2019

PROBING DENATURED STATE CONFORMATIONAL BIAS IN 3-HELIX BUNDLES WITH FOLDABLE OR INTRINSICALLY DISORDERED DOMAINS

Moses Joseph Leavens

PROBING DENATURED STATE CONFORMATIONAL BIAS IN 3-HELIX BUNDLES WITH FOLDABLE OR INTRINSICALLY DISORDERED DOMAINS

By

Moses Joseph Leavens

A.S., General Education, Great Falls College, Montana State University, Great Falls, MT, USA, 2006

B.A., Biology & Chemistry, B.S. Mathematics, University of Providence, Great Falls, MT, USA, 2011

M.S., Biochemistry & Biophysics, University of Montana, Missoula, MT, USA, 2015

Dissertation

Presented in partial fulfillment of the requirements
For the degree of

Doctor of Philosophy
In Biochemistry & Biophysics

The University of Montana
Missoula, MT

May 2019

Approved by:

Dr. Scott Whittenberg, Dean of the Graduate School, Graduate School

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

Dr. Stephen Lodmell, Committee Member, Division of Biological Sciences

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

Dr. Stephen Sprang, Committee Member, Division of Biological Sciences

Dr. Scott Wetzel, Committee Member, Division of Biological Sciences

Dr. Keith Parker, Committee Member, Department of Biomedical & Pharmaceutical Sciences

PROBING DENATURED STATE CONFORMATIONAL BIAS IN 3-HELIX BUNDLES WITH FOLDABLE OR INTRINSICALLY DISORDERED DOMAINS

Chairperson: Bruce E. Bowler

Protein misfolding is associated with several life threatening diseases – Alzheimer's, Parkinson's, Cystic fibrosis, Sickle cell anemia, and the Transmissible spongiform encephalopathies, to include Chronic Wasting disease, Scrapie, and Creutzfeldt-Jakob disease. To provide insight into protein misfolding, it is beneficial to understand how an amino acid sequence encodes structure. Once thought to behave as random coil polymers under denaturing conditions, an abundance of recent evidence suggests denatured proteins retain 'non-random' or residual structure, which may bias a protein chain to fold efficiently to its native conformer. Probing energetics of residual structure in denatured proteins is challenging, given the shortage of available methods to test such questions. Using the metalloprotein Cytochrome *c* (Cyt*c*) as a scaffold, we fuse three highly divergent amino acid sequences with helical topology to the N-terminus of Cyt*c*. Using site-directed mutagenesis, we engineer histidine into highly solvent accessible positions in each of these three folds, creating panels of single histidine variants within each protein sequence. In denaturing conditions, the engineered histidine in each variant serves as ligand to the heme in Cyt*c*, forming a histidine-heme (His-heme) loop with a specific size. Two of the three helical proteins are foldable domains excised from human DNA excision repair protein HHR23A, Ubiquitin-associated domain 1 (UBA(1)) and Ubiquitin-associated domain 2 (UBA(2)). The UBA domains function in Ubiquitin/proteasome signaling. The third protein is mouse cAMP-responsive element binding protein (CBP), an intrinsically disordered protein that serves as a general transcription coactivator involved in hormonal signaling. When bound to its signaling partner activator for thyroid hormone and retinoid receptor (ACTR), CBP folds to an α-helical topology. For UBA(1)-Cyt*c*, UBA(2)-Cyt*c*, and CBP-Cyt*c* fusion proteins, guanidine hydrochloride (GuHCl) denaturation show three state equilibrium unfolding, with the Cyt*c* domain unfolding first. Furthermore, engineered histidine residues in UBA(2) and UBA(1) destabilize the Cyt*c* domain, while CBP has modest effect on Cyt*c*. Equilibrium and kinetic His-heme loop formation measurements in the denatured state at 4 and 6 M GuHCl show that loop stability decreases as the size of the His-heme loop increases, in accord with the Jacobson-Stockmayer equation. However, we observe that the His27-heme loop in UBA(2)-Cyt*c*, the His15-heme & His31-heme loops in UBA(1)-Cyt*c*, are more stable than expected, and break more slowly than expected, relative to His-heme loops exhibiting random coil behavior in the denatured state. Thus, these loops are indicative of non-random behavior in the denatured state. We observe unusual behavior in the biophysical properties of CBP in the denatured state. These results indicate local sequence near His27 in UBA(2), and His15 & His31 in UBA(1), are prone to persistent interactions in the denatured state. When mapped onto their native structures, these regions with denatured state residual structure localize to reverse turns in the UBA domains, and unstructured regions in CBP. Our results on turn bias persisting in the denatured state are consistent with previous findings for the four-helix bundle protein Cyt*c*', where reverse turns may help establish the gross topology of helix-bundles via biasing the conformational distribution of the denatured state. These results could assist the biotechnology and protein engineering fields in developing therapeutics to aid protein misfolding diseases.

Acknowledgements

I would like to acknowledge Bruce and members of our lab for their help, particularly former lab members Levi McClelland (for initial training), Franco Tzul (for his data), Melisa Cherney, and Michaela Finnegan (for their cloning work on this project). Thank you - you made my job a hell of a lot easier! I appreciate your effort. To Bruce, who guided me in this project and helped with manuscript writing. In addition, I would like to extend my gratitude to the following people and organizations: Aaron Thomas, the Montana Established Program to Stimulate Competitive Research, the University of Montana – Missoula (UM), the Sloan Indigenous Graduate Partnership of the Alfred P. Sloan Foundation, the American Indian Science & Engineering Society (AISES), the UM AISES chapter, the National Science Foundation, the Stone Child College Chippewa-Cree Higher Education Program, UL Everclean, and the University of Montana Department of Chemistry & Biochemistry, for their financial support throughout my college education – 14 years! With no financial help, obtaining a college degree would have been more difficult, and perhaps I would have followed in the footsteps of Ray Bradbury, spending three days a week for ten years IN THE LIBRARY! Persistence and dedication, regardless of circumstance or failure, will carry you a long way.

I would like to acknowledge my grandfather Bob Linn and high school running coach Branch Brady, who taught me the importance of hard work, determination, resiliency, accountability, and perseverance. I am thankful for family and friends who encouraged me to change, and reach for my potential—not to compete with others, but to better myself, so I could help. To my wife Raina and daughter Zoey, who have sacrificed their time so I could accomplish this work. Every other weekend for the last 6 years was spent commuting 320 miles for my daughter, driving approximately 8,000 miles per year just for that occasion! Going to school, balancing lab duties, taking courses, learning to parent, and making ends meet, are not easy, and I will not miss that aspect of earning a graduate degree. To my previous research mentors – Laurie P. Shornick, Steve Lodmell, Jordi Mata-Fink, K. Dane Wittrup, Nicole D. Henderson, and Alan D. Christian, thank you for hosting me in your laboratories—providing me an opportunity to do science. To Doug, who reviewed numerous pieces of my writing, and encouraged me to return to graduate school. With the right help, many things are possible. To anyone else I left out, thank you for your help. I am indebted to helping others, because of the people aforementioned who helped me through life.

I hope the following research discoveries herein will shed light on an important fundamental subject in molecular biology:

"How sequence encodes structure remains a mystery." - David Shortle

Table of Contents

List of Figures

List of Tables

List of Equations

Chapter 1: Proteins and the Denatured State

Proteins and protein folding

The word protein stems from the Greek word 'proteios', which means 'of first importance.' These biological molecules are the final product when genes are transcribed to mRNA, and then translated to an amino acid sequence. Proteins carry out essential processes that make life possible—cellular signaling, regulation of gene expression, help transport oxygen in the blood, aid in helping the cytoskeleton give the cell support, and are critical to helping the immune system fight off an infection. Almost every biological event at the molecular level involves a protein. Made up of 20 amino acids, proteins can vary in size, amino acid sequence, functional outcome, and in their final shape or topology.

Cyrus Levinthal was one of the first persons to recognize that the conformational possibilities within even a short peptide could take an numerous amount of time to sample, suggesting that a protein may possibly sample just a subset of these conformations to fold to its final structure. Later on, folding experiments on small domains confirmed that many proteins do indeed fold on a rapid time scale. Shortly afterwards in the 1960s Christian Anfinsen and his colleagues at National Institutes of Health were working on bovine ribonuclease (nuclease). At this time, it was still unclear whether proteins could fold on their own. It was mostly thought at the time that proteins needed some kind of life force within the cellular environment to help them fold. Christian Anfinsen and his coworkers explored this question.[74] The scientific question was determining whether nuclease can denature and upon removing denaturing conditions, refold back to its native structure. If not, then perhaps it needed something else to accommodate its refolding. By reducing disulfide bonds and denaturing nuclease in 8 M urea, then removing those conditions in the reverse order, they discovered that nuclease could regain ~99% of its enzyme activity and thus is refolded to its native shape on its own. Reversing the order of oxidation and

removing the urea resulted in a defective enzyme. Christian Anfinsen's work suggested the native fold of a protein is dictated by its amino acid sequence, not a cellular component, and this amino sequence was later hypothesized to produce the most thermodynamically favorable conformation. Anfinsen's thermodynamic hypothesis assumed that the native structure of a protein is readily accessible (i.e. no barriers to reach native state) and that the primary structure encodes one conformation. Many studies since the 1990s now suggest that many different amino acid sequences can encode the same structure. Studies on T4 lysozyme showed that mutating multiple hydrophobic residues to methionine within T4 lysozyme's hydrophobic core could produce the same structure.[75] In addition, other work has been carried out that demonstrates just one amino acid switch can change a fold entirely to a different conformation.[88] With advances in protein folding research, the question remains as to what biases remain in a protein denatured state that allow a protein to dictate one fold over another. Discovery of these intrinsic conformational biases within a protein's denatured state could provide insight into understanding how evolution works to favor a particular fold, and in addition, elucidate the nature of denatured state conformational biases that could very well provide important information on developing therapeutics for protein misfolding diseases.

Many protein misfolding diseases involve conformational changes that are distinct from the native state, and in many instances, disrupt the normal biological function of that specific protein. Well known examples of this are sickle cell anemia, where red blood cells adopt a sickle shape, where hemoglobin can no longer bind and transport oxygen efficiently. The well-known prion diseases are also another example, where specific mutations, in some cases at or near turns, can switch the prion protein from an α-fold to a β-fold.[77] Furthermore, the energetics of proteins

are often disrupted with single amino acid changes in a protein sequence, and these changes often times do play a role in the denatured state of proteins.[5]

The Denatured State of Proteins

The argument that denatured proteins behave simply as random coils no longer holds ground. The last 25 years of protein folding research have provided sufficient evidence to conclude that several proteins under varying denaturing conditions of urea, guanidine hydrochloride, and temperature retain so called 'non-random' or residual structure. Numerous studies have indicated that the native state topology is present, even at extreme denaturing conditions[6-8,11]. The challenge remains in examining this residual structure in unfolded proteins—where is residual structure located along an unfolded protein chain, and what energetics are different for this part of the chain that allow these regions to behave differently. A few methods have been developed, most notably small angle X-ray scattering (SAXS) and single molecule fluorescence resonance energy transfer (smFRET). On one end of this spectrum, SAXS and smFRET allow calculation of a protein's radius of gyration (R_g) under denaturing conditions, which gives a measure of expansion of a denatured protein. Much debate in the field of protein folding has questioned the effect of fluorophores on the denatured state dimensional properties. Still to this day, debates occur between smFRET and SAXS methods, and often times, the findings reveal differences in R_g.[83] As an alternative, histidine-heme (His-heme) loop formation methods have been developed.[77] The main advantage of His-heme loop formation methods over traditional methods such as SAXS and smFRET for studying the denatured state of proteins, is that His-heme loop formation methods cannot only inform about the bulk properties of a denatured protein, but detect residual structure along certain segments of an unfolded chain. Thus, in some ways, there is an additive advantage with using this methodology.

Therefore, the goal of this dissertation research project are the following: to apply the denatured state His-heme loop formation method to proteins not containing a c-type heme, to evaluate the denatured state conformational bias in divergent amino acid sequences that encode the same overall fold, and to apply His-heme loop formation studies to an intrinsically disordered amino acid sequence that folds to α-helical topology when bound to its signaling partner. In chapter 2, we successfully apply the His-heme loop formation methodology to the Ubiquitin-associated domain 2, UBA(2), from the human DNA excision repair protein HHR23A. We discover that biases towards a specific reverse turn segment of primary structure in the UBA(2) domain remain in the denatured state. We then compared this finding to studies on a related UBA domain, the first UBA domain that resides in the internal portion of HHR23A, the UBA(1) domain in chapter 3. The amino acid sequences are divergent and share ~20% of amino acid identity, but have similar three-helix bundle topology. Both UBA domains possess a common hydrophobic surface patch for biological function. We observe that both reverse turns in the UBA(1) domain are persistent in strong concentrations of increasing guanidine hydrochloride (GuHCl), and thus remain thermodynamically and kinetically biased in the denatured state ensemble (DSE). While the UBA(2) domain shows DSE conformational bias towards the second reverse turn, the UBA(1) domain shows DSE persistent structure with the first and second reverse turn regions. This indicates that the denatured state does not need to be entirely biased towards both reverse turns to fold to its native conformer. Therefore, denatured state conformational biases towards reverse turn regions remain in the DSE for these three-helix bundles with low amino acid sequence identity. To test the role of residual structure in a disordered protein sequence that forms three-helix bundle topology when bound to its common signaling partner, we discovered in chapter 4 that the mouse cAMP-responsive element binding

protein (CBP), an intrinsically disordered protein that serves as a general transcription coactivator involved in hormonal signaling, exhibits unusual thermodynamic and kinetic properties in the denatured state. This strange behavior in the denatured state is vastly different from the foldable UBA domains that encode three-helix bundles. We do not observe necessarily the biases towards reverse turns in the denatured state. Rather, the disordered CBP amino acid sequence has a majority of its chain expanded yet containing biophysical properties consistent with a folded protein.

Chapter 2: Probing Denatured State Conformational Bias in the Three Helix Bundle, UBA(2), Using a Cytochrome *c* Fusion Protein

Probing Denatured State Conformational Bias in the Three Helix Bundle, UBA(2), Using a Cytochrome *c* Fusion Protein[a]

Moses J. Leavens, Melisa M. Cherney, Michaela L. Finnegan and Bruce E. Bowler*

Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Dynamics, University of Montana, Missoula, Montana 59812

*Leavens MJ, Cherney MM, Finnegan ML, and Bowler BE (2018). Probing Denatured State Conformational Bias in a Three-Helix Bundle, UBA(2), Using a Cytochrome *c* Fusion Protein. *Biochemistry*, 2018, *57* (11), pp 1711–1721. DOI: 10.1021/acs.biochem.8b00015

ABSTRACT:

Previous work with the four-helix-bundle protein cytochrome *c′* from *Rhodopseudomonas palustris* using histidine-heme loop formation methods revealed fold-specific deviations from random coil behavior in its denatured state ensemble. To examine the generality of this finding, we extend this work to a three-helix-bundle polypeptide, the second ubiquitin-associated domain, UBA(2), of the human DNA excision repair protein. We use yeast iso-1-cytochrome *c* as a scaffold, fusing the UBA(2) domain at the N-terminus of iso-1-cytochrome *c*. We have engineered histidine into highly solvent accessible positions of UBA(2), creating six single histidine variants. Guanidine hydrochloride denaturation studies show that the UBA(2)-cytochrome *c* fusion protein unfolds in a 3-state process with iso-1-cytochrome *c* unfolding first. Furthermore, engineered histidine residues in UBA(2) strongly destabilize the iso-1-cytochrome *c* domain. Equilibrium and kinetic histidine-heme loop formation measurements in the denatured state at 4 and 6 M guanidine hydrochloride show that loop stability decreases as the size of the histidine-heme loop increases in accord with the Jacobson-Stockmayer equation. However, we observe that the His27-heme loop is both more stable than expected from the Jacobson-Stockmayer relationship and breaks more slowly than expected. These results show that the sequence near His27, which is in the reverse turn between helices two and three of UBA(2), is prone to persistent interactions in the denatured state. Therefore, consistent with our results for cytochrome *c′*, this reverse turn sequence may help to establish the topology of this fold by biasing the conformational distribution of the denatured state.

INTRODUCTION

Residual structure in the denatured state ensemble (DSE) has been observed in many proteins.[1-4] Long range interactions and pockets of hydrophobic clustering have been observed in the DSE, and it is thought that these interactions in the DSE are important in limiting the conformational search and directing the collapse of a protein chain to its native conformer.[5] Some of the early evidence of residual structure stems from NMR studies on staphylococcal nuclease,[6-8] hen lysozyme,[9-11] and the 434-repressor protein[12] in denaturing concentrations of urea. In hen lysozyme, a single mutation from tryptophan to glycine located between two domains in the protein significantly decreased residual structure in the denatured state in 8 M urea, consistent with long range interactions occurring in its DSE.[10] The 434-repressor protein was shown to contain residual structure in 8 M urea localized to a cluster of nonpolar residues running from residues 55-60.[12] Furthermore, it has been shown for several proteins that amino acid substitutions can have a significant effect on the stability of the denatured state,[2, 4] whether through reverse hydrophobic effects,[13-16] or via electrostatic interactions.[17-24] Stabilization of residual structure can in turn affect the conformational distribution in the DSE and thus could bias the conformational search for the native state increasing folding efficiency.[5] Some studies show that residual structure in the denatured state primarily affects the unfolding rather than the folding rate.[19, 25, 26] However, conformational biases have long been viewed to provide nucleation sites for folding a protein.[27] Thus, understanding the nature of conformational biases in the denatured state for different folds can provide important insight into how primary structure evolves to favor a particular fold.

Loop formation probability along the primary structure of a protein in the denatured state can track the propensity for intrinsic conformational bias of a protein sequence. Deviations from

random coil behavior can be detected through comparison with the predictions of the Jacobson-Stockmayer equation (eq. 2-1), which assumes loop formation in a random coil is entirely entropic.[28]

$$(2-1)\ \Delta S_{\mathrm{loop}} = -\nu_3 \mathrm{R} ln(n) + \mathrm{R} ln((3/2\pi C_{\mathrm{n}} l^2)^{\nu_3} \mathrm{V_i})$$

In eq.2-1, ν_3 is the scaling exponent for loop formation, R is the gas constant, *n* is the number of monomers in the loop, C_{n} is Flory's characteristic ratio, *l* is the distance between the monomers forming the loop, and V_i is the approach volume within which the atoms must be constrained for loop formation to occur. Our previous work using histidine-heme (His-heme) loop formation thermodynamic and kinetic methods has shown that foldable protein sequences behave differently than low complexity polypeptide sequences.[29] Some loops are more favored to form than expected under denaturing conditions. Stopped-flow loop breakage kinetics show that loops which are more stable than expected based on the expectations of the Jacobson-Stockmayer equation also have slow breakage rates, indicating that these segments of the sequence are prone to stabilizing interactions. In particular, observations in the DSE of the four-helix bundle protein cytochrome *c′* (Cyt*c′*) from *Rhodopseudomonas palustris*, using His-heme loop formation methods, have revealed deviations from random coil behavior along the primary structure of this four-helix bundle.[30, 31] Segments of the primary structure, which formed more stable than expected denatured state His-heme loops and have slow loop breakage rates, localized near the reverse turns in the tertiary structure of the protein.[31] Molecular dynamics simulations of the DSE of Cyt*c′* showed that these reverse turns were stabilized by dynamic hydrophobic clusters.[31] Thus, residues bracketing reverse turns make persistent contacts in the DSE of this four-helix bundle, which may help to establish the native fold topology in the denatured state.

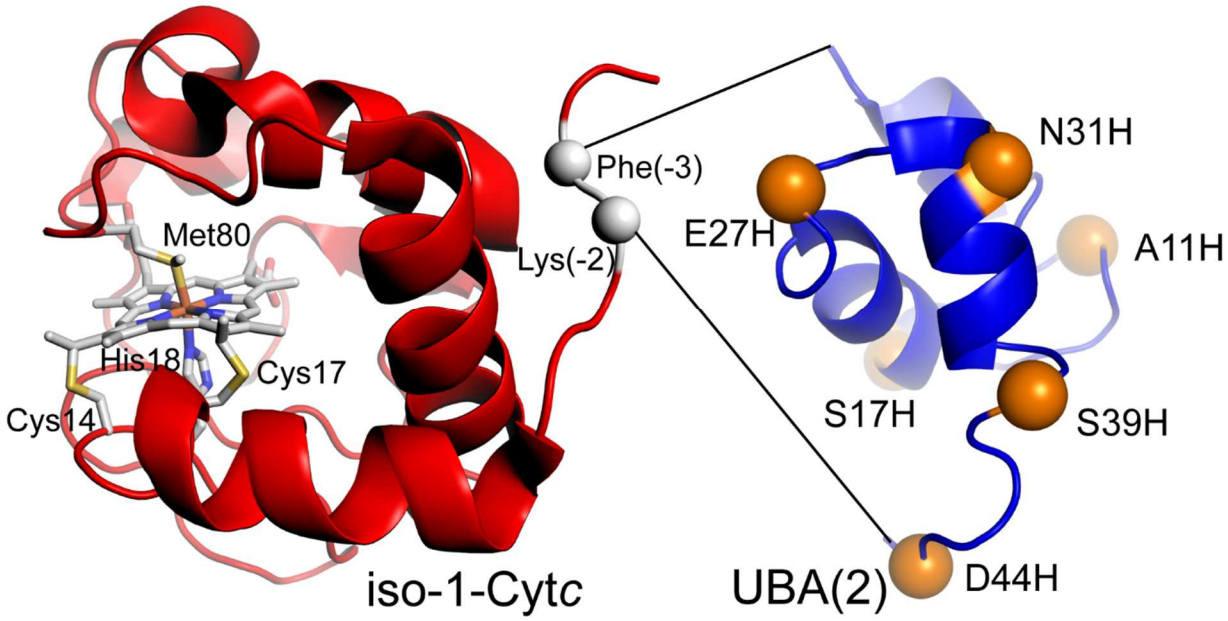

Figure 2-1. Structures of yeast iso-1 cytochrome *c* (red, PDB ID: 2YCC) and the human HHR23A UBA(2) domain (blue, PDB: 1DV0), showing insertion of the UBA(2) domain between Phe(-3) and Lys(-2) (white spheres) near the N-terminus of iso-1-Cyt *c*. Sites of single histidine substitutions are depicted with orange spheres in the UBA(2) domain. The heme and the amino acids that attach the heme to the iso-1-Cyt*c* polypeptide, Cys14, Cys17, His18 and Met80 are shown as stick models colored by element.

Previously, we have only been able to apply the His-heme loop formation method, as a means of characterizing denatured state conformational properties, to proteins that have a naturally occurring *c*-type heme. In the current work, we have developed a version of the pRbs_BTR1 plasmid,[32] pRbs_BTR1fuse, which allows yeast iso-1-Cytochrome *c* (iso-1-Cyt*c*) fusion proteins to be readily generated. We have used this plasmid to fuse a three-helix bundle domain, the human DNA excision repair protein's second Ubiquitin-associated domain, UBA(2), to the N-terminus of iso-1-Cyt*c* (UBA(2) – iso-1-Cyt*c*, Figure 2-1). Assessing His-heme loop equilibria in the denatured state, we observe a strong deviation from random coil behavior in a segment of the primary structure that localizes around turn between helix 2 and helix 3 of

UBA(2). Stopped-flow loop breakage kinetics yield a small His-heme loop breakage rate constant, k_b, for the histidine that probes this region of the primary structure of UBA(2). Thus, the reverse turn located between helices 2 and 3 may be important for establishing the topology of the native fold of UBA(2).

EXPERIMENTAL PROCEDURES

Preparation of the pRbs_BTR1fuse vector. The pRbs_BTR1 vector[32] was used as a starting point for preparation of the pRbs_BTR1fuse plasmid. The pRbs_BTR1 vector is a derivative of the pBTR1 vector[33, 34] with an optimized ribosomal binding site. pRbs_BTR1 co-expresses the genes for yeast iso-1-Cyt*c*, *CYC1*, and the yeast heme lyase, *CYC3*. The pRbs_BTR1 vector used to prepare pRbs_BTR1fuse carries five mutations in five codons relative to wild type iso-1-Cyt*c* leading to H26N, H33N, H39Q, K72A and C102S substitutions.[35] The first three mutations eliminate histidines from iso-1-Cyt*c* so that only the single histidines engineered into the fused protein can form His-heme loops in the denatured state.[36] The K72A mutation prevents Lys72 (trimethylated in *Saccharomyces cerevisiae*) from forming an alkaline conformer.[33] The C102S mutation prevents disulfide dimerization during physical studies. The iso-1-Cyt*c* gene has an *Eco*RI restriction site that includes the codons for Glu(-4) and Phe(-3). To make this site unique, two *Eco*RI sites in the pRbs_BTR1 vector needed to be eliminated (Figure 2-S1). The codons for Ala(-1) and Gly(1) can be converted to an *Ngo*MIV restriction site with a T→C point mutation in the Gly(1) codon (Figure 2-S1). To make this restriction site unique, an *Ngo*MIV restriction site at sequence position 2474 of pRbs_BTR1 must be removed (Figure 2-S1). The sequences for oligonucleotides used for mutagenesis are provided in Table 2-S1. Mutations were introduced using PCR-based mutagenesis in the order in which the oligonucleotides are listed in Table 2-S1. Mutations were confirmed by DNA

sequencing (Genomics Core Facility, University of Montana). Restriction digests with *Eco*RI and *Ngo*MIV on the final pRbs_BTR1fuse vector confirmed the presence of a single *Eco*RI and a single *Ngo*MIV restriction site.

Preparation of the human HHR23A UBA(2) domain fused to the N-terminus of iso-1-Cyt*c*. The gene for the second ubiquitin associated domain, UBA(2), of the human homolog of yeast Rad23A, HHR23A (resides 319 – 363), cloned into the pGEX-2T vector was obtained from the Feigon laboratory at UCLA.[37] The UBA(2) gene was amplified by PCR using a forward primer which placed an *Eco*RI restriction site at the beginning of the gene and a reverse primer which placed Lys(-2) from iso-1-Cyt*c* and an *Ngo*MIV restriction site at the end of the gene (Table 2-S2). The PCR product was purified using a QIAquick PCR purification kit (Qiagen) and characterized by agarose gel electrophoresis. The PCR product was then double-digested with *Eco*RI and *Ngo*MIV restriction enzymes (New England Biolabs) and purified again with the QIAquick kit. The pRbs_BTR1fuse vector was digested first with the less efficient enzyme, *Ngo*MIV, the linearized DNA was purified by gel electrophoresis and then extracted from the gel using the QIAquick kit. Linearized vector was digested with *Eco*RI followed by purification with the QIAquick kit. The UBA(2) fragment was then ligated between the *Eco*RI and *Ngo*MIV sites of the pRbs_BTR1fuse vector using T4 ligase (New England Biolabs) to produce a plasmid carrying the gene for the UBA(2) – iso-1-Cyt*c* fusion protein, pRbs_BTR1(UBA2_Cc). The ligation reaction mixture was transformed into competent TG-1 *Escherichia coli* cells. DNA extracted from a single colony (Wizard Plus DNA miniprep kit, Promega) demonstrated the presence of the gene for the UBA(2) – iso-1-Cyt*c* fusion protein based on the cleavage pattern obtained from an *Eco*RI/*Hind*III double-digest. Sequencing revealed that an error in the reverse primer modified the three C-terminal residues of the UBA(2) domain. These were corrected

using PCR-based mutagenesis (primer sequences are provided in Table 2-S2) and confirmed by dideoxy sequencing of the entire UBA(2) – iso-1-Cyt*c* gene (Genomics Core Facility, University of Montana).

Preparation of UBA(2) – iso-1-Cyt*c* fusion protein variants. Primers used to mutate the pRbs_BTR1(UBA2-Cc) plasmid are listed in Table 2-S3. The UBA(2) domain inserted into the N-terminus of iso-1-Cyt*c* is 45 residues long. The UBA(2) domain has a cysteine at position 26. To avoid problems with disulfide dimerization, the pRbs_BTR1(UBA2_Cc) plasmid was mutated using the QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies) to produce a C26A substitution (pWT variant). The pRbs_BTR1(UBA2_Cc) plasmid carrying the gene for the pWT variant was used as a template to produce A11H, S17H, E27H, N31H, Q39H and D44H variants of UBA(2) – iso-1-Cyt*c* using the QuikChange Lightning kit. All sites for the introduction of histidines were chosen based on their high solvent accessibility using an accessible surface area algorithm from the Center for Informational Biology, Ochanomizu University (http://cib.cf.ocha.ac.jp/bitool/ASA/). The sequences of the gene for each variant were confirmed by dideoxy sequencing (Eurofins Genomics, Louisville, Kentucky).

Expression and Purification of Variants of UBA(2) – iso-1-Cyt*c*. pWT and single histidine variants were expressed from the pRbs_BTR1(UBA2_Cc) plasmid following transformation into ultra BL21 (DE3) *E. coli* competent cells (EdgeBio, Gaithersburg, MD) using the manufacturers protocol. Sterile Fernbach flasks containing 1 L of 2xYT bacterial media were inoculated with 0.5 mL of cells from L-ampicillin plates, after suspending the colonies on the plates in 1.0 mL sterile L-broth. Subsequently, 1.0 mL of a 100 mg/mL ampicillin stock was added to each Fernbach flask. Flasks were placed in an orbital shaker at 150 rpm and 37 °C for

30 hours. Media was spun down at 5,000 rpm for 10 min (Sorvall Lynx 6000, F12 rotor). On average, a 1 L culture yielded 5.1 g of pelleted cells. Cell pellets were stored at -80 ^{o}C.

Protein extraction and purification were carried out as described, previously.[38-41] In brief, cell pellets were thawed and suspended in 50 mM Tris, 500 mM NaCl, 1 mM EDTA, pH 8.0, 500 μL of 100 mM PMSF was added and the cells were lysed using a Qsonica Q700 sonicator. Lysate was spun for 30 min at 10,000 rpm (Sorvall Lynx 6000, F14 rotor). The cleared lysate was adjusted to 8% ammonium sulfate by addition of solid $(NH_4)_2SO_4$ and allowed to equilibrate with stirring overnight at 4 ^{o}C. Protein was subsequently spun for 30 min at 10,000 rpm (Sorvall Lynx 6000, F14 rotor). After removing precipitated protein impurities, the supernatant was dialyzed against two changes of MilliQ water containing 1 mM Na_2EDTA and 1 mM β-mercaptoethanol (βME). A small amount of 50 mM sodium phosphate monobasic was used to adjust the dialysis solution to pH 6.0. After removal from the dialysis tubing, the diasylate was stirred with 100 mL of CM-sepharose (GE Healthcare Life Sciences) to bind UBA(2) – iso-1-Cyt*c* variants to the resin, which was subsequently loaded into a glass column. The CM-sepharose column was then was washed with MilliQ water to remove unbound material and then eluted using a 200 mL linear gradient (0 – 0.8 M NaCl in 50 mM sodium phosphate pH 6.0, 1 mM Na_2EDTA, 2 mM βME). The protein solution was then exchanged into 25 mM sodium phosphate, pH 6.0, 1 mM Na_2EDTA by centrifuge ultrafiltration. Protein was flash frozen in liquid N_2 and stored at -80 ^{o}C until use.

Prior to experiments, all UBA(2) – iso-1-Cyt*c* variants were purified using a HiTrap SP HP 5.0 mL column coupled to an ÄKTAprime plus chromatography system (GE Healthcare Life Sciences). Variants were subsequently oxidized with $K_3[Fe(CN)_6]$, separated from the oxidant,

and exchanged into a buffer appropriate for each experiment using Sephadex G25 (GE Healthcare Life Sciences, superfine grade) chromatography.

Purified fusion proteins were characterized by SDS-PAGE and matrix-assisted laser-desorption time-of-flight (MALDI-TOF) mass spectrometry (Bruker microflex mass spectrometer, Table 2-S4), both before and after experiments, to ensure that no degradation of the hybrid polypeptides occurred during experimentation.

Isothermal Equilibrium Unfolding of UBA(2) – iso-1-Cyt*c* Variants By Guanidine Hydrochloride. The pWT and six single histidine variants were oxidized using $K_3[Fe(CN)_6]$, followed by separation from the oxidizing agent using Sephadex G-25 chromatography and with CD buffer (20 mM Tris, 40 mM NaCl, 1 mM Na_2EDTA, pH 7.0) as the chromatography buffer. Protein stock concentration was subsequently calculated using five absorbance values at 339 nm (20.9×10^3 M^{-1} cm^{-1}), 526.5 nm (11.0×10^3 M^{-1} cm^{-1}), 541.8 nm (9.0×10^3 M^{-1} cm^{-1}), 550 nm (oxidized, 9.1×10^3 M^{-1} cm^{-1}), 550 nm (reduced 28.0×10^3 M^{-1} cm^{-1}).[42] Refractive indices of CD buffer and 6 M guanidine hydrochloride (GuHCl) in CD buffer were measured using a Fisher Scientific refractometer. The GuHCl concentration was evaluated with the empirical equation of Nozaki for the dependence of refractive index on GuHCl concentration.[43] A 4 μM solution of the UBA(2) – iso-1-Cyt*c* in concentrated GuHCl was titrated into a 4 μM solution of UBA(2) – iso-1-Cyt*c* in CD buffer using a Hamilton Microlab 500 Titrator coupled to an Applied Photophysics Chirascan CD Spectrophotometer. The change in ellipticity was monitored at 222 nm using 250 nm as background, $\theta_{222corr}$ (= θ_{222nm} - θ_{250nm}). $\theta_{222corr}$ was plotted versus GuHCl concentration and fit to eq 2-2, which assumes three state unfolding and a linear dependence of the free energies

$$(2-2)\ \theta_{222corr} = \theta_D + m_D[GuHCl] + \frac{\theta_N + m_N[GuHCl] - \theta_D - m_D[GuHCl] + \left[(\theta_I - \theta_D - m_D[GuHCl])e^{\frac{m_{NI}[GuHCl] - \Delta G_{NI}^{o\prime}(H2O)}{RT}}\right]}{\left[1 + \left(e^{\frac{m_{NI}[GuHCl] - \Delta G_{NI}^{o\prime}(H2O)}{RT}}\right)\left(1 + e^{\frac{m_{ID}[GuHCl] - \Delta G_{ID}^{o\prime}(H2O)}{RT}}\right)\right]}$$

of unfolding, ΔG_u, on GuHCl concentration.[44, 45] In eq 2-2, θ_N and m_N are the intercept and the slope of the native state baseline, θ_I is the intermediate state baseline, θ_D and m_D are the intercept and the slope of the denatured state baseline, m_{NI} is the rate of change of the free energy of the native to intermediate state transition (ΔG_{NI}) with respect to GuHCl concentration, m_{ID} is the rate of change of the free energy of the of intermediate to denatured state transition (ΔG_{ID}) with respect to GuHCl concentration and $\Delta G_{NI}^{o'}(H_2O)$ and $\Delta G_{ID}^{o'}(H_2O)$ are the free energies of the native to intermediate and intermediate to denatured state transitions, respectively, in the absence of denaturant. Single histidine variants were fit assuming that the native baseline does not vary with GuHCl concentration because the native baselines were too short to allow reliable evaluation of m_N. Reported parameters are the average and standard deviation of at least three independent trials.

His-Heme Loop Formation Equilibria by Denatured State pH Titration. Histidine-heme loop equilibria were monitored using UV-Vis spectroscopy with a Beckman DU 800 spectrophotometer for all UBA(2) – iso-1-Cyt*c* variants. Denatured state pH titrations were monitored at room temperature (22 ± 1 °C) with 3 μM denatured protein in 15 mM NaCl, 5 mM Na_2HPO_4, 1 mM Na_2EDTA (1x buffer) at the specified GuHCl concentration. The exact concentration of an 8 M GuHCl solution in MilliQ water was determined by refractive index measurements as described above. Titrations were carried out as previously reported.[46] Briefly, pH was changed by mixing 18 μM protein in 6x buffer with 8 M GuHCl and HCl or NaOH

solutions in the appropriate proportions to keep protein concentration at 3 μM, GuHCl at the specified concentration and to change pH by ~ 0.2 units. At each pH, absorbance spectra were acquired from 350 – 450 nm. The absorbance band at 398 nm, A_{398}, using absorbance at 450 nm, A_{450}, as a baseline, $A_{398corr}$ (= $A_{398} - A_{450}$) was plotted versus pH. Plots of $A_{398corr}$ versus pH were used to determine the apparent pK_a, pK_a(obs), by fitting the data to eq 2-3, a modified form of the

$$(2-3)\ \ A_{398\mathrm{corr}} = \frac{A_{LS} + A_{HS} \times 10^{n_p[pK_a(obs) - pH]}}{1 + 10^{n_p[pK_a(obs) - pH]}}$$

Henderson-Hasselbalch equation that evaluates the number of protons, n_p, linked to the His-heme loop formation. In eq 2-3, A_{LS} is $A_{398corr}$ at high pH when the heme is low spin with histidine bound to the heme iron and A_{HS} is $A_{398corr}$ at low pH when the heme is high spin with H_2O bound to the heme iron.

Stopped-Flow Loop Breakage Measurements. Loop breakage measurements were performed with pH jump stopped-flow experiments using an Applied Photophysics SX20 stopped-flow apparatus at 25 °C. For all UBA(2) – iso-1-Cyt*c* variants, the transition from low spin heme (i.e. high pH) to high spin heme (i.e. low pH) was monitored at 398 nm. The UBA(2) – iso-1-Cyt*c* variants were oxidized with $K_3[Fe(CN)_6]$, followed by separation from the oxidizing agent and transfer into MOPS buffer (10 mM MOPS, 40 mM NaCl, 2 mM Na_2EDTA, pH 6.8) using Sephadex G-25 chromatography. Protein stock concentration was determined as described above. The exact concentrations of 6 M or 8 M GuHCl stocks in MOPS buffer pH 6.8 were determined as described above. Denatured 6 μM protein solutions were prepared at 4 M or 6 M GuHCl in 10 mM MOPS, 2 mM Na_2EDTA, pH 6.8 for downward pH jump experiments. For stopped-flow experiments, these solutions were mixed 1:1 with 4 M or 6 M GuHCl, respectively, in 100 mM citrate at pH 3 or pH 3.5 to yield a final protein concentration of 3 μM.

For the pWT variant, the starting pH for downward pH jump experiments was 8 (20 mM Tris, 2 mM Na_2EDTA, pH 8.0). Samples were obtained from the stopped-flow waste line and pH was measured directly after mixing experiments to determine the final pH. The dead time of the stopped-flow was measured by reduction of dichlorophenolindophenol as a function of L-ascorbic acid concentration and found to be 2 ms under our mixing conditions.[47] Data were fit to a single exponential equation after adding 0.002 s to the time at each data point.

RESULTS

Isothermal Equilibrium Unfolding of UBA(2) – iso-1-Cyt*c* Variants by Guanidine Hydrochloride. Global unfolding thermodynamics of pWT and the six single histidine variants of UBA(2) – iso-1-Cyt*c* were monitored by CD at 25 °C and pH 7.0 using GuHCl as denaturant. Figure 2-2 compares the denaturation curves, $\theta_{222corr}$ versus GuHCl concentration, for the pWT

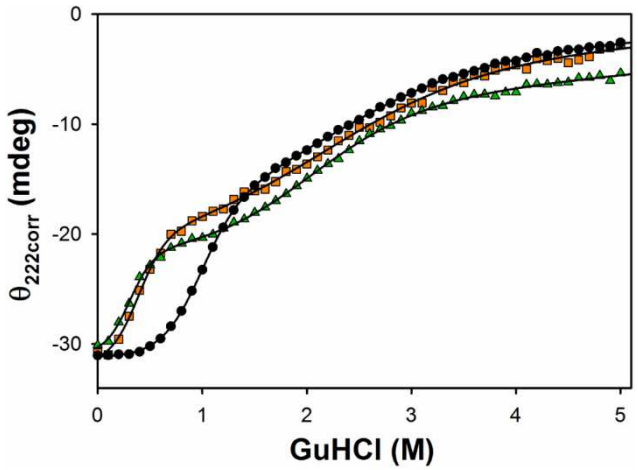

Figure 2-2. GuHCl denaturation curves for UBA(2) – iso-1-Cyt*c* variants, pWT (black circles), S39H, (green triangles) and A11H (orange squares) depicted as plots of corrected ellipticity at 222 nm, $\theta_{222corr}$, versus GuHCl concentration. Solid curves are fits to eq 2-2 in Experimental Procedures. Parameters obtained from the fits are given in Table 2-1. Experiments were performed at 25 °C and pH 7.

variant and two of the six single histidine variants, S39H and A11H. All UBA(2) – iso-1-Cyt*c* fusion proteins show similar 3-state equilibrium unfolding titration curves. To fit the pWT and six single histidine variant data, we used a 3-state equilibrium unfolding model (eq 2-2, Experimental Procedures) to extract the thermodynamic parameters listed in Table 2-1. We were able to assign the thermodynamic parameters to unfolding of the UBA(2) and iso-1-Cyt*c* domains based on previous isothermal equilibrium unfolding experiments with GuHCl as

Table 2-1. Thermodynamic Parameters for GuHCl Unfolding of UBA(2) – iso-1-Cyt*c* at 25 °C and pH 7.0.

UBA(2) – iso-1-Cyt*c* Variants[a]	m_{NI}, kcal mol^{-1} M^{-1}	$\Delta G_{NI}^{o'}(H_2O)$, kcal mol^{-1}	C_{mNI}, M	m_{ID}, kcal mol^{-1} M^{-1}	$\Delta G_{ID}^{o'}(H_2O)$, kcal mol^{-1}	C_{mID}, M
pWT	2.89 ± 0.16	2.97 ± 0.14	1.02 ± 0.02	0.86 ± 0.14	1.86 ± 0.32	2.14 ± 0.05
D44H (18)	4.45 ± 0.28	1.60 ± 0.17	0.36 ± 0.02	1.15 ± 0.05	2.10 ± 0.10	1.80 ± 0.05
S39H (23)	3.9 ± 1.1	1.06 ± 0.41	0.26 ± 0.03	1.17 ± 0.06	2.37 ± 0.18	2.02 ± 0.08
N31H (31)	2.94 ± 0.32	1.02 ± 0.24	0.34 ± 0.05	1.11 ± 0.03	2.22 ± 0.09	1.99 ± 0.01
E27H (35)	3.6 ± 0.5	1.61 ± 0.29	0.44 ± 0.03	0.58 ± 0.20	1.18 ± 0.36	2.07 ± 0.17
S17H (45)	4.3 ± 1.4	2.4 ± 0.8	0.56 ± 0.02	0.8 ± 0.5	1.6 ± 1.0	2.1 ± 0.7
A11H (51)	4.32 ± 0.3	1.54 ± 0.13	0.35 ± 0.03	0.66 ± 0.15	1.6 ± 0.17	2.5 ± 0.5

[a]Denatured state His-heme loop size is given in brackets.

denaturant.[32, 48] Iso-1-Cyt*c* with all histidines except His18 eliminated has an unfolding midpoint near 1 M GuHCl and an *m*-value of 3.8 to 4.0 kcal/mol-M.[32] The *m*-value reported for GuHCl denaturation of UBA(2) is ~1.1 kcal/mol-M.[48] Thus, GuHCl unfolding of the iso-1-Cyt*c* domain precedes unfolding of the UBA(2) domain (Figure 2-2, Table 2-1). For the iso-1-Cyt*c* domain, the denaturant *m*-value for the N→I transition, m_{NI}, increased significantly, relative to the pWT variant, when surface histidines were introduced into the UBA(2) domain, except for

the N31H variant. All single histidine substitutions cause a decrease in the native to intermediate unfolding midpoint, C_{mNI}, by 0.5 – 0.7 M and the free energy of unfolding in the absence of denaturant, $\Delta G_{NI}^{o'}(H_2O)$ by 1 – 2 kcal/mol. Therefore, the presence of a histidine in the UBA(2) domain destabilizes the iso-1- Cyt*c* domain.

The single histidine mutations in the UBA(2) domain have a more modest effect on the stability of the UBA(2) domain itself. For our control experiment with the pWT variant, the intermediate to denatured unfolding midpoint, C_{mID}, is 2.14 M, while the C_{mID} values for all six single histidine variants are in the range 1.8 M – 2.5 M. The *m*-value for GuHCl unfolding of the UBA(2) domain, m_{ID}, increases for some variants and decreases for others relative to pWT. The net effect is that the free energy of unfolding in the absence of denaturant for the UBA(2) domain, $\Delta G_{ID}^{o'}(H_2O)$, except for the E27H variant, is not strongly affected by the single histidine substitutions to the UBA(2) domain.

The parameters in Table 2-2 were used to calculate the population of denatured protein for each variant at 4 M GuHCl and 6 M GuHCl. All variants (including the pWT variant) are greater than 95% unfolded at 6 M GuHCl. At 4 M GuHCl, all variants including pWT are ~95% unfolded, with the exception of the E27H and A11H variants, which are both 83 – 84% unfolded.

Equilibrium Denatured State His-Heme Loop Formation with UBA(2) – iso-1-Cyt*c* Variants. We use His-heme loop formation equilibria to interrogate the denatured state thermodynamic properties along the primary sequence of the three-helix bundle, UBA(2). This methodology requires a heme. To extend the His-heme loop formation technique to UBA(2), which does not have a *c*-type heme, UBA(2) has been fused near the N-terminus of iso-1-Cyt*c* (Figure 2-1). In the denatured state, single histidines engineered into the denatured state can form a loop of a known size. Furthermore, histidine residues native to iso-1-Cyt*c* have been replaced

(except the heme ligand of the CXXCH heme attachment sequence, His18) to eliminate competition with the engineered histidine residues in the three-helix bundle, UBA(2). The closest point of attachment of the iso-1-Cyt*c* polypeptide to the fused UBA(2) domain is Cys14 (see Figure 2-1). Thus, each loop contains 16 residues from the iso-1-Cyt*c* sequence (insertion of the domain is between Lys(-2) and Phe(-3), Figure 2-1). For loop sizes less than 16, chain stiffness effects start to become important.[49] So, for evaluating the adherence of the UBA(2) sequence to random coil properties in the denatured state, insertion between Lys(-2) and Phe(-3) is optimal. This insertion site was also used for previous work from this lab on loop formation in the denatured state with homopolymeric amino acid sequences,[29, 50-52] allowing for ready comparison to these data.

In the denatured state, Met80, a weak ligand for Fe^{3+},[53] will be displaced by water, allowing the engineered histidine (or a lysine for the pWT variant[54]) to bind to the Fe^{3+} of the heme. Because histidine is a side chain that can be readily protonated, the relative stability of a His-heme loop can be measured using a denatured state pH titration, to produce an apparent pK_a, pK_a(obs), and n_p, the number of protons involved in the process (Figure 2-3A).

Figure 2-3B shows denatured state pH titration curves for the pWT, E27H, and the D44H variants of UBA(2) – iso-1-Cyt*c* in 6 M GuHCl. At pH 7, the His-heme loops are formed (pH 9 for the Lys-heme loop of pWT) and the Fe^{3+} in the heme is in the low spin state. As pH decreases, a blue shift occurs as the Fe^{3+}-heme transitions from a low (strong field ligand, His or Lys) to a high spin state (weak field ligand, water). Denatured state titration curves are fit to a modified form of the Henderson-Hasselbach equation (eq 2-3 in Experimental Procedures), to determine the relative stability of the loops, pK_a(obs), and the number of protons, n_p, involved in loop formation (Table 2-2). The pWT variant has pK_a(obs) near 7 at both 4 and 6 M GuHCl,

setting an upper limit for determination of the pK_a(obs) for His-heme loop formation in the denatured state. For the single histidine variants, pK_a(obs) is lowest (most stable His-heme loop)

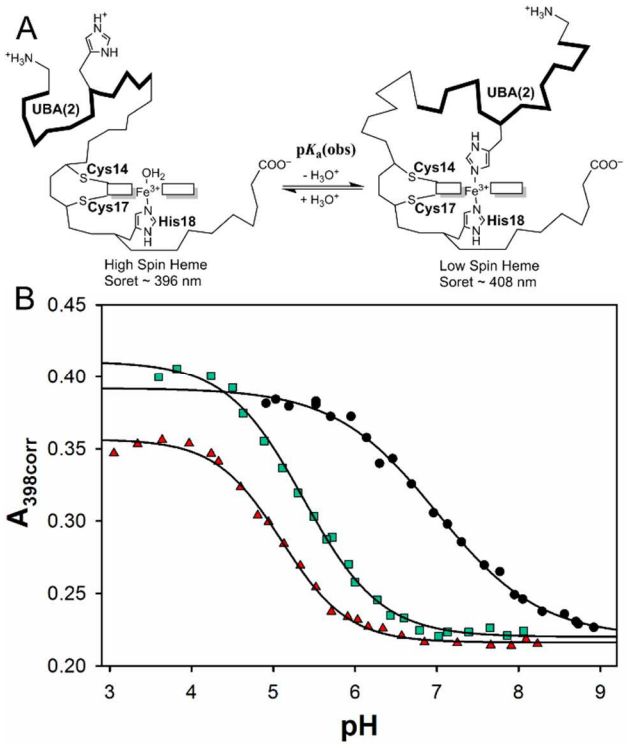

Figure 2-3. (A) Schematic representation of His-heme loop formation with the UBA(2) – iso-1-Cyt*c* fusion protein. (B) Representative curves for denatured state His-heme loop formation equilibria for UBA(2) – iso-1-Cyt*c* variants at 6 M GuHCl. Plots of $A_{398corr}$ versus pH for the pWT, D44H and E27H variants are shown with black circles, red triangles and green squares, respectively. Data were fit to eq 2-3 in Experimental Procedures (solid curves). Thermodynamic parameters obtained from the fits are given in Table 2-2. Experiments were performed at 22 ± 1 °C.

for the shortest loop (D44H variant, loop size, *n*, of 18) at both 4 M and 6 M GuHCl. In general, as loop size increases, pK_a(obs) increases. As expected from the requirement that a single proton

must dissociate from histidine for the loop to form (Figure 2-3A), n_p is also near 1 (Table 2-2). Thus, the longer His-heme loops are less stable under both denaturing conditions.

At 6 M GuHCl, the value of n_p falls to ~0.8 for the A11H variant. Broadening of the pH titration curve for longer loops typically indicates that the His-heme loop has become sufficiently long that it is no longer stable enough to complete the spin state transition. Lys-heme binding at higher pH values is required to complete the spin-state transition. For iso-1-Cyt*c*, biphasic spin state transitions were observed for loop sizes of 56, 72 and 83 in 6 M GuHCl.[46] For cytochrome *c*′, loops sizes of 48 – 111 all lead to biphasic loop formation equilibria at 6 M GuHCl.[31] Thus,

Table 2-2. Thermodynamic Parameters for Denatured State Loop Formation of UBA(2) – iso-1-Cyt*c* Variants in 6 M and 4 M GuHCl Solutions at 22 ± 1˚C

Variant	Loop size (n)	pK_a(obs)	n_p	pK_{loop}(His)
		6 M GuHCl		
pWT	-	7.01 ± 0.07	0.72 ± 0.06	-
D44H	18	5.11 ± 0.02	1.04 ± 0.05	-1.48 ± 0.02
S39H	23	5.42 ± 0.01	1.07 ± 0.02	-1.17 ± 0.01
N31H	31	5.65 ± 0.02	1.02 ± 0.19	-0.95 ± 0.02
E27H	35	5.39 ± 0.08	1.05 ± 0.13	-1.20 ± 0.08
S17H	45	5.86 ± 0.03	0.95 ± 0.09	-0.73 ± 0.03
A11H	51	6.35 ± 0.09	0.79 ± 0.07	-0.24 ± 0.09 (-0.39 ± 0.17)[a]
		4 M GuHCl		
pWT	-	7.04 ± 0.06	0.77 ± 0.01	-
D44H	18	4.63 ± 0.09	1.02 ± 0.07	-1.96 ± 0.09
S39H	23	5.09 ± 0.06	1.20 ± 0.08	-1.51 ± 0.06
N31H	31	5.56 ± 0.04	1.15 ± 0.11	-1.04 ± 0.04
E27H	35	5.14 ± 0.02	1.06 ± 0.08	-1.45 ± 0.02
S17H	45	5.61 ± 0.06	0.99 ± 0.03	-0.99 ± 0.06

A11H	51	5.92 ± 0.03	1.15 ± 0.15	-0.67 ± 0.03

[a]From a fit to eq 2-4. pK_a(HisH$^+$) = 6.76 ± 0.14, and pK_{loop}(Lys) = -3.1 ± 0.7 were the other parameters obtained from the fit of eq 2-4 to the A11H titration data. pK_a(LysH$^+$) was set to 10.5 for these fits.

this behavior is expected for the A11H variant with a loop size of 51 at 6 M GuHCl. At 4 M GuHCl, only loop sizes greater than 70 show evidence for biphasic loop formation.[31, 46] Thus, the observation that the A11H variant fits well to eq 2-3 with $n_p \sim 1$ at 4 M GuHCl is also consistent with previous results. To extract a more accurate loop stability, A11H titration curves at 6 M GuHCl were fit to eq 2-4.[31, 46] In eq 2-4, pK_{loop}(His) and pK_{loop}(Lys) are the pK values for the binding

$$A_{398orr} = \frac{A_{HS} + A_{LS}\left(\left(\frac{10^{-pK_{loop}(His)}}{1+10^{pK_a(HisH^+)-pH}}\right)+\left(\frac{10^{-pK_{loop}(Lys)}}{1+10^{pK_a(LysH^+)-pH}}\right)\right)}{1+\left(\frac{10^{-pK_{loop}(His)}}{1+10^{pK_a(HisH^+)-pH}}\right)+\left(\frac{10^{-pK_{loop}(Lys)}}{1+10^{pK_a(LysH^+)-pH}}\right)} \quad (2\text{-}4)$$

of a fully deprotonated histidine or lysine to the heme, respectively, and pK_a(HisH$^+$) and pK_a(LysH$^+$) are the acid dissociation constants for the histidine and the lysine, respectively. The other parameters are as defined in eq 2-3 (Experimental procedures). The parameters from the fit of the A11H data to this equation are provided in Table 2-1. The pK_a(HisH$^+$) of 6.76 ± 0.14 is consistent with the value of 6.6 ± 0.1 obtained for iso-1-Cyt*c* at GuHCl concentrations from 3 to 6 M.[46]

The His-heme loop formation equilibrium can be broken down into a two-step process, ionization of the histidine followed by binding of the deprotonated histidine to the heme. This two-step process can be represented by eq 2-5.

(2-5) $$pK_a(obs) = pK_a(HisH^+) + pK_{loop}(His)$$

Therefore, subtracting $pK_a(HisH^+)$ (= 6.6)[46] from each $pK_a(obs)$ value in Table 2-2 allows calculation of $pK_{loop}(His)$ for every loop size (Table 2-2). In general, $pK_{loop}(His)$ becomes less negative (less stable loop) as loop size increases.

His-Heme Loop Breakage Kinetic Measurements in the Denatured State of UBA(2) – iso-1-Cyt*c* Variants. Previous work has indicated that the kinetics of His-heme loop breakage and formation adheres to a model involving rapid deprotonation of histidine followed by the binding of the ionized histidine to the heme.[55] In this way, the observed rate constant, k_{obs}, is pH dependent and is governed by eq 2-6,[46] where k_b and k_f are the rate constants for loop breakage and

(2-6) $$k_{obs} = k_b + k_f\left(\frac{K_a(HisH^+)}{K_a(HisH^+) + [H^+]}\right)$$

loop formation, respectively, and $K_a(HisH^+)$ is the dissociation constant for the deprotonation of histidine. If the pH is much lower than $pK_a(HisH^+)$, then $k_{obs} \approx k_b$. Histidine-heme loop breakage kinetics (and Lysine-heme loop breakage kinetics for the pWT variant) at 6 M and 4 M GuHCl solutions were measured by pH jump stopped-flow methods at 25 °C. Loop breakage follows single exponential kinetics (Figure 2-S2). Measurements were made using downward pH jumps to both pH 3.5 and 3.0 to confirm that a pH regime where $k_{obs} \approx k_b$ had been attained (Table 2-S5). k_{obs} values were similar at both pH values implying that k_{obs} had reached the lower limit of $k_{obs} \approx k_b$.

The k_b data for His-heme loops forming between the residues in UBA(2) and the heme of iso-1-Cyt*c* in 6 M and 4 M GuHCl solutions are shown in Figure 2-4. The magnitude of k_b varies in

an irregular manner versus loop size. We have previously shown for polyalanine loops inserted at the N-terminus of iso-1-Cyt*c*, that as the size of the loop increases, k_b decreases then plateaus at a constant value as would be expected for loops exhibiting random coil behavior.[29] Thus, the variation in k_b observed for the single histidine variants of UBA(2) – iso-1-Cyt*c* is indicative of nonrandom behavior in the denatured state along the sequence of UBA(2). The pattern of variation of k_b is similar at both 4 M and 6 M GuHCl denaturing conditions. The kinetic data at 4 M GuHCl in Figure 2-4 are corrected for the viscosity difference between 4 M and 6 M GuHCl (Table 2-S5), as described previously,[29, 31] except 6 M GuHCl was used as the reference state, here. While all of our loops show a decrease in k_b at 4 M versus 6 M GuHCl, some loops have more of a significant decrease in k_b. At 6M GuHCl, the E27H variant has the smallest k_b value (~46 s^{-1}). At 4 M GuHCl the loop breakage rate constant for the E27H variant corrected for viscosity drops to ~29 s^{-1}.

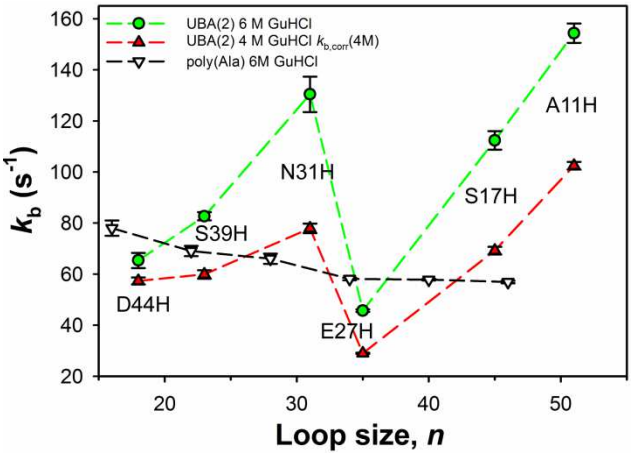

Figure 2-4. Loop breakage kinetics versus loop size, *n*, at 25 °C for all UBA(2) – iso-1-Cyt*c* variants in 6 M and 4 M GuHCl. The magnitudes of k_b at 4 M GuHCl have been adjusted for the difference in viscosity between 4 and 6 M GuHCl, $k_{b,corr}$(4M). Data from homopolymeric inserts

of alanine $(KAAAAA)_n$[29] (n = 1 – 5) inserted between Phe(-3) and Lys(-2) of iso-1-Cyt*c* as for the UBA(2) domain (see Figure 2-1) acquired at 6 M GuHCl are included for comparison.

DISCUSSION

Effect of Single Histidine Substitutions in the UBA(2) domain on the Stability of the UBA(2) – iso-1-Cyt*c* Fusion Protein. Previous global stability experiments using GuHCl on the individual iso-1-Cyt*c* and UBA(2) domains allowed assignment of the first phase of unfolding to the iso-1-Cyt*c* domain and the second phase to the UBA(2) domain (see Results). The pWT yields C_{mNI} = 1.02 for the iso-1-Cyt*c* domain, while all variants with single histidines engineered into the UBA(2) domain have C_m values between 0.26 and 0.56 (Figure 2-2, Table 2-1). Therefore, histidine residues placed at solvent accessible surface positions of the UBA(2) domain strongly destabilize the iso-1-Cyt*c* domain. The effect of these single histidine mutations on the overall global stability of the UBA(2) domain appears to be more modest. The C_m for UBA(2) domain of the pWT variant C_m is 2.14 M. For all single histidine variants, C_m is similar to that of pWT or within error of the pWT value. The resulting values for $\Delta G_{ID}^{o\prime}(H_2O)$ for UBA(2) domain of the pWT variant and the single-histidine variants are similar, too. The sole exception is the E27H variant, which has a significantly reduced m_{ID} value relative to the pWT variant. Even though histidine has a low helical propensity,[56] the substitution sites are either outside of the helical regions or near the ends of helices (A11H, S17H and N31H variants, see Figure 2-1), where the effect of lower helical propensity on helix stability will be smaller.[57]

The observation that the presence of a single histidine in the UBA(2) domain destabilizes the iso-1-Cyt*c* domain (relative to pWT) suggests that unfolding of the iso-1-Cyt*c* domain is driven

by His-heme binding from the histidine in the UBA(2) domain. The C_{mID} for unfolding of the UBA(2) domain is about 1.5 M higher than C_{mNI} for the iso-1-Cyt*c* domain in the single histidine variants. Thus, the UBA(2) domain is predominately folded near the end of the first phase of the biphasic unfolding transitions. Evidently, a folded UBA(2) domain is capable of binding to the heme and unfolding the iso-1-Cyt*c* domain. Given the magnitudes of m_{ID} and $\Delta G_{ID}^{o'}(H_2O)$ for the D44H, S39H and N31H variants, the His-heme binding appears not to perturb the structure or stability of the UBA(2) domain. For the E27H, and to some extent for the A11H variant, m_{ID} is significantly decreased suggesting that binding to the heme of iso-1-Cyt*c* may perturb the structure of the UBA(2). However in the case of the E27H variant, the UBA(2) domain may itself be destabilized by the E27H substitution.

His-Heme Loop Formation in 4 M and 6 M GuHCl Relative to Random Coil Behavior. For a random coil, loop stability can be modeled using the Jacobson-Stockmayer equation (eq 2-1).[28] Assuming the His-heme loops exhibit random coil behavior and that the enthalpy of Fe^{3+}-His bond formation is constant for all His-heme loops, the free energy of loop formation, $\Delta G_{loop}(His)$, can be expressed with eq 2-7.

$$\Delta G_{loop}(His) = ln(10)RTpK_{loop}(His) = -T\Delta S_{loop} \quad (2\text{-}7)$$

Using the expression for ΔS_{loop} in eq 2-1, eq 2-7 can be rewritten as eq 2-8,

$$pK_{loop}(His) = pK_{loop}(His)_{ref} + \nu_3 \log(n) \quad (2\text{-}8)$$

where ν_3, the scaling exponent for loop formation and $pK_{loop}(His)_{ref}$ is nominally $pK_{loop}(His)$ for a loop size of 1. Therefore, $pK_{loop}(His)$ should vary linearly with the logarithm of loop size, *n*, with

a slope equal to ν_3. For a random coil with excluded volume, ν_3 is expected to have values ranging from 1.8 to 2.4.[58-60]

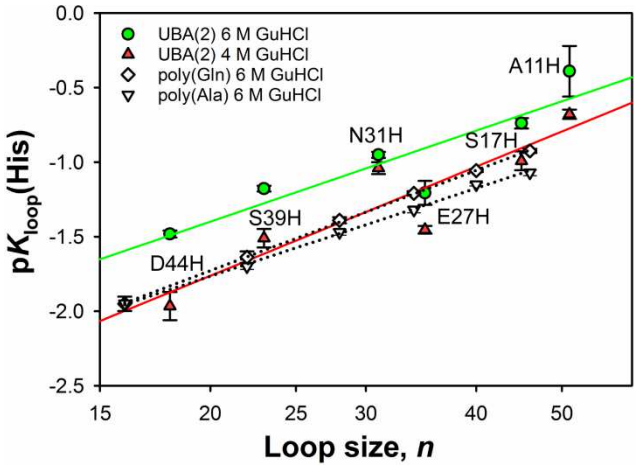

Figure 2-5. Plot of loop stability, pK_{loop}(His), vs. loop size (plotted logarithmically) in the denatured state of the three helix bundle UBA(2) domain fused to the N-terminus of iso-1-Cyt*c* for all histidine variants in 4 M (red) and 6 M (green) GuHCl. Data from homopolymeric inserts of alanine $(KAAAAA)_n$[29] and glutamine $(KQQQQQ)_n$[51] (n = 1 – 5) inserted between Phe(-3) and Lys(-2) of iso-1-Cyt*c* as for the UBA(2) domain (see Figure 2-1) acquired at 6 M GuHCl are included for comparison.

Figure 2-5 shows a plot of pK_{loop}(His) versus the log of loop size, *n*, for data acquired at 4 and 6 M GuHCl. His-heme loop formation with homopolymeric sequences of alanine, glycine or glutamine inserted at the N-terminus of iso-1-Cyt*c* lead to tight correlations between pK_{loop}(His) and log(*n*) (see Figure 2-5).[29, 51, 52] By comparison, there is considerably more scatter about the best fit line to eq 2-8, indicative of local deviations from random coil behavior for UBA(2). There is a particularly strong deviation for the E27H variant. pK_{loop}(His) is 0.28 and 0.30 units

more negative than expected from the best fit lines to eq 2-8 at 4 and 6 M GdnHCl, respectively. Thus, the His27-heme loop deviates from random coil behavior by ~0.4 kcal/mol [$\Delta\Delta G_{loop}$(His) = ln(10)RTΔpK_{loop}(His)].

The loop breakage data in Figure 2-4 also show considerable variation as a function of loop size. The small rate constant for loop breakage, k_b, for the E27H variant indicates that this portion of the UBA(2) can stabilize persistent structure in the denatured state, relative to other segments of the primary structure of UBA(2). Glu27 is part of the turn between helix 2 and 3 of UBA(2) (Figure 2-1). Thus, this result is consistent with our previous work on the four-helix bundle protein, Cyt*c′*, where persistent His-heme loops were found to localize near the turn between helices 1 and 2 and at the ends of the Ω-loop connecting helices 2 and 3.[31] Molecular dynamics (MD) simulations showed that, in the denatured state of Cyt*c′*, a nativelike topology is stabilized for these turns by dynamic hydrophobic clusters. For Cyt*c′*, aromatic amino acid sides chains (Phe and Trp) contribute significantly to these hydrophobic clusters, suggesting that Tyr23 and Phe24, which are near His27 in the UBA(2) domain of the E27H variant, play a role in the persistence of the His27-heme loop. We have shown that an Ala→Trp substitution three residues from the histidine which forms the His-heme loop, stabilizes the loop by ~0.3 kcal/mol in 6 M GuHCl and ~0.7 in 3 M GuHCl.[50] When an Ala four residues from the histidine which forms the His-heme loop is replaced by Phe, Tyr or Trp the His-heme loop is stabilized by 0.4 – 0.5 kcal/mol in 3 M GuHCl, whereas a Leu substitution leads to minimal stabilization of the loop.[61] A number of other studies show that aromatic residues stabilize residual structure and long range interactions in the denatured state of proteins.[9, 10, 62-66] These precedents support the involvement of Tyr23 and Phe24 in the deviation from random coil behavior observed for the His27-heme loop of the UBA(2) domain.

As for Cyt*c′*, persistent loops do not form for histidines near all reverse turns, indicating that the denatured state need not be fully biased toward the native topology to favor efficient folding. For acyl coenzyme A binding protein (ACBP), three of the four helices show nativelike long-range interactions in the acid denatured state.[67] The turn prediction algorithm, NetTurnP,[68] predicts the turn between helices 2 and 3, but not the turn between helices 1 and 2, suggesting that the second reverse turn sequence may be a stronger turn signal.

Bulk properties of the Denatured State of UBA(2). The fits of the plot of pK_{loop}(His) versus Log(n) yield scaling exponents, ν_3, of 2.4 ± 0.6 and 2.0 ± 0.5 at 4 and 6 M GuHCl, respectively. These values are consistent with the range 0f 1.8 – 2.4 expected for a random coil with excluded volume.[58-60] These values are also similar to the scaling exponents of 1.9 to 2.3 observed in 6 M GuHCl for His-heme loop formation with poly(Gly), poly(Gln) and poly(Ala) inserts into the N-terminus of iso-1-Cyt*c* (see Figure 2-5).[29, 51, 52] Simulations have shown that random coil scaling is compatible with local residual structure.[69]

A notable feature of the plot of pK_{loop}(His) versus Log(n) for the UBA(2) domain in 6 M GdnHCl (Figure 2-5) is that loop stability is considerably lower than for loops formed by the homopolymeric inserts of similar length in 6 M GuHCl. In fact, it is the 4 M GuHCl UBA(2) pK_{loop}(His) versus Log(n) data that corresponds well with the poly(Ala) and poly(Gln) 6 M GuHCl data. These data suggest that UBA(2) may have a more extended denatured state than the poly(Gln) or poly(Ala) inserts. Unfolding simulations of the engrailed homeodomain (en-HD), a three-helix bundle with a different topology from the UBA(2) domain, show that it has an extended denatured state.[70, 71] For the c-Myb transforming domain (c-Myb) and the human Rap1 Myb domain (hRAP1), which have the same three-helix bundle topology as en-HD, the denatured states are more compact.[70] The high intrinsic helical propensity of helix 1 of the en-

HD domain (>60% helix near the center of the helix) compared to c-Myb and hRap1 (no helices >20%),[70] as predicted by Agadir[72] appears to cause the more extended structure of the denatured state of en-HD. The center of helix 1 of UBA(2) is predicted to be ~37% helix by Agadir (default parameters, T = 278 K, μ = 0.1 M). Thus, like en-HD, UBA(2) may have a more extended denatured state. The low magnitudes of k_f calculated for loop formation with the UBA(2) domain also are consistent with a more extended denatured state (Figure 2-S3).

In general, GuHCl *m*-values for denatured state His-heme loop formation, m_{eq}, with a few exceptions,[46] are small (<0.2 kcal mol^{-1} M^{-1}).[29, 31, 51] m_{eq} values >0.2 kcal mol^{-1} M^{-1} are observed near the N- and C-terminal ends of the UBA(2) domain (Table 2-S6). The dominant contribution to the larger m_{eq} values is the *m* value associated with the rate constant for loop formation, $m_f^{\ddagger}$. This observation suggests that the ends of the domain are more sensitive to expansion of the denatured state, perhaps due to population of a U-shaped intermediate near the end of the unfolding transition in 4 M GuHCl, as observed for en-HD.[70, 71, 73]

CONCLUSIONS:

Fusion of a UBA(2) domain to the N-terminus of iso-1-Cyt*c* has allowed application of the denatured state His-heme loop formation method to a folded domain that does not contain a *c*-type heme. We find that the bulk properties of the domain under denaturing conditions (4 and 6 M GuHCl) are consistent with a random coil with excluded volume (ν_3 = 2.0 – 2.4). However, there is considerable scatter about the best fit line to the Jacobson-Stockmayer relationship. A histidine at position 27 in the turn between helices 2 and 3 forms a particular persistent loop, consistent with the observation for Cyt*c′* that persistent denatured state loops tend to occur near sequences that form reverse turns in the tertiary structure.[31] Loop formation is less favorable than

for homopolymeric sequences, which may reflect a more extend denatured state as observed for en-HD, which like UBA(2), has a helix with high intrinsic helical propensity.

SUPPORTING INFORMATION

Tables 2-S1 to 2-S3 provide sequences for oligonucleotides used for cloning and mutagenesis. Table 2-S4 contains MALDI-TOF data for the UBA(2) – iso-1-Cyt*c* variants. Tables 2-S5 and 2-S6 contain rate constants and *m*-values for denatured state His-heme loop formation. Figure 2-S1 provides a vector map for pRbs_BTR1 showing sites where restriction sites were added or eliminated to produce pRbs_BTR1fuse. Figure 2-S2 shows a typical kinetic trace for loop breakage with a UBA(2) – iso-1-Cyt*c* variant. Figure 2-S3 provides a plot of Log(k_f) versus loop size for the UBA(2) – iso-1-Cyt*c* variants.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bruce.bowler@umontana.edu

ORCID

Bruce E. Bowler: 0000-0003-1543-2466

Funding

This research was supported by National Science Foundation grant, MCB-1412164 and NIH grant, R01GM074750, to B.E.B. The Bruker microflex MALDI-TOF mass spectrometer was purchased with Major Research Instrumentation Grant CHE-1039814 from the National Science Foundation. M.J.L. acknowledges the Sloan Indigenous Graduate Partnership of the Alfred P.

Sloan Foundation and the National Science Foundation DEB 0614406 and the NSF EPSCoR Track-1 EPS-1101342 (INSTEP 3) for graduate education support.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

iso-1-Cyt*c*, iso-1-cytochrome *c*; CD, circular dichroism; GuHCl, guanidine hydrochloride; UBA(2), human HHR23A ubiquitin-associated domain 2; UBA(2) – iso-1-Cyt*c*, fusion protein with UBA(2) inserted near the N-terminus of iso-1-Cyt*c*; pWT, UBA(2)-iso-1-Cyt*c* carrying a C26A mutation in the UBA(2) domain.

Supporting Information

Probing Denatured State Conformational Bias in the Three Helix Bundle UBA(2) Using a Cytochrome *c* Fusion Protein

Moses J. Leavens, Melisa M. Cherney, Michaela Finnegan and Bruce E. Bowler*

Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Dynamics, University of Montana, Missoula, Montana 59812

*To whom correspondence should be addressed.

Telephone: (406) 282-1883. Fax: (406) 243-4227. E-mail: bruce.bowler@umontana.edu

A. Primers used for cloning and mutagenesis

Table 2-S1. Oligonucleotide primers for preparation of pRbs_BTR1fuse[a]

Primer	Primer Sequence 5′ to 3′
ER1elm	d(GAAACAGCTATGACCATGATTACGA**CC**TCGAGCTCGGTAC)
ER1elm-r	d(GTACCGAGCTCGA**GG**TCGTAATCATGGTCATAGCTGTTTC)
NgoM4add	d(GCTGAATTCAAGGCCGG**C**TCTGCTAAGAAAGGTGC)

NgoM4add -r	d(GCACCTTTCTTAGCAGA**G**CCGGCCTTGAATTCAGC)
ER3elm	d(CTTAGGAGGGTATCGATACTATGAA**C**TCACAAAAAA TGGGTTG)
ER3elm-r	d(CAACCCATTTTTTGTGA**G**TTCATAGTATCGATACCCT CCTAAG)
NgoM4add	d(GCCACGTTCGCC**T**GCTTTCCCCGTC)
NgoM4add -r	d(GACGGGGAAAGC**A**GGCGAAGCTGGC)

[a]Restriction sites are underlined; mutated bases are bolded.

Table 2-S2. Oligonucleotide primers for preparation of UBA(2) □□iso-1-Cyt*c* fusion protein[a]

Primer	Primer Sequence 5′ to 3′
UBA(2)ER1_for	d(ATATGAATTCCAGGAGAAAGAAGCTATA)
UBA(2)NgoM4_ rev	d(TTAAGCCGGCCTTATCAGCGAAGTTCTGACT)

UBA(2)fix	d(CTCCTGAGTCAGAACTT**T**G**A**TGA**C**G**AG**AAGGCCG GCTCTGC)
UBA(2)fix-r	d(GCAGAGCCGGCCTT**CTCG**TCA**T**C**A**AAGTTCTGACT CAGGAG)

[a]Restriction sites are underlined; mutated bases are bolded.

Table 2-S3. Oligonucleotide Primers for Preparation of UBA(2)-iso-1-Cyt*c* Variants[a]

Primer	Primer Sequence 5′ to 3′
C26A	d(TCATCCAGGCCTATTTCGCG**GCT**GAAAAAAATGAGAACTTGG)
C26A-r	d(CCAAGTTCTCATTTTTTTC**AGC**CGCGAAATAGGCCTGGATGA)
A11H	d(GCTATAGAGAGGTTGAAG**CAC**CTGGGCTTCCCAGAGAGC)
A11H-r	d(GCTCTCTGGGAAGCCCAG**GTG**CTTCAACCTCTCTATAGC)
S17H	d(GCCCTGGGCTTCCCAGAG**CAC**CTGGTCATCCAGGCCTATTTC)
S17H-r	d(GAAATAGGCCTGGATGACCAG**GTG**CTCTGGGAAGCCCAGGGC)
E27H	d(CAGGCCTATTTCGCGGCA**CAT**AAAAATGAGAACTTGGC)
E27H-r	d(GCCAAGTTCTCATTTTT**ATG**TGCCGCGAAATAGGCCTG)
N31H	d(GCGGCAGAAAAAAATGAG**CAC**TTGGCTGCCAACTTCCTC)
N31H-r	d(GAGGAAGTTGGCAGCCAA**GTG**CTCATTTTTTTCTGCCGC)

S39H	d(GCTGCCAACTTCCTCCTG**CAT**CAGAACTTTGATGACGAG)
S39H-r	d(CTCGTCATCAAAGTTCTG**ATG**CAGGAGGAAGTTGGCAGC)
D44H	d(CTGAGTCAGAACTTTGAT**CAC**GAGAAGGCCGGCTCTGC)
D44H-r-r	d(GCAGAGCCGGCCTTCTC**GTG**ATCAAAGTTCTGACTCAG)

[a]Mutation sites are bolded and underlined.

B. Characterization of UBA(2)-iso-1-Cyt*c* fusion proteins

Table 2-S4. MALDI-ToF Mass Spectral Data for UBA(2) – iso-1-Cyt*c* variants

Variant	m/z observed	m/z expected[a]
pWT	17,633.55	17,636.71
D44H	17655.66	17,658.77
S39H	17,690.52	17,686.78
N31H	17661.10	17,659.75
E27H	17,649.83	17,644.74
S17H	17,687.63	17,686.78
A11H	17,701.64	17,702.78

[a]Expected m/z obtained using ExPASy PeptideMass tool (http://web.expasy.org/peptide_mass/) using average mass, no cutting and $[M+H]^+$ options.

Table 2-S5. Equilibrium *m*-values and kinetic loop breakage and loop formation *m*-values for UBA(2) – Cyt *c* variants

Variant	m_f (kcalmol^{-1} M^{-1})	m_b (kcalmol^{-1}M^{-1})	m_f+m_b (kcalmol^{-1}M)	m_{eq} (kcalmol^{-1}M^{-1})
D44H	0.291 ± 0.06	0.0388 ± 0.01	0.330 ± 0.06	0.330 ± 0.08
S39H	0.131 ± 0.04	0.0955 ± 0.009	0.226 ± 0.04	0.226 ± 0.05
N31H	0.0919 ± 0.03	0.152 ± 0.01	0.060 ± 0.04	0.0738 ± 0.12
E27H	0.0340 ± 0.05	0.134 ± 0.003	0.168 ± 0.05	0.168 ± 0.07
S17H	0.0278 ± 0.05	0.143 ± 0.01	0.171 ± 0.05	0.171 ± 0.06
A11H	0.171 ± 0.06	0.121 ± 0.008	0.292±0.06	0.292 ± 0.08

Data calculated using viscosity corrected k_b values in 4 M GuHCl conditions. The 6 M GuHCl loop breakage data is uncorrected for viscosity to allow comparison to poly(A) Cyt*c* variants under 6 M GuHCl conditions. Standard propagation of the errors were used to calculate parameteres from k_f, k_b, and pK_a(obs). The m_{eq}, m_f, and m_b values were calculated using the below equations.

$$m_b = \frac{RT}{2} ln\left(\frac{k_b^{6M}}{k_b^{4M}}\right), \qquad m_f = -\frac{RT}{2} ln\left(\frac{k_f^{6M}}{k_f^{4M}}\right)$$

$$m_{eq} = \frac{ln(10)RT}{2}\left(pK_{loop}(His)^{6M} - pK_{loop}(His)^{4M}\right)$$

Table 2-S6. Loop breakage rate constants for UBA(2)-Cyt*c* Variants in 6M and 4M GuHCl at pH 3.5 and pH 3.0

Variant	Loop size	k_b pH 3.41±0.07 (6M)	k_b pH 3.54±0.18 (6M)	k_b pH 3.59±0.10 (4M)	k_b pH 4.11±0.16 (4M)
pWT	-	117.29 ± 2.9	120.1 ± 6.3	88.3 ± 2.4	97.4 ± 3.6
D44H	18	65.33 ± 2.9	65.8 ± 2.3	57.3 ± 1.3	78.6 ± 3.7
S39H	23	82.66 ± 1.6	81.9 ± 1.0	59.8 ± 1.6	77.2 ± 3.2
N31H	31	130.41 ± 6.9	126.1 ± 6.8	77.8 ± 1.8	86.4 ± 1.7
E27H	35	45.67 ± 0.5	45.5 ± 0.5	28.9 ± 0.2	35.4 ± 1.0
S17H	45	112.37 ± 3.6	106.4 ± 2.8	69.18 ± 1.5	77.4 ± 1.5
A11H	51	154.35 ± 3.7	153.8 ± 4.1	102.4 ± 1.5	111.7 ± 4.5

Loop breakage rate constants of at least nine trials showing breakage has gone to completion. The 6 M GuHCl data is uncorrected for viscosity to compare to 6 M poly(A) data for iso-1-Cyt c. The 4 M data has been corrected for viscosity.

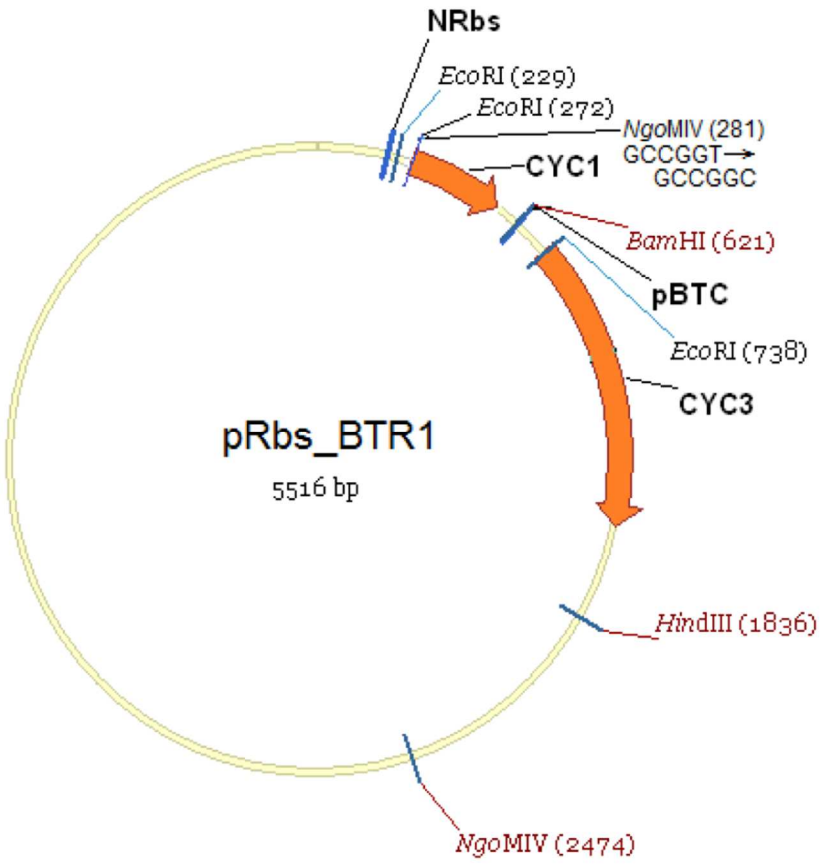

Fig. 2-S1. Map of pRbs_BTR1 plasmid. The locations are shown for the iso-1-cytochrome *c* gene, *CYC1*, and of the heme lyase gene, *CYC3*. The *Eco*R1 restriction sites at positions 229 and 738, and the *Ngo*MIV restriction site at position at position 2474, which are eliminated and the *Ngo*MIV, which is introduced at position 281, to produce the pRbs_BTR1fuse plasmid, also are shown. The *Eco*R1(272) and *Ngo*MIV(281) restriction site allows insertion of the genes for protein

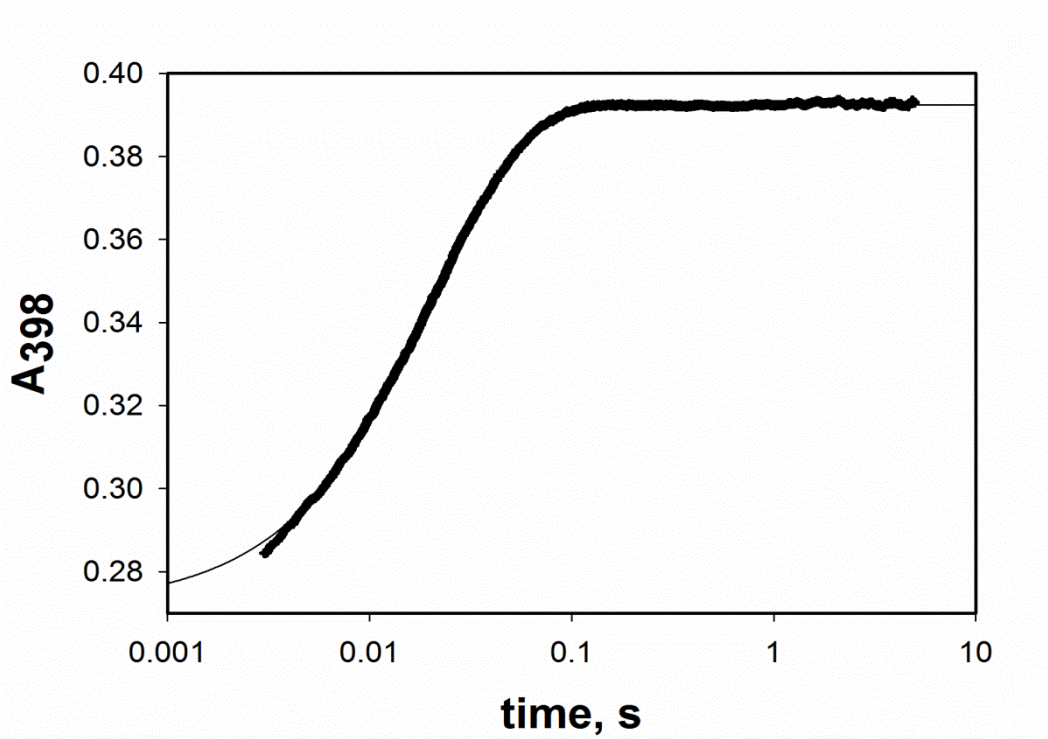

Figure 2-S2. Representative plot of histidine-heme loop breakage in the denatured state at 6 M GuHCl conditions for the UBA(2) E27H – Cytc variant. Data was collected using absorbance at 398 nm, A_{398}, while performing pH jump from pH 6.8 to pH 3.41. The solid curve is a fit of the data to a single exponential equation.

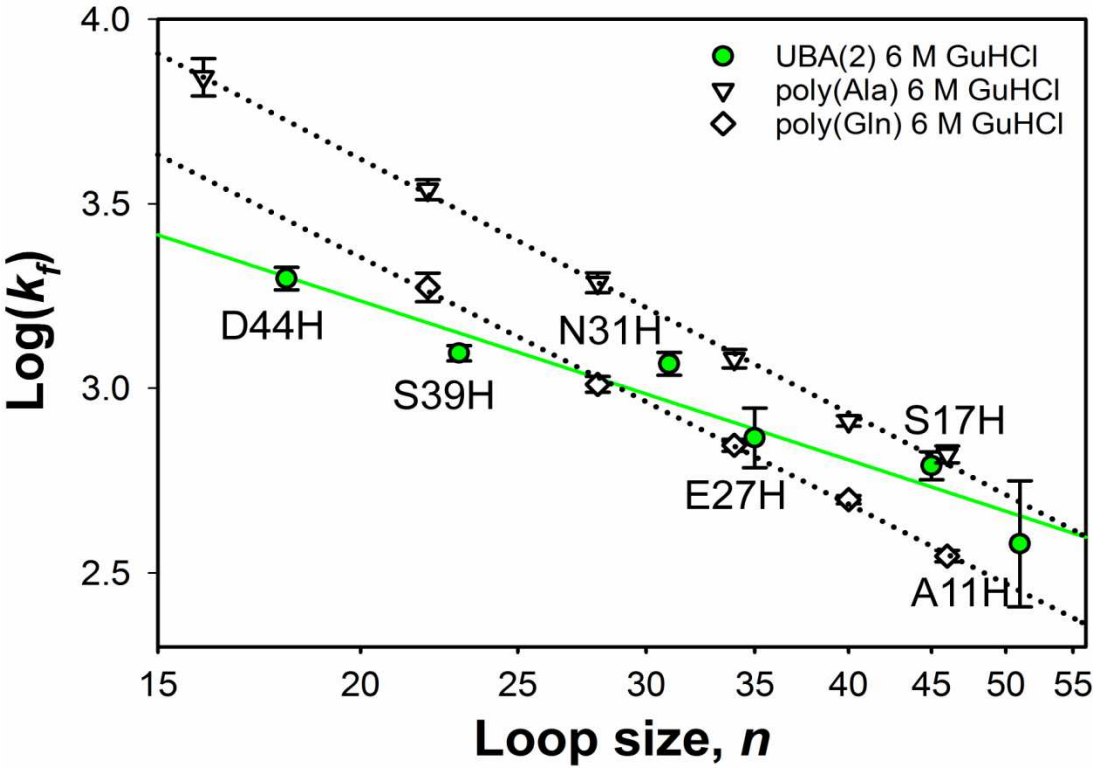

Fig. 2-S3. Plot of loop stability, Log(k_f), vs. loop size (plotted logarithmically) in the denatured state of the three helix bundle UBA(2) domain fused to the N-terminus of iso-1-Cyt*c* for all histidine variants in 6 M GuHCl. k_f was calculated assuming loop formation is two state: $k_f = k_b \times K_{loop}(His)$. k_b at pH 3.41 from Table S6 was used for k_b and $K_{loop}(His)$ was calculated from the p$K_{loop}(His)$ data in Table 2 of the main text. Data from homopolymeric inserts of alanine $(KAAAAA)_n$ [29] and glutamine $(KQQQQQ)_n$ [51] (n = 1-5) inserted between Phe(-3) and Lys(-2) of iso-1-Cyt*c* as for the UBA(2) domain (see Figure 2-1) acquired at 6 M GuHCl are included for comparison.

REFERENCES FOR UBA(2)-CYT C

1. Mittag, T., and Forman-Kay, J. D. (2007) Atomic-level characterization of disordered protein ensembles, *Curr. Opin. Struct. Biol. 17*, 3-14.
2. Cho, J.-H., Sato, S., Horng, J.-C., Anil, B., and Raleigh, D. P. (2008) Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability, *Arch. Biochem. Biophys. 469*, 20-28.
3. Bowler, B. E. (2012) Globular proteins: characterization of the denatured state, In *Comprehensive Biophysics* (Egelman, E., Ed.), pp 72-114, Academic Press, Oxford.
4. Bowler, B. E. (2007) Thermodynamics of protein denatured states, *Mol. BioSyst. 3*, 88-99.
5. Dill, K. A., and Shortle, D. (1991) Denatured states of proteins, *Annu. Rev. Biochem. 60*, 795-825.
6. Zhang, O., Kay, L. E., Shortle, D., and Forman-Kay, J. D. (1997) Comprehensive NOE characterization of a partially folded large fragment of staphylococcal nuclease Δ131Δ, using NMR methods with improved resolution, *J. Mol. Biol. 272*, 9-20.
7. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures, *J. Mol. Biol. 268*, 179-184.
8. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels, *J. Mol. Biol. 268*, 158-169.
9. Wirmer, J., Schloerb, C., Klein-Seetharaman, J., Hirano, R., Ueda, T., Imoto, T., and Schwalbe, H. (2004) Protein interactions: modulation of compactness and long-range

interactions of unfolded lysozyme by single point mutations, *Angew. Chem. Int. Ed. 43*, 5780-5785.

10. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein, *Science 295*, 1719-1722.
11. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea, *Biochemistry 36*, 8977-8991.
12. Neri, D., Billeter, M., Wider, G., and Wüthrich, K. (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor, *Science 257*, 1559-1563.
13. Pakula, A. A., and Sauer, R. T. (1990) Reverse hydrophobic effects relieved by aminoacid substitutions at a protein surface., *Nature 344*, 363–364.
14. Bowler, B. E., May, K., Zaragoza, T., York, P., Dong, A., and Caughey, W. S. (1993) Destabilizing effects of replacing a surface lysine of cytochrome *c* with aromatic amino acids: implications for the denatured state, *Biochemistry 32*, 183-190.
15. Bowler, B. E., Dong, A., and Caughey, W. S. (1994) Characterization of the guanidine hydrochloride-denatured state of iso-1-cytochrome *c* by infrared spectroscopy, *Biochemistry 33*, 2402-2408.
16. Herrmann, L., Bowler, B. E., Dong, A., and 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.

17. Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions, *Biochemistry 38*, 4896-4903.

18. Cho, J.-H., and Raleigh, D. P. (2005) Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins, *J. Mol. Biol. 353*, 174-185.

19. Cho, J.-H., Sato, S., and Raleigh, D. P. (2004) Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state, *J. Mol. Biol. 338*, 827-837.

20. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M., and Pace, C. N. (1999) Increasing protein stability by altering long-range coulombic interactions, *Protein Sci. 8*, 1843-1849.

21. Pace, C. N., Alston, R. W., and Shaw, K. L. (2000) Charge-charge interactions influence the denatured state ensemble and contribute to protein stability, *Protein Sci. 9*, 1395-1398.

22. Tan, Y. J., Oliveberg, M., Davis, B., and Fersht, A. R. (1995) Perturbed pK_A-values in the denatured states of proteins, *J. Mol. Biol. 254*, 980-992.

23. Whitten, S., and Garcia-Moreno E., B. (2000) pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state., *Biochemistry 39*, 14292–14304.

24. Swint-Kruse, L., and Robertson, A. D. Hydrogen bonds and the pH dependence of ovomucoid third domain stability, *Biochemistry 34*, 4724–4732.

25. Trefethen, J. M., Pace, C. N., Scholtz, J. M., and Brems, D. N. (2005) Charge-charge interactions in the denatured state influence the folding kinetics of ribonuclease Sa, *Protein Sci. 14*, 1934-1938.

26. Cho, J.-H., and Raleigh, D. P. (2006) Electrostatic interactions in the denatured state and in the transition state for protein folding: effects of denatured state interactions on the analysis of transition state structure, *J. Mol. Biol. 359*, 1437-1446.

27. Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988) Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding, *Biochemistry 27*, 7167-7175.

28. Jacobson, H., and Stockmayer, W. H. (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems, *J. Chem. Phys. 18*, 1600-1606.

29. Tzul, F. O., and Bowler, B. E. (2010) Denatured states of low complexity polypeptide sequences differ dramatically from those of foldable sequences, *Proc. Natl. Acad. Sci. U.S.A. 107*, 11364-11369.

30. Rao, K. S., Tzul, F. O., Christian, A. K., Gordon, T. N., and Bowler, B. E. (2009) Thermodynamics of loop formation in the denatured state of *Rhodopseudomonas palustris* cytochrome *c*': scaling exponents and the reconciliation problem, *J. Mol. Biol. 392*, 1315-1325.

31. Dar, T. A., Schaeffer, R. D., Daggett, V., and Bowler, B. E. (2011) Manifestations of native topology in the denatured state ensemble of *Rhodopseudomonas palustris* cytochrome *c*', *Biochemistry 50*, 1029-1041.

32. Duncan, M. G., Williams, M. D., and Bowler, B. E. (2009) Compressing the free energy range of substructure stabilities in iso-1-cytochrome *c*, *Protein Sci. 18*, 1155-1164.

33. Pollock, W. B., Rosell, F. I., Twitchett, M. B., Dumont, M. E., and Mauk, A. G. (1998) Bacterial expression of a mitochondrial cytochrome *c*. Trimethylation of Lys72 in yeast iso-1-cytochrome *c* and the alkaline conformational transition, *Biochemistry 37*, 6124-6131.

34. Rosell, F. I., and 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.

35. McClelland, L. J., Seagraves, S. M., Khan, M. K. A., Cherney, M. M., Bandi, S., Culbertson, J. E., and Bowler, B. E. (2015) The response of Ω-loop D dynamics to truncation of trimethyllysine 72 of yeast iso-1-cytochrome *c* depends on the nature of loop deformation, *J. Biol. Inorg. Chem. 20*, 805-819.

36. Godbole, S., and Bowler, B. E. (1997) A histidine variant of yeast iso-1-cytochrome *c* that strongly affects the energetics of the denatured state, *J. Mol. Biol. 268*, 816-821.

37. Withers-Ward, E. S., Mueller, T. D., Chen, I. S. Y., and Feigon, J. (2000) Biochemical and structural analysis of the interaction between the UBA(2) domain of the DNA repair protein HHR23A and HIV-1 Vpr, *Biochemistry 39*, 14103-14112.

38. Cherney, M. M., Junior, C., and Bowler, B. E. (2013) Mutation of trimethyllysine-72 to alanine enhances His79-heme mediated dynamics of iso-1-cytochrome *c*, *Biochemistry 52*, 837-846.

39. Redzic, J. S., and Bowler, B. E. (2005) Role of hydrogen bond networks and dynamics in positive and negative cooperative stabilization of a protein, *Biochemistry 44*, 2900-2908.

40. Wandschneider, E., Hammack, B. N., and Bowler, B. E. (2003) Evaluation of cooperative interactions between substructures of iso-1-cytochrome *c* using double mutant cycles, *Biochemistry 42*, 10659-10666.

41. Goldes, M. E., Jeakins-Cooley, M. E., McClelland, L. J., Mou, T.-C., and Bowler, B. E. (2016) Disruption of a hydrogen bond network in human versus spider monkey cytochrome *c* affects heme crevice stability, *J. Inorg. Biochem. 158*, 62-69.

42. Margoliash, E., and Frohwirt, N. (1959) Spectrum of horse-heart cytochrome *c*, *Biochem. J. 71*, 570-572.

43. Nozaki, Y. (1972) The preparation of guanidine hydrochloride, *Methods Enzymol. 26*, 43-50.

44. Pace, C. N., Shirley, B. A., and Thomson, J. A. (1989) Measuring the conformational stability of a protein, In *Protein structure: a practical approach* (Creighton, T. E., Ed.), pp 311-330, IRL Press at Oxford University Press, New York.

45. Schellman, J. A. (1978) Solvent denaturation, *Biopolymers 17*, 1305-1322.

46. Wandschneider, E., and Bowler, B. E. (2004) Conformational properties of the iso-1-cytochrome *c* denatured state: dependence on guanidine hydrochloride concentration, *J. Mol. Biol. 339*, 185-197.

47. Tonomura, B., Nakatani, H., Ohnishi, M., Yamaguchi-Ito, J., and Hiromi, K. (1978) Test reactions for a stopped-flow apparatus. Reduction of 2,6-dichlorophenolindophenol and potassium ferricyanide by L-ascorbic acid, *Anal. Biochem. 84*, 370-383.

48. Anil, B., Song, B., Tang, Y., and Raleigh, D. P. (2004) Exploiting the right side of the Ramachandran plot: substitution of glycines by D-alanine can significantly increase protein stability, *J. Am. Chem. Soc. 126*, 13194-13195.

49. Smith, C. R., Mateljevic, N., and Bowler, B. E. (2002) Effects of topology and excluded volume on protein denatured state conformational properties, *Biochemistry 41*, 10173-10181.

50. Khan, M. K. A., Miller, A. L., and Bowler, B. E. (2012) Tryptophan significantly stabilizes His-heme loops only when it is near a loop end, *Biochemistry 51*, 3586-3595.

51. Khan, M. K. A., and Bowler, B. E. (2012) Conformational properties of polyglutamine sequences in guanidine hydrochloride solutions, *Biophys. J. 103*, 1989–1999.

52. Finnegan, M. L., and Bowler, B. E. (2012) Scaling properties of glycine-rich sequences in guanidine hydrochloride solutions, *Biophys. J. 102*, 1969-1978.

53. Xu, Y., Mayne, L., and Englander, S. W. (1998) Evidence for an unfolding and refolding pathway in cytochrome *c*, *Nat. Struct. Biol. 5*, 774-778.

54. Smith, C. R., Wandschneider, E., and Bowler, B. E. (2003) Effect of pH on the iso-1-cytochrome *c* denatured state: changing constraints due to heme ligation, *Biochemistry 42*, 2174-2184.

55. Kurchan, E., Roder, H., and Bowler, B. E. (2005) Kinetics of loop formation and breakage in the denatured state of iso-1-cytochrome *c*, *J. Mol. Biol. 353*, 730-743.

56. Rohl, C. A., Chakrabartty, A., and Baldwin, R. L. (1996) Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol, *Protein Sci. 5*, 2623-2637.

57. Chakrabartty, A., Schellman, J. A., and Baldwin, R. L. (1991) Large differences in the helical propnesity of glycine and alanine, *Nature 351*, 586-588.

58. de Gennes, P.-G. (1979) *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, NY.

59. Chan, H. S., and Dill, K. A. (1990) The effect of internal constraints on the configurations of chain molecules, *J. Chem. Phys. 92*, 3118-3135.

60. Redner, S. (1980) Distribution functions in the interior of polymer chains, *J. Phys. A: Math. Gen. 13*, 3525-3541.

61. Finnegan, M. L., and Bowler, B. E. (2010) Propensities of aromatic amino acids versus leucine and proline to induce residual structure in the denatured-state ensemble of iso-1-cytochrome *c*, *J. Mol. Biol. 403*, 495-504.

62. Crowhurst, K. A., and Forman-Kay, J. D. (2003) Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain, *Biochemistry 42*, 8687-8695.

63. Crowhurst, K. A., Tollinger, M., and Forman-Kay, J. D. (2002) Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain, *J. Mol. Biol. 322*, 163-178.

64. Marsh, J. A., and Forman-Kay, J. D. (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints, *J. Mol. Biol. 391*, 359-374.

65. Day, R., and Daggett, V. (2005) Ensemble versus single-molecule protein unfolding, *Proc. Natl. Acad. Sci. U.S.A. 102*, 13445-13450.

66. Wong, K.-B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M. V., Fersht, A. R., and Daggett, V. (2000) Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding, *J. Mol. Biol. 296*, 1257-1282.

67. Bruun, S. W., Iešmantavičius, V., Danielsson, J., and Poulsen, F. M. (2010) Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein, *Proc. Natl. Acad. Sci. U.S.A. 107*, 13306-13311.

68. Petersen, B., Lundegaard, C., and Petersen, T. N. (2010) NetTurnP - neural network prediction of beta-turns by use of evolutionary information and predicted protein sequence features, *PLoS ONE 5*, e15079.

69. Fitzkee, N. C., and Rose, G. D. (2004) Reassessing random-coil statistics in unfolded proteins, *Proc. Natl. Acad. Sci. U.S.A. 101*, 12497-12502.

70. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. N., DeMarco, M. L., Daggett, V., and Fersht, A. R. (2003) Unifying features in protein folding mechanisms, *Proc. Natl. Acad. Sci. U.S.A. 100*, 13286-13291.

71. DeMarco, M. L., Alonso, D. O. V., and Daggett, V. (2004) Diffusing and colliding: the atomic level folding/unfolding pathway of a small helical protein, *J. Mol. Biol. 341*, 1109-1124.

72. Muñoz, V., and Serrano, L. (1994) Elucidating the folding problem of α-helical peptides using empirical parameters III: temperature and pH dependence, *J. Mol. Biol. 245*, 297-308.

73. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V., and Fersht, A. R. (2005) Solution structure of a protein denatured state and folding intermediate, *Nature 437*, 1053-1056.

74. Anfinsen, Christian B. (1973). Principles that Govern the Folding of Protein Chains. *Science, 181*(4096), 223-230.

Chapter 3: Three-Helix Bundle UBA Domains with Similar Native State Topology and Divergent Sequence Contain Denatured State Conformational Bias Bracketing Turns

Three-Helix Bundle UBA Domains with Similar Native State Topology and Divergent Sequence Contain Denatured State Conformational Bias Bracketing Turns

Moses J. Leavens, Lisa E. Spang, Melisa M. Cherney, Michaela L. Finnegan and Bruce E. Bowler*

Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Dynamics, University of Montana, Missoula, Montana 59812

ABSTRACT:

Observations in the four-helix-bundle protein cytochrome *c′* from *Rhodopseudomonas palustris* and the three-helix bundle protein Ubiquitin-associated domain 2 from human DNA excision repair protein HHR23A revealed non-random behavior in their denatured state ensembles. Mapping these regions with denatured state conformational bias onto their native structures show these regions localize to the second reverse turn. To further test this turn hypothesis, we extend this work to a three-helix bundle with divergent sequence, the internal Ubiquitin-associated domain from HHR32A, the first ubiquitin-associated domain, UBA(1). We apply histidine-heme loop formation methodology to probe the denatured state conformational properties of UBA(1), using yeast iso-1-cytochrome *c* as a scaffold, fusing the UBA(1) domain at the N-terminus of iso-1-cytochrome *c*. We engineer histidine into high solvent accessible positions of UBA(1), creating eight single histidine variants. Similar to previous work with UBA(2), guanidine hydrochloride equilibrium unfolding shows the UBA(1)-cytochrome *c* fusion protein unfolds in 3-state manner, with iso-1-cytochrome *c* unfolding first. Moreover, the engineered histidine residues in UBA(1) destabilize the iso-1-cytochrome *c* domain. Histidine-heme loop formation equilibrium and kinetic measurements at 4 and 6 M guanidine hydrochloride show loop stability decreases as the size of the histidine-heme loop increases, in accordance with the Jacobson-Stockmayer equation. Yet we observe the His15-heme and His31-heme loops are more stable than expected from the Jacobson-Stockmayer relationship, and break more slowly than expected, relative to loops exhibiting random coil behavior in the denatured state. These results indicate local sequence near His15 and His31, which are near the reverse turns of UBA(1), contain persistent interactions in the denatured state. Consistent with our results for the four-helix bundle cytochrome *c′*, and the three-helix bundle UBA(2), these reverse turn sequences may establish helix bundle topology, through biasing the conformational distribution of their denatured states.

INTRODUCTION

Levinthal's paradox states unfolded proteins cannot undergo a random search[74]. Years later, proteins folding kinetic experiments on several small domains with varying native topology were discovered to fold within the millisecond and even microsecond time regime. These findings indicate that shortcuts or folding pathways must exist for a protein sequence to collapse efficiently to its native conformer. These regions of interest along a protein chain, thought to contain this energetic bias, are of interest for protein folding evolution and could provide insight into protein misfolding diseases, and thus are termed non-random or 'residual structure'. Residual structure within a protein's amino acid sequence may provide clues to understanding translation of amino acid sequence to structure[75], and shed light on understanding mechanistically how evolution favors a particular fold. At one time, denatured proteins were thought to lack energetically persistent interactions, and in fact are modeled well with the power law relationship[86], where the radius of gyration, R_g, is proportional to the number of monomers in a chain (eq 3-1).

$$(3-1)\ R_g = c \times N^{\upsilon}$$

In eq. 3-1, *c* is a constant, *N* is the number of monomers in the chain, and ν is a scaling exponent. For proteins denatured in guanidine hydrochloride (GuHCl) or urea, it was shown that the majority of these denatured proteins adhered well to this model, yielding scaling exponents consistent with scaling exponents for random coil homopolymers, suggesting unfolded proteins behave as random coils in denaturing conditions and thus are energetically featureless[86]. Although, denatured proteins behave as random coils, they appear to contain pockets of residual structure in the denatured state ensemble (DSE), an observation termed the reconciliation

problem. Work on staphylococcal nuclease has shown using NMR residual dipolar coupling, that denatured nuclease retains native state topology in high concentrations of urea[6-8]. Moreover, long range interactions and hydrophobic stabilization has been observed in the DSE, and these interactions in the DSE may be critical in limiting the conformational search and providing collapse of a protein chain to its native conformer.[5] Several proteins in varying denaturing conditions of urea or guanidine hydrochloride have been observed to possess residual structure.[6-11] For example, the 434-repressor protein denatured in 8 M urea was shown to contain residual structure in hydrophobic residues 55-60 in the amino acid sequence.[12]

To better understand the energetics of residual structure that appear to occur in denatured proteins, loop formation thermodynamic and kinetic methods have been developed[75]. The probability of forming a loop along a protein's primary structure in the denatured state can detect the potential intrinsic conformational bias of a protein sequence. In the DSE, a protein's primary structure is modeled with the Jacobson-Stockmayer equation (eq. 3-2), which predicts loop formation for a random coil is solely limited by entropy.[28]

$$(3-2)\ \Delta S_{\mathrm{loop}} = -\nu_3 \mathrm{R} ln(n) + \mathrm{R} ln((3/2\pi C_{\mathrm{n}} l^2)^{\nu_3} \mathrm{V_i})$$

In eq. 3-2, ν_3 is the scaling exponent for loop formation, R is the gas constant, *n* is the number of monomers in the loop, C_{n} is Flory's characteristic ratio, *l* is the distance between the monomers forming the loop, and $\mathrm{V_i}$ is the approach volume within which the atoms must be constrained for loop formation to occur. We recently applied our denatured state histidine-heme (His-heme) loop formation thermodynamic and kinetic method to a three-helix bundle not containing a *c*-type heme[82], and thus have shown successfully that the denatured state His-heme loop formation method can be applied to any fold.

Specifically, we observed using this denatured state His-heme loop formation thermodynamic and kinetic method, that the three-helix bundle protein UBA(2) fused to yeast iso-1-cytochrome *c*, had a scaling exponent in the range of a random coil with excluded volume. This was previously observed for the four-helix bundle protein Cyt*c*'.[31] While we observed that loop formation decreases as the size of the His-heme loop increases, a subset of denatured state His-heme loops in these three-helix and four-helix bundles are more favorable to form than expected, and they also break more slowly than expected, relative to His-heme loops exhibiting random coil behavior in the denatured state[29]. Thus, these data for Cyt*c*' and UBA(2) are indicative of non-random behavior in the DSE. These results are consistent with the distinct behavior of denatured foldable proteins versus low complexity polypeptide sequences[29]. The His-heme loops which are more stable than expected based on expectations of the Jacobson-Stockmayer equation, and contain slow loop breakage rates, imply that these segments of the primary structure in these three-helix and four-helix-bundles are prone to stabilizing interactions in the denatured state.[30, 31] The portions of primary structure containing this thermodynamic and kinetic bias in the DSE, localize near reverse turns in the tertiary structures of Cyt*c*' and UBA(2).[31,82] In this way, residues bracketing reverse turns make persistent contacts in the DSE in these folds. In the four-helix bundle Cyt*c*', observations with molecular dynamic simulations showed that the persistent turns in this four-helix bundle are stabilized by hydrophobic side chains, suggesting an additive stabilization effect by these aromatics in the DSE[31]. These interactions in the DSE may play an important role in establishing topology of these four and three-helix bundles.

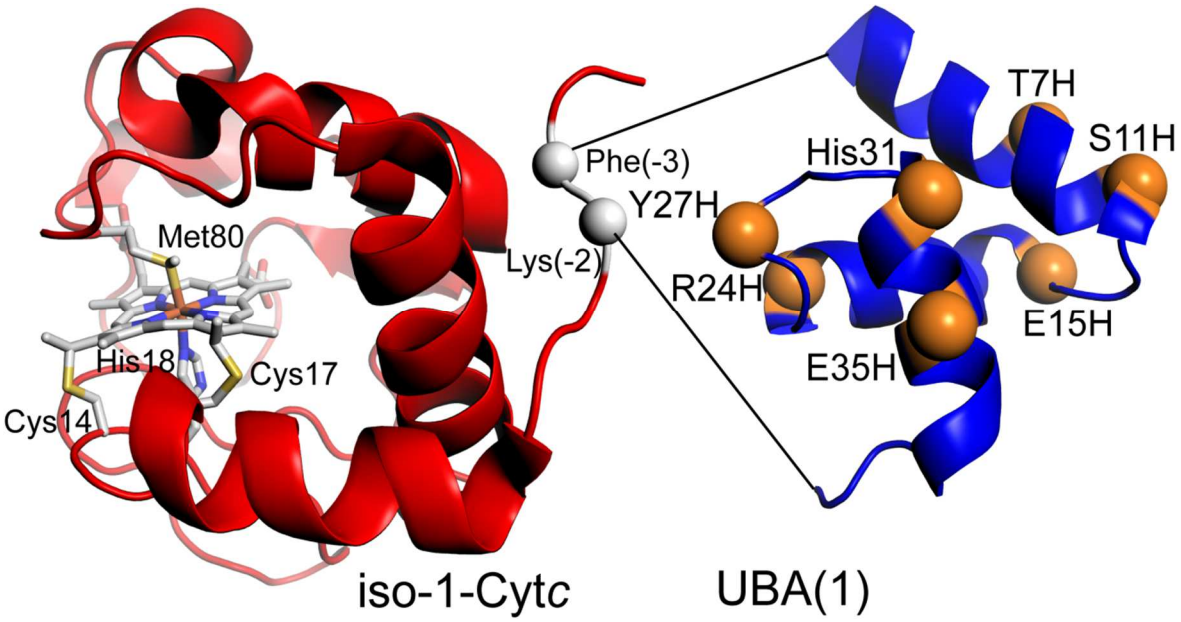

Figure 3-1. Structures of yeast iso-1 cytochrome *c* (red, PDB ID: 2YCC) and the human HHR23A UBA(1) domain (blue, PDB: 1IFY), showing insertion of the UBA(1) domain between Phe(-3) and Lys(-2) (white spheres) near the N-terminus of iso-1-Cyt *c.* Sites of single histidine substitutions are depicted with orange spheres in the UBA(1) domain. The heme and the amino acids that attach the heme to the iso-1-Cyt*c* polypeptide, Cys14, Cys17, His18 and Met80 are shown as stick models colored by element**.**

Based on the hydrophobic clustering observation of within turns in Cyt*c*' in the denatured state, we proposed Tyr23 and Phe24 in the UBA(2) domain may be involved in stabilizing its persistent second reverse turn in the DSE[82]. Herein, we apply our denatured state His-heme loop formation thermodynamic and kinetic method to examine the DSE intrinsic conformational properties of the internal UBA domain from HHR23A, the first Ubiquitin-associated domain, UBA(1), using a UBA(1) –iso-1-Cyt*c* fusion protein (Fig. 3-1). The UBA(1) and UBA(2) amino acid sequences are approximately 20% identical, and have similar native state topology—forming compact three-helix bundles[84]. Examining His-heme loop equilibria in the denatured

state for UBA(1), we observe His-heme loops have greater scatter about the line of best fit relative to UBA(2), where its scaling exponent is within the range for a random coil with excluded volume. Furthermore, we observe that strong deviations from random coil behavior along the primary structure localize to the two reverse turns in the native structure of UBA(1). Denatured state His-heme loop breakage kinetics monitored by stopped-flow spectroscopy indicate small His-heme loop breakage rate constants, k_b, for the histidines that probe these segments of DSE primary structure in UBA(1). To examine the role of a local hydrophobic side chain on persistent His-heme loops in the denatured state, a Tyr→Gln mutation at position 27 in the second reverse turn sequence of UBA(1) was prepared, and discovered to have no significant effect on stabilization of persistent structure when a His31-heme loop forms. However, in general agreement with the three-helix bundle UBA(2) and the four-helix bundle Cyt*c*', reverse turn sequences in the tertiary structure of UBA(1) make persistent contacts in the DSE for a three-helix bundle with low sequence identity. Thus, reverse turns in helix-bundles may help limit the conformational distribution of their denatured states, and perhaps govern formation of their overall topology.

EXPERIMENTAL PROCEDURES

Preparation of the human HHR23A UBA(1) domain fused to the N-terminus of iso-1-Cyt*c*. The pRbs_BTR1 fuse vector[32] was used to fuse the gene for the first ubiquitin associated domain, UBA(1), of the human homolog of yeast Rad23A, HHR23A (resides 7 – 49) at the N-terminus of yeast iso-1-Cyt*c* domain. The UBA(1) gene (cloned into the pGEX-2T vector obtained from the Feigon laboratory at UCLA.[37]) was amplified by PCR using a forward primer which placed an *Eco*RI restriction site at the beginning of the gene and a reverse primer which placed Lys(-2) from iso-1-Cyt*c* and an *Ngo*MIV restriction site at the end of the gene (Table 3-

S1), and purified using a QIAquick PCR purification kit (Qiagen) as described previously for preparation of the UBA(2) – iso -1-Cyt*c* fusion protein.[82] The PCR product was inserted into *Eco*RI and *Ngo*MIV restriction sites of the pRbs_BTR1fuse vector as previously described for the UBA(2) gene[82], to produce a plasmid carrying the gene for the UBA(1) – iso -1-Cyt*c* fusion protein, pRbs_BTR1(UBA1_Cc). In this construct residues 7-49 of UBA(1) are placed between Phe(-3) and Lys(-2) of iso –1-Cyt*c*. The sequence of pRbs_BTR1(UBA1_Cc) was confirmed by dideoxy sequencing of the entire UBA(1) – iso-1-Cyt*c* gene (Genomics Core Facility, University of Montana).

UBA(1) – iso-1-Cyt*c* fusion protein preparation. Primers used to mutate the pRbs_BTR1(UBA1_Cc) plasmid are listed in Table 3-S2. The UBA(1) domain inserted at the N-terminus of iso-1-Cyt*c* is 40 residues long. The UBA(1) domain has a native histidine at position 31, which was mutated to asparagine, to create a pseudo-wild type (pWT) background for creation of our single engineered histidine variants. Thus, the pRbs_BTR1(UBA1_Cc) plasmid carrying the gene for the pWT variant was used as a template to produce T7H, S11H, E15H, R24H, Y27H, and E35H variants of UBA(1) – iso-1-Cyt*c* using the QuikChange Lightning kit (Agilent). The wild type His31 construct was used for His-heme loop formation studies, too. A Tyr→Gln mutation at position 27, to examine role of a local nonpolar side chain on the DSE properties near the second turn of UBA(1) was also prepared. All sites for the introduction of histidine side chains were chosen based on their high solvent accessibility using an accessible surface area algorithm from the Sealy Center for Structural Biology, University of Texas Medical Branch (http://curie.utmb.edu/getarea.html). All mutations were introduced using a PCR-based site-directed mutagenesis kit (QuikChange Lightning, Agilent). The gene sequences

for each variant were confirmed by dideoxy sequencing (Eurofins Genomics, Louisville, Kentucky).

Expression and Purification of UBA(1) – iso-1-Cyt*c* Variants. Expression and purification was carried out as previously described for UBA(2)-iso-1-Cyt*c* for the UBA(1)-iso-1-Cyt*c* fusion protein[82]. Briefly, the wild type His31, His31_Y27Q, pWT, and the six single histidine variants created in the pWT background, were expressed from the pRbs_BTR1(UBA1_Cc) plasmid following transformation into ultra BL21 (DE3) *E. coli* competent cells (New England Biolabs, Ipswich, MA) using the manufacturer's protocol. Sterile 2xYT bacterial media (1 L in 2.8 L Fernbach flasks) were inoculated with 0.5 mL of cells suspended from L-ampicillin plates. A 1.0 mL aliquot of 100 mg/mL ampicillin stock was added to each Fernbach flask. Flasks were placed in an orbital shaker at 150 rpm and 37 °C for 30 hours. Media was spun down at 5,000 rpm for 10 min (Sorvall Lynx 6000, F12 rotor). Cell pellets were used immediately or stored at -80 °C.

Protein extraction and purification has been previously documented[38-41,82]. In summary, cell pellets were thawed and suspended in 50 mM Tris, 500 mM NaCl, 1 mM EDTA, pH 8.0, 500 µL of 100 mM PMSF was added and the cells lysed using a Qsonica Q700 sonicator. Lysate was spun for 30 min at 10,000 rpm (Sorvall Lynx 6000, F14 rotor). The cleared lysate was adjusted to 8% ammonium sulfate and allowed to equilibrate with stirring overnight at 4 °C. The solution was subsequently spun for 30 min at 10,000 rpm (Sorvall Lynx 6000, F14 rotor) to remove precipitated protein impurities, followed by dialysis against two changes of MilliQ water containing 1 mM Na_2EDTA and 1 mM β-mercaptoethanol (βME). Sodium phosphate monobasic (50 mM) was used to adjust the dialysate to pH 6.0, followed by batch loading onto 100 mL of CM-sepharose (GE Healthcare Life Sciences) to bind UBA(1) – iso-1-Cyt*c* variants to the resin. After loading the resin into a glass column, it was washed with MilliQ water to remove unbound

sample then eluted using a 200 mL linear gradient (0 – 0.8 M NaCl in 50 mM sodium phosphate pH 6.0, 1 mM Na_2EDTA, 2 mM βME). Protein solution was exchanged into 25 mM sodium phosphate, pH 6.0, 1 mM Na_2EDTA by centrifuge ultrafiltration. Protein was used immediately or flash frozen in liquid N_2 and stored at -80 ^oC until further use.

Before experiments, UBA(1) – iso-1-Cyt*c* variants were purified using a HiTrap SP HP 5.0 mL column coupled to an ÄKTAprime plus chromatography system (GE Healthcare Life Sciences). Variants were oxidized with $K_3[Fe(CN)_6]$, followed by Sephadex G25 (GE Healthcare Life Sciences, superfine grade) chromatography with running buffer appropriate for each experiment, to remove the ferricyanide.

Purified fusion proteins were characterized by SDS-PAGE and matrix-assisted laser-desorption time-of-flight (MALDI-TOF) mass spectrometry (Bruker microflex mass spectrometer, Table 3-S3), to assess purity and to ensure no cleavage of the hybrid polypeptides had occurred. This characterization was done before and after each experiment.

Guanidine Hydrochloride Unfolding of UBA(1) – iso-1-Cyt*c* Variants. The His31, His31_Y27Q, pWT, and six single histidine variants in pWT background were oxidized and exchanged into CD buffer (20 mM Tris, 40 mM NaCl, 1 mM Na_2EDTA, pH 7.0) as described above. Protein concentration and degree of oxidation were evaluated using absorbance at 339 nm (20.9×10^3 M^{-1} cm^{-1}), 526.5 nm (11.0×10^3 M^{-1} cm^{-1}), 541.8 nm (9.0×10^3 M^{-1} cm^{-1}), 550 nm (oxidized, 9.1×10^3 M^{-1} cm^{-1}), 550 nm (reduced 28.0×10^3 M^{-1} cm^{-1}).[42] The concentration of the ~6 M GuHCl stock in CD buffer was evaluated from refractive index measurements.[43] A 4 μM solution of the UBA(1) – iso-1-Cyt*c* in concentrated GuHCl was titrated into a 4 μM solution of UBA(1) – iso-1-Cyt*c* in CD buffer using a Hamilton Microlab 500 Titrator coupled to an

Applied Photophysics Chirascan CD Spectrophotometer. Ellipticity was monitored at 222 nm (α-helix) and 250 nm (background). $\theta_{222corr}$ (= θ_{222nm} - θ_{250nm}) was plotted versus GuHCl concentration and fit to eq 3-3,

$$(3-3)\ \theta_{222corr} = \theta_D + m_D[GuHCl] + \frac{\theta_N + m_N[GuHCl] - \theta_D - m_D[GuHCl] + \left[(\theta_I - \theta_D - m_D[GuHCl])e^{\frac{m_{NI}[GuHCl] - \Delta G_{NI}^{o\prime}(H2O)}{RT}}\right]}{\left[1 + \left(e^{\frac{m_{NI}[GuHCl] - \Delta G_{NI}^{o\prime}(H2O)}{RT}}\right)\left(1 + e^{\frac{m_{ID}[GuHCl] - \Delta G_{ID}^{o\prime}(H2O)}{RT}}\right)\right]}$$

Equation 3-3 describes 3-state unfolding, where the free energy of unfolding, ΔG_u, is linearly dependent on GuHCl concentration.[44, 45] In eq 2, θ_N and m_N are the intercept and the slope of the native state baseline, θ_I is the intermediate state baseline, θ_D and m_D are the intercept and the slope of the denatured state baseline, m_{NI} is the rate of change of the free energy of the native to intermediate state transition (ΔG_{NI}) with respect to GuHCl concentration, m_{ID} is the rate of change of the free energy of the of intermediate to denatured state transition (ΔG_{ID}) with respect to GuHCl concentration and $\Delta G_{NI}^{o'}(H_2O)$ and $\Delta G_{ID}^{o'}(H_2O)$ are the free energies of the native to intermediate and intermediate to denatured state transitions, respectively, in the absence of denaturant. All variants were fit assuming that the native baseline is invariable with GuHCl concentration because the native baselines were too short to allow reliable evaluation of m_N. Reported parameters are the average and standard deviation of at least three independent trials.

His-Heme Loop Formation Equilibria by Denatured State pH Titration. Histidine-heme loop equilibria were monitored using UV-Vis spectroscopy with a Beckman DU 800 spectrophotometer for every UBA(1) – iso-1-Cyt*c* variant. Denatured state pH titrations were monitored at room temperature (22 ± 1 °C) with 3 μM denatured protein in 15 mM NaCl, 5 mM Na_2HPO_4, 1 mM Na_2EDTA (1x buffer) at 6 M or 4 M GuHCl concentration as described in

chapter 2. The absorbance band at 398 nm, A_{398}, was used to follow His-heme loop formation using absorbance at 450 nm, A_{450}, as a baseline. Plots of $A_{398corr}$ (= $A_{398} - A_{450}$) versus pH were fit to eq. 3-4 to evaluate the apparent pK_a, pK_a(obs).

$$(3-4)\ \ A_{398\mathrm{corr}} = \frac{A_{LS} + A_{HS} \times 10^{n_p[pK_a(obs) - pH]}}{1 + 10^{n_p[pK_a(obs) - pH]}}$$

In eq. 3-3 n_p is the number of protons linked to the His-heme loop formation, A_{LS} is $A_{398corr}$ at high pH when the heme is low spin with histidine bound to the heme iron and A_{HS} is $A_{398corr}$ at low pH when the heme is high spin with H_2O bound to the heme iron.

Stopped-Flow Loop Breakage Measurements. Loop breakage measurements were performed by pH jump stopped-flow using an Applied Photophysics SX20 stopped-flow apparatus at 25 °C. As described in Chapter 2, prior to stopped-flow, all UBA(1) – iso-1-Cyt*c* variants were oxidized with $K_3[Fe(CN)_6]$, and transferred into MOPS buffer (10 mM MOPS, 40 mM NaCl, 2 mM Na_2EDTA, pH 6.8) using Sephadex G-25 chromatography as described above. Protein stock concentration was determined as described above. The exact concentrations of 6 M or 8 M GuHCl stocks in MOPS buffer pH 6.8 were determined as described above. Denatured 6 μM protein solutions were prepared at 4 M or 6 M GuHCl in 10 mM MOPS, 2 mM Na_2EDTA, pH 6.8 for downward pH jump experiments. These solutions were mixed 1:1 with 4 M or 6 M GuHCl, respectively, in 100 mM citrate at pH 3 or pH 3.5 to yield a final protein concentration of 3 μM, using the stopped-flow spectrophotometer. For the pWT variant, the starting pH for downward pH jump experiments was 8 (20 mM Tris, 2 mM Na_2EDTA, pH 8.0). The pH after mixing was assessed by measuring pH after 1:1 manual mixing of the GuHCl citrate and GuHCl MOPS buffer solutions. The stopped-flow dead time determined by reduction of

dichlorophenolindophenol as a function of L-ascorbic acid concentration was 2 ms.[47] Data were fit to a single exponential equation after adding 0.002 s to the time at each data point.

RESULTS

Unfolding of UBA(1) – iso-1-Cyt*c* Variants by Guanidine Hydrochloride. Global unfolding thermodynamics of pWT and all single histidine variants of UBA(1) – iso-1-Cyt*c* variants were measured by CD at 25 °C and pH 7.0 using GuHCl as denaturant. Figure 2 shows the denaturation curves, $\theta_{222corr}$ versus GuHCl concentration, for

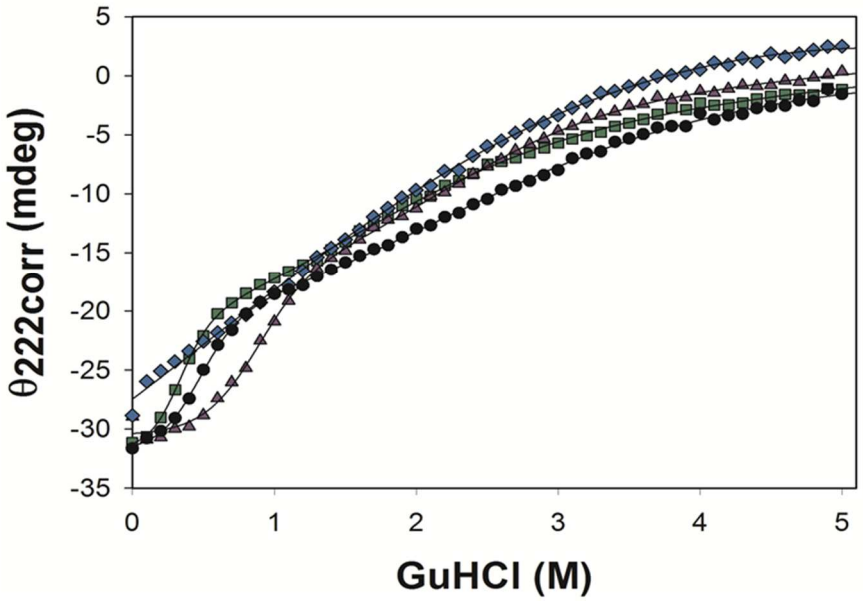

Figure 3-2. GuHCl denaturation curves for UBA(1) – iso-1-Cyt*c* variants, pWT (purple triangles), T7H (black circles), R24H (green squares), and His31 (blue diamonds) shown as plots of corrected ellipticity at 222 nm, $\theta_{222corr}$, versus GuHCl concentration. Solid curves are fits to eq 2 in Experimental Procedures. Parameters obtained from the fits are given in Table 1. Experiments were performed at 25 °C and pH 7.

Table 3-1. Thermodynamic Parameters for GuHCl Unfolding of UBA(1) – iso-1-Cyt*c* at 25 ºC and pH 7.0.

UBA(1) – iso-1-Cyt*c* Variants[a]	m_{NI}, kcal mol^{-1} M^{-1}	$\Delta G_{NI}^{o'}(H_2O)$, kcal mol^{-1}	C_{mNI}, M	m_{ID}, kcal mol^{-1} M^{-1}	$\Delta G_{ID}^{o'}(H_2O)$, kcal mol^{-1}	C_{mID}, M
pWT	2.6 ± 0.4	2.3 ± 0.4	0.91 ± 0.02	0.9 ± 0.09	2.1 ± 0.18	2.3 ± 0.4
E35H (25)	6 ± 1	2.4 ± 0.6	0.42 ± 0.03	0.8 ± 0.05	1.8 ± 0.14	2 ± 0.05
[b]His31 (29)	-	-	-	-	-	-
His31_Y27Q (29)	4.4 ± 0.7	1.1 ± 0.5	0.23 ± 0.07	0.9 ± 0.007	2.2 ± 0.036	2 ± 0.02
Y27H (33)	5 ± 1	3.1 ± 0.5	0.64 ± 0.007	0.8 ± 0.07	1.8 ± 0.18	2 ± 0.04
R24H (45)	4.8 ± 0.5	1.7 ± 0.2	0.43 ± 0.01	0.7 ± 0.2	0.98 ± 0.25	1 ± 0.05
E15H (51)	4.1 ± 0.7	2.01 ± 0.9	0.5 ± 0.1	1.0 ± 0.2	0.6 ± 0.1	0.7 ± 0.03
[c]S11H (49)	-	-	-	-	-	-
T7H (53)	4.8 ± 0.7	2.3 ± 0.4	0.50 ± 0.04	0.4 ± 0.2	0.7 ± 0.4	1.8 ± 0.3

[a]Denatured state His-heme loop size is given in brackets. [b]His31 fit to two-state model with following parameters: $\Delta G_u = 0.63 \pm 0.06$, m = 0.52 ± 0.06, $C_m = 1.22 \pm 0.13$. [c]S11H unable to be fit.

pWT, and three of the single histidine variants (T7H, R24H, and His31). The UBA(1) – iso-1-Cyt*c* fusion proteins show 3-state equilibrium unfolding titration curves as previously observed with the UBA(2) – iso – 1-Cyt*c* fusion protein[82]. The parameters from fits to a 3-state equilibrium unfolding model (eq 3-3, Experimental Procedures) are listed in Table 3-1. We assigned the thermodynamic parameters to unfolding of the UBA(1) and iso-1-Cyt*c* domains based on previous published GuHCl denaturation experiments on the individual domains.[32, 48] In previous work with the UBA(2)-Cyt*c* fusion protein, the Cyt*c* domain unfolded first followed by the UBA domain[82]. Iso-1-Cyt*c* with all histidines removed except His18 has an unfolding midpoint near 1 M GuHCl and an *m*-value of 3.8 to 4.0 kcal/mol-M, close to the m_{NI} values in Table 3-1.[32] The UBA(1) and UBA(2) domains yield *m*-values near 1.2 kcal/mol^{-1}M^{-1} [48], and these values are close to the m_{ID} values in Table 3-1. As observed for UBA(2)-Cyt*c*, the iso-1-Cyt*c* domain denaturant *m*-value for the N→I transition, m_{NI}, increased significantly, relative to

the pWT variant, when surface histidines are introduced into the UBA(1) domain. Each single histidine substitution causes a decrease in the native to intermediate unfolding midpoint, C_{mNI}, by 0.3 – 0.7 M, while the free energy of unfolding in the absence of denaturant, $\Delta G_{NI}^{o'}(H_2O)$ shows little change. Therefore, while single histidine substitutions in the UBA(1) increase the solvent exposed surface area of the Cyt*c* domain, the impact on the global stability of iso-1- Cyt*c* domain is, in general, small.

In addition, many of the single histidine mutations on the UBA(1) domain appear to have little to no effect on the overall stability of UBA(1). In the pWT variant, the intermediate to denatured unfolding midpoint, C_{mID}, is 2.33 M, and a majority of the C_{mID} values for the single histidine variants are in this range. The GuHCl unfolding *m*-values for the UBA(1) domain, m_{ID}, remains more or less constant for most of these variants, but decreases for R24H and T7H when compared to pWT. The cumulative effect is the free energy of unfolding in the absence of denaturant for the UBA(1) domain, $\Delta G_{ID}^{o'}(H_2O)$, except for the R24H, E15H and the T7H variants, are not strongly impacted by the single histidine substitutions to the UBA(2) domain. The parameters in Table 3-1 were used to calculate the population of denatured protein for each variant at 4 M GuHCl and 6 M GuHCl. All variants (including the pWT variant) are at least 95% unfolded at 6 M and 4 M GuHCl.

Equilibrium Denatured State His-Heme Loop Formation with UBA(1) – iso-1-Cyt*c* Variants. Denatured state His-heme loop formation equilibria are used to interrogate denatured state thermodynamic properties along the primary structure of the three-helix bundle, UBA(1). The His-heme methodology has been previously described in Chapter 2. Briefly, the method requires heme, and a single engineered histidine within the fold under study, to form a His-heme loop of a particular size. As with our previous work on the UBA(2)-Cyt*c* fusion protein, we

applied our His-heme loop formation method to UBA(1), which like UBA(2), does not have a *c*-type heme. We fused the UBA(1) domain near the N-terminus of iso-1-Cyt*c* (Figure 3-1). Under denaturing conditions only an engineered histidine in UBA(1) is available, because all the native histidines to iso-1-Cyt*c* have been replaced. The closest point of heme attachment in the iso-1-Cyt*c* domain fused to the UBA(1) domain is Cys14 (see Figures 3-1 and 3-3). Thus, every loop contains 16 residues from the iso-1-Cyt*c* amino acid sequence. With loop sizes less than 16, chain stiffness effects become important.[49] Therefore, for evaluating the UBA(1) domain with respect to random coil properties in the denatured state, insertion between Lys(-2) and Phe(-3) is ideal. Our lab has used this insertion site for previous work on loop formation in the denatured state with homopolymeric amino acid sequences,[29, 50-52] providing straightforward comparison to these data.

Under denaturing conditions, Met80 is a weak ligand for Fe^{3+},[53] being displaced by water, thus permitting the engineered histidine (or a lysine in the pWT variant[54]) to bind to heme Fe^{3+}. Because histidine is a side chain that can be protonated, the relative stability of a His-heme loop can be determined with a denatured state pH titration, to yield an apparent pK_a, pK_a(obs), and n_p, the number of protons involved in His-heme loop formation (Figure 3-3).

Figure 3-4 shows denatured state pH titration curves for His31, T7H, and the E35H variants of UBA(1) – iso-1-Cyt*c* in 6 M GuHCl. At pH 7, the His-heme loops are formed (>pH 7 for the His-heme loop of T7H) and the Fe^{3+} in the heme is in the low spin state. As pH decreases, a blue shift occurs as the Fe^{3+}-heme transitions from a low (strong field ligand, His or Lys) to a high spin state (weak field ligand, water). Denatured state titration curves are fit to a modified form of the Henderson-Hasselbach equation (eq 3-4 in Experimental Procedures), to determine the relative stability of the loops, pK_a(obs), and the number of protons, n_p, involved in loop

formation (Table 3-2). The pWT variant has pK_a(obs) near 7.3 at 6 M GuHCl and 6.7 at 4 M GuHCl, setting an upper limit for determination of the pK_a(obs) for His-heme loop formation in the denatured state. For single histidine variants, the pK_a(obs) is lowest for E35H, with pK_a(obs) at 6 M GuHCl equal to 4.93, and pK_a(obs) at 4 M GuHCl equal to 4.64, respectively (i.e. most stable His-heme loop). Data were fit to eq 3-4 in Experimental Procedures (solid curves).

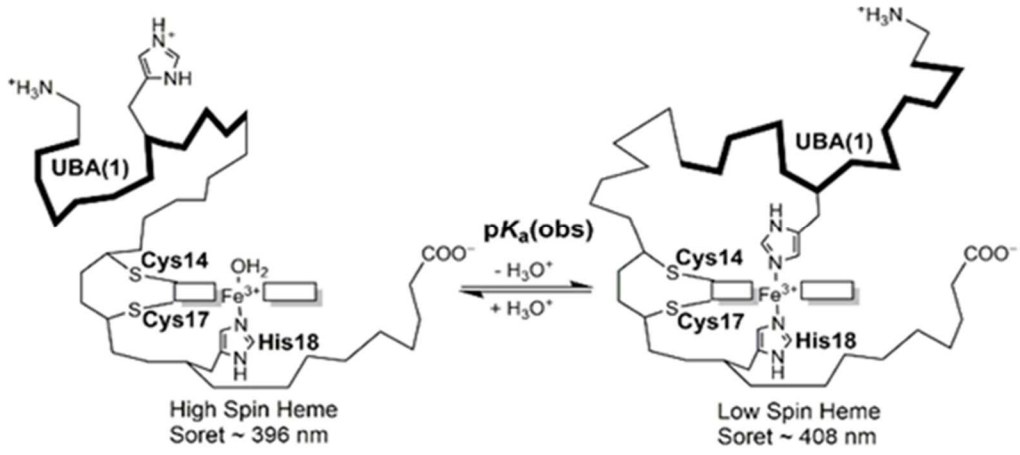

Figure 3-3. Denatured state histidine-heme loop formation schematic for UBA(1)-iso-1-Cyt*c*, showing His-heme loop formed at high pH (right), and loop broken at low pH (left).

Thermodynamic parameters obtained from the fits are given in Table 3-2. In general, as loop size increases, pK_a(obs) increases. As expected from the requirement that one proton must dissociate from histidine for the His-heme loop to form (Figure 3-3), n_p is also near 1 (Table 3-2). In this way, longer His-heme loops are less stable under both denaturing conditions.

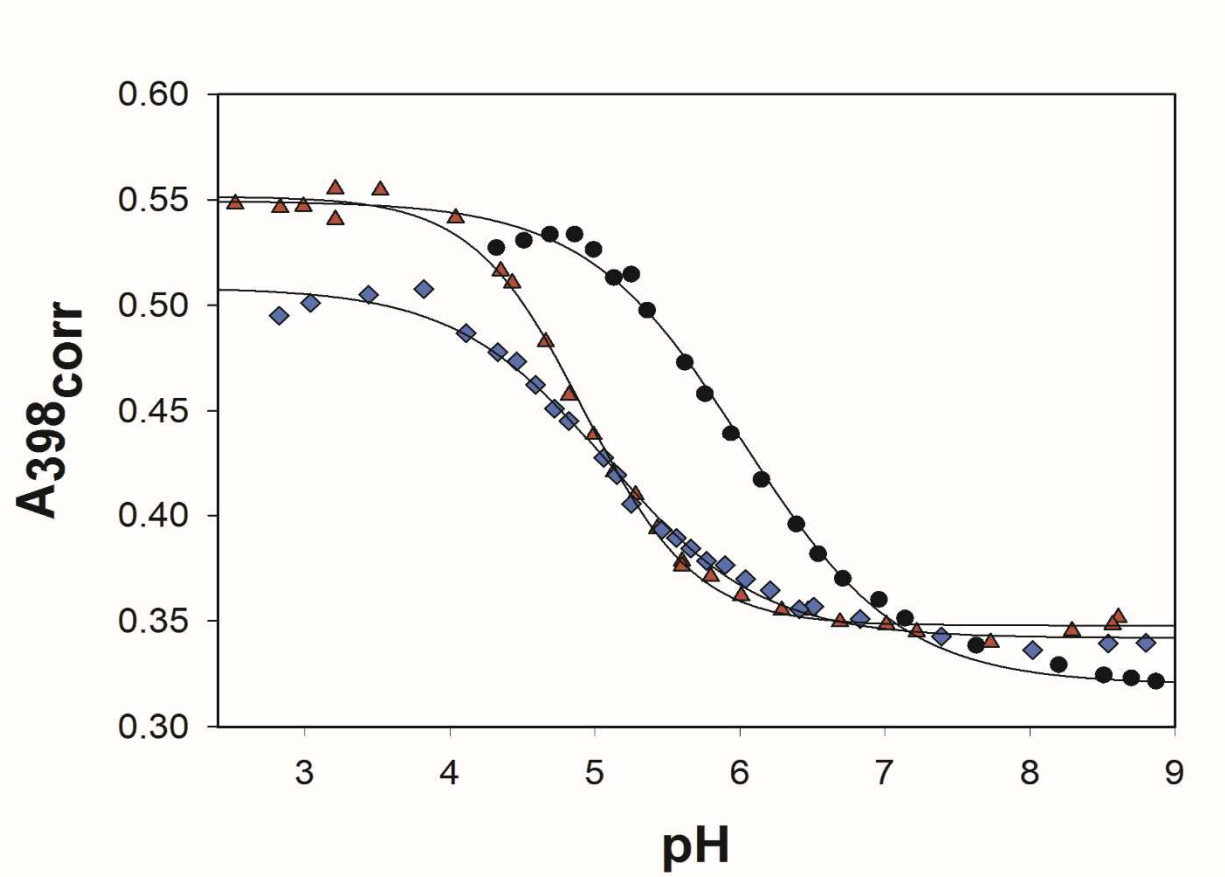

Figure 3-4. Selected curves for denatured state His-heme loop formation equilibria for UBA(1) – iso-1-Cyt*c* variants at 6 M GuHCl. Plots of $A_{398corr}$ versus pH for the T7H, E35H, and His31 variants are shown with black circles, red triangles, and blue squares.

Table 3-2. Thermodynamic Parameters for Denatured State Loop Formation of UBA(1) – iso-1-Cyt*c* Variants in 6 M and 4 M GuHCl Solutions at 22 ± 1˚C

Variant	Loop size (n)	pK_a(obs)	n_p	pK_{loop}(His)
		6 M GuHCl		
pWT	-	7.3 ± 0.04	0.78 ± 0.06	-
E35H	25	4.93 ± 0.01	1.09 ± 0.05	-1.66 ± 0.01
His31	29	5.14 ± 0.007	0.89 ± 0.05	-1.45 ± 0.007
His31_Y27Q	29	5.08 ± 0.01	1.06 ± 0.01	-1.51 ± 0.01
Y27H	33	5.99 ± 0.09	0.86 ± 0.01	-0.6 ± 0.09
R24H	36	5.75 ± 0.01	1.05 ± 0.03	-0.84 ± 0.01
E15H	45	5.18 ± 0.04	0.82 ± 0.04	-1.41 ± 0.04
S11H	49	5.97 ± 0.04	0.94 ± 0.04	-0.63 ± 0.04
T7H	53	6.03 ± 0.01	0.87 ± 0.07	-0.56 ± 0.01
		4 M GuHCl		
pWT	-	6.69 ± 0.04	0.79 ± 0.04	-
E35H	25	4.64 ± 0.04	1.08 ± 0.08	-1.45 ± 0.04
His31	29	4.58 ± 0.08	0.95 ± 0.13	-2.01 ± 0.08
His31_Y27Q	29	4.53 ± 0.04	0.89 ± 0.05	-2.06 ± 0.04
Y27H	33	5.51 ± 0.04	1.15 ± 0.05	-1.08 ± 0.04
R24H	36	5.48 ± 0.02	0.99 ± 0.04	-1.11 ± 0.02
E15H	45	4.99 ± 0.04	0.88 ± 0.01	-1.60 ± 0.04
S11H	49	5.66 ± 0.03	0.95 ± 0.04	-0.93 ± 0.03
T7H	53	5.72 ± 0.04	1.09 ± 0.05	-0.87 ± 0.04

Histidine-heme loop formation equilibria is a two-step process—ionization of the histidine, pK_a(HisH$^+$), and binding of the deprotonated histidine to the heme, pK_{loop}(His). This two-step process is represented by eq 3-5.

(3-5) $$pK_a(obs) = pK_a(HisH^+) + pK_{loop}(His)$$

Therefore, subtracting pK_a(HisH$^+$) (= 6.6)[46] from each pK_a(obs) value in Table 3-2 allows calculation of pK_{loop}(His) for every loop size (Table 3-2). In general, pK_{loop}(His) decreases in negative magnitude (i.e. less stable His-heme loop) as loop size increases.

His-Heme Loop Breakage Kinetic Measurements in the Denatured State for UBA(1) – iso-1-Cyt*c* Variants. We have previously shown that His-heme loop breakage and formation kinetics follows a model with rapid deprotonation of histidine followed by binding of the ionized histidine to the heme.[55] Therefore, the observed rate constant, k_{obs}, depends on pH and is governed by eq 3-6,[46] where k_b and k_f are the rate constants for loop breakage and

$$k_{obs} = k_b + k_f\left(\frac{K_a(\mathrm{HisH^+})}{K_a(\mathrm{HisH^+}) + [\mathrm{H^+}]}\right) \qquad (3\text{-}6)$$

loop formation, while K_a(HisH$^+$) is the dissociation constant for the deprotonation of histidine. When pH<<pK_a(HisH$^+$), $k_{obs} \approx k_b$. Histidine-heme loop breakage kinetics (and Lysine-heme loop breakage kinetics for pWT) at 6 M and 4 M GuHCl solutions were evaluated by pH jump stopped-flow at 25 °C. Histidine-heme loop breakage adheres to single exponential kinetics (Figure 3-S1). Experiments carried out at pH 3.5 and 3.0 confirm a pH regime where $k_{obs} \approx k_b$ had been reached (Table 3-S4). The k_{obs} values were similar at both pH values implying k_{obs} had reached the lower boundary of $k_{obs} \approx k_b$.

The loop breakage data for the UBA(1) and UBA(2) domains with His-heme loops forming between the residues in the UBA domains and the heme of iso-1-Cyt*c* in 6 M GuHCl solution are shown in Figure 3-5. Comparison with k_b values for poly(Ala) His-heme loops between Phe(-2) and Lys(-3) in the Cyt*c* domain is included[29]. As previously observed with the UBA(2)-Cyt*c* fusion protein, the magnitude of k_b increases and decreases substantially versus loop size. As

depicted in Figure 3-5, polyalanine loops inserted at the N-terminus of iso-1-Cyt*c* indicate that as loop size increases, k_b plateaus to a constant value, thus behaving as loops modeling random coil behavior.[29] Therefore, this irregular behavior in k_b values observed for these single histidine variants of UBA – iso-1-Cyt*c* fusion proteins corresponds to regions of residual or non-random structure in the denatured state along UBA domain primary structure. The most persistent His-heme loops in 6 M GuHCl are the E27H variant in UBA(2)-Cyt*c* (~46 s^{-1}), and the His31 (~29 s^{-1}) and His15 (~25 s^{-1}) variants in the UBA(1) domain.

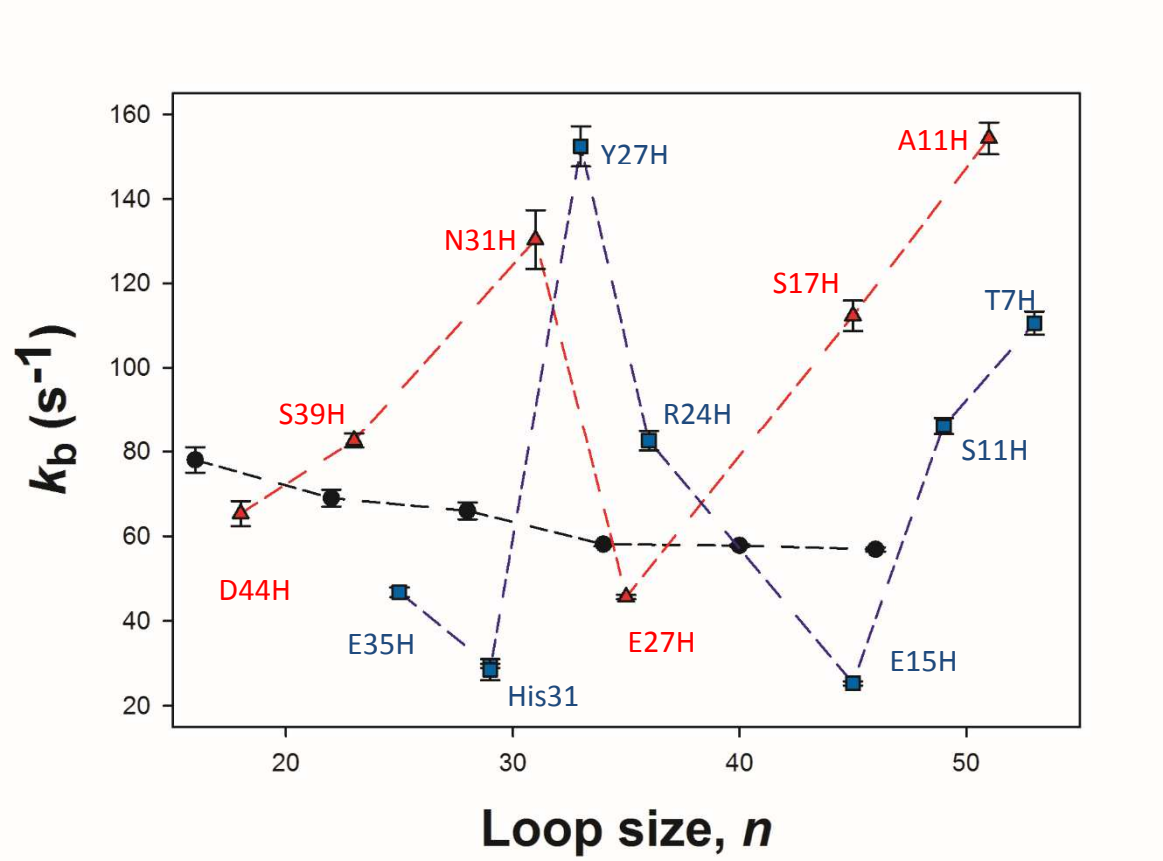

Figure 3-5. Loop breakage kinetics versus loop size, *n*, at 25 °C for UBA(1) – iso-1-Cyt*c* (blue squares) variants, UBA(2) – iso-1-Cyt*c* (red triangles) variants, and homopolymeric inserts of alanine $(KAAAAA)_n$[29] (n = 1 – 5, black circles) in 6 M GuHCl.

DISCUSSION

UBA(1) Single Histidine Mutation Effects On Stability of the UBA(1) – iso-1-Cyt*c* Fusion Protein. We previously fused the UBA(2) domain to the Cyt*c* domain, and upon introducing single surface exposed histidines in UBA(2), discovered they caused a significant destabilization of the iso-1-Cyt*c* domain[82]. The net effect on stability for UBA(1) fused to iso-1-Cyt*c* on destabilization of the iso-1-Cyt*c* domain is less clear. The pWT produces C_{mNI} = 0.9 M for the iso-1-Cyt*c* domain, while all variants with single histidines engineered into the UBA(1) domain have C_m values less than pWT (0.23 to 0.55) (Figure 3-2, Table 3-1). Thus, introduction of histidine residues placed at solvent accessible sites in UBA(1) domain appear to destabilize the iso-1-Cyt*c* domain. However, because of the increase in m_{NI} when histidines are present in UBA(1) the effect in $\Delta G_{NI}^{o\prime}(H_2O)$ is small. As for UBA(2)-Cyt*c*, the impact of these single histidines on the global stability of the UBA(1) domain is relatively weak. The C_m for pWT UBA(1) variant is 2.33 M. Many of the single histidine variants have C_m values within error of pWT; the exceptions are the E15H and R24H variants, which have C_m values equal to 0.7 M and 1.4 M, respectively. The $\Delta G_{ID}^{o\prime}(H_2O)$ values for UBA(1) between the pWT and single-histidine variants are similar, too. The exceptions are R24H, E15H, and the T7H variants, where T7H and R24H have lower values for m_{ID} compared to the pWT variant. Interestingly, the native His31 wild type variant could not be fit to a 3-state equilibrium unfolding model (see Table 3-1). The native histidine at position 31 in UBA(1) destabilized the Cyt*c* domain to such an extent that most of the Cyt*c* domain was unfolded at 0 M GuHCl (see Fig. 3-2).

The C_{mID} for unfolding of UBA(1) is ~1.4 M greater than the C_{mNI} for the Cyt*c* domain in the single histidine variants. Similar to the UBA(2) domain, these UBA domains still appear to

be folded at the end of the first transition for biphasic unfolding. Therefore, folded UBA domains appear to be able to bind to the heme and thus unfold the iso-1-Cyt*c* domain. Engineering in single surface exposed histidines within UBA(1) has slight destabilization in the iso-1-Cyt*c* domain, and is more pronounced in His31 and His31_Y27Q variants. Moreover, because of similar values in m_{ID} and $\Delta G_{ID}^{o'}(H_2O)$ for the E35H, His31_Y27Q, Y27H variants to pWT, the His-heme binding appears to not affect the stability of UBA(1). For R24H and T7H, m_{ID} decreases, implying binding of a histidine in UBA(1) to heme of iso-1-Cyt*c* may impact the structure of UBA(1). The E15H variant has significantly decreased C_{mID} and $\Delta G_{ID}^{o'}(H_2O)$ compared to pWT, but its m_{ID} remains unchanged, suggesting an intact UBA(1) domain can remain folded but has a decrease in stability caused by its histidine substitution.

Denatured State His-Heme Loop Formation in 4 M and 6 M GuHCl in Comparison to Random Coil Behavior and the UBA(2) Domain. Loop stability is modeled using the Jacobson-Stockmayer equation (eq 3-2).[28] Assuming His-heme loops have random coil behavior and the enthalpy of Fe^{3+}-His bond formation is constant for each His-heme loop, the free energy of loop formation, $\Delta G_{loop}(His)$, can be expressed with eq 3-7.

$$\Delta G_{loop}(His) = ln(10)RTpK_{loop}(His) = -T\Delta S_{loop} \quad (3\text{-}7)$$

Using the expression for ΔS_{loop} in eq 3-2, eq 3-7 can be rewritten as eq 3-8,

$$pK_{loop}(His) = pK_{loop}(His)_{ref} + \nu_3 \log(n) \quad (3\text{-}8)$$

where ν_3 is the scaling exponent for loop formation and $pK_{loop}(His)_{ref}$ is $pK_{loop}(His)$ for a loop size equal to 1. Thus, $pK_{loop}(His)$ should be proportional the logarithm of loop size, n, with a

slope equal to ν_3. In a random coil with excluded volume, ν_3 is expected to have values ranging from 1.8 to 2.4.[58-60]

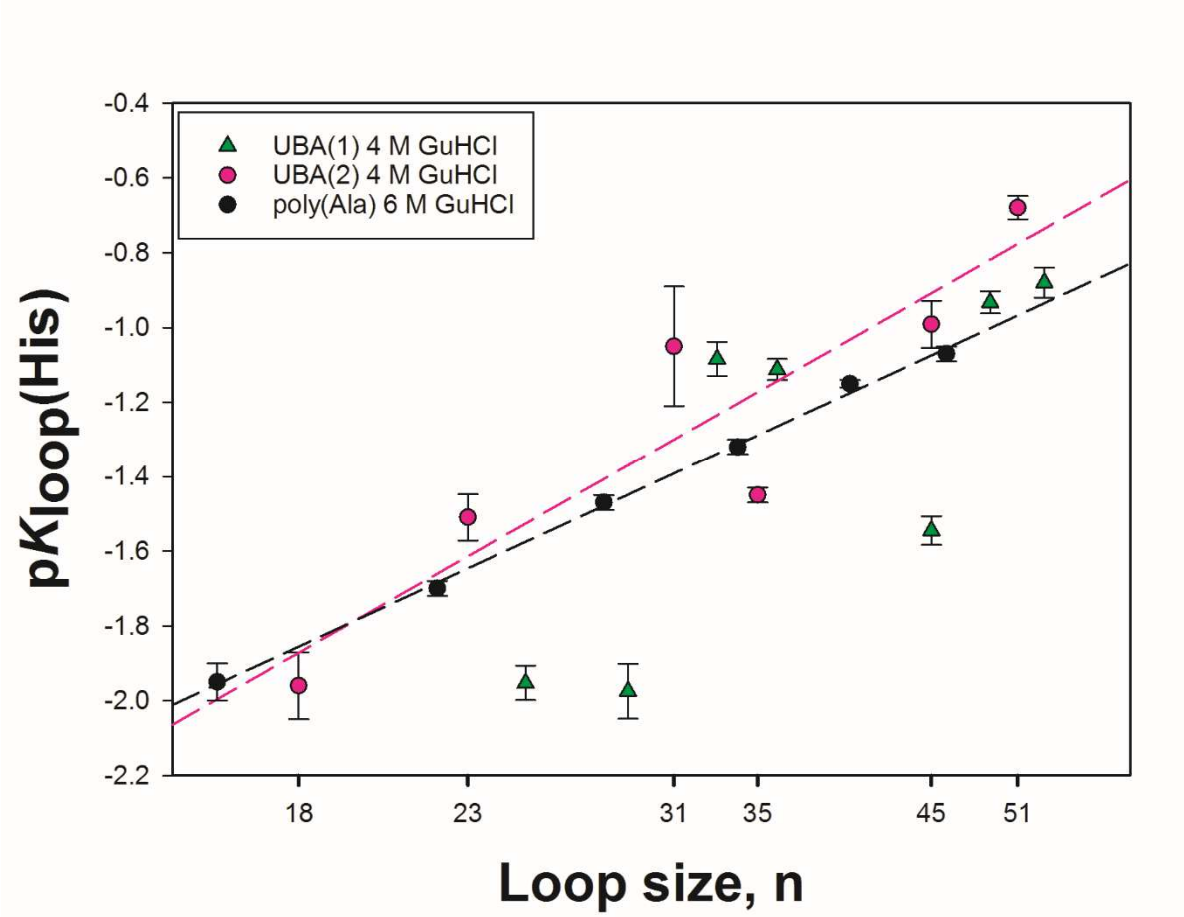

Figure 3-6. Plot of loop stability, pK_{loop}(His), vs. loop size (logarithmic scale) in the denatured state of the three-helix bundles UBA(1) and UBA(2) fused to N-terminus of iso-1-Cyt*c* for histidine variants in 4 M GuHCl. Homopolymeric inserts of alanine $(KAAAAA)_n$[29] inserted between Phe(-3) and Lys(-2) of iso-1-Cyt*c* as for the UBA domains acquired at 6 M GuHCl are displayed for comparison.

Figure 3-6 shows a plot of pK_{loop}(His) versus log of loop size, *n*, for UBA domain data acquired at 4 M GuHCl. Histidine-heme loop formation with homopolymeric sequences of

alanine inserted at the N-terminus of iso-1-Cyt*c* show close adherence to pK_{loop}(His) α log(n) (see Figure 3-6).[29, 51, 52] For the UBA domains, there exists irregular scatter about the best fit line to eq 3-8, implying segments of primary structure in the UBA domains have non-random behavior. Moreover, strong deviations from random coil behavior are observed for E15H and His31 variants in UBA(1). Previously in the UBA(2)-Cyt*c* fusion protein, it was observed that the E27H variant showed strong deviation from random coil behavior. For E27H in UBA(2), pK_{loop}(His) is 0.28 and 0.30 units more negative than expected from the best fit lines to eq 3-8 at 4 and 6 M GdnHCl[82]. Because there is considerably more scatter about the lines of best fit to eq 3-8 for UBA(1) at 6 M and 4 M GuHCl, we used the best fit line for UBA(2) to eq 3-8 at 6 M and 4 M GuHCl for comparing our UBA(1) thermodynamic data. The His31 and E15H variants are ~0.8 (E15H) and ~0.4 (His31) units more negative than expected when compared to the best fit line for UBA(2) at 6 M GuHCl. When comparing these variants to the poly(Ala) best fit line to eq 3-8 at 6 M GuHCl, only E15H deviates from random coil by ~0.3 units, respectively. In 4 M GuHCl, we observe similar deviation from random coil behavior as well. Using the UBA(2) line of best fit to eq 3-8 at 4 M GuHCl, the His15-heme and His31-heme loops are ~0.6 units more negative, thus the His31-heme and His15-heme loops deviate from random coil behavior by ~0.8 kcal/mol, respectively [$\Delta\Delta G_{loop}$(His) = ln(10)RTΔpK_{loop}(His)]. Figure 3-7 shows this similar patterning of deviation from random coil behavior for E15H and His31 at 6 M GuHCl.

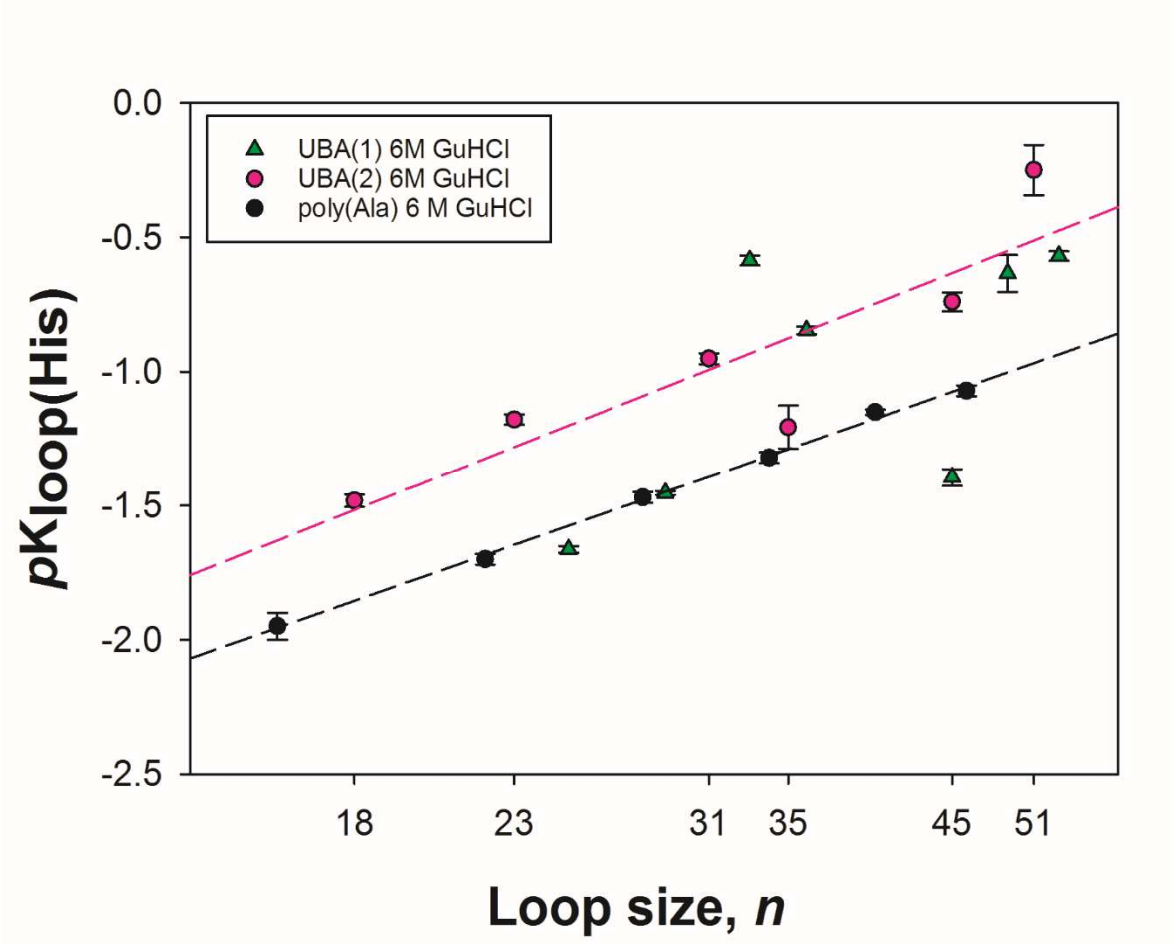

Figure 3-7. Plot of loop stability, pK_{loop}(His), vs. loop size (logarithmic scale) in the denatured state of the three-helix bundles UBA(1) and UBA(2) fused to N-terminus of iso-1-Cyt*c* for histidine variants in 6 M GuHCl. Homopolymeric inserts of alanine $(KAAAAA)_n$[29] inserted between Phe(-3) and Lys(-2) of iso-1-Cyt*c* as for the UBA domains acquired at 6 M GuHCl are displayed for comparison.

The loop breakage data in Figure 3-5 vary in an irregular fashion versus loop size. Small rate constants for loop breakage, k_b, for the E27H variant in the UBA(2) domain was previously observed[82]. For UBA(1), we discover similar behavior for the His31 and E15H variants, where these His-heme loops contain the lowest loop breakage constants. These findings, coupled with

the unusually stable histidine-heme loops evident from thermodynamic data for these variants, indicate the primary structure in UBA(1) in these regions can stabilize persistent structure in the denatured state, relative to other portions of primary structure in UBA(1). For the E27H variant in UBA(2), this residue is part of the reverse turn between helix 2 and 3 of UBA(2)[82]. While in the four-helix bundle protein Cyt*c′*, persistent His-heme loops localize near the reverse turns between helices 1 and 2 and at the ends of the Ω-loop connecting helices 2 and 3.[31] We observe a similar phenomenon here, where both sequences near reverse turns in the UBA(1) domain are discovered to cause persistent structure in its denatured state. Previous findings with Cyt*c*' using molecular dynamic simulations had shown hydrophobic residues make persistent contacts for these His-heme loops at or near reverse turns. In UBA(2), we conjectured that Phe23 and Tyr24 may play a role in stabilizing the second turn between helix 2 and 3. In this study, we considered the possibility that aromatics near the second reverse turn of UBA(1) may be involved in stabilizing this persistent structure in the DSE. Therefore, we carried out a Tyr→Gln mutation in the second reverse turn of UBA(1), to ascertain whether this residue plays a role. As seen in Figure 3-4, His31, there are two data points for His31 loop breakage. The His31_Y27Q mutation and His31 k_b values are quite similar (~25 s^{-1}), as well as the pK_{loop} values at 6 M GuHCl (-1.5 for His31_Y27Q and -1.45 for His31), indicating Tyr27 has no significant impact on the stabilization in second reverse turn for UBA(1) in the denatured state (see Fig. 3-4).

Previous work has shown that an Ala→Trp mutation three residues from the histidine in the His-heme loop, stabilizes the loop by ~0.3 kcal/mol in 6 M GuHCl and ~0.7 in 3 M GuHCl.[50] When Alanine four residues from the histidine which forms the His-heme loop is replaced by Phe, Tyr or Trp, the His-heme loop is stabilized by 0.4 – 0.5 kcal/mol in 3 M GuHCl, whereas a Leu substitution leads to minimal stabilization of the loop.[61] A number of other studies show that

aromatic residues stabilize residual structure and long range interactions in the denatured state of proteins.[9, 10, 62-66] We propose that because the Tyr→Gln mutation is outside this persistent His-heme loop it may behave differently than stabilizing aromatics within the loop. A nearby Tyr in helix 3 just downstream of His31 is 5 residues away and may be a contributing factor to this stabilization.

Similar variation in the irregular trend in k_b is observed at 6 M and 4 M GuHCl. All His-heme loops show a decrease in loop breakage at 4 M GuHCl (when compared to 6 M GuHCl), some with more of a decrease in magnitude. For the His-heme loops probing the reverse turns in UBA(1), the loop breakage rate constant at 4 M GuHCl decreases to ~20 s^{-1}. As a comparison, the persistent His-heme loop in UBA(2) domain drops to ~29 s^{-1}.

As previously observed, for the three-helix bundle UBA(2) and the four-helix bundle Cyt*c′*, His-heme loops containing this persistent residual structure are not found for histidine probes near every reverse turn, suggesting that the denatured state does not need to be entirely biased towards all reverse turns to fold efficiently to their native structures. It is important to note that the UBA(2) domain has a Cys→Ala mutation at position 26, which is present in UBA(2)'s second reverse turn. This substitution may impact the conformational distribution in the denatured state, and may explain why the first reverse turn does not persist in the DSE for UBA(2). The histidine probing the first reverse turn in UBA(2) was positioned on the N-terminal side of UBA(2)'s first reverse turn. We observe that many of the persistent turns, including the persistent turns in UBA(1), contain histidine probes on the C-terminal side of reverse turns. Thus, it is possible that the first reverse turn of UBA(2) was not detected because a histidine probing the C-terminal side of the turn between helix 1 and 2 was not used to probe the DSE.

Biophysical properties of the Denatured State of UBA(1). The fits of the plot of pK_{loop}(His) versus Log(n) for UBA(1) yield scaling exponents, ν_3, of 3.2 ± 1.0 and 2.6 ± 1.1 at 4 and 6 M GuHCl, respectively. These values at 6 M and 4 M GuHCl are within the range for a random coil with excluded volume within error (1.8 – 2.4), however, there is considerable scatter about the line of best fit to eq. 3-8, and therefore, these scaling exponents may not be reliable. We therefore used UBA(2)'s best fit line to eq. 3-8 for comparison to our thermodynamic data for UBA(1) at 6 M and 4 M GuHCl (see Fig 3-6 & 3-7). Part of the reason why scatter is so large for the UBA(1) domain is that there are persistent loops near both reverse turns. Also, we were unable to express a His variant at the C-terminus of the UBA(1) domain. The UBA(2) domain had a small His-heme loop of 18, which primarily contains sequence from iso-1-Cyt*c*, and this may be a reason why scatter is not as large as with UBA(1). Moreover, we do observe consistency with the thermodynamic data with both UBA domains, correlating the most persistent His-heme loops to localize to turns. When UBA(1) pK$_{loop}$ data is compared to UBA(2)'s the line of best fit to eq. 7, significant persistence is observed with histidines that probe the turn regions. Previous studies have shown random coil scaling is compatible with local non-random structure.[69] Thus, these persistent loops near turns are indicative of non-random structure.

Previously it was observed for the UBA(2) domain's pK_{loop}(His) versus Log(n) plot in 6 M GuHCl (Figure 6-7) that loop stability is much lower compared to His-heme loops with homopolymeric inserts of similar length in 6 M GuHCl. The 4 M GuHCl UBA(2) pK_{loop}(His) versus Log(n) data corresponds well with poly(Ala) and poly(Gln) 6 M GuHCl data[82], indicating the UBA(2) domain may have a more expanded denatured state relative to poly(Gln) or poly(Ala) inserts. The pK_{loop}(His) values for 4 of the 7 histidines in the UBA(1) domain are near

the line of the UBA(2) domain suggesting a similar phenomenon for UBA(1) except for the reverse turns.

CONCLUSIONS:

Similar to the UBA(2) domain from the C-terminus of HHR23A, we observe global properties of the internal UBA(1) domain from HHR23A under denaturing conditions (4 and 6 M GuHCl) consistent with a random coil with excluded volume ($\nu_3 = 1.8 – 2.4$). Extensive scatter about the best fit line to the Jacobson-Stockmayer relationship shows that histidine at positions 15 and 31 in turns between the helices form persistent His-heme loops. This result is consistent with the observations for Cyt*c′* and UBA(2)—that persistent denatured state His-heme loops tend to occur at or near local sequence that frames reverse turns in their tertiary structures.[31,82] Therefore, in agreement with the four-helix bundle Cyt*c*' and three-helix bundle UBA(2), denatured state conformational biases in helix-bundles may localize to sequences framing reverse turns, which may help govern the gross topology of α-folds. For UBA(1), most persistent loops align with the UBA(2) domain, implying UBA(1) has a more expanded denatured state like UBA(2), relative to loop formation in homopolymeric sequences.

SUPPORTING INFORMATION

Tables 3-S1 & 3-S2 provide sequences for oligonucleotides used for cloning and mutagenesis. Table 3-S3 contains MALDI-TOF data for the UBA(1) – iso-1-Cyt*c* variants. Table 3-S4 shows loop breakage rate constants for the UBA(1) domain fused to Cyt*c* in 6 M and 4 M GuHCl at pH 3.0. Figure 3-S1 shows a typical kinetic trace for loop breakage with a UBA(1) – iso-1-Cyt*c* variant.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bruce.bowler@umontana.edu

ORCID

Bruce E. Bowler: 0000-0003-1543-2466

Funding

This research was supported by National Science Foundation grant, MCB-1412164 and NIH grant, R01GM074750, to B.E.B. The Bruker microflex MALDI-TOF mass spectrometer was purchased with Major Research Instrumentation Grant CHE-1039814 from the National Science Foundation. M.J.L. acknowledges the Sloan Indigenous Graduate Partnership of the Alfred P. Sloan Foundation and the National Science Foundation DEB 0614406 and the NSF EPSCoR Track-1 EPS-1101342 (INSTEP 3) for graduate education support.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

iso-1-Cyt*c*, iso-1-cytochrome *c*; CD, circular dichroism; GuHCl, guanidine hydrochloride; UBA(1), human HHR23A ubiquitin-associated domain 1; UBA(1) – iso-1-Cyt*c*, fusion protein with UBA(1) inserted near the N-terminus of iso-1-Cyt*c*; pWT, UBA(1)-iso-1-Cyt*c* carrying a H31N mutation in the UBA(1) domain.

Supporting Information

Three-Helix Bundle UBA Domains with Similar Native State Topology and Divergent Sequence Contain Denatured State Conformational Bias Bracketing Turns

Moses J. Leavens, Lisa E. Spang, Melisa M. Cherney, Michaela Finnegan and Bruce E. Bowler*

Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Dynamics, University of Montana, Missoula, Montana 59812

*To whom correspondence should be addressed.

Telephone: (406) 282-1883. Fax: (406) 243-4227. E-mail: bruce.bowler@umontana.edu

Table 3-S1. Oligonucleotide primers for preparation of UBA(1)-iso-1-Cyt*c* fusion protein[a]

Primer	Primer Sequence 5′ to 3′
UBA(1)ER1_for	d(ATATGAATTCGAGTATGAGACGATGCTG)
UBA(1)NgoM4_rev	d(TTAAGCCGGCCTTCCCAGGGATTCCCGTGAG)

[a]Restriction sites are underlined.

Table 3-S2. Oligonucleotide Primers for Preparation of UBA(1)-iso-1-Cyt*c* Variants[a]

Primer	Primer Sequence 5′ to 3′
pWT	d(gctacaacaaccccaaccgagccgtggag)
pWT-r	d(ctccacggctcggttggggttgttgtagc)
His31_Y27Q	d(ccctgagagccagccagaacaaccccccaccg)
His31_Y27Q-r	d(cggtgggggttgttctggctggctctcaggg)
T7H	d(ttcgagtatgagacgatgctgcatgagatcatgtccatgggctat)
T7H-r	d(atagcccatggacatgatctcatgcagcatcgtctcatactcgaa)
S11H	d(tgctgacggagatcatgcacatgggctatgagcgag)
S11H-r	d(ctcgctcatagcccatgtgcatgatctccgtcagca)
E15H	d(gagatcatgtccatgggctatcatcgagagcgggt)
E15H-r	d(acccgctctcgatgatagcccatggacatgatctc)
R24H	d(ggtcgtggccgccctgcatgccagctacaacaacc)
R24H-r	d(ggttgttgtagctggcatgcagggcggccacgacc)

Y27H	d(gccctgagagccagccacaacaaccccaaccga)
Y27H-r	d(tcggttggggttgttgtggctggctctcagggc)
E35H	d(ccgtgagcagatagtgcacggctcggtgg)
E35H-r	d(ccaccgagccgtgcactatctgctcacgg)

Table 3-S3. MALDI-ToF Mass Spectral Data for UBA(1) - iso-1-Cyt*c* variants

Variant	m/z observed	m/z expected[a]
His31 (WT)	17,452.12	17,453.65
H31N (pWT)	17,428.68	17,430.61
E35H	17,439.72	17438.64
Y27H	17,404.14	17,404.58
WT_Y27Q	17,419.56	17,418.6
R24H	17,410.09	17,411.56
E15H	17,439.78	17,438.64
S11H	17,478.23	17,480.67
T7H	17,467.55	17,466.65

[a]Expected m/z obtained using ExPASy PeptideMass tool (http://web.expasy.org/peptide_mass/) using average mass, no cutting and $[M+H]^+$ options.

Table 3-S4. Loop breakage rate constants for UBA(1)-Cyt*c* Variants in 6 M and 4 M GuHCl at pH 3.0.

Variant	Loop size	k_b pH 3 (6M)	k_b pH 3.5 (6M)	k_b pH 3 (4M)	k_b pH 3.5 (4M)
E35H	25	46.7 ± 1.1	-	-	-
H31N (pWT)	-	111.8 ± 5.5	-	107.6 ± 9.0	-
His_31	29	29.4 ± 0.5	-	19.8 ± 0.8	-
Y27H	33	152.4 ± 4.7	-	98.8 ± 2.1	-
WT_Y27Q	29	28.5 ± 2.5	-	23.8 ± 0.4	-
R24H	36	82.6 ± 2.3	-	49.2 ± 1.1	-
E15H	45	25.3 ± 0.5	-	20.4 ± 0.3	-
S11H	49	86.1 ± 2.0	-	46.4 ± 2.2	-
T7H	53	110.6 ± 2.7	-	72.5 ± 1.6	-

Loop breakage rate constants of at least nine trials showing breakage has gone to completion. The 6 M GuHCl data is uncorrected for viscosity for comparison to 6 M poly(A) data for iso-1-Cyt c. The 4 M data has been corrected for viscosity.

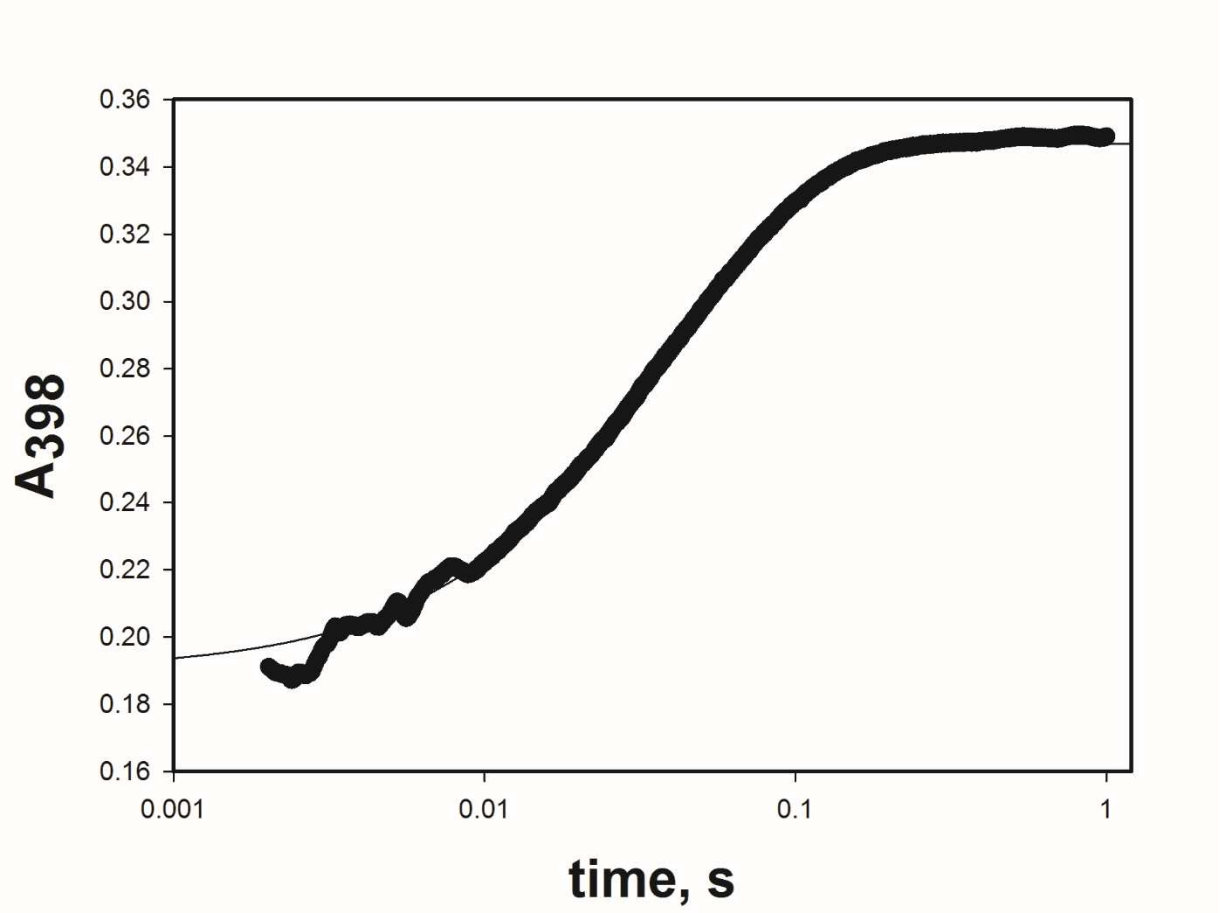

Figure 3-S1. Sample plot of histidine-heme loop breakage in the denatured state at 6 M GuHCl conditions for the His31_Y27Q UBA(1) – Cytc variant. Data collected using absorbance at 398 nm, A_{398}, while performing pH jump from pH 6.8 to pH 3. The solid curve is a fit of the data to a single exponential equation.

REFERENCES FOR UBA(1)-CYT C

1. Mittag, T., and Forman-Kay, J. D. (2007) Atomic-level characterization of disordered protein ensembles, *Curr. Opin. Struct. Biol. 17*, 3-14.

2. Cho, J.-H., Sato, S., Horng, J.-C., Anil, B., and Raleigh, D. P. (2008) Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability, *Arch. Biochem. Biophys. 469*, 20-28.

3. Bowler, B. E. (2012) Globular proteins: characterization of the denatured state, In *Comprehensive Biophysics* (Egelman, E., Ed.), pp 72-114, Academic Press, Oxford.

4. Bowler, B. E. (2007) Thermodynamics of protein denatured states, *Mol. BioSyst. 3*, 88-99.

5. Dill, K. A., and Shortle, D. (1991) Denatured states of proteins, *Annu. Rev. Biochem. 60*, 795-825.

6. Zhang, O., Kay, L. E., Shortle, D., and Forman-Kay, J. D. (1997) Comprehensive NOE characterization of a partially folded large fragment of staphylococcal nuclease D131D, using NMR methods with improved resolution, *J. Mol. Biol. 272*, 9-20.

7. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures, *J. Mol. Biol. 268*, 179-184.

8. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels, *J. Mol. Biol. 268*, 158-169.

9. Wirmer, J., Schloerb, C., Klein-Seetharaman, J., Hirano, R., Ueda, T., Imoto, T., and Schwalbe, H. (2004) Protein interactions: modulation of compactness and long-range interactions of unfolded lysozyme by single point mutations, *Angew. Chem. Int. Ed. 43*, 5780-5785.

10. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein, *Science 295*, 1719-1722.

11. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea, *Biochemistry 36*, 8977-8991.

12. Neri, D., Billeter, M., Wider, G., and Wüthrich, K. (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor, *Science 257*, 1559-1563.

13. Pakula, A. A., and Sauer, R. T. (1990) Reverse hydrophobic effects relieved by amino acid substitutions at a protein surface., *Nature 344*, 363–364.

14. Bowler, B. E., May, K., Zaragoza, T., York, P., Dong, A., and Caughey, W. S. (1993) Destabilizing effects of replacing a surface lysine of cytochrome *c* with aromatic amino acids: implications for the denatured state, *Biochemistry 32*, 183-190.

15. Bowler, B. E., Dong, A., and Caughey, W. S. (1994) Characterization of the guanidine hydrochloride-denatured state of iso-1-cytochrome *c* by infrared spectroscopy, *Biochemistry 33*, 2402-2408.

16. Herrmann, L., Bowler, B. E., Dong, A., and 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.

17. Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions, *Biochemistry 38*, 4896-4903.

18. Cho, J.-H., and Raleigh, D. P. (2005) Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins, *J. Mol. Biol. 353*, 174-185.

19. Cho, J.-H., Sato, S., and Raleigh, D. P. (2004) Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state, *J. Mol. Biol. 338*, 827-837.

20. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M., and Pace, C. N. (1999) Increasing protein stability by altering long-range coulombic interactions, *Protein Sci. 8*, 1843-1849.

21. Pace, C. N., Alston, R. W., and Shaw, K. L. (2000) Charge-charge interactions influence the denatured state ensemble and contribute to protein stability, *Protein Sci. 9*, 1395-1398.

22. Tan, Y. J., Oliveberg, M., Davis, B., and Fersht, A. R. (1995) Perturbed pK_A-values in the denatured states of proteins, *J. Mol. Biol. 254*, 980-992.

23. Whitten, S., and Garcia-Moreno E., B. (2000) pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state., *Biochemistry 39*, 14292–14304.

24. Swint-Kruse, L., and Robertson, A. D. Hydrogen bonds and the pH dependence of ovomucoid third domain stability, *Biochemistry 34*, 4724–4732.

25. Trefethen, J. M., Pace, C. N., Scholtz, J. M., and Brems, D. N. (2005) Charge-charge interactions in the denatured state influence the folding kinetics of ribonuclease Sa, *Protein Sci. 14*, 1934-1938.

26. Cho, J.-H., and Raleigh, D. P. (2006) Electrostatic interactions in the denatured state and in the transition state for protein folding: effects of denatured state interactions on the analysis of transition state structure, *J. Mol. Biol. 359*, 1437-1446.

27. Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988) Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding, *Biochemistry 27*, 7167-7175.

28. Jacobson, H., and Stockmayer, W. H. (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems, *J. Chem. Phys. 18*, 1600-1606.

29. Tzul, F. O., and Bowler, B. E. (2010) Denatured states of low complexity polypeptide sequences differ dramatically from those of foldable sequences, *Proc. Natl. Acad. Sci. U.S.A. 107*, 11364-11369.

30. Rao, K. S., Tzul, F. O., Christian, A. K., Gordon, T. N., and Bowler, B. E. (2009) Thermodynamics of loop formation in the denatured state of *Rhodopseudomonas palustris* cytochrome *c*': scaling exponents and the reconciliation problem, *J. Mol. Biol. 392*, 1315-1325.

31. Dar, T. A., Schaeffer, R. D., Daggett, V., and Bowler, B. E. (2011) Manifestations of native topology in the denatured state ensemble of *Rhodopseudomonas palustris* cytochrome *c*', *Biochemistry 50*, 1029-1041.

32. Duncan, M. G., Williams, M. D., and Bowler, B. E. (2009) Compressing the free energy range of substructure stabilities in iso-1-cytochrome *c*, *Protein Sci. 18*, 1155-1164.

33. Pollock, W. B., Rosell, F. I., Twitchett, M. B., Dumont, M. E., and Mauk, A. G. (1998) Bacterial expression of a mitochondrial cytochrome *c*. Trimethylation of Lys72 in yeast iso-1-cytochrome *c* and the alkaline conformational transition, *Biochemistry 37*, 6124-6131.

34. Rosell, F. I., and 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.

35. McClelland, L. J., Seagraves, S. M., Khan, M. K. A., Cherney, M. M., Bandi, S., Culbertson, J. E., and Bowler, B. E. (2015) The response of Ω-loop D dynamics to truncation of trimethyllysine 72 of yeast iso-1-cytochrome *c* depends on the nature of loop deformation, *J. Biol. Inorg. Chem. 20*, 805-819.

36. Godbole, S., and Bowler, B. E. (1997) A histidine variant of yeast iso-1-cytochrome *c* that strongly affects the energetics of the denatured state, *J. Mol. Biol. 268*, 816-821.

37. Withers-Ward, E. S., Mueller, T. D., Chen, I. S. Y., and Feigon, J. (2000) Biochemical and structural analysis of the interaction between the UBA(2) domain of the DNA repair protein HHR23A and HIV-1 Vpr, *Biochemistry 39*, 14103-14112.

38. Cherney, M. M., Junior, C., and Bowler, B. E. (2013) Mutation of trimethyllysine-72 to alanine enhances His79-heme mediated dynamics of iso-1-cytochrome *c*, *Biochemistry 52*, 837-846.

39. Redzic, J. S., and Bowler, B. E. (2005) Role of hydrogen bond networks and dynamics in positive and negative cooperative stabilization of a protein, *Biochemistry 44*, 2900-2908.

40. Wandschneider, E., Hammack, B. N., and Bowler, B. E. (2003) Evaluation of cooperative interactions between substructures of iso-1-cytochrome *c* using double mutant cycles, *Biochemistry 42*, 10659-10666.

41. Goldes, M. E., Jeakins-Cooley, M. E., McClelland, L. J., Mou, T.-C., and Bowler, B. E. (2016) Disruption of a hydrogen bond network in human versus spider monkey cytochrome *c* affects heme crevice stability, *J. Inorg. Biochem. 158*, 62-69.

42. Margoliash, E., and Frohwirt, N. (1959) Spectrum of horse-heart cytochrome *c*, *Biochem. J. 71*, 570-572.

43. Nozaki, Y. (1972) The preparation of guanidine hydrochloride, *Methods Enzymol. 26*, 43-50.

44. Pace, C. N., Shirley, B. A., and Thomson, J. A. (1989) Measuring the conformational stability of a protein, In *Protein structure: a practical approach* (Creighton, T. E., Ed.), pp 311-330, IRL Press at Oxford University Press, New York.

45. Schellman, J. A. (1978) Solvent denaturation, *Biopolymers 17*, 1305-1322.

46. Wandschneider, E., and Bowler, B. E. (2004) Conformational properties of the iso-1-cytochrome *c* denatured state: dependence on guanidine hydrochloride concentration, *J. Mol. Biol. 339*, 185-197.

47. Tonomura, B., Nakatani, H., Ohnishi, M., Yamaguchi-Ito, J., and Hiromi, K. (1978) Test reactions for a stopped-flow apparatus. Reduction of 2,6-dichlorophenolindophenol and potassium ferricyanide by L-ascorbic acid, *Anal. Biochem. 84*, 370-383.

48. Anil, B., Song, B., Tang, Y., and Raleigh, D. P. (2004) Exploiting the right side of the Ramachandran plot: substitution of glycines by D-alanine can significantly increase protein stability, *J. Am. Chem. Soc. 126*, 13194-13195.

49. Smith, C. R., Mateljevic, N., and Bowler, B. E. (2002) Effects of topology and excluded volume on protein denatured state conformational properties, *Biochemistry 41*, 10173-10181.

50. Khan, M. K. A., Miller, A. L., and Bowler, B. E. (2012) Tryptophan significantly stabilizes His-heme loops only when it is near a loop end, *Biochemistry 51*, 3586-3595.

51. Khan, M. K. A., and Bowler, B. E. (2012) Conformational properties of polyglutamine sequences in guanidine hydrochloride solutions, *Biophys. J. 103*, 1989–1999.

52. Finnegan, M. L., and Bowler, B. E. (2012) Scaling properties of glycine-rich sequences in guanidine hydrochloride solutions, *Biophys. J. 102*, 1969-1978.

53. Xu, Y., Mayne, L., and Englander, S. W. (1998) Evidence for an unfolding and refolding pathway in cytochrome *c*, *Nat. Struct. Biol. 5*, 774-778.

54. Smith, C. R., Wandschneider, E., and Bowler, B. E. (2003) Effect of pH on the iso-1-cytochrome *c* denatured state: changing constraints due to heme ligation, *Biochemistry 42*, 2174-2184.

55. Kurchan, E., Roder, H., and Bowler, B. E. (2005) Kinetics of loop formation and breakage in the denatured state of iso-1-cytochrome *c*, *J. Mol. Biol. 353*, 730-743.

56. Rohl, C. A., Chakrabartty, A., and Baldwin, R. L. (1996) Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol, *Protein Sci. 5*, 2623-2637.

57. Chakrabartty, A., Schellman, J. A., and Baldwin, R. L. (1991) Large differences in the helical propnesity of glycine and alanine, *Nature 351*, 586-588.

58. de Gennes, P.-G. (1979) *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, NY.

59. Chan, H. S., and Dill, K. A. (1990) The effect of internal constraints on the configurations of chain molecules, *J. Chem. Phys. 92*, 3118-3135.

60. Redner, S. (1980) Distribution functions in the interior of polymer chains, *J. Phys. A: Math. Gen. 13*, 3525-3541.

61. Finnegan, M. L., and Bowler, B. E. (2010) Propensities of aromatic amino acids versus leucine and proline to induce residual structure in the denatured-state ensemble of iso-1-cytochrome *c*, *J. Mol. Biol. 403*, 495-504.

62. Crowhurst, K. A., and Forman-Kay, J. D. (2003) Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain, *Biochemistry 42*, 8687-8695.

63. Crowhurst, K. A., Tollinger, M., and Forman-Kay, J. D. (2002) Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain, *J. Mol. Biol. 322*, 163-178.

64. Marsh, J. A., and Forman-Kay, J. D. (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints, *J. Mol. Biol. 391*, 359-374.

65. Day, R., and Daggett, V. (2005) Ensemble versus single-molecule protein unfolding, *Proc. Natl. Acad. Sci. U.S.A. 102*, 13445-13450.

66. Wong, K.-B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M. V., Fersht, A. R., and Daggett, V. (2000) Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding, *J. Mol. Biol. 296*, 1257-1282.

67. Bruun, S. W., Iešmantavičius, V., Danielsson, J., and Poulsen, F. M. (2010) Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein, *Proc. Natl. Acad. Sci. U.S.A. 107*, 13306-13311.

68. Petersen, B., Lundegaard, C., and Petersen, T. N. (2010) NetTurnP - neural network prediction of beta-turns by use of evolutionary information and predicted protein sequence features, *PLoS ONE 5*, e15079.

69. Fitzkee, N. C., and Rose, G. D. (2004) Reassessing random-coil statistics in unfolded proteins, *Proc. Natl. Acad. Sci. U.S.A. 101*, 12497-12502.

70. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. N., DeMarco, M. L., Daggett, V., and Fersht, A. R. (2011) Unifying features in protein folding mechanisms, *Proc. Natl. Acad. Sci. U.S.A. 100*, 13286-13291.

71. DeMarco, M. L., Alonso, D. O. V., and Daggett, V. (2004) Diffusing and colliding: the atomic level folding/unfolding pathway of a small helical protein, *J. Mol. Biol. 341*, 1109-1124.

72. Muñoz, V., and Serrano, L. (1994) Elucidating the folding problem of a-helical peptides using empirical parameters III: temperature and pH dependence, *J. Mol. Biol. 245*, 297-308.

73. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V., and Fersht, A. R. (2005) Solution structure of a protein denatured state and folding intermediate, *Nature 437*, 1053-1056.

74. Arora, P., Oas, T. G., & Myers, J. K. (2004). Fast and faster: A designed variant of the B-domain of protein A folds in 3 microseconds. *Protein Science*, 847-853.

75. Bowler, B. (2012). Residual structure in unfolded proteins. *Current Opinion in Structural Biology*, 4-13.

76. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., . . . Wright, P. E. (2002). Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature*, 549-553.

77. Gregersen N, Bross P. (2010). Protein misfolding and cellular stress: an overview. *Methods Mol Bio, 648*, 3-23.

78. Iesmantacvicius, V., Jensen, M. R., Ozenne, V., Blackledge, M., Poulsen, F. M., & Kjaergaard, M. (2013). Modulation of the Intrinsic Helix Propensity of an Intrinsically Disordered Protein Reveals Long-Range Helix-Helix Interactions. *Journal of the American Chemical Society*, 10155-10163.

79. Jed Long, Thomas P. Garner, Maya J. Pandya, C. Jeremy Craven, Ping Chen, Barry Shaw, Michael P. Williamson, Robert Layfield, Mark S. Searle. (2010). Dimerisation of the UBA Domain of p62 Inhibits Ubiquitin Binding and Regulates NF-kB Signalling. *J. Mol. Biol., 396*, 178-194.

80. K. Sudhindra Rao, Franco O. Tzul, Arwen K. Christian, Tia N. Gordon, Bruce E. Bowler. (2009). Thermodynamics of Loop Formation in the Denatured State of Rhodopseudomonas palustris Cytochrome c': Scaling Exponents and the Reconcilation Problem. *J. Mol. Biol., 392*(5), 1315-1325.

81. Kohn, J. E., & Plaxco, K. W. (2004). Random coil behavior and the dimensions of chemically unfolded proteins. *PNAS*, 12491-12496.

82. Leavens, M. J., Cherney, M. M., Finnegan, M. L., & Bowler, B. E. (2018). Probing Denatured State Conformational Bias in the Three-Helix Bundle, UBA(2), Using a Cytochrome c Fusion Protein. *Biochemistry*, 1711-1721.

83. Riback, J. A., Bowman, M. A., Zmyslowski, A. M., Knoverek, C. R., Jumper, J. M., Hinshaw, J. R., . . . Sosnick, T. R. (2017). Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science*, 238-241.

84. Sugase, K., Dyson, H. J., & Wright, P. E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 1021-1025.

85. Thomas D. Mueller, Juli Feigon. (2002). Solution Structures of UBA Domains Reveal a Conserved Hydrophobic Surface for Protein-Protein Interactions. *J. Mol. Biol., 319*, 1243-1255.

86. Tran , H., Mao, A., & Pappu, R. (2008). Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *JACS, 130*, 7380-7392.

Ch 4: Unusual Behavior in the Denatured State Properties of the Instrinsically Disordered Protein CBP

Unusual Behavior in the Denatured State Properties of the Instrinsically Disordered Protein CBP

Moses J. Leavens, Melisa M. Cherney, and Bruce E. Bowler*

Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Dynamics, University of Montana, Missoula, Montana 59812

ABSTRACT:

Denatured state histidine-heme loop formation methods with the four-helix-bundle protein cytochrome *c′* from *Rhodopseudomonas palustris*, the three-helix bundles from Ubiquitin-associated domains in the human DNA excision repair protein HHR23A, indicate fold-specific deviations from random coil behavior, localizing to reverse turns. To discriminate the denatured state conformational biases between foldable and disordered sequences in three-helix bundles, we extend our denatured state histidine-heme loop formation thermodynamic and kinetic method to mouse cAMP-responsive element binding protein (CBP), an intrinsically disordered protein that serves as general transcription coactivator involved in hormonal signaling. We fuse the CBP domain to the N-terminus of yeast iso-1-cytochrome *c*, engineering single histidine mutations into highly solvent accessible positions within CBP, creating nine single histidine variants. Similar to foldable three-helix bundles, guanidine hydrochloride denaturation shows the CBP-cytochrome *c* fusion protein unfolds in a three-state equilibrium, with the Cyt*c* domain unfolding first. Engineered histidine residues in CBP have minor impact on the stability of the iso-1-cytochrome *c* domain. Histidine-heme loop formation kinetic and thermodynamic measurements at 4 and 6 M guanidine hydrochloride show a weak dependence of the stability of the loop, pK_{loop}, on loop size using the Jacobson-Stockmayer equation. Distinct from foldable three-helix bundles, we observe the His8-heme and the His53-heme loops in unstructured regions are more stable than expected from the Jacobson-Stockmayer relationship, and break more slowly than expected. We do observe unusual behavior with the His38-heme loop in the denatured state that frames a turn. These results show local sequence near His8, His38, and His53, which are on the ends of helices or near turns in CBP, are prone to persistent interactions in the denatured state. These results are different from our work on persistent reverse turns for foldable helix bundles.

INTRODUCTION

Intrinsically disordered proteins (IDPs) and disordered regions (IDRs) are abundant in the eukaryotic proteome. These IDPs and their IDRs are partially identified by their biased amino acid sequence and characterized with protein disorder prediction programs such as DistProt.[88] Because of their abundance in the human proteome, and the key role they play in transcription and cellular signaling, characterizing the residual or non-random structure in IDPs and IDRs has become important. There has been much debate whether IDPs and IDRs can form preformed structure before binding to their numerous partner ligands.[88] These IDRs are often used by larger complexes to serve as scaffolds to initiate a particular response in transcription of genes or cellular signaling. For instance, the nuclear receptor coactivator binding domain (NCBD) of the cyclic AMP (cAMP)-responsive element binding (CREB) protein serves as a downstream signal in the signaling pathway for glucagon.[88] The observations that the same amino acid sequence in an IDP can be applied in many different ways dependent on the signal involved has made this an area of active interest for protein biochemistry.

While residual or non-random structure has been well established to be observed in foldable protein sequences, the presence of such residual structure in unfolded intrinsically disordered protein sequences is less clear. To determine whether residual structure is present in the unfolded state of an IDP, we have examined mouse cAMP-responsive element binding protein (CBP), an intrinsically disordered protein that serves as general transcription coactivator involved in hormonal signaling. To determine whether residual structure is present, and where this residual structure may be located in CBP, we have fused the CBP domain to the N-terminus of our His-heme scaffold protein yeast iso-1-Cyt*c*.

Using His-heme loop formation thermodynamic and kinetic methods in the denatured state, we can track the probability for intrinsic conformational bias along a given protein's primary structure in the denatured state. Deviations from random coil behavior are detected by fitting pK_{loop} (stability of the His-heme loop) data versus loop size, with the expectation that pK_{loop} is proportional to the size of the loop. The Jacobson-Stockmayer equation (eq. 4-1) assumes loop formation in a random coil is solely because of entropy.[28]

$$(4-1)\ \Delta S_{\mathrm{loop}} = -\nu_3 \mathrm{R} ln(n) + \mathrm{R} ln((3/2\pi C_{\mathrm{n}} l^2)^{\nu_3} \mathrm{V_i})$$

In eq.1, R is the gas constant, ν_3 is the scaling exponent for loop formation, *n* is the number of monomers in the loop , *l* is the distance between the monomers forming the loop, C_n is Flory's characteristic ratio, and V_i is the approach volume within which the atoms need to be constrained for loop formation to occur. We have used this histidine-heme (His-heme) loop formation methodology and have applied it to foldable sequences that encode helix bundles. Our lab has previously shown that using this His-heme loop method, that foldable protein sequences act quite differently than low complexity homopolymer sequences such as polyalanine, polyglutamine, and polyglycine.[29] In all cases with His-heme loop formation studies on these homopolymers, it was discovered that these data show tight correlation between pK_{loop} and the log of loop size, n. Morever, using the Jacobson Stockmayer in eq 4-1 to fit our data, we observed that the scaling exponents, ν_3, for loop formation in these homopolymers were in the range expected for a random coil with excluded volume (ν_3 in between 1.8 - 2.4).[29] We extended our His-heme loop studies to compact three-helix bundles—the Ubiquitin-associated domains isolated from the human DNA excision repair protein HHR23A. We discovered as for the four-helix bundle cytochrome *c*' from *Rhodopseudomonas palustris*[30, 31], that persistent His-heme loops are found in the reverse turn regions of the tertiary structures of these proteins. Thus, these His-heme loops

are more favorable than expected based on predictions by the Jacobson-Stockmayer equation (eq 4-1). We subjected these thermodynamically favorable His-heme loops to downward pH jumps monitored by stopped-flow, and discovered that the loop breakage rate constants in these thermodynamically favorable loops were extremely low, indicating these parts of the chain in denaturing conditions of guanidine hydrochloride exhibit persistent structure in the DSE, and perhaps are involved in governing the fold topology in these helix-bundles.

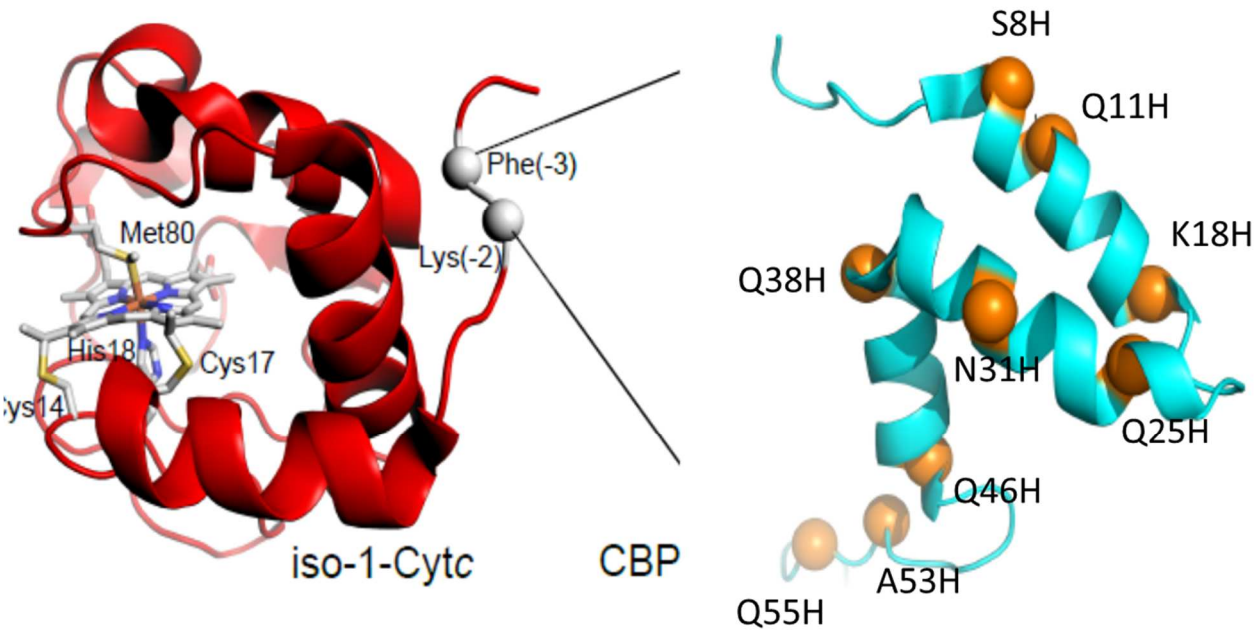

Figure 4-1. Structures of yeast iso-1 cytochrome *c* (red, PDB ID: 2YCC) and the mouse CBP domain (blue, PDB: 2KKJ), showing insertion of the CBP domain between Phe(-3) and Lys(-2) (white spheres) near the N-terminus of iso-1-Cyt *c*. Sites of single histidine substitutions are shown with orange spheres in the CBP domain. The heme and the amino acid residues that attach the heme to the iso-1-Cyt*c* domain, Cys14, Cys17, His18 and Met80 are shown as stick models and colored by element.

To extend our hypothesis that reverse turns remain in the denatured state of helix bundles, we have fused the mouse CBP domain to yeast iso-1-Cyt*c*, where CBP folds to three-helix bundle topology when bound to its signaling partner, activator for thyroid hormone and retinoid receptor (ACTR). We demonstrate that His-heme loop formation methodology can be applied to intrinsically disordered protein and intrinsically disordered region sequences. We have used our pRbs_BTR1fuse vector and have fused this disordered three-helix sequence to the N-terminus of iso-1-Cyt*c* (CBP – iso-1-Cyt*c*, Figure 4-1). We find that histidine-heme loop formation equilibria in the denatured state show a relatively weak dependence on pK_{loop} versus the log of loop size. Despite this weak dependence on loop size and on guanidine hydrochloride concentration, we observe robust deviations from random coil behavior in CBP's primary structure that localizes near the ends of the N and C-termini, and a region between helix 2 and 3 within this disordered chain. Stopped-flow loop breakage kinetics yield very low His-heme loop breakage rate constants, k_b, for the histidine probes in these N and C-terminal regions, and the region between helix 2 and 3 of CBP. Thus, these regions located at the N and C termini, and between helix 2 and 3 of CBP, may be critical for establishing its three-helix bundle topology when bound to its signaling partner, ACTR.

EXPERIMENTAL PROCEDURES

Preparation of mouse CBP fused to the N-terminus of iso-1-Cyt*c*. The cAMP response element binding protein gene (resides 1-59), was cloned into pRbs-BTR1fuse to produce pRbs_BTR1(CBP_Cc). The gene for CBP gene was amplified by PCR using methods as previously described (see chapter 2 experimental procedures). After amplication, the PCR product containing the CBP gene was ligated between the *Eco*RI and *Ngo*MIV restriction sites of the pRbs_BTR1fuse vector using T4 ligase (New England Biolabs) to produce the CBP – iso-1-

Cyt*c* fusion sequence, pRbs_BTR1(CBP_Cc). The entire CBP – iso-1-Cyt*c* sequence was confirmed by dideoxy sequencing (University of Montana Genomics Core Facility).

Preparation of CBP – iso-1-Cyt*c* fusion protein variants. Primers to mutate the pRbs_BTR1(CBP_Cc) plasmid are in Table 4-S2. The CBP domain inserted on the N-terminal of iso-1-Cyt*c* is 59 residues. The wild type sequence of CBP encodes no native histidine. The pRbs_BTR1(CBP_Cc) plasmid containing the gene for CBP was used to produce S8H, Q11H, K18H, Q25H, N31H, Q38H, Q46H, A53H, and Q55H variants of the CBP – iso-1-Cyt*c* fusion protein using QuikChange Lightning mutagenesis (Agilent Technologies). Histidine surface mutations were chosen by the accessible surface area algorithm from the Sealy Center for Structural Biology, University of Texas Medical Branch (http://curie.utmb.edu/getarea.html). Dideoxy sequencing confirmed the presence of each mutation (Eurofins Genomics, Louisville, Kentucky).

CBP – iso-1-Cyt*c* Variant Protein Expression & Purification. WT and single histidine variants were expressed from the pRbs_BTR1(CBP_Cc) plasmid by transformation into ultra BL21 (DE3) *E. coli* competent cells (New England Biolabs) by adhering to the manufacturer's protocol. One liter of sterile 2xYT media was inoculated with ~20 μM PMSF and ~5-10 mL of suspended bacterial cells with 1 mL of 100 mg/mL L-ampicillin and placed in a 2.8 L Fernbach flask. An orbital shaker at 150 rpm, 30 °C, for 24 hours was used to grow cultures. Previous protocols were used to express and purify the CBP-iso-1-Cyt*c* fusion proteins, with two exceptions[38-41]. Cleavage of CBP-iso-1-Cyt*c* fusion proteins initially occurred, so ~20 mM PMSF was added at every purification step to avoid cleavage of the CBP domain. Second, after ammonium sulfate equilibration overnight at 4 °C, the pH in the dialysis buffer was changed to pH 8.5 (MilliQ water with 1 mM Na_2EDTA, 12.5 mM sodium phosphate monobasic, pH 8.5, at

4 °C)[76]. After this step, the protein was bound to CM-Sepharose and the CBP-iso-1-Cyt*c* fusion proteins were eluted with a salt gradient (applying 0 – 0.8 M NaCl in 50 mM sodium phosphate pH 7.5, 1 mM Na_2EDTA) at 4 °C. After ion-exchange chromatography, fusion proteins were oxidized as previously described in chapter 2, and run down a column using G25 Sephadex chromatography to remove oxidant. Matrix-assisted laser-desorption time-of-flight (MALDI-TOF) mass spectrometry (Bruker microflex mass spectrometer, Table 4-S4) was carried out before and after each experiment, to determine that no fragmentation of the fusion proteins had taken place.

Denaturation of CBP – iso-1-Cyt*c* Variants By Guanidine Hydrochloride. The WT CBP-Cyt *c* and eight single histidine variants were oxidized and separated from oxidant as described above into CD buffer (20 mM Tris, 40 mM NaCl, 1 mM Na_2EDTA, pH 7.0). Degree of oxidation and protein concentration were checked using UV/Vis absorbance at the following wavelengths: 339 nm (20.9 x 10^3 M^{-1} cm^{-1}), 526.5 nm (11.0 x 10^3 M^{-1} cm^{-1}), 541.8 nm (9.0 x 10^3 M^{-1} cm^{-1}), 550 nm (oxidized, 9.1 x 10^3 M^{-1} cm^{-1}), 550 nm (reduced 28.0 x 10^3 M^{-1} cm^{-1}).[42] Refractive index measurements to determine concentration of the GuHCl stock in CD buffer and CD buffer were carried out as previously described in chapter 2. The CBP– iso-1-Cyt*c* variants were diluted to 4 μM in concentrated GuHCl and titrated into a 4 μM solution of CBP – iso-1-Cyt*c* in CD buffer with Hamilton Microlab 500 Titrator linked to an Applied Photophysics Chirascan CD Spectrophotometer. Change in ellipticity was monitored at α-helix (222 nm) using 250 nm as background, $\theta_{222corr}$ (= θ_{222nm} - θ_{250nm}). The $\theta_{222corr}$ versus GuHCl concentration was graphed and fit to eq 4-2, that assumes three state unfolding and linear dependence of the free

$$(4-2)\ \theta_{222corr} = \theta_D + m_D[GuHCl] + \frac{\theta_N + m_N[GuHCl] - \theta_D - m_D[GuHCl] + \left[(\theta_I - \theta_D - m_D[GuHCl])e^{\frac{m_{NI}[GuHCl] - \Delta G_{NI}^{o\prime}(H2O)}{RT}}\right]}{\left[1 + \left(e^{\frac{m_{NI}[GuHCl] - \Delta G_{NI}^{o\prime}(H2O)}{RT}}\right)\left(1 + e^{\frac{m_{ID}[GuHCl] - \Delta G_{ID}^{o\prime}(H2O)}{RT}}\right)\right]}$$

energies of unfolding, ΔG_u, on GuHCl concentration.[44, 45] In eq 2, θ_N and m_N are the intercept and the slope of the native state baseline, θ_I is the intermediate state baseline, θ_D and m_D are the intercept and the slope of the denatured state baseline, m_{NI} is the rate of change of the free energy of the native to intermediate state transition (ΔG_{NI}) with respect to GuHCl concentration, m_{ID} is the rate of change of the free energy of the of intermediate to denatured state transition (ΔG_{ID}) with respect to GuHCl concentration and $\Delta G_{NI}^{o'}(H_2O)$ and $\Delta G_{ID}^{o'}(H_2O)$ are the free energies of the native to intermediate and intermediate to denatured state transitions, in aqueous buffer. Single histidine variants were fit assuming that the native baseline is independent of GuHCl concentration since the native baselines were too short for most mutants to reliably evaluate m_N. Reported parameters are the average and standard deviation of at least three independent trials.

His-Heme Loop Equilibria in 6 M and 4 M GuHCl for CBP-Cyt*c* Variants. The denatured state histidine-heme loop formation has been previously described.[31,82] In brief, a UV-Vis spectrophotometer (Beckman DU 800) was used to monitor the spin state shift of the heme in the soret region (398 nm) for all CBP – iso-1-Cyt*c* variants. The His-heme denatured state pH titrations were measured at room temperature (22 ± 1 °C) with 3 μM denatured protein in 15 mM NaCl, 5 mM Na_2HPO_4, 1 mM Na_2EDTA (1x buffer) at 6 M or 4 M GuHCl. Refractive index measurements of GuHCl were done as described above to determine GuHCl concentration. The pH titrations at 6 M or 4 M GuHCl were done by keeping all variables constant except pH.[46] At every pH value, absorbance spectra was recorded from 350 – 450 nm. The absorbance band at

398 nm, A_{398}, using absorbance at 450 nm, A_{450}, as a baseline, $A_{398corr}$ (= $A_{398} - A_{450}$) was graphed versus pH. The $A_{398corr}$ versus pH graphs were fit to eq 4-3 to determine the apparent pK_a, pK_a(obs),

$$(4-3)\ \ A_{398\mathrm{corr}} = \frac{A_{LS} + A_{HS} \times 10^{n_p[pK_a(\mathrm{obs}) - pH]}}{1 + 10^{n_p[pK_a(\mathrm{obs}) - pH]}}$$

and the number of protons, n_p, linked to His-heme loop formation. In eq 4-3, A_{LS} is $A_{398corr}$ at high pH when the heme is low spin with histidine bound to the heme iron and A_{HS} is $A_{398corr}$ at low pH when the heme is high spin with H_2O bound to the heme iron.

Stopped-Flow Loop Breakage Kinetics of CBP-Cyt*c* Variants. These methods have been previously described in chapter 2. Briefly, absorbance spectra were collected at 398 nm using an Applied Photophysics SX20 stopped-flow apparatus with temperature controlled at 25 °C. The CBP – iso-1-Cyt*c* variants move from low spin heme (i.e. high pH) to high spin heme (i.e. low pH) in a downward pH jump experiment. The CBP – iso-1-Cyt*c* variants were oxidized with ferricyanide and separated from oxidant by G-25 Sephadex chromatography into either MOPS buffer (10 mM MOPS, 40 mM NaCl, 2 mM Na_2EDTA, pH 6.8) or a Tris buffer (20 mM Tris, 2 mM Na_2EDTA, pH 8.0), depending on loop size. Typically, larger His-heme loops do not completely form until a higher pH, so a higher pH buffer for these variants with larger loops is appropriate. Protein concentration was determined as described above. Exact concentrations of GuHCl were determined as described above. Denatured CBP-Cyt*c* variants at 6 μM protein were made at 4 M or 6 M GuHCl in the appropriate buffer (as described above), and mixed 1:1 with 4 M or 6 M GuHCl in 100 mM citrate at pH 3 or pH 3.5, to produce a final protein concentration of 3 μM. Final pH after mixing was determined in accordance with the protocol in chapter 3. The dead time of the stopped-flow was measured by reduction of dichlorophenolindophenol as a

function of L-ascorbic acid concentration and is 2 ms under our mixing conditions.[47] Kinetic data were fit to single exponential model after adding 0.002 s to each time point.

RESULTS

Denaturation of CBP – iso-1-Cyt*c* Variants By Guanidine Hydrochloride. The global unfolding of the CBP – iso-1-Cyt*c* variants was monitored with CD at 25 °C and pH 7.0 using GuHCl as denaturant. Figure 4-2 depicts denaturation curves, $\theta_{222corr}$ versus GuHCl concentration, for the WT

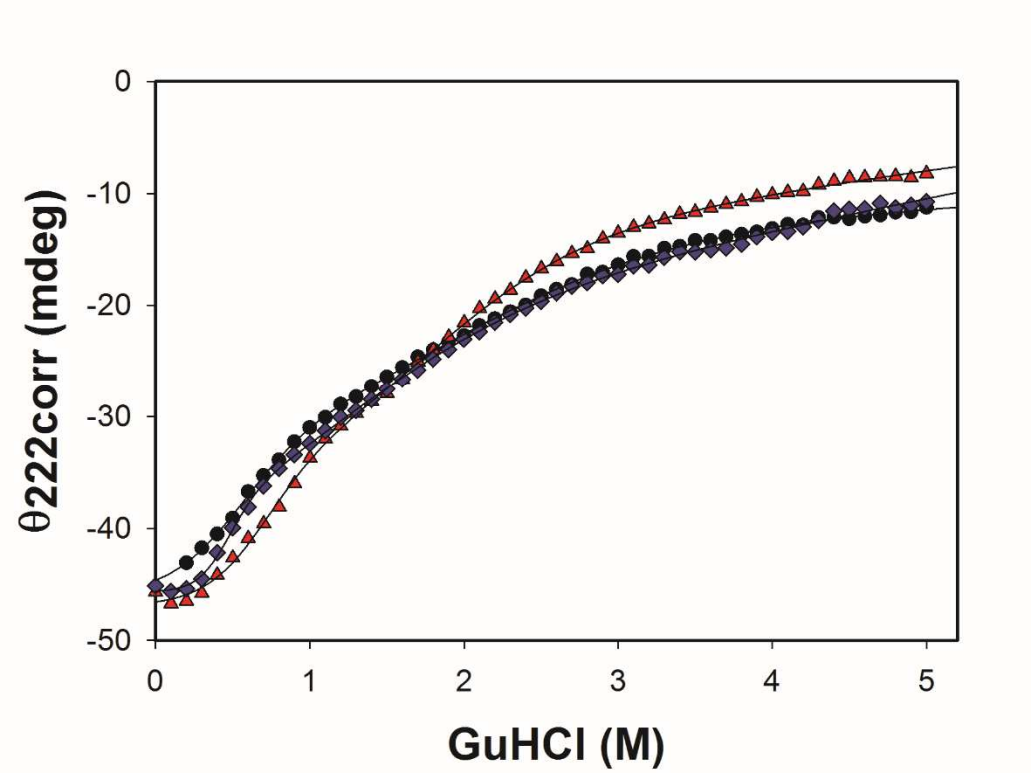

Figure 4-2. GuHCl denaturation curves for CBP – iso-1-Cyt*c* variants, WT (black circles), Q25H (red triangles), and S8H (purple diamonds) represented as plots of corrected ellipticity at 222 nm, $\theta_{222corr}$, versus GuHCl concentration. Solid curves are fits to eq 4-2 in Experimental

Procedures. Parameters obtained from the fits are given in Table 4-1. Experiments were performed at 25 °C and pH 7.

Table 4-1. Thermodynamic Parameters for GuHCl Unfolding of CBP – iso-1-Cyt*c* at 25 ºC and pH 7.0.

CBP – iso-1-Cyt*c* Variants[a]	m_{NI}, kcal mol^{-1} M^{-1}	$\Delta G_{NI}^{o'}(H_2O)$, kcal mol^{-1}	C_{mNI}, M	m_{ID}, kcal mol^{-1} M^{-1}	$\Delta G_{ID}^{o'}(H_2O)$, kcal mol^{-1}	C_{mID}, M
WT	3.7 ± 0.7	2.0 ± 0.4	0.54 ± 0.01	0.5 ± 0.1	0.5 ± 0.4	0.9 ± 0.6
Q55H (21)	5 ± 1	1.5 ± 0.4	0.29 ± 0.01	1.2 ± 0.1	1.4 ± 0.2	1.25 ± 0.03
A53H (23)	4.84 ± 0.05	1.79 ± 0.05	0.37 ± 0.01	0.86 ± 0.02	0.90 ± 0.03	1.04 ± 0.01
Q46H (30)	2.9 ± 0.7	0.9 ± 0.5	0.3 ± 0.10	0.8 ± 0.1	2.0 ± 0.2	2.5 ± 0.2
Q38H (38)	-	-	-	-	-	-
N31H (45)	-	-	-	-	-	-
Q25H (51)	3.4 ± 0.9	1.8 ± 0.4	0.5 ± 0.1	1.3 ± 0.3	2.2 ± 0.6	1.67 ± 0.04
K18H (58)	1.7 ± 0.4	0.6 ± 0.2	0.32 ± 0.09	0.5 ± 0.1	2.5 ± 0.4	5 ± 1
Q11H (65)	-	-	-	-	-	-
S8H (68)	5.0 ± 0.6	2.4 ± 0.2	0.49 ± 0.01	0.91 ± 0.05	1.0 ± 0.1	1.03 ± 0.09

[a]Denatured state His-heme loop size is in brackets. Q38H, N31H, and Q11H could not be fit reliably.

variant and two of the eight single histidine variants, Q25H and S8H. Each CBP – iso-1-Cyt*c* fusion protein shows similar 3-state equilibrium unfolding titration curves as previously observed the UBA– iso-1-Cyt*c* constructs.[82] Thus, eq 2 was used to extract thermodynamic parameters for CBP-iso-1-Cyt*c* listed in Table 4-1. Based on previous studies with GuHCl as denaturant,[32, 48] we assigned the first phase to the iso-1-Cyt*c* unfolding and the second phase of the unfolding transition to the CBP domain (Fig. 4-2.). The iso-1-Cyt*c* domain has an unfolding concentration midpoint near 1 M GuHCl and *m*-value of 3.8 to 4.0 kcal/mol-M.[32] Previous work has shown that CBP cannot be fit to a thermodynamic unfolding model by itself.[88] Given the knowledge of Cyt*c* unfolding, and its behavior with the UBA domains[76] in the UBA-Cyt*c* fusion

proteins, fitting CBP to the second phase of the unfolding transition is reasonable. Thus, in GuHCl unfolding, the iso-1-Cyt*c* domain unfolds first followed by the CBP domain (Fig. 4-2, Table 4-1), as for the UBA-Cyt*c* constructs. For iso-1-Cyt*c*, denaturant *m*-values for the N→I transition, m_{NI}, increased slightly for some of the single histidine mutants (Q55H, A53H, and S8H) compared to WT when surface histidines are engineered into CBP, while others show either little to no effect, or a slight decrease in *m*-values of the Cyt*c* domain (Q46H and K18H). Some histidine mutations lower the native to intermediate unfolding midpoint, C_{mNI}, by ~0.2 M GuHCl, and the free energy of unfolding in the absence of denaturant, $\Delta G_{NI}^{o'}(H_2O)$ by 0.5 – 1.2 kcal/mol. On the other hand, other histidine mutations have little to no effect on these parameters (A53H, Q25H, and S8H), respectively. Thus, impacts on overall stability of the iso-1- Cyt*c* domain are variant dependent. Some of the variants were unable to be fit, because of insufficient baseline (Table 4-1).

The single histidine variant of the CBP domain appears to stabilize this disordered domain. The control experiment with WT in Table 4-1 shows the intermediate to denatured unfolding midpoint, C_{mID}, is 0.9 M, while the C_{mID} values for most of the eight single histidine variants are greater than 1 M, with the sole exception of S8H. The *m*-values for the CBP domain, m_{ID}, increased for all variants except K18H, relative to WT. The midpoint concentration for intermediate to denatured, C_{mID}, increases for all mutations, except for A53H and S8H. Therefore, the net effect is that for a majority of single histidine mutants in the CBP domain, a stabilization of this disordered protein occurs. Parameters in Table 4-1 were used to calculate the population of denatured protein for each variant at 6 M GuHCl, and a majority of mutants are at least 95% unfolded.

Denatured State His-Heme Loop Formation Equilibria in CBP-Cyt*c* Variants. We apply the His-heme loop formation equilibria to investigate the denatured state thermodynamic properties of the disordered three-helix bundle CBP. This methodology has been previously described in chapters 2 and 3, and will be briefly summarized here. We fuse the CBP domain to the N-terminus of Cyt*c*, at the same insertion site as other studied three-helix bundle domains (Figure 1). All competitive histidines that may interfere with our single histidine sites have been eliminated from the iso-1-Cyt*c* domain. In iso-1-Cyt*c*, Cys14 is the closest attachment to the heme (Figure 4-1). Thus, each loop contains 16 residues from the iso-1-Cyt*c* sequence (insertion of CBP domain is between Lys(-2) and Phe(-3), Figure 1). For loop sizes less than 16, chain stiffness effects become important.[49] In denaturing conditions, Met80 is a weak ligand for Fe^{3+},[53] and will be displaced by water, allowing an engineered histidine (or a lysine for the WT variant[54]) to bind Fe^{3+} in the heme. Histidine can be protonated, so the relative stability of a His-heme loop can be measured using a denatured state pH titration, to produce an apparent pK_a, pK_a(obs), and n_p, the number of protons involved during His-heme loop formation.

Figure 4-3 displays denatured state pH titration curves for the WT, A53H, and the S8H variants of CBP – iso-1-Cyt*c* in 6 M GuHCl. At high pH (pH ~7), His-heme loops are formed (~pH 9 for the WT Lys-heme loop) and Fe^{3+} in heme is in low spin state. A blue shift occurs when pH is decreased, and the Fe^{3+}-heme transitions from a low (strong field ligand, His or Lys) to a high spin state (weak field ligand, water). Denatured state titration curves are fit to the Henderson-Hasselbach equation (eq 4-3 in Experimental Procedures), to determine the stability of the loops, pK_a(obs), and the number of protons, n_p, involved in loop formation (Table 4-2). In CBP, the WT variant has pK_a(obs) near 7 at 4 and 6 M GuHCl, establishing an upper limit to

determine the pK_a(obs) for denatured state His-heme loop formation. As previously observed, pK_a(obs) is lowest (most stable His-heme loop) for one of the smallest loop, His53-heme.

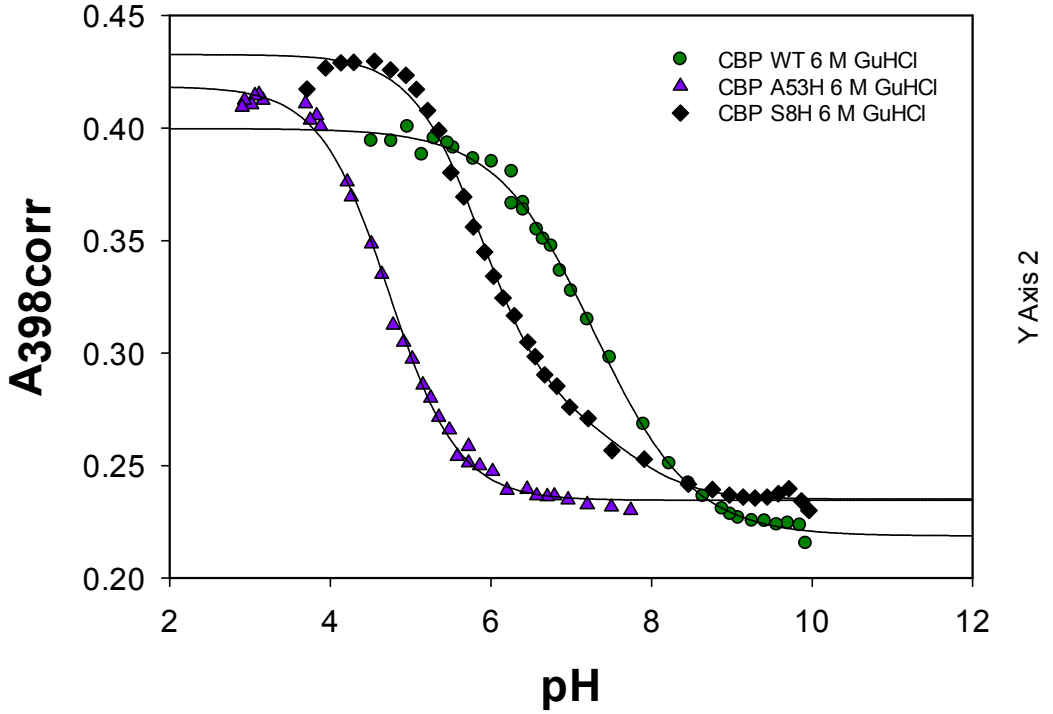

Figure 4-3. Representative curves for denatured state His-heme loop formation equilibria for CBP – iso-1-Cyt*c* variants at 6 M GuHCl. Plots of $A_{398corr}$ versus pH for the WT, A53H, and S8H variants are shown with green circles, purple triangles, and black diamonds, respectively. Data were fit to eq 4-3 in Experimental Procedures (solid curves). Thermodynamic parameters obtained from fits are listed in Table 4-2. All denatured state pH titrations were performed at 22 ± 1 °C.

As loop size increases, pK_a(obs) does not necessarily increase, different from the foldable three-helix bundle sequences. Loop sizes 21 - 30 were able to be fit to Eq. 4-3, but all of the loops after a loop size of 30 were not able to be fit to Eq. 4-3. It has been previously shown that longer

His-heme loops (>50 typically) are less stable in 6 M and 4 M GuHCl conditions. The fact that a loop size of less than 50 (His38-heme and His31-heme) needed to be fit to a two-state model suggests something about the disordered sequence is perturbing the histidine from binding to the heme, and needs a lysine to complete the loop formation process.

At 6 M GuHCl, the value of n_p falls to ~0.8 for the following variants: Q38H, N31H, Q25H, K18H, Q11H, and S8H. Previously, it has been observed for the pH titration curves of longer His-heme loops that the loops have become sufficiently unstable that the loop cannot finish this spin state transition, and Lys-heme binding at higher pH values is needed to complete the spin-state transition. Iso-1-Cyt*c* biphasic spin state transitions have been documented for loop sizes of 56, 72 and 83 in 6 M GuHCl.[46] For cytochrome *c′*, loops sizes of 48 – 111 all lead to biphasic loop formation equilibria at 6 M GuHCl.[31] However, because loop sizes less than 50 need to be fit to Eq. 4-4, it suggests the CBP sequence strongly affects His-heme loop formation.

Table 4-2. Thermodynamic Parameters for Denatured State Loop Formation of CBP – iso-1-Cyt*c* Variants in 6 M and 4 M GuHCl Solutions at 22 ± 1˚C

Variant	Loop size (n)	pK_a(obs)	n_p	pK_{loop}(His)
		6 M GuHCl		
WT	-	7.01 ± 0.07	0.72 ± 0.06	-
Q55H	21	5.01 ± 0.06	1.01 ± 0.07	-1.58 ± 0.06
A53H	23	4.71 ± 0.01	0.97 ± 0.02	-1.88 ± 0.01
Q46H	30	5.96 ± 0.04	0.92 ± 0.06	-0.63 ± 0.04
[a]Q38H	38	-	-	-0.49 ± 0.05
[a]N31H	45	-	-	-0.41 ± 0.03
[a]Q25H	51	-	-	-0.32 ± 0.05
[a]K18H	58	-	-	-0.28 ± 0.02

[a]Q11H	65	-	-	-0.32 ± 0.04
[a]S8H	68	-	-	-0.69 ± 0.09

[a]Fit to eq 4. Fits for following variants: **Q38H**: pK_a(HisH$^+$) = 6.02 ± 0.08, and pK_{loop}(Lys) = -4.36 ± 0.17. **N31H:** pK_a(HisH$^+$) = 6.81 ± 0.002, and pK_{loop}(Lys) = -2.87 ± 0.33 **Q25H:** pK_a(HisH$^+$) = 6.63 ± 0.006, and pK_{loop}(Lys) = -3.18 ± 0.18 **K18H:** pK_a(HisH$^+$) = 6.69 ± 0.05, and pK_{loop}(Lys) = -3.04 ± 0.015 **Q11H:** pK_a(HisH$^+$) = 6.73 ± 0.05, and pK_{loop}(Lys) = -2.87 ± 0.04 **S8H:** pK_a(HisH$^+$) = 6.67 ± 0.1, and pK_{loop}(Lys) = -3.42 ± 0.13 were the other parameters obtained from fit of eq 4 to the 6M GuHCl titration data. pK_a(LysH$^+$) was set to 10.5 for these fits.

4 M GuHCl

WT	-	7.04 ± 0.06	0.77 ± 0.01	-
Q55H	21	4.70 ± 0.07	1.03 ± 0.13	-1.89 ± 0.07
A53H	23	4.37 ± 0.003	0.91 ± 0.02	-2.22 ± 0.003
Q46H	30	5.63 ± 0.05	0.97 ± 0.09	-0.96 ± 0.05
Q38H	38	-	-	-0.42 ± 0.06
N31H	45	-	-	-0.41 ±0.03
Q25H	51	-	-	-0.36 ± 0.05
K18H	58	-	-	-0.45 ± 0.07
Q11H	65	-	-	-0.50 ± 0.007
S8H	68	5.75 ± 0.009	0.91 ± 0.04	-0.84 ± 0.009

[a]Fit to eq 4. Fits for following variants: **Q38H**: pK_a(HisH$^+$) = 5.67 ± 0.08, and pK_{loop}(Lys) = -4.48 ± 0.15. **N31H:** pK_a(HisH$^+$) = 6.59 ± 0.08, and pK_{loop}(Lys) = -3.15 ± 0.06 **Q25H:** pK_a(HisH$^+$) = 6.52 ± 0.05, and pK_{loop}(Lys) = -3.47 ± 0.06 **K18H:** pK_a(HisH$^+$) = 6.36 ± 0.08, and pK_{loop}(Lys) = -3.59 ± 0.11 **Q11H:** pK_a(HisH$^+$) = 6.58 ± 0.11, and pK_{loop}(Lys) = -3.31 ± 0.34 were the other parameters obtained from fit of eq 4 to the 6M GuHCl titration data. pK_a(LysH$^+$) was set to 10.5 for these fits.

At 4 M GuHCl, only loop sizes greater than 70 show evidence for biphasic loop formation.[31, 46] The CBP variants with loop sizes from 38 – 68 were fit to Eq. 4-4. In eq 4-4, pK_{loop}(His) and pK_{loop}(Lys) are the pK values for the binding of a fully deprotonated histidine or lysine to the heme, respectively, and pK_a(HisH$^+$) and pK_a(LysH$^+$) are the acid dissociation constants for the histidine and the lysine.

(4-4) $$A_{398orr} = \frac{A_{HS} + A_{LS}\left(\left(\frac{10^{-pK_{loop}(His)}}{1+10^{pK_a(HisH^+)-pH}}\right)+\left(\frac{10^{-pK_{loop}(Lys)}}{1+10^{pK_a(LysH^+)-pH}}\right)\right)}{1+\left(\frac{10^{-pK_{loop}(His)}}{1+10^{pK_a(HisH^+)-pH}}\right)+\left(\frac{10^{-pK_{loop}(Lys)}}{1+10^{pK_a(LysH^+)-pH}}\right)}$$

The other parameters are as defined in Eq 3 (Experimental procedures). The parameters from the fit of the CBP data to this equation are provided in Table 4-2. The pK_a(HisH$^+$) values vary, but for the most part, are consistent with a pK_a(HisH$^+$) value equal to 6.76 ± 0.14 from 3 to 6 M.[46]

As previously mentioned, His-heme loop formation equilibria is a two-step process, histidine ionization then binding of ionized histidine to heme to form the His-heme loop. This two-step is shown in Eq 4-5.

(4-5) $$pK_a(obs) = pK_a(HisH^+) + pK_{loop}(His)$$

Subtracting pK_a(HisH$^+$) (= 6.6)[46] from each pK_a(obs) value in Table 4-2 or using the fit pK_{loop} values from the two-state fits yields pK_{loop}(His) for every loop size (Table 4-2). Distinct from previous foldable three-helix bundle sequences, pK_{loop}(His) nearly becomes constant for loop sizes that range from 38 - 65.

His-Heme Loop Breakage Kinetics By Stopped-Flow. It has been well documented that histidine-heme loop breakage and loop formation follows a model involving deprotonation of histidine followed by the binding of the ionized histidine to the heme.[55] The observed rate constant, k_{obs}, is pH dependent and is governed by eq 4-6,[46] where k_b and k_f are the rate constants for loop breakage and loop formation.

(4-6) $$k_{obs} = k_b + k_f\left(\frac{K_a(HisH^+)}{K_a(HisH^+)+[H^+]}\right)$$

The acid dissociation constant, $K_a(HisH^+)$ corresponds to histidine deprotonation. When pH<<p$K_a(HisH^+)$, $k_{obs} \approx k_b$. Stopped-flow pH jump experiments track histidine-heme or lysine-heme loop breakage kinetics (and Lysine-heme loop breakage kinetics for the WT variant) at 6 M and 4 M GuHCl at 25 °C. Loop breakage data follows single exponential kinetics (Figure 4-S2). Measurements were made at pH 3.5 and 3.0 to confirm $k_{obs} \approx k_b$ had been attained (Table 4-S5).

The loop breakage data for CBP fused to iso-1-Cyt*c* in 6 M and 4 M GuHCl solutions are depicted in Figure 4-4. Notably, there is no significant change in k_b at 4 M and 6 M GuHCl; however, these His-heme loops in CBP vary in an irregular pattern versus loop size. Polyalanine His-heme loops are shown in Fig. 4-4 as well, showing that as loop size increases for these loops, k_b starts to plateau, as expected for random coil behavior. [29] The independence with respect to GuHCl concentration, suggest a high level of disorder, but the irregular scatter as a function of loop size indicates some local deviation from random coil behavior. Only the His25-heme loop shows significant dependence of k_b on GuHCl concentration. Kinetic data at 4 M GuHCl in Figure 4-4 is corrected for the viscosity change between the two GuHCl solutions (Table 4-S5),[29, 31] with 6 M GuHCl used as the reference state. As mentioned, k_b appears to be insensitive to GuHCl concentration, yet the significant loops are the following at 6 M GuHCl: His53-heme (21 s^{-1}), His38-heme (37 s^{-1}), and His8-heme (63 s^{-1}). The only loop that appears to be sensitive to GuHCl concentration is His25-heme, where k_b transitions from 102 s^{-1} (6 M GuHCl) to 77 s^{-1} (4 M GuHCl).

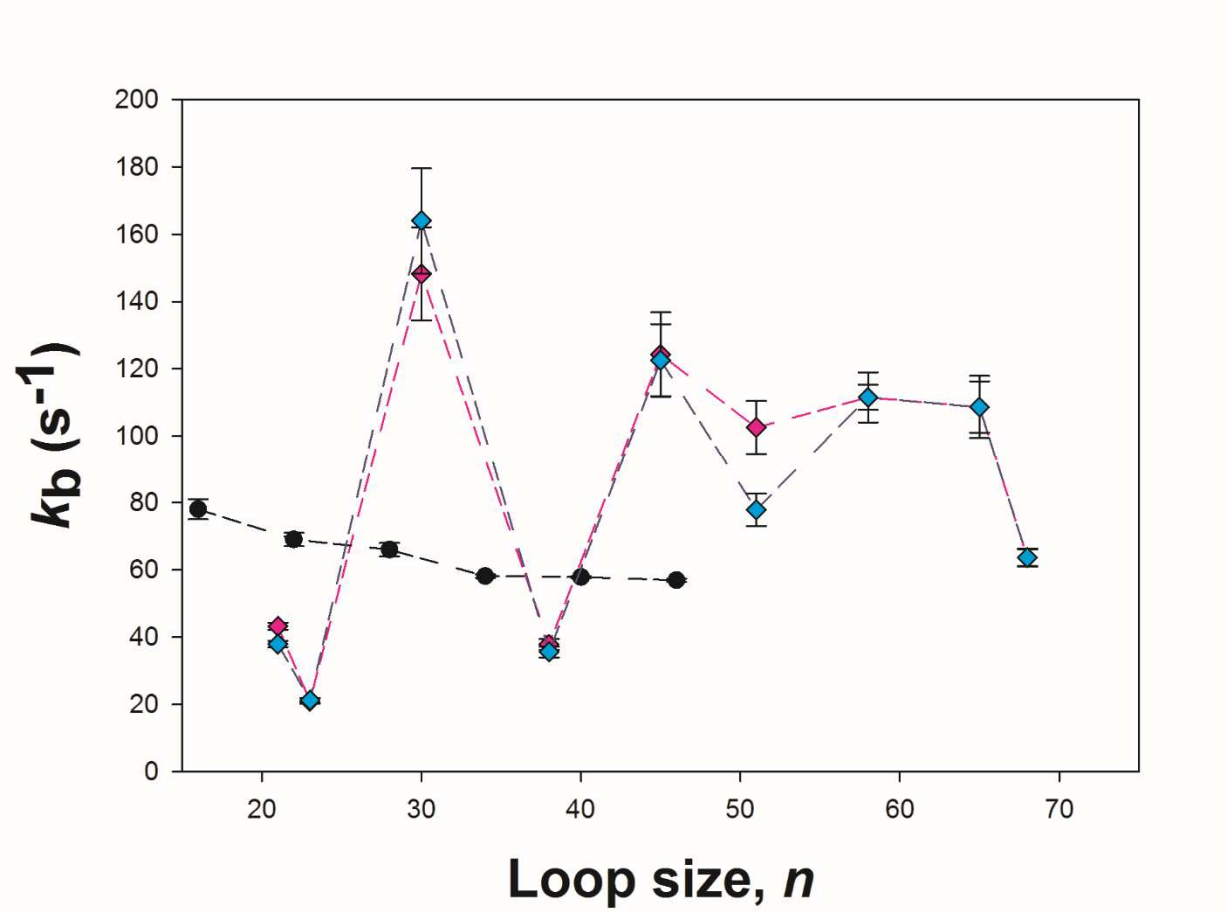

Figure 4-4. Loop breakage kinetics versus loop size, *n*, at 25 °C for CBP – iso-1-Cyt*c* variants in 6 M and 4 M GuHCl. The k_b values at 4 M GuHCl have been corrected for viscosity differences between 4 and 6 M GuHCl, $k_{b,corr}$(4M). Homopolymeric inserts of alanine $(KAAAAA)_n$[29] (n = 1 – 5) inserted between Phe(-3) and Lys(-2) of iso-1-Cyt*c* at 6 M GuHCl are included for comparison.[29]

DISCUSSION

CBP – iso-1-Cyt*c* Fusion Protein Stability Effects With Single Histidine Substitutions in the CBP domain.

Previous global stability experiments using GuHCl on the iso-1-Cyt*c*-UBA domains allowed assignment of the first phase of unfolding to the Cyt*c* domain and the second phase to the UBA domains.[82] We observe similar behavior with the CBP domain, with iso-1-Cyt*c* unfolding first followed by CBP. The WT yields a $C_{mNI} = 0.54$ for the iso-1-Cyt*c* domain, while all variants with single histidines engineered into the CBP domain have C_m values between 0.28 and 0.54 (Figure 4-2, Table 4-1). Therefore, similar to the foldable three-helix bundles fused to Cyt*c*, except for the Q25H and S8H variants, all other variants with histidine residues engineered at solvent accessible surface positions of the CBP domain slightly destabilize the iso-1-Cyt*c* domain. The impact of single histidine mutations on the overall global stability of the CBP domain appears to be mostly a stabilizing one. The CBP domain WT variant has C_m that is 0.9 M, with $\Delta G_{ID}^{o\prime}(H_2O)$ equal to ~0.5 kcal/mol and m_{ID} ~0.5 kcal/molM. For all single histidine variants, C_m is greater than WT. The values for $\Delta G_{ID}^{o\prime}(H_2O)$ for CBP domain of the WT variant is uniformly less than $\Delta G_{ID}^{o\prime}(H_2O)$ values for the single-histidine variants.

The observation that introducing one histidine into the CBP domain mostly does not affect the stability of the iso-1-Cyt*c* domain (relative to WT) for most mutants implies unfolding of the iso-1-Cyt*c* domain is mostly unaffected by His-heme binding from the histidine in the CBP domain. This result suggests that the His-heme loops are uniformly less stable for the CBP-iso-1-Cyt*c* fusion protein versus the UBA-iso-1-Cyt*c* fusion proteins. This has been observed with

foldable three-helix bundles fused to iso-1-Cyt*c*.[82] The C_{mID} for unfolding of the CBP domain is at least ~0.7 M higher than C_{mNI} for the iso-1-Cyt*c* domain in the single histidine variants. This suggests that the CBP domain is still folded near the end of the first phase of the biphasic unfolding transition. This is in agreement with the foldable three-helix bundles, where these domains are folded at the end of the first transition during GuHCl denaturation. This implies CBP can be folded and bind to the heme, although this His-heme loop interaction for the most part, is weak (see Discussion on His-heme loop thermodynamics). Based on increasing loop size corresponding to increase in C_{mID} (see Table 4-1) for the CBP single histidine variants, it appears that the further the histidine is away from the heme in Cyt*c*, the more stable the CBP domain is. For the majority of these variants (except A53H and S8H), $\Delta G_{ID}^{o\prime}(H_2O)$ values are increased, suggesting that the histidine –heme interaction in itself can nucleate structure in the CBP domain. This agrees with many studies on intrinsically disordered proteins, where coupled binding to form a specific topology is often observed.[88] Overall, we observe similarities with disordered and foldable three-helix bundle sequences when fused to Cyt*c*; however, differences are observed as well.

CBP His-Heme Loop Thermodynamics & Kinetics Relative To Random Coil Behavior. We have shown in previous chapters that loop formation probability can be modeled with the Jacobson-Stockmayer equation (eq 4-1, see Chapter 2 for detail).[28] Briefly, the entropy of a loop, ΔS_{loop}, can be equated to ΔG_{loop}(His) using eq 4-7.

$$\Delta G_{loop}(\text{His}) = ln(10)\text{RTp}K_{loop}(\text{His}) = -\text{T}\Delta S_{loop} \quad (4\text{-}7)$$

Using the equation for ΔS_{loop} in eq 4-1, eq 4-7 is simplified to eq 4-8,

(4-8) $pK_{loop}(His) = pK_{loop}(His)_{ref} + \nu_3 \log(n)$

where $pK_{loop}(His)_{ref}$ is nominally $pK_{loop}(His)$ for a loop size of 1 and ν_3 is the scaling exponent for loop formation. Equation 4-8 is the equation of a line, thus for a loop exhibiting random coil behavior, $pK_{loop}(His)$ is proportional to the logarithm of loop size, n, with a slope equal to ν_3. A random coil with excluded volume is expected to have ν_3 between 1.8 - 2.4.[58-60] Figure 4-4 depicts a plot of $pK_{loop}(His)$ versus loop size, n, on a logarithmic scale for pK_{loop} data measured at 4 M and 6 M GuHCl for the CBP domain. Homopolymer His-heme loops composed of alanine and glutamine repeats inserted at same position in iso-1-Cyt*c* as for the CBP domain are included for comparison.[29,51,52] Notably there is not a strong dependence on $pK_{loop}(His)$ versus loop size between loops sizes 38 – 65. Moreover, even in different denaturing conditions of GuHCl, $pK_{loop}(His)$ is not impacted significantly for His-heme loop sizes of 38 – 65. Specifically, there is strong deviation from random coil behavior for the S8H and A53H variants. The $pK_{loop}(His)$ for S8H is 0.57 and 0.52 units more negative than expected from the best fit lines to eq 4-5 at 4 and 6 M GdnHCl, respectively. Moreover, $pK_{loop}(His)$ for A53H is 0.58 and 0.51 units more negative than expected from the best fit lines to eq 4-8 at 4 and 6 M GuHCl In this way, the His8-heme and His53-heme loops deviate from random coil behavior by ~0.78 kcal/mol and ~0.69 kcal/mol, respectively [$\Delta\Delta G_{loop}(His) = ln(10)RT\Delta pK_{loop}(His)$].

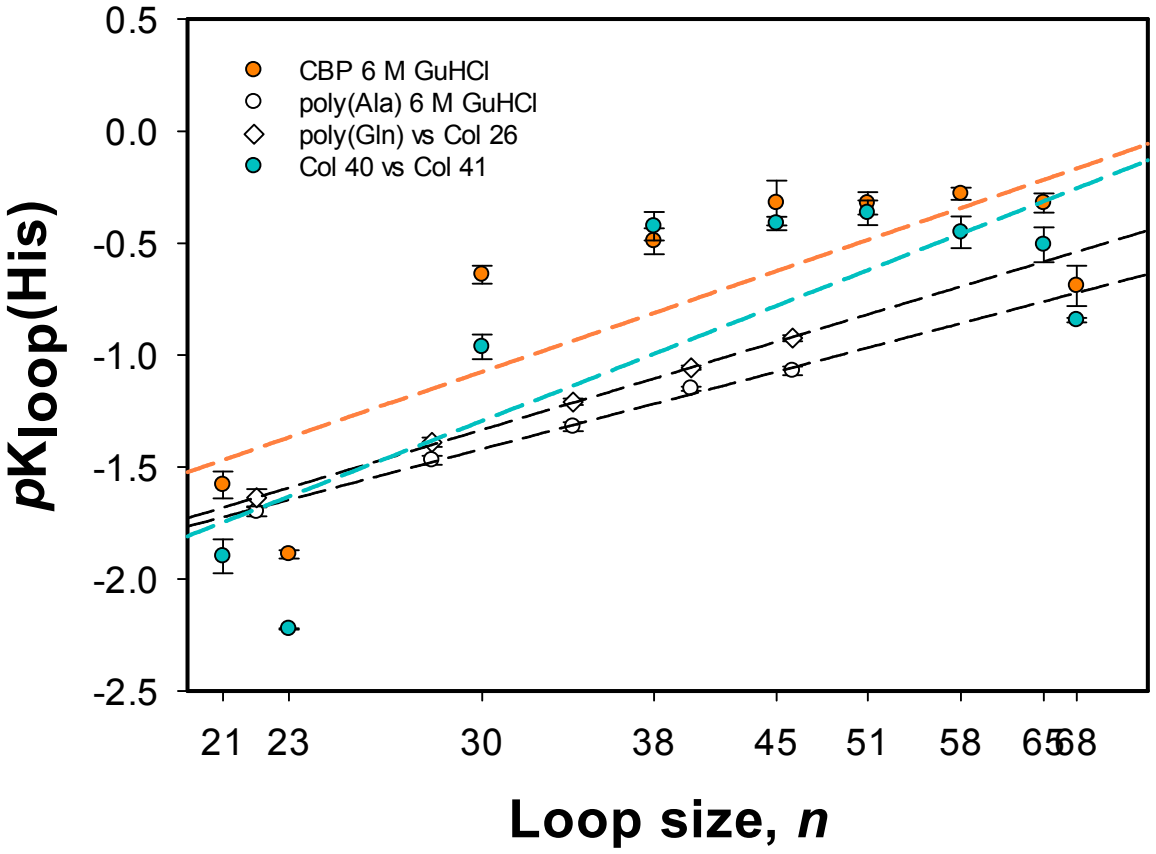

Figure 4-5. Plot of loop stability, pK_{loop}(His), vs. loop size (logarithmic scale) in the denatured state of the intrinsically disordered CBP domain fused to the N-terminus of iso-1-Cyt*c* for each histidine variant in 4 M (cyan) and 6 M (orange) GuHCl. Data from homopolymeric inserts of alanine $(KAAAAA)_n$[29] and glutamine $(KQQQQQ)_n$[51] (n = 1 – 5) inserted between Phe(-3) and Lys(-2) of iso-1-Cyt*c* as for the CBP domain (see Figure 1) measured at 6 M GuHCl are included for comparison.

We have previously observed in foldable three-helix bundles that the plots of pK_{loop}(His) versus Log(*n*) in 6 M GuHCl (Figure 4-5) contain loop stabilities that are weaker, relative to loops with homopolymeric inserts of approximate similar length in 6 M GuHCl.[29] In the three-

helix bundle UBA(2), it was discovered that the 4 M GuHCl pK_{loop}(His) versus Log(n) data overlaid well with poly(Ala) and poly(Gln) at 6 M GuHCl, not the UBA(2) pK_{loop}(His) versus Log(n) data at 6 M GuHCl. We do not observe this with CBP at 6 M nor 4 M GuHCl, as pK_{loop}(His) versus Log(n) data at both GuHCl concentrations is well above the pK_{loop}(His) data for homopolymeric inserts at 6 M GuHCl, and also well above the pK_{loop}(His) data for UBA domains at 6 M GuHCl (see Fig. 3-7). These data suggest that CBP may have a much more extended denatured state than poly(Gln) and poly(Ala) inserts, as well as the foldable UBA domains. These observations are in agreement with previous findings on the denatured state of proteins, where foldable or disordered sequences were observed to remain expanded.[83]

Histidine-Heme loop breakage kinetic data depicted in Figure 4-4 show to a certain extent variation in loop breakage as a function of loop size. Small rate constants for loop breakage, k_b, are observed for the S8H, Q38H, and the A53H variants. The S8H and A53H loop breakage data correspond well with the thermodynamic data with pK_{loop}(His), where decreases in pK_{loop}(His) are observed from 6 M to 4 M GuHCl. On the other hand, the Q38H His-heme loop does not show a drop in pK_{loop}(His) from 6 M to 4 M GuHCl, indicating that the kinetics of this His-heme loop breakage are independent of the equilibria for loop formation. Overall, these His-heme loops located on the N and C-termini are prone to structure stabilization in the denatured state of CBP, while the His38-heme loop may also be involved in formation of residual structure in the denatured state. In CBP, the His8-heme loop resides at the beginning of helix 1, while the His53-heme loop is located at the N-terminus in an unstructured part of the disordered sequence. The His38-heme loop is located near the end of a turn between helix 2 and 3 (Figure 4-1). These results are somewhat consistent with our previous work on the four-helix bundle protein, Cytc', and the three-helix bundle UBA domains, where persistent His-heme loops are found at or near

turn regions between helices (For Cyt*c*', helices 1 and 2 and at the ends of the Ω-loop connecting helices 2 and 3, while for the UBA domains, helices 2 and 3 in UBA(2), and both turns in UBA(1)).

It is worth noting that several of these His-heme loops in these three-helix bundles (UBA(1), UBA(2), and CBP) that are persistent in denaturing conditions are adjacent to or near aromatic or proline residues in these sequences. Proline – aromatic residues in proteins have been known to interact favorably with each other, partially because of proline's restricted conformation.[90] Histidine in different charged states (neutral or positively charged) can form π-π or cation-π interactions.[89] These interactions are known to occur in proteins, and could be a contribution to the stabilization observed in persistent residual structure in these specific sequences of UBA and CBP in denaturing conditions. For the persistent turn in UBA(2), the sequence is Tyr-Phe-Ala-Ala-His-Lys-Asn. For the persistent turns in UBA(1), the sequence for turn 1 (between helices 1 and 2) is Met-Gly-Tyr-Phe-His-Arg, while the sequence for turn 2 (between helices 2 and 3) is Tyr-Asn-Asn-Pro-His-Arg. Based on previous work with the four-helix bundle Cyt*c*', it was observed using molecular dynamics that hydrophobic side chains make continual contacts with the persistent reverse turn structure via temperature unfolding simulations.[31] This evidence of hydrophobic contacts stabilizing turn regions conjectures the participation of these side chains stabilizing persistent structure in the DSE for the foldable three-helix bundles: Tyr23 and Phe24 for the second reverse turn of UBA(2), Tyr14 in the first turn of UBA(1), and Tyr27 and Pro30 in the second turn of UBA(1). In UBA(1), however, Tyr27 was discovered to be insignificant in stabilizing its second reverse turn.

For the CBP domain, the amino acid sequences corresponding to the most persistent regions are the following (Histidines probing these regions in red): Ser-Pro-His-Ala-Leu at the start of

helix 1, Asn-Pro-His-Leu-Met in the turn between helix 2 and 3, and Tyr-Val-His-Asn-Gln-Pro in the unstructured region between two turns. Thus, similar to the UBA domains, nearby proline resides or aromatics such as Pro7 (start of helix 1) or Tyr51 (unstructured region between two turns) could be contributing to the presence of residual structure in the denatured state of CBP. Many of the His-heme loops in the CBP sequence that do not contain this thermodynamic and kinetic bias contain histidine probes further away from aromatic or proline residues.

Previous work in our lab has shown that an Ala→Trp mutation three residues from a histidine probe influences His-heme loop stabilization by ~0.3 kcal/mol in 6 M GuHCl and ~0.7 kcal/mol in 3 M GuHCl.[50] Moreover, other work has shown that in 3 M GuHCl when Ala is replaced by aromatics (Phe, Tyr, or Trp) four residues from the histidine probe, that the His-heme loop increases in stability by 0.4 – 0.5 kcal/mol, while switch to Leu from the same Ala residue shows no significant stabilization in the His-heme loop.[61] Aromatic side chains have been previously observed to strengthen residual structure in denatured proteins and also play a role in long range contacts.[9, 10, 62-66] These observations support the premise that aromatics and proline residues contribute to the stabilization of persistent structure in the DSE of proteins. Although, in polymeric inserts with alanine, proline is found to have little effect on His-heme loop thermodynamics.[29] Perhaps the amino acid sequences within the loop are needed for proline to stabilize a His-heme loop.

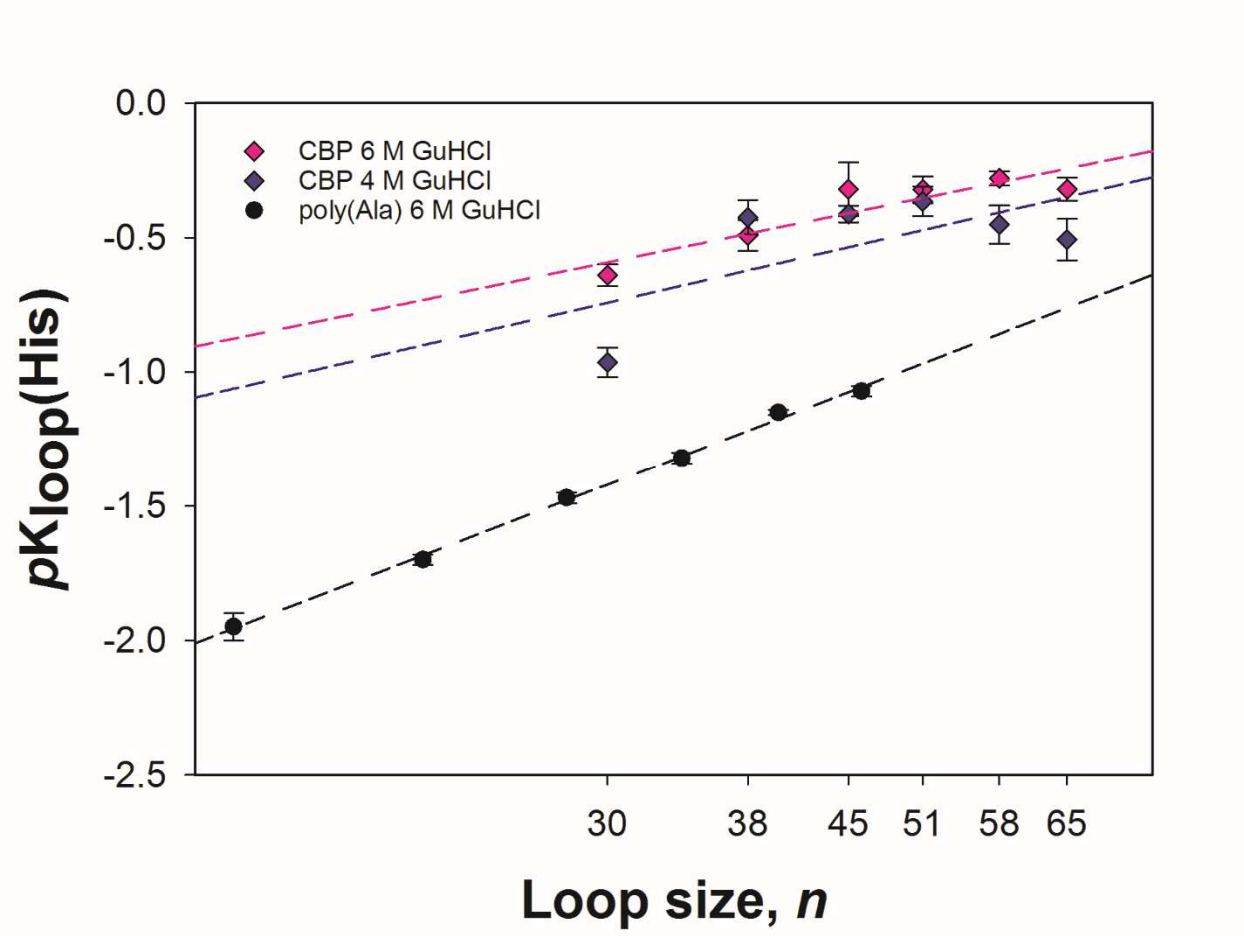

Fig. 4-6. His-heme loop stability plot, pK_{loop}(His) vs. loop size (logarithmic scale) in the denatured state of the intrinsically disordered CBP domain fused to the N-terminus of iso-1-Cyt*c* for subset of histidine variants in 4 M (purple) and 6 M (pink) GuHCl. Data from homopolymer alanine $(KAAAAA)_n$[29] included for comparison.

Foldable Versus Disordered Sequence in the Denatured States of 3-Helix Bundles.

Compared to foldable three-helix bundles, this behavior in pK_{loop}(His) in CBP is drastically different. In the foldable UBA domains (see Figures 2-5 and 3-5), the magnitude of scatter in pK_{loop}(His) between 4 M and 6 M GuHCl for many of the UBA single histidine variants is much greater. Because it appears that effects from the Cyt*c* sequence has an impact on the shorter loops

of CBP (see Fig. 4-5, loops 21 and 23), we graphed a plot of pK_{loop}(His) versus loop size, *n*, on a logarithmic scale, for loops that primarily are composed of the CBP sequence. This plot is displayed in Figure 4-6. For reference, homopolymer inserts of alanine acquired at 6 M GuHCl are included. Fits of pK_{loop}(His) versus Log(*n*) yield scaling exponents, ν_3, of 1.1 and 1.2 at 4 and 6 M GuHCl, respectively, for loops containing mostly the CBP sequence. These values are less than the range of 1.8 – 2.4 expected for a random coil with excluded volume.[58-60] These scaling exponent values for CBP drastically differ from scaling exponents of 1.9 to 2.3 observed in 6 M GuHCl for His-heme loops with poly(Gly), poly(Gln) and poly(Ala) inserts into the N-terminus of iso-1-Cyt*c* (see Figure 5).[29, 51, 52] For the UBA(2) domain, the scaling exponents ν_3 is 2.2 and 2.4 at 6 M and 4 M GuHCl, well within the range expected for a random coil with excluded volume. Given that folded proteins have a radius of gyration that scales with ν equal to approximately 0.3 (i.e. $R_g \alpha N^{0.3}$), the scaling exponent, ν_3, for loop formation with a compact folded protein should be approximately equal to 1. With CBP's scaling exponents equal to 1.1 and 1.2 in 4 M and 6 M GuHCl, this suggests that CBP is mostly compact, thus behaving as a folded protein. Because of this compactness in the denatured state, this may be affecting our thermodynamic and kinetic data observed at 4 M and 6 M GuHCl, as we observe these parameters hardly change between the two conditions. This interpretation would agree with our GuHCl unfolding data that suggests introducing a histidine to disordered sequence initiates structure formation.

CONCLUSIONS

Fusion of the intrinsically disordered domain CBP to the N-terminus of iso-1-Cyt*c* has successfully allowed denatured state His-heme loop formation thermodynamic and kinetic studies on a intrinsically disordered protein, which forms a three-helix bundle when bound to its

signaling partner ACTR. We observe that the dimensional properties of this disordered domain in denaturing conditions (4 and 6 M GuHCl) are vastly different than foldable three-helix bundle sequences, yielding a weak dependence on pK_{loop}(His) versus loop size. For loops that primarily contain the CBP sequence, we do observe that the CBP domain does not adhere to a model for a random coil with excluded volume at 6 M GuHCl (ν_3 = 1.2), with similar behavior in 4 M GuHCl (ν_3 = 1.1). Several portions of the CBP disordered chain are invariant in pK_{loop}(His) versus loop size at 4 M and 6 M GuHCl; however, we do observe His-heme loop formation increasing in strength in loops that are located on the N and C-terminus of CBP. These histidines that probe the N and C terminal ends also have low loop breakage rate constants, indicating these portions of primary structure in the disordered chain may have substantial stabilization in the denatured state. While the His38-heme loop is within a region of CBP where pK_{loop}(His) is invariant with loop size at both 4 M and 6 M GuHCl, its low k_b value (~37 s^{-1}) suggests His38 has both slow kinetics for loop formation and breakage, indicating that the DSE of CBP has unusual properties. These observations are distinct from the four-helix bundle Cyt*c′* and the three-helix bundle UBA domains, where persistent denatured state His-heme loops tend to be near sequences that form turns in their tertiary structures.[31,82] Loop formation is less favorable than foldable and homopolymeric sequences, which implies a large portion of the denatured CBP domain may be expanded. This has been previously observed for other disordered proteins[83].

SUPPORTING INFORMATION

Tables 4-S1 to 4-S3 provide sequences for oligonucleotides used for cloning and mutagenesis. Table 4-S4 contains MALDI-TOF data for the CBP – iso-1-Cyt*c* variants.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bruce.bowler@umontana.edu

ORCID

Bruce E. Bowler: 0000-0003-1543-2466

Funding

This research was supported by National Science Foundation grant, MCB-1412164 and NIH grant, R01GM074750, to B.E.B. The Bruker microflex MALDI-TOF mass spectrometer was purchased with Major Research Instrumentation Grant CHE-1039814 from the National Science Foundation. M.J.L. acknowledges the Sloan Indigenous Graduate Partnership of the Alfred P. Sloan Foundation and the National Science Foundation DEB 0614406 and the NSF EPSCoR Track-1 EPS-1101342 (INSTEP 3) for graduate education support.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

iso-1-Cyt*c*, iso-1-cytochrome *c*; CD, circular dichroism; GuHCl, guanidine hydrochloride; CBP, cAMP-responsive element binding protein ; CBP – iso-1-Cyt*c*, fusion protein with CBP inserted at the N-terminus of iso-1-Cyt*c*; WT, CBP-iso-1-Cyt*c* native sequence with no histidine

Supporting Information

Unusual Behavior in the Denatured State Properties of the Instrinsically Disordered Protein CBP

Moses J. Leavens, Melisa M. Cherney, and Bruce E. Bowler*

Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Dynamics, University of Montana, Missoula, Montana 59812

*To whom correspondence should be addressed.

Telephone: (406) 282-1883. Fax: (406) 243-4227. E-mail: bruce.bowler@umontana.edu

Table 3-S1. Oligonucleotide primers for preparation of CBP-iso-1-Cyt*c* fusion protein[a]

Primer	Primer Sequence 5′ to 3′
CBPER1_for	d(ATATGAATTCCCAAACAGGAGCATCTCG)
CBPNgoM4_rev	d(TTAAGCCGGCCTTCATGCCAGGCTGATTGGC)
CBPfix	d(TCAGCCTGGCATG**CAG**AAGGCCGGCTCTG)
CBPfix-r	d(CAGAGCCGGCCTT**CTG**CATGCCAGGCTGA)

[a]Restriction sites are underlined; inserted bases are bolded.

Table 4-S2. Oligonucleotide Primers for Preparation of CBP-iso-1-Cyt*c* Variants

Primer	Primer Sequence 5′ to 3′
Q55H	d(ccaagtatgtggccaatcaccctggcatgc)
Q55H-r	d(gcatgccagggtgattggccacatacttgg)
A53H	d(gcacagccaagtatgtgcacaatcagcctggcatgc)
A53H-r	d(gcatgccaggctgattgtgcacatacttggctgtgc)
Q46H	d(ctaatggcagctttcatcaaacaccgcacagccaag)
Q46H-r	d(cttggctgtgcggtgtttgatgaaagctgccattag)
Q38H	d(ccttaaatcaaacccacacctaatggcagctttcatc)
Q38H-r	d(gatgaaagctgccattaggtgtgggtttgatttaagg)
N31H	d(gcagcagcaggtgctgcacatccttaaatcaaac)
N31H-r	d(gtttgatttaaggatgtgcagcacctgctgctgc)
Q25H	d(gctctcctcagcaccagcagcaggtgc)
Q25H-r	d(gcacctgctgctggtgctgaggagagc)
K18H	d(ggagagctgggtgagtgtagggtccgtagcagg)

K18H-r	d(cctgctacggaccctacactcacccagctctcc)
Q11H	d(ccaagtgccctgcatgacctgctacggac)
Q11H-r	d(gtccgtagcaggtcatgcagggcacttgg)
S8H	d(cccaaacaggagcatctcgccacatgccctgcaaga)
S8H-r	d(tcttgcagggcatgtggcgagatgctcctgtttggg)

Table 4-S3. MALDI-ToF Mass Spectral Data for UBA(1) - iso-1-Cyt*c* variants

Variant	m/z observed	m/z expected[a]
WT	19,096.51	19,082.61
Q55H	19,085.13	19,091.63
A53H	19,167.17	19,148.68
Q46H	19,091.99	19,091.63
Q38H	19,098.99	19,091.63
N31H	19,115.27	19,105.65
Q25H	19,096.13	19,091.63
K18H	19,081.43	19,091.58
Q11H	19,091.50	19,091.63
S8H	19,129.01	19,132.68

[a]Expected m/z obtained using ExPASy PeptideMass tool (http://web.expasy.org/peptide_mass/) using average mass, no cutting and $[M+H]^+$ options.

Table 4-S4. Loop breakage rate constants for CBP-Cyt*c* Variants in 6 M and 4 M GuHCl at pH 3.0.

Variant	Loop size	k_b pH 3 (6M)	k_b pH 3.5 (6M)	k_b pH 3 (4M)	k_b pH 3.5 (4M)
WT	-	103.6 ± 9.4	-	116 ± 5.9	-
Q55H	21	43.1 ± 1.1	-	48.2 ± 1	-
A53H	23	21.0 ± 0.8	-	27.0 ± 0.7	-
Q46H	30	148.3 ± 13.8	-	208.7 ± 15.6	-
Q38H	38	37.8 ± 1.6	-	45.2 ± 1.6	-
N31H	45	124.3 ± 12.5	-	156.1 ± 10.6	-
Q25H	51	102.6 ± 8	-	99.0 ± 4.9	-
K18H	58	111.6 ± 7.4	-	72.1 ± 3.8	-
Q11H	65	108.7 ± 7.6	-	102.6 ± 9.23	-
S8H	68	63.6 ± 2.5	-	68.9 ± 2.8	-

REFERENCES FOR CBP-CYT C

1. Mittag, T., and Forman-Kay, J. D. (2007) Atomic-level characterization of disordered protein ensembles, *Curr. Opin. Struct. Biol. 17*, 3-14.

2. Cho, J.-H., Sato, S., Horng, J.-C., Anil, B., and Raleigh, D. P. (2008) Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability, *Arch. Biochem. Biophys. 469*, 20-28.

3. Bowler, B. E. (2012) Globular proteins: characterization of the denatured state, In *Comprehensive Biophysics* (Egelman, E., Ed.), pp 72-114, Academic Press, Oxford.

4. Bowler, B. E. (2007) Thermodynamics of protein denatured states, *Mol. BioSyst. 3*, 88-99.

5. Dill, K. A., and Shortle, D. (1991) Denatured states of proteins, *Annu. Rev. Biochem. 60*, 795-825.

6. Zhang, O., Kay, L. E., Shortle, D., and Forman-Kay, J. D. (1997) Comprehensive NOE characterization of a partially folded large fragment of staphylococcal nuclease D131D, using NMR methods with improved resolution, *J. Mol. Biol. 272*, 9-20.

7. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures, *J. Mol. Biol. 268*, 179-184.

8. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels, *J. Mol. Biol. 268*, 158-169.

9. Wirmer, J., Schloerb, C., Klein-Seetharaman, J., Hirano, R., Ueda, T., Imoto, T., and Schwalbe, H. (2004) Protein interactions: modulation of compactness and long-range interactions of unfolded lysozyme by single point mutations, *Angew. Chem. Int. Ed. 43*, 5780-5785.

10. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein, *Science 295*, 1719-1722.

11. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea, *Biochemistry 36*, 8977-8991.

12. Neri, D., Billeter, M., Wider, G., and Wüthrich, K. (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor, *Science 257*, 1559-1563.

13. Pakula, A. A., and Sauer, R. T. (1990) Reverse hydrophobic effects relieved by amino acid substitutions at a protein surface., *Nature 344*, 363–364.

14. Bowler, B. E., May, K., Zaragoza, T., York, P., Dong, A., and Caughey, W. S. (1993) Destabilizing effects of replacing a surface lysine of cytochrome *c* with aromatic amino acids: implications for the denatured state, *Biochemistry 32*, 183-190.

15. Bowler, B. E., Dong, A., and Caughey, W. S. (1994) Characterization of the guanidine hydrochloride-denatured state of iso-1-cytochrome *c* by infrared spectroscopy, *Biochemistry 33*, 2402-2408.

16. Herrmann, L., Bowler, B. E., Dong, A., and 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.

17. Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions, *Biochemistry 38*, 4896-4903.

18. Cho, J.-H., and Raleigh, D. P. (2005) Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins, *J. Mol. Biol. 353*, 174-185.

19. Cho, J.-H., Sato, S., and Raleigh, D. P. (2004) Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state, *J. Mol. Biol. 338*, 827-837.

20. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M., and Pace, C. N. (1999) Increasing protein stability by altering long-range coulombic interactions, *Protein Sci. 8*, 1843-1849.

21. Pace, C. N., Alston, R. W., and Shaw, K. L. (2000) Charge-charge interactions influence the denatured state ensemble and contribute to protein stability, *Protein Sci. 9*, 1395-1398.

22. Tan, Y. J., Oliveberg, M., Davis, B., and Fersht, A. R. (1995) Perturbed pK_A-values in the denatured states of proteins, *J. Mol. Biol. 254*, 980-992.

23. Whitten, S., and Garcia-Moreno E., B. (2000) pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state., *Biochemistry 39*, 14292–14304.

24. Swint-Kruse, L., and Robertson, A. D. Hydrogen bonds and the pH dependence of ovomucoid third domain stability, *Biochemistry 34*, 4724–4732.

25. Trefethen, J. M., Pace, C. N., Scholtz, J. M., and Brems, D. N. (2005) Charge-charge interactions in the denatured state influence the folding kinetics of ribonuclease Sa, *Protein Sci. 14*, 1934-1938.

26. Cho, J.-H., and Raleigh, D. P. (2006) Electrostatic interactions in the denatured state and in the transition state for protein folding: effects of denatured state interactions on the analysis of transition state structure, *J. Mol. Biol. 359*, 1437-1446.

27. Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988) Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding, *Biochemistry 27*, 7167-7175.

28. Jacobson, H., and Stockmayer, W. H. (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems, *J. Chem. Phys. 18*, 1600-1606.

29. Tzul, F. O., and Bowler, B. E. (2010) Denatured states of low complexity polypeptide sequences differ dramatically from those of foldable sequences, *Proc. Natl. Acad. Sci. U.S.A. 107*, 11364-11369.

30. Rao, K. S., Tzul, F. O., Christian, A. K., Gordon, T. N., and Bowler, B. E. (2009) Thermodynamics of loop formation in the denatured state of *Rhodopseudomonas palustris* cytochrome *c*': scaling exponents and the reconciliation problem, *J. Mol. Biol. 392*, 1315-1325.

31. Dar, T. A., Schaeffer, R. D., Daggett, V., and Bowler, B. E. (2011) Manifestations of native topology in the denatured state ensemble of *Rhodopseudomonas palustris* cytochrome *c*', *Biochemistry 50*, 1029-1041.

32. Duncan, M. G., Williams, M. D., and Bowler, B. E. (2009) Compressing the free energy range of substructure stabilities in iso-1-cytochrome *c*, *Protein Sci. 18*, 1155-1164.

33. Pollock, W. B., Rosell, F. I., Twitchett, M. B., Dumont, M. E., and Mauk, A. G. (1998) Bacterial expression of a mitochondrial cytochrome *c*. Trimethylation of Lys72 in yeast iso-1-cytochrome *c* and the alkaline conformational transition, *Biochemistry 37*, 6124-6131.

34. Rosell, F. I., and 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.

35. McClelland, L. J., Seagraves, S. M., Khan, M. K. A., Cherney, M. M., Bandi, S., Culbertson, J. E., and Bowler, B. E. (2015) The response of Ω-loop D dynamics to truncation of trimethyllysine 72 of yeast iso-1-cytochrome *c* depends on the nature of loop deformation, *J. Biol. Inorg. Chem. 20*, 805-819.

36. Godbole, S., and Bowler, B. E. (1997) A histidine variant of yeast iso-1-cytochrome *c* that strongly affects the energetics of the denatured state, *J. Mol. Biol. 268*, 816-821.

37. Withers-Ward, E. S., Mueller, T. D., Chen, I. S. Y., and Feigon, J. (2000) Biochemical and structural analysis of the interaction between the UBA(2) domain of the DNA repair protein HHR23A and HIV-1 Vpr, *Biochemistry 39*, 14103-14112.

38. Cherney, M. M., Junior, C., and Bowler, B. E. (2013) Mutation of trimethyllysine-72 to alanine enhances His79-heme mediated dynamics of iso-1-cytochrome *c*, *Biochemistry 52*, 837-846.

39. Redzic, J. S., and Bowler, B. E. (2005) Role of hydrogen bond networks and dynamics in positive and negative cooperative stabilization of a protein, *Biochemistry 44*, 2900-2908.

40. Wandschneider, E., Hammack, B. N., and Bowler, B. E. (2003) Evaluation of cooperative interactions between substructures of iso-1-cytochrome *c* using double mutant cycles, *Biochemistry 42*, 10659-10666.

41. Goldes, M. E., Jeakins-Cooley, M. E., McClelland, L. J., Mou, T.-C., and Bowler, B. E. (2016) Disruption of a hydrogen bond network in human versus spider monkey cytochrome *c* affects heme crevice stability, *J. Inorg. Biochem. 158*, 62-69.

42. Margoliash, E., and Frohwirt, N. (1959) Spectrum of horse-heart cytochrome *c*, *Biochem. J. 71*, 570-572.

43. Nozaki, Y. (1972) The preparation of guanidine hydrochloride, *Methods Enzymol. 26*, 43-50.

44. Pace, C. N., Shirley, B. A., and Thomson, J. A. (1989) Measuring the conformational stability of a protein, In *Protein structure: a practical approach* (Creighton, T. E., Ed.), pp 311-330, IRL Press at Oxford University Press, New York.

45. Schellman, J. A. (1978) Solvent denaturation, *Biopolymers 17*, 1305-1322.

46. Wandschneider, E., and Bowler, B. E. (2004) Conformational properties of the iso-1-cytochrome *c* denatured state: dependence on guanidine hydrochloride concentration, *J. Mol. Biol. 339*, 185-197.

47. Tonomura, B., Nakatani, H., Ohnishi, M., Yamaguchi-Ito, J., and Hiromi, K. (1978) Test reactions for a stopped-flow apparatus. Reduction of 2,6-dichlorophenolindophenol and potassium ferricyanide by L-ascorbic acid, *Anal. Biochem. 84*, 370-383.

48. Anil, B., Song, B., Tang, Y., and Raleigh, D. P. (2004) Exploiting the right side of the Ramachandran plot: substitution of glycines by D-alanine can significantly increase protein stability, *J. Am. Chem. Soc. 126*, 13194-13195.

49. Smith, C. R., Mateljevic, N., and Bowler, B. E. (2002) Effects of topology and excluded volume on protein denatured state conformational properties, *Biochemistry 41*, 10173-10181.

50. Khan, M. K. A., Miller, A. L., and Bowler, B. E. (2012) Tryptophan significantly stabilizes His-heme loops only when it is near a loop end, *Biochemistry 51*, 3586-3595.

51. Khan, M. K. A., and Bowler, B. E. (2012) Conformational properties of polyglutamine sequences in guanidine hydrochloride solutions, *Biophys. J. 103*, 1989–1999.

52. Finnegan, M. L., and Bowler, B. E. (2012) Scaling properties of glycine-rich sequences in guanidine hydrochloride solutions, *Biophys. J. 102*, 1969-1978.

53. Xu, Y., Mayne, L., and Englander, S. W. (1998) Evidence for an unfolding and refolding pathway in cytochrome *c*, *Nat. Struct. Biol. 5*, 774-778.

54. Smith, C. R., Wandschneider, E., and Bowler, B. E. (2003) Effect of pH on the iso-1-cytochrome *c* denatured state: changing constraints due to heme ligation, *Biochemistry 42*, 2174-2184.

55. Kurchan, E., Roder, H., and Bowler, B. E. (2005) Kinetics of loop formation and breakage in the denatured state of iso-1-cytochrome *c*, *J. Mol. Biol. 353*, 730-743.

56. Rohl, C. A., Chakrabartty, A., and Baldwin, R. L. (1996) Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol, *Protein Sci. 5*, 2623-2637.

57. Chakrabartty, A., Schellman, J. A., and Baldwin, R. L. (1991) Large differences in the helical propnesity of glycine and alanine, *Nature 351*, 586-588.

58. de Gennes, P.-G. (1979) *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, NY.

59. Chan, H. S., and Dill, K. A. (1990) The effect of internal constraints on the configurations of chain molecules, *J. Chem. Phys. 92*, 3118-3135.

60. Redner, S. (1980) Distribution functions in the interior of polymer chains, *J. Phys. A: Math. Gen. 13*, 3525-3541.

61. Finnegan, M. L., and Bowler, B. E. (2010) Propensities of aromatic amino acids versus leucine and proline to induce residual structure in the denatured-state ensemble of iso-1-cytochrome *c*, *J. Mol. Biol. 403*, 495-504.

62. Crowhurst, K. A., and Forman-Kay, J. D. (2003) Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain, *Biochemistry 42*, 8687-8695.

63. Crowhurst, K. A., Tollinger, M., and Forman-Kay, J. D. (2002) Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain, *J. Mol. Biol. 322*, 163-178.

64. Marsh, J. A., and Forman-Kay, J. D. (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints, *J. Mol. Biol. 391*, 359-374.

65. Day, R., and Daggett, V. (2005) Ensemble versus single-molecule protein unfolding, *Proc. Natl. Acad. Sci. U.S.A. 102*, 13445-13450.

66. Wong, K.-B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M. V., Fersht, A. R., and Daggett, V. (2000) Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding, *J. Mol. Biol. 296*, 1257-1282.

67. Bruun, S. W., Iešmantavičius, V., Danielsson, J., and Poulsen, F. M. (2010) Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein, *Proc. Natl. Acad. Sci. U.S.A. 107*, 13306-13311.

68. Petersen, B., Lundegaard, C., and Petersen, T. N. (2010) NetTurnP - neural network prediction of beta-turns by use of evolutionary information and predicted protein sequence features, *PLoS ONE 5*, e15079.

69. Fitzkee, N. C., and Rose, G. D. (2004) Reassessing random-coil statistics in unfolded proteins, *Proc. Natl. Acad. Sci. U.S.A. 101*, 12497-12502.

70. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. N., DeMarco, M. L., Daggett, V., and Fersht, A. R. (2011) Unifying features in protein folding mechanisms, *Proc. Natl. Acad. Sci. U.S.A. 100*, 13286-13291.

71. DeMarco, M. L., Alonso, D. O. V., and Daggett, V. (2004) Diffusing and colliding: the atomic level folding/unfolding pathway of a small helical protein, *J. Mol. Biol. 341*, 1109-1124.

72. Muñoz, V., and Serrano, L. (1994) Elucidating the folding problem of a-helical peptides using empirical parameters III: temperature and pH dependence, *J. Mol. Biol. 245*, 297-308.

73. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V., and Fersht, A. R. (2005) Solution structure of a protein denatured state and folding intermediate, *Nature 437*, 1053-1056.

74. Arora, P., Oas, T. G., & Myers, J. K. (2004). Fast and faster: A designed variant of the B-domain of protein A folds in 3 microseconds. *Protein Science*, 847-853.

75. Bowler, B. (2012). Residual structure in unfolded proteins. *Current Opinion in Structural Biology*, 4-13.

76. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., . . . Wright, P. E. (2002). Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature*, 549-553.

77. Gregersen N, Bross P. (2010). Protein misfolding and cellular stress: an overview. *Methods Mol Bio, 648*, 3-23.

78. Iesmantacvicius, V., Jensen, M. R., Ozenne, V., Blackledge, M., Poulsen, F. M., & Kjaergaard, M. (2013). Modulation of the Intrinsic Helix Propensity of an Intrinsically Disordered Protein Reveals Long-Range Helix-Helix Interactions. *Journal of the American Chemical Society*, 10155-10163.

79. Jed Long, Thomas P. Garner, Maya J. Pandya, C. Jeremy Craven, Ping Chen, Barry Shaw, Michael P. Williamson, Robert Layfield, Mark S. Searle. (2010). Dimerisation of the UBA Domain of p62 Inhibits Ubiquitin Binding and Regulates NF-kB Signalling. *J. Mol. Biol., 396*, 178-194.

80. K. Sudhindra Rao, Franco O. Tzul, Arwen K. Christian, Tia N. Gordon, Bruce E. Bowler. (2009). Thermodynamics of Loop Formation in the Denatured State of Rhodopseudomonas palustris Cytochrome c': Scaling Exponents and the Reconcilation Problem. *J. Mol. Biol., 392*(5), 1315-1325.

81. Kohn, J. E., & Plaxco, K. W. (2004). Random coil behavior and the dimensions of chemically unfolded proteins. *PNAS*, 12491-12496.

82. Leavens, M. J., Cherney, M. M., Finnegan, M. L., & Bowler, B. E. (2018). Probing Denatured State Conformational Bias in the Three-Helix Bundle, UBA(2), Using a Cytochrome c Fusion Protein. *Biochemistry*, 1711-1721.

83. Riback, J. A., Bowman, M. A., Zmyslowski, A. M., Knoverek, C. R., Jumper, J. M., Hinshaw, J. R., . . . Sosnick, T. R. (2017). Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science*, 238-241.

84. Sugase, K., Dyson, H. J., & Wright, P. E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 1021-1025.

85. Thomas D. Mueller, Juli Feigon. (2002). Solution Structures of UBA Domains Reveal a Conserved Hydrophobic Surface for Protein-Protein Interactions. *J. Mol. Biol., 319*, 1243-1255.

86. Tran , H., Mao, A., & Pappu, R. (2008). Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *JACS, 130*, 7380-7392.

87. Tuefel, D., Johnson, C., Lum , J., & Neuweiler, H. (2011). Backbone driven collapse in unfolded protein chains. *J Mol Bio, 409*, 250-262.

88. Wright, P. E., & Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews*, 18-29.

89. Huang, R.B., & Liao, S.M. (2013). The multiple roles of histidine in protein interactions. *Chem Central Journal.*

90. Zondlo, Neal J. (2013). Aromatic-Proline Interactions: Electronically Tunable CH/pi Interactions. *Acc. Chem. Res.*, 1039-1049.

Chapter 5: Conclusions and Future Directions

A long standing paradigm in molecular biology is sequence → structure → function. The folding code—translation of amino acid sequence to structure, is difficult to comprehend because protein structure is far more conserved than sequence[75]. Using the denatured His-heme loop formation method developed in this laboratory, we applied this methodology by fusing three

three-helix bundles with divergent sequences to the N-terminus of Cyt*c*. Three large scale objectives in this dissertation were met: applying the His-heme denatured state loop formation method to proteins not containing a c-type heme (chapter 2, UBA(2) domain), application of this method to foldable three-helix bundles with divergent amino acid sequences (chapter 3, UBA(2) and UBA(1) domains), and applying this method to a disordered amino acid sequence that forms three-helix bundle topology when bound to its signaling partner ACTR, for direct comparison to foldable three-helix bundle sequences (chapter 4, CBP domain).

The overall results in this dissertation support the hypothesis that biases towards reverse turns in helix bundles remain in the denatured state. These results on three-helix bundles with foldable amino acid sequences are in agreement with the four-helix bundle Cyt*c*'[31]. In this way, denatured state conformational biases toward reverse turns may be a general phenomenon in all α-folds, despite divergent amino acid sequence composition for helix-bundles. In Cyt*c*' and UBA(2), it was observed that not every turn sequence under denaturing conditions contains a persistent His-heme loop, suggesting that the denatured state does not necessarily need to be fully biased towards all turns to serve as nucleation site for efficient folding to their native conformers[31,82].

The UBA(2)-Cyt*c* fusion protein was observed to contain one persistent His-heme loop, His27-heme. The His27-heme loop corresponds to the second reverse turn of UBA(2), with the amino acid sequence Phe-Ala-Ala-Glu-Lys. This region is located between helix 2 and 3. A cysteine at position 26 in the UBA(2) domain was mutated to alanine, and this mutation may have affected its conformational distribution of its denatured state. This substitution perhaps would have affected the DSE conformational distribution in the first turn of UBA(2). Previous work has shown the HHR23A UBA(2) domain GuHCl unfolding $\Delta G_u(H_2O)$ to be 1.34 ± 0.06

kcal/mol[76]. Previously, we carried out GuHCl unfolding with HHR23A UBA(2) C26A by itself, and observed a substantial increase in $\Delta G_u(H_2O)$ stability (~2.66 kcal/mol), suggesting the Cys→Ala mutation at position 26 between helices 2 and 3 stabilizes the fold of UBA(2). The second turn of UBA(2) is flanked by a hydrophobic environment, with Tyr23 and Phe24 on the N-terminus of the second turn. We have previously observed in Cyt*c*', that turns persistent in the denatured state make persistent contacts with hydrophobic residues in the denatured state, and in UBA(2), the cysteine at 26 may be a part of the hydrophobic core of HHR23A UBA(2) domain. The cysteine residue does not appear to be near the surface of this protein[84].

For UBA(1), both turn regions were detected to contain persistent structure in the denatured state. Why turn 1 in the UBA(2) domain was not detected is interesting, because the amino acid sequences of turn 1 in UBA domains across all species are highly conserved (i.e. LGY or MGY motif). Both UBA(2) and UBA(1) were isolated from the human DNA excision repair protein, HHR23A, a human DNA excision repair protein. The UBA(2) domain is the C-terminal UBA domain in HHR23A, while the UBA(1) domain is the internal UBA domain in HHR23A. Both UBA domains function in Ubiquitin/proteasome signaling. Perhaps there is some slight difference in the functional aspects of these domains as to why the DSE conformational bias is different, or, as mentioned above, our histidine probe was not in an ideal position to detect the bias in turn 1 in UBA(2), or the Cys→Ala mutation at position 26 in UBA(2) had an impact on the denatured state conformational distribution.

Moreover, there appears to be a relationship between reverse turn denatured state conformational bias, and the role of these turns in biological function for the UBA domains. Juli Feigon et al. demonstrate hydrophobic surface patches for both HHR23A domains, which are considered to be binding sites for protein-protein interaction (e.g. ubiquitin)[84]. The degree of

amino acid sequence conservation in turn 1 in all UBA domains across multiple species suggests this turn sequence may be important not just to restraining the conformational distribution of these UBA domains in folding, but may play an important role in biological function. For example, in the UBA(2) domain, it was previously observed that the first reverse turn contains amino acid sequences that are involved in binding to the Vpr-1 protein in the human immunodeficiency virus type 1 (HIV-1)[77]. Furthermore, mutation of proline to glutamic acid at the end of this turn sequence abolishes binding to the Vpr-1 protein in HIV-1. Thus, it would be worth testing whether this proline plays a direct role in biasing the conformational distribution of UBA(2). Similar methods used in this dissertation could be applied to test this question. The amino acid sequences of UBA and CBP domains showing persistent turn regions are highlighted in red below (from N-terminus on left).

CBP

PNRSISPSALQDLLRTLKSPSSPQQQQQVLNILKSNPQLMAAFIKQRTAKYVANGMQ

UBA(1)

TLVTGSEYETMLTEIMSMGYERERVVAALRASYNNPHRAVEYLLTGIPG

UBA(2)

GSQEKEAIERLKALGFPESLVIQAYFACEKNENLAANFLLSQNFDDE

While the UBA domains show denatured state persistent structure at the sequences between helices, the intrinsically disordered protein CBP shows something unusual. The most persistent His-heme loops in CBP were discovered to correspond to sequences at unstructured region-helix boundaries located at the N and C-termini, although, the interpretation of this data is difficult to discern. Additionally, a region at the beginning of helix 3 containing possibly denatured state conformational bias (when bound to ACTR) was also detected. Given the strange

behavior with our pK_{loop} data with CBP, it is unclear currently what is happening with CBP in the denatured state using our His-heme method. On one hand, CBP appears to behave more like a folded protein when bound to iso-1-Cyt*c*, as our GuHCl unfolding data suggest most single histidine mutations enhance the stability of the CBP domain. How CBP can be well folded and exhibit such weak pK_{loop} values in most parts of the disordered chain is a conundrum. In this regard, it is showing opposite behavior of what foldable three-helix bundle sequences do to the Cyt*c* domain (i.e. single histidine mutations have little effect on UBA domain stability, but bind well to the heme in Cyt*c*, whereas CBP binds weakly to the heme in Cyt*c*, where the presence of single histidine mutations in the disordered sequence initiate structure formation).

Future directions – manipulation of turns, examining two-helix bundle turn biases, and mutations to test relationship between DSE conformational bias and biological function

We will extend the His-heme loop formation method to two-helix bundles, to determine whether a two-helix bundle's single turn sequence has similar thermodynamic and kinetic bias in denaturing conditions as observed for four-helix and three-helix bundles. Thus, His-heme loop formation studies in the denatured state of two-helix bundles with varying helical propensity will be carried out.

In addition, site-directed mutagenesis on turns which are persistent in the DSE for foldable three helix bundles (i.e. UBA domains) will be studied. This will include phi-value analysis on the persistent turn sequences within the UBA domains, where careful stepwise deconstruction of these turns will be studied using stopped-flow refolding/unfolding kinetic experiments in denaturing conditions. In addition, carrying out His-heme loop formation thermodynamic and kinetic experiments will supplement the data with phi-value analysis. These

experiments will provide information on how a single mutation will affect the behavior of a nucleation sequence (i.e. folding kinetics will increase or decrease relative to the wild type variant). This information will elucidate what sequence information is essential to increase the folding efficiency of these UBA domains. In practical terms, these types of studies would be beneficial to protein biotechnology and could help in therapeutic development for protein misfolding diseases.

Last but not least, we have subcloned the gene for ACTR (the putative binding partner for CBP) into the pRbs_BTR1fuse vector. Initial expression and purification of ACTR suggests this protein is heavily disordered compared to CBP, given the multiple attempts to express and purify the wild type ACTR. Mass spectrometry data shows cleavage at the N-terminus of ACTR during the purification process. For future work, ACTR will be expressed and purified using >20 mM Phenylmethylsulfonyl fluoride (PMSF) at each step of purification, to protect the protein from proteolysis near its N-terminus. We have had limited success expressing and purifying the ACTR-Cyt*c* fusion protein, but we were able to carry out a GuHCl unfolding experiment. As seen previously, ACTR fused to Cyt*c* shows 3-state equilibrium unfolding behavior (Fig. 5-1).

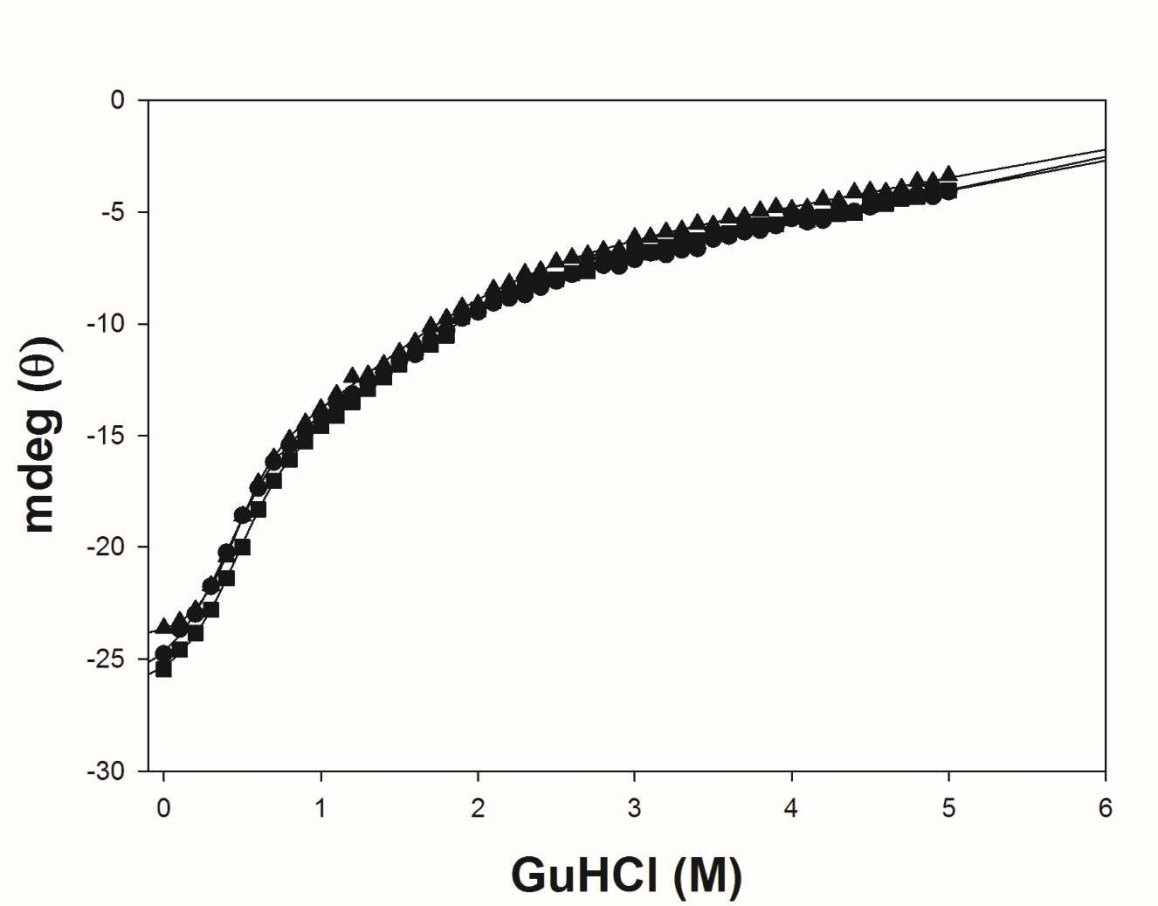

Fig. 5-1. GuHCl melting curve of ACTR-Cyt*c* fusion construct. As observed for other fusion proteins, Cyt*c* unfolds first followed by the ACTR domain.