The 22 residue loop we study here is attached to the heme at Cys14 and there are glycines in the natural sequence at positions 1 and 6. With the six amino acid insert, the Histidine is at position -8 (horse numbering) and the insert runs from positions -3 to -7. Flory's theoretical treatment assumes an even (random) distribution of glycines, whereas in our system as we increase % Gly, the glycine content becomes skewed toward the N-terminus of our 22 residue loop.

It is evident from our results, that even with the skewed distribution of glycine, k_f increases monotonically as the glycine content increases (Figure 2.11B, Table 2.6). The magnitude of k_f increases by a factor of ~ 1.3 going from the NH5A variant (% Gly = 9)

to the Gly 5 variant (% Gly = 32). Using Flory's results on the effects of % Gly and chain length on C_n , we can estimate C_n at a chain length of 22, $C_n(22)$.⁸⁶ For the NH5A variant, $C_n(22) \sim 5.5$ and for the Gly5 variant $C_n(22) \sim 3.5$. Using the Jacobson-Stockmayer equation,^{85; 138} we estimated that this change in $C_n(22)$ would lead to a decrease in the $pK_a(\text{obs})$ of 0.41 units⁸³ which corresponds to an increase by a factor of ~ 2.6 in the equilibrium constant for loop formation. If we make the simplifying assumption that k_b is independent of sequence composition for a non-interacting random coil, then k_f should increase by a factor of ~ 2.6 for the Gly 5 variant *versus* the NH5A variant. If we account for the observed changes in k_b (Table 2.6), an increase in k_f by a factor of ~ 3.1 would be expected. Thus, the skewed distribution of glycine in our insert dampens the decrease in the end-to-end distance distribution of the Gly 5 variant relative to the NH5A variant compared to what would be expected for an even distribution of glycine in the loop. However, even for this skewed glycine distribution, compaction of the chain by the more flexible glycine is not completely eliminated.

2.4.6 Sequence composition effects on the loop breakage rate constant, k_b

The most notable observation in these data is the degree to which k_b varies. There is a factor of 1.45 difference between the largest and the smallest k_b (pH 3.00 data in Table 2.6). This variation is larger than that for k_f , indicating that k_b is more important in controlling the loop formation equilibrium.

As noted above, k_b is larger for the His_Gly than for the His_Ala series of variants. There is a reasonably constant factor of 1.20 ± 0.03 increase in k_b for the His_Gly *versus* the His_Ala variants for all values of % Gly in the insert (Figure 2.11A).

This ~20% increase in k_b is likely due to lower rotational barriers about the main chain of glycine, as discussed earlier. However, the unusual dependence of k_b on % Gly in the insert indicates that there is more to the effect of glycine on chain dynamics than main chain rotational barriers. In both the His_Ala and His_Gly series of variants, k_b initially decreases. This observation is counter to what might be expected if glycine acted only to increase conformational space and the rate at which it is explored (rotational barriers).¹¹⁰ The initial decrease in k_b also argues against nucleation of α -helical structure when histidine binds to the heme as seen in several peptide model systems,¹³⁹⁻¹⁴¹ since glycine is a helix breaker.¹⁴² The observed decrease in k_b for the first two Ala→Gly replacements suggests that the conformational flexibility of glycine allows the polypeptide chain to relax to a more stable conformer (perhaps allowing better interactions with the heme) than is possible with the all Ala insert. It is also possible that the first two glycines relieve strain in the loop, thus slowing loop breakage (see Chapter 3).

The increase in k_b for inserts with more than two glycines might be attributable to the increased rate of exploration of conformational space overtaking the effects of reduced loop strain or of conformational relaxation of the closed loop form of the chain to conformations with better stabilization by van der Waals interactions. It is also possible that decrease in the net stabilization possible by van der Waals interactions in the closed loop for the small glycine side chain *versus* the larger alanine side chain contributes to the increase in k_b as the number of glycines in the insert increases.

2.4.7 Compensation between k_f and k_b

In Figure 2.10B for both the His_Gly and the His_Ala variants, the $pK_a(\text{obs})$ decreases initially, but as the number of glycines in the insert increases above 2, $pK_a(\text{obs})$ levels out. The data in Figure 2.11 shows that both k_b and k_f are increasing as the $pK_a(\text{obs})$ levels out. Thus, at higher % Gly in the insert, the increases in k_f and k_b exactly compensate for each other such that the equilibrium for loop formation remains unchanged. For the range of overall percent glycine within the 22-residue loops studied here (9 to 32 %), this observation is inconsistent with the predictions of the Jacobson-Stockmayer equation.^{85; 138} This leveling out in $pK_a(\text{obs})$ is not expected to occur until above 30% glycine content. This inconsistency is likely due to the skewed distribution of glycines in our insert.

However, our observation of this compensation between k_f and k_b points out an interesting aspect of the effects of increased polypeptide chain flexibility (decreased C_n) on the probability of loop formation as given by the Jacobson-Stockmayer equation. As the chain becomes more flexible, initially, compaction of the chain dimension increases the probability of loop formation for a given loop size. However, it appears that as conformational space increases with increased glycine content, the time that any given contact can persist diminishes due to the proliferation of nearby conformations when the contact is broken. Compaction and diminished persistence appear to compensate nearly exactly at high flexibilities. Thus, the advantage conferred by increased glycine content for rapid formation of contacts and compaction of a polypeptide so that it can fold, saturates rapidly due to decreased contact persistence.

2.4.8 Guiding the initial stages of folding

The observation that loop breakage rates play a dominant role in modulating loop equilibria is consistent with our previous results on the kinetics of His-heme loop formation for a set of His-heme loops in the denatured state of iso-1-cytochrome *c* ranging in size from 16 to 83 amino acids.⁸⁹ We found that loop breakage rates varied by a factor of 5 and were primarily responsible for the deviations from random coil behavior for equilibrium loop formation in the denatured state in 3 M gdnHCl.

Rates of loop formation are sensitive to sequence composition, but appear to be insensitive to sequence order (Figure 2.11B). This observation argues that the speed at which contacts form is unimportant to the “folding code”. Rather what appears to be important to the “folding code” is whether a contact persists once it forms. Thus, the key to the “folding code” may be in understanding how local sequence order modulates rates of breakage of contacts as a protein folds.

From the perspective of a folding funnel, it is likely that many contacts form and break rapidly near the lip of the funnel. Early in folding, it will be the contacts with a greater tendency to persist that will begin the descent down the funnel. Experimental data for poly(Ser) or poly(Gly-Ser) peptides indicates that loops ranging in size from 3 to 11 residues form on 5 to 25 ns timescales. Protein folding at its fastest occurs on an ~ 1 μ s time scale.^{9; 10} Thus, simple loop formation is not what limits folding efficiency. Our data indicate the key to efficiency and specificity of folding is in the relative persistence of early contacts. Thus, in considering a speed limit for folding, it is not just the forward rates of loop formation that must be considered but also the backward rates that allow these primitive contacts to persist.³⁷

2.5 Conclusion

Experimentation on protein-denatured states can still be considered a “frontier” area of research. There are many factors that influence protein folding, some of which are still not well understood. Characterization of the stabilizing interactions and structures in the denatured state are considerably more difficult to assign than in the native state. Using poly-Ala/Gly inserts near the N-terminus of iso-1-cytochrome *c*, we were able to observe significant effects of local sequence on the kinetics and equilibria of loop formation. In particular the amino acid – Ala *versus* Gly – next to the histidine which forms the loop has a large effect on loop equilibria. The rate of loop breakage appears to play a dominant role in this effect. Our data indicate that increasing glycine content – even when skewed to one end of the residues involved in loop formation – increases the rate of loop formation. This observed increase in k_f does not provide support to the possibility that the main chain of poly-Gly sequences are more extended than previously thought due to backbone hydration or main chain sterics. A number of studies suggest that rapid conformational equilibration occurs during protein folding.^{77;}
^{143; 144} Thus, both speed of contact formation and the persistence of a contact may be important in selecting which structures form during folding. Our current findings support this idea since, unlike the rate of loop breakage; the rates of loop formation are insensitive to local sequence effects. This observation suggests that the key to deciphering the “folding code” may lie in defining which of the contacts that can form actually persist, rather than in trying to discern what factors control how fast these contacts form.

CHAPTER 3

THERMODYNAMICS AND KINETICS OF LOOP FORMATION WITH HOMOPOLYMERIC POLYALANINE SEQUENCES FROM THE N-TERMINUS OF ISO-1-CYTOCHROME C IN POOR AND GOOD SOLVENT CONDITIONS

3.1 Introduction

Chemical denaturants such as urea and guanidine hydrochloride (gdnHCl) have been used to unfold proteins since the early 1900's.¹⁴⁵ They were found to work by stabilizing the denatured state *versus* the native state.^{146; 147} Their exact mode of action is still debated. However, Charles Tanford's experimental results^{33; 42; 148} led him to hypothesize that the favorable unfolded population in a denaturant is due to the higher transfer free energies of the numerous amino acids into denaturant *versus* water. In other words, more regions of the protein's polypeptide chain are soluble in these denaturants. This model has been termed "The Transfer Model" and is widely accepted. Another school of thought is from the works of John Schellman who hypothesized that the unfolding process is primarily due to the available denaturant binding sites on a protein. He reasoned that the unfolded state would have a lot more available denaturant binding sites – generally thought to be the protein backbone¹⁴⁹ – than the compact, native state of the protein.^{46; 132; 150; 151} Unfolding would occur *via* a mass action effect when these denaturants are present. This model is known as the "Binding Model" and has garnered a lot of support in recent years. Experimentally, it is observed that the transition region between the folded and unfolded states of a protein has a linear dependence on denaturant concentration. Both the "Transfer Model" and the "Binding Model" are able to explain the cooperativity of this linear transition phase in the observable region, going from

native to denatured state, during an unfolding experiment.¹⁵² This observation has fostered the usage of linear extrapolation methodology (LEM)^{132; 153; 154} to obtain thermodynamic information from unfolding experiments. Thus, the use of chemical denaturants has positively impacted our ability to probe the nature of the denatured state.

Hence, we are able to use gdnHCl as a perturbant in assessing the thermodynamic stability of homo and heteropolymer denatured states in this Chapter. However, under “denaturing conditions” a protein can have different distributions of various denatured state ensemble (DSE) conformations. The more common DSE conformations are the extended and the compact states. The degree of compaction that exists in the DSE is due to the number of hydrophobic interactions from various nonpolar amino acid residues^{87; 91} and the effect the solvent has on the polymer chain. This interaction is what is responsible for the well known “hydrophobic collapse” in the initial stages of protein folding.^{1; 8} We know from polymer theory and solvent denaturation studies that placing a polymer in different solvents can enable an increase in the extended DSE over the compact DSE.^{33; 42; 85; 87; 91} The degree to which a solvent can accomplish complete solvation of the polymer categorizes it as one of three solvent types – Theta, Poor or Good Solvent.^{87; 91; 92} The solvent classification is based on the statistical behavior of the average polymer chain radius as a function of the number of amino acids in the chain.⁹¹ At the molecular level, this statistical behavior is a direct result of the interplay between local interactions (residue-residue, solvent-residue) and nonlocal interactions (excluded volume effects). A “Theta” solvent is an unfavorable solvent in which the local interactions of the polymer’s monomer units, which normally lead to compaction of the chain, exactly cancel out the non-local interactions of excluded volume and steric

repulsion, which normally leads to expansion of the chain. Therefore, the polymer chain adopts an unperturbed distribution and appears to be governed only by local interactions. The polymer chain in this solvent is said to have “random-flight” dimensions.⁹¹ A “Poor” solvent is one in which the local monomer-monomer attractive interactions, of the polymer chain, dominate the local solvent-monomer interactions. This causes a solvent-driven contraction which dominates the nonlocal excluded volume expansion, leading to a compact polymer chain. The size of the polymer in this solvent is said to be less than “random-flight” dimensions.⁹¹ Lastly, a “Good” solvent is one in which the local solvent-monomer interactions are more favorable and hence, dominate the local monomer-monomer interactions. This results in the nonlocal excluded volume expansion interactions dominating the solvent-driven contraction interactions, leading to an expanded polymer chain.^{87; 91} The polymer chain is said to have more than “random-flight” dimensions. Water is considered a very poor solvent, while 0-4 M gdnHCl and urea are considered poor solvents. In order to achieve good solvent conditions, the concentrations of the denaturants must be at least 6 M and 8 M for gdnHCl and urea, respectively.⁸⁷

Investigations in this section focus on subjecting homo and heteropolymeric protein sequences to poor and good solvents. The Jacobson-Stockmayer equation^{84; 85} (Eq (3.1)) below is commonly used to predict loop formation equilibria under these denaturing conditions, where proteins are expected to exist as random coils.

$$\Delta S_{loop} = -v_3 R \ln(n) + R \ln[(3 / 2\pi C_n l^2)^{v_3} V_i] \quad (3.1)$$

where n is the number of monomers in the loop, R is the gas constant, C_n is Flory's characteristic ratio, which is sensitive to chain flexibility,⁸⁶ l is the distance between monomers, V_i is the approach volume of the atoms involved in loop formation and ν_3 is the scaling exponent for loop formation. In Chapter 2, we focused on Flory's characteristic ratio C_n , to assess the flexibility effect of increasing glycine content in a fixed loop size. As mentioned previously, loop formation is affected by three factors. First, chain stiffness due to steric bulk can increase C_n and thus the stiffness of a loop. To relieve this chain stiffness we can decrease the steric bulk of the individual monomer units.⁸³ For short chain lengths, chain stiffness can also prevent loop formation. This constraint can be relieved by increasing loop size. There is a caveat to increasing the length of the chain because it is coupled to the second factor that affects loop formation – loop entropy. Loop entropy is not manifested in the flexibility of a polymer chain until the chain reaches a sufficiently long length. After this threshold is met, the entropy of a very large polypeptide chain begins to influence the ability for the chain to make end-to-end contacts or residue-residue contacts. The third factor that affects loop formation is residual structure. This factor does not come from polymer theory but is a physically restrictive interaction within a polymer chain that prevents it from behaving like a freely jointed random coil. Such residual structure, if present, reduces the conformational space of a polymer chain thereby allowing quicker contact formation between sequentially distant monomers in the chain. In the case of a protein's folding polypeptide chain, this can be a good thing. Another informative component of the Jacobson-Stockmayer equation is the scaling factor, ν_3 , which indicates whether a polymer chain behaves like a freely jointed “unperturbed” random coil. For a freely jointed random coil, the scaling

factor is 1.5.^{37; 84; 85; 87} When excluded volume is taken into account, the value of v_3 increases to a range of 1.8 to 2.4.³⁷ Studies on other proteins and DNA indicated that under denaturing conditions, these natural polypeptide chains behave as a random coil with excluded volume having scaling factors in the range of 1.5 to 2.1.^{111; 155-157} Previous work on simple loops which form from the C-terminus of iso-1-cytochrome *c* suggested that the natural sequence has a lot of residual structure because scaling factors were on the order of ~ 4 to ~ 4.5 in poor solvents and only really decreased to ~ 2.0 when subjected to good solvent conditions (5 – 6 M *gdnHCl*).^{80; 82} Work presented in this chapter provides additional insight into these previous observations.

Additionally, work presented in this chapter will shed some light on previous work done by Kiefhaber and co-workers using triplet-triplet energy transfer on host-guest peptides, which indicated that contact formation over short distances ($N < 5$) does not depend on the loop size. Rather, the limiting factor in loop contact formation was the nature of the amino acid residues in the chain. Less bulky glycines produced more flexible chains which had faster contact formation times compared to chains with more bulky residues such as prolines.¹¹¹ However, the rates of contact formation for longer loops ($N > 30$) were found to decrease with increasing loop size and demonstrated behavior expected for a random coil.¹¹¹

Therefore, working with homopolymeric and heteropolymeric sequences will shed some light on why previous work in the lab was not in accord with the findings on scaling factors of other groups. Subjecting homopolymeric and heteropolymeric sequences to different solvent conditions will enable us to determine the degree of compactness of their denatured states as well as make comparisons of polymer theory to

the behavior of natural foldable polymers, specifically proteins. The homopolymeric protein models used were fashioned by engineering five-alanine repeats at the N-terminal end of yeast iso-1-cytochrome *c*. The initial five-alanine insertion is between F(-3) and K(-2) of the iso-1-cytochrome *c* sequence. For solubility reasons, a lysine residue was added to each five-alanine insert; thus, each insert is of the form KAAAAA, which results in a six-residue extension. All insertions are in a relatively disordered region of the protein as previously described (see Figure 2.2).⁸³ Therefore, mutations in this region are not expected to have significant effects on the overall structure or stability of the native fold. Conversely, the heteropolymeric protein models used for comparison were based on loops formed from the natural sequence of yeast iso-1-cytochrome *c*⁸² and cytochrome *c*' using an engineered histidine along the natural sequence of the proteins for His-heme loop ligation in the denatured state. The homopolymeric loops ranged in size from 16 to 46 residues, while the heteropolymeric counterparts were 10 to 111 residues for cytochrome *c*' and 9 to 83 residues for iso-1-cytochrome *c*. Under good *versus* poor solvent conditions, we aimed to see a weakening of our simple loops indicating loss of residual structure. The differences in the resulting nature of the denatured state will be assessed thermodynamically through equilibrium loop formation measurements for both classes of polymers. Additionally, we extend our investigations into the kinetics of the homopolymeric loops under both poor and good solvent conditions through stopped-flow loop breakage measurements. We have seen previously, from Chapter 2 that the rates of loop formation (closure) for simple loops are much faster than can be measured with a stopped-flow apparatus. However, under low pH

conditions, these loop formation rates (k_f) can be extracted using the equilibrium constant for loop formation ($K_{loop} = k_f/k_b$) since only k_b contributes to $k_{(obs)}$ at low pH.⁸³

3.2 Materials and Methods

3.2.1 Mutagenesis of NH5A-X inserts

The NH5A variant was produced as previously described in Section 2.2.1. The subsequent variants, NH5A-2, NH5A-3 and NH5A-4, were made using the unique restriction site elimination method¹²³ as previously described.¹²⁴ NH5A-2 was made using single-stranded pBTR1 vector DNA carrying the NH5A variant of iso-1-cytochrome *c* as the template and the following primer: 5'-d(GAACCGGCCTTCGCCGCCGCCGCCGCTTTAGCAGCAGCAG)-3'. NH5A-3 was made using single-stranded pBTR1 vector DNA carrying the NH5A-2 variant of iso-1-cytochrome *c* as the template and the following primer: 5'-d(CAGCAGCAGCCTTGGCGGCGGCGGCGGCGTGGGAATTCAG)-3'. NH5A-4 was made using single-stranded pBTR1 vector DNA carrying the NH5A-3 variant of iso-1-cytochrome *c* as the template and the following primer: 5'-d(GGCGGCGGCGGCTTTTGGCGCAGCGGCTGCGTGGGAATTCAGT)-3'; a mispriming of the oligonucleotide primer resulted in a six-alanine insert instead of the required five-alanine insert, thus, the following primer was used to remove the additional alanine: 5'-d(CCTTGGCGGCGGCGGCGGCTTTTGCC)-3'. Insertion sequence is underlined in each oligonucleotide.

Due to difficulty with ssDNA-based mutagenesis of GC-rich sequences, a PCR based mutagenesis system was pursued for continued insertions. The NH5A-5 variant was made by designing primers using Stratagene primer design criteria for the QuikChange II PCR based mutagenesis kit. The forward primer used was 5'-(CTTCGCTGCGGCAGCGCGGTGGAATTCAGTCAT)-3', while the reverse primer used was 5'-(GCGGCTGCCGCAGCGAAGGCCGCTGCGGCA)-3'. Underlined sequences are the anchoring sequences. The insertion sequences are the rest of the primers. Modifications made to the standard PCR protocol were (i) changing the annealing temperature from 55 °C to 42 °C – making it the same as the primers melting temperature and (ii) increasing the primers' "anchor" sequence to 15 base pairs. PCR product was transformed into the *Escherichia coli* TG-1 cell line and DNA was extracted from overnight cultures using a DNA Wizard Plus Miniprep kit from Promega. Extracted DNA's were initially screened via *Bam*HI restriction digests and then subjected to gel electrophoresis on a 1% agarose gel using *Bam*HI digested NH5A-4 template as a control. Promising mutagenic sequences were confirmed using standard DNA sequencing methods for those DNA samples that had longer retention times on the agarose gel, compared to the control.

3.2.2 Mutagenesis of cytochrome *c'* variants

Pseudo wild type (pWT, Q1A mutation) was obtained from Jay Winkler, at the Beckman Institute, California Institute of Technology, in the pETcp expression vector.¹⁵⁸ The cytochrome *c'* D3H variant was made using Stratagene primer design criteria for the QuikChange II PCR based mutagenesis kit with pWT cytochrome *c'* as the template and

the following primers: Forward Primer – GGTCGCGGCGACCCCAGTGATTGCGCAGC and Reverse Primer - GCTGCGCAATCACGTGGGTCGCCGCGACC. Mutations are underlined in the oligonucleotide sequence. Mutagenesis reactions were carried out as described by Stratagene’s QuikChange II protocol. The only modification made to the standard PCR protocol was to change the annealing temperature from 55 °C to 68 °C – making it the same as the extension temperature. PCR product was transformed into the *E. coli* TG-1 cell line and DNA was extracted from overnight cultures using a DNA Wizard Plus Miniprep kit from Promega. Extracted DNA’s were sequenced (Murdock Sequencing Facility, University of Montana) to confirm the desired mutagenic sequences. All other variants (unpublished data – Dr. Sudhindra Rao) were prepared in the exact same manner as described above.

3.2.3 Protein expression and purification of poly(Ala) cytochrome *c* variants

Iso-1-cytochrome *c* poly(Ala) variants were expressed and purified as previously described in Section 2.2.4 above.

3.2.4 Protein expression and purification of cytochrome *c*’ variants

Confirmed DNA sequences for the pWT, A104H, D3H and K31H cytochromes *c*’ in the pETcp vector were transformed into Novagen BL21-DE3 competent cells along with the pEC86 vector.¹⁵⁹ The pEC86 vector contains the *ccm*ABCDEFGH genes required for covalent attachment of heme in the periplasm of *E. coli*.¹⁵⁹ The vector also contains the gene for chloramphenicol resistance. Selection was obtained by incubation

in the presence of ampicillin and chloramphenicol antibiotics on LB-plates. Protein expression was carried out by suspending cells on the plate with 4 mL of L-broth and using this to inoculate 1 L of Terrific Broth rich media containing ampicillin and chloramphenicol. The 1 L culture was grown for 22 hrs. at 37 °C with shaking at 150 rpm. Cells were then harvested using a GS-3 rotor in a Sorvall RC 5C Plus centrifuge at 5000 rpm for 10 minutes.

Harvested cell pellets were lysed using a combination of three freeze/thaw cycles and osmotic shock.¹⁶⁰ Cell pellets were resuspended in lukewarm lysis buffer (30% sucrose, 1 mM EDTA, 30 mM Tris-HCl, pH 8.0) on an orbital shaker at room temperature for 20 minutes @ 200 rpm. Approximately 75 mL of lysis buffer per liter of culture were used for cell pellet resuspension. Dislodging of the cell pellet from container walls initially required rigorous shaking! Resuspended pellet was then centrifuged at 5500 rpm for 10 min. The sucrose buffer containing most of the protein was then poured off and treated with 1 mM PMSF protease inhibitor. The resulting pellet was then resuspended in equal volumes of **cold** 5 mM magnesium sulfate (MgSO₄) solution if the pellet still had color. If the pellet was beige to white in color then purification continued on the “sucrose cut” only. The 5 mM MgSO₄ resuspension was then centrifuged at 8500 rpm for 1 hour. Supernatant was then collected and treat with 1 mM PMSF.

The crude protein extract was diluted to 5 mM ionic strength with **cold** water before CM sepharose purification. The pH of the crude protein extract was adjusted to 5.0 with concentrated acetic acid and the suspension was centrifuged for 30 minutes at 10,000 rpm to remove any precipitates that may have developed after pH adjustment.

The cleared lysate was **immediately** top loaded onto a 100 mL CM Sepahrose column pre-equilibrated with CM Low buffer (5 mM NaOAc, pH 5.0). Protein was eluted with a 500 mL low to high salt gradient using CM High buffer (5 mM NaOAc, 500 mM NaCl, pH 5.0). Eluted protein was exchanged in low salt HPLC Buffer A (10 mM NaP_i, pH 6.0) *via* centrifuge ultrafiltration. Yields at this point were typically 7 mg/L of culture. Before experimentation, the protein was additionally HPLC purified, in its oxidized state, using a Waters AP-1 ProteinPak SP-8HR cation exchange column with high salt HPLC Buffer B (10 mM NaP_i, pH 6.0, 500 mM NaCl) in the following stepwise gradient: 0-10 min. 0% B, 10-11 min. 4.5% B, 11-21 min. linear gradient to 5.1% B, 21-31 min. hold at 5.1% B, 31-51 min. linear gradient to 12% B, 51-71 min. hold at 12% B, 71-72 min. 100% B, 72-87 min. hold at 100% B, 87-88 min. 0% B, 88-102 min. hold at 0% B. This gradient successfully removed an impurity with a mass of ~ 9 kDaltons commonly observed in the MALDI-TOF mass spectrum. The molecular ion observed for all variants was that expected for each variant within the accuracy of the Applied Biosystems Voyager – DE PRO Biospectrometry Workstation mass spectrometer.

3.2.5 Protein stability measurements

Protein stability measurements were done at 25 °C in the presence of 20 mM Tris, 40 mM NaCl, pH 7.0 for iso-1-cytochrome *c* variants using an Applied Photophysics Chirascan circular dichroism spectrometer coupled to a Hamilton MICROLAB 500 Titrator as previously described in Section 2.2.5. Stability measurements on cytochrome *c*' variants were done at 25 °C in the presence of 20 mM MES, 40 mM NaCl, pH 6.50 using the same apparatus as above.

Cytochrome *c'* protein concentrations were first determined using 100 mM NaP_i buffer at pH 7.05 because the spectrum is pH dependent. Scans were taken from 250 to 750 nm and absorbance at 398 nm was used with a molar extinction coefficient of 85000 M⁻¹ cm⁻¹ to calculate concentrations.¹⁶¹ Iso-1-cytochrome *c* concentrations were obtained as previously described.⁶⁹ The data for iso-1-cytochrome *c* were fitted to a linear free energy relationship, using Eq 3.2, as described previously,^{80; 117} to extract the free energy of unfolding in the absence of denaturant, ΔG_u^{o'}(H₂O), and the *m*-value. Two-state folding is assumed in these fits.

$$\theta = \frac{\theta_N^o + ((\theta_D^o + m_D[\text{gdnHCl}]) \exp\{m[\text{gdnHCl}] - \Delta G_u^o'(H_2O) / RT\})}{1 + \exp(m[\text{gdnHCl}] - \Delta G_u^o'(H_2O) / RT)} \quad (3.2)$$

where θ is the ellipticity of the sample, θ_N^o is the ellipticity of native protein at 0 M gdnHCl, θ_D^o is the ellipticity of denatured protein at 0 M gdnHCl, m_D is the denaturant dependence of the ellipticity of the denatured state, m is the gdnHCl concentration dependence of the free energy of unfolding, ΔG_u, and ΔG_u^{o'}(H₂O) is the free energy of unfolding extrapolated to 0 M gdnHCl. However, unlike iso-1-cytochrome *c* variants, cytochrome *c'* does not demonstrate slight increases in ellipticity at 222 nm at low gdnHCl concentrations, previously attributed to possible specific Cl⁻ binding¹⁶² as described in Section 2.2.5. Thus, no modification was made to the standard linear free energy fitting equation,¹⁶² which naturally uses a floating native baseline ($\theta_N = \theta_N^o + m_N[\text{gdnHCl}]$) to fit the data instead of a constant native state baseline. Reported parameters are the average and standard deviation of three independent trials.

3.2.6 Equilibrium loop formation in the denatured state

Equilibrium loop formation in the denatured state was monitored in 3 M gdnHCl at 22 ± 1 °C as previously described in Section 2.2.6 for all iso-1-cytochrome *c* variants. Data were fitted to a modified form of the Henderson-Hasselbalch equation (Eq 3.3) as previously described.

$$A_{398} = \frac{(A_{LS} + A_{HS} 10^{n(pK_a(obs) - pH)})}{(1 + 10^{n(pK_a(obs) - pH)})} \quad (3.3)$$

Additionally, equilibrium loop formation in a good solvent (6 M gdnHCl) was monitored at 22 ± 1 °C as previously described,⁸² for the NH(-2) variant along with the other NH5A-X variants. Data were fitted to Eq 3.3.

Cytochrome *c'* variants were also monitored under similar denatured state conditions (3 M gdnHCl, 5 mM Na₂HPO₄, 15 mM NaCl) at 22 ± 1 °C. Some cytochrome *c'* variants required fitting the data to a two-ligand binding model (Eq 3.4)⁸² for adequate fits. Reported parameters are the average and standard deviation of three independent trials.

$$A_{398} = \frac{A_{HS} + A_{LS} \left(\left(\frac{10^{-pK_{loop}(His)}}{1 + 10^{pK_a(HisH^+) - pH}} \right) + \left(\frac{10^{-pK_{loop}(Lys)}}{1 + 10^{pK_a(LysH^+) - pH}} \right) \right)}{1 + \left(\frac{10^{-pK_{loop}(His)}}{1 + 10^{pK_a(HisH^+) - pH}} \right) + \left(\frac{10^{-pK_{loop}(Lys)}}{1 + 10^{pK_a(LysH^+) - pH}} \right)} \quad (3.4)$$

where A_{LS} is the absorbance at 398 nm in the low spin form of the heme with a strong field ligand bound to the heme, A_{HS} is the absorbance at 398 nm in the high spin form of the heme with the ligand displaced by acid, $pK_{loop}(His)$ is the pK for loop formation with histidine as the ligand to the heme, $pK_{loop}(Lys)$ is the pK for loop formation with lysine as the ligand to the heme, $pK_a(HisH^+)$ is the pK_a for histidine ionization, and $pK_a(LysH^+)$ is the pK_a for lysine ionization, which is assumed to be 10.5 in fitting biphasic pH titration curves.

3.2.7 pH dependent stopped-flow kinetic measurements

pH dependent stopped-flow kinetic measurements were used to monitor His-heme loop breakage as previously described in Section 2.2.7 above in both 3 M and 6 M gdnHCl at 25 °C for the iso-1-cytochrome *c* variants only. Additionally, variants with higher $pK_a(obs)$, namely NH5A-3, NH5A-4 and NH5A-5, loop breakage was measured at pH 4.0. Therefore, additional 100 mM acetate buffers in both 3 M and 6 M gdnHCl were prepared for the above mentioned variants.

3.3 Results

3.3.1 Global Stability Measurements

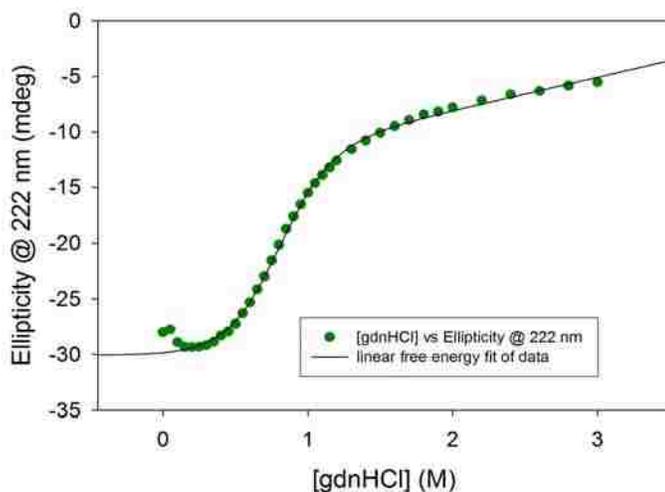


Figure 3.1: A typical gdNHCl protein denaturation titration monitored at 222 nm and 25 °C using an Applied Photophysics Chirascan circular dichroism spectrometer for the homopolymeric poly(Ala) variants (NH5A-5 variant Run #1). The solid line is a fit of the data using Eq 3.2.

Figure 3.1 shows a typical gdNHCl unfolding titration for the poly(Ala) homopolymeric protein variants. Global stability measurements on the initial polyalanine insert (NH5A) show that it is significantly lower in stability relative to the H(-2)I52 [written as NH(-2)] variant. This destabilization is due primarily to the lack of the global stabilizing mutation N52I,¹⁶³ in the NH5A variant. With continued AAAAAK insertions, there is a subtle but noticeable increase in the stability of the variants, which is confirmed in the highly reproducible C_m (midpoint) values. The C_m value indicates the [gdNHCl] at which half the protein population is unfolded. There is also a noticeable decrease in the m -values.

Table 3.1: Thermodynamic parameters for gdnHCl unfolding at 25 °C and pH 7 for N-terminal loops of iso-1-cytochrome *c* variants

Loop size	Variant	$\Delta G_u^{\circ}(\text{H}_2\text{O})^a$ (kcal/mol)	m -value ^a (kcal/(mol*M))	C_m Midpoint (M)
16	NH(-2) ^b	5.57 ± 0.14	4.80 ± 0.17	1.20 ± 0.07
22	NH5A	2.00 ± 0.08	4.52 ± 0.18	0.44 ± 0.03
28	NH5A-2	2.00 ± 0.41	4.32 ± 0.60	0.46 ± 0.04
34	NH5A-3	2.20 ± 0.19	4.43 ± 0.44	0.50 ± 0.01
40	NH5A-4	2.31 ± 0.03	3.38 ± 0.13	0.69 ± 0.02
46	NH5A-5	2.55 ± 0.03	3.37 ± 0.14	0.76 ± 0.02

^aValues obtained by fitting the data to a linear free energy relationship, as described previously,^{80,117}.

Values are the averages and standard deviations of three independent trials.

^bData value obtained from Ph. D. thesis of Eydjejo Kurchan.⁹⁰ This variant contains a stabilizing N52I mutation for improved yields.¹⁶³

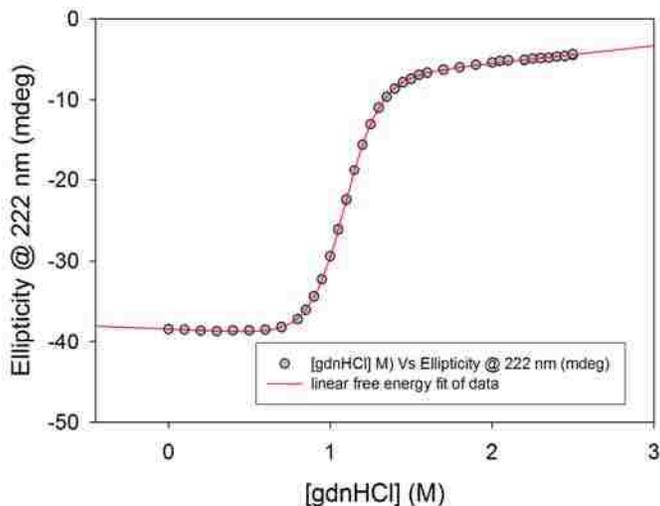


Figure 3.2: A typical gdnHCl protein denaturation titration monitored at 222 nm and 25 °C using an Applied Photophysics Chirascan circular dichroism spectrometer for a cytochrome *c*' variant (A104H variant Run #2).

Figure 3.2 represents a typical trace of gdnHCl denaturation for a cytochrome *c*' variant. Immediately obvious from Table 3.2, is the fact that cytochrome *c*' variants are much more stable than the homopolymeric iso-1-cytochrome *c* variants. Also obvious

from Figure 3.2 is the fact that the more stable cytochrome *c'* variants do not suffer from anion binding effects at low concentrations of gdnHCl as do the poly(Ala) variants of iso-1-cytochrome *c*. Thus, linear fits of the free energy relationship equation were done using a floating native baseline instead of a fixed native baseline (Eq 3.2) – as in the case of the latter protein. A summary of the data is given in Table 3.2 and Figure 3.3.

Table 3.2: Thermodynamic parameters for gdnHCl unfolding of cytochrome *c'* variants at pH 6.5 and 25 °C

Cytochrome <i>c'</i> variant	$\Delta G_u^{0'}(\text{H}_2\text{O})^a$ (kcal/mol)	<i>m</i> -value ^a (kcal/(mol*M))	(<i>C_m</i>) Midpoint ^a (M)
A104H (10)	6.08 ± 0.26	5.60 ± 0.12	1.09 ± 0.05
K97H (17) ^b	6.80 ± 0.09	5.28 ± 0.07	1.29 ± 0.03
K91H (23) ^b	7.07 ± 0.08	5.75 ± 0.16	1.23 ± 0.02
K84H (30) ^b	5.89 ± 0.16	5.32 ± 0.14	1.11 ± 0.06
E73H (41) ^b	5.76 ± 0.13	5.08 ± 0.14	1.13 ± 0.05
A66H (48) ^b	6.79 ± 0.09	4.88 ± 0.06	1.39 ± 0.04
D58H (56) ^b	6.72 ± 0.09	4.80 ± 0.11	1.40 ± 0.04
K49H (65) ^b	6.75 ± 0.10	4.66 ± 0.07	1.45 ± 0.03
K39H (75) ^b	6.91 ± 0.15	4.70 ± 0.10	1.47 ± 0.05
K31H (83)	7.41 ± 0.29	5.05 ± 0.20	1.47 ± 0.03
K20H (94) ^b	5.89 ± 0.13	4.17 ± 0.09	1.41 ± 0.06
K13H (101) ^b	8.17 ± 0.11	4.91 ± 0.07	1.66 ± 0.05
D3H (111)	7.41 ± 0.17	5.08 ± 0.15	1.46 ± 0.05
pWT	8.81 ± 0.22	5.14 ± 0.13	1.71 ± 0.04

Values in parenthesis represent the loop size formed *via* His-heme ligation.

^aValues obtained by fitting data to a linear free energy equation without fixing the native baseline. Values are the averages and standard deviations of three independent trials.

^bUnpublished data (Dr. Sudhindra Rao).

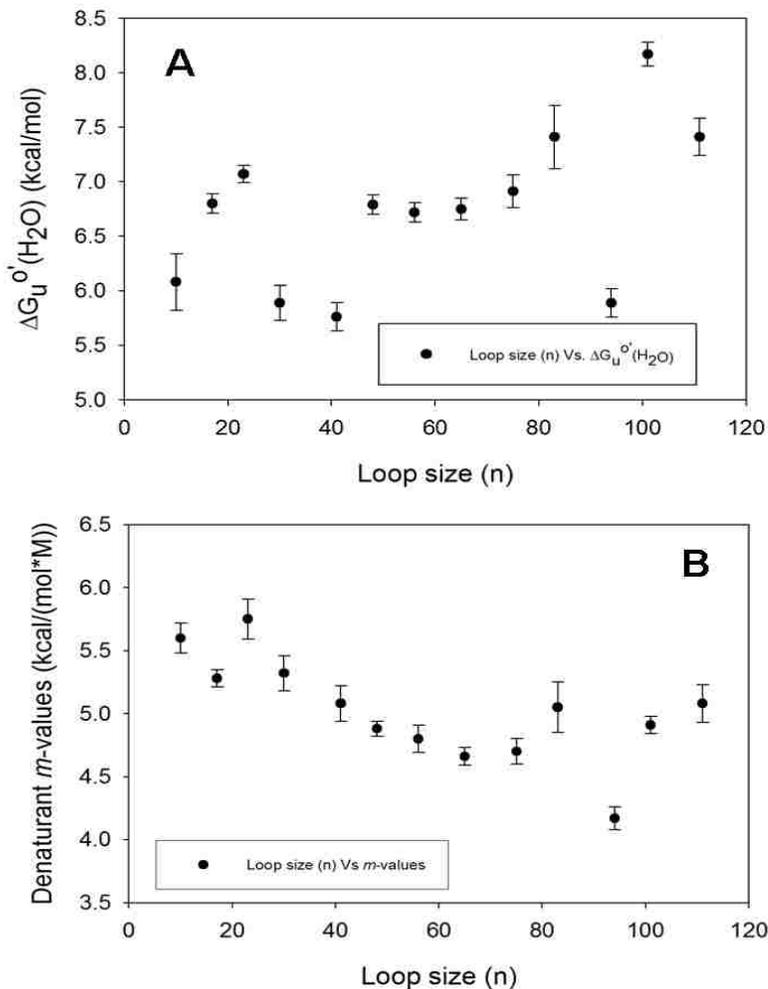


Figure 3.3: Plots of loop size (n) versus (A) free energy of unfolding and (B) denaturant m -values (indicating degree of hydrophobic exposure upon unfolding) for cytochrome c' variants.

Comparison of the thermodynamic unfolding data for both Poly(Ala) iso-1-cytochrome c and cytochrome c' variants, from Tables 3.1 and 3.2 and Figure 3.3, show a general parallel trend of increasing ΔG_u^o and C_m midpoints with increasing loop size. The increase is not uniform for cytochrome c' as it is for the poly(Ala) iso-1-cytochrome c variants. The former actually seem to have two distinct groups of variants following this increasing trend as seen from Figure 3.3A. This trend of increasing

stabilities is also accompanied by a decrease in denaturant m -values for both foldable and non-foldable sequences with increasing loop size. After a loop size of about 56, there is an increase in the m -values for the foldable sequences as seen in Figure 3.3B. This trend was previously seen for loops formed from the C-terminal side of iso-1-cytochrome c .¹²⁴

3.3.2 Equilibrium loop formation measurements for poly(Ala) sequences

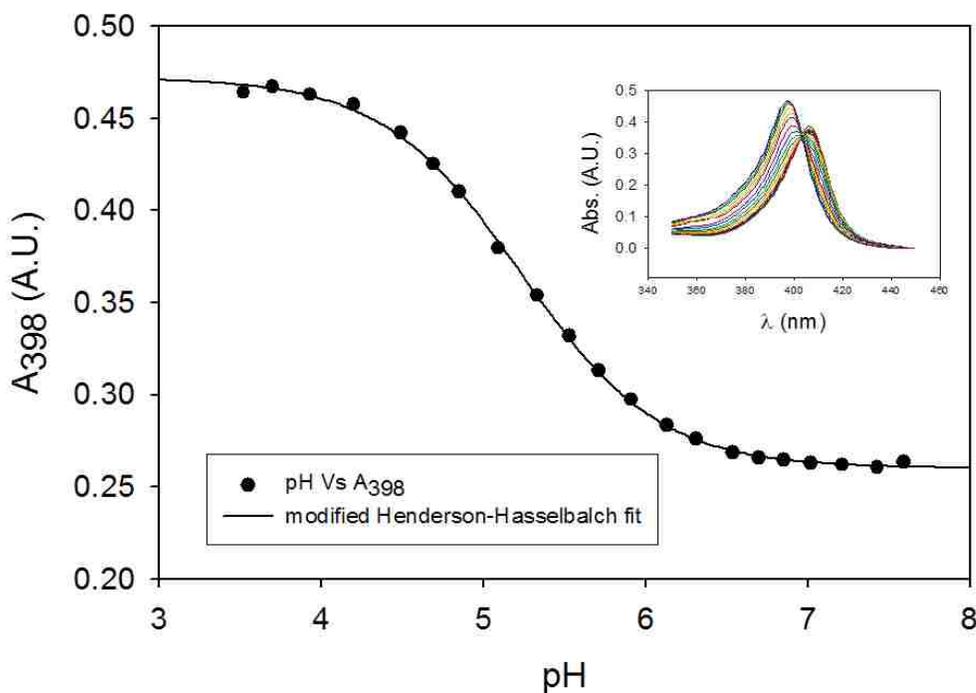


Figure 3.4: A typical equilibrium loop formation titration representative of the poly(Ala) variants in 3 M gdnHCl (NH5A-4 variant). Fit is based on a modified form of the Henderson-Hasselbalch equation. See Figure 2.1 and Eq 3.3. Inset is a plot of absorbance from 350 to 450 nm going from pH ~ 7 to ~ 3 in increments of ~ 0.2 pH units at 22 ± 1 °C.

Figure 3.4 is a typical equilibrium loop formation titration for the homopolymeric poly(Ala) variants in 3 M gdnHCl. Immediately obvious from the plots is the well-defined two-state nature of the transition, which is confirmed by the single

isosbestic point seen in the plot of absorbance *versus* wavelength as a function of pH in the inset of Figure 3.4. The n-value of ~ 1 is consistent with the one-ligand, one-proton process in Figure 2.1 for the His-heme ligation. Figure 3.5 is an additional plot of data for both 3 M and 6 M gdnHCl denaturant for the same NH5A-4 variant. Thermodynamic parameters for loop formation for all poly(Ala) variants of iso-1-cytochrome *c* are collected in Table 3.3.

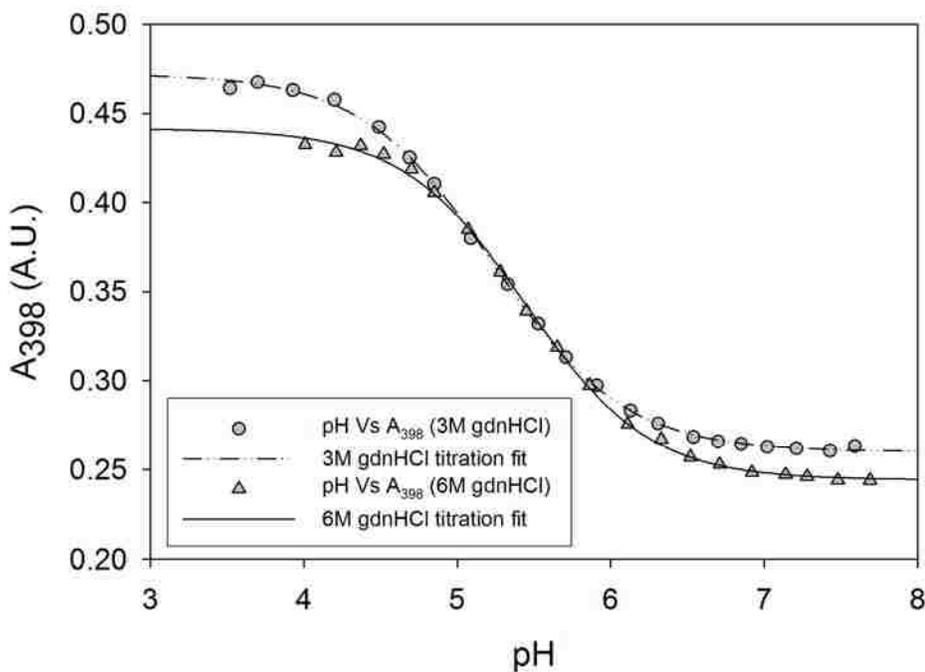


Figure 3.5: Typical equilibrium loop formation traces for poly(Ala) variants in both 3 M and 6 M gdnHCl solvent conditions (NH5A-4 variant) at 22 ± 1 °C.

Table 3.3: Thermodynamic parameters for equilibrium loop formation at 22 ± 1 °C in 3 M and 6 M guanidine hydrochloride for poly(Ala) iso-1-cytochrome *c* variants

Loop size	Variant	3 M gdnHCl Values		6 M gdnHCl Values	
		pK _a (obs)	n-value	pK _a (obs)	n-value
16	NH(-2)	4.41 ± 0.01	1.04 ± 0.04	4.65 ± 0.05	1.14 ± 0.04
22	NH5A	4.62 ± 0.04	1.03 ± 0.05	4.90 ± 0.02	1.15 ± 0.01
28	NH5A-2	4.97 ± 0.04	1.05 ± 0.13	5.13 ± 0.02	1.11 ± 0.01
34	NH5A-3	5.16 ± 0.07	1.04 ± 0.12	5.28 ± 0.02	1.06 ± 0.06
40	NH5A-4	5.27 ± 0.03	1.03 ± 0.05	5.45 ± 0.01	1.07 ± 0.01
46	NH5A-5	5.41 ± 0.03	1.11 ± 0.06	5.53 ± 0.02	1.08 ± 0.05

Data are based on three individual trials. Errors are standard deviations.

n-values of ~ 1 indicate release of one proton and are consistent with the His-heme ligation scheme in Figure 2.1.

Plots of the pK_a(obs) against log of loop size under “poor” and “good” solvent conditions shed light on the very subtle changes in the scaling factors (v_3). For a freely jointed random coil polymer chain configuration, v_3 is predicted to be ~ 1.5 without accounting for excluded volume.^{37; 87} Accounting for excluded volume – as is necessary in the the case of proteins – drives the value of v_3 to ~ 1.8 to 2.4 .^{37; 87} We can see from the slopes of the plots in Figure 3.6, that the homopolymeric poly(Ala) sequences satisfy the prediction of a random coil with excluded volume under poor solvent conditions (3 M gdnHCl). Upon further solvation in a good solvent (6 M gdnHCl), the scaling factor decreases significantly from 2.26 to 1.97, but still is in line with a homopolymeric chain exhibiting excluded volume interactions. The deviation from v_3 for a random coil with excluded volume in 3 M gdnHCl is primarily due to the NH(-2) and NH5A variants. Their loops are dominated by the natural protein sequence (16 of 22 and 16 of 28 residues, respectively). Removal of the NH(-2) and NH5A variants from consideration in the v_3 calculation results in a scaling factor of 2.00 ± 0.10 (dashed line in Figure 3.6) – exactly parallel to that in 6 M gdnHCl.

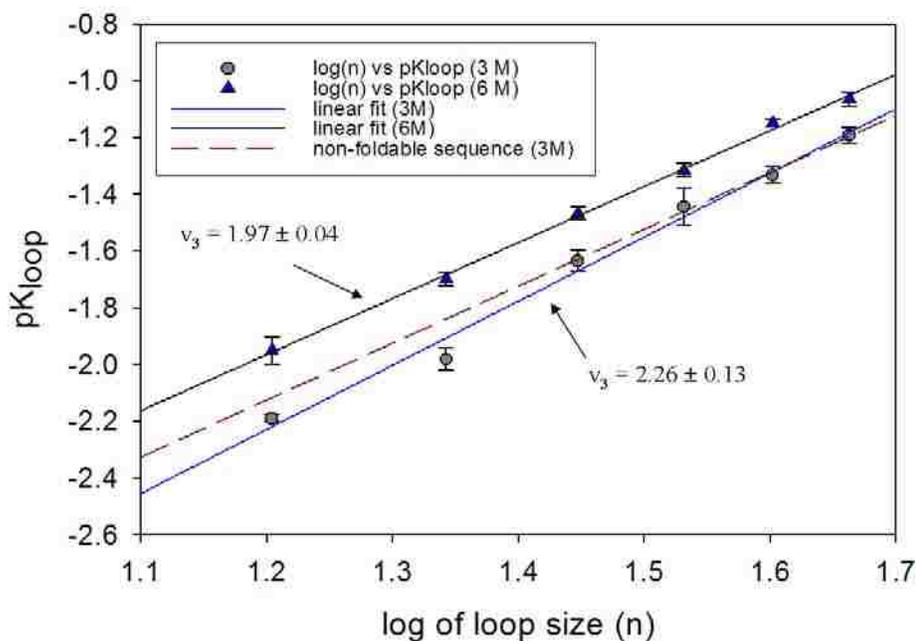


Figure 3.6: Linear plots of 3 M and 6 M equilibrium loop formation data for poly(Ala) variants with the resulting scaling factors under each solvent condition.

3.3.3 Equilibrium loop formation for heteropolymeric sequences of cytochrome *c*'

The variants of cytochrome *c*' behaved differently from those for iso-1-cytochrome *c* during pH titrations. For most variants, fits of the data to a one-ligand His-heme binding process produced $pK_a(\text{obs})$ that were lower than the expected 6.6 for free histidine. However, a handful of variants produced $pK_a(\text{obs})$ larger than 6.6, which indicated that another ligand is required to drive the titration to completion at higher pH's. More than likely, the other ligand is a lysine.⁸² Therefore, these unconventional variants were fitted to a two-ligand version of the Henderson-Hasselbalch equation (see Eq 3.4). The output parameters were the pK_{loop} of the His-heme loop, $pK_{\text{loop}}(\text{His})$, the

pK_a of the HisH^+ , $pK_a(\text{HisH}^+)$, and the pK_{loop} of the Lys-heme loop, $pK_{\text{loop}}(\text{Lys})$. The $pK_a(\text{LysH}^+)$ was assumed to be 10.5 in this two-ligand equation (see Eq 3.4).^{82; 164}

Figure 3.7 shows a typical trace for this unconventional two-ligand driven equilibrium loop formation. The inset in Figure 3.7 clearly indicates that there is not just a single isosbestic point present, which would indicate a two-state process. Conversely, Figure 3.8 typifies most of the equilibrium loop formation traces for the cytochrome c' variants. A single isosbestic point in the inset of Figure 3.8 indicates a simple two-state process. A summary of the equilibrium loop formation data in 3 M gdnHCl is presented in Table 3.4 below.

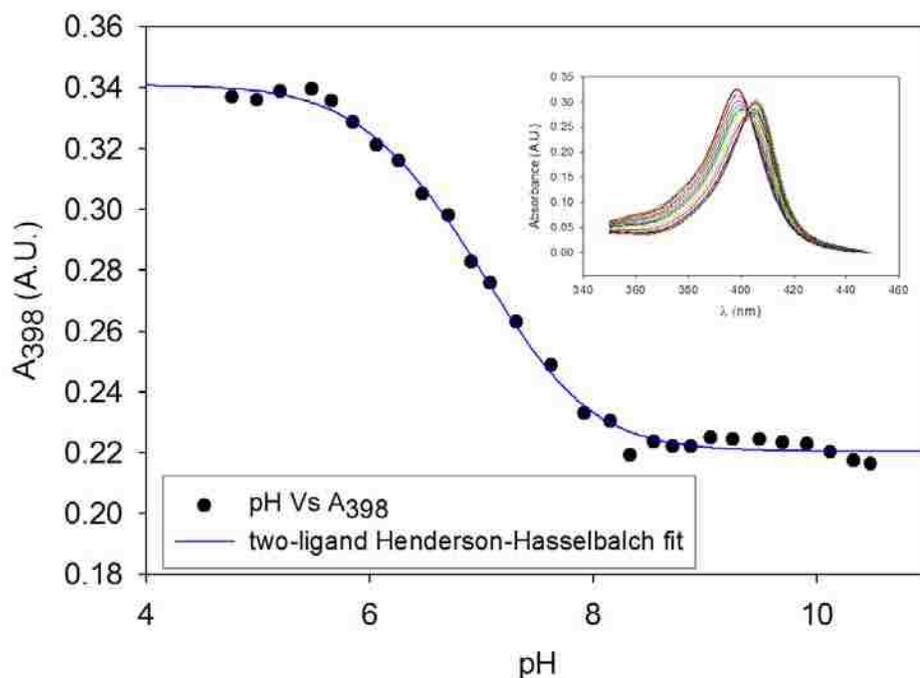


Figure 3.7: A representative trace for unconventional two-ligand driven equilibrium loop formation of some cytochrome c' variants (D3H variant). Inset is a plot of absorbance from 350 to 450 nm from pH ~ 10.5 to 5 in increments of ~ 0.2 pH units at 22 ± 1 °C.

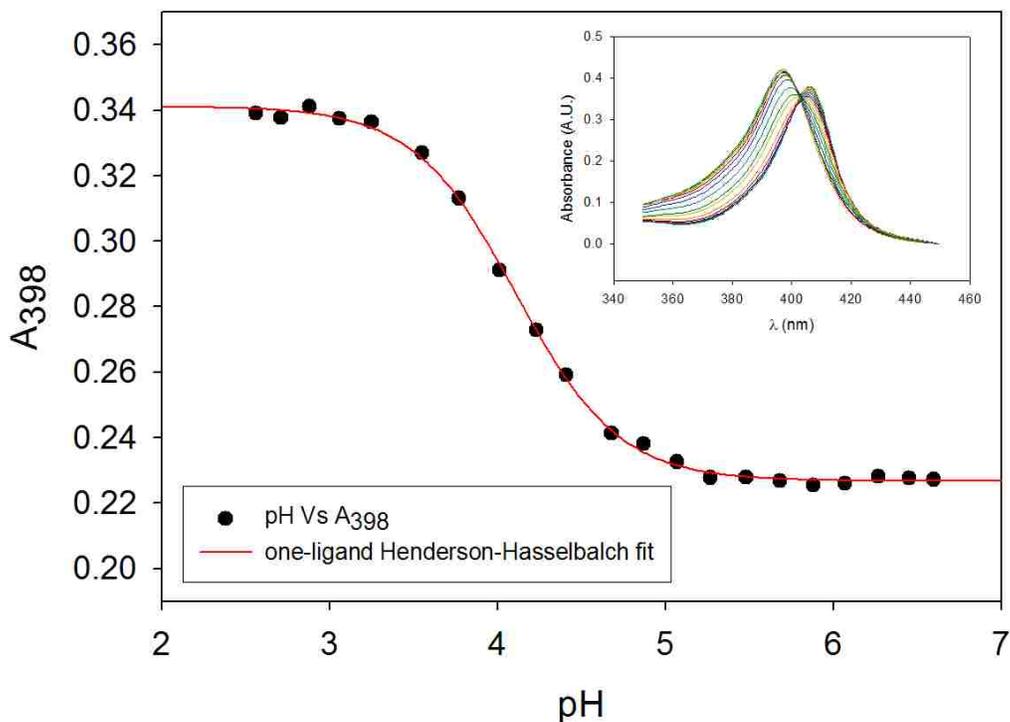


Figure 3.8: A representative trace for a normal one-ligand driven equilibrium loop formation observed for most cytochrome *c'* variants (A104H variant). Inset is a plot of absorbance from 350 to 450 nm from pH ~6.5 to ~2.5 in increments of ~0.2 pH units at 22 ± 1 °C.

$pK_{loop}(His)$ (pK for loop formation with a fully deprotonated His) values have been extracted either directly from the two-ligand fits for some variants or by calculation from the apparent $pK_a(obs)$ for other variants, using Eq 3.5 .

$$pK_a(obs) = pK_a(HisH^+) + pK_{loop}(His) \quad (3.5)$$

These values are plotted against log of loop size (n) to obtain scaling factors (v_3) which can be used for comparisons of our polymer chains to random coils (Figure 3.9). Immediately obvious from Figure 3.9 is that the scaling factors of both the cytochrome *c'*

and the poly(Ala) variants are not very different from each other. Nonetheless, the homopolymeric sequences do produce a scaling factor of 2.26, which is well within the predicted range for a random coil with excluded volume. The heteropolymeric sequences produce a scaling factor of 2.5 ± 0.3 , which higher but also within error of the predicted range for a random coil with excluded volume. Also evident from the plot is that the data for the homopolymeric variants fit exactly on the line for a random coil. While the heteropolymeric variants have considerable scatter about their linear fit. This is indicative of the presence of residual structure.

Table 3.4: Thermodynamic parameters for loop formation in 3 M gdnHCl for cytochrome *c'* variants at 22 ± 1 °C

Cytochrome <i>c'</i> variant	pK _a (obs) ^a	pK _a (HisH ⁺) fitted ^b	pK _{loop} (Lys) ^b	pK _{loop} (His) or calculated ^c
A104H (10)	4.12 ± 0.07			-2.57
K97H ^d (17)	4.42 ± 0.07			-2.27
K91H ^d (23)	5.08 ± 0.05			-1.61
K84H ^d (30)	4.93 ± 0.04			-1.76
E73H ^d (41)	5.07 ± 0.04			-1.62
A66H ^d (48)	5.92 ± 0.09			-0.77
D58H ^d (56)	5.68 ± 0.04			-1.01
K49H ^d (65)	6.19 ± 0.02			-0.50
K39H ^d (75)		6.83 ± 0.09	-2.57 ± 0.14	-0.81
K31H (83)	6.19 ± 0.05			-0.50
K20H ^d (94)		6.81 ± 0.05	-3.25 ± 0.17	-0.77
K13H ^d (101)		6.72 ± 0.09	-3.12 ± 0.07	-0.22
D3H (111)		6.41 ± 0.33	-3.47 ± 0.06	0.53
WT	7.30 ± 0.05			---

Values and errors are the averages and standard deviations of three independent trials.

^aValues are from a one-ligand fit of data using a modified Henderson-Hasselbalch equation (see Eq 3.3).

^bValues are from a two-ligand fit of data to Equation 3.5. The pK_a(LysH⁺) was assumed to be 10.5 in the equation.

^cCalculated values were obtained by averaging the pK_a(HisH⁺) obtained [6.7 ± 0.2] and subtracting it from pK_a(obs).

^dUnpublished Data (Dr. Sudhindra Rao)

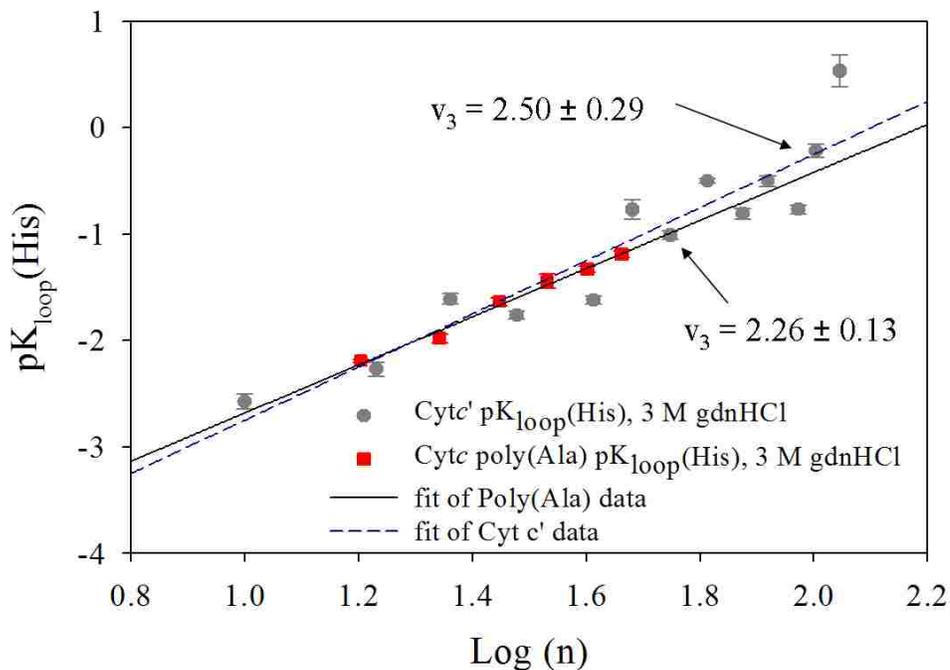


Figure 3.9: Plots of log of loop size (n) versus $pK_{loop}(His)$ for both homopolymeric poly(Ala) sequences (red squares; solid line – linear fit) and heteropolymeric foldable cytochrome c' sequences (gray circles; dashed line – linear fit). The linear fits to the 3 M gdnHCl data produce scaling factors (v_3) from the slopes.

3.3.4 Poly(Ala) stopped-flow loop breakage in poor and good solvents

Loop formation and breakage kinetics are consistent with a model involving a rapid protonation equilibrium (of histidine) followed by His-heme loop formation.⁸⁹ This model predicts that k_{obs} has the pH dependence given by Eq 3.6,

$$k_{obs} = k_b + k_f \left(\frac{K_a(HisH^+)}{[H^+] + K_a(HisH^+)} \right) \quad (3.6)$$

where $K_a(\text{HisH}^+)$ is the ionization constant of the histidine involved in loop formation and k_f and k_b are the rate constants for loop formation and loop breakage, respectively. Thus, if $\text{pH} \ll \text{p}K_a(\text{HisH}^+)$, k_b can be obtained. We have used this model in obtaining k_b rates from our data fits. Representative data, fitted to a single exponential rise to maximum equation, are shown in Figure 3.10.

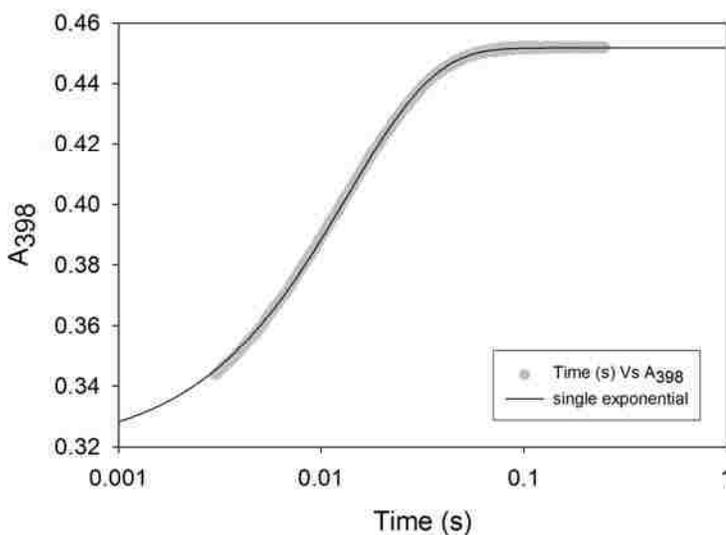


Figure 3.10: Representative data for loop breakage in 3 M gdnHCl. Data are fitted to a single exponential rise to maximum equation (solid line).

Table 3.5 summarizes the kinetic parameters from loop breakage in both 3 M and 6 M gdnHCl denaturant concentrations for all poly(Ala) variants. We measured k_{obs} under 3 M gdnHCl conditions for the variants [NH(-2), NH5A, NH5A-2] at $\text{pH} 3.10 \pm 0.02$ and for variants [NH5A-3, NH5A-4, NH5A-5] at $\text{pH} 3.75 \pm 0.12$. These pH's were also verified by immediately collecting the effluent of the breakage experiments and measuring the pH. Errors in pH are the standard deviations among different variants in

the grouping. We subsequently measured k_{obs} under 6 M gdnHCl conditions for the variants [NH(-2), NH5A, NH5A-2] at $\text{pH } 3.16 \pm 0.03$ and for variants [NH5A-3, NH5A-4, NH5A-5] at $\text{pH } 4.01 \pm 0.02$. These pH's were also verified by immediately collecting the effluent of the breakage experiments and checking them. Errors are the standard deviations among different variants in the grouping. We note that with $\text{pK}_a(\text{His}) = 6.6 \pm 0.1$, Eq 2.1 indicates that the contribution of k_f to k_{obs} at pH [3.10, 3.75] for variants [NH(-2), NH5A, NH5A-2, NH5A-3, NH5A-4, NH5A-5] is about 1 to 5 s^{-1} in 3 M gdnHCl. While the contribution of k_f to k_{obs} at pH [3.75, 4.01] for variants [NH(-2), NH5A, NH5A-2, NH5A-3, NH5A-4, NH5A-5] is about 0.7 to 3 s^{-1} in 6 M gdnHCl. Thus, equating k_{obs} to k_b at these respective pH's is a reasonable assumption consistent with the small contribution expected from k_f at the pH used to monitor loop breakage (Eq 3.6). Loop breakage was also done at a higher pH on the above groups of variants (see Table A1 in Appendix A). The difference in the k_{obs} rate constants from higher to lower pH is small, which confirms that it is appropriate to equate k_{obs} to k_b at low pH.

Loop formation rates are much faster than the deadtime of stopped-flow instrumentation.⁸⁹ Thus, these rate constants were calculated by extracting the pK_{loop} (pK for loop formation with a fully deprotonated His) from the apparent $\text{pK}_a(\text{obs})$, using Eq 3.5. This equation is a reasonable approximation if $\text{pK}_a(\text{obs})$ is at least one unit less than $\text{pK}_a(\text{HisH}^+)$. Since $\text{pK}_a(\text{HisH}^+)$ equals 6.6 ± 0.1 in 3 M gdnHCl,⁸² the approximation is reasonable for the data presented here (Table 3.3). pK_{loop} was then used in conjunction with the loop breakage rate constants, k_b , to extract loop formation rate constants (Table 3.6). The k_b and k_f values presented in Tables 3.5 and 3.6 do not account for viscosity

effects at 6 M gdnHCl.^{42; 149; 165} Corrected values obtained using Eq 3.7¹⁴⁹ are presented in Table 3.7.

$$k'_c = k_c \left(\frac{\eta}{\eta_o} \right)^{-\beta} \quad (3.7)$$

where k'_c is the viscosity-corrected rate constant for loop formation at a given viscosity (η), k_c is the rate constant at viscosity (η_o), and β is either -1 for k_c being inversely proportional to viscosity or 0 for viscosity independence.¹⁴⁹ Viscosity values at 3 and 6 M gdnHCl concentration were obtained from a fourth-order polynomial fit to the data of Kawahara and Tandford.¹⁶⁶ The corrected k_b rate constants are only about 6% higher in 6 M gdnHCl than in 3 M gdnHCl. Conversely, the viscosity corrected k_f rate constants showed a 19% decrease in 6 M gdnHCl *versus* 3 M gdnHCl.

Table 3.5: Kinetic parameters for loop breakage at 25 °C in 3 M and 6 M guanidine hydrochloride of poly(Ala) iso-1-cytochrome *c* variants

Loop size	Variant	3 M gdnHCl		6 M gdnHCl	
		k_b (pH 3.10 ± 0.02) (s ⁻¹)	k_b (pH 3.75 ± 0.12) (s ⁻¹)	k_b (pH 3.16 ± 0.03) (s ⁻¹)	k_b (pH 4.01 ± 0.02) (s ⁻¹)
16	NH(-2)	102.3 ± 2.8		78.4 ± 3.3	
22	NH5A	93.0 ± 0.2		68.9 ± 2.4	
28	NH5A-2	79.9 ± 0.7		65.7 ± 2.0	
34	NH5A-3		74.3 ± 1.0		58.6 ± 0.2
40	NH5A-4		74.3 ± 1.2		57.8 ± 0.2
46	NH5A-5		73.0 ± 1.2		57.4 ± 0.2

Data are based on three individual trials and errors are from their standard deviations.

$k_{(obs)}$ from loop breakage experiments are presented as k_b since the calculated contribution from k_f at these low pH's is 1 – 5 s⁻¹ and 0.7 – 3 s⁻¹ under 3 M and 6 M gdnHCl, respectively.

Table 3.6: Calculated kinetic parameters for loop formation at 25 °C in 3 M and 6 M guanidine hydrochloride N-terminal loops of iso-1-cytochrome *c* variants

Loop size	Variant	3 M gdnHCl		6 M gdnHCl	
		k_f (pH 3.10 ± 0.02) (s ⁻¹)	k_f (pH 3.75 ± 0.12) (s ⁻¹)	k_f (pH 3.16 ± 0.03) (s ⁻¹)	k_f (pH 4.01 ± 0.02) (s ⁻¹)
16	NH(-2)	15840 ± 640		6990 ± 830	
22	NH5A	8910 ± 830		3450 ± 230	
28	NH5A-2	3440 ± 300		1920 ± 120	
34	NH5A-3		2060 ± 320		1210 ± 70
40	NH5A-4		1590 ± 110		810 ± 30
46	NH5A-5		1140 ± 80		670 ± 30

Data are based on three individual trials and errors are from their standard deviations.

k_f values presented here are not yet corrected for viscosity at the higher 6 M gdnHCl concentrations.

Table 3.7: Viscosity corrected kinetic parameters for loop breakage and loop formation at 25 °C in 6 M guanidine hydrochloride for iso-1-cytochrome *c* poly(Ala) variants

Loop size	Variant	Viscosity Corrected k_b	Viscosity Corrected k_f
		k_b [pH 3.16 ± 0.03; pH 4.01 ± 0.02] (s ⁻¹)	k_f [pH 3.16 ± 0.03; pH 4.01 ± 0.02] (s ⁻¹)
16	NH(-2)	108.5 ± 4.5	9680 ± 1150
22	NH5A	95.4 ± 3.4	4780 ± 310
28	NH5A-2	91.0 ± 2.8	2660 ± 170
34	NH5A-3	81.1 ± 0.4	1680 ± 100
40	NH5A-4	80.0 ± 0.4	1130 ± 40
46	NH5A-5	79.5 ± 0.3	920 ± 50

Data are based on propagation of error in uncorrected k_b and k_f .

k_f values presented here are corrected for viscosity at the higher 6 M gdnHCl concentration.

3.4 Discussion

3.4.1 Global protein stability of homopolymeric poly(Ala) iso-1-cytochrome *c* variants and heteropolymeric cytochrome *c*' variants

For both the iso-1-cytochrome *c* and cytochrome *c*' variants (Tables 3.1 and 3.2), there is a general decreasing trend in *m*-values (indicates a lesser degree of solvent exposed surface area upon unfolding) and an increasing trend in $\Delta G_u^{0'}(\text{H}_2\text{O})$ as loop size increases. For the heteropolymeric cytochrome *c*' variants the decreasing trends are not immediately obvious as they are for the iso-1-cytochrome *c* variants. However, a simple plot of loop size *versus* $\Delta G_u^{0'}(\text{H}_2\text{O})$ and *m*-values (see Figure 3.3), shows two groups of variants following these general trends. The decrease in the *m*-values suggests that the larger loops may be more compact due to formation of residual structure as more of the protein is constrained into a loop in the denatured state. This trend was previously seen as well for loops formed from the C-terminal side of iso-1-cytochrome *c*.¹²⁴ If this is the case, then the denatured state would be increasingly stabilized relative to the native state and we would expect to see a decrease in the $\Delta G_u^{0'}(\text{H}_2\text{O})$ as loop size increases.

However, we see an increase in $\Delta G_u^{0'}(\text{H}_2\text{O})$ with increasing loop size. This is confirmed in the highly reproducible C_m values (indicates the [gdnHCl] at which half the protein population is unfolded). Since equilibrium loop formation pK_{loop} values (see Tables 3.3 and 3.4 and Figures 3.6 and 3.9) become less negative with increasing loop size, this indicates that the loops are becoming less stable as loop size increases. Destabilization of the larger loops would translate to a less stabilized denatured state and we would expect to see an increase in the $\Delta G_u^{0'}(\text{H}_2\text{O})$ as loop size increases. Thus, our equilibrium loop formation data indicates that weaker loops, which may be due to the

increased loop entropy of these larger loops, is the dominant influence on the observed increase in $\Delta G_u^{0'}(\text{H}_2\text{O})$.

For cytochrome *c'* many of the histidine mutations are in α -helices since this protein is a four-helix bundle. Histidine has a low helix propensity *versus* lysine or alanine;¹⁴² therefore, native state effects cannot be ruled out in the $\Delta G_u^{0'}(\text{H}_2\text{O})$ trends seen for cytochrome *c'* variants.

3.4.2 Equilibrium loop formation in the denatured state for homopolymeric poly(Ala) iso-1-cytochrome *c* and heteropolymeric cytochrome *c'* variants

One of the advantages of using His–heme ligation as a probe is that a direct correlation can be made between the apparent pK_a , $\text{pK}_a(\text{obs})$, of titrating the histidine off the heme and the strength of that ligand interaction. Typically, the apparent pK_a is lower than the pK_a of free histidine because deprotonation of histidine is coupled to favorable formation of a bond between Fe^{3+} in the heme and the histidine imidazole side-chain. The stability of the His–heme bond is dependent on three main factors as previously described in the *Introduction*. First, chain stiffness can decrease the stability if the polypeptide chain forming the loop is too short or too sterically restrictive, thereby hindering effective His–heme molecular orbital overlap.^{80; 81} These hindrances can be relieved to increase the bond strength by either increasing loop size or decreasing residue bulk. Second, loop entropy decreases the His–heme bond strength as the loop becomes larger. Too long a loop increases the conformational space for the histidine to search, thereby decreasing the probability of making an actual His–heme bond. Third, residual structure formation increases the His–heme bond strength due to the presence of

additional stabilizing contributions in the loop, if the structure is induced by loop formation. The summed stability of these contributors is conferred onto the His–heme bond. Lower pK_a values for the His–heme ligation are a direct indication that the interaction is stronger and thus requires more acid to break the bond by protonating the imidazole ring of the histidine.

The above knowledge of loop behavior enables a straightforward analysis of our data on equilibrium loop formation. The steady increase in the $pK_a(\text{obs})$ with increasing loop size, as seen in Figure 3.6, indicates that for N-terminally formed loops of iso-1-cytochrome *c*, loop entropy is the dominating factor that affects loop stability. When compared to previous data⁸² from our lab on loops formed from the C-terminal end of iso-1-cytochrome *c* (see Figure 3.11), it is clear that loops formed from the N-terminal side of this protein do not suffer from loop strain. There is no comparable drop in $pK_a(\text{obs})$ to indicate relief of loop strain at an optimal loop size as is the case for C-terminal loops (see Figure 3.11).

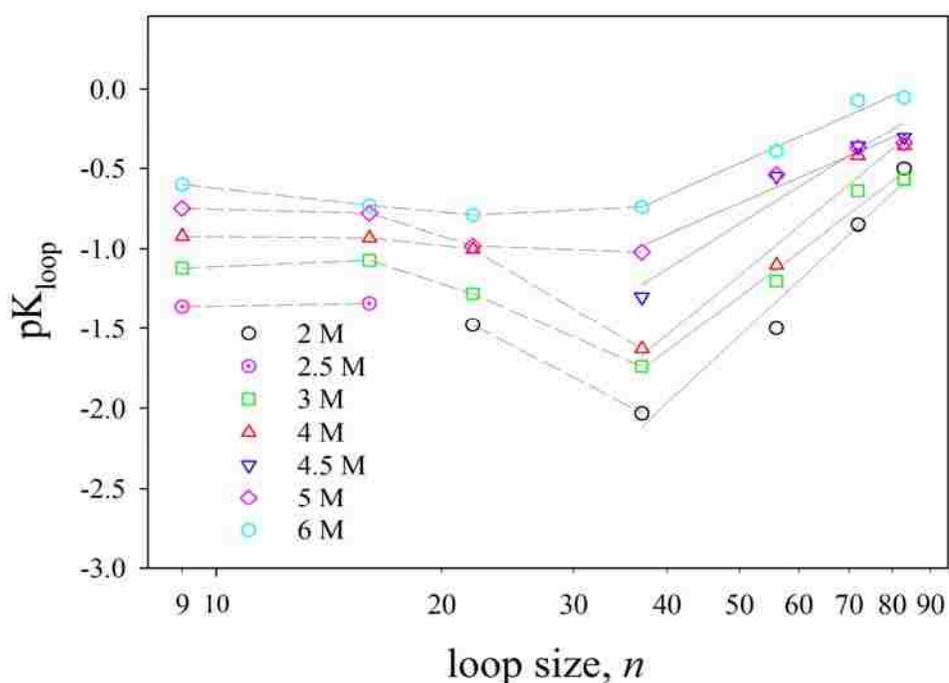


Figure 3.11: pK_{loop} versus log of loop size (n) for iso-1-cytochrome c loops formed from the C-terminal end under poor and good solvent conditions. The solid lines represent linear fits of the data and therefore their slopes are scaling factors for comparison to a random coil. Figure taken from Wandschneider and Bowler.⁸²

The resulting scaling factors from Figure 3.6 indicate that homopolymeric non-foldable protein sequences follow the theories of polymer science. The scaling factor of 2.26 for these types of protein sequences in a poor solvent (3 M gdnHCl) indicates that the local interactions of the individual monomer units are still not strong enough to dominate the local solvent-monomer interactions as well as the nonlocal excluded volume interactions of the polymer chain. Hence, we see a denatured state ensemble behavior akin to random coil with excluded volume. In 6 M gdnHCl solvent conditions the scaling factor drops to 1.97 which indicates very minimal, if any, structural interactions occurring in the denatured state of homopolymeric sequences. Curiously, if

the 3 M gdnHCl data is plotted without the NH(-2) and NH5A variants, the slope of the fit (see dashed line in Figure 3.6) is exactly parallel to that of the variants in 6 M gdnHCl. This shows that deviation from random coil behavior is due solely to the foldable natural sequence. The non-foldable poly(Ala) sequences behave like a random coil regardless. Conversely, for heteropolymeric sequences, as seen from previous data⁸² and from our work on cytochrome *c*' (see Figure 3.9), there can be considerable residual structure. It is a bit surprising that the natural heteropolymeric sequence of the iso-1-cytochrome *c* protein produces such high scaling factors ($v_3 \sim 4$ in 3 M gdnHCl) compared to the natural heteropolymeric sequence of cytochrome *c*' ($v_3 = 2.5$). We note however, that the data around the linear fit of the cytochrome *c*' data shows considerable scatter. This may indicate the presence of periodic clusters of residual structure along the polypeptide chain. Further experimentation under good solvent conditions should provide further insight into the behavior of denatured cytochrome *c*'.

This result mostly confirms the conclusion of previous data,⁸² as depicted by Figure 3.11, that the high scaling factor in poor solvent conditions (2-4 M gdnHCl) for a natural protein polypeptide is definitely due to the presence of residual structure. The reason for other naturally occurring polymeric chains, such as DNA, having random coil like scaling factors^{156; 157} may be attributable to the lower complexity conferred to a polymer having an assortment of only four different nucleotide bases as compared to twenty different amino acids for proteins.

3.4.3 Kinetic analyses of homopolymeric protein sequences in poor and good solvent

Previous work done by Kiefhaber and co-workers using triplet-triplet energy transfer on host-guest peptides indicated that contact formation over short distances ($N < 5$) does not depend on the loop size.¹¹¹ Rather, the limiting factor in loop contact (formation) for such short loops depended on the nature of the amino acid residue in the chain. Less bulky glycines produced more flexible chains which had faster contact formation times compared to chains with more constrained residues such as prolines.¹¹¹ For such short loops this effect is primarily due to chain stiffness. This effect has also been seen in equilibrium data from our lab (see Figure 3.11) presumably due to the loop formation from the chain having to “wrap around” the heme to form a His-heme bond.^{81;}⁸² Furthermore, Kiefhaber demonstrated that the rates of contact formation for longer loops ($N > 30$) were found to decrease with increasing loop size. He also found that this decrease in contact formation rates is attributable to the increased end-to-end distances of a polymer chain in a good solvent.¹¹¹ Thus, this decrease in contact (loop) formation is primarily due to diffusion across a larger distance which can be demonstrated directly from the effect that the solvent viscosity has on the system.

We investigated these findings with our own system. For the most part, our findings are in agreement with Kiefhaber’s. Our loop breakage rate constants were much slower in 6 M gdnHCl *versus* 3 M gdnHCl. Since Kiefhaber’s work suggested that this decrease is viscosity related, we did a viscosity correction^{42; 149; 165} on our loop breakage data. The results are summarized in Table 3.7 and are almost accounted for by viscosity when compared to the 3 M data from Table 3.5. However, there is a 6 – 9% higher loop

breakage rate in 6 M gdnHCl *versus* 3 M gdnHCl – even with the correction. This may be within the error of the calculation but it may also be the result of disruption or weakening of minimal amounts of residual structure in our homopolymeric sequences being in a “goodd” solvent. A plot of the 3 M gdnHCl k_b data against the 6 M gdnHCl viscosity corrected data, seen in Figure 3.12, argues for the latter case.

More interestingly, Figure 3.12 shows that the k_b rate constants for variants NH5A-3 to NH5A-5 (loop sizes 34-46), level off in both 3 M and 6 M gdnHCl. To explain this we must remember that loops do not behave as a random coil until they reach a certain length several times the persistence length of the chain. At short lengths there may be too much chain stiffness to enable random coil behavior in a good solvent.⁹² Thus, our data indicate that the loop is strained for shorter loops, < 34 residues. Chain lengths of at least 30 residues were found to be necessary to display random coil behavior with an excluded volume component.^{92; 111} Therefore, our data are exactly in accordance with previous results.^{92; 111}

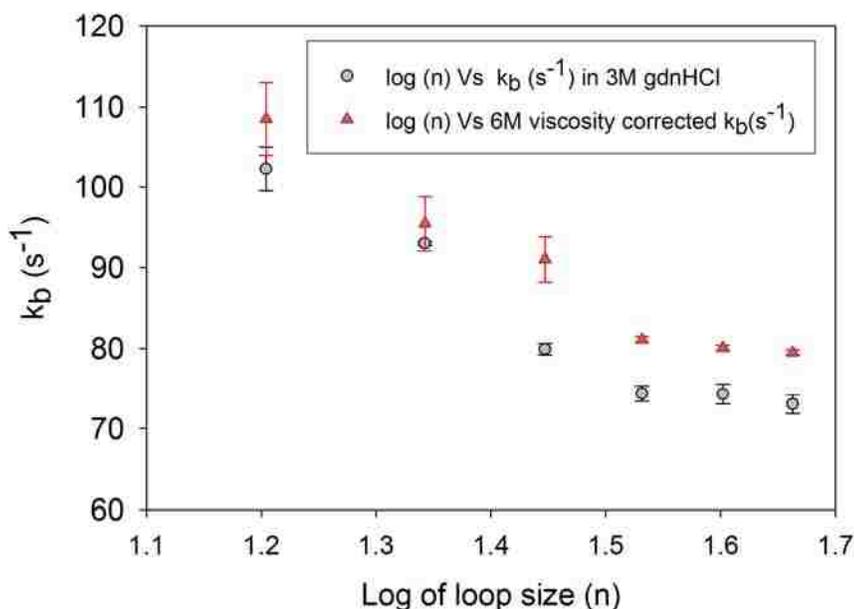


Figure 3.12: Loop breakage rate constants as a function of log of loop size (n) for poly(Ala) loops in 3 M and 6 M viscosity corrected gdnHCl solvent conditions.

An increase to 6 M gdnHCl from 3 M gdnHCl conditions should produce a decrease in the contact formation rates, accounted for entirely by viscosity and the expanded end-to-end contact distances associated with good solvents, according to Keifhaber.¹¹¹ Therefore we also did a viscosity correction^{42; 149; 165} on our k_f rate constants. The results are summarized in Table 3.7. When compared to their 3 M gdnHCl loop formation rates (see Table 3.6), the 6 M gdnHCl viscosity corrected values are 38 – 19% lower than expected. The difference is depicted in Figure 3.13. Thus, Keifhaber was correct about the decrease in contact formation rates in good solvents being due to increased end-to-end distance, since a significant fraction of the decreased rate of contact formation is not accounted for by viscosity. Thus, our loop formation findings are in accord with Keifhaber's.

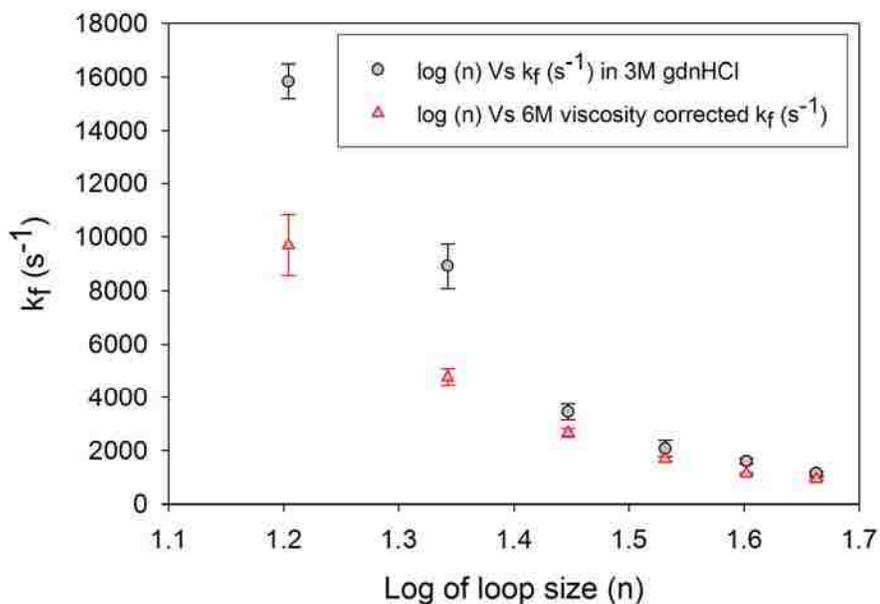


Figure 3.13: Calculated loop formation rate constants as a function of log of loop size (n) for poly(Ala) loops in 3 M and 6 M viscosity corrected gdnHCl solvent conditions.

A plot of log of loop size (n) *versus* log k_f should produce a linear relationship from which scaling factors, v_3 , can be verified kinetically as well. Figure 3.14 shows that these poly(Ala) sequences in 3 M gdnHCl, are just outside the value for a random coil with excluded volume while in a “good” solvent (6 M gdnHCl) that value drops to 2.28 – well within the range for a random coil with excluded volume. The two shortest loops (16 and 22), which have the most natural foldable sequence, seem to influence the higher v_3 value in 3 M gdnHCl. This is expected in “poor” solvents if small amounts of residual structure are present.

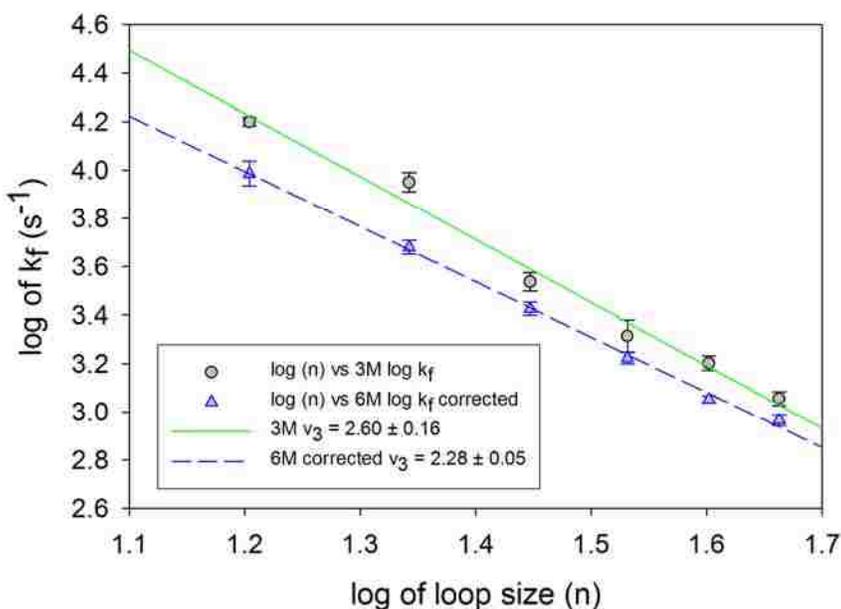


Figure 3.14: Calculated log of loop formation rate constants as a function of log of loop size (n) for poly(Ala) loops in 3 M and 6 M viscosity corrected gdnHCl solvent conditions. The slope provides scaling factors for comparison to a random coil.

3.5 Conclusion

The effect of sequence composition and, in some cases, sequence order, can have a significant impact on the thermodynamic and kinetic behavior of a natural polymeric sequence. Over a decade of research on protein denatured states has established that such naturally occurring polymers do not behave the same as random polymers. The nature of this disparity is believed to be due to naturally occurring hydrophobic clusters present in a protein's natural sequence which consequently is suspected to guide the initial stages of the protein folding process leading down the folding funnel. The resilient nature of these hydrophobic clusters is such because of the need to prevent breakage of the crucial seeding structure (see Chp 2). Hence, these residual structures are not easily solvated,

even under denaturing conditions. Numerous studies have shown residual structure persists even under 8 M urea denaturing conditions.⁹⁶

The data for cytochrome *c'* presented here is unlike any of our previously studied heteropolymeric protein sequences.⁸² Smooth fits to a linear equation of $\text{pK}_{\text{loop}}(\text{His})$ to log of loop size are not seen as compared to previously studied loop variants in iso-1-cytochrome *c*.⁸² This scatter about linearity is more than likely due to the presence of irregular clusters of residual structure in cytochrome *c'*. For iso-1-cytochrome *c* hydrophobic collapse was more dominant in shorter compared to longer denatured state loops (see Figure 3.11). These new data are exciting and await future kinetic studies which may establish the role of residual structure in speeding up rates of loop formation or slowing loop breakage for cytochrome *c'* variants.

Comparison of thermodynamic data on scaling factors of heteropolymeric sequences from both iso-1-cytochrome *c*⁸² and cytochrome *c'* with current homopolymeric engineered sequences has provided stronger evidence supporting the presence of robust residual structures in a natural protein sequence. We have also convincingly demonstrated through thermodynamic and kinetic studies that homopolymeric protein sequences exhibit the behavior of a random coil with excluded volume for loop lengths 30 residues and greater. This finding is in accord with previously published data.¹¹¹

CHAPTER 4

COMPETITION BETWEEN REVERSIBLE AGGREGATION AND LOOP FORMATION IN DENATURED ISO-1-CYTOCHROME C

4.1 Introduction

Slow phases in protein folding are usually associated with proline isomerization.¹⁶⁷ However, slow folding phases can also be caused by aggregation.¹⁶⁸⁻¹⁷⁰ Interest in protein aggregation and its causes is high due to its role in a number of human diseases characterized by irreversible aggregation.^{26; 171} Several recent studies demonstrate reversible aggregation during folding.^{168; 169; 172-175} The state that mediates aggregation during folding varies. In some cases, the denatured state is implicated,^{168; 172; 173} and in other cases, partially folded intermediates are implicated.¹⁷⁵ The reversibility of aggregation during folding can depend on protein concentration.¹⁷⁴ Studies on the dynamics of denatured proteins have shown that proline residues can dramatically affect the conformational properties of the denatured state.^{110; 121} In the case of iso-2-cytochrome *c*, proline isomerization has been shown to modulate non-native His-heme ligation.¹²¹ In this study, we investigate the effects of proline on aggregation in the denatured state of iso-1-cytochrome *c* (we note that iso-1 and iso-2 cytochromes *c* are naturally occurring variants of cytochrome *c* in *Saccharomyces cerevisiae* that are ~80% identical in sequence).¹²¹

To probe the conformational properties of the denatured state, we have developed methods for measuring His-heme loop formation using variants of iso-1-cytochrome *c* containing a single histidine capable of loop formation.^{37; 38} In general, we find that loop

formation in the denatured state is independent of protein concentration.^{81; 82; 176} Similarly, the kinetics of His-heme loop formation and breakage are consistent with two-state intramolecular binding of histidine to heme.⁸⁹ However, when the histidine is <10 amino acids from the point of attachment of the heme to the polypeptide chain, we observe significant concentration dependence for equilibrium loop formation.^{81; 82; 176} The AcH26I52 variant (has the H33N and H39Q mutations, an acetylated N-terminus and the N52I mutation to stabilize the protein) of iso-1-cytochrome *c* has a histidine at position 26, which is nine residues from the nearest point of attachment to the heme (His 18). There is also a proline at position 25. Equilibrium loop formation is strongly dependent on protein concentration for this variant,⁸² consistent with intermolecular His-heme binding. Interestingly, when the histidine is at position 27 (AcH27 variant), such that it is not immediately adjacent to Pro25, equilibrium loop formation is not significantly concentration dependent.¹⁷⁶ This result suggests that the position of proline relative to histidine in short loops can have a dramatic effect on intermolecular aggregation.

Studies with a variant of iso-2-cytochrome *c* having a single histidine at position 26 (N26H, H33N, H39K) show that denatured state loop formation is slowed by a factor of 140 relative to the wild-type protein.¹²¹ The lifetime of 800 ms for His26-heme loop formation suggested that loop formation was gated by *trans* to *cis* isomerization of the Gly24-Pro25 peptide bond in this iso-2-cytochrome *c* variant. Thus, it appears that both intermolecular aggregation and proline isomerization could influence the properties of the denatured state of cytochrome *c* when only a small number of residues separate the histidine from the heme.

To probe the effects of Pro25 on the denatured state of iso-1-cytochrome *c*, we use two variants, AcH26I52¹²⁴ and AcA25H26I52 (adds the mutation P25A to the AcH26I52 variant). Both have a histidine at position 26 capable of forming a nine-residue His-heme loop in the denatured state. In the latter variant, Pro25 has been mutated to Ala. We characterize denatured state loop formation using both equilibrium and kinetic methods. The equilibrium data demonstrate that replacement of Pro25 weakens intermolecular His-heme binding. Kinetic data show that the rate of intramolecular loop formation and breakage is not strongly influenced by this mutation. However, in the absence of Pro25, the bimolecular rate constant for intermolecular His26-heme binding is decreased and the unimolecular rate for breakage of the intermolecular His26-heme bond is increased. Thus, Pro25 acts primarily to promote intermolecular interactions in the denatured state of iso-1-cytochrome *c*.

We also develop an equilibrium model to analyze amplitude data from our kinetics experiments and contrast it to a previous kinetic model¹⁷⁷ for predicting yields of intramolecular (folding or loop formation) *versus* intermolecular (aggregation) products. The comparison suggests that equilibrium control of folding *versus* aggregation may be more adaptive than kinetic control *in vivo*.

4.2 Materials and Methods

4.2.1 Preparation of the AcH26I52 and AcA25H26I52 variants

The AcH26I52 variant of iso-1-cytochrome *c*, which contains the T(-5)S and K(-2)L mutations (horse cytochrome *c* numbering is used; thus, the first five amino acids of

iso-1-cytochrome *c* are numbered -5 to -1) to give an acetylated N-terminus (eliminates competition between the N-terminal amino group and histidine for binding to the heme under denaturing conditions),¹¹⁹ the H33N and H39Q mutations so that His26 is the only histidine capable of denatured state loop formation,^{80; 117} the N52I mutation for stability¹⁶³ and the C102S mutation to prevent disulfide dimerization during physical studies, was prepared and purified from *Saccharomyces cerevisiae* as previously described.¹²⁴ The AcA25H26I52 variant adds the P25A mutation to the AcH26I52 variant. It was prepared using single-stranded pRS/C7.8 vector DNA containing the AcTM variant [T(-5)S, K(-2)L, H26N, H33N, H39Q and C102S relative to wild type] of iso-1-cytochrome *c*⁸¹ as template for the unique restriction site elimination method.¹²³ The SacI+II- and SacI-II+ selection oligonucleotides⁸¹ were used, as appropriate. The SacI+II- oligonucleotide eliminates a unique *Sac*II site and creates a *Sac*I restriction enzyme site upstream from the iso-1-cytochrome *c* gene from yeast (*CYCI*).¹⁷⁸ The SacI-II+ oligonucleotide does the opposite. The Asn52→Ile mutation (N52I) was introduced using the N52I oligonucleotide, 5'-d(TTTCTTGATGATGCCATCTGTGT)-3' (site of mutation is underlined), and the SacI+II- selection oligonucleotide. The Pro25→Ala mutation was then added using the P25A oligonucleotide, 5'-d(CAACCTTGTTGGCGCCACCCTTT)-3' and the SacI-II+ selection oligonucleotide. Finally, the Asn26→His mutation was added using the A25H26 oligonucleotide 5'-d(AACCTTGTGGGCGCCACCCTT)-3' and the SacI+II- selection oligonucleotide. All mutations were confirmed by dideoxy sequencing. The pRS/C7.8 vector containing the AcA25H26I52 variant was transformed into the GM-3C-2 *S. cerevisiae* cell line

(cytochrome *c* deficient)¹⁷⁹ and the transformants characterized, as described previously.⁸⁰ Expression and purification were carried out as for the AcH26I52 variant.

4.2.2 Protein stability measurements

The stability of the AcA25H26I52 variant was monitored as a function of gdnHCl concentration using an Applied Photophysics Chirascan circular dichroism spectrometer coupled to a Hamilton MICROLAB 500 Titrator using methods described previously.¹³¹ Data were acquired at 25 °C and pH 7.0 in the presence of 20 mM Tris, 40 mM NaCl as buffer. The data were fit to a linear free energy relationship ($\Delta G_u^{\circ}(\text{H}_2\text{O}) - m[\text{gdnHCl}]$), as described previously,⁸⁰ to extract the free energy of unfolding in the absence of denaturant, $\Delta G_u^{\circ}(\text{H}_2\text{O})$, and the *m*-value (rate of change of free energy of unfolding as a function of denaturant concentration). We use a constant native state baseline in fitting our data, as discussed previously.⁸⁰ Reported parameters are the average and standard deviation of three independent trials.

4.2.3 Equilibrium loop formation in the denatured state

pH titrations for monitoring His-heme binding in the denatured state (3 M gdnHCl, 5 mM Na₂HPO₄, 15 mM NaCl) of the AcA25H26I52 variant at 1, 3, 7.5 and 15 μM protein concentration were carried out with a Beckman DU-800 UV-Vis spectrometer. Titrations were carried out at room temperature (22 ± 1 °C). Spectra from 350 to 450 nm were acquired at each pH. Titration procedures have been described previously.⁸² Data at 398 nm *versus* pH were fit to a modified form of the Henderson-Hasselbalch equation, allowing extraction of the apparent pK_a for loop formation,

$pK_a(\text{obs})$, and the number of protons, n , involved in the process. Reported parameters are the average and standard deviation of three independent trials.

The concentration dependence of $pK_a(\text{obs})$ was fit to a model that assumes a competition between intramolecular His-heme loop formation, K_C , and His-heme dimerization, K_A (Eq 4.1, see Figure 4.2 in Results).

$$pK_a(\text{obs}) = -\log\left(\frac{K_C + \sqrt{K_C^2 + 4K_A[\text{Cyt}c]_t}}{2}\right) \quad (4.1)$$

A derivation of this equation is provided in Appendix B.

4.2.4 pH dependent stopped-flow kinetic measurements

To monitor the breakage and formation of the His-heme bond in the denatured state of the Ach26I52 variant, stopped flow mixing methods were used, and reactions were monitored by absorbance spectroscopy (Applied Photophysics PiStar 180 spectrometer) at 398 nm and 25 °C to observe the Soret band shift resulting from His-heme bond formation or breakage.⁸² For pH dependent His-heme bond formation reactions, a 2 mm pathlength and 2 μL flow cell were used, and the final reaction mixture was obtained from 1:1 mixing of 30 μM Ach26I52, 3 M gdnHCl, and 5 mM acetate (pH 3.10) with 3 M gdnHCl and 100 mM buffer to achieve the desired ending pH (MES, pH 5.5-7.0; MOPS, pH 6.5-8.0). The loop breakage reaction, using a pathlength of 10 mm and a 20 μL flow cell, was initiated by 1:1 mixing of 6 μM Ach26I52, 3 M gdnHCl, and 5 mM MOPS (pH 7.10) with 3 M gdnHCl and 100 mM buffer to achieve the desired

ending pH (acetate, pH 3.7-5.5). Final reaction pH was determined by collecting the product of the mixing reaction and immediately measuring pH. Using the method of reduction of 2,6-dichlorophenolindophenol,¹³⁵ dead times of 0.7 ms and 1.2 ms were determined for the 2 and 20 μ L flow cells, respectively, under our mixing conditions.

4.2.5 Continuous flow measurements as a function of pH

The rates of denatured state His-heme bond formation for the Ach26I52 variant were on the edge of the range accessible to stopped-flow mixing methods at higher pH. Therefore, continuous flow mixing methods were used to confirm stopped-flow mixing results. Continuous flow mixing methods with an efficient capillary mixer have been described in detail previously.¹⁸⁰⁻¹⁸³ Briefly, stock solutions of the Ach26I52 variant at 50 μ M were prepared in 3 M gdnHCl and 5 mM acetate buffer (pH 4.0). The loop formation reaction was initiated by a 5-fold dilution of the protein stock solution by 200 mM buffer and 3 M gdnHCl, producing a final reaction mixture of 3 M gdnHCl and 10 μ M protein at the desired ending pH (MES pH 6.0; MOPS pH 7.75). Final reaction pH was determined by collecting the product of the mixing reaction and immediately measuring the pH. Reactions were monitored using continuous flow ultrafast mixing absorbance spectroscopy^{180; 182} with a 0.9 mL/s flow rate at 395 nm and 25 °C to observe the Soret band shift which indicates His-heme bond formation. A dead time of 60 ± 10 μ s was measured using ascorbate reduction of 2,6-dichlorophenolindophenol as a test reaction.^{135; 183} Data reduction and analysis were carried out as described previously.^{89; 90}

4.2.6 Sequential stopped flow mixing methods

Due to the presence of a double exponential with 1:1 stopped flow mixing, sequential mixing (Applied Photophysics PiStar 180 spectrometer) was used to separate the slow and fast phases in both the formation and breakage reactions. The initial “pre-mix” reaction mixture was a 1:1 mix of 60 μM protein, 3 M gdnHCl, 1 mM MOPS (pH 7.0, formation), or 1 mM acetate (pH 3.75, breakage) with 10 mM acetate (pH 4.1, formation) or 10 mM MOPS (pH 7.1, breakage) and 3 M gdnHCl. This “pre-mix” was held for a range of aging times (5 – 5000 ms). After the appropriate aging time, another 1:1 mixing reaction was carried out with the “pre-mix” [30 μM protein and 3 M gdnHCl (pH 4.1 or 7.1)] with 3 M gdnHCl and 100 mM buffer at the desired ending pH (acetate, pH 3.75, for breakage; MOPS, pH 7.1, for formation). The final product includes 15 μM protein and 3 M gdnHCl at the desired ending pH. The reaction pH is measured by immediately taking the final product and measuring the pH. The pathlength was 2 mm and the volume of the flow cell was 2 μL . The dead time for the reaction was measured by the instrument during the experiment and was approximately 1 ms. Data were fit to double exponential decay or rise to maximum equations, as appropriate.

4.2.7 Stopped-flow mixing as a function of protein concentration

To monitor the concentration dependence of breakage and formation of the histidine-heme bond in the denatured state for the AcH26I52 and AcA25H26I52 variants, stopped-flow mixing methods were used, and reactions were monitored by absorbance spectroscopy (Applied Photophysics SX20 stopped-flow spectrometer) at 398 nm and 25 $^{\circ}\text{C}$ to observe the Soret band shift.⁸² A 5 μL flow cell was used for both His-heme bond

formation and breakage reactions. The 5 mm pathlength of the flow cell was used for low protein concentrations (2, 3.75 and 7.5 μM), while the 1 mm pathlength was used for higher protein concentrations (15, 30, 45 and 60 μM). The starting buffer for His-heme bond formation reactions was 10 mM acetate (pH 4.10) in 3 M gdnHCl. The ending buffer was 100 mM MOPS (pH 7.10) in 3 M gdnHCl. Final reaction mixtures for His-heme bond formation reactions were obtained by a 1:1 mixing of starting buffer (with a 2X protein concentration) and ending buffer. The 3 M gdnHCl concentration was verified by refractive index measurements.¹⁸⁴ The starting buffer for His-heme bond breakage reactions was 10 mM MOPS (pH 7.1) in 3 M gdnHCl. The ending buffer was 100 mM acetate (pH 3.75) in 3 M gdnHCl. For both upward and downward pH-jump experiments, the final reaction pH was determined by collecting the product of the mixing reaction and immediately measuring the pH. The starting and ending pH values for these experiments were selected on the basis of the pH-dependent loop formation kinetics such that loop formation or breakage would go to completion. Reduction of 2,6-dichlorophenolindophenol by ascorbate¹³⁵ gave a 0.7 ms dead time for the 5 μL flow cell, under our mixing conditions. The time axis of the data was adjusted for the dead time prior to fitting to triple- or quadruple-exponential functions. The appropriateness of the fitting function was judged by comparing residuals (see Figures 4.8 to 4.11 in Results Section).

4.3 Results

4.3.1 Global stability of the AcA25H26I52 variant

The global stability of the AcA25H26I52 variant was measured by gdnHCl denaturation, yielding $\Delta G_u^{\circ}(\text{H}_2\text{O}) = 7.8 \pm 0.5$ kcal/mol (Figure 4.1).

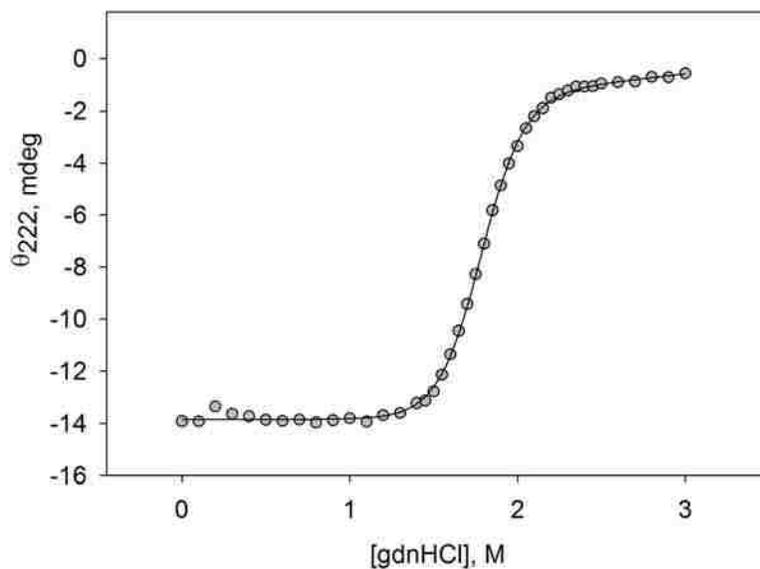


Figure 4.1: Plot of ellipticity *versus* gdnHCl concentration for the AcA25H26I52 variant at 25 °C. The solid line is a fit of the data to a two state model assuming a linear dependence of ΔG_u on gdnHCl concentration. The fit yields $m = 4.4 \pm 0.3$ kcal mol⁻¹ M⁻¹ and a titration midpoint, C_m , of 1.79 ± 0.02 M. $\Delta G_u^{\circ}(\text{H}_2\text{O})$ is reported in the text. The errors are the standard deviation of parameters from three experiments.

The P25A mutation destabilizes the AcA25H26I52 variant by ~ 1.5 kcal/mol relative to the AcH26I52 variant ($\Delta G_u^{\circ}(\text{H}_2\text{O}) = 9.37 \pm 0.24$).¹²⁴ It is clear from Figure 4.1 that the AcA25H26I52 is fully unfolded in 3 M gdnHCl, the conditions used for denatured state His-heme bond formation and breakage experiments.

4.3.2 Denatured state His-heme bond formation for the AcA25H26I52 variant

Intramolecular His-heme bond formation in the denatured state (3 M gdnHCl) leads to a nine-residue loop when a histidine is at position 26 of iso-1-cytochrome *c* (Figure 4.2).

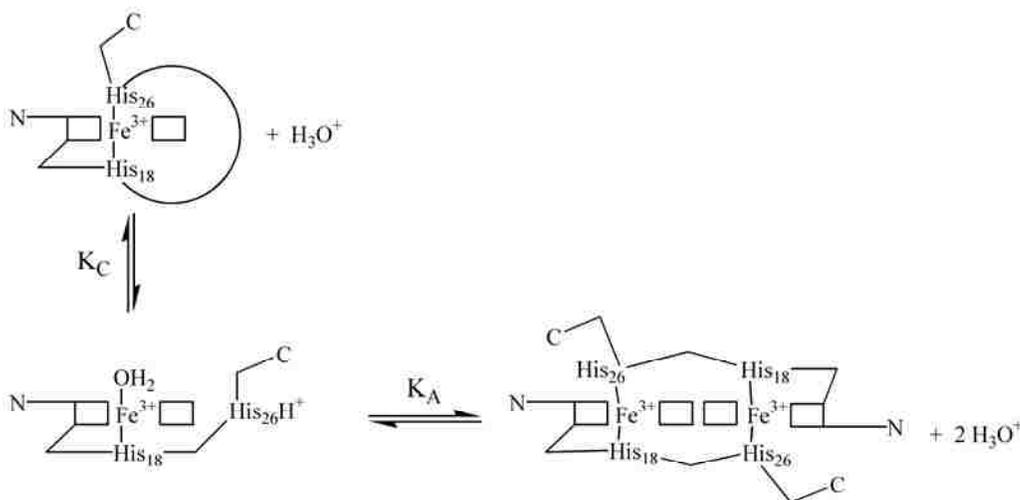


Figure 4.2: Reaction scheme for intramolecular loop formation in competition with intermolecular dimerization.

Spectroscopically monitored pH titrations yield an apparent pK_a, pK_a(obs), and the number of protons, *n*, involved in the process (inset of Figure 4.3 and Table 4.1). For intramolecular loop formation, the pK_a(obs) should be independent of protein concentration. However, if there is an intermolecular component (Figure 4.2), His-heme bond formation will be favored by mass action as cytochrome *c* concentration increases. Thus, intermolecular His-heme bond formation will cause the pK_a(obs) to decrease as cytochrome *c* concentration increases.

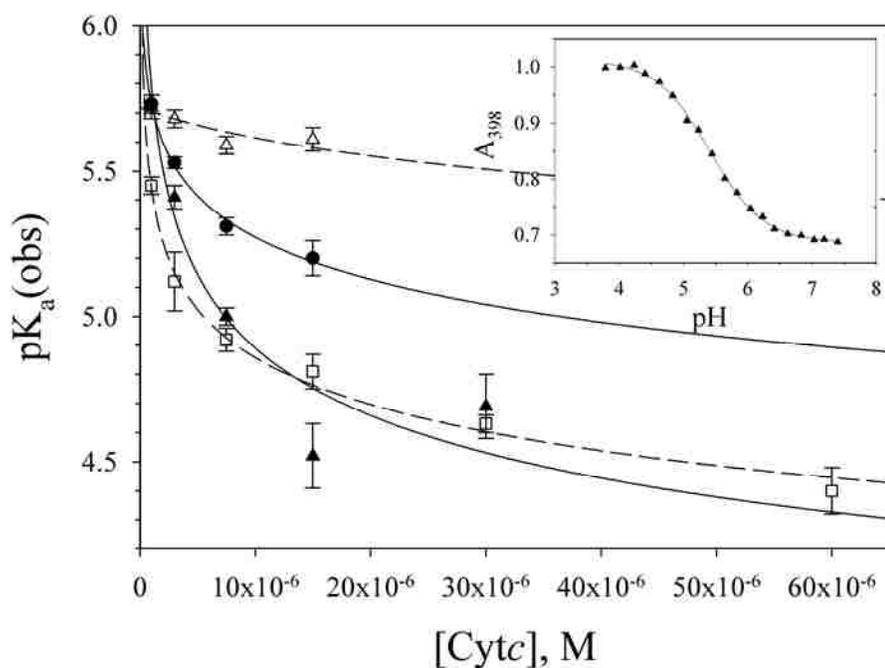


Figure 4.3: Plot of $pK_a(\text{obs})$ versus concentration for the AcA25H26I52 (●) and Ach26I52 (▲) variants and two other previously reported variants, Ach22 (□) and Ach27 (Δ).^{82; 176} Error bars are the standard deviation of $pK_a(\text{obs})$ from three independent experiments. All data were acquired at 22 ± 1 °C in 3 M gdnHCl containing 5 mM Na_2HPO_4 and 15 mM NaCl. Solid (Ach26I52 and AcA25H26I52) and dashed lines (Ach22 and Ach27) are shown for fits to the equilibrium model in Figure 4.2 (Eq 4.1, Materials and Methods; Appendix B). Inset: pH titration of AcA25H26I52 at 7.5 μM protein concentration in 3 M gdnHCl. The solid curve is a fit of the data to the Henderson-Hasselbalch equation with the number of protons released allowed to vary. The values of $pK_a(\text{obs})$ and the number of protons, n , for the AcA25H26I52 variant at concentrations ranging from 1 to 15 μM are collected in Table 4.1.

Table 4.1: Thermodynamic parameters for equilibrium His-heme bond formation in 3 M gdnHCl at 22 ± 1 °C for the AcA25H26I52 variant

Protein Concentration μM	$pK_a(\text{obs})$	n
1	5.73 ± 0.03	1.20 ± 0.13
3	5.53 ± 0.02	1.16 ± 0.12
7.5	5.41 ± 0.03	1.04 ± 0.01
15	5.20 ± 0.06	0.98 ± 0.06

Figure 4.3 shows the concentration dependence of the denatured state $pK_a(\text{obs})$ (3 M gdnHCl) for the AcA25H26I52 variant and the AcH26I52 variant. Data for two other variants,¹⁷⁶ AcH22 (His22, five-residue intramolecular loop) and AcH27 (His27, 10-residue intramolecular loop) are provided for comparison. For the AcH27 variant, the concentration dependence of $pK_a(\text{obs})$ is slight. The $pK_a(\text{obs})$ is strongly concentration dependent for the AcH26I52 and AcH22 variants and moderately so for the AcA25H26I52 variant. The data fit well to the equilibrium model in Figure 4.2 (Eq 4.1, Materials and Methods; Appendix B), yielding apparent binding constants (Table 4.2) for intramolecular His-heme loop formation (K_C) and intermolecular His-heme binding (K_A).

Table 4.2: Apparent intramolecular and intermolecular His-heme binding constants in the denatured state for iso-1-cytochrome *c* variants^a

Variant	Intramolecular loop size	K_C (M)	K_A (M)	$pK_a(\text{obs})^{0b}$
AcH22	5	– ^c	$2.3 \pm 0.3 \times 10^{-5}$	– ^c
AcH26I52	9	– ^c	$7 \pm 13 \times 10^{-5}$	– ^c
AcA25H26I52	9	$4.4 \pm 2.4 \times 10^{-7}$	$2.6 \pm 0.3 \times 10^{-6}$	6.35 ± 0.24
AcH27	10	$1.9 \pm 0.2 \times 10^{-6}$	$1.3 \pm 0.7 \times 10^{-7}$	5.72 ± 0.04

^aData were acquired in 3 M gdnHCl at 22 ± 1 °C.

^bApparent pK_a extrapolated to 0 M cytochrome *c* concentration.

^cMeaningful values could not be obtained for these variants from the fit to Eq 4.1.

For the AcH27 variant the apparent pK_a extrapolated to 0 M protein concentration, $pK_a(\text{obs})^0$ (Table 4.2) is identical to the $pK_a(\text{obs})$ at 1 μM protein concentration,¹⁷⁶ consistent with the minor intermolecular component observed for this variant. K_A varies

over two orders of magnitude (Table 4.2). Surprisingly, the single-site mutation P25A in the AcA25H26I52 variant *versus* the AcH26I52 variant decreases K_A for the His26 as the heme ligand, by ~ 10 -fold, indicating that aggregation in the denatured state can be strongly dependent on local sequence.

In the sections that follow, we use kinetic methods to probe the factors that cause this difference in denatured state aggregation between the AcH26I52 and AcA25H26I52 variants due to the mutation of Pro25 \rightarrow Ala. We first probe the pH dependence of His 26-heme bond formation and breakage as an initial step in characterizing intramolecular *versus* intermolecular kinetic phases using the AcH26I52 variant. We then confirm these assignments for the AcH26I52 variant with double-jump stopped-flow experiments. Finally, we characterize the concentration dependence of the kinetics for both variants over a 30-fold concentration range.

4.3.3 His26-heme bond formation and breakage kinetics as a function of pH for the AcH26I52 variant

Rates of His26-heme bond formation and breakage in the denatured state (3 M gdnHCl) as a function of pH were measured by both stopped-flow and continuous-flow mixing methods. Fast and slow phases were observed for both upward and downward pH-jump experiments (Table 4.3). The rate constant for the fast phase increases with increasing pH (Figure 4.4) as observed for intramolecular loop formation with other single-histidine variants of iso-1-cytochrome *c*.⁸⁹

Table 4.3: Rate constants, k_{obs} , and amplitudes for pH jump experiments as a function of pH at 25 °C and 3 M gdnHCl for the AcH26152 variant

pH	Fast phase		Slow phase		F:S Amp ^d
	k_{obs} (s ⁻¹)	Amplitude	k_{obs} (s ⁻¹)	Amplitude	
3.70 ^a	291 ± 13	8.0 ± 0.4 x10 ⁻²	3.84 ± 0.02	9.73 ± 0.04 x10 ⁻²	0.83
3.92 ^a	282 ± 10	8.3 ± 0.6 x10 ⁻²	5.57 ± 0.06	7.6 ± 0.1 x10 ⁻²	1.13
4.00 ^a	269 ± 18	7 ± 1 x10 ⁻²	4.8 ± 0.5	10.37 ± 0.02 x10 ⁻²	0.66
4.20 ^a	293 ± 28	7.2 ± 0.8 x10 ⁻²	3.16 ± 0.03	9.44 ± 0.05 x10 ⁻²	0.77
4.84 ^a	457 ± 47	5 ± 1 x10 ⁻²	1.6 ± 0.1	3.31 ± 0.04 x10 ⁻²	1.59
5.40 ^a	483 ± 50	2.9 ± 0.9 x10 ⁻²	1.35 ± 0.05	2.84 ± 0.02 x10 ⁻²	1.05
5.83 ^b	585 ± 70	6 ± 1 x10 ⁻²	1.1 ± 0.1	3.7 ± 0.7 x10 ⁻²	1.51
6.00 ^c	647 ± 2	5.73 ± 0.01 x10 ⁻²			
6.02 ^b	647 ± 90	11 ± 3 x10 ⁻²	0.88 ± 0.04	3.21 ± 0.05 x10 ⁻²	3.37
6.32 ^b	972 ± 22	11 ± 2 x10 ⁻²	0.70 ± 0.02	2.81 ± 0.04 x10 ⁻²	3.96
6.91 ^b	1160 ± 130	11 ± 3 x10 ⁻²	0.49 ± 0.02	2.1 ± 0.1 x10 ⁻²	5.00
7.75 ^c	1149 ± 6	27.4 ± 0.1 x10 ⁻²			

^aDownward stopped-flow pH jumps; final protein concentration of 3 μM.

^bUpward stopped-flow pH jumps; final protein concentration of 15 μM.

^cValues obtained by continuous flow at a final protein concentration of 10 μM; time scale is too fast to see slow phase.

^dF:S Amp is the ratio of amplitudes for the Fast:Slow kinetic phases.

Data taken from the Ph. D. thesis of Eydiejo Kurchan.^{90; 185}

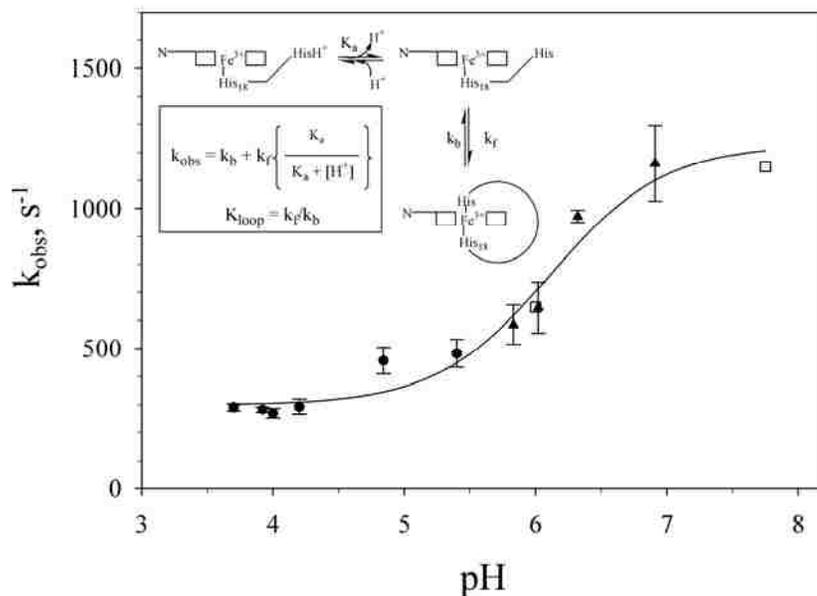


Figure 4.4: Plot of k_{obs} versus pH for the fast phase of His26-heme bond formation and breakage for the Ach26152 variant. The stopped flow data for downward (\bullet) and upward (\blacktriangle) pH jumps are shown with error bars corresponding to one standard deviation. For continuous flow upward pH jump data (\square) the error is smaller than the size of the symbol. The inset shows the mechanism for loop formation used to fit the pH dependence of k_{obs} .⁸⁹ The solid curve is a fit of the data to this mechanism. As previously,⁸⁹ the k_b reported in the text is the average and standard deviation of the four lowest pH data points.

The fit of the data in Figure 4.4 to the mechanism in the inset to Figure 4.4 yields $k_f = 930 \pm 60 \text{ s}^{-1}$ and $k_b = 280 \pm 10 \text{ s}^{-1}$. These values of k_f and k_b give $K_{\text{loop}} (= k_f/k_b)$ of 3.3 ± 0.3 . The $\text{p}K_a$ obtained from the fit is 6.1 ± 0.1 , reasonably consistent with deprotonation of His26 prior to loop formation. Thus, the fast phase appears to be attributable to intramolecular His-heme loop formation in the denatured state of this protein.

4.3.4 Sequential stopped-flow mixing

If the fast His26-heme bond formation and breakage phases in the denatured state are truly linked to the same reaction and are distinct from the slow phases, double jump stopped-flow experiments should confirm this assignment. Based on the data in Figure

4.4, we have used a pH of 7 to form His-heme bonds and a pH of 4 to break His-heme bonds in the denatured state (3 M gdnHCl).



Figure 4.5: Sequential mixing data for formation of His-heme bonds at different aging times for the Ach26152 variant in the denatured state (3 M gdnHCl) of iso-1-cytochrome *c*. Protein at pH 7 is premixed to pH 4 and held for the indicated aging time. The protein is then mixed to raise the pH back to 7 and data are acquired at 398 nm as a function of time. Data were obtained at 25 °C. The first ~1 ms of data are acquired prior to the stop. The pre-stop data are not used in fitting. Data taken from Ph. D. thesis of Eydiejo Kurchan.^{90; 185}

Table 4.4: Data for His-heme bond formation (end pH 7.10) from two independent sequential mixing experiments at 25 °C and 3 M gdnHCl for the AcH26I52 variant

Age Time (ms)	Slow Phase		Fast Phase	
	Rate (s ⁻¹)	Amplitude	Rate (s ⁻¹)	Amplitude
50	0.59 ± 0.01	-1.8 ± 0.05 x10 ⁻³	966 ± 77	0.03 ± 0.006
100	1.80 ± 0.09	1.8 ± 0.03 x10 ⁻³	768 ± 23	0.03 ± 0.013
200	2.04 ± 0.06	7.2 ± 0.07 x10 ⁻³	683 ± 32	0.03 ± 0.002
500	1.19 ± 0.01	12.2 ± 0.04 x10 ⁻³	1133 ± 34	0.08 ± 0.006
750	1.17 ± 0.01	13.4 ± 0.04 x10 ⁻³	801 ± 19	0.06 ± 0.002
1000	0.96 ± 0.01	12.7 ± 0.04 x10 ⁻³	1014 ± 25	0.08 ± 0.004
2500	1.02 ± 0.008	13.2 ± 0.03 x10 ⁻³	944 ± 17	0.08 ± 0.003
5000	1.30 ± 0.02	14.2 ± 0.06 x10 ⁻³	713 ± 24	0.03 ± 0.001

Age Time (ms)	Slow Phase		Fast Phase	
	Rate (s ⁻¹)	Amplitude	Rate (s ⁻¹)	Amplitude
5	0.83 ± 0.02	-4.2 ± 0.03 x10 ⁻³	1335 ± 35	0.18 ± 0.01
10	0.76 ± 0.03	-3.8 ± 0.05 x10 ⁻³	2798 ± 197	0.39 ± 0.12
25	0.66 ± 0.01	-3.9 ± 0.02 x10 ⁻³	3210 ± 310	1.4 ± 0.8
50	0.65 ± 0.03	-1.8 ± 0.03 x10 ⁻³	1596 ± 27	0.10 ± 0.004
75	-----	-----	1485 ± 78	0.10 ± 0.01
150	0.87 ± 0.02	4.2 ± 0.04 x10 ⁻³	2086 ± 103	0.18 ± 0.03
200	1.3 ± 0.001	9.8 ± 0.03 x10 ⁻³	1746 ± 65	0.09 ± 0.001
300	1.6 ± 0.2	9.4 ± 0.04 x10 ⁻³	1329 ± 35	0.08 ± 0.004
400	1.2 ± 0.02	11 ± 0.05 x10 ⁻³	1489 ± 40	0.12 ± 0.01
500	1.2 ± 0.01	12 ± 0.05 x10 ⁻³	915 ± 66	0.02 ± 0.002
1000	1.4 ± 0.01	14 ± 0.05 x10 ⁻³	1115 ± 28	0.07 ± 0.003

Data taken from the Ph. D. thesis of Eydiejo Kurchan.^{90, 185}

Double-jump experiments for His26-heme bond formation (pH 7 to pH 4 to pH 7) were carried out with aging times from 5 to 5000 ms at pH 4 (Figure 4.5, Table 4.4). At short aging times, a fast (1000-2000 s⁻¹) His26-heme bond formation phase dominates. At short aging times, the amplitude of the slow phase (~1 s⁻¹) is initially small and negative. The amplitude of the slow phase goes through zero at an aging time of ~75 ms

and then becomes positive (Figure 4.6A). A fit of the amplitude *versus* time data yields a rate constant $\sim 5.5 \text{ s}^{-1}$ (see legend to Figure 4.6A).

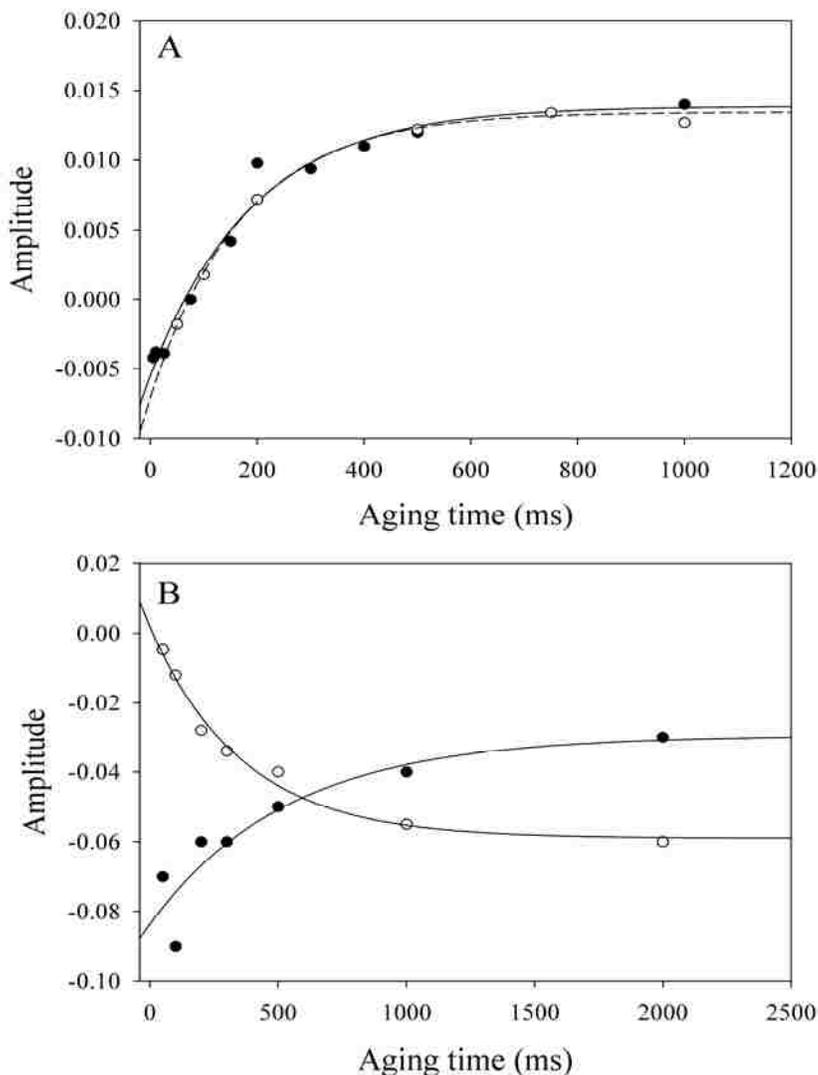


Figure 4.6: (A) Plot of amplitude *versus* aging time for the slow phase formation of the histidine-heme bond for two separate experiments (●, ○) for the AcH26I52 variant. The lines are fits to a single exponential rise to maximum equation. The rate constant for the growth of the slow phase amplitude for each data set is $k = 5.2 \pm 0.8 \text{ s}^{-1}$ (●, solid line) and $k = 5.8 \pm 0.5 \text{ s}^{-1}$ (○, dashed line). (B) Plot of amplitude *versus* aging time for the slow (○) and fast (●) phases for breakage of the His26-heme bond for the AcH26I52 variant. Solid lines are fits to a single exponential equation and give rate constants of $k = 2.8 \pm 0.4 \text{ s}^{-1}$ and $1.9 \pm 1.2 \text{ s}^{-1}$ for the slow and fast phase amplitudes, respectively. Data taken from the Ph. D. thesis of Eydiejo Kurchan.^{90; 185}

The negative amplitude observed for the slow phase with short aging times can be explained by a reversal of intermolecular His26-heme binding as the protein concentration is rapidly decreased from 60 to 15 μM in the sequential mixing experiment. Double-jump mixing experiments with an aging time of 2000 ms demonstrate that the fast phase is independent of protein concentration, whereas the slow phase is concentration-dependent and thus must be intermolecular (Table 4.5).

Table 4.5: Concentration dependence of His-heme bond formation for the AcH26I52 variant in sequential mixing experiments^a

[AcH26I52] ^b	Slow phase		Fast phase	
	Amplitude	Rate (s^{-1})	Amplitude	Rate (s^{-1})
7.5 μM	$7.1 \pm 0.05 \times 10^{-3}$	0.88 ± 0.02	0.10 ± 0.01	1294 ± 57
15 μM	$10.9 \pm 0.01 \times 10^{-3}$	1.2 ± 0.02	0.15 ± 0.01	1006 ± 29
30 μM	$16.9 \pm 0.01 \times 10^{-3}$	2.1 ± 0.02	0.21 ± 0.02	1259 ± 32
60 μM	$31.1 \pm 0.10 \times 10^{-3}$	2.42 ± 0.02	0.38 ± 0.02	1356 ± 45

^aExperiments were done with an aging time of 2000 ms at 25 °C in 3 M gdnHCl. Initial pH was 3.75 and final pH after mixing was 7.1.

^bFinal protein concentration after mixing.

Data taken from the Ph. D. thesis of Eydiejo Kurchan.^{90; 185}

In double-jump His-heme bond breakage experiments [pH 4 to 7 to 4 (Figure 4.7 and Table 4.6)], a fast phase ($300\text{-}400 \text{ s}^{-1}$) dominates at short aging times. The amplitude for the slow phase ($\sim 5 \text{ s}^{-1}$) increases as aging time increases in parallel with a decrease in the fast phase amplitude (Figure 4.6B). The rate constants for the growth of the slow phase amplitude and the reduction of the fast phase amplitude are within error the same ($\sim 2.5 \text{ s}^{-1}$, see the legend to Figure 4.6B).



Figure 4.7: Sequential mixing data for breakage of His-heme bonds at different aging times for the AcH26I52 variant in the denatured state of iso-1-cytochrome *c* (3 M gdnHCl). Protein at pH 4 is premixed to pH 7 and held for the indicated aging time. The protein is then mixed to lower the pH back to 4 and data are acquired at 398 nm as a function of time. Data were obtained at 25 °C. The first ~1 ms of data are acquired prior to the stop. The pre-stop data were not used in fitting. Data obtained from the Ph. D. thesis of Eydiejo Kurchan.^{90, 185}

Table 4.6: Data from His-heme bond breakage (end pH 3.75) sequential mixing experiments at 25 °C and 3 M gdnHCl for the AcH26I52 variant

Age Time (ms)	Slow Phase		Fast Phase	
	Rate (s ⁻¹)	Amplitude	Rate (s ⁻¹)	Amplitude
50	3.84 ± 0.19	-4.7 ± 0.08 x10 ⁻³	346 ± 4	-7.0 ± 0.1 x10 ⁻²
100	6.06 ± 0.08	-1.2 ± 0.1 x10 ⁻²	418 ± 4	-9.0 ± 0.1 x10 ⁻²
200	5.06 ± 0.05	-2.8 ± 0.01 x10 ⁻²	302 ± 3	-6.0 ± 0.05 x10 ⁻²
300	7.14 ± 0.03	-3.4 ± 0.01 x10 ⁻²	360 ± 4	-6.0 ± 0.07 x10 ⁻²
500	6.60 ± 0.02	-4.0 ± 0.01 x10 ⁻²	307 ± 3	-5.0 ± 0.05 x10 ⁻²
1000	7.0 ± 0.02	-5.5 ± 0.01 x10 ⁻²	308 ± 5	-4.0 ± 0.05 x10 ⁻²
2000	6.61 ± 0.02	-6.0 ± 0.01 x10 ⁻²	266 ± 4	-3.0 ± 0.04 x10 ⁻²

We note that the observed rate constant of $5.0 \pm 1.2 \text{ s}^{-1}$ for the slow phase in the double jump His26-heme bond breakage experiments (Table 4.6) matches the rate constant of $\sim 5.5 \text{ s}^{-1}$ at which the slow amplitude grows in double-jump His-heme bond formation experiments. Similarly, the rate constant for slow His26-heme bond formation at $30 \text{ }\mu\text{M}$ protein concentration ($2.1 \pm 0.02 \text{ s}^{-1}$, see Table 4.5) is similar to the rate constant of $\sim 2.5 \text{ s}^{-1}$ for the growth of the slow phase amplitude in double-jump His26-heme bond breakage experiments (aging at pH 7 is conducted at $30 \text{ }\mu\text{M}$ AcH26I52 in the double-jump His26-heme bond breakage experiments). Thus, growth of the amplitude of slow His26-heme bond breakage requires formation of the His26-heme bonds that form at a slow rate and growth of the amplitude of slow His26-heme bond formation requires breakage of the His26-heme bonds that break at a slow rate. Therefore, double-jump experiments establish that slow His26-heme bond formation is linked to slow His26-heme bond breakage and fast His26-heme bond formation is linked to fast His26-heme bond breakage.

4.3.5 Effect of Pro25 on the concentration dependence of the rates of denatured state His26-heme bond formation and breakage

The equilibrium data in Figure 4.3 and Table 4.2 show that simple mutation of Pro25→Ala diminishes the intermolecular component of His26-heme bond formation in the denatured state (3 M gdnHCl). To probe the basis for this difference, the concentration dependence of both His26-heme bond formation and breakage was examined using pH-jump experiments in 3 M gdnHCl for both the AcH26I52 and AcA25H26I52 variants. The observed kinetics is complex, requiring three to four

exponential components to provide satisfactory fits to the data (Tables 4.7 to 4.10 and Figures 4.8 to 4.11).

Table 4.7: Kinetic phases for His-heme bond formation at 25 °C and 3 M gdnHCl for AcH26I52 iso-1-cytochrome *c* as a function of protein concentration

[AcH26I52] ^a (μM)	Very Fast Phase		Fast phase		Medium phase		Slow phase	
	Amplitude	Rate Constant (s ⁻¹)	Amplitude	Rate Constant (s ⁻¹)	Amplitude	Rate Constant (s ⁻¹)	Amplitude	Rate Constant (s ⁻¹)
2 ^b	0.005 ± 0.003	1140 ± 540	0.0004 ± 0.0001	120 ± 50	0.00062 ± 0.00004	0.9 ± 0.1	0.00088 ± 0.00006	0.21 ± 0.01
3.75 ^b	0.011 ± 0.009	1460 ± 480	0.0009 ± 0.0003	170 ± 70	0.0016 ± 0.0003	1.1 ± 0.4	0.0017 ± 0.0003	0.30 ± 0.03
7.5 ^b	0.030 ± 0.015	1570 ± 410	0.0011 ± 0.0003	150 ± 40	0.0055 ± 0.0005	1.6 ± 0.2	0.0050 ± 0.0005	0.50 ± 0.04
15	0.039 ± 0.019	1720 ± 390	0.0015 ± 0.0006	180 ± 100	0.0098± 0.0016	2.6 ± 0.4	0.011 ± 0.001	0.72 ± 0.05
30	0.082 ± 0.082	1540 ± 650	0.0019 ± 0.0004	100 ± 70	0.0229 ± 0.0015	3.3 ± 0.4	0.019 ± 0.003	0.95 ± 0.06
45	0.155 ± 0.065	1900 ± 270	0.0030 ± 0.0018	75 ± 74	0.033 ± 0.002	3.9 ± 0.6	0.026 ± 0.004	1.1 ± 0.1
60	0.213 ± 0.104	2260 ± 450	0.002 ± 0.000	85 ^c	0.048 ± 0.005	4.9 ± 0.4	0.033 ± 0.003	1.33 ± 0.06

^aFinal concentration after 1:1 mixing.

^bAmplitude data at 2, 3.75 and 7.5 μM, which were collected with a 5 mm pathlength, have been divided by 5 to normalize them relative to the higher concentration data which was collected with a 1 mm pathlength.

^cA triple exponential provides an adequate fit at 60 μM. For consistency with other data fits a fit to a quadruple exponential was carried out with the fast phase constrained to a value in between that observed at 30 and 45 μM. The parameters from the triple exponential fit at 60 μM AcH26I25 are $A_{\text{fast}} = 0.18 \pm 0.09$, $k_{\text{fast}} = 2070 \pm 450 \text{ s}^{-1}$, $A_{\text{med}} = 0.040 \pm 0.006$, $k_{\text{med}} = 6.4 \pm 0.7 \text{ s}^{-1}$, $A_{\text{slow}} = 0.041 \pm 0.003$, $k_{\text{slow}} = 1.50 \pm 0.07 \text{ s}^{-1}$.

Table 4.8: Kinetic phases for His-heme bond breakage at 25 °C and 3 M gdnHCl for AcH26I52 iso-1-cytochrome *c* as a function of protein concentration

[AcH26I52] ^a (μM)	Fast Phase		Intermediate phase		Slow phase	
	Amplitude	Rate Constant (s^{-1})	Amplitude	Rate Constant (s^{-1})	Amplitude	Rate Constant (s^{-1})
2 ^b	0.0033 \pm 0.0002	269 \pm 21	0.0004 \pm 0.0001	56 \pm 21	0.00432 \pm 0.00006	6.32 \pm 0.06
3.75 ^b	0.0054 \pm 0.0005	288 \pm 12	0.0009 \pm 0.0001	44 \pm 11	0.0105 \pm 0.0002	6.23 \pm 0.07
7.5 ^b	0.0086 \pm 0.0005	276 \pm 22	0.0018 \pm 0.0002	44 \pm 10	0.0255 \pm 0.0003	6.32 \pm 0.03
15	0.0143 \pm 0.0007	259 \pm 22	0.0058 \pm 0.0003	32 \pm 5	0.0657 \pm 0.0007	6.45 \pm 0.05
30	0.021 \pm 0.002	241 \pm 23	0.0150 \pm 0.0004	28 \pm 3	0.113 \pm 0.002	6.62 \pm 0.06
45	0.027 \pm 0.001	229 \pm 16	0.030 \pm 0.001	26 \pm 1	0.157 \pm 0.001	6.86 \pm 0.04
60	0.036 \pm 0.004	214 \pm 24	0.062 \pm 0.003	24 \pm 2	0.205 \pm 0.006	7.01 \pm 0.10

^aFinal concentration after 1:1 mixing.

^bAmplitude data at 2, 3.75 and 7.5 μM which were collected with a 5 mm pathlength have been divided by 5 to normalize them relative to the higher concentration data which was collected with a 1 mm pathlength.

Table 4.9: Kinetic phases for His-heme bond formation at 25 °C and 3 M gdnHCl for AcA25H26I52 iso-1-cytochrome *c* as a function of protein concentration

[AcA25H26I52] ^a (μ M)	Very Fast Phase		Fast phase		Medium phase		Slow phase	
	Amplitude	Rate Constant (s ⁻¹)	Amplitude	Rate Constant (s ⁻¹)	Amplitude	Rate Constant (s ⁻¹)	Amplitude	Rate Constant (s ⁻¹)
2 ^b	0.007 \pm 0.010	1540 \pm 910	0.0010 \pm 0.0006	210 \pm 130			0.00063 \pm 0.00008	0.63 \pm 0.13
3.75 ^b	0.009 \pm 0.004	1220 \pm 310	0.0008 \pm 0.0003	160 \pm 50			0.0012 \pm 0.0002	0.69 \pm 0.13
7.5 ^b	0.030 \pm 0.017	1600 \pm 380	0.0013 \pm 0.0003	170 \pm 70			0.0041 \pm 0.0003	0.82 \pm 0.07
15 ^c	0.063 \pm 0.029	1780 \pm 420	0.0025 \pm 0.0022	300 \pm 150	0.0016 \pm 0.0002	4.0 \pm 0.8	0.0122 \pm 0.0004	0.87 \pm 0.02
30 ^c	0.105 \pm 0.041	1840 \pm 390	0.0030 \pm 0.0021	390 \pm 180	0.0032 \pm 0.0015	4.6 \pm 1.2	0.0219 \pm 0.0011	1.16 \pm 0.10
45	0.125 \pm 0.043	1450 \pm 170			0.0109 \pm 0.0026	4.0 \pm 1.0	0.0312 \pm 0.0031	1.26 \pm 0.08
60	0.158 \pm 0.050	1610 \pm 200			0.0219 \pm 0.0064	4.0 \pm 0.9	0.0383 \pm 0.0071	1.41 \pm 0.14

^aFinal concentration after 1:1 mixing.

^bAmplitude data at 2, 3.75 and 7.5 μ M which were collected with a 5 mm pathlength have been divided by 5 to normalize them relative to the higher concentration data which were collected with a 1 mm pathlength.

^cParameters at 15 and 30 μ M are from fitting the data to a quadruple exponential decay function.

Parameters at other concentrations are from fitting the data to a triple exponential decay function.

Table 4.10: Kinetic phases for His-heme bond breakage at 25 °C and 3 M gdnHCl for AcA25H26I52 iso-1-cytochrome *c* as a function of protein concentration

[AcA25H26I52] ^a (μM)	Fast Phase		Intermediate phase		Slow phase	
	Amplitude	Rate Constant (s^{-1})	Amplitude	Rate Constant (s^{-1})	Amplitude	Rate Constant (s^{-1})
2 ^{b,c}	0.0067 \pm 0.0002	256 \pm 9			0.00315 \pm 0.00004	22.6 \pm 0.6
3.75 ^b	0.0099 \pm 0.0004	258 \pm 9	0.0009 \pm 0.0004	37 \pm 8	0.0067 \pm 0.0004	22.0 \pm 0.5
7.5 ^b	0.0184 \pm 0.0008	240 \pm 6	0.006 \pm 0.002	34 \pm 6	0.019 \pm 0.003	21.8 \pm 0.3
15	0.026 \pm 0.001	229 \pm 8	0.017 \pm 0.013	33 \pm 6	0.033 \pm 0.013	18.6 \pm 2.9
30	0.035 \pm 0.001	238 \pm 13	0.0312 \pm 0.0005	40 \pm 4	0.063 \pm 0.005	19.4 \pm 0.5
45	0.045 \pm 0.002	238 \pm 19	0.046 \pm 0.005	46 \pm 5	0.091 \pm 0.007	20.1 \pm 0.6
60	0.050 \pm 0.002	240 \pm 18	0.062 \pm 0.005	51 \pm 5	0.110 \pm 0.007	20.6 \pm 0.5

^aFinal concentration after 1:1 mixing.

^bAmplitude data at 2, 3.75 and 7.5 μM which were collected with a 5 mm pathlength have been divided by 5 to normalize them relative to the higher concentration data which was collected with a 1 mm pathlength.

^cThere was no significant difference in the fit to a double *versus* a triple exponential function, therefore the parameters from a double exponential fit are reported.

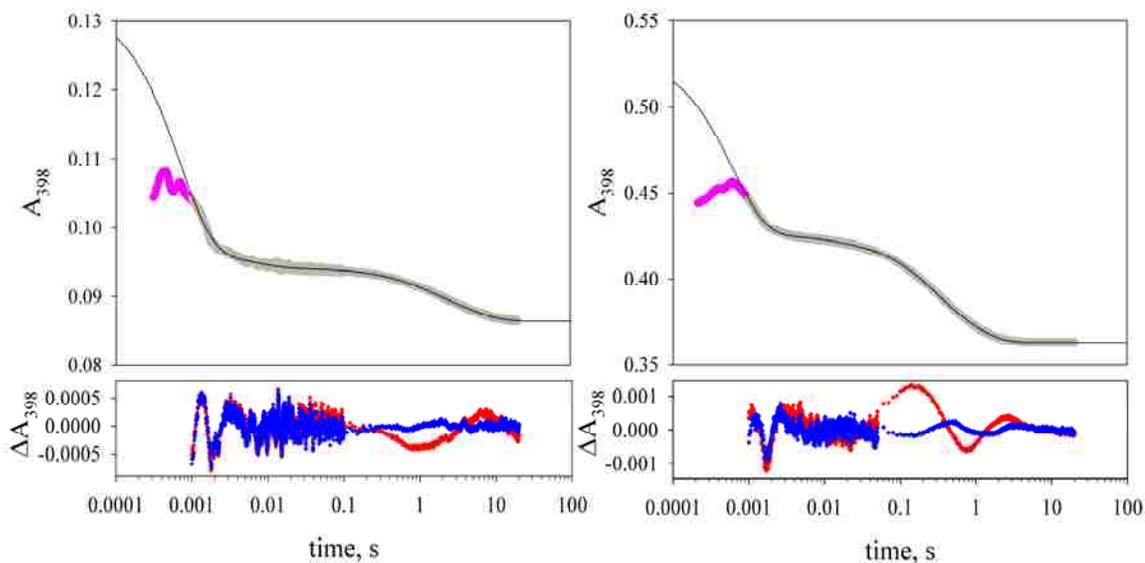


Figure 4.8: A_{398} versus time data for upward pH jumps collected for the Ach26I52 variant (upper panels) at final concentrations of 2 μM (left panels) using the 5 mm pathlength and 45 μM (right panels) using the 1 mm pathlength of the 5 μL flow cell. Data collected with logarithmic time sampling on a 100 (2 μM) or 50 (45 μM) ms time scale with pressure hold were merged with data collected with linear time sampling on a 20 s time scale. The fits shown are to a quadruple exponential decay function. The gray data points were used in the fit. The magenta data are pre-stop data. A small signal glitch is seen in the data near 2 ms (upper panels, see also residuals). The lower panels compare the residuals for fits of the data to a triple (red) versus a quadruple (blue) exponential decay function. Parameters from fits are reported in Table 4.7.

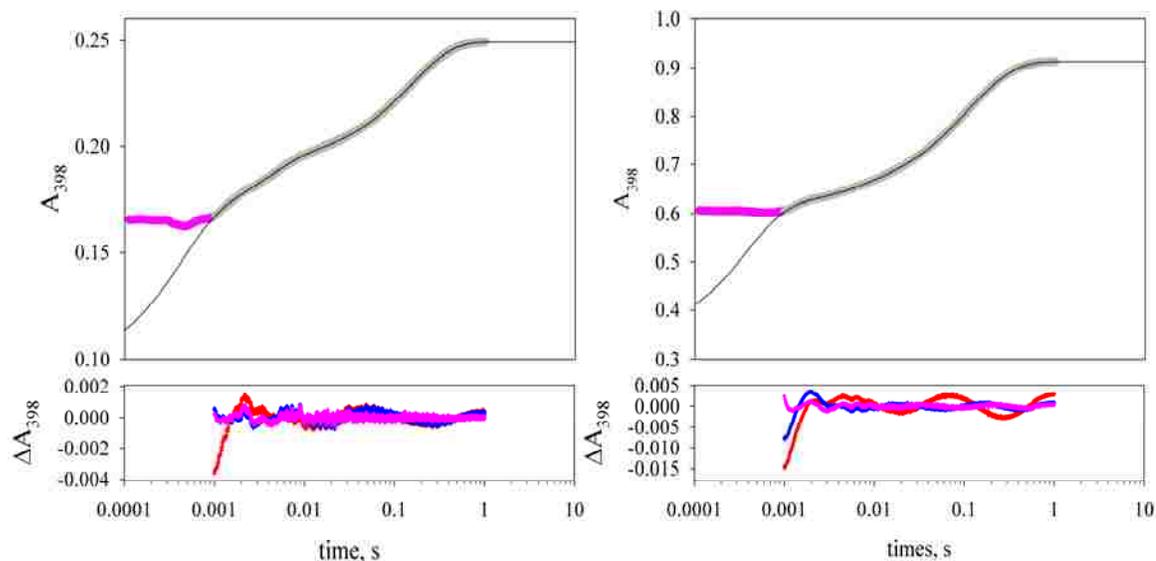


Figure 4.9: A_{398} versus time data for downward pH jumps collected for the AcH26I52 variant (upper panels) at final concentrations of $3.75 \mu\text{M}$ (left panels) using the 5 mm pathlength and $60 \mu\text{M}$ (right panels) using the 1 mm pathlength of the $5 \mu\text{L}$ flow cell. Data were collected with logarithmic time sampling on a 1 s time scale with pressure hold. The fits shown are to a quadruple exponential rise to maximum function. The gray data points were used in the fit. The magenta data are pre-stop data. The lower panels compare the residuals for fits of the data to double (red), triple (blue) and quadruple (magenta) exponential decay functions. In these examples the glitch in the data near 2 ms (upper panels, see also residuals) is fairly prominent, and a quadruple exponential provides a better fit to the data beyond 5 ms. The fastest of the four phases is likely an artifact due to the presence of the glitch. Traces where the glitch is less prominent (data not shown) provide nearly identical parameters for the three slower phases with either the triple or quadruple exponential functions, indicating that the strategy of using a quadruple exponential function to deal with the data glitch is reasonable. Parameters from fits are reported in Table 4.8.

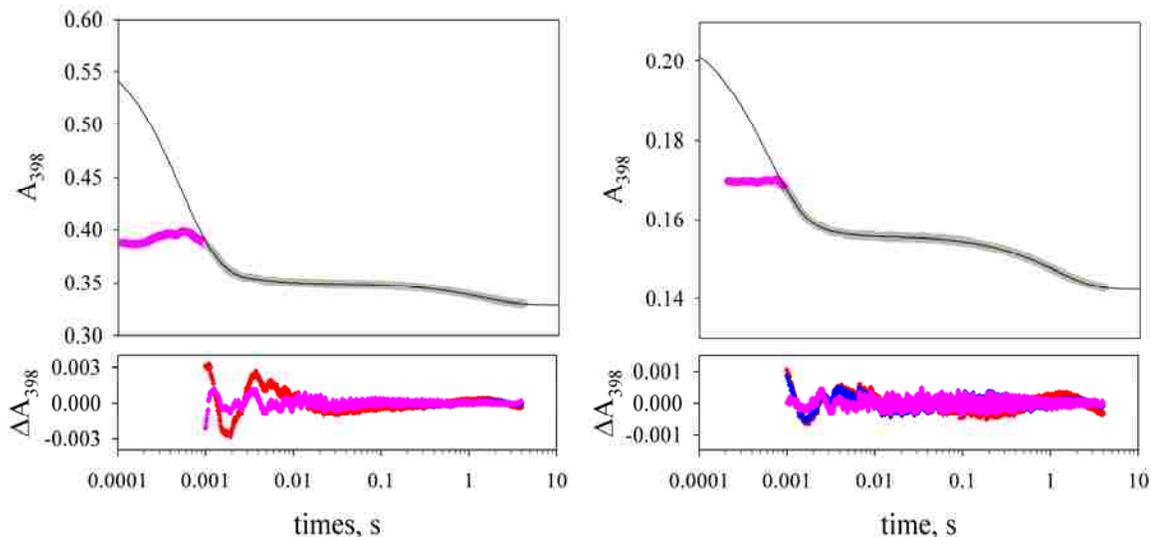


Figure 4.10: A_{398} versus time data for upward pH jumps collected for the AcA25H26I52 variant (upper panels) at final concentrations of 7.5 μM using the 5 mm pathlength and 15 μM using the 1 mm pathlength of the 5 μL flow cell. Data were collected with logarithmic time sampling on a 5 s time scale with pressure hold. The fit shown at 7.5 μM is to a triple exponential decay function and that at 15 μM is a quadruple exponential decay function. The gray data points were used in the fit. The magenta data are pre-stop data. The lower panels compare the residuals for fits of the data to double (red), triple (blue) and quadruple (magenta) exponential decay functions. A small signal glitch is observed near 2 ms (upper panels, see also residuals). The improvement in the data fit was marginal for the quadruple versus the triple exponential decay function at 7.5 μM protein concentration, so parameters from the triple exponential fit are reported. The improvement in the data fit was significant for the quadruple versus the triple exponential decay function at 15 μM protein concentration, so parameters from the quadruple exponential fit are reported. Parameters from fits are reported in Table 4.9.

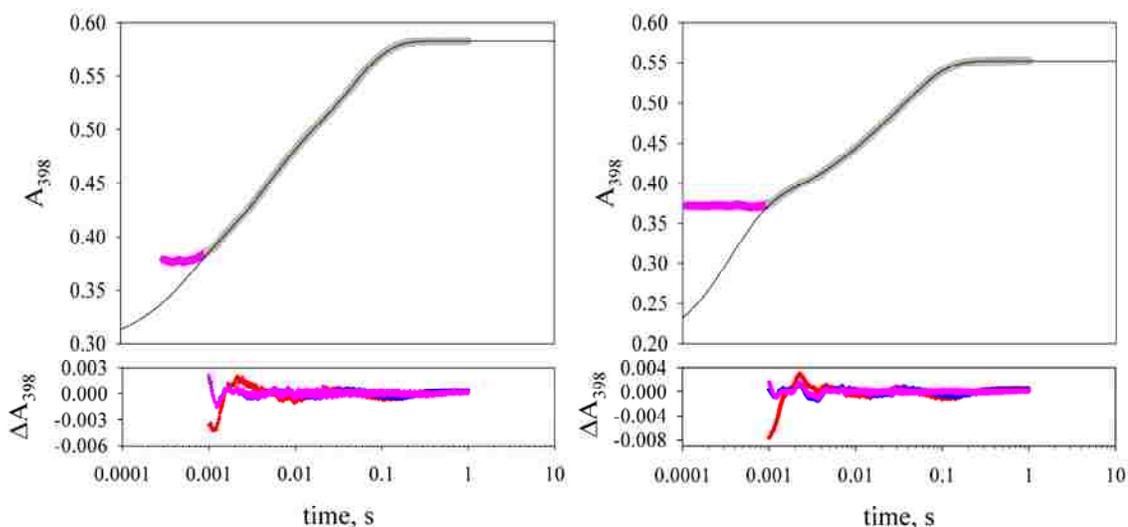


Figure 4.11: A_{398} versus time data for downward pH jumps collected for the AcA25H26I52 variant (upper panels) at final concentrations of $7.5 \mu\text{M}$ using the 5 mm pathlength and $45 \mu\text{M}$ using the 1 mm pathlength of the $5 \mu\text{L}$ flow cell. Data were collected with logarithmic time sampling on a 1 s time scale with pressure hold. The fits shown are to a quadruple exponential rise to maximum function. The gray data points were used in the fit. The magenta data are pre-stop data. The lower panels compare the residuals for fits of the data to double (red), triple (blue) and quadruple (magenta) exponential decay functions. Due to the small glitch in the data near 2 ms (upper panels, see also residuals), a quadruple exponential provides a better fit to the data beyond 5 ms. The fastest of the four phases is likely an artifact due to the presence of the glitch. Traces where the glitch is less prominent (data not shown) provide nearly identical parameters for the three slower phases with either the triple or quadruple exponential functions, indicating that the strategy of using a quadruple exponential function to deal with the data glitch is reasonable. Parameters from fits are reported in Table 4.10.

For upward pH-jump data, the best fit for both variants is obtained using a quadruple-exponential function. The fractional amplitudes of three of the phases are nearly invariant with protein concentration (Figures 4.12 and 4.13). The fourth phase diminishes in fractional amplitude as concentration increases and is likely due to a small artifact observed in absorbance data near 2 ms (see traces in Figures 4.8 to 4.11). If this phase is disregarded, a fast phase with a rate constant near 1600 s^{-1} (AcH26I52, $1650 \pm 350 \text{ s}^{-1}$; AcA25H26I52, $1580 \pm 200 \text{ s}^{-1}$) and two slow phases are observed. The ability to discern two slow phases in these experiments *versus* the double-jump experiments likely results from the superior signal to noise of the SX20 stopped-flow spectrometer used in

the single jump experiments. Both slow phases are concentration-dependent for the AcH26I52 variant (Figure 4.14), with k_{obs} leveling off at higher protein concentration. For the AcA25H26I52 variant, the faster of the two slow phases is not observed at lower protein concentration and within error is independent of protein concentration [$4 \pm 1 \text{ s}^{-1}$ (Figure 4.14)]. The slower of the two phases is concentration-dependent.

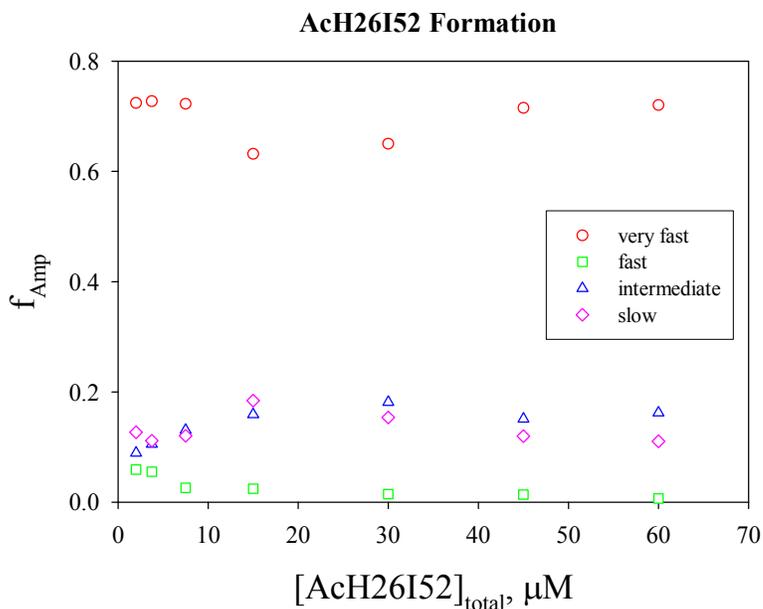


Figure 4.12: Concentration dependence of the fractional amplitudes, f_{Amp} , of the kinetic phases for loop formation for the AcH26I52 variant of iso-1-cytochrome *c*. Data were acquired at 25 °C in 3 M gdnHCl as described in Materials and Methods.

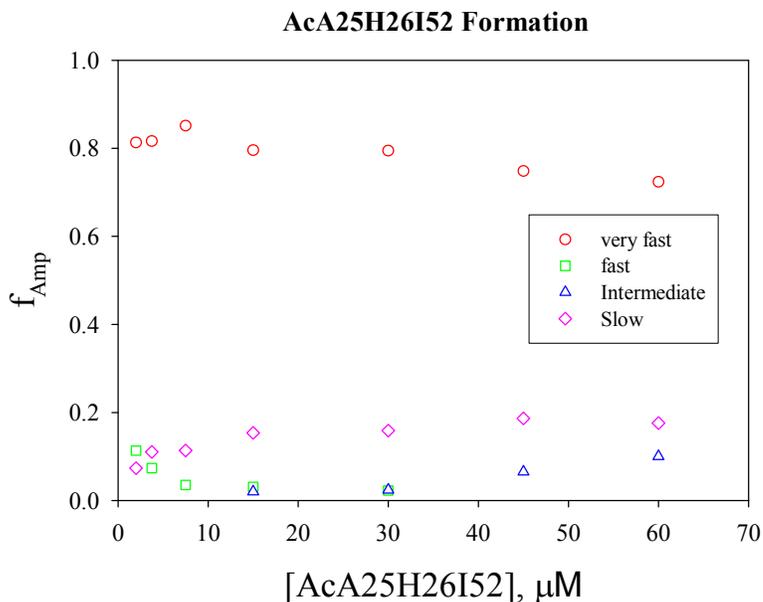


Figure 4.13: Concentration dependence of the fractional amplitudes, f_{Amp} , of the kinetic phases for loop formation for the AcA25H26I52 variant of iso-1-cytochrome *c*. Data were acquired at 25 °C in 3 M gdnHCl as described in Materials and Methods.

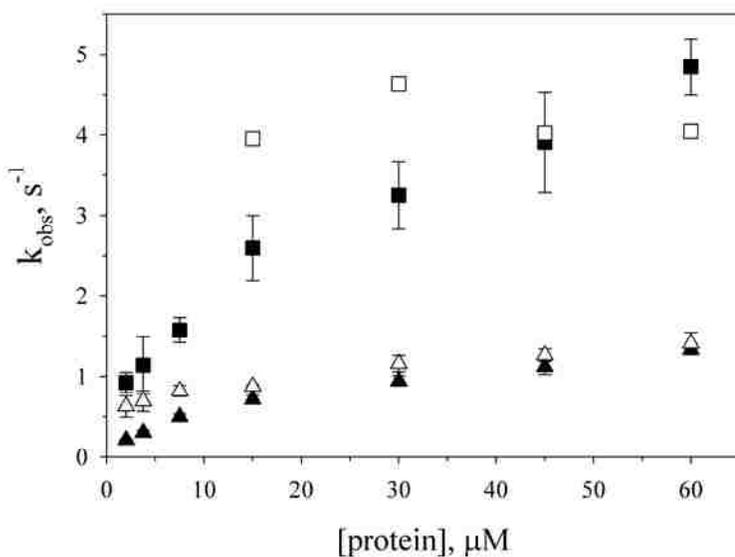


Figure 4.14: Concentration dependence of the slow phase rate constants for His-heme bond formation with the AcH26I52 (\blacksquare , \blacktriangle) and AcA25H26I52 (\square , \triangle) variants. Concentration is final concentration after mixing. Data were acquired at 25 °C in 3 M gdnHCl as described in the Materials and Methods. The error bars are one standard deviation of the mean. For clarity, errors bars of $\sim \pm 1 \text{ s}^{-1}$ are omitted for the faster of the two slow phases (\square) for the AcA25H26I52 variant.

For downward pH-jump data (His-heme bond breakage), three kinetic phases are observed for both proteins. The rate constants for these phases appear to be independent of protein concentration (AcH26I52, $254 \pm 27 \text{ s}^{-1}$; $36 \pm 12 \text{ s}^{-1}$ and $6.5 \pm 0.3 \text{ s}^{-1}$; AcA25H26I52, $243 \pm 10 \text{ s}^{-1}$; $40 \pm 7 \text{ s}^{-1}$ and $20.7 \pm 1.5 \text{ s}^{-1}$). At the lowest concentration of protein used, the intermediate phase cannot be discerned for the AcA25H26I52 variant. For both variants, at low protein concentration, the fractional amplitude, f_{Amp} , for the fast phase for loop breakage is similar to or larger than f_{Amp} for the slow phase (Figure 4.15). At higher protein concentrations, the slow phase dominates the amplitude for the breakage reaction, although the concentration dependence of the fractional amplitudes of the three phases saturates at an initial protein concentration of greater than $\sim 30 \mu\text{M}$. The total amplitudes for His26-heme bond breakage and His26-heme bond formation are similar to each other and are linearly dependent on protein concentration (insets in Figure 4.15), indicating that the full reaction is being detected in both directions and at all protein concentrations.

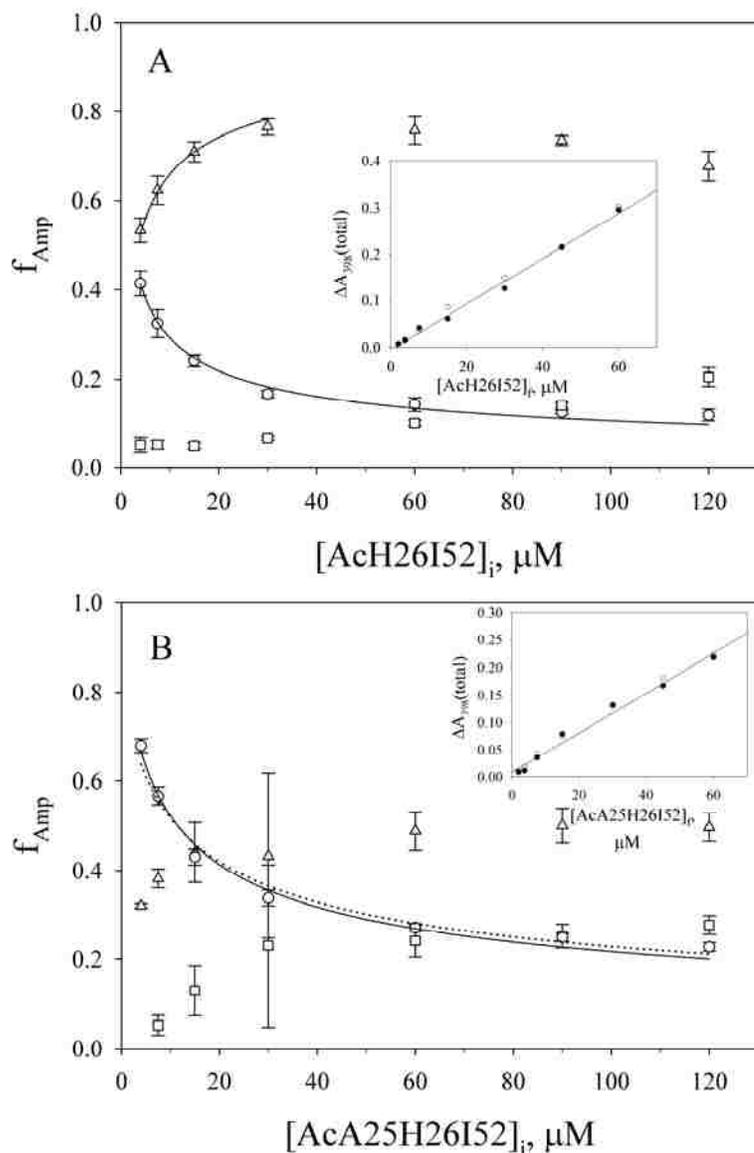


Figure 4.15: Concentration dependence of the fractional amplitudes, f_{Amp} , of the kinetics phases for His26-heme bond breakage for the (A) AcH26I52 and (B) AcA25H26I52 variants of iso-1-cytochrome *c*. Data were acquired at 25 °C in 3 M gdnHCl as described in Materials and Methods. The f_{Amp} for the fast phase (\circ), the intermediate phase (\square) and the slow phase (Δ) are shown with error bars derived from standard propagation of the error in the observed amplitudes. The concentrations prior to mixing, $[\text{AcH26I52}]_i$ and $[\text{AcA25H26I52}]_i$, are plotted on the x-axes, since f_{Amp} reflects the equilibrium distribution of species prior to mixing. The solid lines are fits to the equilibrium model in Appendix C as described in the Discussion. The dotted line in panel B is a fit to Eq 4.2 (Discussion) with the $K_{\text{loop}}(\text{obs})$ constrained to 4.65. The inset in each panel shows the total amplitude, $\Delta A_{398}(\text{total})$, versus protein concentration after mixing, $[\text{AcH26I52}]_f$ and $[\text{AcA25H26I52}]_f$, for the His26-heme bond formation (\bullet) and breakage (\circ) reactions. The solid line is a linear fit to $\Delta A_{398}(\text{total})$ versus protein concentration for the His26-heme bond formation amplitudes.

Comparison of the fast phases (due to intramolecular loop formation and breakage) from these concentration-dependent kinetic experiments with the fast phases from the pH-dependent His26-heme bond formation/breakage experiments in Figure 4.4 shows that the rate constants obtained for loop formation, k_f , and loop breakage, k_b , are similar. The magnitudes of k_b obtained for His26-heme bond breakage at pH 3.75 for the AcH26I52 and AcA25H26I52 variants ($\sim 250 \text{ s}^{-1}$) are similar to $k_b \sim 280 \text{ s}^{-1}$ obtained for the AcH26I52 variant in Figure 4.4. On the basis of the mechanism in the inset of Figure 4.4, k_f for intramolecular loop formation from the protein concentration-dependent kinetic experiments can be approximated as the difference between the fast rate constant for His26-heme bond formation at pH 7.1 and the fast breakage rate constant at pH 3.75. This approximation yields k_f values of $1400 \pm 380 \text{ s}^{-1}$ and $1340 \pm 210 \text{ s}^{-1}$ for the AcH26I52 and AcA25H26I52 variants, respectively. Again, these values are reasonably similar to $k_f = 930 \pm 60 \text{ s}^{-1}$ observed with the pH dependent data for the AcH26I52 variant in Figure 4.4.

4.4 Discussion

Our equilibrium data on the effect of protein concentration on His-heme bond formation in the denatured state of iso-1-cytochrome *c* demonstrate that the contribution of the intermolecular reaction drops off rapidly with the distance of the histidine from the heme (Table 4.2). His22, which is five residues from the heme, has a 100-fold larger intermolecular equilibrium constant, K_A , than His27 which is 10 residues from the heme. We also find that the intermolecular reaction is very sensitive to local sequence. The

Pro25→Ala mutation decreases K_A for His26 by ~10-fold. Comparison with data for iso-2-cytochrome *c* also supports this sensitivity to local sequence.¹²¹ For iso-2-cytochrome *c*, $pK_a(\text{obs}) = 5.3$ at a protein concentration of 60 μM , much higher than would be expected if an intermolecular reaction was occurring. Thus, the small sequence differences between iso-1- and iso-2- cytochrome *c* for the residues in between His26 and the heme (Val20→Ile and Lys22→Glu) clearly affect intermolecular aggregation as is observed for the Pro25→Ala mutation. Aggregation is also strongly dependent on solution conditions. Data for the AcH26I52 variant in 6 M gdHCl ⁸² yield a K_A of $(1.3 \pm 1.0) \times 10^{-6}$ M, 1 order of magnitude lower than in 3 M gdHCl . Thus, the intermolecular aggregation we observe here is likely to be more prominent under folding conditions.

An interesting feature of our concentration-dependent kinetic data is that the amplitudes of His26-heme bond formation phases are independent of protein concentration (Figures 4.12 and 4.13), whereas the amplitudes of the His26-heme bond breakage phases are strongly concentration-dependent (Figure 4.15). For His26-heme bond formation, the fast intramolecular phase always dominates. By contrast, for His26-heme bond breakage, the fast intramolecular phase diminishes in importance rapidly as protein concentration increases. Thus, it is evident that aggregation is under equilibrium control. Previous models for the outcome of aggregation *versus* folding are based on kinetic partitioning between irreversible folding and aggregation reactions.¹⁷⁷ We develop a model for equilibrium control of partitioning between folding (or loop formation) and aggregation, compare the predictions of the two models, and discuss the implications for evolution of proteins. First we discuss the nature of the slow aggregation

phases and then we use our data to extract equilibrium and kinetic parameters necessary to compare kinetically irreversible and equilibrium models for aggregation.

4.4.1 Nature of the slow phases for loop formation and breakage

The concentration dependence of the slow phases for His26-heme bond formation clearly indicates an intermolecular process. For the AcH26I52 variant, adequate fits of the slow phase data are obtained with two similar-amplitude exponential components with rate constants that differ by a factor of ~ 3 (Figure 4.14, Table 4.7). This method of fitting the data is clearly an over-simplification. However, attempts to fit the data to a simple second-order rate process produced worse fits. The two slow phases for His26-heme bond formation are matched by two slow phases for His-heme bond breakage. In Figure 4.15, the amplitude of the slowest His26-heme bond breakage rate ($6.5 \pm 0.3 \text{ s}^{-1}$) starts to lose amplitude to the intermediate breakage rate ($36 \pm 12 \text{ s}^{-1}$) at high concentrations of the AcH26I52 variant. This observation suggests that at high concentration, higher-order (less kinetically stable) aggregates have formed. It is also possible that the two different slow phases represent intermolecular His26-heme bond formation and breakage with the Gly24-Pro25 peptide bond in *cis* versus *trans* conformations. However, this is inconsistent with the observed decrease in the amplitude of the slower breakage rate at higher high concentrations of the AcH26I52 variant. Thus, the data support a mechanism involving His-heme dimerization at lower protein concentrations with contributions from higher-order aggregation at higher concentrations.

For the AcA25H26I52 variant, two slow phases for His26-heme bond formation and breakage are also observed. The rate of the dominant slow phase ($0.6 - 1.4 \text{ s}^{-1}$) for

His26-heme bond formation depends on the concentration of the AcA25H26I52 variant, whereas the lower-amplitude slow phase ($\sim 4 \text{ s}^{-1}$), which only appears at a final concentration of $\geq 15 \text{ }\mu\text{M}$, does not. The reason for the lack of concentration dependence is unclear (although the large error bars on this rate constant may obscure concentration dependence). The rate constants for the two slow His26-heme breakage phases are too similar ($40 \pm 7 \text{ s}^{-1}$ and $20.7 \pm 1.5 \text{ s}^{-1}$) to be fully resolved.

When AcH26I52 is compared to AcA25H26I52, both have a slow phase for His26-heme bond breakage with a rate near 40 s^{-1} . However, the slower of the two His26-heme bond breakage phases is ~ 3 -fold faster for the AcA25H26I52 variant. Thus, both have intermolecular processes attributable to dimerization and higher-order aggregation. The higher-order aggregates appear to have similar kinetic stability.

4.4.2 Extraction of equilibrium parameters from the kinetics of intramolecular loop formation

For both variants, the fractional amplitudes of the fast and slow phases of His26-heme bond formation are nearly invariant with protein concentration (Figures 4.12 and 4.13), indicating that the amplitudes are controlled by the relative rates of the individual processes. The average fractional amplitude of the fast His26-heme bond formation phase, f_{fast} , across all protein concentrations studied is 0.70 ± 0.04 for the AcH26I52 variant and 0.79 ± 0.04 for the AcA25H26I52 variant. Thus, the fractional amplitude of the fast phase assigned to intramolecular loop formation can be used to obtain an apparent equilibrium constant for loop formation [$K_{\text{loop}}(\text{obs}) = f_{\text{fast}}/(1-f_{\text{fast}})$] at pH 7.1

(Table 4.11). The values are similar to the intramolecular equilibrium constant obtained for the AcH26I52 variant derived from the pH dependence of k_{obs} (Figure 4.4).

Table 4.11: Equilibrium parameters for intramolecular and intermolecular His26-heme bond formation at pH 7.1 and 25 °C

Parameter and Method of Evaluation	Variant	
	AcH26I52	AcA25H26I52
$K_{\text{loop}}(\text{obs})$		
pH dependent kinetics	$3.3 \pm 0.3^{\text{a}}$	–
f_{fast} : His-heme bond formation amplitudes	2.3 ± 0.3	3.8 ± 0.8
$k_{\text{f}}(\text{fast})/k_{\text{b}}(\text{fast})$	5.5 ± 1.6	5.5 ± 0.9
f_{fast} : His-heme bond breakage amplitudes	8.7 ± 0.3	16.1 ± 0.9
f_{slow} : His-heme bond breakage amplitudes	10.6 ± 0.3	– ^b
$K_{\text{inter}}(\text{obs})$		
f_{fast} : His-heme bond breakage amplitudes	$3.00 \pm 0.08 \times 10^7 \text{ M}^{-1}$	$2.1 \pm 0.2 \times 10^7 \text{ M}^{-1}$
f_{slow} : His-heme bond breakage amplitudes	$3.00 \pm 0.07 \times 10^7 \text{ M}^{-1}$	– ^b

^a $K_{\text{loop}}(\text{obs})$ is for fully deprotonated His26 in this case, i.e., K_{loop} . Correcting to pH 7.1 with $K_{\text{loop}}(\text{pH}) = K_{\text{loop}}/(1 + 10^{(\text{pK}_{\text{a}} - \text{pH})})$ using $\text{pK}_{\text{a}} = 6.6$ for His26⁸² yields $K_{\text{loop}}(\text{obs}) \sim 2.5$.

^bThe data deviate too strongly from a simple dimerization in competition with intramolecular loop formation equilibrium to fit f_{slow} to this model.

$K_{\text{loop}}(\text{obs})$ at pH 7.1 can also be estimated from the measured rate constants for the fast forward and back reactions [$k_{\text{f}}(\text{fast})/k_{\text{b}}(\text{fast})$] for intramolecular loop formation obtained from our kinetic studies on the AcH26I52 and AcA25H26I52 variants *versus*

protein concentration. The values for $K_{loop}(obs)$ obtained in this way are similar to those obtained from the fast amplitude data, f_{fast} , for His26-heme bond formation (Table 4.11).

4.4.3 Extraction of equilibrium properties for intermolecular association from the kinetics of His26-heme bond breakage

The fractional amplitude data for loop breakage as a function of protein concentration (f_{Amp} for the fast phase in Figure 4.15) provide information about the relative contributions of intra- and intermolecular His26-heme bond formation, since these species are at equilibrium prior to the downward pH jump. Since all downward pH-jump experiments start at pH 7.1, we can evaluate apparent equilibrium constants, $K_{loop}(obs)$ and $K_{inter}(obs)$, for denatured state intramolecular loop formation and intermolecular association, respectively, at this pH. We approximate the equilibrium with the simplest possible model: His-heme dimerization in competition with intramolecular loop formation. The fraction of protein which forms an intramolecular loop, f_{loop} , is given by Eq 4.2 for this model (see Appendix C).

$$f_{loop} = \frac{-\left(1 + \frac{1}{K_{loop}(obs)}\right) + \sqrt{\left(1 + \frac{1}{K_{loop}(obs)}\right)^2 + 8\left(\frac{K_{inter}(obs)}{K_{loop}(obs)^2}\right)[Cytc]_t}}{4\left(\frac{K_{inter}(obs)}{K_{loop}(obs)^2}\right)[Cytc]_t} \quad (4.2)$$

This equation was used to fit the fast phase amplitude for the loop breakage reaction as a function of $[\text{Cyt}c]_i$ (\circ in Figure 4.15 A, B) to yield $K_{\text{loop}}(\text{obs})$ and $K_{\text{inter}}(\text{obs})$ (Table 4.11). $K_{\text{inter}}(\text{obs})$ is of the same order of magnitude for both variants, in contrast to the results from the concentration dependence of $\text{p}K_{\text{a}}(\text{obs})$ (Table 4.2). However, the parameters obtained from fits to Eq 4.2 are sensitive to small errors in the data. If we constrain $K_{\text{loop}}(\text{obs})$ to 4.65 (the average value from amplitude and rate constant data for the fast phase of His-heme bond formation and breakage for the AcA25H26I52 variant), the quality of the fit is only slightly diminished (dotted line Figure 4.15B), and we obtain a $K_{\text{inter}}(\text{obs})$ of $(1.5 \pm 0.1) \times 10^6 \text{ M}^{-1}$ which is more in line with the results from the concentration-dependent $\text{p}K_{\text{a}}(\text{obs})$ data. The poorer separation of dimerization and higher-order aggregation for the AcA25H26I52 variant relative to the AcH26I52 variant (Figure 4.15) probably also affects the accuracy of $K_{\text{inter}}(\text{obs})$ obtained for the AcA25H26I52 variant.

For the AcH26I52 variant, the rise in the fractional amplitude of the slow breakage phase, f_{slow} [Figure 4.15A (Δ)], can be fit to the equilibrium model (for equations see Appendix B) out to a protein concentration of 30 μM . The values for $K_{\text{loop}}(\text{obs})$ and $K_{\text{inter}}(\text{obs})$ are similar to those obtained by fitting Eq 4.2 to f_{fast} (Table 4.11). Thus, our simple dimerization model is adequate for describing the data at relatively low concentrations of AcH26I52. For the AcA25H26I52 variant, dimerization and higher-order aggregation are poorly separated. Thus, f_{slow} cannot be fit to a model involving simple loop formation in competition with dimerization.

4.4.4 How fast is bimolecular aggregation in the denatured state?

Reversible intermolecular aggregation has been found to affect the folding kinetics of a number of proteins.^{168; 169; 172-175} In these cases, aggregates accumulate during folding and are not pre-existing in the denatured state in a strong denaturant, although for the proteins CI2 and U1A, aggregation is believed to occur from the denatured state under folding conditions.^{168; 172; 173} The bimolecular rates for association of denatured protein molecules under folding conditions can be very fast, ranging from $3 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ to $4 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$ for CI2 and U1A, respectively.^{168; 172; 173} We can estimate the bimolecular association rate constant, k_{inter} , for His26-heme bond formation at pH 7.1 using $K_{\text{inter}}(\text{obs})$ and the slowest rate constant for His-heme bond breakage. For the AcH26I52 variant, we find $k_{\text{inter}} = (2.0 \pm 0.2) \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$. For the AcA25H26I52 variant, we find $k_{\text{inter}} = (3.0 \pm 0.3) \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$ [using $K_{\text{inter}}(\text{obs}) = (1.5 \pm 0.1) \times 10^6 \text{ M}^{-1}$]. Thus, as for the U1A protein,^{168; 173} intermolecular association in the denatured state of these variants of iso-1-cytochrome *c* is very fast.

4.4.5 Equilibrium *versus* irreversible kinetic control of aggregation

The impact of pathological protein aggregation in living cells depends on the competition between productive intramolecular folding and deleterious intermolecular aggregation. As seen here and in previous studies, the bimolecular rate for reversible intermolecular association can be very fast. Thus, competition between protein folding and protein aggregation has the potential to be problematic in living cells particularly if both processes are irreversible and thus kinetically-controlled. Therefore, it is instructive

to compare the advantages and disadvantages of kinetic *versus* equilibrium control of this competition.

For the iso-1-cytochrome *c* variants studied here, both intermolecular His26-heme bond formation and intramolecular loop formation are reversible, and thus under equilibrium control. We examine this competition with the data for the AcH26I52 variant. Using the intramolecular loop formation-intermolecular dimerization equilibrium model described above and the parameters $K_{\text{intra}}(\text{obs})$ and $K_{\text{inter}}(\text{obs})$ obtained from the fits to the data in Figure 4.15A, we show the relative proportions of intramolecular loop formation *versus* intermolecular aggregation as a function of protein concentration (Figure 4.16, solid lines). For comparison the predictions of an irreversible kinetic model (Eq 4.3) for the competition between folding and aggregation¹⁷⁷ are shown (Figure 4.16, dashed line).

$$f_{\text{loop}} = \frac{k_{\text{intra}}(\text{obs})}{[\text{Cyt}c]_t \times k_{\text{inter}}(\text{obs})} \times \ln \left(1 + \frac{[\text{Cyt}c]_t \times k_{\text{inter}}(\text{obs})}{k_{\text{intra}}(\text{obs})} \right) \quad (4.3)$$

Since $k_{\text{intra}}(\text{obs})$ at pH 7.1 is large for loop formation ($\sim 1400 \text{ s}^{-1}$), in fact, irreversible kinetic control would have led to a considerably higher yield of the intramolecular product for the AcH26I52 variant over the concentration range of our experiments (Figure 4.16).

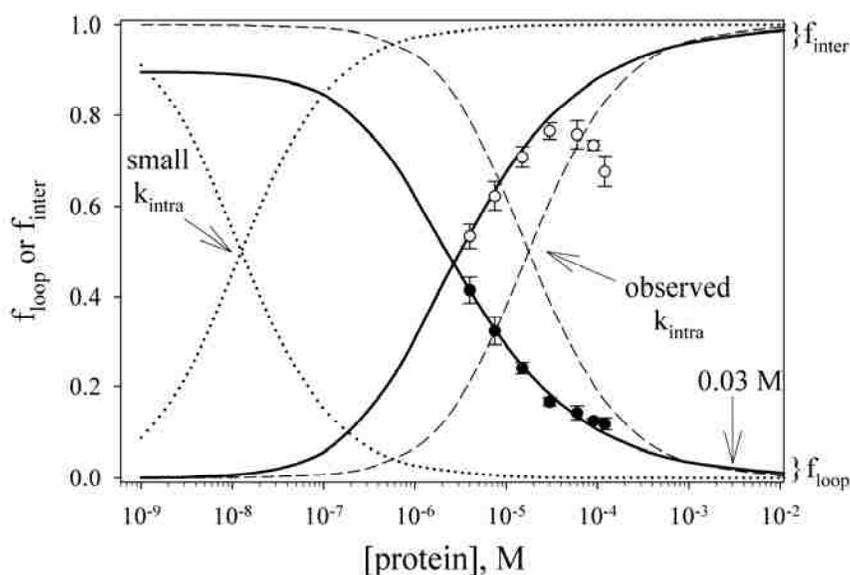


Figure 4.16: Comparison of equilibrium *versus* kinetic control for the production of intramolecular loops, f_{loop} , *versus* intermolecular His-heme dimers, f_{inter} , for the ACh26I52 variant. Curves for f_{loop} decrease as protein concentration increases, while curves for f_{inter} increase as protein concentration increases. The solid curves show f_{loop} (Eq 4.2) and f_{inter} for the equilibrium model of intramolecular *versus* intermolecular competition using the parameters from the fit of f_{fast} for His-heme bond breakage to this model given in Table 4.11. Experimental data for intramolecular loop breakage (\bullet , f_{fast} in Fig. 4.15A) and breakdown of intermolecular His-heme dimers (\circ , f_{slow} in Fig. 4.15A) for the ACh26I52 variant are shown against these curves. The deviation of these data from the solid curves at higher concentration is due to higher order aggregation. The calculated concentration dependencies for f_{loop} (Eq 4.3) and f_{inter} ($1-f_{\text{loop}}$) assuming kinetic control of the partitioning between intramolecular and intermolecular reactions are shown with dashed curves. The pH 7.1 values, $k_{\text{intra}}(\text{obs}) = 1400 \text{ s}^{-1}$ and $k_{\text{inter}}(\text{obs}) = 2.0 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ for the ACh26I52 variant, were used to generate these curves. The effect of reducing $k_{\text{intra}}(\text{obs})$ to 1 s^{-1} on the concentration dependencies of f_{loop} and f_{inter} using the kinetic control model is shown with dotted curves.

However, for many protein folding reactions the rate of folding is considerably slower than the millisecond time scale.¹⁸⁶ For comparison, the outcome for irreversible kinetic control is shown in Figure 4.16 (dotted line) if loop formation occurs on a 1 s time scale ($k_{\text{intra}} = 1 \text{ s}^{-1}$) in competition with the very fast bimolecular His26-heme reaction observed here. Clearly, kinetic control is highly disadvantageous if the intramolecular reaction is slow. Another aspect of equilibrium control is that the square root dependence of f_{loop} on $[\text{Cyt}c]_t$ for equilibrium control *versus* the logarithmic dependence of f_{loop} on

[Cyt c]_t for the kinetic model leads to a less abrupt fall off in f_{loop} at higher protein concentrations (see Figure 4.16). This would be advantageous for the yield of folded *versus* aggregated protein, as well.

In vivo, proteins must fold and operate in crowded environments, where concentrations of macromolecules approach 350 mg/mL.¹⁸⁷ For a small protein like cytochrome *c* this corresponds to a concentration of ~0.03 M. Under these conditions, fast folding kinetics (~1000 s⁻¹) will not permit competition with the very fast bimolecular aggregation we observe in the denatured state (see arrow in Figure 4.16). However, if intermolecular interactions are kept modest and reversible [similar to $K_{inter(obs)} = 3 \times 10^7 \text{ M}^{-1}$ observed for the AcH26I52 variant (Table 4.11)], a protein of average stability (~5 kcal/mol) would form less than 0.1% aggregate even at a concentration of macromolecules of 0.1 M if the partitioning between folding and aggregation were under equilibrium control.

Thus, there are advantages for the folding *versus* aggregation competition to have evolved with equilibrium rather than kinetics controlling this partitioning *in vivo*. The observation that aggregation during folding can be reversible^{168; 169; 172-175} and that protein sequences appear to have evolved to minimize aggregation-prone sequences¹⁷¹ supports this contention. The very fast bimolecular rate constants, seen for oligomerization of denatured states,^{168; 172; 173} indicate that it would be difficult for folding to prevail over aggregation if kinetic control of this partitioning operated in the crowded environment of a cell. Clearly, chaperones have evolved to assist in making aggregation reversible when necessary and thus allow for productive folding for protein sequences that are not optimized for reversible aggregation.¹⁸⁸

4.4.6 Aggregation during the folding of cytochrome *c*

For horse heart cytochrome *c*, reversible aggregation occurs during folding at high protein concentration.¹⁶⁹ Given the results on intermolecular His26-heme binding presented here, it is possible that the aggregation is mediated through His26, although the presence of lysine at position 25 in the horse protein appears to make this process much less favorable than for Pro25 or Ala25, since aggregation is observed only above 30 μM protein concentration during folding. At the highest concentrations used in the horse cytochrome *c* study (500 μM), irreversible kinetic control would have decreased the yield of folded protein to $\sim 50\%$. With weak reversible aggregation, all of the protein reached the native state.

4.5 Conclusion

We have shown that reversible His26-heme aggregation competes with intramolecular His26-heme loop formation in the denatured state of iso-1-cytochrome *c*. The bimolecular step in aggregation is very fast, indicating that intermolecular contacts between denatured proteins occur rapidly even in dilute solution. The thermodynamics of intermolecular His26-heme binding depends strongly on the sequence proximity of the histidine to the heme and on the identity of residues adjacent to the histidine. For His26, Pro25 strongly enhances the intermolecular interaction relative to Ala25, primarily by slowing breakdown of what appears to be an intermolecular dimer. We have developed an equilibrium model to analyze aggregation in the denatured state of these variants. Comparison of the predictions of this model with a previous irreversible kinetic model

for aggregation shows that equilibrium control has evolutionary advantages in terms of productive partitioning between folding and aggregation.

CHAPTER 5

PROBING DENATURED STATE DISTRIBUTIONS OF THE INITIAL EVENTS OF PROTEIN FOLDING USING TIME-RESOLVED FÖRSTER RESONANCE ENERGY TRANSFER (TR-FRET)

5.1 Introduction

Proteins are naturally occurring heteropolymers. This heterogeneity, more than likely, is a naturally engineered attribute, which supposedly imparts unprecedented control over the polypeptide sequence in folding to its fully functional native state. In Chapter 3 we delved into the comparison of naturally occurring protein sequences from yeast iso-1-cytochrome *c*, as well as cytochrome *c'*, *via* scaling factor assessments of the denatured state. We compared these naturally occurring sequences to synthetically engineered homopolymeric protein sequences and saw that the non-random-coil nature of the former was replaced by random-coil behavior with excluded volume for the latter. There are numerous accounts of non-random-coil nature for natural polypeptide sequences^{74; 76; 77; 80; 82; 117; 144; 160; 189} but there are also several accounts of random-coil properties¹⁹⁰⁻¹⁹². Thus, the obvious question is, “What is the function of heterogeneity for a natural protein sequence”? We suspect this heterogeneity is a necessary property for proteins in establishing hydrophobic contacts during the theorized hydrophobic collapse stage of protein folding. This brings us to the second question. How can we probe such interactions at the timescale of their occurrence? To do this, we focus on the denatured state and the events leading to the formation of one of the most primitive contacts possible at this level – a simple loop – using Förster resonance energy transfer (FRET).

The denatured state is not just a single conformation of the polypeptide chain. The denatured state ensemble (DSE) is composed of rapidly exchanging conformations.^{193; 194} Therefore, techniques such as NMR, CD, steady-state fluorescence and absorbance, which just measure the ensemble-averaged spectroscopic property of the DSE will not be able to provide information on the underlying complexity of the DSE.^{74-76; 189; 195} A technique sensitive to these complexities is needed. We have chosen to use a prominent method, time-resolved Förster resonance energy transfer (TR-FRET), currently being used in the field. FRET is a property of a fluorescent molecule (fluorophore) to undergo resonance energy transfer with an acceptor molecule if there is sufficient overlap between the donor emission spectrum and the acceptor absorbance spectrum. The donor's excited-state decay rate constant (k), for transfer of energy to the acceptor, scales as the inverse to the sixth power of the donor-acceptor distance (r) according to Eq 5.1:^{75; 196}

$$k = k_o \left(1 + \left(\frac{R_o}{r} \right)^6 \right) \quad (5.1)$$

where k_o is the donor's excited state decay rate in the absence of a quencher and R_o is the Förster distance. The Förster distance (critical distance) is dependent on the spectroscopic properties of the donor and acceptor and is given by Eq 5.2:^{75; 196}

$$R_o = 8.8 \times 10^{-5} \kappa^2 n^{-4} \Phi_D J \quad (5.2)$$

where κ is the orientation factor ($2/3$ for random orientation), n is the refractive index of the solution, Φ_D is the donor fluorescence quantum yield, and J is the overlap integral, which describes the spectral overlap between the donor fluorescence spectrum and the acceptor molar absorption spectrum.^{75; 196} This resonance energy transfer is distance dependent as seen from (Eq 5.1) and the possible distances probed can be ascertained from the R_0 – usually $0.3R_0 < r < 1.5R_0$.⁷⁵ Thus, unlike basic fluorescence, which only reports on the presence or absence of a specifically probed attribute, FRET can give distance information in relation to the donor and acceptor. FRET should therefore easily report on the complexity of conformational states that exist in the DSE. Unfortunately, since the DSE has been found to be a collection of rapidly exchanging conformations, a steady-state FRET analysis would simply provide information on the average of the ensemble, which is not particularly useful for discerning various conformations in the ensemble.^{74-76; 195} However, lifetime measurements of the ensemble are dependent on the excited-state decay of the donor fluorophore. Time-resolved detection of various DSE conformational states is possible, if the rate of interconversion between substates within the DSE is greater than k_0 . Furthermore, since the donor excited-state decay rate is controlled through distance-dependent quenching by the acceptor, we can obtain distance distributions of the various conformational states in the DSE through an analysis of the FRET efficiency. Additionally, anisotropic measurements can be made to give information on the relative conformations of the donor fluorophore and how it is affected by the tumbling motions of the protein. With this level of control and information return from TR-FRET, we can probe the nature of the conformational distributions present in the DSE of iso-1-cytochrome *c* and how they change during the initial stages of a

supposed nucleation event – loop formation. An understanding of the conformational changes to the DSE of iso-1-cytochrome *c* as a loop is formed will provide a greater understanding of how nature has engineered the polypeptide chain of this protein, and possibly proteins in general, to proceed on its “conformational walk” across the energy landscape down towards the bottom of the folding funnel.

Previous TR-FRET experimentation on iso-1-cytochrome *c* showed that the sequence is prone to misligation under denaturing conditions.⁷⁴ The native axial Met 80 heme ligand is not favored under oxidizing and low pH conditions. However at pH 7, various other ligands can bind to the heme in the denatured state.^{80; 117; 119} Therefore, care must be taken in mutant design to avoid misligation complications. Previous studies have also highlighted a region near the hydrophobic heme group that displays surprising stable distributions of conformers in the DSE at high gdnHCl concentrations.^{74; 76} This is of great interest to us since previous studies on loop formation from the C-terminal side of iso-1-cytochrome *c* has also provided a maximally stabilized loop (variant AcH54I52) under various gdnHCl concentrations – even at 6 M.⁸² Probing the variants used in our previous study⁸² would enable a greater understanding of the factors that affect loop formation. Hence, we focus on that protein sequence region using fluorescent donor probes at various positions to enable a greater elucidation of the complexity of sequence-heme interactions and their role in loop formation and stabilization. Thus, in this chapter we revisit loop formation of various iso-1-cytochrome *c* loop forming variants ranging from a 37-residue to an 83-residue loop, each having a fluorescence donor probe attached approximately in the center of the loop. Quenching from the heme acceptor leading up to loop closure will provide distance distributions of the conformational states

present in each of these loops formed under denaturing conditions. We will thereby gain direct insight into the mysterious DSE and explain the loop stability trends we previously saw under equilibrium conditions.⁸²

5.2 Materials and Methods

5.2.1 Preparation of cysteine variants for a eukaryotic expression system

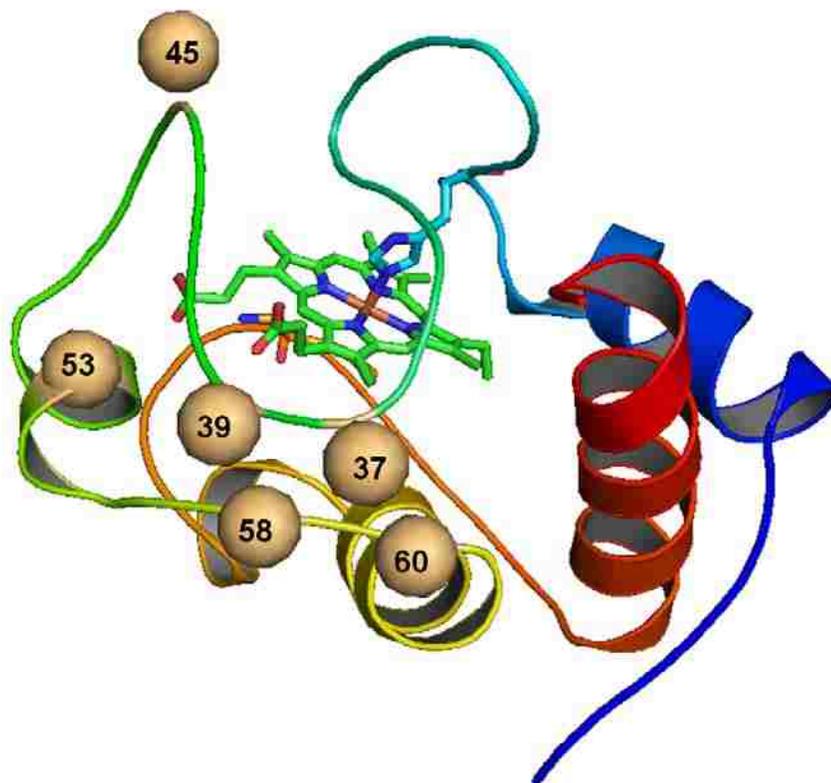


Figure 5.1: Cartoon diagram of the various cysteine mutations in relation to the heme and the rest of the protein. Pymol was used in conjunction with the 2YCC PDB file to generate cartoon.

The following variants were prepared for this study: AcH54I52C37, AcH54I52C39, AcH73I52C45, AcH89I52C53, AcH100I52C58 and AcH100I52C60.

The locations of the engineered cysteines relative to the heme and the rest of the protein are shown in Figure 5.1. All variants contained appropriate mutations to remove the possibility of misligation. They all contain the T(-5)S and K(-2)L mutations (horse cytochrome *c* numbering is used; thus, the first five amino acids of iso-1-cytochrome *c* are numbered -5 to -1) to give an acetylated N-terminus (eliminates competition between the N-terminal amino group and histidine for binding to the heme under denaturing conditions)¹¹⁹ as well as the H26N, H33N and H39Q^{80; 117} mutations which remove the wild type surface histidines so that the desired engineered histidine is the only ligand capable of denatured state loop formation. In addition, they contain the N52I mutation for stability¹⁶³ and improved protein yields as well as the C102S mutation to prevent disulfide dimerization during physical studies. The AcH54I52C37 (Gly37 is 52.9% solvent exposed) and AcH54I52C39 (His39 is 50.3% solvent exposed) variants added the G37C [5'-d(GACCAGATTGTCTACAAAAGATACCG)-3'] and Q39C [5'-d(AGCTTGACCAGAAACATCTGCCAAAGAT)-3'] mutations, respectively to the AcH54I52 variant (37-residue loop). The AcH73I52C45 (Gly45 is 88.1% solvent exposed) and AcH89I52C53 (Ile53 is 50.4% solvent exposed) variants added the G45C [5'-d(TGTGTACGAATAACATTCAGCTTGACC)-3'] and I53C [5'-d(CACGTTTTTCTTACAGATGGCATCTGT)-3'] mutations to the AcH73I52 (56-residue loop) and AcH89I52 (72-residue loop) variants, respectively. The AcH100I52C58 (Leu58 is 52.1% solvent exposed) and AcH100I52C60 (Asp60 is 50.8% solvent exposed) variants added the L58C [5'-d(ATTTTCCTCCCAACACACGTTTTTCTT)-3'] and D60C [5'-d(CATGTTATTTTCACACCACAACACGTTTTT-3')] mutations to the AcH100I52

variant (83-residue loop). Solvent exposed surface area for each residue position was calculated based on the 2YCC protein data bank file using the GetArea web-based program.¹⁹⁷⁻¹⁹⁹ Each of the aforementioned mutations is in a solvent exposed area of the native conformation under oxidizing conditions; therefore, we do not expect drastic perturbations to global protein stabilities after chemical attachment of the fluorophore donor. Although, if glycines in tight turns are replaced it can be destabilizing. All cysteine variants were prepared using single-stranded pRS/C7.8 vector DNA containing the appropriate AcHXI52 variant of iso-1-cytochrome *c* as template for the unique restriction site elimination method.¹²³ The SacI+II- selection oligonucleotide⁸¹ was used for all variants. The SacI+II- oligonucleotide eliminates a unique *SacII* site and creates a *SacI* restriction enzyme site upstream from the iso-1-cytochrome *c* gene (*CYCI*).¹⁷⁸ All mutations were confirmed by dideoxy sequencing. The pRS/C7.8 vector containing each of the individually confirmed mutations was transformed into the GM-3C-2 *Saccharomyces cerevisiae* cell line (cytochrome *c* deficient)¹⁷⁹ and the transformants characterized, as described previously.⁸⁰ Expression and purification were carried out for all variants from *S. cerevisiae* as previously described.¹²⁴

5.2.2 Preparation of cysteine variants for a bacterial expression system

Bacterial counterparts to the AcH54I52C37 and AcH54I52C39 variants from section 5.2.1 above were prepared as a fall back if the yeast expression system was not compatible with the mutations made. Bacterial variants were based on the NK5A template.⁹⁰ This template is similar in function to the AcTM variant. It has the H26N, H33N and H39Q mutations to prevent competitive binding to the heme in the denatured

state. Unlike acetylation in the AcTM template, the NK5A template minimizes N-terminal competition by pushing the N-terminus farther away from the heme *via* a KAAAAA insertion between residues F(-3) and K(-2).^{83; 90} The resulting NK5A template was used to incorporate the G37C and Q39C mutations using single-stranded pBTR1 vector DNA carrying the NK5A variant of iso-1-cytochrome *c* as the template and the following primers: 5'-d(GACCAGATTGTCTACAAAAAGATACCG)-3', and 5'-d(AGCTTGACCAGAACATTCTGCCAAAGAT)-3', respectively. The Aat2- selection primer, 5'-d(GTGCCACCTGATATCTAAGAAACC)-3' was used together with these mutagenic primers. It converts a unique *Aat2* restriction site in the pBTR1 vector carrying the NK5A variant into an *EcoRV* restriction site. Subsequently, the stabilizing N52I¹⁶³ as well as the K54H mutations were engineered into the resulting NK5AC37 or NK5AC39 templates to make them comparable to the yeast system. This set of mutations was accomplished using either sequential N52I then K54H primers (using NK5AC39 template) or with the single N52IK54H [5'-d(GCCTTACCTGCACCAGCACC GTGGA)-3'] primer (using NK5AC37 template). The *EcoRV*- selection primer, 5'-d(GTGCCACCTGACGTCTAAGAAACC)-3' or the previously listed Aat2- selection primer, as appropriate, was used together with single-stranded pBTR1 vector DNA containing the NK5AC37 or the NK5AC39 variants and the respective mutagenic primer. The *EcoRV*- selection primer converts a unique *EcoRV* restriction site in the pBTR1 vector carrying the NK5AC37 or NK5AC39 variant into an *Aat2* restriction site. The resultant variants were NK5AC37I52H54 (*EcoRV*-) and NK5AC39I52H54 (*Aat2*-). All variants were expressed and purified as described in Section 2.2.4.

5.2.3 Chemical attachment of 1,5 I-AEDANS fluorophore to cysteine variant(s)

Crude protein from CM Sepahrose ion exchange purification was further purified by HPLC on a BioRad UNO-S6 ion exchange column. The desired variant was eluted with a salt gradient (HPLC B: 50 mM sodium phosphate, 1 M NaCl, pH 7) after a 10 minute pre-equilibration with low salt buffer (HPLC A: 50 mM sodium phosphate, pH 7). Purified protein was then desalted in low salt buffer and exchanged into a 100 mM sodium phosphate, pH 7.2 buffer using Amicon centriprep ultrafiltration devices (MWCO 10 kDa). The protein's buffer volume was then reduced in this manner to achieve 100 – 200 μ M concentrations before pre-treatment with 5X molar excess of a 2 mM TCEP solution (prepared with 100 mM sodium phosphate, pH 7.2 buffer). A 40 mM stock of 5-(2-((iodoacetyl)amino)ethyl)aminonaphthalene-1-sulfonic acid (1,5 I-AEDANS) fluorophore labeling agent (prepared in 100 mM sodium phosphate, pH 7.2 buffer) was prepared and both solutions were deoxygenated separately with a vacuum line and placed under an argon atmosphere at room temperature for an hour. A 400X molar excess of the labeling agent was then added dropwise to the reduced protein solution, while gently stirring under an argon atmosphere. The 1,5 I-AEDANS undergoes a covalent attachment^{200; 201} between the free sulfhydryl group of cysteine and its iodoacetamidyl group *via* an S_N2 substitution reaction as shown in Figure 5.2 below.



Figure 5.2: Schematic representation of Cysteine-AEDANS conjugation on free sulfhydryl groups of a protein.

The 1,5 I-AEDANS reagent is very thiol selective and generally does not react with histidines or methionines at physiological pH.²⁰² The charged sulfonate group on the 1,5 I-AEDANS usually restricts the fluorophore to cysteines on the surface of the protein *versus* those in a more hydrophobic environment.^{201; 202} The reaction was then allowed to proceed for 1 to 2 hours in the dark at room temperature. The unreacted labeling agent was then quenched by consumption with 10X molar excess of β -mercaptoethanol (β -ME). The labeled protein reaction mixture was then immediately and quickly subjected to ultracentrifugation (250 ml Amicon) in HPLC A (50 mM sodium phosphate, 1 mM EDTA, pH 7) under dark conditions to remove the deleterious iodide ion and its photoproducts. After this essential step, the labeled protein was then either flash frozen and stored, or purified by HPLC for experimentation. HPLC purification, using a Waters Protein-Pak Cation Exchange column, enabled separation with good baseline resolution of the labeled from the unlabeled protein under reducing condition (sodium dithionite addition). EDTA chelating agent had to be added to all HPLC buffers and subsequent

experimentation buffers from this point forward so as to prevent cleavage of the fluorophore from the protein variant. After HPLC purification, the protein was rapidly desalted *via* short ultracentrifugation spins and the AEDANS conjugation and protein integrity were confirmed using MALDI-TOF analysis.

5.2.4 Characterization of AEDANS-labeled protein

The presence of the chemically attached AEDANS fluorophore was first confirmed by MALDI-TOF analysis. The AEDANS fluorophore should add an additional 307 mass units to the protein. Global stability (gdnHCl denaturation monitored by circular dichroism) and equilibrium loop formation (pH titrations in 3 M gdnHCl) measurements were performed, as previously described in section 2.2.6, on the AcH54I52C37 variant in order to determine if the bulky fluorophore perturbed the protein significantly.

5.2.5 Fluorescence lifetime measurements with intrinsic Trp59 fluorophore

Initially, time-resolved fluorescence lifetime measurements were evaluated using the intrinsic fluorophore Trp59, found in the naturally occurring sequence of iso-1-cytochrome *c*. The variant used in these preliminary studies is an 83-residue loop forming variant – AcH100I52. Under loop closure conditions, the Trp59 would be positioned exactly in the center of the loop similar to the AcH100I52C58 variant. Loop closure and opening was controlled by the presence or absence of 50 mM imidazole in the protein mixture. These experiments were done at 20 °C under both native and denaturing conditions using a monochromator to select for emission at 350 nm following

excitation of the Trp59 with the third harmonic of a Mira 900 Ti:Sapphire laser (Coherent, Santa Clara, CA).²⁰³ The lack of satisfactory results prompted an extrinsic fluorophore approach. See Results section for details.

5.2.6 Fluorescence lifetime and anisotropy measurements with the 1,5 I-AEDANS fluorophore

All AEDANS labeled protein variants were oxidized with potassium ferricyanide to ensure 100% oxidized protein before experimentation. The oxidized protein was separated from the oxidizing agent by size exclusion chromatography (G-25 Sephadex beads), pre-equilibrated with 3xPi (15 mM Na₂HPO₄, 45 mM NaCl), 1 mM EDTA, pH 7 buffer. All protein variants were exchanged in this buffer for time-resolved fluorescence lifetime and anisotropy measurements.

Lifetime data were collected by time-correlated single-photon counting (TCSPC) using the FLASC 1000 sample chamber (Quantum Northwest, Liberty Lake, WA) thermostatted to 20 °C.²⁰³ For the AEDANS fluorophore donor, pulsed excitation was used at 375 nm employing a repetition rate of 4.7 MHz from a frequency-doubled ps Mira 900 Ti:Sapphire laser (Coherent, Santa Clara, CA).²⁰³ The emission components were separated using bandpass filters 500/40 nm (Chroma, Rockingham VT) and the decay curves were collected using the TimeHarp 200 PCI board (PicoQuant, Berlin) until 4×10^4 counts were obtained at the maximum of the curve.²⁰³

Anisotropy measurements were also carried out using the FLASC 1000 sample chamber thermostatted to 20 °C. This sample chamber has the ability to simultaneously collect vertical (V), horizontal (H) and variable polarization components of the emission

due to its unique T format.²⁰³ A beam-splitting Glan-Thompson polarizer (Karl Lambrecht, Chicago, IL) was used to separate the H and the V components.²⁰³ Typically, the H, V and magic angle components can be detected simultaneously with this arrangement by using separate photon multipliers.²⁰³ This feature thereby guarantees that all data collection occurs under identical excitation conditions. The H and V components were collected using matched bandpass filters and for equal lengths of time until the decay curve of the V component reached a maximum of 4×10^4 counts.²⁰³ The time-resolution of the instrumental setup was 35 ps/channel.²⁰³ Anisotropy measurements were only carried out on the end-states of the loop forming variants (i.e. open and closed loop forms).

All data were fitted using the Fluofit windows-based software program from PicoQuant GmbH (Berlin, Germany). All time-resolved FRET plots were exported as image files using this program.

5.3 Results

5.3.1 Results of mutagenesis and protein expression

Both cysteine labeled bacterial variants (NK5AC37I52H54 and NK5AC39I52H54) were successfully engineered and expressed. However, no experimentation has been done with them to date. Available protein aliquots have been flash frozen and stored at -80 °C.

All cysteine engineered yeast variants have been successfully made and their sequences confirmed. However, due to complications encountered with the *CYCI* gene

functionality testing,⁸⁰ only the AcH54I52C37 variant was successfully characterized in the *S. cerevisiae* GM-3C-2 cell line. Therefore, the focus of this chapter is based on the experiments done with this variant.

5.3.2 Characterization of AcH54I52C37_AEDANS labeled variant

The presence of the AEDANS attached to the protein was confirmed by MALDI-TOF analysis of the protein before and after labeling. Global stability measurements using circular dichroism (see Figure 5.3), provided thermodynamic stability parameters for the AcH54I52C37_AEDANS variant ($\Delta G_u^{\circ}(\text{H}_2\text{O}) = 1.75 \pm 0.04$ kcal/mol, m -value = 2.93 ± 0.09 , midpoint = 0.60 ± 0.03). Comparing these results to those of the AcH54I52 variant ($\Delta G_u^{\circ}(\text{H}_2\text{O}) = 3.93 \pm 0.02$ kcal/mol, m -value = 3.96 ± 0.02 , midpoint = 0.99 ± 0.01),¹²⁴ shows that the addition of the fluorophore significantly decreases the overall stability of the protein. However, equilibrium loop formation experiments (see Figure 5.4) show that the strength of the His-heme bond is the same for both the AcH54I52 variant ($\text{pK}_a(\text{obs}) = 4.80 \pm 0.02$, n -value = 0.83 ± 0.03)⁸² and the AcH54I52C37_AEDANS variant ($\text{pK}_a(\text{obs}) = 4.81 \pm 0.03$; n -value = 0.73 ± 0.05). Thus, this indicates that the loop thermodynamics is not affected by the bulky fluorophore. FRET measurements should therefore accurately report on the nature of the loop in the denatured state.

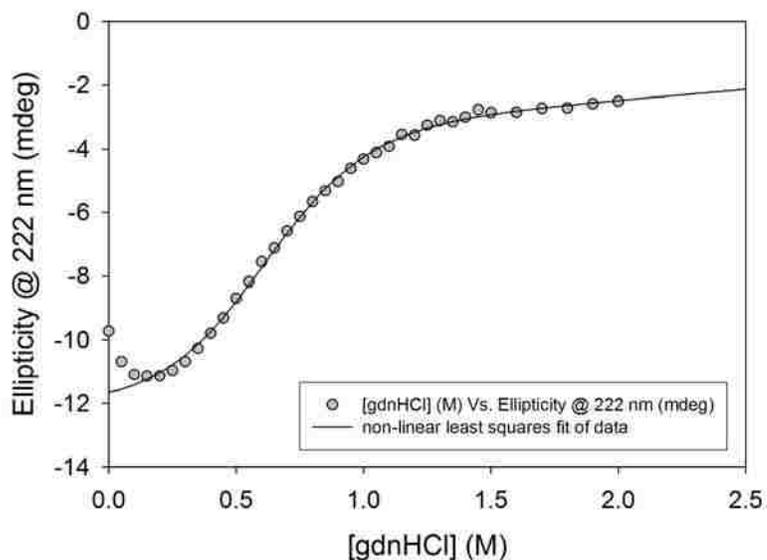


Figure 5.3: A typical gdnHCl protein denaturation titration monitored at 222 nm and at 25 °C using an Applied Photophysics Chirascan circular dichroism spectrometer for the Ach54I52C37_AEDANS variant. Fitted line is a non-linear least squares fit to a linear free energy equation.

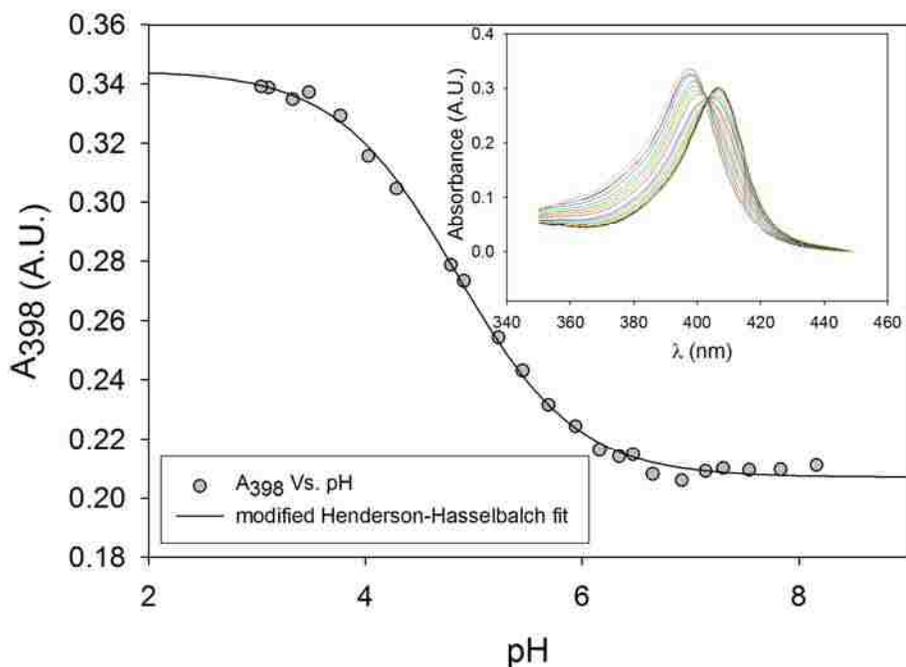


Figure 5.4: A typical equilibrium loop formation titration representative of the Ach54I52C37_AEDANS variant in 3 M gdnHCl at 22 ± 1 °C. Fit is based on a modified form of the Henderson-Hasselbalch equation. See Figure 2.1. Inset is a plot of absorbance from 350 to 450 nm going from pH ~8 to ~3 in increments of ~0.2 pH units. The single isosbestic point indicates two-state behavior.

5.3.3 Fluorescence lifetime measurements of AcH100I52 variant

Initially, fluorescence energy transfer kinetic measurements were carried out on the AcH100I52 variant using the intrinsic Trp59 fluorescent donor and the heme as the quencher. Figure 5.5 is representative of the data obtained for both opened and closed loop forms under both native and denatured conditions. The entire results are summarized in Table 5.1. Those experiments produced similar lifetimes under both native and denaturing conditions (3 M gdnHCl), both in the absence and presence of imidazole. In the denatured state imidazole should break the His-heme loop. Further analysis of individual components using fractional amplitudes crudely implied minor changes in Trp59 populations, quenched ($\tau \sim 0.95$ ns) and unquenched ($\tau \sim 2.3$ ns) of closed *versus* opened loop forms in the denatured state. The small differences seen were within the error of the instrument; thus, it is inappropriate to draw conclusions from these data. Seemingly, the donor probe is equally quenched, relative to τ_D (2.4 ns)⁷⁵, under all conditions, thus these results provide no information. The R_o of the Trp-Heme pair (34 Å)⁷⁵ is therefore inappropriate for this distance range; the 1,5, I-AEDANS may be a better fluorophore.

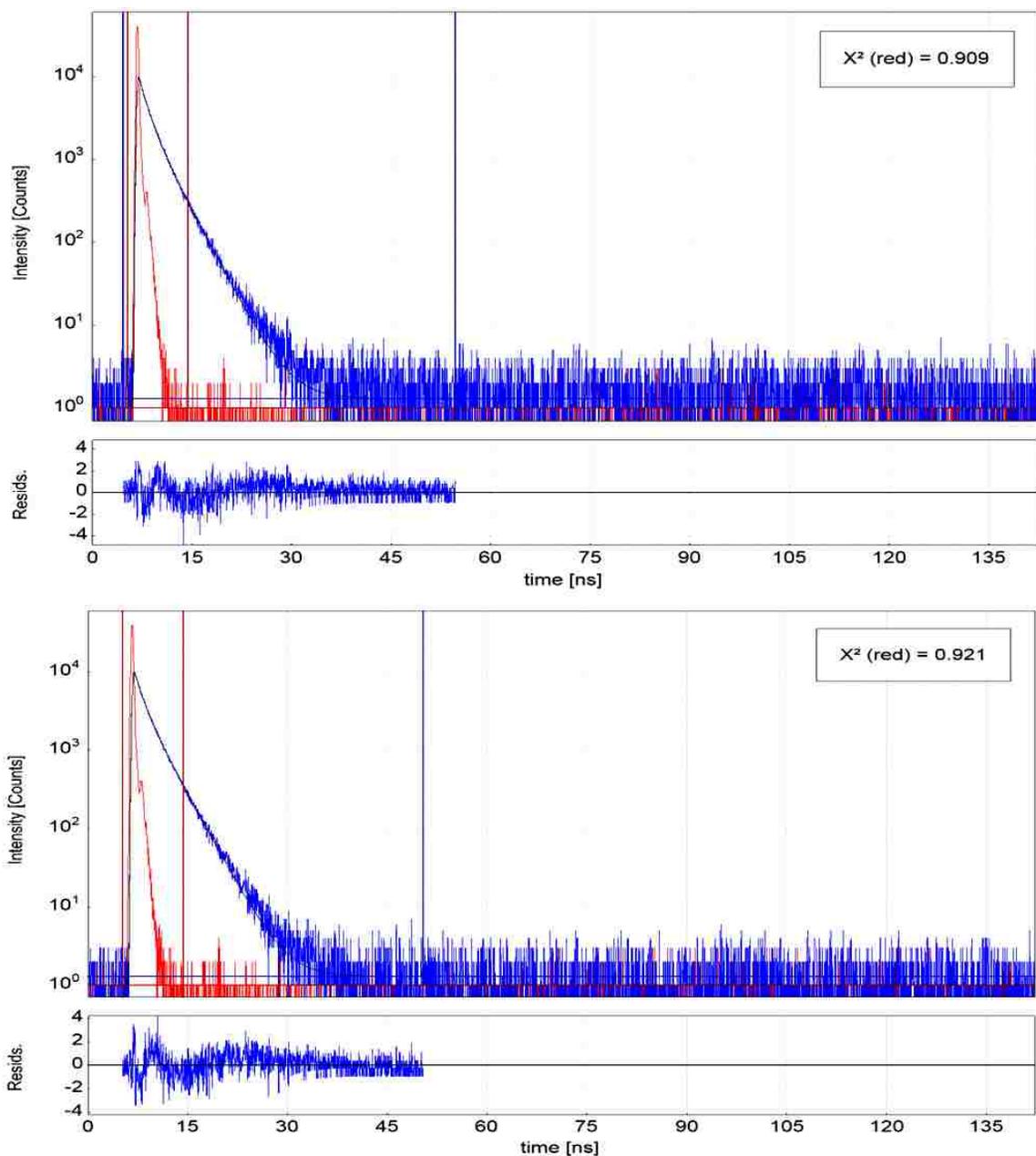


Figure 5.5: Time-resolved fluorescence measurements of AcH100I52 under native conditions in the presence and absence of imidazole at 20 °C. The native Trp59 fluorophore was centrally positioned in the closed loop form. (A) In the absence of the imidazole heme-competing ligand and (B) in the presence of 50 mM imidazole heme-competing ligand. Data were fit to two exponentials. Data are shown in blue; Instrument response function is shown in red; Black solid line is a double exponential fit to the data. Residuals below show how well the data fit to a two exponential equation.

Table 5.1: Fluorescence lifetime measurements of AcH100I52 variant under both native and denaturing conditions in the presence and absence of imidazole

	(Native) ¹ Imidazole Absent ^a	(Native) ¹ Imidazole Present ^b	(Denatured) ² Closed Loop ^a	(Denatured) ² Opened Loop ^b
Intensity Averaged Lifetime (ns)	2.13	2.15	2.27	2.29
Amplitude (counts)	4480 ± 50	5710 ± 50	5080 ± 50	5760 ± 50
Component Lifetime (ns)	2.83 ± 0.02	2.76 ± 0.02	2.90 ± 0.02	2.81 ± 0.02
Amplitude (counts)	7920 ± 140	8600 ± 140	7260 ± 140	6720 ± 140
Component Lifetime (ns)	0.99 ± 0.02	0.90 ± 0.02	0.99 ± 0.02	0.95 ± 0.02
Fractional Amplitude (Fast Component)	0.361 ± 0.004	0.399 ± 0.004	0.412 ± 0.004	0.462 ± 0.004
Fractional Amplitude (Very Fast Component)	0.64 ± 0.01	0.60 ± 0.01	0.59 ± 0.01	0.54 ± 0.01

¹Native Conditions: Protein in 3xPi buffer at physiological pH at 20 °C.
²Denatured Conditions: Protein in 3xPi, 3M gdnHCl buffer at 20 °C.
^aNo imidazole added to solution.
^bImidazole present at 50 mM final concentrations.

5.3.4 Fluorescence lifetime and anisotropy measurements of AcH54I2C37 variant

Time-resolved fluorescence measurements were subsequently done using the AcH54I52C37_AEDANS variant. A control experiment was carried out to assess the maximal donor lifetime and to determine if the donor has a multi-component lifetime. For the control experiment, the 1,5 I-AEDANS fluorophore was reacted with β -mercaptoethanol (β -ME) to form the β -ME-AEDANS complex. Time-resolved

measurements of this compound under closed loop (pH 6.36) and opened loop (pH 3.2) conditions resulted in similar lifetimes of 9.88 ns in 3 M gdnHCl. There was also a very low population (~0.6%) of a very fast component (~ 0.5 ns) detected. However, this is well within the error of our measurements – a representative trace of this control is shown in Figure 5.6.

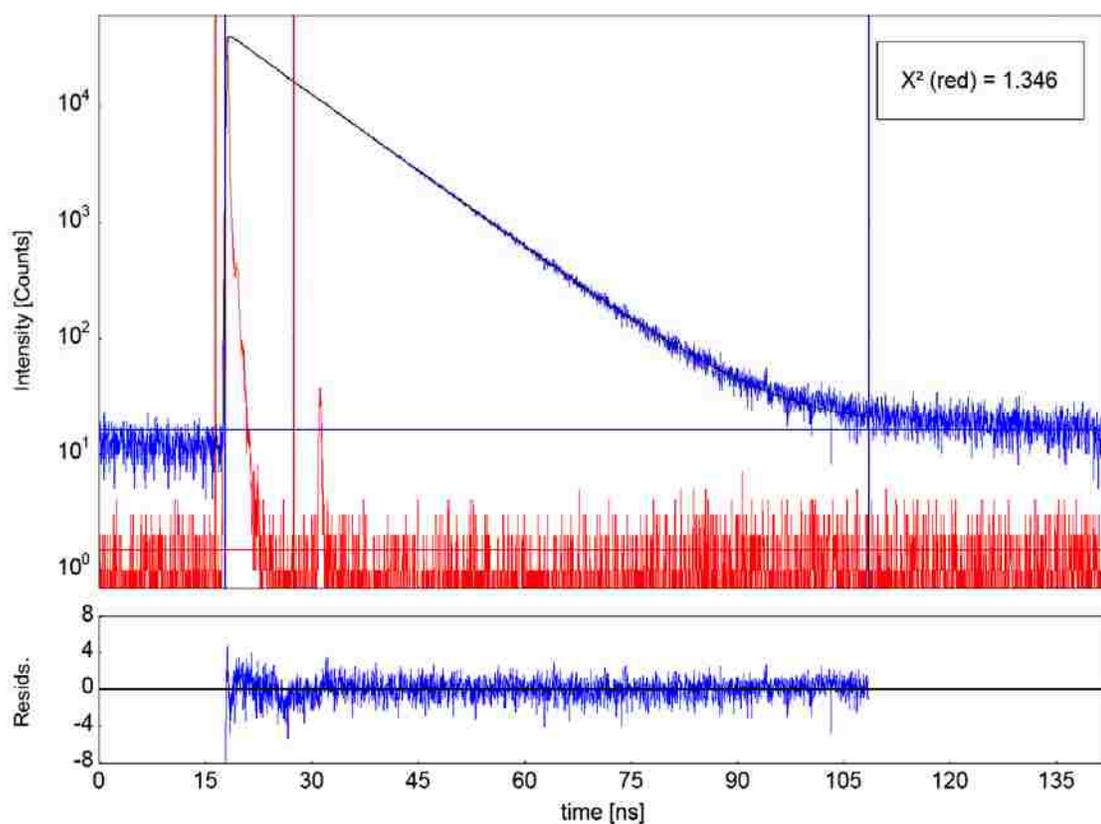


Figure 5.6: β -ME-AEDANS used as a control and measured at both pH's 6.36 (shown) and 3.20 for the closed and the opened loop conditions, respectively, in 3 M gdnHCl/3xPi at 20 °C. Probe lifetimes at both pHs were 9.88 ns. Data are shown in blue; Instrument response function is shown in red; Black solid line is a double exponential fit to the data. Residuals below show how well the data fit to a two exponential equation.

Time-resolved measurements were then taken for the AcH54I52C37_AEDANS labeled variant at various pH's based on the equilibrium loop formation plot shown in Figure 5.4. Various points along the pH curve were chosen so as to obtain a range of data having different populations of open and closed loops. The results of the data are summarized in Tables 5.2 and 5.3 and Figures 5.7 to 5.9.

Table 5.2: Individually fitted time-resolved fluorescence lifetimes and component lifetimes and amplitudes for the AcH54I52C37_AEDANS Variant

Component Lifetimes and Amplitudes							
pH	τ_{avg} (INT) (ns)	Amp1 (% cts)	τ_1 (ns)	Amp2 (% cts)	τ_2 (ns)	Amp3 (% cts)	τ_3 (ns)
3.15	5.35	64.8 ± 0.5	6.82 ± 0.02	25.6 ± 1.3	3.28 ± 0.04	9.6 ± 2.5	0.87 ± 0.02
3.76	5.24	64.7 ± 0.5	6.77 ± 0.02	24.9 ± 1.4	3.11 ± 0.04	10.4 ± 2.4	0.80 ± 0.02
4.26	5.18	63.9 ± 0.6	6.80 ± 0.00	26.1 ± 1.3	2.90 ± 0.00	10.0 ± 2.5	0.73 ± 0.02
4.79	4.90	57.4 ± 0.6	6.86 ± 0.03	29.9 ± 1.3	2.91 ± 0.03	12.7 ± 2.2	0.72 ± 0.02
5.46	4.28	42.6 ± 0.9	7.10 ± 0.04	38.7 ± 1.1	2.91 ± 0.03	18.7 ± 1.9	0.71 ± 0.01
6.59	2.90	20.9 ± 1.0	7.23 ± 0.08	45.9 ± 1.2	2.58 ± 0.02	33.2 ± 1.6	0.60 ± 0.01

Data are based on 3 exponential fits of each pH data set individually.
 τ_{avg} (INT) is Intensity Averaged Lifetimes.

Table 5.2 contains the TR-FRET results for data fitted at individual pH's. A representative trace of the data is shown in Figure 5.7 for the variant at the pH

corresponding to its $pK_a(\text{obs})$ value. Thus, this sample is representative of equal amounts of opened *versus* closed loop forms. The decay rate constants for all pH's were adequately fitted to three exponentials as seen from the residuals in Figure 5.7. Plots of the intensity weighted lifetimes, component lifetimes and their fractional amplitudes *versus* pH are given in Figure 5.8 (A-C).

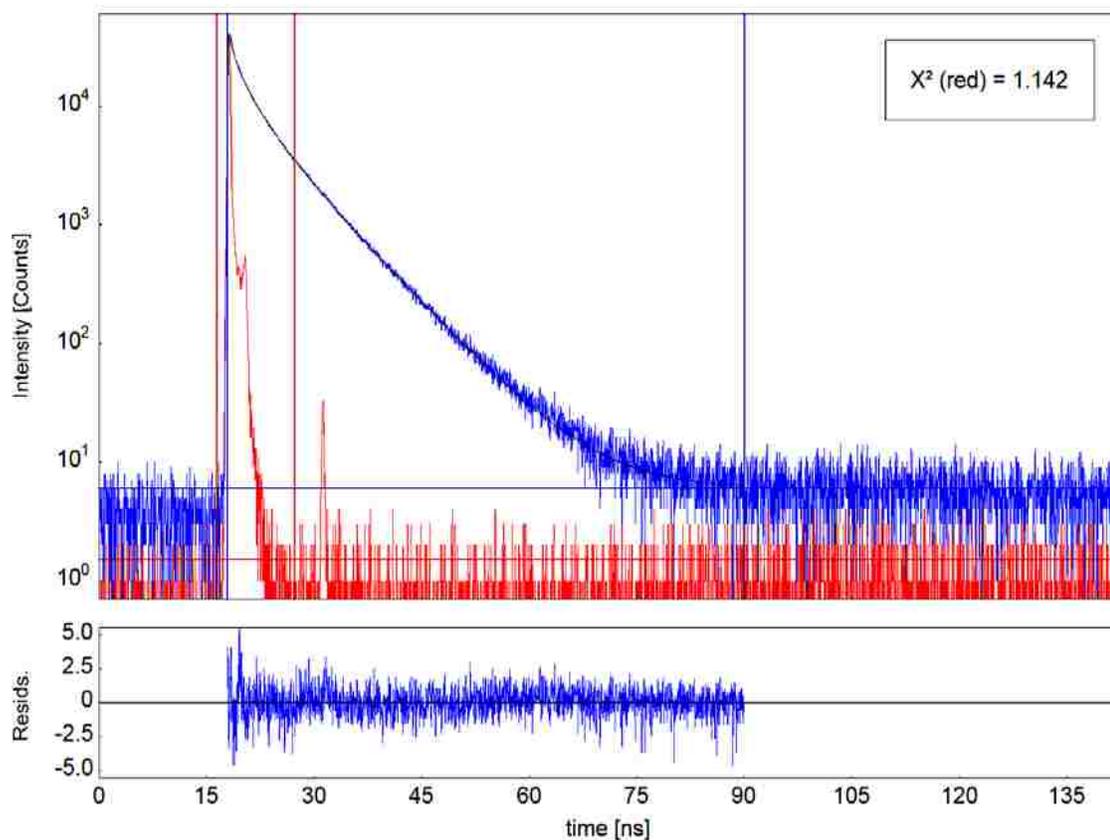


Figure 5.7: AcH54I52C37_AEDANS variant at pH 4.79 in 3 M gdnHCl at 20 °C fitted individually with three exponentials. Residuals indicate an adequate fit of the data. Data are shown in blue; Instrument response function is shown in red; Black solid line is a triple exponential fit to the data. Residuals below show how well the data fit to a three exponential equation.

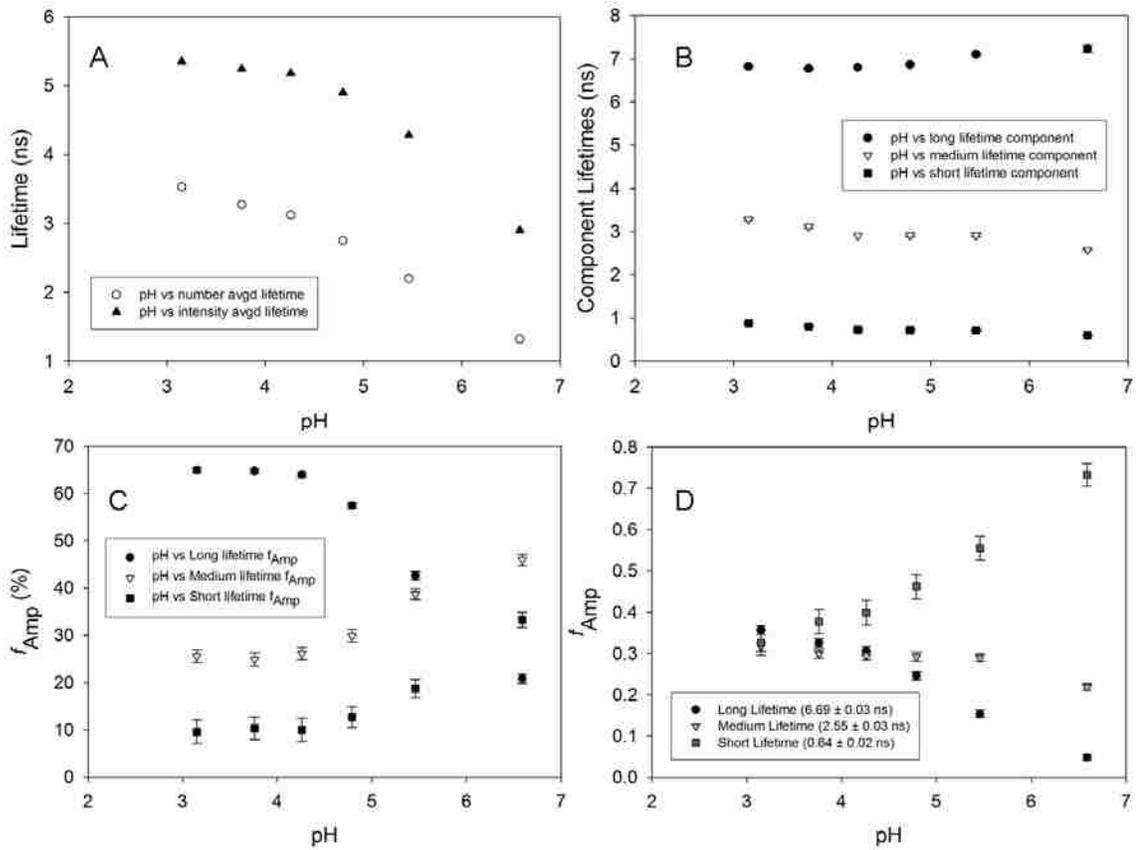


Figure 5.8: Parameters from individually fitting each data set to three exponentials are shown in (A) – (C). (A) Intensity weighted and Amplitude weighted averaged lifetime as a function of pH. (B) Individual lifetime components as a function of pH. (C) Fractional amplitudes for each component lifetime as a function of pH. (D) Fractional amplitudes from global fitting of the data showing long, medium and short lifetime components as a function of pH. Some errors are smaller than the symbol used and may not be visible.

The intensity weighted lifetimes as seen in Figure 5.8 (B) were relatively constant across the pH range. Therefore, the data were globally fitted across all pH data sets. Figure 5.9 show the outcome of the global fit at pH 4.79. As seen from the residuals in Figure 5.9, this three exponential global fitting routine adequately described the data (see Figure 5.8 (D)).

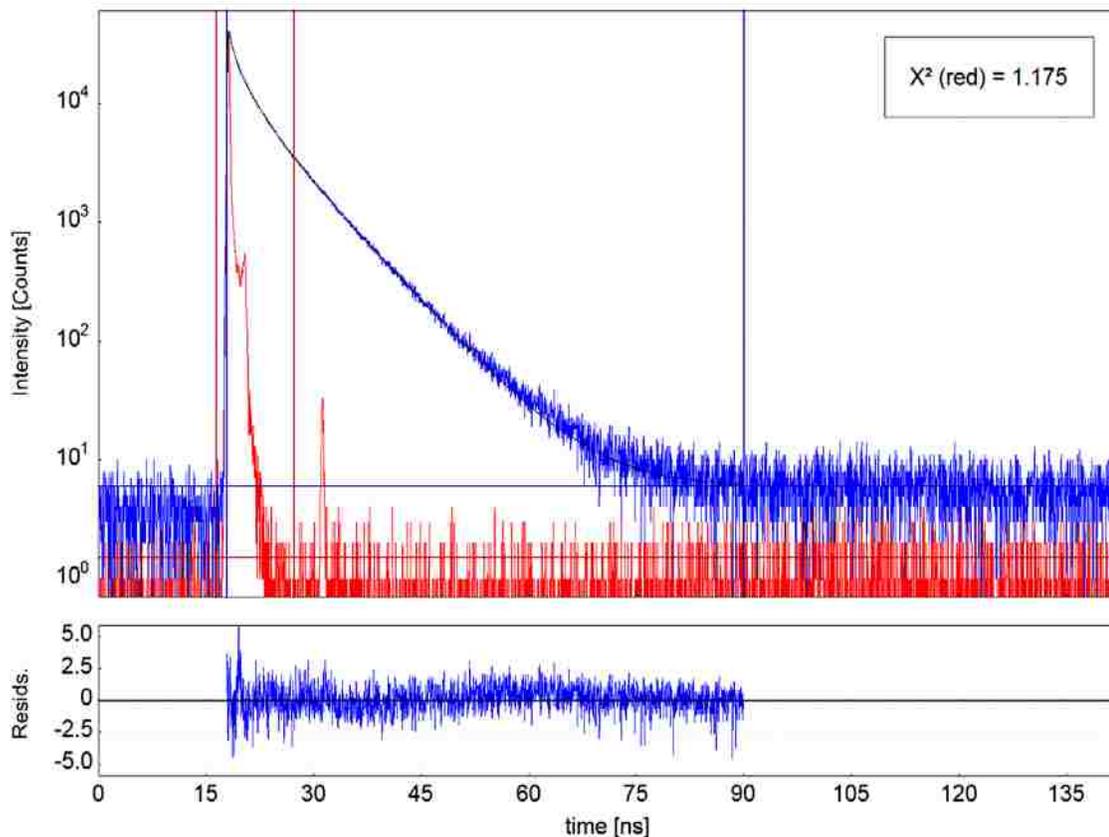


Figure 5.9: AcH54I52C37_AEDANS variant at pH 4.79 in 3 M gdnHCl at 20 °C fitted globally with three exponentials. Data are shown in blue; Instrument response function is shown in red; Black solid line is a triple exponential fit to the data. Residuals below show how well the data fit to a three exponential equation.

Table 5.3 summarizes the results of this global fitting routine. The changes in fractional amplitudes for each component lifetime as a function of pH are depicted in Figure 5.8 (D). It is evident from the divergence in Figure 5.8 (D) that there is a shift in the populations of species or “states” of a loop in the denatured state as expected with varying pH. This result is clear evidence of the ensemble nature of the denatured state and further analysis should provide insights into the dynamic behavior of this ensemble. The constant presence of the medium timescale decay lifetime is of intriguing interest.

Table 5.3: Globally fitted time-resolved fluorescence lifetimes and fractional amplitudes for the AcH54I52C37 AEDANS Variant

pH	Global Lifetimes and Fractional Amplitudes					
	f_{Amp1}	τ_1 (ns)	f_{Amp2}	τ_2 (ns)	f_{Amp3}	τ_3 (ns)
3.15	0.357 ± 0.004	6.69 ± 0.03	0.32 ± 0.01	2.55 ± 0.03	0.33 ± 0.03	0.64 ± 0.02
3.76	0.325 ± 0.004	6.69 ± 0.03	0.30 ± 0.01	2.55 ± 0.03	0.38 ± 0.03	0.64 ± 0.02
4.26	0.306 ± 0.004	6.69 ± 0.03	0.30 ± 0.01	2.55 ± 0.03	0.40 ± 0.03	0.64 ± 0.02
4.79	0.246 ± 0.004	6.69 ± 0.03	0.29 ± 0.01	2.55 ± 0.03	0.46 ± 0.03	0.64 ± 0.02
5.46	0.154 ± 0.003	6.69 ± 0.03	0.29 ± 0.01	2.55 ± 0.03	0.56 ± 0.03	0.64 ± 0.02
6.59	0.048 ± 0.002	6.69 ± 0.03	0.22 ± 0.01	2.55 ± 0.03	0.73 ± 0.03	0.64 ± 0.02

Data parameters are based on triple-exponential fits of globally fitted pH data sets.
 f_{AMP} is the fractional amplitude of each globally fitted lifetime.

Anisotropy measurements were carried out for the end-states of the system being studied (i. e. the opened and closed forms of the loop); however, those data have not yet been analyzed. The main reason for anisotropy measurements was to determine whether the probe is randomly oriented under 3 M gdnHCl conditions. Anisotropy measurements have been previously carried out for I-AEDANS attached to a similar Cys-labeled cytochrome *c* model under similar denaturing conditions. The results were consistent with the orientation factor (κ^2) being 2/3.^{74; 76} It is generally accepted that an attached fluorophore on a fully unfolded protein is sufficiently random in motion to justify the use of 2/3.^{74; 78} Therefore, we confidently assume the κ^2 of our probe to be 2/3 for these measurements in calculating R_0 and distance extractions.

5.4 Discussion

5.4.1 Overview

The mechanism by which proteins fold from their polypeptide chain to their compact, functional native structures has been a long sought after goal. Great advancements have been made toward understanding this life-essential mechanism. The heterogeneity of the protein sequence has complicated much of our efforts at a straightforward resolution to this folding mechanism. There have been many different physical techniques employed to try to shed light on the underlying key aspects of protein folding. Elucidating the conformational dynamics of the polypeptide chain seems to be one of the better approaches. The complexity of the rapidly exchanging denatured state ensemble^{193; 194} requires a specialized approach. As such, experimental techniques that probe the radius of gyration (R_g), such as small angle X-ray and neutron scattering, as well as those that look at point-to-point distances (r_{D-A}) and distributions such as time-resolved FRET decay, have been the most powerful.⁷⁸ We have employed the latter of these techniques in an attempt to understand the behavior of previously studied loops of various lengths.⁸² Since FRET is sensitive to donor-acceptor distances, it reports on the distance distribution of the DSE. Thus, FRET should enable us to spectroscopically view the conformational changes a polypeptide chain goes through when it forms a primitive loop. These interactions made on this basic conformational walk are essential to the understanding of how initial contacts made in loop formation lead to higher order structures. The presence or increase in the presence of a given conformation can shape our view of this complex process. Our venture into probing this “conformational walk”

of the polypeptide chain has unearthed some intriguing properties of the denatured state and may hold vital information toward cracking the “folding code”.

5.4.2 Extracting discrete distances from observed lifetimes

Time-resolved FRET experimentation on the AcH54I52C37_AEDANS variant allowed us to monitor the behavior of three distinct species in the denatured state going from an opened to a closed loop. Using Eqs 5.1 and 5.2, the Förster critical distance (R_0) was calculated to be 35 Å for our AEDANS-Heme (D-A) pair. A plot of FRET efficiency as a function of the critical distance is shown in Figure 5.10. This value of R_0 is a bit lower than published values of 39/40 Å for the same D-A pair.^{76; 101} This may be due to the use of the β -ME-AEDANS control, which gives a 9.88 ns lifetime *versus* the routinely used N-acetylcysteine-AEDANS control which gives a 10.2 ns lifetime.⁷⁶ Nonetheless, using this R_0 value, we extracted distances corresponding to the three different species seen in pH-dependent denatured state loop formation. The globally fitted time-resolved FRET lifetimes of 6.69, 2.55 and 0.64 ns correspond to distances of 40, 29 and 22 Å, respectively.

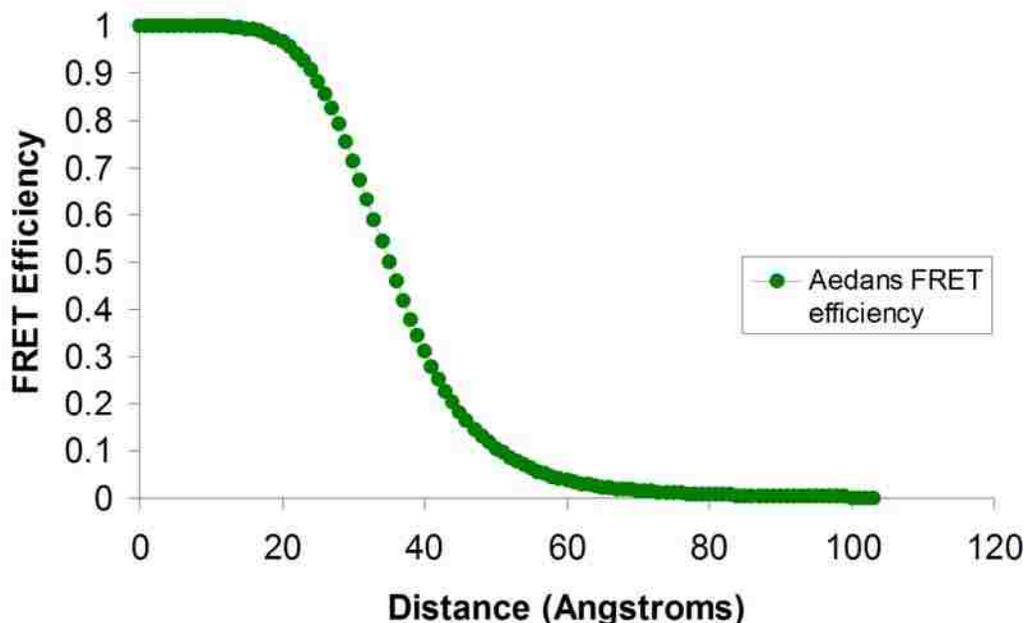


Figure 5.10: Plot of FRET efficiency *versus* distance for the AEDANS-Heme pair. R_0 at the 50% efficiency mark is 35 Å for the pair. Equations 5.1 and 5.2 were used in this calculation.

5.4.3 Comparison with donor-acceptor distances predicted for a random coil

We interpret these extracted distances by first analyzing the expected distances from theory for a random coil and a random coil with excluded volume. According to Goldenberg,²⁰⁴ random coil polymers are easily characterized by distributions of distances that are Gaussian in nature. These distance distributions are described by a mean squared end-to-end distance, $\langle r^2 \rangle$, which scales according to the number of residues (n) in the polymer and the length of the bond (l) between the residues.²⁰⁴ The general form of the equation is given in Eq 5.3²⁰⁴ below:

$$\langle r^2 \rangle = nl^2 \quad (5.3)$$

For a random polymer with no excluded volume, Goldenberg's simulations yield Eq 5.4 for the root mean square distance, RMS(r):

$$RMS(r) = (5.82 \pm 0.02)n^{(0.495 \pm 0.001)} \quad (5.4)$$

Using Eq 5.4²⁰⁴ we calculate the random coil distance for the opened loop (20 monomers; His18-Cys37) and the closed loop (18 monomers; Cys37-His54) to be ~ 26 Å and ~ 24 Å, respectively. These distances are practically the same and definitely not what we see from our AcH54I52 variant. Proteins and any other polymer are more accurately described by taking into account excluded volume effects. Goldenberg's simulations yield Eq 5.5 when excluded volume is accounted for:

$$RMS(r) = (5.68 \pm 0.08)n^{(0.583 \pm 0.003)} \quad (5.5)$$

Using Eq 5.5²⁰⁴, we calculate distances, for a random coil with excluded volume taken into account, for the opened and closed loops to be ~ 33 Å and ~ 31 Å, respectively. These distances are not very different from each other yet again.

To provide an estimate of the maximal donor-acceptor distance possible, we use an extended β -strand (3.5 Å/residue) as a model. This model produces distances of 70 Å and 63 Å for open (20 residues) and closed (18 residues) loops, respectively. The Gaussian distance distribution used by Goldenberg to model the denatured state is inconsistent with the nearly equal contributions we observe from three subpopulations of the DSE in the open loop form at low pH. The 22 Å distance is shorter than, the 29 Å

distance is near the magnitude of, and the 40 Å distance is longer than the predicted random separation of the donor and acceptor. The longest observed distance of 40 Å is considerably shorter than the maximum possible donor-acceptor separation.

5.4.4 Comparison with previous FRET data on the denatured state of iso-1-cytochrome *c*

Previously, Pletneva, *et. al.*, reported distance distributions from TR-FRET measurements on a yeast iso-1-cytochrome *c* variant labeled with the AEDANS fluorophore at position 39.⁷⁴ Imidazole was used to probe the open loop form at pH 7. At 2.8 M gdnHCl, excess imidazole, pH 7 conditions, three distinct distance distributions were present at ~18, 25 and 40 Å.⁷⁴ The variants used in this study do not have a sole histidine to form discrete loops. In fact, the known heme competing ligands, His26, His33 and the N-terminal amine group were all present. Thus, misligation of various heme competing ligands to the heme was possible. Addition of excess imidazole to counteract distributions due to misligation resulted in a 10% decrease in the distributions of the two compact populations but did not eliminate them.⁷⁴ Our AcH54I52C37 variant does not suffer from misligation because it has the H26N, H33N, and H39Q mutations as well as an acetylated N-terminus. Additionally, our variant forms a single His54-heme loop compared to the work of Pletneva *et. al.*. Thus, the populations we see at pH ~3 (open loop form) should be directly those of the unfolded DSE and not those from misligated species.

5.4.5 What causes compact subpopulations in the denatured state probed at positions 37 and 39?

The question then is, why are three different subpopulations observed for the DSE when it is probed at positions 37 and 39? To answer this question we must consider the region of our labeling site, the amino acid sequence in that area and the spatial arrangement with respect to the heme and the other substructures in the fully folded protein. An analysis of sequence hydrophobicity of the AcTMI52 variant (only sequence differences compared to AcH54I52 are the naturally-occurring residues Gly37 and Lys54) of iso-1-cytochrome *c*, using the Eisenberg hydrophobicity scoring scale, is shown in Figure 5.11.

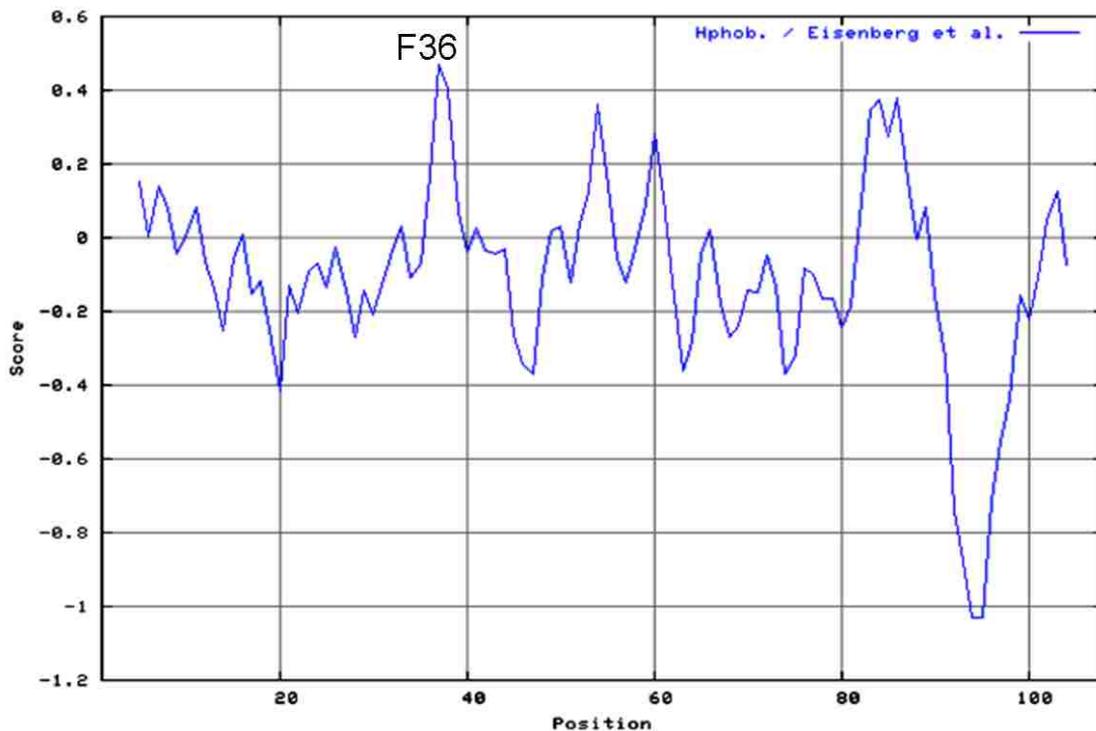


Figure 5.11: Hydrophobicity plot of AcTMI52 variant amino acid sequence using the Eisenberg scoring scale with a window of 9 and default values of the ProtScale program from the ExpASY website.

Evident from Figure 5.11 is that residue 37 is in a region of the polypeptide chain that is strongly hydrophobic; there is a particularly hydrophobic phenylalanine right next to it at position 36. It is widely accepted that the formation of hydrophobic clusters^{51; 205} is a major stabilizing force that drives protein folding. Thus, in the unfolded form of the protein, the tendency for these exposed hydrophobic residues to interact would be high.

TR-FRET studies on cytochrome *c'*, which has significantly more hydrophobic residues than iso-1-cytochrome *c*, 54.4% *versus* 49.1%, have shown that it forms more compact structures (50%) than iso-1-cytochrome *c* (10%) in the denatured state at *gdnHCl* concentrations well beyond the unfolding transition.^{75; 77} Additionally, as seen in Figure 5.1, the Cys37 labeling site is relatively near to the heme. Therefore, it is very likely that this hydrophobic cofactor confers significant stability to this region of the protein. Studies on heme-peptide complexes, having a ligand environment similar to *b* type cytochromes, have shown that hydrophobic amino acids, particularly alanines and phenylalanines, confer significant binding and stabilizing energy to the complexes.^{139; 140} Therefore, having a hydrophobic phenylalanine in the vicinity of position 37 may quite possibly be a naturally engineered attribute that is present in the iso-1-cytochrome *c* protein to assist in the initial stages of protein folding.¹³⁹ Hence, hydrophobic clustering may adequately explain the behavior seen in our TR-FRET data for the Cys37 labeled variant, providing transient stabilization of compact subpopulations in the DSE.

The existence of degenerate compact and extended populations in the DSE has been suggested as essential to protein folding to prevent “energetic frustration” of the polypeptide chain during folding.^{74; 75} The rapid exchange between these degenerate populations may enable a misfold to be dynamically repaired before leading to disease-

causing aggregates. In the case of cytochrome *c*, the hydrophobic heme seems to influence the persistence of these compact conformations when hydrophobic residues are near it in the denatured state;^{74; 76} conversely, that observation is not seen with barstar,⁷⁸ presumably because it does not have such a hydrophobic cofactor to influence its hydrophobic amino acids in the unfolded state. Thus, the DSE populations seen in barstar are more extended with minimal population of compact forms.⁷⁸

5.4.6 Impact of loop formation on the denatured state ensemble

The formation of a discrete loop and probing its impact on the DSE is unique to work being done in this field. Tracking the distributions of distances as this discrete loop is formed in the DSE is novel and has clearly provided significant contributions to elucidating the dynamics of compact and extended conformations present in the heterogeneous DSE. Figure 5.12 is a representation of the transition in populations seen in the DSE going from the open loop form to the closed loop form in 3 M gdnHCl. Creating a closed loop reduces the number of residues between the heme and the AEDANS probe from 20 to 18, respectively. These distances are essentially the same, yet the extended form at 40 Å practically disappears. This result demonstrates that a simple long range contact formed in the DSE dramatically shifts the DSE toward more compact subpopulations. Therefore, early events in protein folding might be expected to have a similar effect which would drastically reduce the conformational search for the stable native fold, resulting in efficient protein folding.

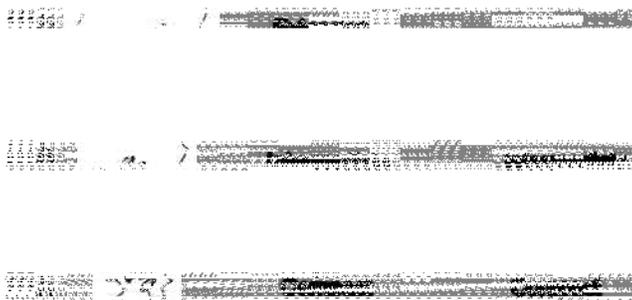


Figure 5.12: Schematic representation of the effects of loop formation on the DSE.

However, we cannot completely rule out the possibility that the multi-exponential emission decay component of the AEDANS donor at position 37 of iso-1-cytochrome *c* is due to local sequence effects – possibly quenching by the adjacent Arg 38. It is not possible to detect this effect using the β -ME-AEDANS as a control. The more peptide-like N-acetylcysteine-AEDANS control also shows only single exponential decay character.⁷⁶ However, Saxena, *et. al.* claim it is not possible to see multi-exponential emission decay with N-acetylcysteine-AEDANS as the control.⁷⁸ Thus, further experiments with a short peptide model are necessary, to clarify whether local sequence affects the decay properties of AEDANS.⁷⁸

5.5 Conclusion

The results of this chapter should still be considered preliminary, even though there is strong evidence of the behavior of denatured state conformers. A peptide based control needs to be done to account for the possibility of multi-exponential AEDANS

fluorescence decay due to local influences of adjacent residues. Also, additional experimentation is needed on different loop sizes with donor labels at different positions relative to the heme. The set of variants currently prepared are of an appropriate variety to provide meaningful information. This additional information will clarify some of the interactions seen in this preliminary study.

The Cys37 is located on the relatively stable green foldon as defined by Englander, *et. al.*,⁷² thus, it is not surprising to see compact structures present in DSE of AcH54I52C37. In other TR-FRET studies on iso-1-cytochrome *c*, increasing gdnHCl concentrations from 2.7 M to 4.4 M only decreased the distribution of compact structures from 60% to 30% but never completely eliminated them.⁷⁴ Therefore, possible additional experimentation could involve TR-FRET at different gdnHCl concentrations for our sets of variants to probe the breakup of assumed hydrophobic clusters around the heme. However, at this time it is believed that the highly hydrophobic nature of the heme seems to be extremely influential in attracting hydrophobic residues and stabilizing the resulting “supercompact” DSE conformations. However, a recent alternate explanation implied that these “compact” conformations are due to stabilizing residual structures from local electrostatic interactions.¹⁰¹ These interactions are only effective over short distances and may be the stabilizing forces responsible for maintaining the initial contacts which guide the polypeptide chain down the folding funnel.¹⁰¹ Also, loop formation studies with peptide models have implied that stabilizing forces for initial contacts made are simply the abundant intramolecular hydrogen bonding of the peptide backbone.¹¹³

Previous studies of equilibrium loop formation at different gdnHCl concentrations showed that the AcH54I52 variant had the most stable loop [lowest

$pK_a(\text{obs})$].⁸² The m -value (degree of hydrophobic burial) for this variant was the highest of several other loop forming variants.⁸² This variant therefore had an optimal hydrophobic content for efficient loop collapse and residual structure formation. It was suggested that this behavior may have evolved because a 35 residue segment is near optimal for the formation of supersecondary structure.⁸² Our current TR-FRET data on this same AcH54I52 variant provide additional insights in explaining why this 37-residue loop was the most stable loop forming variant among all the other variants previously studied.⁸² The increased stability of loop formation for this 37-residue loop may have been due to residual structure in the sequence forming the loop rather than from an optimal loop size for hydrophobic collapse by mass action. Future TR-FRET studies on the remaining variants, especially the AcH73I52C45 and the AcH89I52C53 variants should be extremely informative in clarifying our interpretations of the effect of loop formation on the DSE. The Cys45 variant is away from the heme therefore stable “supercompact” conformer populations might be expected to be drastically reduced, or completely converted to extended conformers if the hydrophobic heme is the stabilizing entity. In addition, the attached fluorophore would be in a far less hydrophobic region as seen from the Eisenberg hydrophobicity plot in Figure 5.11. The Cys53 is also intriguing because it is a part of a hydrophobic region as well (see Figure 5.11), yet it is away from the heme in the native state. Thus, the influence from intra-residue hydrophobic interactions should be high in the unfolded state, but the impact of the heme might be diminished.

References

1. Berg, J. M., Tymoczko, J. L. & Stryer, L. (2002). *Biochemistry: Protein Structure and Function*. 5th edit, Freeman and Company, New York.
2. Nakai, S. & Modler, W. (1996). *Food Proteins: Properties and Characterization*, Wiley-VCH, Weinheim, Germany.
3. Anfinsen, C. B., Haber, E., Sela, M. & White Jr., F. H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences, USA* **47**, 1309-1314.
4. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.
5. Onuchic, J., Luthey-Schulten, Z. & Wolynes, P. G. (1997). Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry* **48**, 545-600.
6. Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique* **65**, 44-45.
7. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural Biology* **4**, 10-19.
8. Nelson, D. L. & Cox, M. M. (2004). *Lehninger Principles of Biochemistry*. 4th edit, W. H. Freeman, New York.
9. Kubelka, J., Chiu, T. K., Davies, D. R., Eaton, W. A. & Hofrichter, J. (2006). Submicrosecond protein folding. *Journal of Molecular Biology* **359**, 546-553.
10. Arora, P., Oas, T. G. & Myers, J. K. (2004). Faster and faster: A designed variant of the B-domain of protein A folds in 3 μ sec. *Protein Science* **13**, 847-853.
11. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology* **277**, 985-994.
12. Chan, H. S., Bromberg, S. & Dill, K. (1995). Models of co-operativity in protein folding. *Philosophical Transactions of the Royal Society of London, Series B* **348**, 61-70.
13. Chan, H. S., Shimizu, S. & Kaya, H. (2004). Cooperativity principle in protein folding. *Methods in Enzymology* **380**, 350-379.
14. Baldwin, R. L. & Rose, G. D. (1999). Is protein folding hierarchic? I. Local structure and peptide folding. *Trends in Biochemical Sciences* **24**, 26-33.

15. Baldwin, R. L. & Rose, G. D. (1999). Is protein folding hierarchic? I. Folding intermediates and transition states. *Trends in Biochemical Sciences* **24**, 77-83.
16. Popot, J.-L. & Engelman, D. M. (1990). Membrane protein folding and oligomerization: The two-stage model. *Biochemistry* **29**, 4031-4037.
17. Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proceedings of the National Academy of Sciences, USA* **70**, 697-701.
18. Fersht, A. R. (1995). Optimisation of rates of protein folding - The nucleation-condensation mechanism and its implications. *Proceedings of the National Academy of Sciences, USA* **92**, 10869-10873.
19. Fersht, A. R. (1997). Nucleation mechanisms in protein folding. *Current Opinion in Structural Biology* **7**, 3-9.
20. Kuwajima, K. (1989). The molten globular state as a clue for understanding the folding cooperativity of globular protein structure. *Proteins: Structure, Function and Genetics* **6**, 87-103.
21. Dill, K. A., Bromberg, S., Yue, K. Z., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995). Principles of protein folding - a perspective from simple exact models. *Protein Science* **4**, 561-602.
22. Ptitsyn, O. B. (1996). How molten is the molten globule? *Nature Structural Biology* **3**, 488-490.
23. Bedard, S., Krishna, M. M. G., Mayne, L. & Englander, S. W. (2008). Protein folding: Independent unrelated pathways or predetermined pathway with optional errors. *Proceedings of the National Academy of Sciences, USA* **105**, 7182-7187.
24. Pande, V. S., Grosberg, A. Y., Tanaka, T. & Rokhsar, D. S. (1998). Pathways for protein folding: Is a new view needed? *Current Opinion in Structural Biology* **8**, 68-79.
25. Rumbley, J., Hoang, L., Mayne, L. & Englander, S. W. (2001). An amino acid code for protein folding. *Proceedings of the National Academy of Sciences, USA* **92**, 105-112.
26. Dobson, C. M. (2003). Protein folding and misfolding. *Nature* **426**, 884-890.
27. Thomas, P. J., Qu, B. H. & Pedersen, P. L. (1995). Defective protein folding as a basis of human disease. *Trends in Biochemical Sciences* **20**, 456-459.
28. Horwich, A. (2002). Protein aggregation in disease: A role for folding intermediates forming specific multimeric interactions. *Journal of Clinical Investigation* **110**, 1221-1232.

29. Dobson, C. M. (2001). The structural basis of protein folding and its links with human disease. *Philosophical Transactions of the Royal Society of London, Series B* **356**, 133-145.
30. Tan, S. Y. & Pepys, M. B. (1994). Amyloidosis. *Histopathology* **25**, 403-414.
31. Kelly, J. W. (1998). Alternative conformation of amyloidogenic proteins and their multi-step assembly pathways. *Current Opinion in Structural Biology* **8**, 101-106.
32. Jahn, T. R. & Radford, S. E. (2005). The Yin and Yang of protein folding. *FEBS Journal* **272**, 5962-5970.
33. Tanford, C. (1968). Protein denaturation. *Advances in Protein Chemistry* **23**, 121-282.
34. Cho, J., Saito, S., Horng, J., Anil, B. & Raleigh, D. (2008). Electrostatic interactions in the denatured state ensemble: Their effect upon protein folding and protein stability. *Archives of Biochemistry and Biophysics* **469**, 20-28.
35. Shortle, D. (1996). Structural analysis of non-native states of proteins by NMR methods. *Current Opinion in Structural Biology* **6**, 24-30.
36. Whitten, S. T. & Garcia-Moreno, E. B. (2000). pH dependence of the stability of staphylococcal nuclease: Evidence of substantial electrostatic interactions in the denatured state. *Biochemistry* **39**, 14292-14304.
37. 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, Inc., Hauppauge, NY.
38. Bowler, B. E. (2007). Thermodynamics of protein denatured states. *Molecular BioSystems* **3**, 88-99.
39. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chemical Reviews (Washington, D. C.)* **104**, 3607-3622.
40. Shortle, D. (1996). The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB Journal* **10**, 27-34.
41. Dyson, H. J. & Wright, P. E. (1998). Equilibrium NMR studies of unfolded and partially folded proteins. *Nature Structural Biology* **5**, 499-503.
42. Tanford, C., Kawahara, K. & Lapanje, S. (1966). Proteins in 6 M guanidine hydrochloride: Demonstration of random coil behavior. *Journal of Biological Chemistry* **241**, 1921-1923.

43. Tanford, C., Kawahara, K. & Lapanje, S. (1967). Proteins as random coils. I. Intrinsic viscosities and sedimentation coefficients in concentrated guanidine hydrochloride. *Journal of the American Chemical Society* **89**, 729-736.
44. Tanford, C., Kawahara, K., Lapanje, S., Hooker, T. M., Zarlengo, M. H., Salahuddin, A., Aune, K. C. & Takagi, T. (1976). Proteins as random coils. III. Optical rotary dispersion in 6 M guanidine hydrochloride *Journal of the American Chemical Society* **89**, 5023-5029.
45. Fleming, P. J. & Rose, G. D. (2005). Conformational properties of unfolded proteins. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 1, pp. 710-736. 6 vols. Wiley-VCH, Weinheim, Germany.
46. Aune, K. C., Salahuddin, A., Zarlengo, M. H. & Tanford, C. (1967). Evidence for residual structure in acid- and heat-denatured proteins. *Journal of Biological Chemistry* **242**, 4486-4489.
47. Tsong, T. Y. (1974). The Trp-59 fluorescence of ferricytochrome *c* as a sensitive measure of the over-all protein conformation. *Journal of Biological Chemistry* **249**, 1988-1990.
48. Tsong, T. Y. (1975). An acid induced conformational transition of denatured cytochrome *c* in urea and guanidine hydrochloride solution. *Biochemistry* **7**, 1542-1547.
49. Tiffany, M. L. & Krimm, S. (1972). Effect of temperature on the circular dichroism spectra of polypeptides. *Biopolymers* **11**, 2309-2316.
50. Tiffany, M. L. & Krimm, S. (1973). Extended conformation of polypeptides and proteins in urea and guanidine hydrochloride. *Biopolymers* **12**, 575-587.
51. Neri, D., Billeter, M., Wider, G. & Wüthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **257**, 1559-1563.
52. Crowhurst, K. A. & Foman-Kay, J. D. (2003). Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain. *Biochemistry* **42**, 8687-8695.
53. Marsh, J. A., Neale, C., Jack, F. E., Choy, W.-Y., Lee, A. Y., Crowhurst, K. A. & Foman-Kay, J. D. (2007). Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. *Journal of Molecular Biology* **367**, 1494-1510.
54. Sosnick, T. R. & Trewhella, J. (1992). Unfolded states of ribonuclease A have compact dimensions and residual secondary structure. *Biochemistry* **31**, 8329-8335.

55. Garcia, P., Serrano, L., Durand, D., Rico, M. & Bruix, M. (2001). NMR and SAXS characterization of the denatured state of the chemotactic protein Che Y: Implications for protein folding initiation. *Protein Science* **10**, 1100-1112.
56. Feng, 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.
57. Wrabl, J. & Shortle, D. (1999). A model of the changes in the denatured state structure underlying *m* value effects in staphylococcal nuclease. *Nature Structural Biology* **6**, 876-883.
58. Nicholson, M. M., Mo, H., Prusiner, S. B., Cohen, F. E. & Marqusee, S. (2002). Differences between the prion protein and its homolog dopel: A partially structured state with implications for scrapie formation. *Journal of Molecular Biology* **316**, 807-815.
59. Bhutani, N. & Udgaonkar, J. B. (2003). Folding domains of thioredoxin characterized by native-state hydrogen exchange. *Protein Science* **12**, 1719-1731.
60. Shortle, D. & Meeker, A. K. (1986). Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. *Proteins: Structure, Function and Genetics* **1**, 81-89.
61. Shortle, D. (1995). Staphylococcal nuclease: A showcase of *m*-value effects. *Advances in Protein Chemistry* **46**, 217-245.
62. Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. (1999). pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. *Biochemistry* **38**, 4896-4903.
63. Cho, J., Sato, S. & Raleigh, D. (2004). Thermodynamics and Kinetics in Protein Folding: A single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state. *Journal of Molecular Biology* **338**, 827-837.
64. Cho, J. H., Saito, S. & Raleigh, D. P. (2008). *Non-native electrostatic interactions in the denatured state ensemble: Effects on protein stability and folding*. Unfolded proteins: from denatured to intrinsically disordered (Creamer, T. P., Ed.), Nova Science Publishers, Inc., Hauppauge, New York.
65. Cho, J. & Raleigh, D. (2005). Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins. *Journal of Molecular Biology* **353**, 174-185.
66. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999). Increasing protein

- stability by altering long range Coulombic interactions. *Protein Science* **8**, 1843-1849.
67. Pace, C. N., Alston, R. W. & Shaw, K. L. (2000). Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Science* **9**, 1395-1398.
 68. Pakula, A. A. & Sauer, R. T. (1990). Reverse hydrophobic effects relieved by amino-acid substitutions at a protein surface. *Nature* **344**, 363-364.
 69. Bowler, B. E., May, K., Zaragoza, T., York, P., Dong, A. & Caughey, W. S. (1993). Destabilizing effects of replacing a surface lysine cytochrome *c* with aromatic amino acids: Implications for the denatured state. *Biochemistry* **32**, 183-190.
 70. Herrmann, L., Bowler, B. E., Dong, A. & Caughey, W. S. (1995). The effects of hydrophilic to hydrophobic surface mutations on the denatured state of iso-1-cytochrome *c*: Investigation of aliphatic residues. *Biochemistry* **34**, 3040-3047.
 71. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: Native-state hydrogen exchange. *Science* **269**, 192-197.
 72. Krishna, M. M. G., Hoang, L., Lin, Y. & Englander, S. W. (2004). Hydrogen exchange methods to study protein folding. *Methods* **34**, 51-64.
 73. Zhang, M.-M., Ford, C. D. & Bowler, B. E. (2004). Estimation of the compaction of the denatured state by a protein variant involved in a reverse hydrophobic effect. *Protein Journal* **23**, 119-126.
 74. Pletneva, E. V., Gray, H. B. & Winkler, J. R. (2005). The many faces of the unfolded state: Conformational heterogeneity in denatured yeast cytochrome *c*. *Journal of Molecular Biology* **345**, 855-867.
 75. 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.
 76. Lyubovitsky, J. G., Gray, H. B. & Winkler, J. R. (2004). Conformations of unfolded *S. cerevisiae* cytochrome *c* probed by fluorescence energy transfer kinetics. *Israel Journal of Chemistry* **44**, 263-269.
 77. Lyubovitsky, J. G., Gray, H. B. & Winkler, J. R. (2002). Mapping the cytochrome *c* folding landscape. *Journal of the American Chemical Society* **124**, 5481-5485.
 78. Saxena, A. M., Udgaonkar, J. B. & Krishnamoorthy, G. (2006). Characterization of intra-molecular distances and site-specific dynamics in chemically unfolded

- barstar: Evidence for denaturant-dependent non-random structure. *Journal of Molecular Biology* **359**, 174-189.
79. Tripp, K. W. & Barrick, D. (2008). Rerouting the folding pathway of the Notch Ankyrin domain by reshaping the energy landscape. *Journal of the American Chemical Society* **130**, 5681-5688.
 80. 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.
 81. 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.
 82. 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.
 83. Tzul, F. O., Kurchan, E. & Bowler, B. E. (2007). Sequence composition effects on denatured state loop formation in iso-1-cytochrome *c* variants: Polyalanine versus polyglycine inserts. *Journal of Molecular Biology* **371**, 577-584.
 84. Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry. The Behavior of Biological Molecules*. 1st edit, vol. 3, W. H. Freeman and Co., San Francisco.
 85. Flory, P. J. (1969). *Statistical Mechanics of Chain Molecules*, John Wiley & Sons, Inc., New York.
 86. Miller, W. G., Brant, D. A. & Flory, P. J. (1967). Random coil configurations of polypeptide copolymers. *Journal of Molecular Biology* **23**, 67-80.
 87. Chan, H. S. & Dill, K. (1991). Polymer principles in protein structure and stability. *Annual Review of Biophysics and Biophysical Chemistry* **20**, 447-490.
 88. Chan, H. S. & Dill, K. A. (1990). The effect of internal constraints on the configurations of chain molecules. *Journal of Physical Chemistry* **92**, 3118-3135.
 89. 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.
 90. Kurchan, E. (2005). Kinetics and thermodynamics of cytochrome *c*. Ph. D. Dissertation, University of Denver.
 91. Dill, K. & Shortle, D. (1991). Denatured states of proteins. *Annual Review of Biochemistry* **60**, 795-825.

92. Tran, H. T., Mao, A. & Pappu, R. V. (2008). Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *Journal of the American Chemical Society* **130**, 7380-7392.
93. Shi, Z., Chen, K., Lim, Z. & Kallenbach, N. R. (2006). Conformation of the backbone in unfolded proteins. *Chemical Reviews (Washington, D. C.)* **106**, 1877-1897.
94. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S., Wirmer, J., Duchardt, E., Ueda, T., Smith, L. J., Dobson, C. M. & Schwalbe, H. (2002). Long-range interactions within a non-native protein. *Science* **295**, 1719-1722.
95. Bond, C. J., Wong, K., Clarke, J., Fersht, A. R. & Daggett, V. (1997). Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: Description of the folding pathway. *Proceedings of the National Academy of Sciences, USA* **94**, 13409-13413.
96. Shortle, D. & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293**, 487-489.
97. Kazmirski, S. I., Wong, K., Freund, M. V., Tan, Y., Fersht, A. R. & Daggett, V. (2001). Protein Folding from a highly disordered denatured state: The folding pathway of chymotrypsin 2 inhibitor at atomic resolution. *Proceedings of the National Academy of Sciences, USA* **98**, 4349-4354.
98. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillion, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Zahid Hasan, M., Pande, V. S., Ruczinski, I., Doniach, S. & Plaxco, K. W. (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings of the National Academy of Sciences, USA* **101**, 12491-12496.
99. Anil, B., Craig-Shapiro, R. & Raleigh, D. P. (2006). Design of a hyperstable protein by rational consideration of unfolded state interactions. *Journal of the American Chemical Society* **128**, 3144-3145.
100. Trefethen, J. M., Pace, C. N., Scholtz, J. M. & Brems, D. N. (2005). Charge-charge interactions in the denatured state influence the folding kinetics of Ribonuclease Sa. *Protein Science* **14**, 1934-1938.
101. Weinkam, P., Pletneva, E. V., Gray, H. B., Winkler, J. R. & Wolynes, P. G. (2009). Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles. *Proceedings of the National Academy of Sciences, USA* **106**, 1796-1801.
102. Eaton, W. A., Munoz, V., Thompson, P. A., Henry, E. R. & Hofrichter, J. (1998). Kinetics and dynamics of loops, α -helices, β -hairpins, and fast folding proteins. *Accounts of Chemical Research* **31**, 745-753.

103. Chan, H. S. & Dill, K. A. (1989). Intrachain loops in polymers: Effects of excluded volume. *Journal of Chemical Physics* **90**, 492–509.
104. Weikl, T. R. & Dill, K. A. (2003). Folding rates and low entropy-loss routes of two-state proteins. *Journal of Molecular Biology* **329**, 585-598.
105. Fersht, A. R. (2000). Transition-state structure as a unifying basis in protein folding mechanisms: Contact order, chain topology, stability and the extended nucleus mechanism. *Proceedings of the National Academy of Sciences, USA* **97**, 1525-1529.
106. Roder, H., Elöve, G. A. & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome *c* by H-exchange labeling and protein NMR. *Nature* **335**, 700-704.
107. Krishna, M. M. G. & Englander, S. W. (2005). The N-terminal and C-terminal motif in protein folding and function. *Proceedings of the National Academy of Sciences, USA* **102**, 1053-1058.
108. Fierz, B. & Kiefhaber, T. (2005). Dynamics of unfolded polypeptide chains. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 1, pp. 809-855. 6 vols. Wiley-VCH, Weinheim, Germany.
109. Pierce, M. M. & Nall, B. T. (1997). Fast folding of cytochrome *c*. *Protein Science* **6**, 618-627.
110. Krieger, F., Möglich, A. & Kiefhaber, T. (2005). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *Journal of the American Chemical Society* **127**, 3346–3352.
111. Krieger, F., Fierz, B., Bieri, O., Drewello, M. & Kiefhaber, T. (2003). Dynamics of unfolded polypeptide chains as models for the earliest steps in protein folding. *Journal of Molecular Biology* **332**, 265-274.
112. Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2002). Measuring dynamic flexibility of the coil state of a helix forming peptide. *Journal of Molecular Biology* **319**, 19-25.
113. Möglich, A., Joder, K. & Kiefhaber, T. (2006). End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proceedings of the National Academy of Sciences, USA* **103**, 12394-12399.
114. Singh, V. R. & Lapidus, L. J. (2008). The intrinsic stiffness of polyglutamine peptides. *Journal of Physical Chemistry B* **112**, 13172-13176.

115. Robinson, C. R. & Sauer, R. T. (1998). Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proceedings of the National Academy of Sciences, USA* **95**, 5929–5934.
116. Ohnishi, S., Kamikubo, H., Onitsuka, M., Kataoka, M. & Shortle, D. (2006). Conformational preference of polyglycine in solution to elongated structure. *Journal of the American Chemical Society* **128**, 16338-16344.
117. 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.
118. Godbole, S. & 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.
119. Hammack, B. N., 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.
120. Berghuis, A. M. & Brayer, G. D. (1992). Oxidation state-dependent conformational changes in cytochrome *c*. *Journal of Molecular Biology* **223**, 959–976.
121. Pierce, M. M. & Nall, B. T. (2000). Coupled kinetic traps in cytochrome *c* folding: His-heme misligation and proline isomerization. *Journal of Molecular Biology* **298**, 955-969.
122. Onishi, S., Kamikubo, H., Onitsuka, M., Kataoka, M. & Shortle, D. (2006). Conformational preference of polyglycine in solution to elongated structure. *Journal of the American Chemical Society* **128**, 16338-16344.
123. Deng, W. P. D. & Nickoloff, J. A. (1992). Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Analytical Biochemistry* **200**, 81-88.
124. Wandschneider, E., Hammack, B. N. & Bowler, B. E. (2003). Evaluation of cooperative interactions between substructures of iso-1-cytochrome *c* using double mutant cycles. *Biochemistry* **42**, 10659-10666.
125. Russel, M., Kidd, S. & Kelley, M. R. (1986). An improved filamentous helper phage for generating single-stranded plasmid DNA. *Gene* **45**, 333-338.
126. Rosell, F. I. & Mauk, A. G. (2002). Spectroscopic properties of a mitochondrial cytochrome *c* with a single thioether bond to the heme prosthetic group. *Biochemistry* **41**, 7811 - 7818.

127. Pollack, W. B. R., Rosell, F. I., Twitchett, M. B., Dumont, M. E. & Mauk, A. G. (1998). Bacterial expression of mitochondrial cytochrome *c*. Trimethylation of Lys 73 in yeast iso-1-cytochrome *c* and the alkaline conformational transition. *Biochemistry* **37**, 6124-6131.
128. Humphreys, D. P., King, L. M., West, S. M., Chapman, A. P., Sehdev, M., Redden, M. W., Glover, D. J., Smith, B. J. & Stephens, P. E. (2000). Improved efficiency of site-specific copper (II) ion-catalyzed protein cleavage effected by mutagenesis of cleavage sites. *Protein Engineering* **13**, 201-206.
129. Sereikaite, J., Jachno, J., Santockyte, R., Chmielewski, P., Bumelis, V. & Dienys, G. (2006). Protein scission by metal ion-ascorbate system. *Protein Journal* **25**, 369-378.
130. Buettner, G. R. (1988). In the absence of catalytic metals ascorbate does not autoxidize at pH 7: Ascorbate as a test for catalytic metals. *Journal of Biochemical and Biophysical Methods* **16**, 27-40.
131. Kristinsson, R. & Bowler, B. E. (2005). Communication of stabilizing energy between substructures of a protein. *Biochemistry* **44**, 2349-2359.
132. Schellman, J. A. (1978). Solvent denaturation. *Biopolymers* **17**, 1305-1322.
133. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods in Enzymology* **131**, 266-280.
134. Hagihara, Y., Tan, Y. & Goto, Y. (1994). Comparison of the conformational stability of the molten globule and native states of horse cytochrome *c*: Effects of acetylation, heat, urea and guanidine-hydrochloride. *Journal of Molecular Biology* **237**, 336 - 348.
135. 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 - 383.
136. Anil, B., Song, B., Tang, Y. & Raleigh, D. P. (2004). Exploiting the right side of the Ramachandran plot: Substitution of glycines by D-alanine can significantly increase protein stability. *Journal of the American Chemical Society* **126**, 13194-13195.
137. Hagen, S. J., Hofrichter, J. & Eaton, W. A. (1997). Rate of intrachain diffusion of unfolded cytochrome *c*. *Journal of Physical Chemistry B* **101**, 2352-2365.
138. Jacobson, H. & Stockmayer, W. H. (1950). Intramolecular reaction in polycondensations. I. The theory of linear systems. *Journal of Chemical Physics* **18**, 1600-1606.

139. Huffman, D. L., Rosenblatt, M. M. & Suslick, K. S. (1998). Synthetic heme-peptide complexes. *Journal of the American Chemical Society* **120**, 6183-6184.
140. Huffman, D. L. & Suslick, K. S. (2000). Hydrophobic interactions in metalloporphyrin-peptide complexes. *Inorganic Chemistry* **39**, 5418-5419.
141. Liu, D., Williamson, D. A., Kennedy, M. L., Williams, T. D., Morton, M. M. & Benson, D. R. (1999). Aromatic side chain-porphyrin interactions in designed hemoproteins. *Journal of the American Chemical Society* **121**, 11798-11812.
142. Chakrabartty, A., Kortemme, T. & Baldwin, R. L. (1994). Helix stabilizing propensities of the amino acids measured in alanine-based peptides without helix stabilizing side-chain interactions. *Protein Science* **3**, 843-852.
143. Yeh, S. & Rousseau, D. L. (1998). Folding intermediates in cytochrome *c*. *Nature Structural Biology* **5**, 222-228.
144. Lee, J. C., Chang, I., Gray, H. B. & Winkler, J. R. (2002). The cytochrome *c* folding landscape revealed by electron-transfer kinetics. *Journal of Molecular Biology* **320**, 159-164.
145. Ramsden, W. (1902). Some new properties of urea. *Journal of Physiology (London)* **28**, 23-27.
146. Simpson, R. B. & Kauzman, W. (1953). The kinetics of protein denaturation. I. The behavior of the optical rotation of ovalbumin in urea solutions. *Journal of the American Chemical Society* **75**, 5139-5152.
147. Frensdorff, K. H., Watson, M. T. & Kauzman, W. (1953). The kinetics of protein denaturation. V. The viscosity of urea solutions of serum albumin. *Journal of the American Chemical Society* **75**, 5167-5172.
148. Tanford, C. (1964). Isothermal unfolding of globular proteins in aqueous urea solutions. *Journal of the American Chemical Society* **86**, 2050-2059.
149. Möglich, A., Krieger, F. & Kiefhaber, T. (2005). Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. *Journal of Molecular Biology* **345**, 153-162.
150. Schellman, J. A. (2002). Fifty years of solvent denaturation. *Biophysical Chemistry* **96**, 91-101.
151. Schellman, J. A. (1987). Selective binding and solvent denaturation. *Biopolymers* **26**, 549-559.
152. Makhatadze, G. I. (1999). Thermodynamics of protein interactions with urea and guanidine hydrochloride. *Journal of Physical Chemistry B* **103**, 4781-4785.

153. Greene Jr., R. F. & Pace, C. N. (1974). Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α -Chymotrypsin, and β -Lactoglobulin. *Journal of Biological Chemistry* **249**, 5388-5393.
154. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein folding. *Protein Science* **4**, 2138-2148.
155. Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2002). Measuring the rate of intramolecular contact formation in polypeptides. *Proceedings of the National Academy of Sciences, USA* **97**, 7220-7225.
156. Kuznetsov, S. V., Shen, Y., Benight, A. S. & Ansari, A. (2001). A semiflexible polymer model applied to loop formation in DNA hairpins. *Biophysical Journal* **81**, 2864-2875.
157. Shen, Y., Kuznetsov, S. V. & Ansari, A. (2001). Loop dependence of the dynamics of DNA hairpins. *Journal of Physical Chemistry B* **105**, 12202-12211.
158. McGuirl, M. A., Lee, J. C., Lyubovitsky, J. G., Thanyakoo, C., Richards, J. H., Gray, H. B. & Winkler, J. R. (2003). Cloning, heterologous expression, and characterization of recombinant class II cytochromes *c* from *Rhodospseudomonas palustris*. *Biochimica et Biophysica Acta* **1619**, 23-28.
159. Arslan, E., Schulz, H., Zufferey, R., Künzler, P. & Thöny-Meyer, L. (1998). Overproduction of the *Bradyrhizobium japonicum* *c*-type cytochrome subunits of *cbb₃* oxidase in *Escherichia coli*. *Biochemical and Biophysical Research Communications* **251**, 744-747.
160. Pletneva, E. V., Zhao, Z., Kimura, T., Petrova, K. V., Gray, H. B. & Winkler, J. R. (2007). Probing the cytochrome *c'* folding landscape. *Journal of Inorganic Biochemistry* **101**, 1768-1775.
161. Kimura, T., Lee, J. C., Gray, H. B. & Winkler, J. R. (2007). Site-specific collapse dynamics guide the formation of the cytochrome *c'* four helix bundle. *Proceedings of the National Academy of Sciences, USA* **104**, 117-122.
162. 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.
163. Das, G., Hickey, D. R., McLendon, G. & Sherman, F. (1989). Dramatic thermostabilization of yeast iso-1-cytochrome *c* by an asparagine to isoleucine replacement at position 57. *Proceedings of the National Academy of Sciences, USA* **86**, 496-499.
164. Nelson, C. J. & Bowler, B. E. (2000). pH dependence of formation of a partially unfolded state of a Lys 73 \rightarrow His variant of iso-1-cytochrome *c*: Implications for

- the alkaline conformational transition of cytochrome *c*. *Biochemistry* **39**, 13584-13594.
165. Sato, S., Sayid, C. J. & Raleigh, D. P. (2000). The failure of simple empirical relationships to predict the viscosity of mixed aqueous solutions of guanidine hydrochloride and glucose has important implications for the study of protein folding. *Protein Science* **9**, 1601-1603.
 166. Kawahara, K. & Tanford, C. (1966). Viscosity and density of aqueous solutions urea and guanidine hydrochloride. *Journal of Biological Chemistry* **241**, 3228-3232.
 167. Balbach, J. & Schmid, F. X. (2002). *Proline isomerization and its catalysis in protein folding*. 2nd edit. Mechanisms in Protein Folding (Pain, R. H., Ed.), Oxford University Press Inc., New York.
 168. Oliveberg, M. (1998). Alternative explanations for "multistate" kinetics in protein folding: Transient aggregation and changing transition state ensembles. *Accounts of Chemical Research* **31**, 765-772.
 169. Nawroki, J. P., Chu, R.-A., Pannell, L. K. & Bai, Y. (1999). Intermolecular aggregations are responsible for the slow kinetics observed in the folding of cytochrome *c* at neutral pH. *Journal of Molecular Biology* **293**, 991-995.
 170. Yang, W. & Gruebele, M. (2006). Binary and ternary aggregation within tethered protein constructs. *Biophysical Journal* **90**, 2930-2937.
 171. Rousseau, F., Schymkowitz, J. & Serrano, L. (2006). Protein aggregation and amyloidosis. *Current Opinion in Structural Biology* **16**, 118-126.
 172. Silow, M. & Oliveberg, M. (1997). Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proceedings of the National Academy of Sciences, USA* **94**, 6084-6086.
 173. Silow, M., Tan, Y.-J., Fersht, A. R. & Oliveberg, M. (1999). Formation of short-lived aggregates directly from the coil in two-state folding. *Biochemistry* **38**, 13006-13012.
 174. Finke, J. M. & Jennings, P. A. (2001). Early aggregated states in the folding of interleukin-1 β . *Journal of Biological Physics* **27**, 119-131.
 175. Ganesh, C., Zaidl, F. N., Udgaonkar, J. B. & Varadarajan, R. (2001). Reversible formation of on-pathway macroscopic aggregates during the folding of maltose binding protein. *Protein Science* **10**, 1635-1644.
 176. Smith, C. R., Wandschneider, E. & Bowler, B. E. (2003). Effect of pH on the iso-1-cytochrome *c* denatured state: Changing constraints due to heme ligation. *Biochemistry* **42**, 2174-2184.

177. Kiefhaber, T., Rudolph, R., Kohler, H. H. & Buchner, J. (1991). Protein aggregation *in vitro* versus *in vivo*: A quantitative model of the kinetic competition between folding and aggregation. *Nature Biotechnology* **9**, 825-829.
178. Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. C. & Hall, B. D. (1979). Identification and isolation of the cytochrome *c* gene. *Cell* **16**, 753-761.
179. Faye, G., Leung, D. W., Tatchell, K., Hall, B. D. & Smith, M. (1981). Deletion mapping of sequences essential for *in vivo* transcription of the iso-1-cytochrome *c* gene. *Proceedings of the National Academy of Sciences, USA* **78**, 2258-2262.
180. Regenfuss, P., Clegg, R. M., Fulwyler, M. J., Barrantes, F. J. & Jovin, T. M. (1985). Mixing liquids in microseconds. *Review of Scientific Instruments* **56**, 283-290.
181. Shastry, M. C. R., Luck, S. D. & Roder, H. (1998). A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. *Biophysical Journal* **74**, 2714-2721.
182. Shastry, M. C. R. & Roder, H. (1998). Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Natural Structural Biology* **5**, 385-392.
183. Roder, H., Maki, K., Cheng, H. & Shastry, M. C. R. (2004). Rapid mixing methods for exploring the kinetics of protein folding. *Methods* **34**, 15-27.
184. Nozaki, Y. (1972). The preparation of guanidine hydrochloride. *Methods in Enzymology* **26**, 43-50.
185. Tzul, F. O., Kurchan, E., Roder, H. & Bowler, B. E. (2009). Competition between reversible aggregation and loop formation in denatured iso-1-cytochrome *c*. *Biochemistry* **48**, 481-491.
186. Jackson, S. E. (1992). How do small single-domain proteins fold? *Folding and Design* **3**, R81-R91.
187. Davis-Searles, P. R., Saunders, A. J., Erie, D. A., Winzor, D. J. & Pielak, G. J. (2001). Interpreting the effects of small uncharged co-solutes on protein-folding equilibria. *Annual Review of Biophysics and Biomolecular Structure* **30**, 271-306.
188. Feldman, D. E. & Frydman, J. (2000). Protein folding *in vivo*: The importance of molecular chaperones. *Current Opinion in Structural Biology* **10**, 26-33.
189. Lyubovitsky, J. G., Gray, H. B. & Winkler, J. R. (2002). Structural features of the cytochrome *c* molten globule revealed by fluorescence energy transfer kinetics. *Journal of the American Chemical Society* **124**, 14840-14841.

190. Kuznetsov, S. V., Hilario, J., Keiderling, T. A. & Ansari, A. (2003). Spectroscopic studies of structural changes in two β -sheet-forming peptides show an ensemble of structures that unfold non-cooperatively. *Biochemistry* **42**, 4321-4332.
191. Calmettes, P., Durand, D., Desmadril, M., Minard, P., Receveur, V. & Smith, J. C. (1994). How random is a highly denatured protein? *Biophysical Chemistry* **53**, 105-114.
192. Lakowicz, J. R., Gryczynski, I., Laczko, G., Wicz, W. & Johnson, M. L. (1994). Distribution of distances between the tryptophan and the N-terminal residue of melittin in its complex with cadmodulin, troponin C, and phospholipids. *Protein Science* **3**, 628-637.
193. Schuler, B., Lipman, E. A. & Eaton, W. A. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743-747.
194. Xie, Z., Srividya, N., Sosnick, T. R., Pan, T. & Scherer, N. F. (2004). Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme. *Proceedings of the National Academy of Sciences, USA* **101**, 534-539.
195. Crick, S. L., Jayaraman, M., Frieden, C., Wetzel, R. & Pappu, R. V. (2006). Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proceedings of the National Academy of Sciences, USA* **103**, 16764-16769.
196. Lakowicz, J. R. (1999). *Principles of Fluorescence Spectroscopy*. 2nd edit, Kluwer Academic/Plenum Publishers, New York, NY.
197. Fraczekiewicz, R. & Braun, W. (2008). Calculation of solvent accessible surface areas, atomic solvation energies and their gradients for macromolecules. In <http://curie.utmb.edu/getarea.html> (Negi, S. S., ed.). Sealy Center for Structural Biology, University of Texas Medical Branch, Galveston, TX 77555, USA.
198. Fraczekiewicz, R. & Braun, W. (1998). Exact and efficient analytical calculation of the the accessible surface area and their gradients for macromolecules. *Journal of Computational Chemistry* **19**, 319-333.
199. Fraczekiewicz, R. & Braun, W. (1996). A new efficient algorithm for calculating solvent accessible surface area of macromolecules. In *Third Electronic Computational Chemistry Conference*, Northern Illinois University.
200. Siefried, S. E., Wang, Y. & von Hippel, P. H. (1988). Fluorescent modification of the cysteine 202 residue of *Escherichia coli* transcription termination factor Rho*. *Journal of Biological Chemistry* **263**, 13511-13514.

201. Hudson, E. N. & Weber, G. (1973). Synthesis and characterization of two fluorescent sulfhydryl reagents. *Biochemistry* **12**, 4154-4161.
202. Yan, Y. & Marriott, G. (2003). Analysis of protein interactions using fluorescence technologies. *Current Opinion in Chemical Biology* **7**, 635-640.
203. Minazzo, A. S., Darlington, R. C. & Ross, J. B. (2009). Loop dynamics of the extracellular domain of Human Tissue Factor and activation of Factor VIIa. *Biophysical Journal* **96**, 681-692.
204. Goldenberg, D. P. (2003). Computational simulation of the statistical properties of unfolded proteins. *Journal of Molecular Biology* **326**, 1615-1633.
205. Garcia, P., Merola, F., Receveur, V., Blandin, P., Minard, P. & Desmadril, M. (1998). Steady state and time-resolved fluorescence study of residual structure in an unfolded form of yeast phosphoglycerate kinase. *Biochemistry* **37**, 7444-7455.

Appendices

Appendix A: Table containing loop breakage rate constants at higher pH's than used for calculation of k_f rate constants

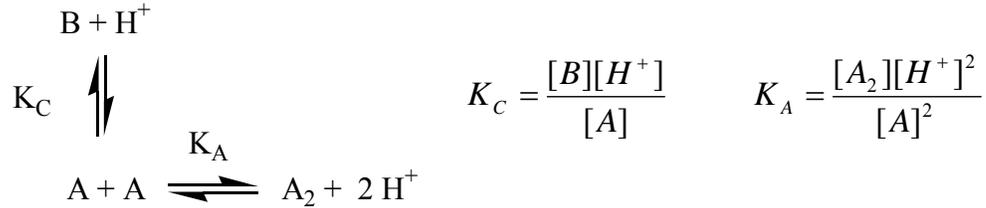
Table A1: Kinetic parameters for loop breakage at 25 °C in 3 M and 6 M guanidine hydrochloride of poly(Ala) iso-1-cytochrome *c* variants at a higher pH

Loop size	Variant	3 M gdnHCl		6 M gdnHCl	
		k_b (pH 3.61 ± 0.09) (s ⁻¹)	k_b (pH 3.96 ± 0.01) (s ⁻¹)	k_b (pH 3.86 ± 0.09) (s ⁻¹)	k_b (pH 4.01 ± 0.02) (s ⁻¹)
16	NH(-2)	112.0 ± 4.0		80.0 ± 2.1	
22	NH5A	100.5 ± 1.0		69.7 ± 1.5	
28	NH5A-2	78.0 ± 0.5		60.0 ± 1.4	
34	NH5A-3		76.6 ± 0.2		58.6 ± 0.2
40	NH5A-4		76.5 ± 1.3		57.8 ± 0.2
46	NH5A-5		76.0 ± 1.9		57.4 ± 0.2

Data are based on three individual trials and errors are from their standard deviations.

$k_{(obs)}$ from loop breakage experiments are presented as k_b , however the contribution from k_f at these higher pH's ranges from ~3 – 18 s⁻¹ and ~2 – 13 s⁻¹ under 3 M and 6 M gdnHCl, respectively.

Appendix B: Derivation of the pH Dependence of $pK_a(\text{obs})$ for Competing Intra- and Intermolecular Reactions



A is the protonated monomer (heme---HisH⁺)
 B is the species with an intramolecular loop
 A₂ is the intermolecular dimer

Conservation of mass gives: $[\text{A}]_t = [\text{A}] + [\text{B}] + 2[\text{A}_2]$,

where $[\text{A}]_t$ is the total concentration of A in all species.

Using $[\text{A}_2] = \text{K}_A[\text{A}]^2/[\text{H}^+]^2$ and $[\text{B}] = \text{K}_C[\text{A}]/[\text{H}^+]$ we obtain:

$$[\text{A}]_t = [\text{A}] + (\text{K}_C/[\text{H}^+])[\text{A}] + 2 (\text{K}_A/[\text{H}^+]^2)[\text{A}]^2$$

If we assume that each monomer in A₂ and B are spectroscopically equivalent, then at the midpoint of a titration, $[\text{A}] = [\text{A}]_t/2$.

The conservation of mass equation then becomes:

$$[\text{A}]_t = [\text{A}]_t/2 + (\text{K}_C/[\text{H}^+])([\text{A}]_t/2) + 2(\text{K}_A/[\text{H}^+]^2)([\text{A}]_t/2)^2$$

Multiplying both sides of the equation by $(2/[\text{A}]_t)[\text{H}^+]^2$ and rearranging we obtain:

$$[\text{H}^+]^2 = \text{K}_C[\text{H}^+] + \text{K}_A[\text{A}]_t$$

Thus,

$$[\text{H}^+]^2 - \text{K}_C[\text{H}^+] - \text{K}_A[\text{A}]_t = 0$$

At the midpoint, $pK_a(\text{obs}) = \text{pH}$, so $pK_a(\text{obs})$ can be obtained by obtaining the roots of this equation using the quadratic formula:

$$[H^+] = \frac{K_C \pm \sqrt{K_C^2 + 4K_A[A]_t}}{2}$$

Since $[H^+]$ must be a positive quantity:

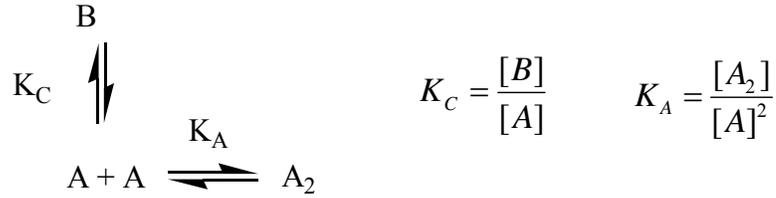
$$[H^+] = \frac{K_C + \sqrt{K_C^2 + 4K_A[A]_t}}{2}$$

And:

$$pK_a(\text{obs}) = -\log\left(\frac{K_C + \sqrt{K_C^2 + 4K_A[A]_t}}{2}\right)$$

This is Eq 4.1 in the text used to fit data in Figure 4.3 of the text.

Appendix C: General Derivation of the Fractional Occupancy of Intra- and Intermolecular Forms for Competing Intra- and Intermolecular Reactions as a Function of Total Reactant Concentration



A is the monomer (heme---His)
 B is the species with an intramolecular loop
 A₂ is the intermolecular dimer

Conservation of mass gives: $[\text{A}]_t = [\text{A}] + [\text{B}] + 2[\text{A}_2]$,

where $[\text{A}]_t$ is the total concentration of A in all species.

Using $[\text{A}_2] = K_A[\text{A}]^2$ and $[\text{B}] = K_C[\text{A}]$ we obtain:

$$[\text{A}]_t = [\text{A}] + K_C[\text{A}] + 2 K_A[\text{A}]^2$$

rearranging gives:

$$2 K_A[\text{A}]^2 + (1+K_C)[\text{A}] - [\text{A}]_t = 0$$

Solving for $[\text{A}]$ using the quadratic formula:

yields:

$$[\text{A}] = \frac{-(1+K_C) \pm \sqrt{(1+K_C)^2 + 8K_A[\text{A}]_t}}{4K_A}$$

Since concentrations cannot be negative only the positive root of the equation is possible:

$$[\text{A}] = \frac{-(1+K_C) + \sqrt{(1+K_C)^2 + 8K_A[\text{A}]_t}}{4K_A}$$

This provides an expression for [A] in terms of [A]_t which is known.

The fraction monomer, f_A, then is:

$$f_A = \frac{[A]}{[A]_t} = \frac{-(1 + K_C) + \sqrt{(1 + K_C)^2 + 8K_A[A]_t}}{4K_A[A]_t}$$

A similar treatment based on conservation of mass can yield the fraction of the intermolecular product, f_B, *versus* [A]_t

Using [A₂] = K_A[A]² and [A] = [B]/K_C we obtain:

$$[A]_t = [B] + ([B]/K_C) + 2 K_A([B]/K_C)^2$$

rearranging gives:

$$2 (K_A/K_C^2)[B]^2 + (1+(1/K_C))[B] - [A]_t = 0$$

Solving for [B] using the quadratic formula yields:

$$[B] = \frac{-\left(1 + \frac{1}{K_C}\right) \pm \sqrt{\left(1 + \frac{1}{K_C}\right)^2 + 8\left(\frac{K_A}{K_C^2}\right)[A]_t}}{4\left(\frac{K_A}{K_C^2}\right)}$$

Since concentrations cannot be negative only the positive root of the equation is possible:

$$[B] = \frac{-\left(1 + \frac{1}{K_C}\right) + \sqrt{\left(1 + \frac{1}{K_C}\right)^2 + 8\left(\frac{K_A}{K_C^2}\right)[A]_t}}{4\left(\frac{K_A}{K_C^2}\right)}$$

This provides an expression for [B] in terms of [A]_t which is known.

The equation for fraction of intramolecular product, f_B, then is

$$f_B = \frac{[B]}{[A]_t} = \frac{-\left(1 + \frac{1}{K_C}\right) + \sqrt{\left(1 + \frac{1}{K_C}\right)^2 + 8\left(\frac{K_A}{K_C^2}\right)[A]_t}}{4\left(\frac{K_A}{K_C^2}\right)[A]_t}$$

This equation is a general form of Eq 4.2 in the text and was used to fit the data for the decrease in intramolecular product at pH 7.1 in Figure 4.15 parts A and B of the text.

The fraction dimer, f_{A_2} is given by:

$$f_{A_2} = 1 - f_A - f_B$$

Using the expressions for f_A and f_B , this equation was used to evaluate the growth in the fraction of intermolecular product in Fig. 4.15A of the text.