The Pennsylvania State University

The Graduate School

Eberly College of Science

MODULATION OF THERMODYNAMIC AND KINETIC PROPERTIES IN *b* HEMOPROTEINS: CHARACTERIZATION OF A FAMILY OF PsaE-CYTOCHROME b_5 CHIMERAS AND *SYNECHOCYSTIS* sp. PCC 6803 HEMOGLOBIN

A Thesis in

Chemistry

by

Jane A. Knappenberger

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2006

The thesis of Jane A. Knappenberger was reviewed and approved* by the following:

Juliette T. J. Lecomte
Associate Professor of Chemistry
Thesis Advisor
Chair of Committee

Philip C. Bevilacqua
Associate Professor of Chemistry

Alan J. Benesi
Director of the NMR Facility
Lecturer in Chemistry

B. Tracy Nixon
Associate Professor of Biochemistry and Molecular Biology

Ayusman Sen
Professor of Chemistry
Head of the Department of Chemistry

*Signatures are on file in the Graduate School

ABSTRACT

Many *b* hemoproteins display significant increases in secondary and tertiary structure as well as free energy of unfolding upon cofactor binding. The extent to which the flexible portions of the apoprotein or the anchors connecting this region to the folded core control its thermodynamic and kinetic properties is unknown. In addition, the role played by interactions in the proximal heme pocket in controlling holoprotein stability and ligand-binding kinetics is not well understood. Although the magnitude of these interactions varies from case to case, important structure-function relationships for this class of proteins can be obtained by examining representative systems. To that end, the properties of two such proteins, rat hepatic cytochrome b_5 (cyt b_5) and the truncated hemoglobin (trHb) from the cyanobacterium *Synechocystis* sp PCC 6803 (S6803), have been investigated.

The electron-transport protein cytochrome b_5 (cyt b_5) contains two hydrophobic cores in the holo state. Core 1 participates in most of the heme-protein interactions and folds only in the presence of the prosthetic group; by comparison, core 2 is present in the apoprotein, makes relatively few heme contacts, and provides a scaffold that orients properly the binding region. To examine the autonomy of core 1 and the extent to which loop closure and the nature of the loop anchor affect folding properties, a class of chimeric proteins was constructed. These proteins, termed EbE1-4, comprised the heme-binding loop of cyt b_5 inserted in place of an existing loop within the folded scaffold of PsaE, a protein with an SH3-domain fold from a cyanobacterial photosystem I. The various EbE constructs differed only in the residues used to join the two proteins' primary structures. For those proteins that were water-soluble, 2-dimensional

homonuclear NMR data confirmed that the insertion did not disturb significantly PsaE's SH3 fold. Thermal and chemical denaturation experiments monitored by CD and absorption spectroscopies indicated that the thermodynamic stability of PsaE was not compromised greatly by more than tripling the size of the flexible loop. The residues chosen to join the flexible and structure regions influenced the sign of $\Delta\Delta G°_{unf}$, the folding and unfolding kinetics in the presence of urea, the solubility of the protein, and the number of native states populated by the protein. Although none of the proteins binds heme specifically, autonomous function of core 1 cannot be discounted, since strain in the tertiary structure caused by ordering both the cyt b_5 binding loop and the PsaE SH3 fold may prevent binding in this case. If the EbE results can be extrapolated to the cyt b_5 case, it appears that restraint of the large, unstructured binding loop may not destabilize greatly the apoprotein fold. This may be a general property of loops in folded, globular proteins. In addition, the results suggest that the interface between the inserted loop and its supporting scaffold can control many properties of the protein.

The cyanobacterium *Synechocystis* sp. PCC 6803 (S6803) expresses a two-on-two globin that uses His46 (distal side) and His70 (proximal) as axial ligands to the heme iron. His46 can be displaced by O_2, CO, and CN^-, among others, whereas His70 is not labile under native conditions. The residue preceding the proximal histidine has been implicated in the control of axial ligand reactivity in globins; the details of the mechanism, however, are not well understood, and little information exists for *bis*-histidyl hexacoordinate proteins. In many vertebrate hemoglobins and in the *Synechocystis* protein, the position is occupied by an alanine whereas, in myoglobins, it is a serine involved in an intricate hydrogen bond network. We examined the role of Ala69

in S6803 hemoglobin through the effects of the Ala → Ser replacement. The substitution resulted in minor structural perturbations, but the holoprotein's response to temperature-, urea-, and acid-induced denaturation was measurably affected. Enhanced three-state behavior was manifested in the decoupling of heme binding and secondary structure formation. Urea-gradient gel experiments revealed that the stability of the apoprotein was unchanged by the replacement, and that a slight alteration of the folding kinetics occurred in the holoproteins. Cyanide-binding experiments were performed to assess *trans* effects. The apparent rate constant for association decreased two-fold upon Ala69Ser replacement. This deceleration was attributed to a change in the lifetime of a state containing a decoordinated His46. The results demonstrated that, as in vertebrate globins and leghemoglobin, proximal influences operate to determine fundamental dynamic and thermodynamic properties of the protein.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Δδ	change in chemical shift in ppm
APS	ammonium persulfate
bp	base pair
CD	circular dichroism
DQF-COSY	double-quantum-filtered correlated spectroscopy
EbE1-4	chimeric proteins whose N- and C-terminal stretches are composed of residues from PsaE and whose intervening residues come from cytochrome b_5
EDTA	ethylenediaminetetraacetic acid
Fe-PPIX	iron(III) protoporphyrin IX
Fe-TCP	iron(III) tetracarboxyphenyl porphine
Hb	hemoglobin
HMQC	heteronuclear multiple quantum coherence
HSQC	heteronuclear single quantum coherence
IPTG	isopropylthio-β-D-galactoside
LB	Luria-Bertani medium
M9	minimal medium
Mb	myoglobin
metMb	ferric myoglobin
metMbCN	metcyano myoglobin
MRE	molar residual ellipticity
NOE	nuclear Overhauser effect
NOESY	two-dimensional nuclear Overhauser effect spectroscopy
OD_{600}	optical density at 600 nm
PAGE	polyacrylamide gel electrophoresis
PDB	protein data bank
PMSF	phenylmethylsulfonyl fluoride
PsaE	photosystem I accessory protein E
rHb	recombinant hemoglobin
rHb-A	recombinant hemoglobin with covalently linked heme
rHb-R	recombinant heme-reconstituted hemoglobin
S6803	*Synechocystis* sp. PCC 6803
S7002	*Synechococcus* sp. PCC 7002
SDS	sodium dodecylsulfate
SH3	Src homology domain 3
TEMED	N,N,N',N'-tetramethylethylenediamine
TOCSY	totally correlated spectroscopy
trHb	truncated hemoglobin
Tris	tris(hydroxymethyl)aminomethane

Preface

Information regarding multiple-author work:

Chapter 3

Knappenberger, J.A., Kraemer-Pecore, C.M., and Lecomte, J.T.J. Insertion of the cytochrome b_5 heme-binding loop into an SH3 domain. Effects on structure and stability, and clues about the cytochrome's architecture. (2004) *Prot. Sci.* 13: 2899-2908.

Division of labor:

Knappenberger, J.A.
- Cloning of EbE1 gene
- Protein purification
- Collection of spectroscopic data (NMR, CD, UV-Vis)
- Data analysis and interpretation

Kraemer-Pecore, C.M.
- Cloning of EbE1 gene

Chapter 5

Knappenberger, J.A., Kuriakose, S.A., Nothnagel, H.J., Vu, B.C., Vuletich, D.A., and Lecomte, J.T.J. Proximal influences in two-on-two globins: Effect of the Ala69Ser replacement on *Synechocystis* sp. PCC 6803 hemoglobin. *Biochemistry*. (submitted)

Division of labor:

Knappenberger, J.A.
- Apo- and holoprotein purification
- Collection of thermal and chemical denaturation data
- Analysis and interpretation of denaturation data
- Development and implementation of urea-gradient gel electrophoresis protocol
- Analysis and interpretation of urea-gradient gel data
- Interpretation of acid-denaturation data
- Interpretation of cyanide-binding kinetic data

Kuriakose, S.A.
- Holoprotein purification
- Collection of NMR data
- Collection of chemical denaturation data
- Collection of acid-denaturation data

Nothnagel, H.J.
- Holoprotein purification
- Collection of cyanide-binding kinetic data
- Analysis and interpretation of cyanide-binding kinetic data

Vu, B.C.

Plasmid preparation
Collection of NMR data
Collection and analysis of acid-denaturation data

Vuletich, D.A.

Collection of ^{15}N NMR data on wild-type and A69S proteins containing ^{15}N-labeled bound cyanide

Lecomte, J.T.J.

Analysis and interpretation of NMR data
Analysis and interpretation of cyanide-binding kinetic data

Acknowledgements

First and foremost, I would like to acknowledge my advisor, Juliette Lecomte. It has been a pleasure to work for and learn from her. Her insight, advise, guidance, and encouragement have been invaluable. My committee members, especially Phil Bevilacqua, have also been a valuable resource. Next, I would like to thank Chris Falzone; his wit and sense of humor have made the past five years manageable. I would also like to thank the present and past members of the Lecomte Lab, especially Kunal Mukhopadhyay, Nancy Scott, Christie Vu, and Dave Vuletich, for insightful research discussions as well as camaraderie. Members of the Bevilacqua lab, Ellen (Moody) Kehres and D.J. Proctor in particular, also provided helpful discussions. Finally, I must thank my family, for none of this would have been possible without their support. My husband, daughter, parents, and siblings were always there when I needed a break from science.

Funding for the work discussed in this thesis has been provided by National Science Foundation grants MCB-091182 and MCB-0349409 as well as National Institutes of Health grant GM-054217.

Chapter 1

Introduction

1.1. *b* Hemoproteins

b Hemoproteins comprise a class of proteins that bind iron protoporphyrin IX (Fe(III)-PPIX or *b* heme (Fe(II)), Figure 1.1) and perform a variety of functions. In these proteins, the protoporphyrin IX skeleton is intact. Contact between the protein and the cofactor involves hydrogen bonding, van der Waals and electrostatic interactions, and hydrophobic forces. In addition, these proteins form either one or two coordination bonds with the iron; a common participant in this ligation scheme is the Nε of a histidine side chain. The relative contribution of each type of interaction to the binding energy varies from case to case as does the affinity for the heme group.

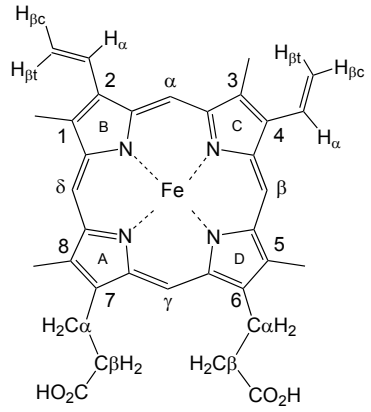


Figure 1.1. Structure of iron protoporphyrin IX (Fe-PPIX). The peripheral substituents are numbered according to the Fischer nomenclature.

Although all *b* hemoproteins bind the same cofactor, this does not translate into identical function. Proteins involved in O_2 transport or storage include hemoglobin (Hb)

(*1*), myoglobin (Mb), leghemoglobin (Lb) (*2, 3*), and possibly neuroglobin (Ngb) (*4, 5*). Cytochrome b_5 participates in electron transport and is unable to bind exogenous ligands (*6*). Cytochrome P450 functions as an oxygenase, and peroxidases such as glutathione peroxidase, cytochrome *c* peroxidase, and horseradish peroxidase, among others, provide yet another functional class. FixLH (*7*) and cytoglobin (Cygb) (*8, 9*) have been implicated in signaling and O_2 sensing. The wide array of chemical processes performed by these proteins indicates that the intricacies of the protein environment, rather than the heme group, control such properties as redox potential, ability to bind various ligands, and propensity for generating or removing reactive oxygen species.

This thesis focuses on the effects of structured and unstructured *b* hemoprotein components in controlling thermodynamic and kinetic properties of both the apo- and holoprotein states. In order to explore these issues, two representative members of this class of proteins were chosen: microsomal cytochrome b_5 and a cyanobacterial "truncated" hemoglobin.

1.2. Cytochrome b_5

Cytochrome b_5 is an electron-transport protein that is found in animals, plants, fungi, and some bacteria. Microsomal and mitochondrial isoforms have been characterized, and both are membrane bound. Additionally, a non-membranous form of the protein has been detected in erythrocytes (*10*). The work presented here involves microsomal cytochrome b_5, a 16-kDa protein that is tail-anchored to the endoplasmic reticulum. It couples with NADH and cytochrome b_5 reductase to supply electrons for processes such as fatty acid desaturation and cytochrome P450 monooxygenase reactions (*11*). In addition, it is thought to interact with cytochrome *c* in the apoptotic signaling pathway

(*6*). Most studies have focused on the ~11-kDa soluble portion of the protein, which was originally obtained via proteolytic cleavage of the full-length cytochrome.

The X-ray crystal structure of the soluble portion of the protein ((*12, 13*), Figure 1.2A) reveals a complicated α, β topology and indicates the presence of two hydrophobic cores, core 1 and core 2. Core 1 is involved in most of the heme-protein contacts. This portion of the protein supplies the two endogenous iron ligands (His39 and His63), which are strictly conserved across species. Extensive mutational analysis has been used to investigate the role of residues in the heme-binding region in controlling the stability and function of the holoprotein. However, the effects of this region on the properties of the apoprotein have not been examined as closely.

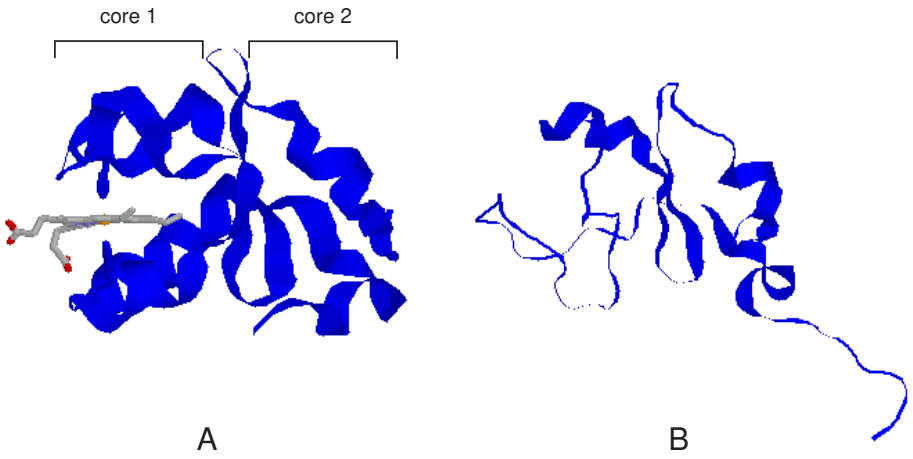


Figure 1.2. (A) X-ray crystal structure of the water-soluble domain of bovine microsomal holocytochrome b_5 (1cyo) showing endogenous completion of the hexacoordination scheme of the heme iron. The regions of the protein containing the two hydrophobic cores are labeled. (B) Solution structure of the water-soluble domain of rat microsomal apocytochrome b_5 (1i8c) demonstrating the loss of structure in the heme-binding domain in the absence of the cofactor.

The solution structure of apocytochrome b_5 ((*14*), Figure 1.2B) reveals that core 1 is devoid of extensive secondary and tertiary structure in the absence of cofactor binding. Rather, it comprises a disordered loop that samples multiple conformations. In contrast, core 2 does not need the heme group to fold (*15*). A previous study on a cytochrome b_5 deletion variant has shown that the core-2 fold can also be formed in the absence of the binding loop (*16*). This region is thought to play a mainly structural role, properly orienting the heme-binding region to accept the cofactor. However, the level of core 1 functional autonomy is unknown. Whether or not core 2 is required for heme binding, it could be necessary to regulate the dynamics or thermodynamic stability of core 1 in order to ensure physiological function. This type of interaction has been proposed for myoglobin in light of the large-scale conformational changes observed upon ligand binding by minimyoglobin, a proteolytic fragment of the full-length protein (*17*).

1.3. Loop effects

Apocytochrome b_5 can be thought of as a small, globular scaffold supporting a large, flexible loop. The ordering of loop termini is accompanied by an entropic penalty, and several theoretical models are available to calculate the magnitude of this effect (*18, 19*). These results differ depending on the model used and the nature of the loop itself. Although many systems are well described by theory, several are not. One striking example of this dichotomy is that of the tenth fibronectin type III domain of human fibronectin (*20*). Depending on the location within the protein, a four-glycine insertion either destabilizes the protein or has no effect. This indicates that the consequence of a particular loop cannot yet be predicted with certainty, and an assessment of the effect of

the unstructured apocytochrome b_5 heme-binding region on its scaffold requires an experimental investigation.

The free energy associated with folding of the core 2 region of the molecule is small, only ~ 6 kJ/mol. By comparison, other proteins of similar size are characterized by higher thermodynamic stabilities. It was previously unknown whether the low stability of the apoprotein was an intrinsic property or a result of supporting the disordered heme-binding region. In an attempt to discover if this particular loop necessarily destabilized its scaffold, it was used to replace another loop in a folded protein of similar size and architecture to the folded core of the apocytochrome.

1.4. Intrinsically unstructured proteins

For years, it was thought that, because a protein's function was dictated by its structure, the population of a well-folded native state was a prerequisite for performing a cellular task. Much effort was put into structure prediction as a step toward function prediction. Recently, however, it has become apparent that a large number of proteins carry out their physiological roles from an unstructured state (*21*). These proteins often participate in protein-protein or protein-nucleic acid complexes that are involved in transcriptional or translational regulation or cell signaling pathways (*21-23*). The extent of disorder within these proteins varies from case to case. Some are always unstructured; others fold upon encountering a binding partner. Still others contain well-defined domains in which the unfolded region appears as a large, flexible loop. It has been hypothesized that this final subset of proteins signifies a method of evolution of additional functional domains within a preexisting protein (*21*). Thus, in addition to any thermodynamic consequences it may

have, the presence of the disordered heme-binding region of apocytochrome b_5 is of general interest in light of this class of proteins.

Among proteins that contain both ordered and disordered segments, it is unknown how critical the evolution of the residues at the interface of these two regions is in maintaining proper function of the preexisting domain. It is also unknown how the loop affects the characteristics of the protein as a whole. Although these properties may vary from protein to protein and from loop to loop, monitoring the biophysical traits of a protein as a function of loop nature can provide insight into the possible outcome of inserting a long stretch of unstructured amino acids. Several proteins were created in order to examine the effect of minor modifications within the cytochrome b_5 heme-binding loop on the thermodynamic and kinetic properties of the protein into which it was inserted. The first step was to identify a suitable scaffold.

1.5. PsaE

Photosystem I is membrane-bound complex involved in electron transport that is found in photosynthetic organisms such as cyanobacteria. It contains many components including chlorophyll, β-carotene, phyloquinones, [4Fe-4S] clusters, and polypeptides (*24, 25*). The single photosystem I accessory protein E (PsaE) molecule found in the complex contacts other proteins within photosystem I including PsaA, PsaB, and PsaC (*26*). In addition, PsaE has also been implicated in controlling cyclic electron transport indirectly via its effects on the association of ferrodoxin with other photosystem I components (*27*).

The solution structure of PsaE ((*28*), Figure 1.3) from the cyanobacterium *Synechococcus* sp. PCC 7002 (S7002) portrays an SH3-domain fold containing a five-

stranded β sheet and one turn of a 3_{10} helix. In addition, the protein supports a 14-residue loop between the C and D β strands. The structures of PsaE proteins from other organisms have since become available (*29, 30*), and the SH3-like fold has been established for each of them. One difference among these proteins exists at the level of the size of the CD loop. This indicates that a variable loop length can be accommodated by the PsaE scaffold. Both the size of the folded region and the orientation of the loop termini approximate those found in apocytochrome b_5. Furthermore, the PsaE fold is maintained in the presence of physiologically relevant interactions with the other photosystem I proteins. Although the extent of loop reorganization is not as great as that observed for the cytochrome b_5 core 1 region upon heme binding (*31*), the fold preservation of the scaffold represents another similarity between these two proteins. In light of these facts, the SH3 core of S7002 PsaE was chosen as an alternate scaffold in which to examine the properties of the cytochrome b_5 heme-binding region.

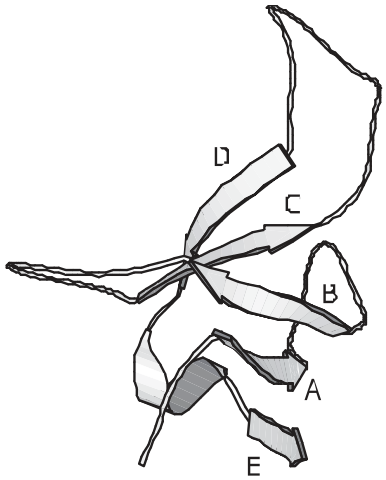


Figure 1.3. Solution structure of *Synochococcus* sp. PCC 7002 photosystem I accessory protein E (PsaE, 1psf). The protein is characterized by the SH3 fold containing a 5-stranded β sheet and one turn of a 3_{10} helix. The strands are labeled A-E.

1.6. The EbE family

A family of four chimeric proteins containing the heme-binding region of the cytochrome inserted into the SH3 fold of PsaE was created to investigate various loop effects. The proteins were named EbE1-4 to indicate that the N and C termini originated from PsaE and the intervening portion contained residues from cytochrome b_5. The four members of the family differed only in the amino acids chosen to join the primary structures of the two parent proteins. In the EbE1 and EbE4 cases, the residues at the interface came directly from the PsaE scaffold and cytochrome b_5 loop. By comparison, EbE2 and EbE3 contained glycine insertions at the N or N and C termini of the loop, respectively. In the creation of these proteins, the size of the disordered loop supported by the PsaE scaffold increased by a minimum of 29 residues.

Chapter 3 describes the side-by-side comparison of EbE1 and PsaE. The effects of loop replacement on the docking of the β strands within the SH3-like fold were explored. In addition, the relative resistance to temperature- and urea-induced denaturation was determined and compared to that calculated using the available theoretical models. Finally, the ability of the cytochrome b_5 heme-binding region to bind heme in a non-native scaffold was examined via a series of attempted reconstitutions with both the physiologically relevant cofactor (*b* heme) and a water-soluble derivative.

Chapter 4 describes work done on the EbE2, EbE3, and EbE4 proteins and explores the consequences of the choice of loop anchor on various thermodynamic and kinetic properties. NMR experiments were performed to examine the side chain and backbone chemical environments, and comparisons to EbE1 and PsaE were made. When possible, the resistance to temperature- and urea-induced denaturation was also determined. In

addition, EbE4 was subjected to an extensive set of kinetic folding and unfolding experiments to inspect the folding pathway of the protein. The EbE family of proteins allowed for an investigation of the effects of loop elongation as well as alterations at the interface between unstructured and structured portions on the properties of the apo state of one *b* hemoprotein; another representative member of this class of proteins was chosen to probe the role played by structured elements in determining holoprotein characteristics.

1.7. Truncated hemoglobins

Hemoglobins form a well-studied class of proteins that are present across all kingdoms. The holoprotein is characterized by the classical globin fold ((*32*), Figure 1.4), which contains eight helices, labeled A-H, arranged in either a three-on-three or two-on-two helical grouping around the heme (*33*). Under most conditions, the protein supplies a

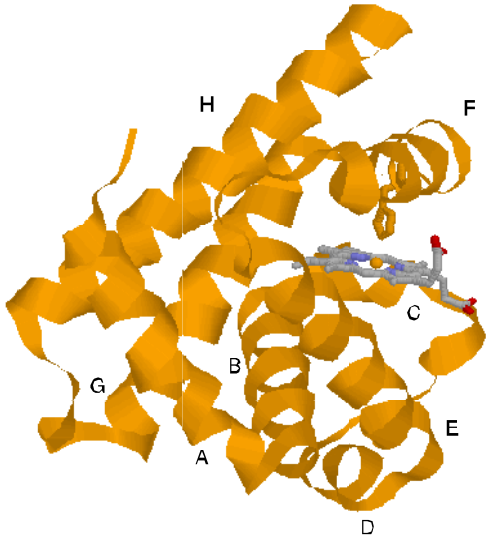


Figure 1.4. Crystal structure of sperm whale myoglobin (4mbn) showing the pentacoordination scheme of the heme iron. The helices are labeled A-H according to the notation of Perutz.

single histidine ligand to the heme iron. Recently, a number of globins have been discovered in a variety of organisms including humans, cyanobacteria, and higher plants (*34-36*). Some of these newly-uncovered proteins contain 20-30 residues fewer than the classical vertebrate specimens and have been termed truncated hemoglobins (trHbs) (*37, 38*). The abbreviated primary structure leads to shortening of one helix and removal of another, resulting in a two-on-two rather than a three-on-three fold. Sequence alignments of the trHbs demonstrate the division of these proteins into three groups, termed I, II, and III or N, O, and P (*38, 39*). These proteins are proposed to carry out a variety of functions and, like their full-length relatives, are capable of binding small exogenous ligands to the heme iron. In the absence of ligand binding, some trHbs retain the classical five-coordination scheme of the heme iron in which four ligands are supplied by the nitrogens within the porphyrin ring and the fifth is the Nε atom of the strictly conserved histidine in the F helix (this position is labeled F7 in the Perutz notation). By comparison, other trHbs, including those from the cyanobacteria *Synechococcus* sp. PCC 7002 and *Synechocystis* sp. PCC 6803, as well as other globins such as cytoglobin and neuroglobin are able to complete endogenously the octahedral coordination scheme of the iron.

1.8. *Synechocystis* sp. PCC 6803 rHb-R

The solution structure of the truncated hemoglobin from the non-nitrogen-fixing cyanobacterium *Synechocystis* sp. PCC 6803 (S6803) is shown in Figure 1.5A (*40*); it depicts *bis*-histidine coordination of the heme iron and displays the typical two-on-two trHb fold. The holoprotein has been termed rHb-R to signify that it is a recombinant heme-reconstituted hemoglobin. One unusual aspect of this protein is its ability to

modify covalently the prosthetic group using the side chain of a non-ligating histidine residue ((*41*), (*42*), Figure 1.5B). Neither the mechanism nor the physiological relevance of this cross link has been determined. In addition, the cellular function of the protein is

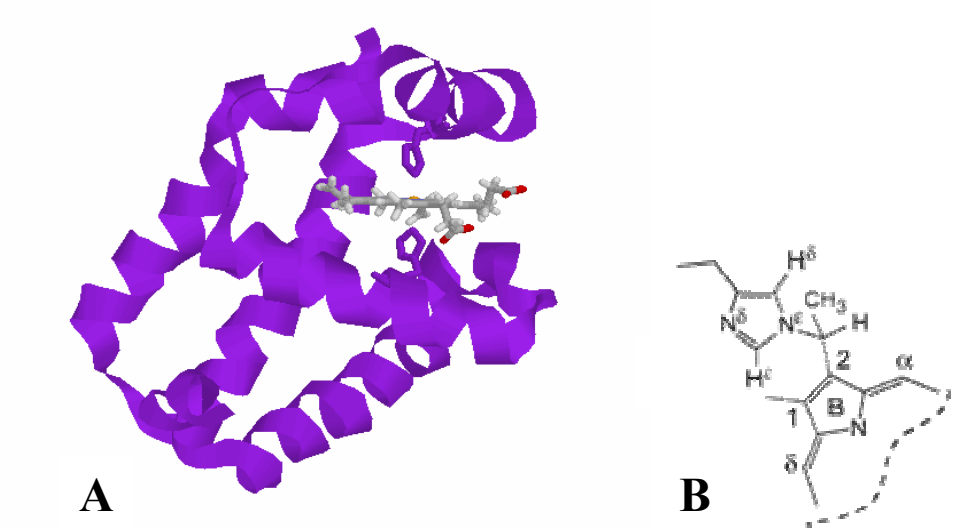


Figure 1.5. (A) Solution structure of the truncated hemoglobin from the cyanobacterium *Synechocystis* sp. PCC 6803 (1mwb) showing endogenous hexacoordination of the heme iron via the Nε atoms of His 46 and His70. (B) Covalent linkage formed between the Nε of His177 and the heme 2-vinyl group under certain reducing conditions.

unknown, and *in vivo* expression of the *Synechococcus* rHb-R, which also forms the heme-protein cross link, was only recently confirmed (Scott, N.L., personal communication). Kinetic and thermodynamic analyses of S6803 rHb-R (*43-46*) have led to a determination of the affinities of the protein for the heme group and various small ligands as well as the stability of the holoprotein fold in both the hemichrome (*bis*-histidine coordination of the ferric iron) and ligand-bound states. Further examination of this protein and comparison to other, more well-known systems could provide clues about its cellular role and the ability of specific structural components to affect holoprotein properties.

1.9. The role of S6803 rHb-R residue 69

Within the holoprotein heme cavity of globins, two regions have been extensively characterized: the proximal and distal pockets. In wild-type proteins, exogenous ligand binding occurs in the distal pocket either directly or by replacement of a protein side chain. The roles of specific distal residues in controlling the rate and affinity with which small ligands bind have been explored at length. Although proximal residues have not been studied as thoroughly, it is known that the hydrophobic packing and hydrogen-bonding interactions in which these side chains participate contribute to coordination bond lengths, positioning of the heme relative to the protein, and heme affinity (*47, 48*).

In most myoglobins, the residue immediately preceding the proximal histidine (His70) is a serine, which forms a side-chain side-chain hydrogen bond with the ligating histidine (*49*). This residue is capable of modulating the relative orientations of the histidine and the heme planes, altering ligand affinity in a modest *trans* effect, controlling heme affinity, and facilitating solvent access in the heme pocket (*48, 50*). By comparison, the amino acid at the analogous position in leghemoglobin and most vertebrate hemoglobins is not capable of participating in hydrogen-bonding interactions. S6803 rHb-R, along with other group I trHbs (*39*), contains an alanine immediately before the proximal histidine (*51*).

Chapter 6 examines the role of this residue in controlling proximal pocket structure, holoprotein stability, heme affinity, ligand-binding properties, and folding kinetics. NMR experiments were performed to compare the chemical environment of the proximal heme pocket of a Ala69Ser variant of S6803 rHb-R to that of the wild-type protein and to establish if the serine side chain had formed the hydrogen bond found in myoglobin

systems. In addition, the wild-type and A69S proteins were subjected to temperature-, urea-, and acid-induced denaturation to determine the magnitude of any change in the specific heme affinity. Urea-gradient gel electrophoresis (*52*) was used to determine the effect of the replacement on the stability of the apoprotein and the folding kinetics of the holoprotein. Finally, the kinetics of cyanide binding to the two proteins was examined to inspect whether or not a *trans* effect was exerted by this residue.

1.10. Summary

Many factors contribute to the function of hemoproteins. Although extreme examples of total disorder (cytochrome *c*, (*53*)) and complete structure formation (HasA, (*54*)) in the absence of the heme group exist, most members of this class of proteins display moderate amounts of preorganization. Heme binding therefore depends critically on the ability of the protein to fold into and to remain in the appropriate conformation to be able to accept the cofactor; the thermodynamic and kinetic properties of the apoprotein are significant in this regard. In addition, remaining in the holo state requires both protein folding and greater specific than non-specific heme affinity; residues in close proximity to the heme group often exert control over holoprotein stability.

In examining both apo- and holoproteins, it is unclear exactly which interactions within the protein matrix control biophysical properties such as thermodynamic stability, folding and unfolding kinetics, heme affinity, kinetics of exogenous-ligand binding, and relative stabilities of the bound and hemichrome states. In light of this, we have chosen two representative systems, rat hepatic cytochrome b_5 and the truncated hemoglobin from the cyanobacterium *Synechocystis* sp. PCC 6803, to examine the effects of both unstructured and structured protein components in determining these characteristics.

The heme-binding region of cytochrome b_5 was used to investigate two aspects of the protein: the consequences of loop closure and the nature of the loop anchor on the kinetic and thermodynamic properties of the apoprotein and the ability of the core-1 region to bind heme in the absence of core 2. This work contributes to the understanding of loop-scaffold interactions and their effects on protein folding and stability. In addition, it adds to the limited available knowledge base regarding the size requirements of functional *b* hemoproteins.

Only a small fraction of the genetically-encoded putative members of the trHb class of proteins have been structurally characterized. The exact functions and expression levels of each of these proteins have yet to be clarified. Investigation into the structural, thermodynamic, and kinetic properties of S6803 rHb-R and comparison to well-characterized systems such as hemoglobins, myoglobins, and peroxidases could aid in the formulation of a possible function for this protein. An Ala69Ser variant of S6803 rHb-R was used probe whether or not the presence of a potential hydrogen-bonding partner for the proximal histidine would result in myoglobin-like structural properties or change significantly the stability and ligand-binding properties of the holoprotein.

References

(1) Perutz, M. F. (1976) Structure and mechanism of haemoglobin. *Br Med Bull 32*, 195-208.

(2) Wittenberg, J. B. (1974) Facilitated oxygen diffusion. The role of leghemoglobin in nitrogen fixation by bacteroids isolated from soybean root nodules. *J Biol Chem 249*, 4057-66.

(3) Wittenberg, J. B., Appleby, C. A., Bergersen, F. J., and Turner, G. L. (1975) Leghemoglobin: the role of hemoglobin in the nitrogen-fixing legume root nodule. *Ann N Y Acad Sci 244*, 28-34.

(4) Ostojic, J., Sakaguchi, D. S., de Lathouder, Y., Hargrove, M. S., Trent, J. T., 3rd, Kwon, Y. H., Kardon, R. H., Kuehn, M. H., Betts, D. M., and Grozdanic, S. (2006) Neuroglobin and cytoglobin: oxygen-binding proteins in retinal neurons. *Invest Ophthalmol Vis Sci 47*, 1016-23.

(5) Sun, Y., Jin, K., Peel, A., Mao, X. O., Xie, L., and Greenberg, D. A. (2003) Neuroglobin protects the brain from experimental stroke in vivo. *Proc Natl Acad Sci U S A 100*, 3497-500.

(6) Schenkman, J. B., and Jansson, I. (2003) The many roles of cytochrome b5. *Pharmacol Ther 97*, 139-52.

(7) Perutz, M. F., Paoli, M., and Lesk, A. M. (1999) Fix L, a haemoglobin that acts as an oxygen sensor: signalling mechanism and structural basis of its homology with PAS domains. *Chem Biol 6*, R291-7.

(8) Fordel, E., Geuens, E., Dewilde, S., De Coen, W., and Moens, L. (2004) Hypoxia/ischemia and the regulation of neuroglobin and cytoglobin expression. *IUBMB Life 56*, 681-7.

(9) Hankeln, T., Ebner, B., Fuchs, C., Gerlach, F., Haberkamp, M., Laufs, T. L., Roesner, A., Schmidt, M., Weich, B., Wystub, S., Saaler-Reinhardt, S., Reuss, S., Bolognesi, M., De Sanctis, D., Marden, M. C., Kiger, L., Moens, L., Dewilde, S., Nevo, E., Avivi, A., Weber, R. E., Fago, A., and Burmester, T. (2005) Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. *J Inorg Biochem 99*, 110-9.

(10) Lederer, F. (1994) The cytochrome b5-fold: an adaptable module. *Biochimie 76*, 674-92.

(11) Porter, T. D. (2002) The roles of cytochrome b5 in cytochrome P450 reactions. *J Biochem Mol Toxicol 16*, 311-6.

(12) Durley, R. C., and Mathews, F. S. (1996) Refinement and structural analysis of bovine cytochrome b5 at 1.5 A resolution. *Acta Crystallogr D Biol Crystallogr 52*, 65-76.

(13) Mathews, F. S., and Strittmatter, P. (1969) Crystallographic study of calf liver cytochrome b5. *J Mol Biol 41*, 295-7.

(14) Falzone, C. J., Mayer, M. R., Whiteman, E. L., Moore, C. D., and Lecomte, J. T. (1996) Design challenges for hemoproteins: the solution structure of apocytochrome b5. *Biochemistry 35*, 6519-26.

(15) Moore, C. D., and Lecomte, J. T. (1993) Characterization of an independent structural unit in apocytochrome b5. *Biochemistry 32*, 199-207.

(16) Constans, A. J., Mayer, M. R., Sukits, S. F., and Lecomte, J. T. (1998) A test of the relationship between sequence and structure in proteins: excision of the heme binding site in apocytochrome b5. *Protein Sci 7*, 1983-93.

(17) Di Iorio, E. E., Yu, W., Calonder, C., Winterhalter, K. H., De Sanctis, G., Falcioni, G., Ascoli, F., Giardina, B., and Brunori, M. (1993) Protein dynamics in minimyoglobin: is the central core of myoglobin the conformational domain? *Proc Natl Acad Sci U S A 90*, 2025-9.

(18) Brant, D. A., and Flory, P. J. (1965) The configuration of random polypeptide chains. II. Theory. *J. Am. Chem. Soc. 87*, 2791-2800.

(19) Chan, H. S., and Dill, K. A. (1989) Intrachain loops in polymers: effects of excluded volume. *J. Chem. Phys. 90*, 492-509.

(20) Batori, V., Koide, A., and Koide, S. (2002) Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain. *Protein Eng 15*, 1015-20.

(21) Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol 6*, 197-208.

(22) Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol 12*, 54-60.

(23) Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol 293*, 321-31.

(24) Chitnis, P. R., and Nelson, N. (1991), in Cell culture and somatic cell genetics of plants (Bogorad, L. and Vasil, I.K. Eds) Vol. 7B, pp 177-224, Academic Press, New York.

(25) Golbeck, J. H. (2003) The binding of cofactors to photosystem I analyzed by spectroscopic and mutagenic methods. *Annu. Rev. Biophys. Biomol. Struct. 32*, 237-56.

(26) Bottcher, B., Graber, P., and Boekema, E. J. (1992) The structure of Photosystem I from the thermophilic cyanobacterium Synechococcus sp. determined by electron microscopy of two-dimensional crystals. *Biochim Biophys Acta 1100*, 125-36.

(27) Rousseau, F., Setif, P., and Lagoutte, B. (1993) Evidence for the involvement of PSI-E subunit in the reduction of ferredoxin by photosystem I. *EMBO J 12*, 1755-65.

(28) Falzone, C. J., Kao, Y. H., Zhao, J., Bryant, D. A., and Lecomte, J. T. (1994) Three-dimensional solution structure of PsaE from the cyanobacterium Synechococcus sp. strain PCC 7002, a photosystem I protein that shows structural homology with SH3 domains. *Biochemistry 33*, 6052-62.

(29) Barth, P., Savarin, P., Gilquin, B., Lagoutte, B., and Ochsenbein, F. (2002) Solution NMR structure and backbone dynamics of the PsaE subunit of photosystem I from Synechocystis sp. PCC 6803. *Biochemistry 41*, 13902-14.

(30) Mayer, K. L., Shen, G., Bryant, D. A., Lecomte, J. T., and Falzone, C. J. (1999) The solution structure of photosystem I accessory protein E from the cyanobacterium Nostoc sp. strain PCC 8009. *Biochemistry 38*, 13736-46.

(31) Falzone, C. J., Wang, Y., Vu, B. C., Scott, N. L., Bhattacharya, S., and Lecomte, J. T. (2001) Structural and dynamic perturbations induced by heme binding in cytochrome b5. *Biochemistry 40*, 4879-91.

(32) Perutz, M. F. (1964) The Structure Of Myoglobin And Haemoglobin. *Dan Tidsskr Farm 38*, 113-22.

(33) Vinogradov, S. N., Hoogewijs, D., Bailly, X., Arrendondo-Peter, R., Guertin, M., Gough, J., Dewilde, S., Moens, L., and Vanfleteren, J. R. (2005) Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. *Proc Natl Acad Sci U S A 102*, 11385-89.

(34) Choudhary, M. L., Jawaid, S., Ahuja, M. K., Shiva, N. K., Gupta, P., Bhuyan, A. K., and Khatri, G. S. (2005) Open reading frame yjbI of Bacillus subtilis codes for truncated hemoglobin. *Protein Expr Purif 41*, 363-72.

(35) Roesner, A., Fuchs, C., Hankeln, T., and Burmester, T. (2005) A globin gene of ancient evolutionary origin in lower vertebrates: evidence for two distinct globin families in animals. *Mol Biol Evol 22*, 12-20.

(36) Scott, N. L., Falzone, C. J., Vuletich, D. A., Zhao, J., Bryant, D. A., and Lecomte, J. T. (2002) Truncated hemoglobin from the cyanobacterium Synechococcus sp. PCC 7002: evidence for hexacoordination and covalent adduct formation in the ferric recombinant protein. *Biochemistry 41*, 6902-10.

(37) Pesce, A., Couture, M., Dewilde, S., Guertin, M., Yamauchi, K., Ascenzi, P., Moens, L., and Bolognesi, M. (2000) A novel two-over-two alpha-helical sandwich fold is characteristic of the truncated hemoglobin family. *Embo J 19*, 2424-34.

(38) Wittenberg, J. B., Bolognesi, M., Wittenberg, B. A., and Guertin, M. (2002) Truncated hemoglobins: a new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants. *J Biol Chem 277*, 871-4.

(39) Vuletich, D. A., and Lecomte, J. T. (2006) A phylogenetic and structural analysis of truncated hemoglobins. *J Mol Evol 62*, 196-210.

(40) Falzone, C. J., Christie Vu, B., Scott, N. L., and Lecomte, J. T. (2002) The solution structure of the recombinant hemoglobin from the cyanobacterium Synechocystis sp. PCC 6803 in its hemichrome state. *J Mol Biol 324*, 1015-29.

(41) Vu, B. C., Jones, A. D., and Lecomte, J. T. (2002) Novel histidine-heme covalent linkage in a hemoglobin. *J Am Chem Soc 124*, 8544-5.

(42) Hoy, J. A., Kundu, S., Trent, J. T., 3rd, Ramaswamy, S., and Hargrove, M. S. (2004) The crystal structure of Synechocystis hemoglobin with a covalent heme linkage. *J Biol Chem 279*, 16535-42.

(43) Hvitved, A. N., Trent, J. T., 3rd, Premer, S. A., and Hargrove, M. S. (2001) Ligand binding and hexacoordination in synechocystis hemoglobin. *J Biol Chem 276*, 34714-21.

(44) Kundu, S., Premer, S. A., Hoy, J. A., Trent, J. T., 3rd, and Hargrove, M. S. (2003) Direct measurement of equilibrium constants for high-affinity hemoglobins. *Biophys J 84*, 3931-40.

(45) Smagghe, B. J., Sarath, G., Ross, E., Hilbert, J. L., and Hargrove, M. S. (2006) Slow ligand binding kinetics dominate ferrous hexacoordinate hemoglobin reactivities and reveal differences between plants and other species. *Biochemistry 45*, 561-70.

(46) Vu, B. C., Nothnagel, H. J., Vuletich, D. A., Falzone, C. J., and Lecomte, J. T. (2004) Cyanide binding to hexacoordinate cyanobacterial hemoglobins: hydrogen-bonding network and heme pocket rearrangement in ferric H117A Synechocystis hemoglobin. *Biochemistry 43*, 12622-33.

(47) Knapp, J. E., Bonham, M. A., Gibson, Q. H., Nichols, J. C., and Royer, W. E., Jr. (2005) Residue F4 plays a key role in modulating oxygen affinity and cooperativity in Scapharca dimeric hemoglobin. *Biochemistry 44*, 14419-30.

(48) Smerdon, S. J., Krzywda, S., Wilkinson, A. J., Brantley, R. E., Jr., Carver, T. E., Hargrove, M. S., and Olson, J. S. (1993) Serine92 (F7) contributes to the control of heme reactivity and stability in myoglobin. *Biochemistry 32*, 5132-8.

(49) Cheng, X. D., and Schoenborn, B. P. (1991) Neutron diffraction study of carbonmonoxymyoglobin. *J Mol Biol 220*, 381-99.

(50) Shiro, Y., Iizuka, T., Marubayashi, K., Ogura, T., Kitagawa, T., Balasubramanian, S., and Boxer, S. G. (1994) Spectroscopic study of Ser92 mutants of human myoglobin: hydrogen bonding effect of Ser92 to proximal His93 on structure and property of myoglobin. *Biochemistry 33*, 14986-92.

(51) Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res. 3*, 109-136.

(52) Creighton, T. E., and Shortle, D. (1994) Electrophoretic characterization of the denatured states of staphylococcal nuclease. *J Mol Biol 242*, 670-82.

(53) Hamada, D., Hoshino, M., Kataoka, M., Fink, A. L., and Goto, Y. (1993) Intermediate conformational states of apocytochrome c. *Biochemistry 32*, 10351-8.

(54) Debarbieux, L., and Wandersman, C. (2001) Folded HasA inhibits its own secretion through its ABC exporter. *Embo J 20*, 4657-63.

Chapter 2

Materials and Methods

2.1. EbE Gene Construction and Cloning

The pET3d vector containing the gene for PsaE from the cyanobacterium *Synechococcus* sp. strain PCC 7002 (*psaE*, a gift of Dr. D. A. Bryant) was cut with BamHI and NcoI, and the ends were not dephosphorylated. The vector was purified on a 1% SeaPlaque gel. The gene for EbE1 (PsaE 1-41, cytochrome b_5 33-75, PsaE 56-69) was constructed from overlapping oligonucleotides with 12 bp overhangs. DNA oligonucleotides ranging in length from 43 to 57 nucleotides were ordered from Operon Technologies. The oligos were purified on a denaturing polyacrylamide gel. The DNA was eluted out of the crushed gel with a high-salt, TE buffer and recovered via EtOH precipitation. Next, the oligos were phosphorylated with polynucleotide kinase. The oligos were then annealed by heating the mixture to 90 °C and slowly cooling to room temperature, and the nicks were sealed using T4 DNA ligase. The completed insert was purified on a 3% NuSieve, 1% SeaKem gel, run onto a DEAE membrane, eluted with TE buffer, and recovered by EtOH precipitation. The vector and insert were ligated with T4 DNA ligase, and the resulting plasmid was transformed into competent *E. coli* DH5α cells. The genes for EbE2 (PsaE 1-41, Gly, cyt b_5 33-75, PsaE 56-69) and EbE3 (PsaE 1-41, Gly, cyt b_5 33-75, Gly, PsaE 56-69) were constructed with the Stratagene QuikChange protocol and primers containing additional Gly codons and the EbE1 gene as the template. The gene for EbE4 (PsaE 1-40, cyt b_5 29-77, PsaE 59-69) was created with the same technique as that used for EbE1. All plasmids were sequenced by the Penn

State Nucleic Acid Facility. Primary structures for all primers, genes, and proteins can be found in Appendix A.

2.2. Protein Overexpression and Purification

2.2.1. EbE1-3

The plasmids harboring the genes for EbE1-3 were used to transform competent *E. coli* BL21(DE3) cells, which were grown in LB medium at 37 °C. Expression of the protein was induced with IPTG when the OD_{600} was ~0.8. The cells were harvested 4.5 h post-induction and subjected to three rounds of lysing by sonication in the presence of PMSF. The majority of the protein was found to be in the soluble portion of the cell. The protein was eluted from a DEAE Sephacel anion exchange column with a 50 mM Tris, 1 mM EDTA buffer at pH 7.5 and a 0 to 0.5 M NaCl gradient. The fractions containing the protein were then concentrated and purified on a Sephadex G50-Fine column using a buffer of 50 mM Tris, 1 mM EDTA, and 20 mM NaCl at pH 7.5. Fractions that contained >95% pure protein (as evidenced by a single band on a denaturing polyacrylamide gel) were exchanged into 20 mM phosphate buffer, pH 7.5. Although the EbE proteins have masses ~11 kDa (as confirmed by electrospray mass spectrometry), all run anomalously, appearing to weigh ~19 kDa by SDS-PAGE.

2.2.2. Cytochrome b_5

The pET3d plasmid harboring the gene for the soluble portion of rat microsomal cytochrome b_5 (part of *cytb*$_5$ (*1*), a gift of Dr. S. G. Sligar) was used to transform competent *E. coli* BL21(DE3) cells, which were grown in M9 medium at 37 °C. Expression of the protein was induced with IPTG when the OD_{600} was ~0.8. The cells were harvested 4 to 6 h post-induction and subjected to three rounds of lysing by

sonication in the presence of PMSF. The majority of the protein was found in the soluble portion of the cell. The protein was eluted from a DEAE Sephacel anion exchange column with a 50 mM Tris, 1 mM EDTA buffer at pH 7.5 and a gradient of 0 to 0.5 M NaCl. The fractions containing the apoprotein were then concentrated and purified on a Sephadex G50-Fine column using a buffer of 50 mM Tris and 1 mM EDTA at pH 7.5. Fractions that contained >95% pure protein (as evidenced by a single band on a denaturing polyacrylamide gel) were exchanged into 20 mM phosphate buffer, pH 7.5.

2.2.3. EbE4, PsaE

E. coli transformation and cell growth conditions were identical to those used for EbE1-3, with the exception that PsaE growths were harvested 5-6 h post-induction. Cells were subjected to three rounds of sonication. Both EbE4 and PsaE were found to be isolated in inclusion bodies, which were solubilized with 8 M urea. The protein was then purified and refolded on a Sephadex G50-Fine sizing column. Pure fractions of PsaE were exchanged into 20 mM phosphate buffer. EbE4 fractions were purified further on a DEAE anion exchange column and then exchanged into 20 mM phosphate buffer on a Sephadex G25 sizing column. Uniformly ^{15}N-labeled EbE4 was obtained via an additional growth in M9 medium containing ^{15}N ammonium chloride. The cells were harvested 14 h after the onset of IPTG-induced protein production, and the labeled protein was purified as detailed above.

2.2.4. wild-type and A69S S6803 rHb

E. coli BL21(DE3) cells were transformed with the pET3c plasmid harboring the gene for the wild-type (*glbN*, a gift of Dr. D. A. Bryant) or variant globin (obtained by performing the Stratagene QuikChange protocol on the wild-type gene) and grown in M9

medium. Protein expression was induced with IPTG when the OD_{600} was ~0.8; cells were harvested 5 to 6 h later and stored at −20 °C until use. Protein retrieval was possible after two rounds of sonication were used to disrupt cell membranes. The majority of the protein was found in insoluble inclusion bodies, which were solubilized with 8 M urea. The protein was then purified and renatured on a Sephadex G50 sizing column and eluted in pH 8.0 buffer containing 50 mM Tris and 1 mM EDTA. Overnight incubation with a three-fold excess of hemin chloride yielded holoproteins, which were purified further on a DEAE anion exchange column containing a pH 8.0 buffer containing 50 mM Tris, 1 mM EDTA, and a 0 to 0.5 M NaCl gradient. Fractions that were to remain in the apo state were simply purified on the anion exchange column without heme addition. Pure proteins (as determined by a single band on an SDS-PAGE gel) were exchanged into water, lyophilized, and stored at -20 °C until use.

2.3. Nuclear Magnetic Resonance Spectroscopy

NMR data were collected on a Bruker DRX-600 spectrometer (14.1 T, operating either at a ^{1}H frequency of 600.05 MHz or 600.18 MHz). Protein concentration ranged from 1-2 mM. Samples contained a minimum of 10% $^{2}H_2O$ for locking purposes, and were buffered with 20 mM phosphate at pH 7-7.5 (without isotope correction). ^{1}H chemical shifts were referenced to DSS through the water line with correction for pH and temperature (*2*). Unless otherwise noted, all data were collected at 25 °C. For 1D data, 16 k complex data points were collected over 12 kHz with presaturation of the water line as needed. A phase-shifted squared-sinusoidal window function was applied to the data before transforming. Homonuclear NOESY (*3, 4*), DQF-COSY (*5*), and TOCSY (*6, 7*) data were collected for EbE1-4. Typically, 2 k data points were collected in the direct

dimension (512 points in the indirect dimension) over an 8 kHz spectral width in each dimension. TQF-COSY (*8, 9*) and 2Q (*8, 9*) data were collected on the EbE4 sample. ^{1}H-^{15}N HSQC (*10*), ^{15}N-separated TOCSY, ^{15}N-separated NOESY, and ^{1}H-^{15}N HMQC (*10*, *11*) data were collected for a uniformly ^{15}N-labeled EbE4 sample. For the ^{1}H-^{15}N HMQC experiment, a 15-ms delay was used to account for 6- to12-Hz two-bond ^{1}H-^{15}N couplings ($^{2}J_{NH}$) in the histidine imidazole ring (*12*). A homonuclear ^{15}N-decoupled NOESY data set was also acquired on EbE4. In addition, the magnitude of the {^{1}H}-^{15}N NOE (*13*) was determined from the ratio of peak volumes in two HSQC-type experiments, one of which was collected with a 3-s amide proton saturation pulse during the recycle delay by use of the GARP-1 sequence (*14*). For the two-dimensional data sets, typically 2 K data points were collected over 8 kHz, and the spectra were centered at the water line (4.76 ppm) in the proton dimension and 122 ppm in the nitrogen dimension (if applicable). For three-dimensional experiments, data were acquired over a spectral width of 7 and 3.5 kHz in the proton dimensions and 2 kHz in the nitrogen dimension. 100-ms mixing times were used for NOESY experiments; TOSCY data were collected with a 45-ms mixing time with DIPSI-2 pulse train (10 kHz field) (*15*). Water suppression was accomplished using either presaturation of the water line or the WATERGATE (*16*) pulse sequence; quadrature detection was obtained via either the States (*17*) or TPPI (*18, 19*) method. ^{15}N frequencies were referenced indirectly to the proton frequency. Most data were collected at 25 °C, however, some variable temperature (5 to 45 °C) experiments were performed. The probe temperature was calibrated using the known temperature dependence of the difference in ^{1}H chemical shifts of neat solutions of methanol (low temperatures) or ethylene glycol (high

temperatures) (*20*). NMR data were processed using NMRpipe (*21*) or XWIN-NMR (Bruker BioSpin) and analyzed using XWIN-NMR and Sparky (*22*). Processing parameters and more detailed experimental explanations have been published previously (*23, 24*).

2.4. Thermal Denaturation

The resistance of each protein to temperature-induced unfolding was determined by monitoring the CD and absorption spectra of the protein in the range from 25 to 95 °C. Data were collected each 2 °C after a 5-min equilibration time. The solutions (5 to 10 μM holoprotein or 50 to 100 μM apoprotein, 20 mM phosphate buffer, pH 7.2-7.5) were then cooled to examine reversibility of the unfolding reaction.

Data were collected on an Aviv 14DS spectrophotometer and either an Aviv 62DS or a Jasco J-810 spectropolarimeter. All instruments were equipped with a Peltier device for temperature control. CD data (300 to 200 nm) yielded global information regarding the conformation of the protein backbone, and absorbance data informed on the local environment of either the aromatic residues (apoproteins; 320 to 260 nm) or the heme group (holoproteins; 650 – 350 nm). Results from both methods were compared to determine if the proteins displayed behavior consistent with a two-state unfolding mechanism.

Thermal denaturation data were fit with the program NFIT (University of Texas, Galveston) or SAVUKA (*25*) to a Gibbs-Helmholtz equation for a two-state process (*26*):

$$\Delta G°(T) = \Delta H°_{Tm}(1 - T/T_m) - \Delta C°p[(T_m - T) + T \ln(T/T_m)]$$

where $\Delta H°_{Tm}$ represents the enthalpy of denaturation at the midpoint temperature (T_m). $\Delta C°p$ values obtained in this manner have limited accuracy. Throughout the experiment,

the observed signal at any temperature T (Y(T)) is a function of the fractional population of the native ($F_N(T)$) and unfolded ($F_U(T)$) states and the spectral properties of these states ($Y_N(T)$ and $Y_U(T)$):

$$Y(T) = F_N(T)Y_N(T) + F_U(T)Y_U(T)$$

The fractional population of the native state is given by:

$$F_N(T) = 1/(1 + K_U(T))$$

where $K_U(T)$ is the equilibrium constant for unfolding. This constant is related to $\Delta G°(T)$ via $-RT\ln K_U(T)$. Although not all proteins displayed coincidence of unfolding profiles monitored by CD and absorbance spectroscopy, each individual trace was well fit using this two-state model.

2.5. Chemical Denaturation

In order to determine the free energy of unfolding, proteins were subjected to urea-induced denaturation. Ultra-pure urea was purified further on a mixed-bed resin column before use. Proteins were exposed to urea concentrations ranging from 0 to 9.6 M in steps of 0.2 M; the experiments were controlled by an automatic titrator. To obtain each of these concentrations, native (20 mM phosphate buffer, pH 7.2 to 7.4) and denatured (9.5 to 10 M urea, 20 mM phosphate buffer, pH 7.2 to 7.4) protein solutions were mixed. The protein concentration typically ranged from 5 to 40 μM and was identical in the native and denatured solutions for each experiment. Urea concentrations of the final and denatured stock solutions were determined by measuring the index of refraction and using the following equation (*26, 27)*:

$$[\text{urea}] = 117.66*\Delta n_D + 29.753*(\Delta n_D)^2 + 185.56*(\Delta n_D)^3$$

The pH of the initial and final solutions was also determined to ensure that it had not drifted throughout the course of the titration. Reversibility of the urea-induced unfolding reaction and the time necessary to reestablish equilibrium were determined by diluting into buffer a concentrated protein solution containing 8 to 9 M urea and monitoring the absorbance spectrum of the protein as a function of time.

Data were collected at 25 °C on an Aviv 14DS spectrophotometer, and Aviv 62DS spectropolarimeter, or a Jasco J-810 spectropolarimeter. Absorbance data were collected from 320 nm to 260 nm in steps of 1 nm for apoproteins and 650 nm to 350 nm in steps of 2 nm for holoproteins. CD data were collected from 270 nm to 200 nm in steps of 1 nm to monitor the degree of secondary structure present throughout the titration. On some samples, kinetic CD data were also collected at 221 nm for 400 s. Averaging times for both experiments ranged from 2 to 5 s. Equilibration time at each urea concentration ranged from 5 to 20 min, depending on the protein.

Data were fit with the program Savuka (*25*). The free energy of unfolding was determined according to the equation

$$\Delta G^\circ_{H2O} = \Delta G^\circ_D + m[D]$$

where ΔG°_{H2O} is the free energy of unfolding in water, ΔG°_D is the free energy of unfolding at a denaturant concentration of D, and m is the dependence of the free energy of unfolding on the denaturant concentration. ΔG°_D depends on the equilibrium constant (or fractional population of the native and denatured states) at each urea concentration and is determined based on the spectral properties of these states as a function of denaturant.

2.6. Urea-gradient Gel Electrophoresis

As another method of determining resistance to urea-induced denaturation, proteins were subjected to urea-gradient gel electrophoresis (*28*). In general, when a protein experiences high enough urea concentrations, the equilibrium between the folded and unfolded states of the proteins shifts to favor the denatured state. This transition is often monitored by absorbance spectroscopy, and, therefore, the ability to observe each of the states populated along the reaction coordinate depends on the extinction coefficient of each species. For example, a three-state unfolding reaction that involves two states with identical (or very similar) epsilon values may appear to be a two-state process. One of the advantages of gradient gels is the ability to monitor electrophoretic mobility under increasingly denaturing conditions. As the protein unfolds, the hydrodynamic volume of the protein increases, thus slowing the migration of the molecule through the gel matrix. In this manner, two states of different size or shape can be differentiated, even if they have the same extinction coefficient. In addition, low-resolution structural information can be obtained by comparing the hydrodynamic volume of a new protein to that of a well-studied protein whose three-dimensional structure has been determined previously.

A two-chamber gradient maker and a peristaltic pump set to 1 mL/min were used to create a linear gradient of urea ranging from 0 to 8 M. In order to account for the higher viscosity of the medium as a function of increasing urea concentration, a reverse gradient of acrylamide (11 to 7 %) was also introduced. Gels were cast perpendicular to the direction of electrophoresis. Of the 10-cm width of the gel (as oriented for electrophoresis), the central 6 cm were composed of the linear gradient and 2 cm on either side were used as 0-M and 8-M fringe regions as appropriate. In order to prevent

the gels from polymerizing while being poured and thereby disrupting the gradient, a light-induced polymerization method was employed in which methylene blue was used as the free-radical generator and sodium toluene sulfinate and diphenyliodonium chloride functioned as the oxidizing and reducing agents, respectively. Gels were cast in the dark then exposed to light for ~30 min. All gels were stored at 4 °C for 12-24 hours prior to use.

In order to ensure that any change in observed electrophoretic mobility arose from protein unfolding rather than gel non-uniformity, two control experiments were performed. As the polymerization reaction proceeds, the methylene blue chromophore is destroyed and the gel undergoes a transition from blue to colorless. In order to verify the linearity of the gradient, a dye that is not involved in the polymerization (in this case bromophenol blue, 0.005 %) was added to the 8 M urea, 7 % acrylamide solution. The polymerized gel had identically positioned color and urea gradients, and a linear transition from colorless to blue indicated that the appropriate urea concentrations had been established across the gel. Once the linearity of the urea gradient was established, the viscosity of the gel matrix was examined. An appropriate reverse acrylamide gradient should counteract the increasing urea concentration and maintain a constant viscosity across the gel. To test this, three different molecules were loaded onto the gel: xylene cyanol FF, bromophenol blue, and a protein that populates the folded state throughout the 0 to 8 M urea range (a form of the truncated hemoglobin from the cyanobacterium *Synechococcus* sp. PCC 7002 in which the heme group has been covalently attached to the protein; the midpoint of the urea-induced denaturation is >9.5 M). Each of these

molecules produced lines with zero slope throughout the region of interest, indicating that a satisfactory acrylamide gradient had been established.

After 30 – 60 min of preelectrophoresis to remove any urea contaminates, 75 – 100 μL samples were loaded across the top surface of the gel. Native samples contained 10 % glycerol, 0.08 % bromophenol blue, 0.08 % xylene cyanol FF, and ≥ 50 μg of protein; denatured samples also contained 8 M urea. The denatured samples were used to verify reversibility of the unfolding reaction. A continuous buffer system (50 mM Tris-acetate, pH 7.5) was used throughout the gel and the tank. Gels were subjected to ~4 hr of electrophoresis (100 V, 4 °C) before being examined. Protein was visualized with Coomassie brilliant blue; gels on which holoprotein samples were loaded were also subjected to heme staining.

2.7. Kinetic Folding and Unfolding Experiments

Manual mixing experiments were performed to investigate the slow processes involved in EbE4 folding and unfolding. In the absence of urea, the time constants for these folding events were 73 s and 57 s, respectively. 234-nm CD data were collected for 600-900 s with a 2-s averaging time at 25 °C using a Jasco J-810 spectropolarimeter. In general, 1:10 or 1:20 dilutions of 100-200 μM protein were achieved by adding the solution to 1.8-1.9 mL of dilutant that was stirring in a 1-cm cell, resulting in a 10-μM protein solution. The dead time of the experiment ranged from 4 to 14 s, and an additional 5 s of data were discarded to account for instrument response. A variety of initial and final urea concentrations were employed to determine the folding mechanism. Refolding data were collected in two manners: constant initial urea concentration (9.12 M) and varying final urea concentrations (0.5 to 3.0 M) or varying initial urea

concentrations (4.6 to 7.9 M) and constant final urea concentration (0.7 M). Unfolding data were collected similarly (0 M → 8.25 to 10.14 M or 2.0 to 7.5 M → 9.1 M). In addition, double-jump kinetics experiments were also performed. Native protein (20 mM phosphate buffer, 0 M urea, pH 7.4, 25 C) was exposed to buffered ~8.9 M urea for times varying from 10 to 900 s and then diluted with buffer, yielding a final urea concentration of ~0.35 M. Data collection was initiated upon final dilution into buffer. Savuka was used to fit the data to a model comprised of a constant and a sum of exponentials according to the equation

$$A(t) = A(\infty) + \sum A_i * \exp(-t/\tau_i)$$

where $A(\infty)$ is the constant corresponding to the final ellipticity, A_i is the pre-exponential amplitude for function i, and τ_i is the corresponding time constant. In all cases, only a single exponential function was required to fit the data.

NOTE: The following methods were applied by others (see preface on page xix of this thesis for explanation of multi-author work).

2.8. Heme Release upon Solution Acidification

Release of the heme group from the protein matrix was monitored by absorbance spectroscopy. Hydrochloric acid was used to adjust the pH of a 10-μM holoprotein solution (5 mM phosphate buffer) from 7.4 to 2.6 in steps of 0.1 to 0.3 pH units. In order to minimize volume loss, the pH of the solution was determined without removing the solution from the cuvette. Spectra were collected between 260 nm and 700 nm at each pH. Refolding of the protein was inspected by returning the solution to neutrality with a

molar equivalent of potassium hydroxide at the end of the titration. The absorbance data at 410 nm were fit with Nfit (University of Texas, Galveston) to a modified Henderson-Hasselbalch equation accounting for two distinct transitions:

$$Y = Y_0 + (Y_1 - Y_0)\frac{10^{n_1(pK_1 - pH)}}{1 + 10^{n_1(pK_1 - pH)}} + (Y_2 - Y_1)\frac{10^{n_2(pK_2 - pH)}}{1 + 10^{n_2(pK_2 - pH)}}$$

where Y_0 is the absorbance of the protein at neutral pH, Y_i is the absorbance value of protein having gone through the *i*th protonation transition, and n_i is the Hill coefficient for that transition.

2.9. Kinetics of Cyanide Binding

Cyanide binding was monitored by visible spectroscopy. Data were collected with protein concentrations of ~10 μM and a 200-, 500- or 1000-fold excess of cyanide. The absorbance at 423 nm (wavelength of maximum difference between free and bound states) was monitored throughout the reaction at 30-s intervals up to a total time affording an absorbance estimate for the bound state. This manual mixing method had a dead-time of 30 s at the shortest. Data were fit to a sum of exponential functions with Nfit. An approximate pseudo-first-order rate constant was obtained for the phase that accounted for the highest fraction of the population, typically > 75%. The values at the three cyanide concentrations were used to obtain an apparent bimolecular rate constant as in previous work on the wild-type and H117A rHb-Rs (*29*). All measurements were performed in triplicate at pH 7.2 (100 mM phosphate) and were repeated in D_2O (pH 7.2, uncorrected for isotope effect). A triplicate data set was also collected on wild-type S6803 rHb-R with a 200-fold excess of cyanide at pH 8.2. Owing to the multiple exponentials required to fit the data, the apparent rate constants were used only for a semi-qualitative comparison.

References

1. Beck von Bodman, S., Schuler, M.A., Jollie, D.R., and Sligar, S.G. (1986) Synthesis, bacterial expression, and mutagenesis of the gene coding for mammalian cytochrome b5. *Proc. Natl. Acad. Sci. USA 83*, 9443-9447.

2. Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., and Sykes, B.D. (1995) ^{1}H, ^{13}C and ^{15}N Chemical Shift Referencing in Biomolecular NMR *J. Biomol. NMR 6,* 135-140.

3. Bodenhausen, G., Kogler, H., and Ernst, R.R., (1984) Selection of Coherence Transfer Pathways in NMR Pulse Experiments. *J. Magn. Reson. 58*, 370-388.

4. Kumar, A., Ernst, R.R., and Wuthrich, K (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun. 95*, 1-6.

5. Rance, M., and Byrd, R.A. (1983) Obtaining high fidelity spin-1/2 powder spectra in anisotropic media: phase-cycled Hahn echo spectroscopy. *J. Magn. Reson. 52*, 221-240.

6. Braunschweiler, L., and Ernst, R.R. (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson. 53*, 521-528.

7. Rance, M. (1987) Improved techniques for homonuclear rotating-frame and isotropic mixing experiments. *J. Magn. Reson. 74*, 557-564.

8. Braunschweiler, L., Bodenhausen, G., and Ernst, R.R. (1983) Analysis of networks of coupled spins by multiple quantum NMR. *Mol. Phys. 48*, 535-560.

9. Rance, M., and Wright, P.E. (1986) Analysis of ^{1}H NMR spectra of proteins using multiple quantum coherence *J. Magn. Reson. 66*, 372-378.

10. Bodenhausen, G., and Ruben, D.G. (1980) Natural abundancy nitrogen-15 NMR by enhanced heteronuclear. spectroscopy. *Chem. Phys. Lett. 69*, 185-189.

11. Mueller, L. (1979) Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence *J. Am. Chem. Soc. 101*, 4481-4484.

12. Pelton, J.G., Torchia, D.A., Meadow, N.D., and Roseman, S. (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques *Protein Sci. 2*, 543 - 558.

13. Farrow, N.A., Muhandriam, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N relaxation, *Biochemistry 33*, 5984-6003.

14. Shaka, A.J., Barker, P.B., and Freeman, R. (1988) Computer-optimized decoupling scheme for wideband applications and low-level operation *J. Magn. Reson. 64*, 547-552.

15. Shaka, A.J., Lee, C.J., and Pines, A. (1988) Iterative Schemes for Bilinear Operators: Application to Spin Decoupling. *J. Magn. Reson. 77*, 274-293

16. Sklenar, V., Piotto, M., Lepik, R., and Saudek, V. (1992) Gradient-tailored water suppression for ^{1}H-^{15}N HSQC experiments optimized to retain full sensitivity *J. Magn. Reson. 102*, 241-245.

17. States, D.J., Haberkorn, R.A., and Ruben, D.J. (1982) A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants *J. Magn. Reson. 48*, 286-292.

18. Drobny, G., Pines, A., Sinton, S., Weitekamp, D.P., and Wemmer, D. (1979) Fourier transform multiple quantum nuclear magnetic resonance. *Symp. Faraday Soc. 13*, 49-55.

19. Cavanagh, J., Fairbrother, W.J., Palmer, A.G.I., and Skelton, N.J. (1996) *Protein NMR Spectroscopy. Principles and Practice*, Academic Press, San Diego, USA.

20. Marion, D., and Wuthrich, K. (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins *Biochem. Biophys. Res. Commun. 113*, 967-974.

21. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995) nmrPipe - a Multidimensional Spectral Processing System Based On Unix Pipes *J. Biomol. NMR 6*, 277-293.

22. Goddard, T.D., and Kneller, D.G. (2006) SPARKY3, *University of California, San Francisco*.

23. Bhattacharya, S., Falzone, C.J., and Lecomte, J.T.J. (1999) Backbone dynamics of apocytochrome b_5 in its native, partially folded state. *Biochemistry 38*, 2577-2589.

24. Falzone, C.J., Kao, Y.-H., Zhao, J., MacLaughlin, K.L., Bryant, D.A., and Lecomte, J.T.J. (1994) ^{1}H and ^{15}N NMR assignments of PsaE, a Photosystem I subunit from the cyanobacterium *Synechococcus* sp. Strain PCC 7002. *Biochemistry 33*, 6043-6051.

25. Bilsel, O., Zitzewitz, J.A., Bowers, K.E., and Matthews, C.R. (1999) Folding mechanism of the alpha-subunit of tryptophan synthase, an alpha/beta barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry 38*, 1018-1029.

26. Pace, C.N., Shirley, B.A., and Thomson, J.A. (1989) *Protein structure: a practical approach*, IRL Press, Oxford; New York.

27. Pace, C.N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol. 131*, 266-280.

28. Creighton, T.E., and Shortle, D. (1994) Electrophoretic characterization of the denatured states of staphylococcal nuclease. *J. Mol. Biol. 242*, 670-682.

29. Vu, B.C., Nothnagel, H.J., Vuletich, D.A., Falzone, C.J., and Lecomte, J.T.J. (2004) Cyanide binding to hexacoordinate cyanobacterial hemoglobins: hydrogen-bonding network and heme pocket rearrangement in ferric H117A Synechocystis hemoglobin *Biochemistry 43*, 12622-12633.

Chapter 3

Insertion of the Cytochrome b_5 Heme-binding Loop into an SH3 Domain. Effects on Structure and Stability, and Clues about the Cytochrome's Architecture

Portions of the following have been previously published and were reproduced with permission from *Protein Science*. Knappenberger, J.A., Kraemer-Pecore, C.M., and Lecomte, J.T.J. (2004) *Prot. Sci.* 13: 2899-2908.

3.1. Introduction

In the fashion of many ligand-binding proteins, the water-soluble domain of rat hepatic cytochrome b_5 undergoes induced refolding upon association with its heme cofactor. In the holoprotein state, the three-dimensional structure is well represented by a narrow ensemble of conformations. As illustrated in Figure 3.1, the fold contains six helices and five strands (*1*). The heme is fastened in its cavity by coordination bonds ligating the Fe ion to the imidazole Nε atoms of His 39 (end of α2) and His 63 (beginning of α4). Two hydrophobic regions, termed core 1 and core 2, are also apparent in the cytochrome structure. The residues of core 1 are located in the central part of the amino acid sequence; they establish most of the protein-heme contacts and therefore play a functional role. In contrast, the residues comprising core 2 are found within the N- and C-terminal segments. Core 2 does not require the heme group to fold correctly (*2*) and is thought to serve a structural purpose. In the apoprotein, core 1 can be viewed as a large, flexible loop inserted into a stable scaffold (*3*). In previous studies aimed at analyzing the structural determinants of apocytochrome b_5, an abridged version of the protein was prepared that contained core 2 and a short linker in place of the heme-binding segment (*4*). The main tertiary features of core 2 appeared to be specified by the shortened sequence, in support of a certain degree of structural independence of the two regions. In

the present work, we focus on the heme-binding segment and its role in the apocytochrome's architecture.

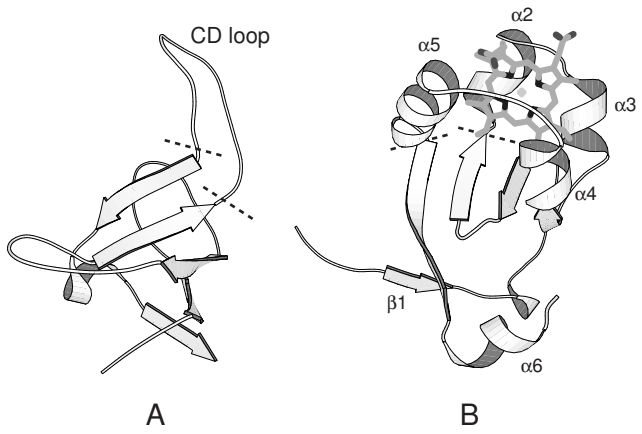


Figure 3.1. Ribbon structures of (A) *Synechococcus* sp. PCC 7002 photosystem I accessory protein E (S7002 PsaE) (1psf) and (B) the water-soluble domain of bovine microsomal cytochrome b_5 (1cyo). Lines indicate the points at which the loops in the parent proteins begin and end. The chimeric protein EbE1 contains the folded scaffold of PsaE (portion below lines) and the heme-binding region of cytochrome b_5 (portion above lines).

Because a compromise between conformational flexibility and stability often needs to be achieved for a protein's proper function, it is of interest to establish the effect of local disorder on the thermodynamic properties of naturally occurring and designed sequences containing unstructured loops. The thermal and chemical denaturations of apocytochrome b_5 from various sources are cooperative, two-state processes (*4-8*). Specifically, the apparent standard Gibbs free energy of denaturation for rat microsomal apocytochrome b_5 is only ~6 kJ/mol at room temperature and neutral pH, such that a small fraction (~10%) of the molecules sample the unfolded state under those conditions. By comparison, many examples of water-soluble proteins the size of core 2 exist that are

more stable than apocytochrome b_5. It is not known whether the low stability is intrinsic to the folded core or due to the unstructured segment.

A theoretical framework is available to predict the entropic effect of disordered loops on fold stability. The contribution is associated with constraining the ends of a chain and is evaluated at

$$\Delta S = -k\,R\,\ln(l) + c$$

where k is a scaling factor equal to 3/2 in the three-dimensional freely-jointed case, l is the linking number or number of peptide bonds between the beginning and the end of the loop, c is a constant related to bond length and distance of approach, and R is the gas constant (*9*). Several experimental studies have indeed shown that introduction of flexible segments in certain regions of proteins can lead to a marked thermodynamic destabilization of the fold. This is the case with the addition of just four Gly residues into the EF loop of the tenth fibronectin type III domain of human fibronectin (FNfn10) (*10*) and that of four Gly or four Thr residues into two long βα loops in yeast phosphoglycerate kinase (yPGK) (*11*). Replacement of a flexible region in bovine α-lactalbumin with the D helix from equine lysozyme results in a chimera of increased stability with respect to the wild-type protein (*12*) in an illustration of the reverse process, that is, the conversion of a disordered element into a rigid one. These specific instances suggest a quantifiable destabilization associated with restraining the ends of an inserted or elongated loop.

In contrast to the above findings, a number of cases exist for which the effects are significantly smaller than predicted. In the fibronectin example, 4-Gly insertions can also be accommodated without consequence at different locations in the protein (*10*).

Likewise, yPGK tolerates 4-Gly and 4-Thr insertions in one of its short loops (*11*). Up to 13 additional residues in a loop of a truncated version of chymotrypsin inhibitor-2 (CI2) affect stability and folding rates only weakly (*13*). Oftentimes, larger loops also fail to comply with the Jacobson-Stockmayer Gaussian model. Insertions of 60-80 random amino acids into an SH2 domain do not markedly affect the protein's thermodynamic stability or ability to bind a phosphotyrosyl peptide (*14*). In a similar manner, some *E. coli* RNase HI mutants retain enzymatic activity despite the insertion of 120-130 random, unstructured residues (*15*). These combined results indicate that the consequence of loop elongation is dependent upon the nature and the site of the insertion. In light of this, it is not yet possible to predict what effect a particular loop will have at a particular position.

In cytochrome b_5, the ends of the segment encompassing the disordered region rejoin after a span of 43 residues. According to equation 1 and with k = 2.41 and c = 4.28 (*9*) derived from ribonuclease T1 experimental data (*16*) rather than the freely-jointed chain model, the entropic penalty for end joining is expected to be on the order of 80 J/mol/K, which corresponds to a Gibbs free energy of 24 kJ/mol at room temperature. This large destabilization is comparable to the measured difference between apo- and holoprotein denaturation free energies (*6*). To explore whether a scaffold is necessarily destabilized to this extent by supporting the cytochrome b_5 core 1 region, this portion of the sequence was inserted in place of an existing 14-residue loop in *Synechococcus* sp. PCC 7002 (S7002) PsaE. The PsaE scaffold was chosen for its small size and SH3 domain structure (*17*). The target insertion points are in the turn connecting the C and D strands. Naturally occurring PsaEs display variability in this region, which suggests that the site is tolerant of alterations and can be used for engineering purposes. In addition, the α-

spectrin and p85α subunit of phosphatidylinositol 3 kinase SH3 domain proteins have previously been subjected to loop-swapping, circular permutation, and loop elongation (*18, 19*).

Using the same Eq. 1 parameters as above to anticipate the entropic penalty of loop closure, the replacement of 14 residues by 43 residues should lead to a decrease in stability (ΔΔG°) of ~6.4 kJ/mol. The thermodynamic stability of SH3 domains is in the range of 12-21 kJ/mol (*18, 20, 21*), with PsaE at the lower limit (*22*). As a result, if the cytochrome b_5 loop destabilizes the PsaE scaffold by the calculated ΔΔG°, a noticeable effect should be detected in the equilibrium populations of molecules sampling the native and denatured conformations under native conditions.

In addition to determining the thermodynamic consequences of the heme-binding region on the apoprotein, it was also of interest to determine the whether or not core 1 was capable of forming specific heme-protein contacts in the background of a non-native scaffold. It has recently been proposed that one method of evolving new, functional domains is the insertion of long disordered segments of amino acids into existing proteins (*23*). It is not understood whether or not function of the new domain requires evolution of both the loop and the supporting scaffold or just the insert or how autonomous these functional domains are. The protein did not bind specifically either *b* heme or a water-soluble derivative. This did not rule out autonomous function of the heme-binding region *per se*, but it did indicate that the heme-protein interactions in the parent cytochrome had been modified in the chimera.

3.2. Results

The primary structures of the wild-type proteins used in this work (S7002 PsaE and the soluble domain of rat hepatic cytochrome b_5) and the hybrid protein are shown in Figure 3.2. The portion of cytochrome b_5 to be inserted into PsaE and the insertion points were chosen by inspection of the three-dimensional structures (Fig 3.1). The resulting

```
Cyt b5 ---------A EQSDKDVKYY TLEEIQKHKD SKSTWVILHH
PsaE   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF
EbE1   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF

Cyt b5 KVYDLTKFLE EHPGGEEVLR EQAGGDATEN FEDVGHSTDA
PsaE   N---KVNYNG FSGSAGGL-- ---------- ----------
EbE1   N---LTKFLE EHPGGEEVLR EQAGGDATEN FEDVGHSTDA

Cyt b5 RELSKTYIIG ELHPDDRSKI AKPSETL
PsaE   ---------N TNNFAEHELE VVG----
EbE1   RELSKTY--N TNNFAEHELE VVG----
```

Figure 3.2. Primary structures of rat hepatic cytochrome b_5 (cyt b5), S7002 PsaE (PsaE), and the resulting chimeric protein, EbE1. The heme-binding region of cytochrome b_5 and the three-dimensional scaffold provided by PsaE into which the loop was inserted are shown in red and blue letters, respectively. EbE1 is so named because its N and C termini are comprised of residues from PsaE, the intervening residues come from cytochrome b_5, and it is the first such chimera. To facilitate the comparison of EbE1 to PsaE and cytochrome b_5, the original amino acid numbering was conserved in both portions of the hybrid. The letter E was appended to the residues originating from PsaE and the letter b to those from the cytochrome.

polypeptide was called EbE1 to reflect that the N and C termini of the protein are comprised of residues from PsaE, the intervening portion is taken from cytochrome b_5, and it is the first such chimera constructed. The gene encoding the 98-residue EbE1 was prepared using the overlapping oligonucleotides method (*24*). The purity of the protein was determined by SDS-PAGE to be greater than 95%. Electrospray ionization mass

spectrometry confirmed the high purity and returned an observed mass of 10,929.6 ± 0.3 Da, consistent with the expected theoretical mass of 10,930.0 Da.

3.2.1. Structural Characterization of EbE1

The structural properties of EbE1 were studied by optical methods. The far-UV CD spectra of PsaE and the hybrid protein are shown in Figure 3.3. Unlike PsaE, EbE1 did not exhibit positive ellipticity values between 220 nm and 228 nm. However, inspection of the spectra for the folded and unfolded states of both proteins (Figure 3.3) revealed a comparable change in ellipticity over this wavelength range upon unfolding. Although not typical of β-sheet proteins, the native spectra in Figure 3.3 are characteristic of SH3 domains (*21, 25-30*). In the absence of α-helical structure, CD bands in the region of 225 nm have been attributed to aromatic side chains (*31*). In SH3 domains, the positive ellipticity in the region of 220 nm stems from interactions among clustered aromatic residues (*25, 27*). PsaE and EbE1 contain Tyr16-Trp17-Tyr18, Phe 40, and Phe 60 (Figure 3.2) and have the potential for distinct clustering.

NMR spectroscopy was used to characterize the hybrid further. Figure 3.4 presents the ^{1}H spectra of PsaE and EbE1, which indicated that the hybrid was folded under normal conditions of pH and temperature. Well-resolved features that were nearly identical in the PsaE and EbE1 spectra included the two upfield peaks and a cluster of resonances downfield from 9.4 ppm. Examination of correlation data confirmed that one γ-CH_3 of Val38E and the δ-CH_3 of Ile10E gave rise to the upfield peaks, in agreement with the PsaE spectrum (*17*). The downfield-shifted peaks corresponded to the backbone NHs of Lys9E, Ala25E, Asp28E, Phe60E, Ala61E, and Glu62E, as well as the Trp17E indole NεH. The latter peak was shifted upfield by 0.48 ppm compared to its position in

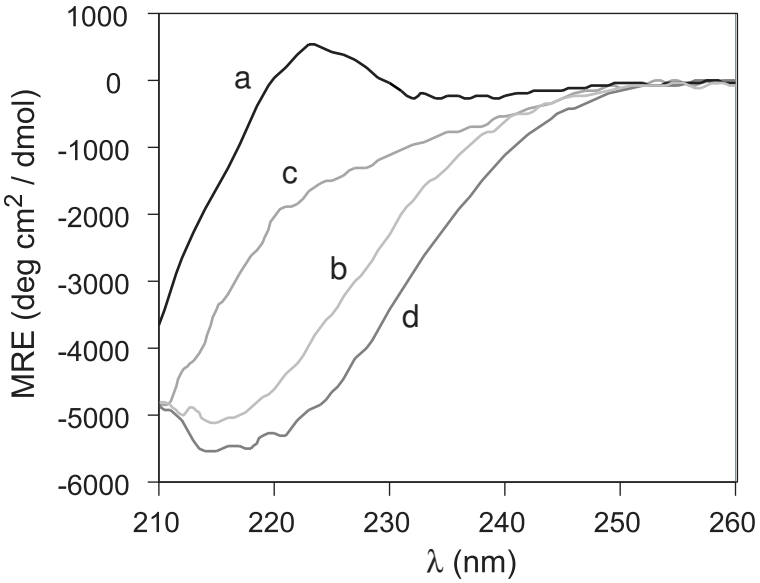


Figure 3.3. Far-UV CD spectra of native (*a*, black) and denatured (*b*, light gray) PsaE and native (*c*, gray) and denatured (*d*, dark gray) EbE1 at pH 7.4 in 20 mM phosphate buffer. Native spectra were obtained at 25 °C, denatured at 95 °C. Both proteins show similar loss of ellipticity upon heating.

the PsaE spectrum, and signified a subtle alteration in the environment of Trp17, the only Trp in the protein. Fluorescence spectra of the two proteins were collected under native conditions to confirm the change (Figure 3.5). The emission maximum of EbE1 was blue-shifted by 2 nm, and its intensity decreased by 15% compared to PsaE, in support of a minor local perturbation and in agreement with the NMR and far-UV CD data. Trp17 is positioned such that its NεH is less than 4 Å away from the 14-residue loop replaced in the creation of the hybrid protein, and the environment adjustment was not surprising.

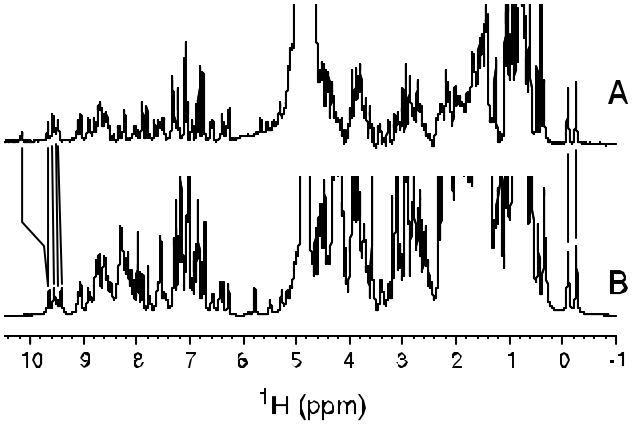


Figure 3.4. ^{1}H NMR spectra of (A) PsaE (25 °C, pH 7.2) and (B) EbE1 (19 °C, pH 7.2, 20 mM phosphate buffer) in 90% ^1H_2O/10% ^2H_2O. Many of the resolved features are nearly identical as shown by the connecting lines.

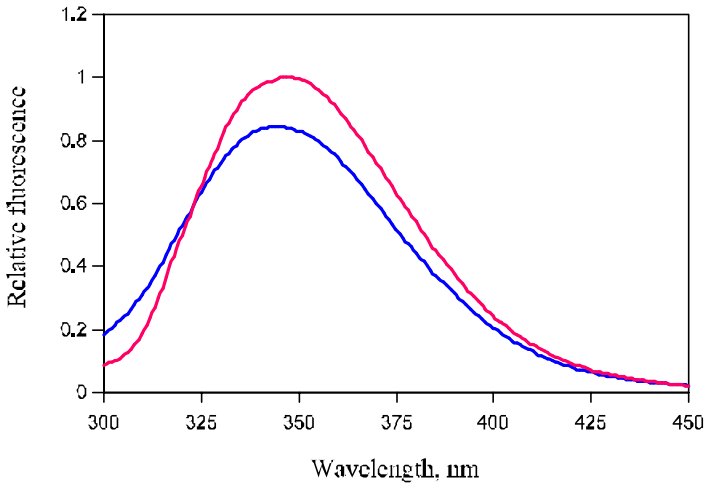


Figure 3.5. Fluorescence spectra of EbE1 (blue) and PsaE (pink) under native conditions (25 °C, pH 7.4). The emission is a result of excitation at 280 nm.

Backbone chemical shifts can be used qualitatively to assess structural similarity. Figure 3.6 presents a comparison of EbE1 to PsaE and, to gauge the significance of the deviations, EbE1 to random coil. The absolute value of Δδ for random coil and EbE1 ranged from 0 ppm to 1.65 ppm, whereas |Δδ| for PsaE and EbE1 ranged only from

0 ppm to 0.12 ppm. Significant spectral alterations (either shift or broadening) on going from PsaE to EbE1 were observed in the proximity of the foreign cytochrome b_5 portion of the molecule at the level of Arg39E, Phe40E, and Asn41E. These residues comprise the final three on the N-terminal side of the insertion. Noticeable changes near the foreign loop and minimal differences elsewhere have been reported for circular permutants of another SH3 domain with varying loop lengths (*19*). Similar results were obtained for Asn56E through Asn59E on the C-terminal side of the EbE1 loop. As mentioned above, the environment surrounding Trp17 has been affected, and this was reflected in the Δδs observed for backbone signals in the PsaE and EbE1 backgrounds. On average, the pronounced chemical shift similarity of the two proteins indicated that EbE1 closely resembled PsaE.

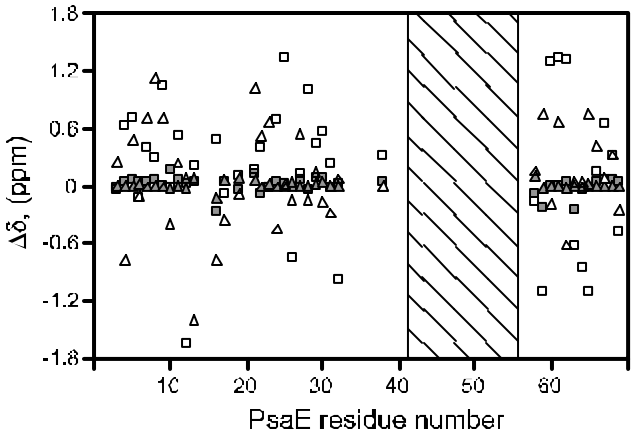


Figure 3.6. Chemical shift differences between PsaE and EbE1 and random coil and EbE1 plotted as a function of PsaE residue number. Squares represent backbone NH (filled, PsaE-EbE1; open, random coil – EbE1), and triangles represent CαH (filled, PsaE-EbE1; open, random coil – EbE1). The hatched region marks the location of the CD and cytochrome b_5 loops within the PsaE scaffold. Several backbone signals of EbE1 were not detected in the region preceeding and following the loop. Chemical shift differences between the two proteins are otherwise small (see Appendix A), indicating that the backbone protons are in nearly identical fold and environment.

Two-dimensional homonuclear NMR experiments were analyzed to inspect the details of the EbE1 fold. The quality of the data is illustrated in Figure 3.7 with the footprint region of the EbE1 NOESY data. The β-strand secondary structure of the hybrid protein was determined by following scalar (NH_i to $C\alpha H_i$, obtained from COSY and TOCSY spectra) and dipolar ($C\alpha H_i$ to NH_{i+1}) connectivities; its topology is summarized in Figure 3.8. EbE1 and PsaE had many common long-range NOEs, including cross-strand effects establishing the register of the sheet. Among the characteristic pairwise CαH interactions were Ile10E to Leu65E and Val8E to Val67E; these confirmed that PsaE strands A and E were antiparallel as in the original protein (*17*). Other NOEs between strands A and B, B and C, C and D, D and E, and E and A were also preserved, in support of the SH3 fold. However, cross-strand NOEs between Arg39E and Thr57E were not found, which verified that the three-dimensional structure leading into and out of the loop had been altered.

The NMR signals contributed by the cytochrome b_5 portion of the hybrid protein were not well resolved, and a minimal number of assignments were obtained. Interestingly, multiple sets of cross-peaks were detected for the Val61b-Gly62b-His63b stretch, and this segment of the protein, therefore, adopted at least two conformations exchanging on the chemical shift time scale. The chemical shifts for one of these conformations agreed well with those previously determined for apocytochrome b_5 (*32*), whereas the second set of observed chemical shifts differed by approximately 0.1 ppm.

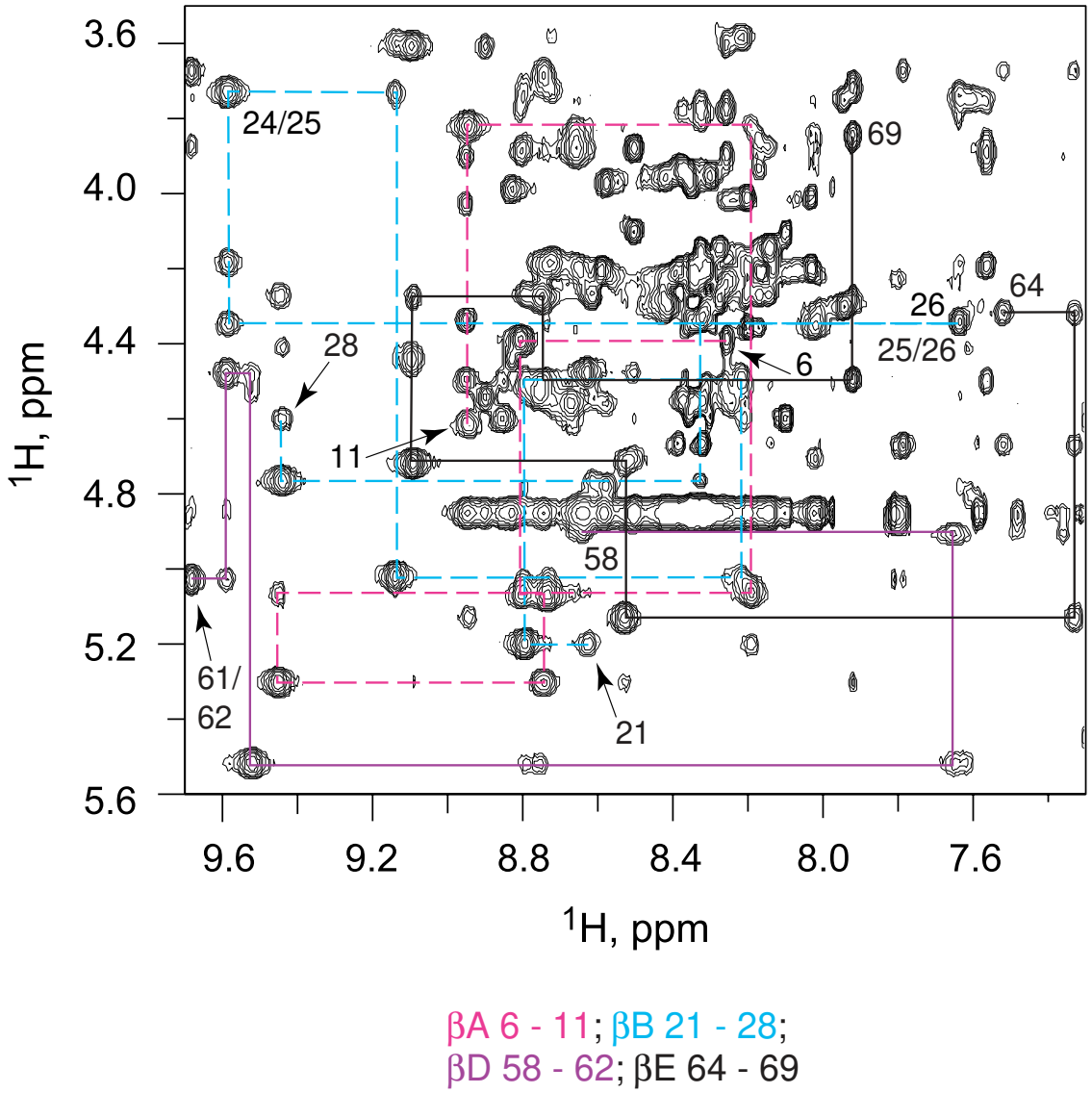


Figure 3.7. Footprint region of the NOESY spectrum of EbE1 obtained in 90% ^1H_2O/10% ^2H_2O at 19 °C (pH 7.2). Many of the NOEs observed for EbE1 are identical to those obtained by PsaE. The β-strand secondary structure of the hybrid protein was determined by following the scalar (NH_i to αH_i, obtained from COSY and TOCSY spectra) and dipolar (αH_i to NH_{i+1}) connectivities. Lines connect cross-peaks corresponding to strands in the SH3 fold (pink, βA, 6 to 11; cyan, βB, 21 to 28; purple, βD 58 to 62; black, βE, 64 to 69). Cross-peaks are identified either with the number *i* for NH_i-αH_i effects or with the numbers *i*/*i*+*1* for sequential αH_i-NH_{i+1} effects.

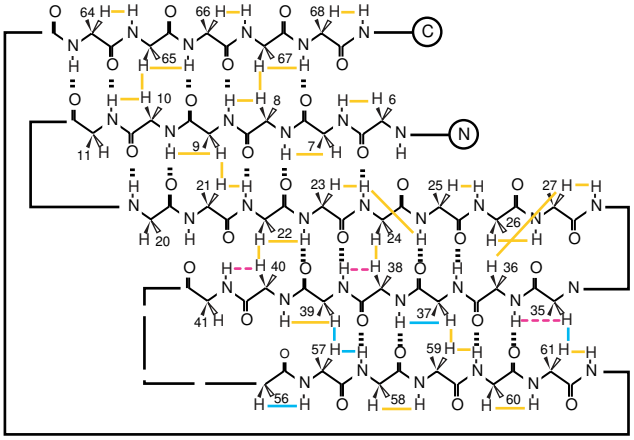


Figure 3.8. Layout of the β-sheet structure of PsaE and EbE1. Black lines symbolize the loops contained in the two proteins: solid black lines, loops common to PsaE and EbE1; dashed black line, CD loop of PsaE replaced by the heme-binding region of cytochrome b_5. Colored lines symbolize inter-residue NOE cross-peaks: gold, NOEs observed in both proteins; cyan, NOEs observed in PsaE but not EbE1; pink, NOEs observed in PsaE that were obscured by overlap in the EbE1 spectrum.

One-dimensional ^{1}H NMR spectra of EbE1 were collected as a function of temperature to monitor the loss of structure upon heating (Figure 3.9). Resolved peaks broadened and decreased in height throughout the heating process, but chemical shifts remained unchanged. This indicated that the equilibrium between N and U was slow with respect to the chemical shift timescale, and implied the existence of an energetic barrier to unfolding. The spectrum of EbE1 obtained upon returning the sample to room temperature exhibited some broad features underneath the sharp, native signals, in a hint that a portion of the sample had been denatured irreversibly via aggregation.

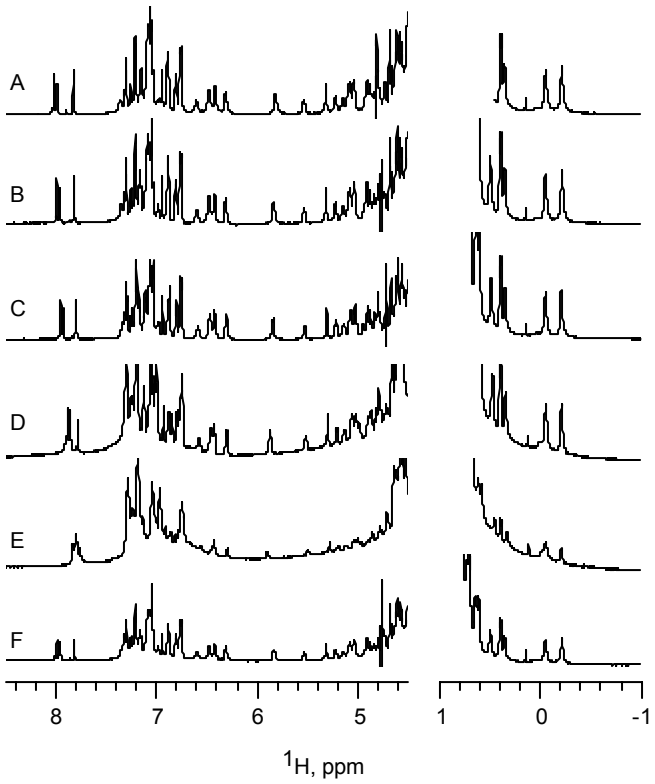


Figure 3.9. ^{1}H NMR spectra of EbE1 in $^{2}H_2O$ as a function of temperature: (A) 21 °C, (B) 25 °C (first heating), (C) 29 °C, (D) 37 °C, (E) 45 °C, (F) 25 °C (after cooling).

3.2.2 Thermal Denaturation of EbE1

The structural and variable-temperature data demonstrated that the PsaE scaffold could tolerate the long insertion, and EbE1 was therefore deemed suitable to inspect the thermodynamic consequences of loop replacement. The thermal denaturation of PsaE and EbE1 was followed by far-UV CD spectroscopy and absorbance spectroscopy. The data were fit to a two-state equilibrium-unfolding model and yielded T_ms for EbE1 and PsaE of 52.5 ± 0.3 °C and 60.8 ± 0.2 °C, respectively. The results are shown in Figure 3.10 and summarized in Table 3.1. Because the difference in folded and unfolded signal was small and the transition for PsaE was not fully reversible, values for the Gibbs free

energy difference obtained from curve fitting were not accurate. However, the presence of a native baseline in the EbE1 unfolding curve was consistent with at most a moderate destabilization compared to PsaE.

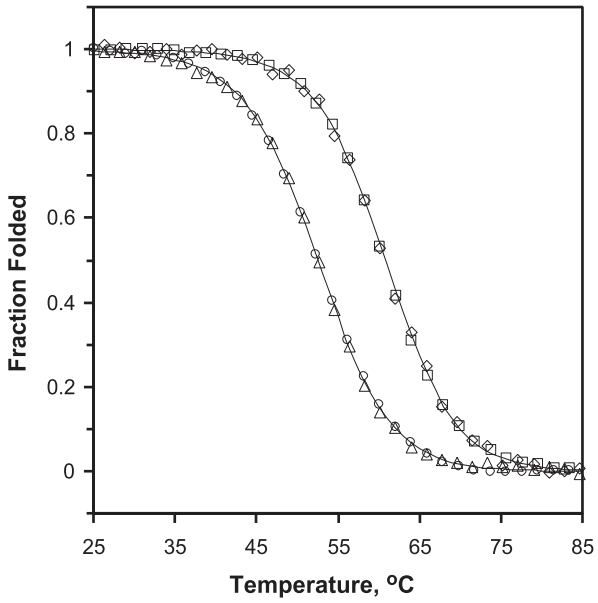


Figure 3.10. Thermal denaturation of EbE1 and PsaE followed by absorbance and CD spectroscopy. Open symbols represent data (Δ, EbE1 CD; ○, EbE1 UV-Vis; ◊, PsaE CD; □, PsaE UV-Vis), and solid lines represent the result of fitting to a two-state model as described in the Materials and methods chapter. Parameters obtained from the best fit of the data are given in Table 3.1.

3.2.3. Chemical Denaturation of EbE1

In contrast to thermal denaturation, the urea-induced process was reversible, and the data (Fig 3.11) could be used for quantitative purposes. Fitting of the EbE1 urea unfolding curves to a two-state model yielded a standard Gibbs free energy of unfolding of 10.27 ± 0.14 kJ/mol. The midpoint of the unfolding transition occurred at ~3.2 M urea, and the *m* value was 3.2 ± 1.0 kJ/mol/M. In comparison, the standard Gibbs free energy of unfolding for PsaE was 14.47 ± 0.03 kJ/mol, with a midpoint of 4.74 M urea

and an *m* value of 3.04 ± 0.10 kJ/mol/M (data are summarized in Table 3.1). As the *m* parameter is generally associated with the amount of nonpolar surface area exposed upon unfolding of the protein, obtaining similar values for two proteins that differed only in the size of a solvent-exposed loop was expected. The numbers above evaluated the difference in the thermodynamic stabilities of the two proteins (ΔΔG°) at ~4.2 kJ/mol.

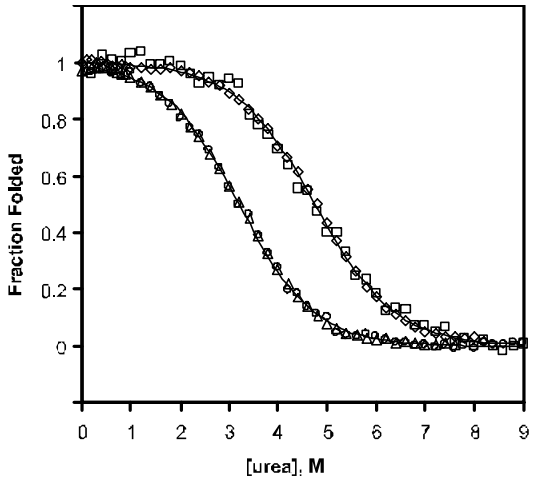


Figure 3.11. Urea denaturation of EbE1 and PsaE followed by absorbance and CD spectroscopy. Open symbols represent data (Δ, EbE1 CD; ○, EbE1 UV-Vis; ◊, PsaE CD; □, PsaE UV-Vis), and solid lines represent the result of fitting to a two-state model as described in the Materials and methods chapter. Parameters obtained from the best fit of the data are given in Table 3.1.

3.3.4. Reconstitution of EbE1

The ability of the heme-binding region to form coordination bonds with the iron of the physiological cofactor in the absence of its native scaffold was investigated by incubating EbE1 with iron protoporphyrin IX (Fe-PPIX). For comparison purposes, the UV-Visible absorption spectra of EbE1 that had been exposed to Fe-PPIX and that of free porphyrin are shown in Figure 3.12A. Other than the protein peak in the 280-nm range,

Table 3.1. Thermodynamic parameters for the denaturation of PsaE and EbE1

	T_m [a]	$\Delta H°$ [b]	$\Delta C°_p$	$\Delta G°_U$ [c]	m	C_m [d]
	°C	kJ mol^{-1}	kJ mol^{-1} K^{-1}	kJ mol^{-1}	kJ mol^{-1} M^{-1}	M
PsaE	60.8 ± 0.2	216 ± 10	2.5 ± 0.1	14.47 ± 0.03	3.04 ± 0.10	4.7
EbE1	52.5 ± 0.3	196 ± 18	2.5 ± 0.8	10.27 ± 0.14	3.2 ± 1.0	3.2

[a] Midpoint of thermal denaturation in 20 mM phosphate buffer at pH 7.4. Errors are standard deviations from globally fitting multiple data sets. [b] Change in enthalpy evaluated at the midpoint of the thermal unfolding transition. [c] Free energy of unfolding (25 °C, 20 mM phosphate buffer, pH 7.4). [d] Midpoint of urea denaturation (25 °C, 20 mM phosphate buffer, pH 7.4), calculated from ΔG° and m.

the two spectra were identical, indicating that EbE1 was not able to bind the prosthetic group specifically. Furthermore, the one-dimensional ^{1}H NMR spectra of EbE1 solutions in the presence and absence of heme were identical (not shown), even when the spectral width was increased in an attempt to observe the population of any high-spin species. This indicated that less than 5% of the protein was able to populate the holoprotein state.

Lack of specific binding could be explained in several ways, one of which being that the protein has a lower specific heme affinity than the parent cytochrome. If the affinity decreased enough to make heme aggregation more favorable than specific heme binding, the equilibrium would favor the unbound state. If heme-heme interactions were, in fact,

the reason for the apparent lack of binding, a water-soluble derivative of the prosthetic group would be expected to produce different results. The porphyrin group chosen was Fe(III) meso-Tetra(4-carboxyphenyl)porphine chloride (Fe-TCP, Figure 3.12B). Because this molecule is significantly bulkier than Fe-PPIX, cytochrome b_5 was first reconstituted

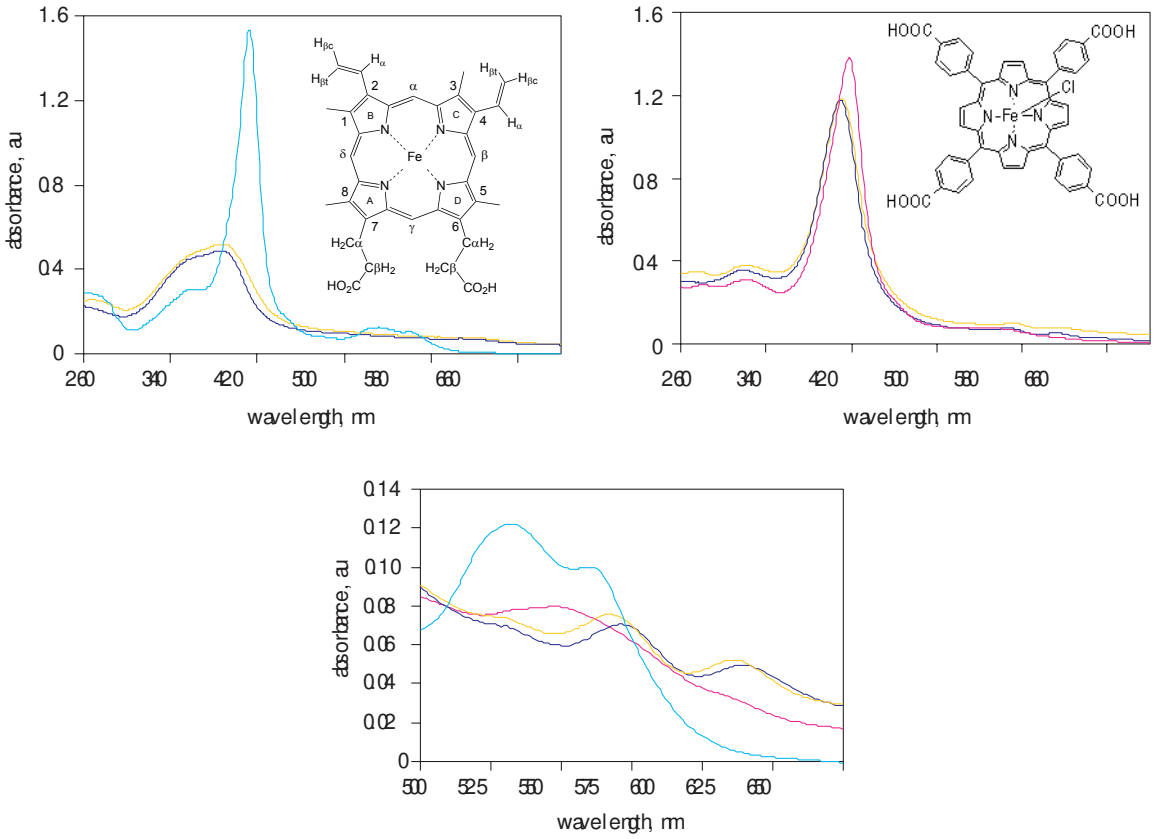


Figure 3.12. (A) UV-Visible spectra of EbE1(gold) and cytochrome b_5 (cyan) after reconstitution with iron protoporphyrin IX (Fe-PPIX, shown as inset). (B) UV-Visible spectra of EbE1(gold) and cytochrome b_5 (cyan) after reconstitution with iron (III) meso-Tetra(4-carboxyphenyl) porphine chloride (Fe-TCP, shown as inset). In (A) and (B) the spectrum of the free porphyrin is shown in blue, and a constant has been added to the spectrum of EbE1 for ease of viewing. (C) Blow-up of the visible region of the spectra of free Fe-TCP (gold), Fe-PPIX holo cyt b_5 (cyan), Fe-TCP holo cyt b_5 (pink), and EbE1 incubated with Fe-TCP (blue). The Fe-TCP holo cyt b_5 traces displays similar α and β bands as Fe-PPIX holo cyt b_5, consistent with bis-histidine coordination of the heme iron. By comparison, the spectra of free Fe-TCP and EbE1 incubated with Fe-TCP are similar and characteristic of a high-spin heme iron. Cytochrome b_5 spectra supported specific cofactor association in both cases; EbE1 did not bind specifically either porphyrin.

with the prosthetic group to prove that binding was still possible. The UV-Visible spectrum of cytochrome b_5 incubated with Fe-TCP (Figure 3.12B,C) suggested that the protein was able to bind the modified porphyrin; the Soret maximum was shifted by 5 nm relative to the cytochrome b_5-Fe-PPIX complex. As evidenced by the α and β bands present in the 500 nm to 650 nm range (Fig 3.12C), a low-spin complex was obtained. This was consistent with bis-histidine coordination of the iron as observed for the physiological prosthetic group. However, the affinity for the porphyrin was diminished as evidenced by regeneration of the apoprotein via ion-exchange chromatography. When the same reconstitution was attempted with EbE1, no evidence for specific binding was observed.

In an attempt to investigate further the requirements for formation of specific protein-cofactor association involving the cytochrome b_5 heme-binding loop, a Zn-substituted porphyrin was used (Zn-PPIX). Although exceptions have been noted, Zn typically maintains a pentacoordinate ligation scheme. Owing to this property, the cofactor, if it bound, was expected to do so in a manner such that only one of the axial coordination bonds between the protein and the porphyrin's metal center was formed. The one-dimensional ^{1}H NMR spectrum of cytochrome b_5 upon incubation with Zn-PPIX showed neither new protein peaks nor resonances of the free porphyrin (data not shown), indicating that the two species were non-specifically associated. Hence, it appeared that the cytochrome needed to participate in two coordination bonds in order to bind specifically its cofactor.

3.3. Discussion

3.3.1 Loop Effects

The denaturation experiments indicated that the destabilization caused by replacing the 14-residue CD loop of PsaE with the 43-residue heme-binding loop of cytochrome b_5 was small. The structural deviations imposed on the original SH3 domain by the insertion, though modest, involved the disruption of H-bonds between the C-terminal end of the βC strand and the N-terminal end of the βD strand; these disruptions were likely to contribute an enthalpic component to the observed thermodynamic effect.

There are several difficulties in comparing the experimental ΔΔG° with theoretical estimates of TΔΔS°. First, given the structural distortions at the base of the EbE1 loop, there is some uncertainty in the linking numbers that should be used in equation 1. Second, the enthalpic contribution could not be obtained accurately from the thermal denaturation data. The third difficulty, and the most daunting, rests with the scaling factor k in equation 1. The freely-jointed three-dimensional chain has a k of 1.5, but excluded volume effects are expected to raise this value (*9, 33*). Calculations for a self-avoiding chain in a face-centered cubic lattice return $k = 2.16$ (*33*). The location of the loop within the polymer also influences the anticipated extent of destabilization. In apocytochrome b_5, the flexible region spans the central portion of the primary structure, whereas the CD loop of PsaE is closer to the C than the N terminus. This discrepancy in loop position would decrease the estimated ΔΔS°. Interestingly, surveys of structures contained in the PDB show that the probability of contact between two amino acids, i and j, decreases linearly as $|i\text{-}j|^k$, where k is 1.5 – 1.64 (*34, 35*). However, experimental data collected on disulfide-containing proteins suggest k values close to 2.4 and 2.5 (*9, 36*),

whereas certain iso-1-cytochrome *c* loops yield a higher value of 4.2 (*37*). A recent analysis (*38*) demonstrates the need to incorporate corrections for excluded volume into the value of *k* as well as the assumptions made regarding the differences between the reference and inserted loops.

As discussed by Bowler and coworkers (*37*), several sequence characteristics impinge on the scaling factor. For example, a rise in proline content will increase the persistence length of the chain and its *k* value, whereas an increase in glycine content will have the opposite effect (*39*), with composition having the largest effect for loops with $l < 40$ (*40*). Regarding the PsaE-to-EbE1 change, the proline content is pratically constant, rising from 0% to 2%, but the glycine content decreases from 29% to 12%. The CD loop and the heme-binging loop should therefore be treated with different *k*s, reflecting a stiffer chain in the latter case. Residual structure can also affect the dependence of $\Delta S°$ on loop length; it may either facilitate closure by compacting the chain in a favorable conformation, or hinder it by increasing the persistence length. Although no regular structure is observed in the CD loop (*17*), the 43-residue cytochrome segment forms a couple of helical turns in its wild-type context (*3*). Whether or not a similar conformation was sampled by the EbE1 loop could not be ascertained.

Overall, the calculated entropic destabilization relying on the same *k* value for 14- and 43-residue loops ranged between 4 and 11 kJ/mol; these estimates are of limited usefulness considering the above reservations. Regardless of the agreement between calculated and observed values, the EbE1 insertion is of little thermodynamic consequence, and this chimera adds to the list of proteins for which loop elongation is minimally perturbing.

3.3.2. Attempted Reconstitution of EbE1

The cytochrome b_5 binding loop was unable to bind either the modified or physiologically relevant hemes in the context of the PsaE scaffold. This could result from a lack of flexibility at the interface of the PsaE and cytochrome b_5 portions of EbE1 that would be required to allow for folding of both regions of the protein. It has been shown that cytochrome b_5, unlike myoglobin and some hemoglobins (*41, 42*), cannot bind specifically protoporphyrin IX (PPIX) or zinc protoporphyrin IX (Zn-PPIX) (current work). The inability to bind cofactors with which it can form zero (PPIX) or one (Zn-PPIX) coordination bonds rather than two as it does with its physiological binding partner (Fe-PPIX) indicates that this interaction is the main contributor to the enthalpy in the specific heme-protein contact and that hydrogen bonding, van der Waals and electrostatic interactions, and hydrophobic forces are significantly less important. Hence, specific heme binding by EbE1 is not likely to be observed unless the protein folds such that both coordination bonds can be formed. The possibility that an increase in hinge-region flexibility could lead to specific binding of the prosthetic group was investigated by the insertion of a Gly at the N-terminus of the cyt b_5 loop (EbE2) or a Gly at both the N and C termini (EbE3), and the results will be discussed in the following chapter. In addition, it is possible that core 1-core 2 communication is necessary for physiological function. Such long-range interactions in cytochrome b_5 have been demonstrated previously by the ability of core 2 variants to affect the heme affinity of the protein (*43*).

3.4. Conclusions

The results obtained with EbE1 lead to a better understanding of the properties of the parent proteins. In photosystem I, the CD loop of PsaE plays a functional role; it makes

contact with PsaC (another extrinsic protein) and the stromal loops of PsaA and PsaB. These interactions naturally result in conformational rearrangement compared to the family of solution structures (*44*), but there is no alteration of the H-bonding network within the resilient scaffold. Interestingly, the thermodynamic stability of a related PsaE protein (*Nostoc* sp. PCC 8009) with a 7- rather than 14-residue loop is similar to that reported here (*22*). The ability of PsaE to accommodate 7-36 extra residues without extreme consequences also supported that this particular insertion location was ideal.

Extrapolation of the EbE1 results to the architecture of the cytochrome can be contemplated in light of the above. If the CD loop has a minor effect on its own scaffold, then its removal and replacement with a foreign loop should inform principally on the effect of the latter. Assuming that the PsaE and core 2 backgrounds are equally receptive of the heme-binding region, the observations made for EbE1 lend support to the view that core 2 has a low intrinsic stability. We conclude that the thermodynamics of the folded protein of apocytochrome b_5 need not be severely affected by the size of the heme-binding loop. In addition, the protein architecture may result in communication between the folded cores in the holoprotein. The inability of the isolated heme-binding region to bind heme specifically may indicate that core 2 has a dual role rather than serving merely a structural purpose.

References

(1) Mathews, F. S., Czerwinski, E. W., and Argos, P. (1979) The X-ray crystallographic structure of calf liver cytochrome b_5, in *The Porphyrins* (Dolphin, D., Ed.) pp 107-147, Academic Press, New York.

(2) Moore, C. D., and Lecomte, J. T. J. (1993) Characterization of an independent structural unit in apocytochrome b_5. *Biochemistry 32*, 199-207.

(3) Falzone, C. J., Mayer, M. R., Whiteman, E. L., Moore, C. D., and Lecomte, J. T. J. (1996) Design challenges for hemoproteins: the solution structure of apocytochrome b_5. *Biochemistry 35*, 6519-6526.

(4) Constans, A. J., Mayers, M. R., Sukits, S. F., and Lecomte, J. T. J. (1998) A test of the relationship between sequence and structure in proteins: excision of the heme binding site in apocytochrome b_5. *Protein Sci. 7*, 1983-1993.

(5) Cowley, A. B., Altuve, A., Kuchment, O., Terzyan, S., Zhang, X., Rivera, M., and Benson, D. R. (2002) Toward engineering the stability and hemin-binding properties of microsomal cytochromes b_5 into rat outer mitochondrial membrane cytochrome b_5: examining the influence of residues 25 and 71. *Biochemistry 41*, 11566-11581.

(6) Pfeil, W. (1993) Thermodynamics of apocytochrome b_5 unfolding. *Protein Sci. 2*, 1497-1501.

(7) Storch, E. M., Daggett, V., and Atkins, W. M. (1999) Engineering out motion: introduction of a de novo disulfide bond and a salt bridge designed to close a dynamic cleft on the surface of cytochrome b5. *Biochemistry 38*, 5054-64.

(8) Storch, E. M., Grinstead, J. S., Campbell, A. P., Daggett, V., and Atkins, W. M. (1999) Engineering out motion: a surface disulfide bond alters the mobility of tryptophan 22 in cytochrome b_5 as probed by time-resolved fluorescence and ^{1}H NMR experiments. *Biochemistry 38*, 5065-5075.

(9) Chan, H. S., and Dill, K. A. (1989) Intrachain loops in polymers: effects of excluded volume. *J. Chem. Phys. 90*, 492-509.

(10) Batori, V., Koide, A., and Koide, S. (2002) Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain. *Protein Eng. 15*, 1015-1020.

(11) Collinet, B., Garcia, P., Minard, P., and Desmadril, M. (2001) Role of loops in the folding and stability of yeast phosphoglycerate kinase. *Eur. J. Biochem. 268*, 5107-5118.

(12) Tada, M., Kobashigawa, Y., Mizuguchi, M., Miura, K., Kouno, T., Kumaki, Y., Demura, M., Nitta, K., and Kawano, K. (2002) Stabilization of protein by replacement of a fluctuating loop: structural analysis of a chimera of bovine alpha-lactalbumin and equine lysozyme. *Biochemistry 41*, 13807-13.

(13) Ladurner, A. G., and Fersht, A. R. (1997) Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J. Mol. Biol. 273*, 330-337.

(14) Scalley-Kim, M., Minard, P., and Baker, D. (2003) Low free energy cost of very long loop insertions in proteins. *Protein Sci. 12*, 197-206.

(15) Doi, N., Itaya, M., Yomo, T., Tokura, S., and Yanagawa, H. (1997) Insertion of foreign random sequences of 120 amino acid residues into an active enzyme. *FEBS Lett. 402*, 177-180.

(16) Pace, C. N., Grimsley, G. R., Thomson, J. A., and Barnett, B. J. (1988) Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J Biol Chem 263*, 11820-5.

(17) Falzone, C. J., Kao, Y.-H., Zhao, J., Bryant, D. A., and Lecomte, J. T. J. (1994) Three-dimensional solution structure of PsaE from cyanobacterium *Synechococcus* sp. strain PCC 7002, a photosystem I protein that shows structural homology with SH3 domains. *Biochemistry 33*, 6052-6062.

(18) Martinez, J. C., Viguera, A. R., Berisio, R., Wilmanns, M., Mateo, P. L., Filimonov, V. V., and Serrano, L. (1999) Thermodynamic analysis of alpha-spectrin SH3 and two of its circular permutants with different loop lengths: discerning the reasons for rapid folding in proteins. *Biochemistry 38*, 549-559.

(19) Viguera, A. R., and Serrano, L. (1997) Loop length, intramolecular diffusion and protein folding. *Nat. Struct. Biol. 4*, 939-946.

(20) Filimonov, V. V., Azuaga, A. I., Viguera, A. R., Serrano, L., and Mateo, P. L. (1999) A thermodynamic analysis of a family of small globular proteins: SH3 domains. *Biophys. Chem. 77*, 195-208.

(21) Lim, W. A., Fox, R. O., and Richards, F. M. (1994) Stability and peptide binding affinity of an SH3 domain from the *Caenorhabditis elegans* signaling protein Sem-5. *Protein Sci. 3*, 1261-1266.

(22) Mayer, K. L., Shen, G., Bryant, D. A., Lecomte, J. T. J., and Falzone, C. J. (1999) The solution structure of photosystem I accessory protein E from the cyanobacterium *Nostoc* sp. strain PCC 8009. *Biochemistry 38*, 13736-13746.

(23) Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol 6*, 197-208.

(24) Gupta, N. K., Ohtsuka, E., Sgaramella, V., Buchi, H., Kumar, A., Weber, H., and Khorana, H. G. (1968) Studies on polynucleotides, 88. Enzymatic joining of chemically synthesized segments corresponding to the gene for alanine-tRNA. *Proc Natl Acad Sci U S A 60*, 1338-44.

(25) Bousquet, J. A., Garbay, C., Roques, B. P., and Mely, Y. (2000) Circular dichroic investigation of the native and non-native conformational states of the growth factor receptor-binding protein 2 N-terminal src homology domain 3: effect of binding to a proline-rich peptide from guanine nucleotide exchange factor. *Biochemistry 39*, 7722-7735.

(26) Knapp, S., Karshikoff, A., Berndt, K. D., Christova, P., Atanasov, B., and Ladenstein, R. (1996) Thermal unfolding of the DNA-binding protein Sso7d from the hyperthermophile *Sulfolobus solfataricus*. *J. Mol. Biol. 264*, 1132-1144.

(27) Okishio, N., Tanaka, T., Fukuda, R., and Nagai, M. (2003) Differential ligand recognition by the Src and phosphatidylinositol 3-kinase Src homology 3 domains: circular dichroism and ultraviolet resonance Raman studies. *Biochemistry 42*, 208-216.

(28) Renzoni, D. A., Pugh, D. J. R., Siligardi, G., Das, P., Morton, C. J., Rossi, C., Waterfield, M. D., Campbell, I. D., and Ladbury, J. E. (1996) Structural and thermodynamic characterization of the interaction of the SH3 domain from Fyn with the proline-rich binding site on the p85 subunit of PI3-kinase. *Biochemistry 35*, 15646-15653.

(29) Viguera, A. R., Arrondo, J. L., Musacchio, A., Saraste, M., and Serrano, L. (1994) Characterization of the interaction of natural proline-rich peptides with five different SH3 domains. *Biochemistry 33*, 10925-10933.

(30) Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L., and Serrano, L. (1994) Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry 33*, 2142-2150.

(31) Schmid, F. X. (1989) Spectral methods of characterizing protein conformation and conformational changes, in *Protein structure. A practical approach* (Creighton, T. E., Ed.) pp 251-285, IRL Press, New York.

(32) Falzone, C. J., Wang, Y., Vu, B. C., Scott, N. L., Bhattacharya, S., and Lecomte, J. T. J. (2001) Structural and dynamic perturbations induced by heme binding in cytochrome b_5. *Biochemistry 40*, 4879-4891.

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

(34) Berezovsky, I. N., Grosberg, A. Y., and Trifonov, E. N. (2000) Closed loops of nearly standard size: common basic element of protein structure. *FEBS Lett 466*, 283-6.

(35) Dokholyan, N. V., and Shakhnovich, E. I. (2001) Understanding hierarchical protein evolution from first principles. *J Mol Biol 312*, 289-307.

(36) Darby, N. J., and Creighton, T. E. (1993) Dissecting the disulphide-coupled folding pathway of bovine pancreatic trypsin inhibitor. Forming the first disulphide bonds in analogues of the reduced protein. *J Mol Biol 232*, 873-96.

(37) Hammack, B. N., Smith, C. R., and Bowler, B. E. (2001) Denatured state thermodynamics: residual structure, chain stiffness and scaling factors. *J Mol Biol 311*, 1091-104.

(38) Wang, L., Rivera, E. V., Benavides-Garcia, M. G., and Nall, B. T. (2005) Loop entropy and cytochrome c stability. *J. Mol. Biol. 353*, 719-729.

(39) Schimmel, P. R., and Flory, P. J. (1968) Conformational energy and configurational statistics of poly-L-proline. *Proc. Natl. Acad. Sci. U.S.A. 58*, 52-59.

(40) Conrad, J. C., and Flory, P. J. (1976) Moments and distribution functions for polypeptide chains. Poly-alanine. *Macromolecules 9*, 41-47.

(41) Sil, S., Bose, T., Roy, D., and Chakraborti, A. S. (2004) Protoporphyrin IX-induced structural and functional changes in human red blood cells, haemoglobin and myoglobin. *J Biosci 29*, 281-91.

(42) Sil, S., and Chakraborti, A. S. (1996) Comparative studies on the interaction of protoporphyrin with hemoglobin and myoglobin. *Indian J Biochem Biophys 33*, 285-91.

(43) Mukhopadhyay, K., and Lecomte, J. T. (2004) A relationship between heme binding and protein stability in cytochrome b5. *Biochemistry 43*, 12227-36.

(44) Fromme, P., Jordan, P., and Krauss, N. (2001) Structure of photosystem I. *Biochim. Biophys. Acta 1507*, 5-31.

Chapter 4

Control of Thermodynamic and Kinetic Properties in a Family of PsaE-cytochrome b_5 Chimeras via Loop Anchor Modification

4.1. Introduction

Based on recent prediction algorithms (*1-3*), it has been proposed that a large number of proteins encoded in eukaryotic genomes (as well as a smaller percentage of prokaryotic proteins (*4, 5*)) either contain extended regions of unstructured amino acids or are disordered altogether. Among those proteins about which functional information is known, many are involved in signal transduction (*6*) or regulation on the transcriptional (*4*) or translational levels (*7, 8*). The lack of secondary and tertiary structure has been rationalized in terms of additional accessible surface area for binding partners, ability to bind multiple targets, and coupling of enthalpic and entropic components of protein-protein interactions (*9-12*). Proteins performing a known function with little or no high-order structure include the C-terminal portion of the yeast transcriptional activator Bob-1 (*13, 14*), the cyclin-dependent kinase inhibitor $p21^{Waf1/Cip1/Sdi1}$ (*15*), and the N-terminal region of the bacterial sigma factor inhibitor FlgM (*16*). Many other proteins fold only upon binding either a small molecule or a protein or nucleic acid target. The basic DNA-binding region of the leucine zipper protein GCN4 (*17*), the yeast zinc-finger DNA-binding domain ADR1 (*18*), the acidic activation domain of the herpes simplex virus VP16 protein (*19, 20*), the transactivation domain of the nuclear tumor suppressor phosphoprotein p53 (*21*), and the transcriptional coactivator CREB binding protein (CBP) (*22*) belong to this group of proteins.

In addition to proteins that are either mostly or entirely unstructured, another class of proteins has garnered significant interest lately: proteins that are largely structured but contain long, flexible loops. These loops are often found on the surface of intracellular proteins and display low sequence complexity. Thousands of proteins are predicted to fall into this group (*4, 23*). Examples include several viral coat proteins (*24*), thyroid transcription factor (*24*), beef liver catalase (*25*), and formate dehydrogenase H (*26*). The locations of the loop codons within the gene often coincide with exon boundaries, implying that the flexible segment had been inserted into the coding region of a pre-existing protein (*27*).

Among homologous proteins, sequence conservation in the loop appears to be lower than in the structured portion of the protein. Because large unstructured loops do not participate in cooperatively folded protein cores, amino acid replacements within them typically do not compromise the structure and stability of the scaffold. This decoupling mechanism provides the conformational adaptability necessary to respond to different ligands and is seen in germline antibodies (*28*). In contrast, modifications within the core of a protein, unless carefully designed to preserve or improve packing and electrostatic interactions, tend to decrease the stability of the fold. The interface of these two dynamically distinct regions is composed of a set of residues with side chains involved in relatively few intramolecular interactions and backbones changing from well constrained in the core to varying broadly in the unstructured segment. This interfacial region may be amenable to structural readjustment at an energetic cost that can be tolerated by the core.

Synechococcus sp. PCC 7002 (S7002) PsaE provides a convenient β scaffold for detailed study of loop anchor effects. PsaE has the fold of an SH3 domain (Figure 4.1),

although it shares no primary structure identity with these proteins (*29*). The thermodynamic stability of S7002 PsaE is ~14 kJ mol^{-1} (*30*) and sufficient to withstand a range of perturbation caused by amino acid replacement and loop insertion. PsaE strands C and D are connected by a 14-residue disordered loop. To create a model of a loop-scaffold protein, the CD segment can be swapped for a longer one. As a substitute, the heme-binding loop of microsomal cytochrome b_5 (cyt b_5, Figure 2) has the required characteristics of length (30-40 residues) and, in the apoprotein state, lacks predicted (*1, 31*) and observed (*32-34*) rigid structure. In addition, a cellular environment does not force the loop to refold or collapse onto itself (*35*). These properties suggest that the loop sequence was selected through evolution not only for adopting a specific fold when exposed to the heme cofactor, but also to avoid the formation of energetically favorable contacts in its absence.

In prior work, we used a PsaE-cyt b_5 hybrid to evaluate the contribution of the heme-binding loop to the thermodynamic properties of the apocytochrome protein (see Chapter 3 of this thesis). Replacement of the CD loop of PsaE with residues 32-74 from rat microsomal cyt b_5 causes a destabilization of the PsaE scaffold that is consistent with the lower limit of expectation based on estimates of loop closure entropy (*36*). In this chimera, referred to as EbE1 for the Psa<u>E</u>-cyt <u>b_5</u>-Psa<u>E</u> origin of the sequence, minor structural perturbations, including separation of the C and D strands as they lead into the loop, are observed (*30*).

In the present work, we explored further the effects of loop insertion on the SH3-like scaffold. The primary structure of the original chimera was altered with either Gly insertions or variation of the sequence at the termini of the loop. We found that the interfacial residues have the capacity to control the standard free energy and kinetics of folding, the aggregation state of the protein, and the conformation of the scaffold. We propose that the residues at the boundary between structured and unstructured regions are important for the maintenance of essential biophysical properties of the protein.

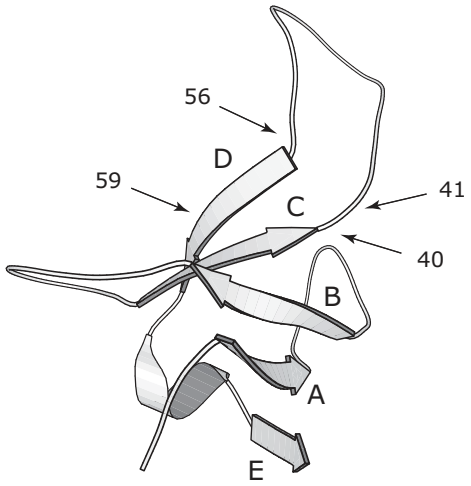


Figure 4.1. Ribbon diagram of PsaE from *Synechococcus* sp. PCC 7002 (1psf, (*37*)). The numbers indicate the boundaries of the scaffold, which includes residues 1–41 and 56–69 (EbE1-3) or 1–40 and 59–69 (EbE4) and excludes the CD loop.

4.2. Results

4.2.1. Design of the EbE Family

Three chimeric proteins were created from the PsaE–cyt b_5 background, and their primary structures are shown in Figure 4.2. The loop insertion replaces the native CD loop of the PsaE protein with the cytochrome binding loop, which encompasses α and β elements in the order (β3)–α2–α3–β5–α4–α5–(β2), where the parentheses indicate

partial inclusion. The structure adopted by this loop in the holoprotein is shown in Figure 4.3. EbE2 was closely related to the EbE1 protein (PsaE 1-41, cyt b_5 32-74, PsaE 56-69), but bore a Gly insert at the N-terminal side of the cyt b_5 region to provide flexibility and possibly resolve structural strains that prevented some of the inter-strand H-bonds from being formed in the EbE1 scaffold (*30*). EbE3 comprised the same PsaE and cyt b_5 residues as EbE1 plus Gly residues at the N and C termini of the loop, again to allow for a readjustment of the dihedral angles at the interface and a decoupling of the loop from the scaffold. The primary structure of EbE4 was PsaE 1-40, cyt b_5 28-76, PsaE 59-69. Compared to EbE1, the PsaE portion of EbE4 was truncated by one residue on the N-terminal side of the loop and three residues on the C-terminal side, two of which were part of the D strand in the parent protein. Compensating, but asymmetrical changes were made to the loop with an extension by three residues on the N-terminal side and one residue on the C-terminal side. These changes at the interface of the protein sequences had the potential to establish a hydrogen-bond competition between the scaffold and the inserted loop. Possible strand pairings in EbE4 included β3 with β2, at the expense of the C/D interactions native to PsaE, and β3 with D, at the expense of β3/β2 native to cyt b_5. In order to facilitate comparison of hybrid protein residues to the corresponding positions in the parent proteins, the original numbering was preserved and the letter E or b was appended to denote the origin of the amino acid. For example, Ile10 from PsaE and Phe58 from cytochrome b_5 became Ile10E and Phe58b, respectively, in the EbE proteins.

```
Cyt b5 ---------A EQSDKDVKYY TLEEIQKHKD SKSTWVILHH KVYD-LTKFL EEHPGGEEVL
PsaE   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF N----KVNYN GFSGSAGGL-
EbE1   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF N----LTKFL EEHPGGEEVL
EbE2   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF N---GLTKFL EEHPGGEEVL
EbE3   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF N---GLTKFL EEHPGGEEVL
EbE4   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF KVYD-LTKFL EEHPGGEEVL

Cyt b5 REQAGGDATE NFEDVGHSTD ARELSKTY-I IGELHPDDRS KIAKPSETL
PsaE   ---------- ---------- ---------- -NTNNFAEHE LEVVG----
EbE1   REQAGGDATE NFEDVGHSTD ARELSKTY-- -NTNNFAEHE LEVVG----
EbE2   REQAGGDATE NFEDVGHSTD ARELSKTY-- -NTNNFAEHE LEVVG----
EbE3   REQAGGDATE NFEDVGHSTD ARELSKTYG- -NTNNFAEHE LEVVG----
EbE4   REQAGGDATE NFEDVGHSTD ARELSKTY-I I---NFAEHE LEVVG----
```

Figure 4.2. Primary structures of rat hepatic cytochrome b_5 (cyt b_5), S7002 PsaE (PsaE) and the chimeric proteins EbE1-4 containing the scaffold of PsaE and the heme-binding region of cytochrome b_5. The portions of the parent proteins that were used in the construction of the various EbE proteins are shown in blue (PsaE) and red (cyt b_5).

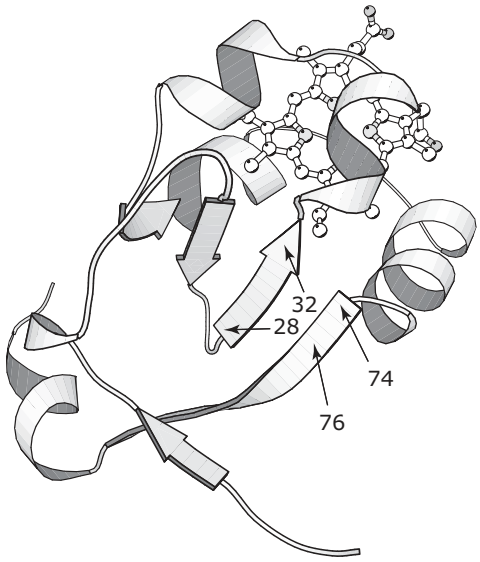


Figure 4.3. Ribbon diagram of the heme domain of cytochrome b_5 (1cyo, (*38*)). The numbers indicate the boundaries of the inserted loop, which includes residues 32–74 (EbE1-3) or 28–76 (EbE4). Residues 28 and 32 are located in strand β3; residues 74 and 76 are in strand β2.

4.2.2. Structural Effects

The backbone conformations of the proteins were examined by far-UV CD spectroscopy. EbE2 (Figure 4.4) and EbE3 (not shown) exhibited properties that were almost identical to those of EbE1, indicating that, as expected, the insertion of glycine residue(s) at the loop termini had little effect on the average secondary structure of the protein. EbE4, in contrast, displayed a spectrum slightly different over the whole wavelength range, hinting at an alteration of secondary and tertiary structure.

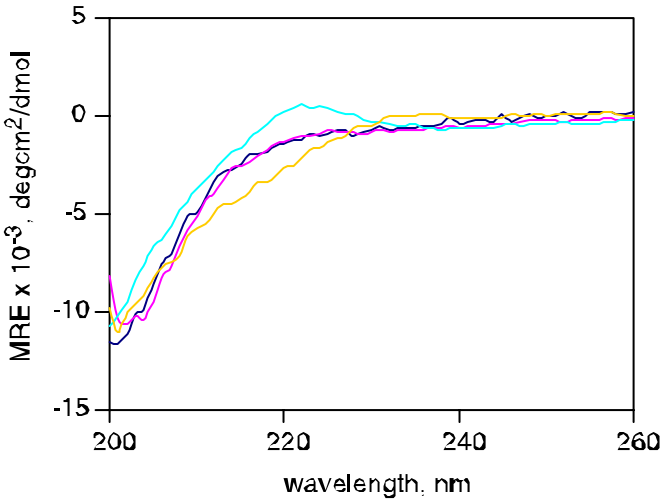


Figure 4.4. Far-UV CD spectra of EbE1 (blue), EbE2 (pink), EbE4 (gold), and PsaE (cyan). The EbE3 spectrum was nearly identical to that of EbE2 and was omitted for clarity. Data were collected at 25 °C in 20 mM phosphate buffer, pH 7.4.

NMR spectroscopy was used to evaluate the structural similarities of EbE2-4 to the parent PsaE scaffold (*39*)and the previously-described EbE1 protein (*30*). Figure 4.5 compares the 1D ^{1}H spectra of all five proteins. In addition to indicating that the proteins were folded under the experimental conditions, the similarities in the resolved peaks in the upfield and downfield regions signified that the overall fold of the PsaE SH3 domain had been preserved. The upfield resonances in the EbE4 spectrum were identified as the

protons from a γ-CH_3 group of Val38E (-0.19 ppm) and the δ-CH_3 group of Ile10E (-0.06 ppm). The downfield peaks were mostly backbone NHs, with the exception of the indole NεH of Trp17E at 9.50 ppm. EbE2 and EbE3 aggregated at NMR concentrations, and were not used for further structural studies. EbE4 was soluble at concentrations necessary for multidimensional NMR experiments, and was subjected to additional characterization.

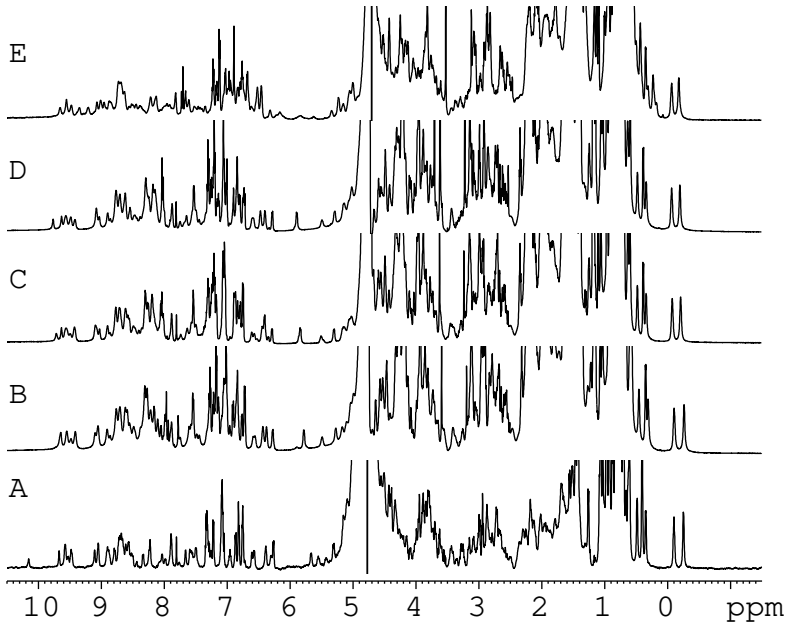


Figure 4.5. ^{1}H NMR spectra of (A) PsaE (25 °C, pH 7.2), (B) EbE1 (19 °C, pH 7.2, 20 mM phosphate buffer, (C) EbE2 (25 °C, pH 7.4, 20 mM phosphate buffer), (D) EbE3 (25 °C, pH 7.4, 20 mM phosphate buffer), and (E) EbE4 (25 °C, pH 7.4, 20 mM phosphate buffer) in 90% $^{1}H_2O$/10% $^{2}H_2O$. Many of the resolved resonances are the same among the family of proteins.

The ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N-labeled EbE4 (Figure 4.6) contained significantly more resonances than would be predicted based on the number and nature of amino acids comprising the EbE4 construct. Slow sampling of at least two

conformations on the chemical shift timescale has been observed for both EbE1 (*30*) and peptides that contain residues from the heme-binding region of cyt b_5 (Davis, R.B., personal communication). However, only small stretches of amino acids were found to be conformationally exchanging in these cases. By comparison, the 100-residue EbE4 protein gave rise to more than 170 resolved peaks and an unknown number of unresolved peaks in the ^{1}H-^{15}N HSQC spectrum. This indicated that the interconversion affected the

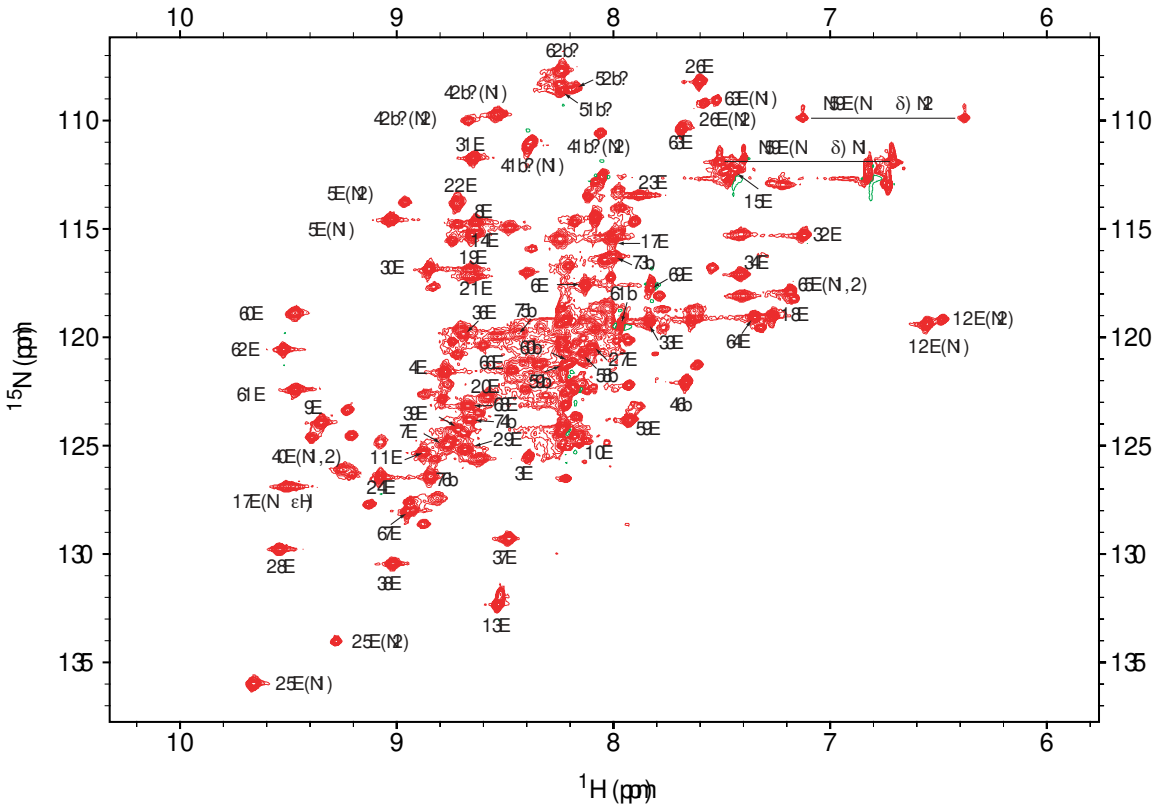


Figure 4.6. ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N-labeled EbE4 at 25 °C in 20 mM phosphate buffer (pH 7.4) and 90% $^{1}H_2O$/10% $^{2}H_2O$. Resolved resonances are labeled according to residue and species.

entire protein rather than just a small segment of it. The chemical shifts for the second set of resonances were not consistent with random coil values, suggesting an equilibrium between multiple native species (N1 and N2). By integration of multiple cross peaks in homonuclear and heteronuclear data, an approximate 3:1 ratio was obtained.

Two- and three-dimensional homonuclear and heteronuclear NMR experiments were performed on EbE4 to acquire chemical shift assignments. Owing to the ~75% population of the N1 state and its similarity to PsaE (see below), the most complete set of assignments were obtained for that form. Backbone ^{15}N, N^1H, and $C\alpha^1H$ shifts for EbE4 N1 were readily determined by comparison to PsaE and EbE1 and by tracing the chain via NH_i-$C\alpha H_i$ (COSY and TOCSY data) and $C\alpha H_i$-NH_{i+1} (NOESY data) connectivities and by comparison to PsaE (*39*) and EbE1 (*30*). Of special significance were the sequential NOEs along Tyr74b-Ile75b-Ile76b-Asn59E, corresponding to the extension of strand D with β2 residues. The fingerprint region of the EbE4 NOESY spectrum showing these connectivities is depicted in Figure 4.7. Side-chain chemical shifts were obtained primarily using correlation data to identify spin systems. Slices of several planes of the three-dimensional TOCSY data, which allowed for the ^1H assignments of residues Ile10E, Val38E, and Leu65E, are shown in Figure 4.8.

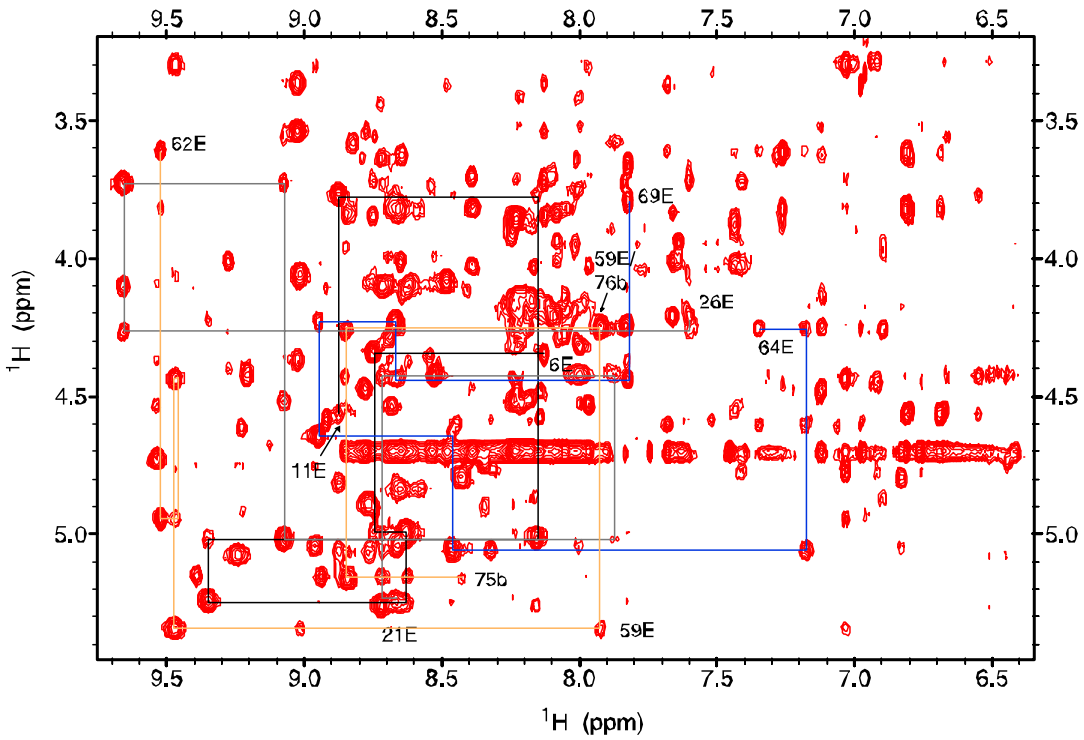


Figure 4.7. Footprint region of the NOESY spectrum of EbE4 obtained in 90% ^1H_2O / 10% ^2H_2O (20 mM phosphate buffer, pH 7.4, 25 °C, 100-ms mixing time). Many of the NOEs observed for EbE4 N1 are the same as those for EbE1 and PsaE. Lines connect cross-peaks corresponding to strands in the SH3 fold (βA: 6E-11E, black; βB: 21E-28E, gray; βD 75b-62E, orange; βE: 64E-69E, blue). Cross peaks are labeled according to residue number and protein of origin.

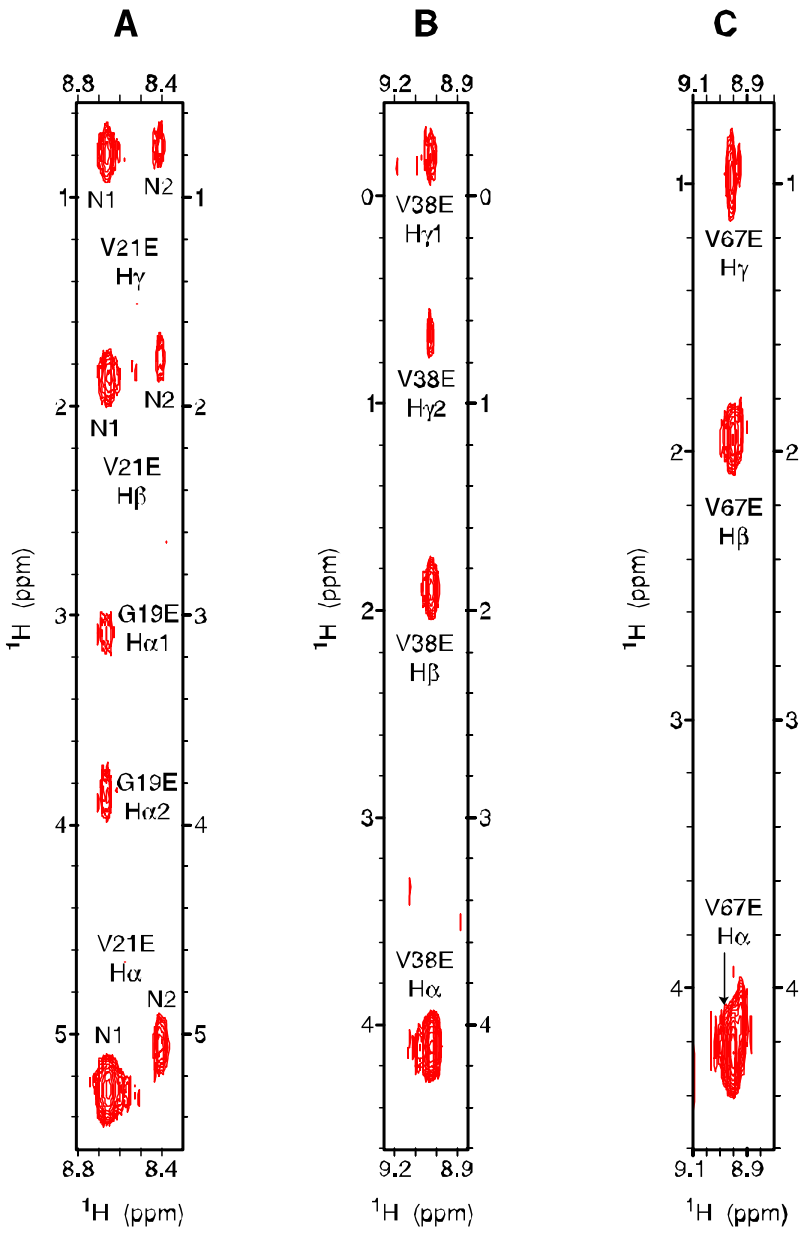


Figure 4.8. Strips in selected planes of the ^{15}N-separated TOCSY spectra of EbE4 (90% ^1H_2O / 10% ^2H_2O, 20 mM phosphate buffer, pH 7.4, 25 °C, 45-ms mixing time) that allowed for side-chain ^{1}H assignments of (A) Val21E and Gly19E (^{15}N = 117.1 ppm), (B) Val38E (^{15}N = 130.5 ppm), (C) Val67E (^{15}N = 128.2 ppm). Tentative ^{15}N assignments were obtained from the ^{1}H-^{15}N HSQC spectrum by comparison to the HSQC spectrum of PsaE and confirmed by following the backbone in the manner described for NOESY assignments.

In agreement with the far-UV CD data, the ^{1}H and ^{15}N chemical shifts obtained for EbE4 N1 suggested minor changes in the backbone of EbE4 N1 relative to that of PsaE. Differences with respect to the random shift calculated using primary structure corrections (*40*) are shown in Figure 4.9. The absolute value of the difference in backbone 15N chemical shift (Δδ) between EbE4 N1 and PsaE ranged from 0 to 0.53 ppm, with the average value being only 0.08. By comparison |Δδ| was as high as 1.63 ppm between EbE4 N1 and the random coil values.

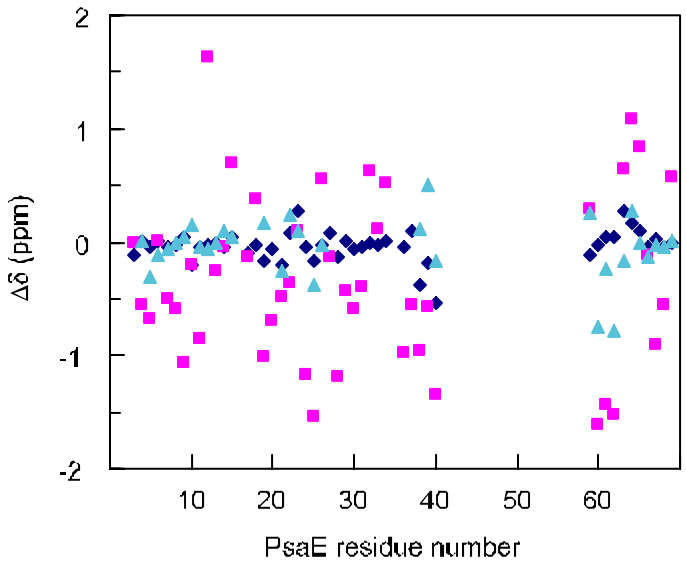


Figure 4.9. Differences in backbone ^{15}N chemical shift assignments between PsaE and EbE4 N1 (blue), random coil and EbE4 N1 (pink), and EbE4 N2 and EbE4 N1 (cyan) plotted as a function of PsaE residue number. The open region from residue 40 to 59 marks the location of the original CD loop and the grafted cytochrome b_5 loop within the SH3 scaffold. The low overall deviation between PsaE and EbE4 N1 and EbE4 N2 and EbE4 N1 supported a maintenance of the fold and the chemical environment.

After backbone chemical shifts were determined, homonuclear and ^{15}N-separated NOESY spectra were used to explore the details of the EBE4 N1 β sheet and to confirm the directionality and alignment of the strands within the sheet. A two-dimensional

layout of the SH3 fold, along with characteristic NOEs, is shown in Figure 4.10. Many of the PsaE contacts are conserved in EbE4 N1 including the CαH interactions between Ile10E and Leu65E, Val24E and Val38E, Ile37E and Asn59E, and Val67E and Lys9E. In addition, Arg39E-Ile75b and Asp28b-Thr73b CαH-CαH contacts demonstrated the docking of cyt b_5 strand β2 on PsaE strand C, indicating that the β2/β3 pairing native to

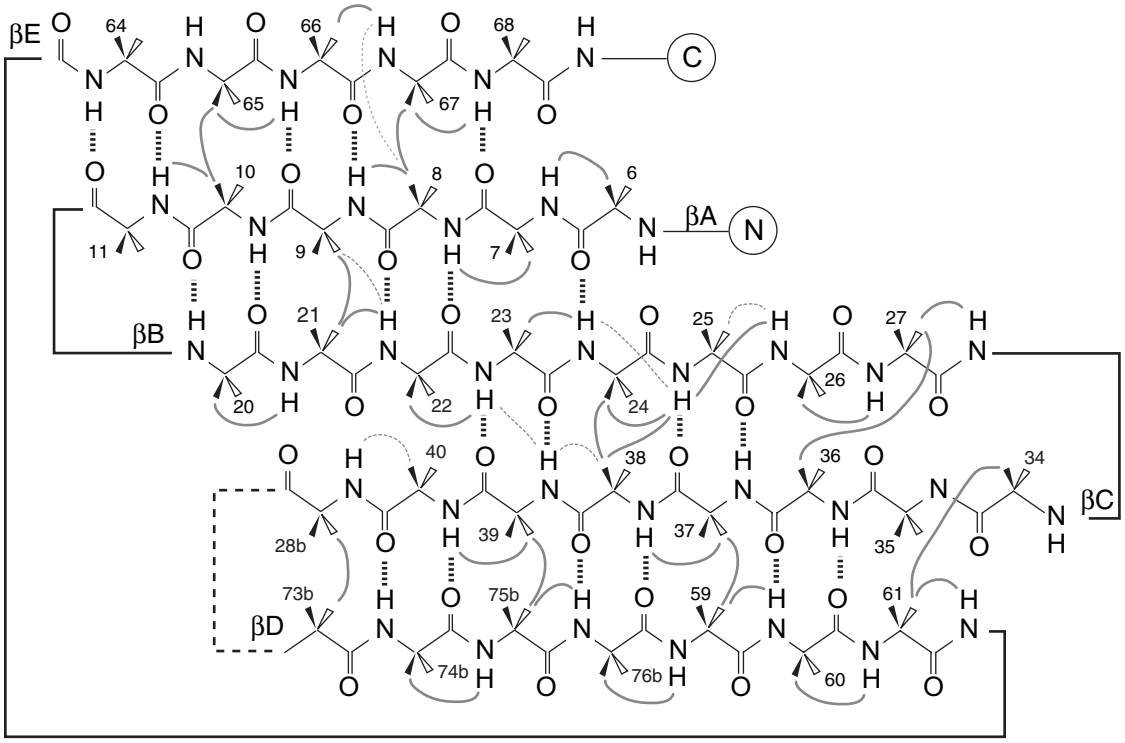


Figure 4.10. The β sheet layout of PsaE, with residues from EbE4. Extended pairing of the modified strands C and D is shown. An alternative pairing modeled after the cytochrome structure would have H-bond interactions between Asp31b NH (in β3) and Tyr74b CO (in β2), resulting in a different structure. Hydrogen bonds are marked with dashed lines. Some of the key NOEs defining the topology of the major native state are indicated with lines (strong: ——; weak: - - - -).

the cytochrome loop had been disrupted. Thus, although half of the PsaE D strand had been removed in the creation of the EbE4 protein, the NOE cross peaks mentioned above confirmed that the C-terminal portion of the protein had been able to complete the SH3 fold of PsaE.

In agreement with the far-UV CD data, the ^{1}H and ^{15}N chemical shifts obtained for EbE4 N1 suggested minor changes in the environment for the backbone relative to that of PsaE. The main difference occurred at the level of the docking of the C-terminal end of strand C with the N-terminal end of strand B. In the PsaE structure, Phe40E is stacked against Trp17E, in an interaction that anchors the AB loop onto the end of strand C. In EbE4, this structural feature appeared to be lost. Instead, Trp17E was found in contact with Tyr74b. This residue of the cytochrome's β2 strand replaces Asn56E in the PsaE structure, a residue that points in the opposite direction. Further evidence for an extension of the β structure was provided by NOEs between Val29b (occupying the position of Lys42E) and Tyr74b. By comparison, the side chains of Asn56 and Lys42 are not in contact in the wild-type PsaE. The distortion of the hydrophobic core was manifested in the large chemical shift differences observed for Trp17E and Phe40E and the contacts between Phe40E and Val29b.

EbE4 contains two prolines, one of which (Pro35E) is found in the BC loop immediately before the C strand. The Tyr34-Pro35 peptide bond in PsaE adopts the *trans* configuration, as shown by ^{1}H-^{1}H NOEs and ^{13}C shifts (*37, 39*). In addition to close dipolar contact between Pro35 CδH_2 and Tyr34 CαH, strong effects were also observed between Pro35 CδH_2 and Ala61 CαH. The same ^{1}H-^{1}H NOEs were detected in the N1 state of EbE4, which supported a *trans* configuration for the Tyr34E-Pro35E peptide

bond as well as a conserved supersecondary structure with the N-terminal end of the C strand positioned correctly with respect to the C-terminal end of the D strand.

In contrast to the folded core of N1, most of the cytochrome loop presented chemical shift degeneracies that could not be resolved with ^{15}N labeling alone. By elimination, it was possible to assign tentatively residues 57b through 63b, a segment of particular interest because it encompasses one of the heme-ligating histidines in the cytochrome (His63) and exhibits conformational multiplicity in EbE1 (*30*). The backbone chemical shift differences in this region were small when calculated both by reference to the random coil value (*40*) and by reference to apocyt b_5 (*41*). Evidence for sampling a turn, however, was noted through weak NH_i–NH_{i+1} NOEs in apocyt b_5 and EbE1. The $\{^1H\}$-^{15}N NOE values for the few loop backbone NHs that could be assigned and were sufficiently resolved were systematically smaller than the 0.8 NOE value observed for residues in the folded core of the protein (data not shown). EbE4's second proline (Pro40b) originates from cyt b_5 and resides in the inserted loop; its assignment and the assignment of the rest of the loop will require ^{13}C data.

Changes were also apparent in the spectrum of the N2 state of EbE4; for the few resonances that were assigned with certainty, the difference in chemical shifts between N1 and N2 was occasionally large ($|\Delta\delta|(^1H) > 0.1$ ppm; $|\Delta\delta|(^{15}N) > 1$ ppm, Figure 4.9), but N2 was a folded species exhibiting $C\alpha H_i$–NH_{i+1} NOEs expected of β structure. $\{^1H\}$-^{15}N NOE data confirmed that both N1 and N2 had PsaE core residues tumbling in the slow motion limit (not shown). The difference between N1 and N2 appeared to be further disruption of the hydrophobic core. The isopropyl group of Val38E, which is normally shifted upfield because of contact with Phe40E and Phe60E, experienced a

reduced ring current shift. The shifts of Ile10E, which also participates in this core, also suggested diminished influence of the rings and an altered geometry. The ring protons of Trp17E yielded sharp lines at random shift values and did not display NOEs to the rest of N2, which further confirmed the disruption of the core. In addition, the configuration of Tyr34E-Pro35E peptide bond could not be established; rearrangement in this region of the protein was emphasized by the inability to assign N2 residues in the 28-36 stretch and the large $\Delta\delta$ in the C-terminal part of the scaffold, in particular at position 61E. The NOEs corresponding to Pro35 CδH_2–Ala61 CαH were not identified.

4.2.3. Thermodynamic Characterization

In light of the finding that the EbE4 protein sampled multiple native conformations (N1 and N2), ^{1}H-^{15}N HSQC data sets were collected between 5 and 40 °C to determine the relative populations of these states as a function of temperature. The resulting spectra are depicted in Figure 4.11. The changes in the relative intensities of the two sets of peaks as well as the chemical shifts of these resonances are evident and reflect the hydrogen-bonding status of the amides and structural changes (*42*). Throughout the temperature range, the interconversion rate remained slow on the chemical shift time scale. At room temperature, the absence of exchange cross peaks in TOCSY and NOESY data confirmed the observed rate constant to be well below 30 s^{-1}. At 5 °C, the protein populated the N1 state almost exclusively; by 40 °C, the N1 and N2 populations were nearly equal. In order to determine the energy involved in the transition between N1 and N2, peaks corresponding to these two states were integrated to obtain equilibrium constants and $\Delta G°$ values; the results are shown in Figure 4.12. The free energy ranged

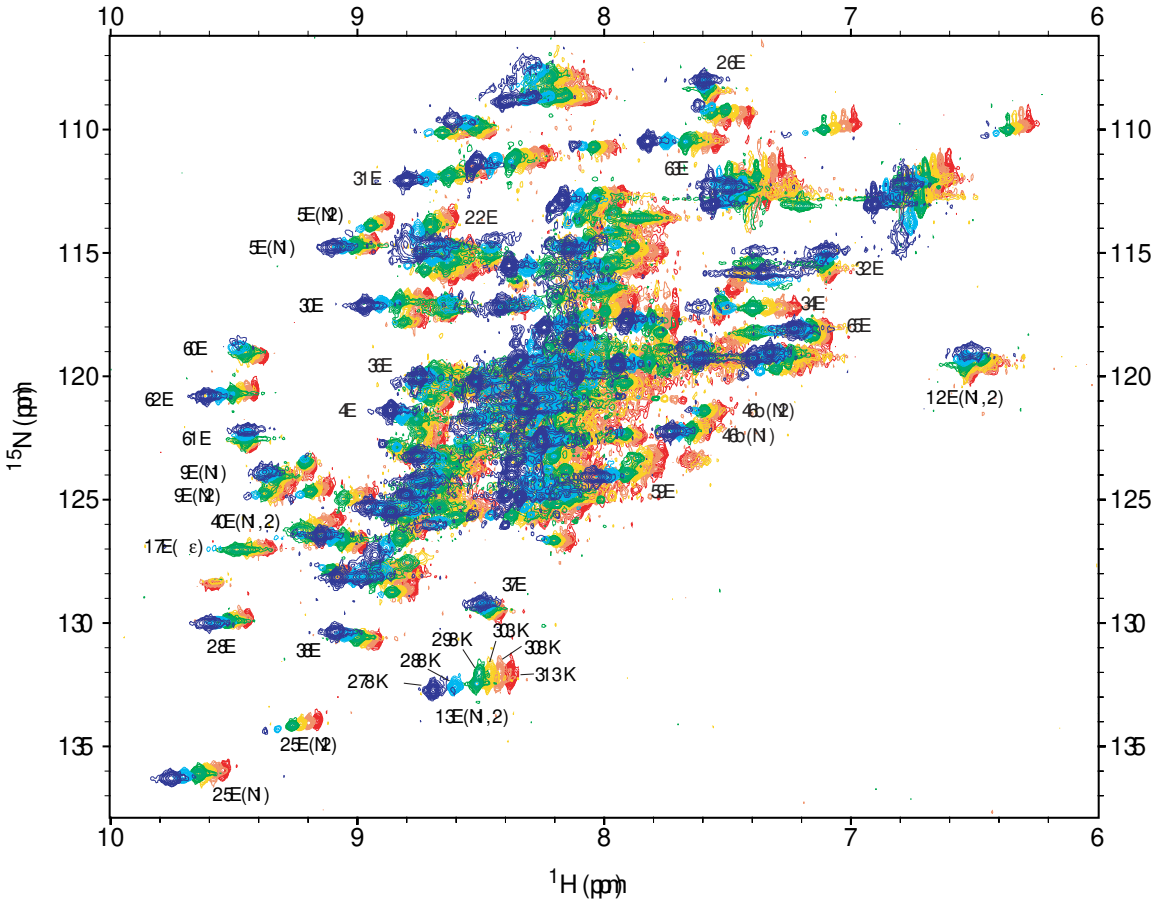


Figure 4.11. ^{1}H-^{15}N HSQC spectra of uniformly ^{15}N-labeled EbE4 (20 mM phosphate buffer, pH 7.4, 90% $^{1}H_2O$ / 10% $^{2}H_2O$) as a function of temperature. Blue: 278 K; cyan: 288 K; green: 298 K; gold: 303 K; salmon: 308 K; and red: 313 K. Resolved resonances are labeled according to residue number and species.

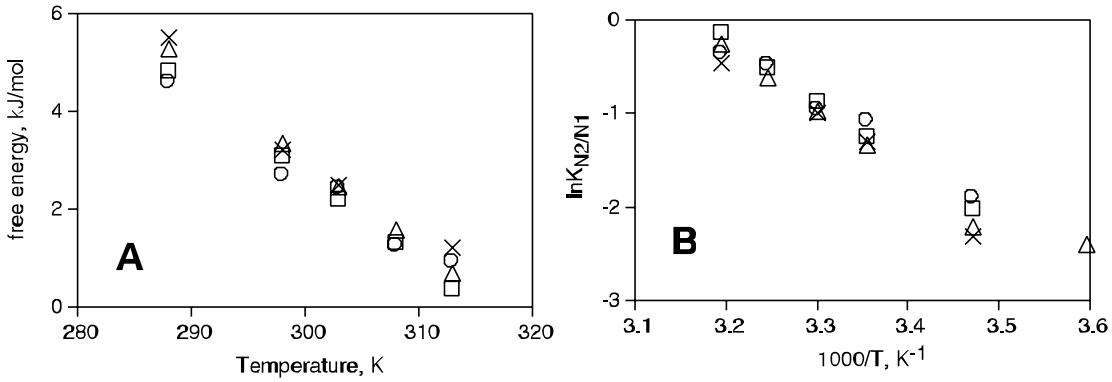


Figure 4.12. (A) Free energy involved in the N1, N2 interconversion as a function of temperature. (B) van't Hoff analysis of the equilibrium constant $K_{N2/N1}$ as a function of temperature, yielding an average a ΔH° of ~45 ± 5 kJ/mol. Values shown are for residues 5E (○), 24E (×), 63E (□), and 46b (Δ).

from ~5 kJ/mol at 5 °C to ~1 kJ/mol at 40 °C. In addition, at 25 °C, the temperature used for the majority of the experiments presented here, the fractional populations of N1 and N2 were ~0.75 and ~0.25, respectively. van't Hoff analysis returned a ΔH° of ~45 kJ/mol associated with the N1 → N2 conversion, indicating that the rearrangement involved several interactions. The extrapolated T_m for the N1-N2 transition was near 43 °C.

In an investigation of the effect the nature of the loop anchor on the ability of the PsaE scaffold to resist temperature-induced unfolding, the thermal denaturations of EbE2 and EbE4 were followed by far-UV CD spectroscopy and absorbance spectrophotometry (Figure 4.13A). EbE3 aggregated even at μM concentrations, and, therefore, was not subjected to thermodynamic characterization. The midpoint of the EbE2 and EbE4 curves were 52.6 and 57.9 °C, respectively. The insertion of a single Gly residue at the N terminus of the cyt b_5 loop did not change T_m. By comparison, EbE4 displayed a midpoint that was 5.4 °C higher than EbE1 and only 2.9 °C lower than PsaE, suggesting

that altering the amino acids at the interface of the PsaE and cyt b_5 sequences had stabilized the protein relative to EbE1. However, the most striking feature of the EbE4 denaturation was its complete and reproducible lack of reversibility. In an alternate experiment, a solution of EbE4 was heated to 57 °C, and the absorbance at 295 nm was monitored as a function of time (Figure 4.13B). The absorbance value decayed to that of the unfolded state with a complex set of time constants. These data indicated that the thermally denatured protein aggregated irreversibly. This precluded the determination of EbE4 thermodynamic parameters.

In order to inspect the free energy of folding, EbE4 was subjected to urea-induced denaturation. Urea titration experiments were followed by far-UV CD and absorbance spectroscopy, and the results are shown in Figure 4.14A. Unlike the thermal unfolding reaction, the chemical unfolding was completely reversible. Although the data for the two methods were non-coincident, each individual trace could be fit using a two-state model. The CD data yielded a $\Delta G°$ of unfolding of 19.7 ± 1.2 kJ/mol, with an *m* value of 3.1 ± 1.2 kJ/mol/M, and a C_M of 6.3 M urea; fitting of UV data resulted in a $\Delta G°$ of unfolding of 21.6 ± 0.8 kJ/mol, with an *m* value of 3.7 ± 0.1 kJ/mol/M, and a C_m of 5.8 M urea. Although these $\Delta G°_U$'s were within error of each other, C_m and *m* values (Table 4.1) distinguished the two denaturation curves and indicated the presence of a third equilibrium state. The most likely explanation is that the N1 and N2 states possess slightly different spectroscopic properties, and the two methods were reporting on the presence of the N1, N2, and U states throughout the unfolding process. Regardless of the

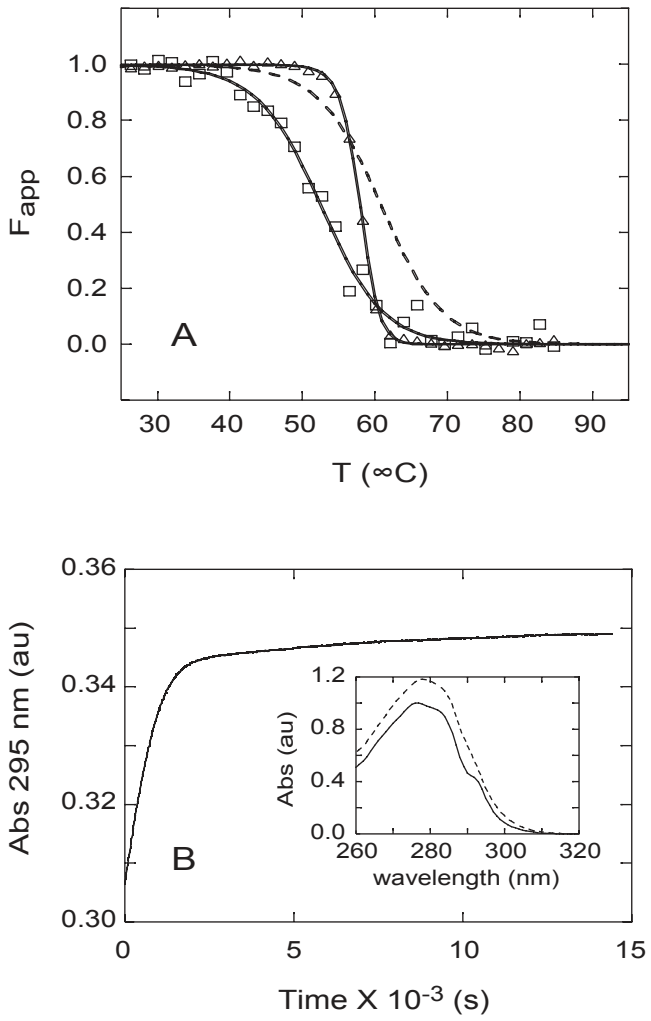


Figure 4.13. (A) Thermal denaturation of EbE2 (□) and EbE4 (△) in 20 mM phosphate buffer (pH 7.4) monitored by CD and UV-Vis spectrometry. The solid line represents the best fit of the data to a two-state model. Parameters are listed in Table 2. For comparison purposes, the curve of PsaE (- - -) is also shown. The EbE1 curve is coincident with the EbE2 curve and has been omitted for clarity. (B) Absorbance at 295 nm as a function of time for a sample of EbE4 that was incubated at 57 °C after being heated from 25 °C to 57 °C in steps of 2 °C with a 5-min equilibration and collection of a 320 nm to 260 nm spectrum at each temperature. Inset: native and thermally-denatured spectra of EbE4.

method used, the free energies obtained for EbE4 represented a marked increased in the stability of the protein with respect to EbE1 ($\Delta\Delta G° \sim 9$ kJ/mol) and PsaE ($\Delta\Delta G° \sim 5$ kJ/mol).

EbE4 was further subjected to urea-induced denaturation at a similar protein concentration as that used for the other equilibrium experiments within the confines of an acrylamide gel (Figure 4.14B) to determine if the free energy of unfolding was artificially inflated by a multimerization event (i.e., N2 = $(N1)_n$). For a size reference, apocytochrome b_5, which contains the same number of residues and the same large loop,

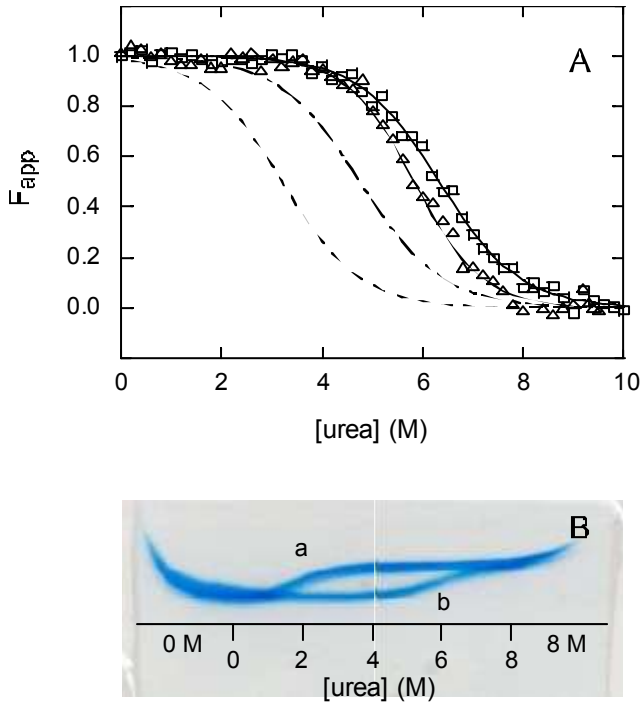


Figure 4.14. (A) Urea-induced denaturation of EbE4 (20 mM phosphate buffer, pH 7.4) monitored by CD (□) and UV-Vis spectroscopy (△). The solid line represents the best fit of the data to a two-state model. Parameters are listed in Table 2. The non-coincidence of the EbE4 traces indicated a minimum of three states. For comparison purposes, the curves for EbE1 (- - -) and PsaE (– — –) are also shown. (B) Urea-gradient gel electrophoresis of apocytochrome b_5 (a) and EbE4 (b) loaded in the native state. The baselines represent folded and unfolded protein; overlap in these regions indicated that EbE4 was a monomer and that its N1 and N2 species were characterized by a similar size and shape. The scale portrays the approximate urea concentrations throughout the gel.

was also run on the gel. The midpoint of the transition in the gel was consistent with that obtained via solution experiments. Most importantly, the protein that gave rise to the native baseline of the EbE4 curve had the same electrophoretic mobility as apocytochrome b_5, which is known to be a monomer. Although the lack of coincidence of the CD and UV titration data was not consistent with two-state behavior, the EbE4 transition in the gel showed no obvious sign of a folding intermediate. This along with the single band observed for the native baseline indicated that the overall size and shape of the N1 and N2 states were similar.

^{1}H-^{15}N HSQC spectra were also collected in the presence of increasing amounts of urea (not shown). At concentrations corresponding to the baseline of the denaturation curves shown in Figure 7A, $K_{N1\text{-}N2}$ remained practically constant and the exchange process was slow on the chemical shift time scale. At the onset of the transition (4.7 M), the population of N2 started to increase relative to that of N1, and unfolded signals became evident. The chemical shifts of the HSQC cross peaks exhibited variable responses to the increase in urea. Most did not shift by a large amount, but a few did, indicating the presence of a complex energy landscape even under conditions that were native according to the optical data.

4.2.4. Folding and Unfolding Kinetics

Despite the appearance of EbE4 in the urea-gradient gel, it became clear during the course of control experiments for the urea titration studies that EbE4 exhibited altered folding kinetics, including a phase extending over minutes. As a result, the apparent rates of EbE4 folding and unfolding as a function of urea concentration were investigated with a focus on slow processes. CD data were collected at 234 nm as a function of time.

Table 4.1. Thermodynamic parameters for the denaturation of various target polypeptides

	T_m [a]	$\Delta H°$ [b]	$\Delta C°_p$	$\Delta G°_U$ [c]	m	C_m [d]
	°C	kJ mol^{-1}	kJ mol^{-1} K^{-1}	kJ mol^{-1}	kJ mol^{-1} M^{-1}	M
PsaE	60.8 ± 0.2	216 ± 10	2.5 ± 0.1	14.47 ± 0.03	3.04 ± 0.10	4.7
apocyt b_5	46.0 ± 0.6	142 ± 6	4.2 (fixed)	6.4 ± 0.4	–	–
EbE1	52.5 ± 0.3	196 ± 18	2.5 ± 0.8	10.27 ± 0.14	3.2 ± 1.0	3.2
EbE2	52.5 ± 0.5	202 ± 15	2.5 (fixed)			
EbE4 (CD)	57.9 ± 0.2	–	–	19.7 ± 1.2	3.1 ± 0.2	6.3
EbE4 (UV)	57.9 ± 0.2	–	–	21.6 ± 0.8	3.7 ± 0.1	5.8

[a] Midpoint of thermal denaturation in 20 mM phosphate buffer at pH 7.4. All errors are standard deviations from global fitting of multiple data sets. [b] Change in enthalpy evaluated at the midpoint of the thermal unfolding transition. [c] Free energy of unfolding at 25 °C in 20 mM phosphate buffer at pH 7.4. [d] Midpoint of urea denaturation at 25 °C in 20 mM phosphate buffer at pH 7.4, calculated from ΔG° and m.

Native EbE4 was manually diluted to final urea concentrations ranging from 8.4 to 9.7 M; alternately, EbE4 in 9.12 M urea was diluted to final conditions between 0.4 and 3.6 M urea. Representative temporal profiles are shown in Figure 4.15A. Time constants (Figure 4.15B) were obtained by fitting of the data to the sum of a constant and a single exponential.

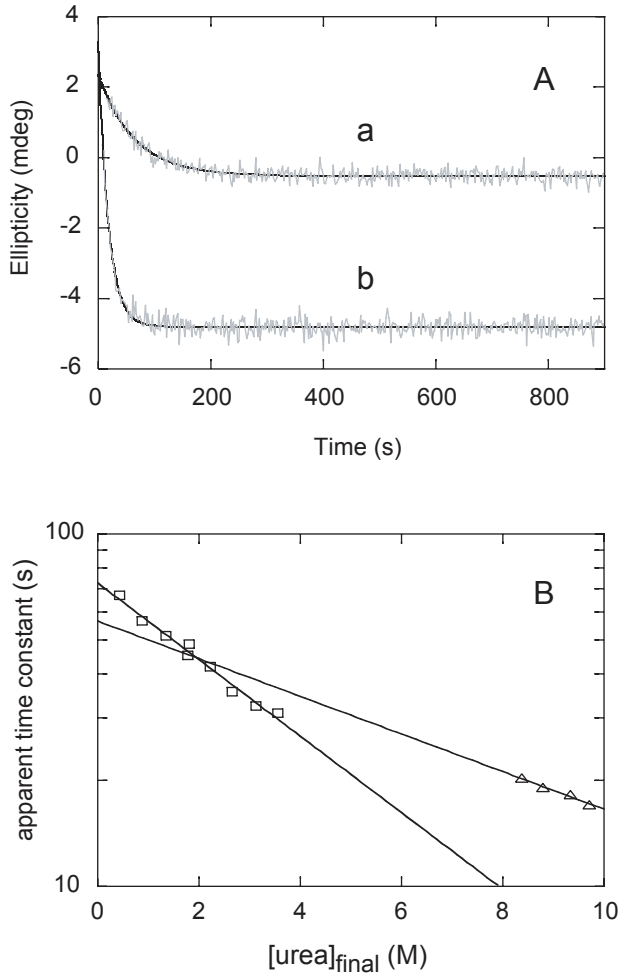


Figure 4.15. (A) Temporal profiles of 10 μM EbE4 in 20 mM phosphate buffer (pH 7.4) refolding from 9.12 M to 0.44 M urea (a) or unfolding from 0 M to 9.7 M urea (b). The solid lines are the result of a fit of the data to the sum of a constant and a single exponential function. (B) Unfolding (△) and refolding (□) relaxation times obtained from exponential fits of the unfolding and refolding data as a function of final urea concentration. Solid lines indicate extrapolation back to 0 M urea.

The kinetic data exhibited three unusual properties. First, both the unfolding and refolding traces depicted decay from a state with positive ellipticity at 234 nm; neither the native nor denatured spectra of EbE4 display this property. Rather, the native spectrum displayed zero ellipticity, and the denatured protein gave rise to negative values. Second, extrapolation of the apparent time constants back to 0 M urea yielded apparent τ_{fold} and

τ_{unfold} values of 72.8 s and 56.8 s, respectively. The ratio of these values returned an apparent free energy of unfolding of approximately −0.6 kJ/mol, which was inconsistent with equilibrium results. Finally, the refolding time constant decreased with moderate increases in urea concentration. However, each of these properties can be explained when the N1, N2 interconversion is considered.

The relative volumes of resonances corresponding to the two native forms in the ^{1}H-^{15}N HSQC spectra indicated that the ratio of molecules populating the N1 and N2 states was ~3:1 at room temperature. Since the multi-dimensional NMR data supported little change in the overall protein fold upon going from EbE1 to EbE4 N1, the assumption could be made that the two proteins have similar far-UV CD spectra. Using the relationship $\Theta_{obs} = 0.75\ \Theta_{N1} + 0.25\ \Theta_{N2}$, a theoretical CD spectrum for EbE4 N2 was calculated (Figure 4.16), which contained a region of positive ellipticity that included 234 nm, the observation wavelength for the kinetics experiments. At this wavelength, the native and denatured states of PsaE display near zero and negative ellipticity, respectively. This supported that the transitions reported on in both the refolding and unfolding experiments were originating from the N2 state. The reason for the observation of the refolding to begin with N2 was clear upon inspection of the apparent unfolding time constant. Based on this value and the $\Delta G°_{unf}$ obtained from equilibrium experiments, the refolding time constant was expected to be on the order of 0.02 s, which is significantly shorter than the dead time of the manual mixing experiment and the limits of N1 and N2 life time determined by NMR spectroscopy. The two distinct sets of resonances for N1 and N2 in the NMR data demonstrated that the two states were in slow exchange on the chemical shift timescale. Rather than folding of the protein, the

observed refolding kinetic trace depicted the interconversion between the N2 and N1 states, thus explaining the disagreement between the equilibrium constant obtained from the titration experiments and that derived from the ratio of the apparent time constants extrapolated back to 0 M urea. The single exponential behavior supported that, under native conditions, rapid and complete formation of N2 occurred during the dead time of the experiment

A decrease in the apparent refolding time constant with increasing urea concentration is most often seen for proteins that populate an off-pathway intermediate; however, this can, under certain circumstances, be consistent with an on-pathway intermediate (*43*). The variable-temperature ^{1}H-^{15}N HSQC data indicated that under native conditions, the N1 ⇆ N2 step involved a small apparent ΔG° and a high activation energy barrier. EbE4 N2 became increasingly populated at increasing temperatures conditions suggesting that

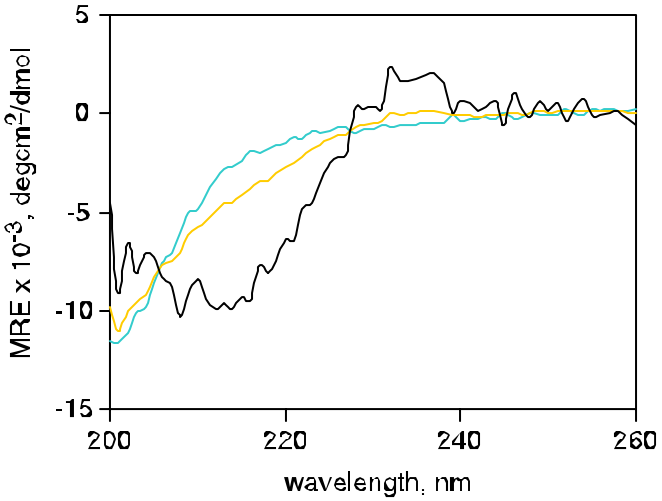


Figure 4.16. Theoretical far-UV CD spectrum of the EbE4 N2 state (black) calculated in the manner described in the text. For comparison purposes the native spectra of EbE1 (aqua) and EbE4 (gold) are also shown.

N2 was a state between N1 and U on the unfolding pathway, or at least that this route competed efficiently with one taking N1 directly to U. The acceleration of the folding reaction with increasing and moderate urea concentrations indicated that the height of this barrier, or the energy of the transition state, was highly susceptible to urea. Since this phenomenon was observed at low urea concentrations, it could be assumed that, under strongly denaturing conditions, the barrier all but disappeared, and the interconversion between N1 and N2 was rapid. The unfolding kinetic traces were obtained by subjecting native protein to urea concentrations between 8.3 and 10.1 M. Under these conditions, the N1 to N2 transition occurred during the dead time of the experiment, and the observed process was the slow unfolding of N2.

In a further investigation of the kinetic behavior of EbE4, refolding and unfolding experiments were performed in which the final urea concentration was held constant and the initial conditions were varied. Kinetic profiles corresponding to refolding jumps into 0.7 M urea from 4.5-7.9 M urea and unfolding jumps into 9.1 M urea from 2.0-7.5 M urea were collected at 234 nm; representative traces are depicted in Figure 4.17A. The data were again fit to the sum of a constant and a single exponential. Global fitting yielded a τ_{fold} of 52.2 ± 0.5 s at 0.7 M urea and a τ_{unfold} of 15.02 ± 0.13 s at 9.1 M urea. These values agreed well with those obtained from experiments outlined above in which the final urea concentration was varied. The pre-exponential amplitudes as a function of initial urea concentration are shown in Figure 4.17B (refolding) and Figure 4.17C (unfolding). The shapes and midpoints of these profiles are similar to those obtained via the CD equilibrium urea titration experiment (Figure 4.14). Fitting of these data to a two-state equilibrium unfolding model was not possible owing to the lack of baselines, but the

points did lie on theoretical curves created using the parameters obtained from fitting the equilibrium CD data.

Finally, EbE4 was subjected to a series of double-jump refolding experiments. Native protein was exposed to ~9 M urea, and, after delay times ranging from 10 to 900 s, the protein was diluted into buffer, with a final urea concentration of ~0.35 M. The data were fit in the same manner as the other kinetic data. The major exponential of the profile of the pre-exponential amplitudes as a function of delay time (Figure 4.18) yielded a time constant of 14.5 ± 1.8 s; thus the apparent time required for the recovery of amplitude was within error of that observed for unfolding experiments at the urea concentration to which the protein was subjected during the time delay. The time constants obtained from global fitting of the individual refolding profiles were approximately 60 s, and agreed well with other data.

4.2.5. Reconstitution of EbE2-4

As mentioned in Chapter 3, one goal of creating the EbE proteins was to determine whether or not the cytochrome b_5 heme-binding region was capable of participating in specific protein-cofactor interactions in the absence of its native scaffold. The inability of EbE1 to bind specifically the physiologically relevant heme (Fe-PPIX) or a water-soluble derivative (Fe-TCP) may have been a result of conformational strain at the interface of the cytochrome b_5 and PsaE sequences that would not allow both regions of the protein to fold simultaneously. EbE2 and EbE3 contain glycine insertions on the N or N and C end(s) of the loop, and were designed to alleviate such strain if it existed. UV-Vis absorbance spectra of the EbE proteins incubated with Fe-PPIX and Fe-TCP are

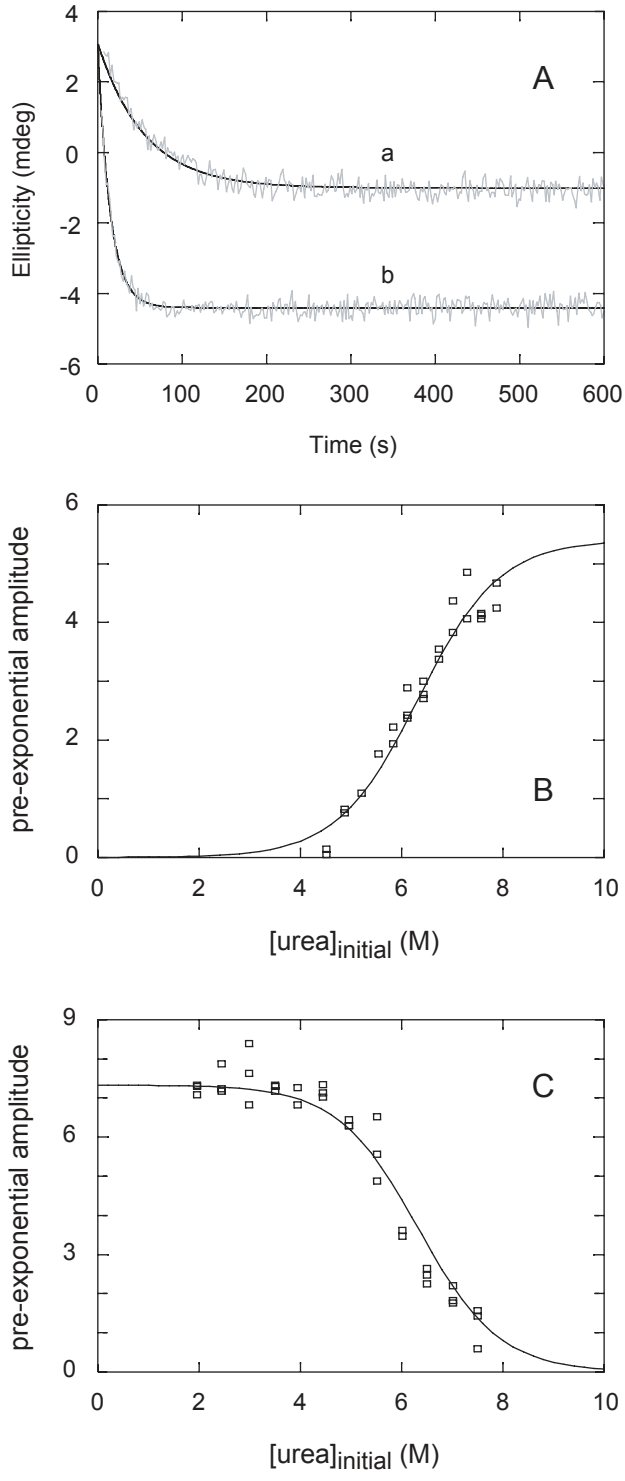


Figure 4.17. legend on following page

Figure 4.17. (A) Representative temporal profiles of 10 mM EbE4 (20 mM phosphate buffer, pH 7.4) refolding from 7.0 to 0.7 M urea (a) or unfolding from 2.0 to 9.1 M urea (b). The solid lines represent the fit of the data to the sum of a constant and a single exponential function. (B) Pre-exponential amplitudes obtained for kinetic traces of protein being refolded into a final urea concentration of 0.7 M as a function of initial urea concentration. (C) Pre-exponential amplitudes obtained for kinetic traces of protein being unfolded into a final urea concentration of 9.1 M as a function of initial urea concentration. Solid lines in (B) and (C) represent theoretical curves created using the parameters listed in Table 2.

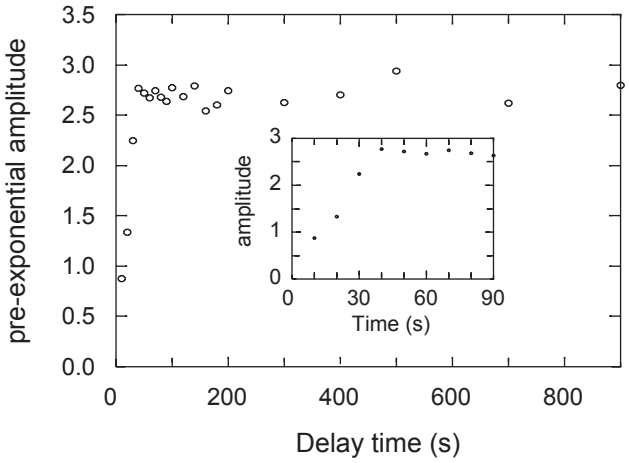


Figure 4.18. Pre-exponential amplitudes of the exponential function describing the signal obtained for double-jump refolding experiments as a function of delay time. Native protein was introduced to 9.1 M urea then diluted to final urea concentration of 0.7 M after delay times ranging from 10 to 900 s. Solutions contained 20 mM phosphate buffer, pH 7.4 under all conditions. An enlargement of the amplitudes for traces corresponding to delay times of 10 to 90 s is inlaid.

shown in Figure 4.19; for comparison, the spectra of free heme and holocytochrome b_5 are also shown. The data do not support specific heme binding. Rather, they indicate that the cytochrome region chosen in the creation of these proteins was still unable to undergo the apo- to holoprotein conversion characteristic of the wild-type protein.

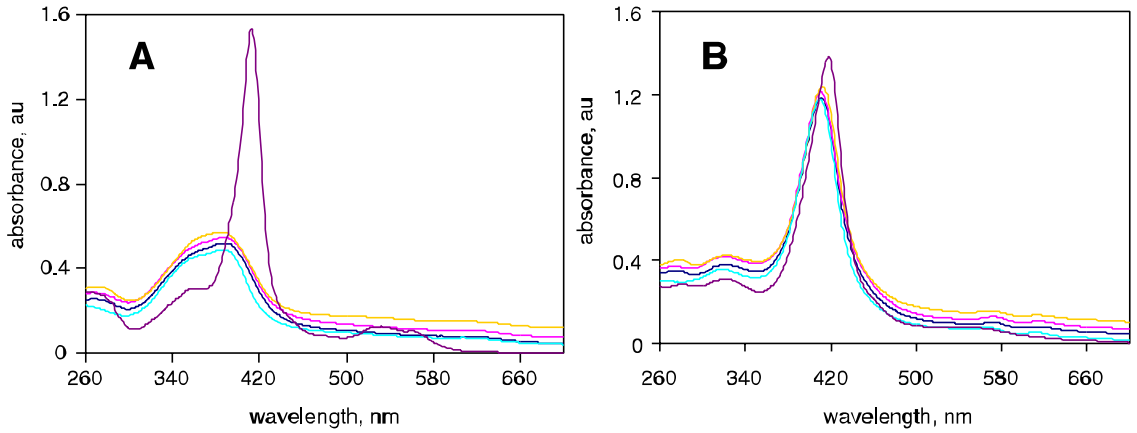


Figure 4.19. (A) Absorbance spectra of EbE2 (blue), EbE3 (pink), and EbE4 (gold) incubated with Fe-PPIX. (B) Absorbance spectra of EbE2 (blue), EbE3 (pink), and EbE4 (gold) incubated with Fe-TCP. For both figures, the spectra of the free porphyrin (cyan) and holocytochrome b_5 (purple) are shown for comparison. Constants were added to offset the EbE2-4 traces and ease viewing. The data do not support specific heme binding.

4.3. Discussion

The effect of loop insertion or elongation has been studied in a variety of protein backgrounds. Stretches of unstructured amino acids have the capacity to destabilize the scaffold to which they are attached, and, in some cases, they do so in a manner that is predictable with loop entropy calculations (for an analysis of this problem, see (*44*)). This behavior, however, is far from general, and the effect of an insertion or deletion is often difficult to anticipate. For example, insertion of a constant loop at different sites in a protein may lead to different energetic contribution, illustrating the importance of the anchor point. This is confirmed by the present study, which investigated the consequences of single Gly insertions and restructuring of the loop-scaffold interface on the biophysical properties of an SH3-like fold.

4.3.1. Structural Effects

The CD and NMR data obtained for the EbE2-4 proteins indicated that the insertion of the cyt b_5 heme-binding region into the folded PsaE core had little effect on the topology of the protein. Differences between the spectra obtained for EbE4 N1 and those obtained for EbE1 implied a restructuring of interactions in the vicinity of the loop termini, including inter-aromatic contacts. EbE4 N2 differed further, but still maintained the SH3 fold, albeit separated from N1 by a large activation energy barrier. The appellation "N2" for the second form was preferred over "I" because N2 had well-defined secondary and tertiary features, existed in significant concentration under native conditions, and exhibited an extended native baseline with respect to urea denaturation. Thus, a major consequence of the modification of EbE1 into EbE4 was the competitive population of original and alternative scaffold conformations.

Slow interconversion between two states is characteristic of Xxx-Pro peptide bond isomerization, a phenomenon that is a well-known source of native state heterogeneity in proteins. EbE4 contains two prolines, one of which is located in the SH3 scaffold and may affect its structure. The urea-independent equilibrium constant in the native baseline was consistent with a *cis*-*trans* isomerization. The transition state dependence, however, was not, as it appeared to entail considerable exposure of the protein to solvent. Several other arguments can be articulated against a *cis*-*trans* isomerization being solely responsible for the slow exchange. For example, EbE1 contains the same prolines and does not exhibit this phenomenon. Thus, stabilization of *cis* Xxx-Pro bond(s) in N2 would have to originate from the changed anchor points and presumably involve modified interactions relayed from those locations. It is also unlikely that the *trans*

Tyr34E-Pro35E bond would require isomerization to the *cis* form prior to unfolding to the U state. ^{13}C labeling and production of a Pro35E variant of EbE4 may address this aspect of N2. Whether or not the N1-N2 equilibrium was connected to the state of the Tyr34E-Pro35E bond, N2 appeared to involve a rearrangement of the β structure in the C-terminal region. Large differences in chemical shifts extending far beyond the putative isomerization site, as required to account for the doubling of nearly every peak in the ^{1}H-^{15}N HSQC spectrum, could be due in part to alterations of the hydrophobic core (and therefore changes in the ring current contribution) and partial strand distortion.

Most naturally-occurring globular proteins populate a single native state endowed with the required functional properties. An exception of interest is dihydrofolate reductase (DHFR), which, in the absence of substrate and cofactor, partitions into two states (N1 or E1 and N2 or E2) (*45*). The two native states exchange slowly on the NMR chemical shift time scale, resulting in a doubling of resonances across the whole protein (*46*). The structural differences between N1 and N2 are minor and involve a perturbation of the hydrophobic core; only one is capable of binding substrate. In spite of this, the population of both native states by DHFRs from various species (*47*) suggests a physiological role for the alternative form. The folding mechanism of *Escherichia coli* DHFR has been studied extensively (*48-50*), and it has been established that both N1 and N2 can be directly populated from the unfolded state.

Slow exchange between multiple native conformations as a result of loop alteration has been observed for the apoflavodoxin from *Anabaena* PCC 7119 (*51*). In this case, the observation of a second native state was dependent upon the presence of a 20-residue loop. Slow internal reorganization was also found for a variant of ubiquitin that had been

engineered to produce a more structured transition state (*52*). A more radical example of conformational equilibrium is provided by a chimeric protein containing ubiquitin inserted into a surface loop of barnase such that only one of the domains can fold at a time (*53*). Temperature and ligand binding dramatically affect the relative populations of the two folded states, one containing folded ubiquitin and an unfolded barnase, the other containing folded barnase and unfolded ubiquitin.

The ability of a polypeptide chain to reverse its direction and form a stable turn is a necessary characteristic of a globular protein. Loop insertion studies have tested the robustness of folded cores and revealed that most proteins behave in their own fashion because the location as well as the nature of the loop contributes to the behavior of the whole. Here we have shown with EbE1 and EbE2 that if the attachment points are not modified, consistent thermodynamic data can be obtained that presumably reflect loop properties, such as its propensity to form residual structure. In contrast, thermodynamic data obtained on constructs with modified attachment points reflect loop and scaffold properties. In EbE4, N1 and N2 can be considered to represent an initial step in the conversion of one protein topology into another. Although the change is not as drastic as that observed for the barnase-ubiquitin chimera, the EbE4 protein demonstrates that alteration of just a few loop-anchoring residues is enough to induce native state heterogeneity.

4.3.2. Thermodynamic Effects

It was clear that the nature of the loop anchor played a large role in dictating the thermodynamic stability of the protein. The choice of just a few residues at the N- and C-terminal sides of the cyt b_5 loop altered the midpoints of both the temperature- and urea-

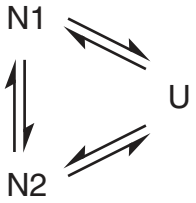
induced denaturations of the hybrid proteins. In addition, the loop termini affected the free energy of unfolding by as much as ~10 kJ/mol. The source of this stabilization is unclear, although it is possible that some structure was formed in the EbE4 loop, the alleviation of strain present at the EbE1 junction of the cyt b_5 and PsaE sequences, efficient pairing between PsaE strand C and the cytochrome's strand β2, or any combination of these factors. Strain relief appeared more likely than structure formation in the loop as this structure would have to participate in the cooperative core of the protein to alter the measured $\Delta G°_{unf}$. In addition, it is well documented that restructuring of the hydrophobic core can impact dramatically the thermodynamic stability of a folded protein (*54-58*). In the EbE4 case, evidence for modest reorganization relative to the PsaE scaffold was found in the NMR data.

Minor loop modifications also altered the reversibility of the thermal unfolding reaction leading to aggregation of the protein. Protein aggregation is difficult to predict, and the fact that the insertion of a single residue was able to induce this behavior is remarkable, especially when the similarities of the NMR spectra are considered. Algorithms such as TANGO (*59, 60*) are available to predict the propensity of amino acid sequences to form aggregates of both α-helical and β-sheet nature. Calculation of the aggregation propensities of EbE1-4 returned negligible population of this state by EbE1-3 and a high likelihood for EbE4. The EbE4 results were consistent with the inability of the thermally unfolded protein to renature upon cooling. However, the aggregation behavior of EbE2-3 was not predicted by this algorithm, indicating that all of the factors leading to nonreversible multimerization have yet to be elucidated.

Unlike for PsaE and EbE1, the non-coincidence of the EbE4 equilibrium unfolding curves supported the population of a third state in the folding pathway. This third species was determined to be an alternate native conformation based on its population at 25 ˚C and 0 M urea, its nonrandom chemical shifts and extensive NOEs, and the narrow native baseline observed in the urea-gradient gel. The lack of overlap of the CD and UV-Vis absorbance data indicated that the N1 and N2 species were differentially monitored by these two methods.

4.3.3. Kinetic Effects

The choice of loop termini affected the observed folding time constants of the protein. Although extensive kinetic experiments were not performed on EbE1 or PsaE, no slow event was observed during the course of the rescue experiment designed to test the reversibility of the urea unfolding transition. In the case of EbE4, a slow refolding was detected and attributed to the participation of N2 in the mechanism. In general, three-state situations can be described with the following diagram:



This triangular scheme can be simplified to linear schemes presenting N2 either as an on-pathway or an off-pathway species. As discussed in the results section, this conformational exchange led to several anomalies in the kinetic behavior observed for EbE4. Similar complication of the folding kinetics has been observed for the apoflavodoxin (*51*) and ubiquitin (*52*) examples described above as well as a salt-induced

collapsed state of the 101-residue ribosomal protein S6 from *Thermus thermophilus* (*61*). In these cases, however, the folding pathway was described by the triangular model portrayed above in which some portion of the protein was able to populate N1 directly from the unfolded state rather than proceeding slowly through the N2 state. The ability of a single exponential function to fit the EbE4 kinetic folding and unfolding profiles combined with the decay from positive ellipticity in both regimes indicated that the behavior of this protein could not be explained by a similar model. Rather, the linear model (N1 ⇆ N2 ⇆ U) portraying obligatory population of N2 best describes the available data. In addition, the retrieval of the same apparent time constant for the double-jump refolding data regardless of the delay time is consistent with the absence of an additional stable folding intermediate. In spite of these observations, it is possible that an inspection of more rapid events that occur within the dead-time of the manual mixing experiment would reveal a small fraction of EbE4 populating N1 directly from the U state in a manner similar to that found for DHFR.

4.3.4. Attempted Heme Reconstitution of the EbE Chimeras

The inability of EbE2-4 to bind heme specifically supported that, in the context of the PsaE scaffold, the particular portion of the cytochrome b_5 binding loop chosen for the construct was unable to participate in holoprotein-like specific contacts with the heme. This could stem from steric or entropic penalties that would be present at the interface of the PsaE and cytochrome b_5 portions of the EbE proteins as a result of folding of both regions of the protein. If this were so, the relative free energy gain associated with heme binding and loop folding would have to exceed that present in the PsaE scaffold in order for observable population of the holoprotein state much in the fashion of the barnase-

ubiquitin chimera. Alternately, these data may suggest that the cytochrome b_5 core 1 region cannot bind heme in the absence of its native scaffold, and core 1-core 2 communication is necessary for physiological function.

4.4. Conclusions

The effects of long, unstructured regions on globular proteins are hard to predict. It has become apparent that many gene products contain such segments, and the importance of understanding their role has grown. We have shown here that the nature of the interface between structured and disordered protein segments has the ability to dictate several biophysical properties including free energy of folding, protein aggregation propensity, population of multiple native states, and folding kinetics. The ability to populate N1 and N2 states may be of functional relevance to the subclass of intrinsically unstructured proteins that contain both ordered and disordered elements. Many of these proteins are involved in more than one type of protein-protein interaction, and evolution of the loop termini may allow for differential folding to accommodate multiple binding partners. Regardless of the role or conformation of the loop, it is clear that the anchors responsible for joining of the loop and the scaffold can influence the packing of structured residues remote from them. Modest realignment of the EbE4 hydrophobic core in this manner provides yet another example of such long-distance communication.

References

(1) Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics 21*, 3433-4.

(2) Garner, E., Cannon, P., Romero, P., Obradovic, Z., and Dunker, A. K. (1998) Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization. *Genome Inform Ser Workshop Genome Inform 9*, 201-213.

(3) Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform Ser Workshop Genome Inform 10*, 30-40.

(4) Liu, J., Tan, H., and Rost, B. (2002) Loopy proteins appear conserved in evolution. *J Mol Biol 322*, 53-64.

(5) Sim, K. L., and Creamer, T. P. (2002) Abundance and distributions of eukaryote protein simple sequences. *Mol Cell Proteomics 1*, 983-95.

(6) Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol 323*, 573-84.

(7) Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit 18*, 343-384.

(8) Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem Sci 27*, 527-33.

(9) Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol 12*, 54-60.

(10) Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol 6*, 197-208.

(11) Verkhivker, G. M. (2005) Protein conformational transitions coupled to binding in molecular recognition of unstructured proteins: deciphering the effect of intermolecular interactions on computational structure prediction of the p27Kip1 protein bound to the cyclin A-cyclin-dependent kinase 2 complex. *Proteins 58*, 706-16.

(12) Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol 293*, 321-31.

(13) Chang, J. F., Phillips, K., Lundback, T., Gstaiger, M., Ladbury, J. E., and Luisi, B. (1999) Oct-1 POU and octamer DNA co-operate to recognise the Bob-1 transcription co-activator via induced folding. *J Mol Biol 288*, 941-52.

(14) Chasman, D., Cepek, K., Sharp, P. A., and Pabo, C. O. (1999) Crystal structure of an OCA-B peptide bound to an Oct-1 POU domain/octamer DNA complex: specific recognition of a protein-DNA interface. *Genes Dev 13*, 2650-7.

(15) Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A 93*, 11504-9.

(16) Daughdrill, G. W., Hanely, L. J., and Dahlquist, F. W. (1998) The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations. *Biochemistry 37*, 1076-82.

(17) Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990) Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature 347*, 575-8.

(18) Hyre, D. E., and Klevit, R. E. (1998) A disorder-to-order transition coupled to DNA binding in the essential zinc-finger DNA-binding domain of yeast ADR1. *J Mol Biol 279*, 929-43.

(19) Donaldson, L., and Capone, J. P. (1992) Purification and characterization of the carboxyl-terminal transactivation domain of Vmw65 from herpes simplex virus type 1. *J Biol Chem 267*, 1411-4.

(20) O'Hare, P., and Williams, G. (1992) Structural studies of the acidic transactivation domain of the Vmw65 protein of herpes simplex virus using 1H NMR. *Biochemistry 31*, 4150-6.

(21) Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J., and Pavletich, N. P. (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science 274*, 948-53.

(22) Radhakrishnan, I., Perez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., and Wright, P. E. (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell 91*, 741-52.

(23) Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998) Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput*, 437-48.

(24) Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, C. J., Williams, C. J., and Dunker, A. K. (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol 55*, 104-10.

(25) Fita, I., and Rossmann, M. G. (1985) The NADPH binding site on beef liver catalase. *Proc Natl Acad Sci U S A 82*, 1604-8.

(26) Boyington, J. C., Gladyshev, V. N., Khangulov, S. V., Stadtman, T. C., and Sun, P. D. (1997) Crystal structure of formate dehydrogenase H: catalysis involving Mo, molybdopterin, selenocysteine, and an Fe4S4 cluster. *Science 275*, 1305-8.

(27) Fukuchi, S., Homma, K., Minezaki, Y., and Nishikawa, K. (2006) Intrinsically disordered loops inserted into the structural domains of human proteins. *J Mol Biol 355*, 845-57.

(28) Yin, J., Beuscher, A. E. t., Andryski, S. E., Stevens, R. C., and Schultz, P. G. (2003) Structural plasticity and the evolution of antibody affinity and specificity. *J. Mol. Biol. 330*, 651-6.

(29) Falzone, C. J., Kao, Y. H., Zhao, J., Bryant, D. A., and Lecomte, J. T. (1994) Three-dimensional solution structure of PsaE from the cyanobacterium Synechococcus sp. strain PCC 7002, a photosystem I protein that shows structural homology with SH3 domains. *Biochemistry 33*, 6052-62.

(30) Knappenberger, J. A., Kraemer-Pecore, C. M., and Lecomte, J. T. (2004) Insertion of the cytochrome b5 heme-binding loop into an SH3 domain. Effects on structure and stability, and clues about the cytochrome's architecture. *Protein Sci 13*, 2899-908.

(31) Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J. E. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput*, 473-84.

(32) Bhattacharya, S., Falzone, C. J., and Lecomte, J. T. (1999) Backbone dynamics of apocytochrome b5 in its native, partially folded state. *Biochemistry 38*, 2577-89.

(33) Falzone, C. J., Mayer, M. R., Whiteman, E. L., Moore, C. D., and Lecomte, J. T. (1996) Design challenges for hemoproteins: the solution structure of apocytochrome b5. *Biochemistry 35*, 6519-26.

(34) Falzone, C. J., Wang, Y., Vu, B. C., Scott, N. L., Bhattacharya, S., and Lecomte, J. T. (2001) Structural and dynamic perturbations induced by heme binding in cytochrome b5. *Biochemistry 40*, 4879-91.

(35) Bryant, J. E., Lecomte, J. T., Lee, A. L., Young, G. B., and Pielak, G. J. (2005) Protein dynamics in living cells. *Biochemistry 44*, 9275-9.

(36) Chan, H. S., and Dill, K. A. (1989) Intrachain loops in polymers: effects of excluded volume. *J. Chem. Phys. 90*, 492-509.

(37) Falzone, C. J., Kao, Y.-H., Zhao, J., Bryant, D. A., and Lecomte, J. T. J. (1994) Three-dimensional solution structure of PsaE from cyanobacterium *Synechococcus* sp. strain PCC 7002, a photosystem I protein that shows structural homology with SH3 domains. *Biochemistry 33*, 6052-6062.

(38) Durley, R. C. E., and Mathews, F. S. (1996) Refinement and structural analysis of bovine cytochrome b_5 at 1.5 angstrom resolution. *Acta Crystallogr. D Biol. Crystallogr. 52*, 65-76.

(39) Falzone, C. J., Kao, Y.-H., Zhao, J., MacLaughlin, K. L., Bryant, D. A., and Lecomte, J. T. J. (1994) ^{1}H and ^{15}N NMR assignments of PsaE, a photosystem I subunit from the cyanobacterium *Synechococcus* sp. Strain PCC 7002. *Biochemistry 33*, 6043-6051.

(40) Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc. 123*, 2970-2978.

(41) Falzone, C. J., Mayer, M. R., Whiteman, E. L., Moore, C. D., and Lecomte, J. T. J. (1996) Design challenges for hemoproteins: the solution structure of apocytochrome b_5. *Biochemistry 35*, 6519-6526.

(42) Baxter, N. J., and Williamson, M. P. (1997) Temperature dependence of ^{1}H chemical shifts in proteins. *J. Biomol. NMR 9*, 359-369.

(43) Capaldi, A. P., Shastry, M. C., Kleanthous, C., Roder, H., and Radford, S. E. (2001) Ultrarapid mixing experiments reveal that Im7 folds *via* an on-pathway intermediate. *Nat. Struct. Biol. 8*, 68-72.

(44) Wang, L., Rivera, E. V., Benavides-Garcia, M. G., and Nall, B. T. (2005) Loop entropy and cytochrome c stability. *J. Mol. Biol. 353*, 719-729.

(45) Cayley, P. J., Dunn, S. M., and King, R. W. (1981) Kinetics of substrate, coenzyme, and inhibitor binding to Escherichia coli dihydrofolate reductase. *Biochemistry 20*, 874-9.

(46) Falzone, C. J., Wright, P. E., and Benkovic, S. J. (1991) Evidence for two interconverting protein isomers in the methotrexate complex of dihyrofolate reductase from Eschericia coli. *Biochemistry 30*, 2184-91.

(47) Wallace, L. A., and Matthews, C. R. (2002) Highly divergent dihyrofolate reductases conserve complex folding mechanism. *J. Mol. Biol. 315*, 193-211.

(48) Ionescu, R. M., Smith, V. F., O'Neill, J. C., Jr., and Matthews, C. R. (2000) Multistate equilibrium unfolding of Escherichia coli dihydrofolate reductase: thermodynamic and spectroscopic description of the native, intermediate, and unfolded ensembles. *Biochemistry 39*, 9540-50.

(49) Jones, B. E., Beechem, J. M., and Matthews, C. R. (1995) Local and global dynamics during the folding of Escherichia coli dihydrofolate reductase by time-resolved fluorescence spectroscopy. *Biochemistry 34*, 1867-77.

(50) Jones, B. E., Jennings, P. A., Pierre, R. A., and Matthews, C. R. (1994) Development of nonpolar surfaces in the folding of Escherichia coli dihydrofolate reductase detected by 1-anilinonaphthalene-8-sulfonate binding. *Biochemistry 33*, 15250-8.

(51) Lopez-Llano, J., Maldonado, S., Jain, S., Lostao, A., Godoy-Ruiz, R., Sanchez-Ruiz, J. M., Cortijo, M., Fernandez-Recio, J., and Sancho, J. (2004) The long and short flavodoxins: II. The role of the differentiating loop in apoflavodoxin stability and folding mechanism. *J Biol Chem 279*, 47184-91.

(52) Bofill, R., and Searle, M. S. (2005) Engineering stabilising beta-sheet interactions into a conformationally flexible region of the folding transition state of ubiquitin. *J Mol Biol 353*, 373-84.

(53) Radley, T. L., Markowska, A. I., Bettinger, B. T., Ha, J. H., and Loh, S. N. (2003) Allosteric switching by mutually exclusive folding of protein domains. *J Mol Biol 332*, 529-36.

(54) Halle, B. (2002) Flexibility and packing in proteins. *Proc Natl Acad Sci U S A 99*, 1274-9.

(55) Holder, J. B., Bennett, A. F., Chen, J., Spencer, D. S., Byrne, M. P., and Stites, W. E. (2001) Energetics of side chain packing in staphylococcal nuclease assessed by exchange of valines, isoleucines, and leucines. *Biochemistry 40*, 13998-4003.

(56) Jain, R. K., and Ranganathan, R. (2004) Local complexity of amino acid interactions in a protein core. *Proc Natl Acad Sci U S A 101*, 111-6.

(57) Rose, G. D., and Wolfenden, R. (1993) Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu Rev Biophys Biomol Struct 22*, 381-415.

(58) Willis, M. A., Bishop, B., Regan, L., and Brunger, A. T. (2000) Dramatic structural and thermodynamic consequences of repacking a protein's hydrophobic core. *Structure 8*, 1319-28.

(59) Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol 22*, 1302-6.

(60) Linding, R., Schymkowitz, J., Rousseau, F., Diella, F., and Serrano, L. (2004) A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J Mol Biol 342*, 345-53.

(61) Otzen, D. E., and Oliveberg, M. (1999) Salt-induced detour through compact regions of the protein folding landscape. *Proc Natl Acad Sci U S A 96*, 11746-51.

Chapter 5

Proximal influences in two-on-two globins: Effect of the Ala69Ser replacement on *Synechocystis* sp. PCC 6803 hemoglobin

Note: Multiple authors contributed to this work. For a detailed description of the division of labor, see the Preface located on page xix of this thesis.

5.1. Introduction

In many *b* hemoproteins the heme iron is endogenously hexacoordinated, with the axial positions of its octahedral geometry filled by two histidine side chains (*1*). The electron-transfer protein cytochrome b_5, which has been discussed in previous chapters, offers one example of this coordination scheme. *Bis*-histidine ligation of the heme group is also observed in a small number of proteins belonging to the hemoglobin (Hb) superfamily, which, unlike cytochrome b_5, are able to bind exogenous ligands. Recently discovered instances include *Drosophila* hemoglobin Hb (*2*), certain cyanobacterial globins (*3-5*), neuroglobin (*6*), and cytoglobin (*7-9*). The α subunit of adult human Hb is also capable of *bis*-histidine coordination of the ferric iron, particularly when in complex with a protective protein (*10, 11*). In these hexacoordinate globins, the axial histidines belong to the F helix (proximal side of the heme) and the E helix (distal side). It has been proposed that hexacoordination in globins facilitates reduction of the iron (*12*), moderates the reactivity of the heme (*11*), and enhances thermal stability (*13*).

Despite much effort applied to the study of proteins and model compounds (*1*), the mechanism by which iron redox potential and axial bond strength are regulated is not entirely understood. Likewise, the role of the coordination bonds in stabilizing the protein fold is not generally predictable. The globins mentioned above offer new ground for detailed analysis. In each case, coordination of the distal histidine is reversible, and an exogenous ligand can bind with measurable affinity to the heme. This property brings into focus the accessibility of the iron, the relative axial bond strength, and the respective role of proximal and distal effects in controlling heme reactivity. In this work, we used a cyanobacterial globin to explore the effects of a conservative substitution in the F helix in a *bis*-histidine protein.

Proximal effects have been investigated in pentacoordinate globins, and several studies provide context for the cyanobacterial system presented here. In pentacoordinate globins, substitutions at positions other than the proximal histidine (labeled F8 in the Perutz nomenclature), for example at F4 and F7, generally do not change the coordination and spin state of the iron or the fold of the protein. These residues, however, are often involved in packing interactions (*14, 15*) and hydrogen-bonding networks (*16*) that adjust essential characteristics of the protein, such as proximal Fe-N bond length and heme reactivity (*17, 18*). Proximal residues control solvent access (*19*) and contribute to the positioning of the proximal histidine and the iron relative to the heme plane, which, in turn, relate to the affinity for exogenous ligands (*20*) and the affinity for the heme group (*21*). The latter property has been shown to control the thermodynamic stability of holomyoglobin (holoMb) (*22*). In comparison to distal effects, proximal effects are subtle (*21*), but complex. This is well illustrated by F-helix swapping experiments in Hb and Mb and their consequences on heme spectroscopic

properties (*23, 24*); whereas Mb can be converted into an α-type globin by insertion of α-Hb residues, such conversion does not take place in β-Hb. Thus, proximal interactions, coupled or not with subunit interactions, play a role in dictating the behavior of pentacoordinate globins.

Among the many differences in sequence observed throughout the globin family, the identity of the amino acid at the position preceding the proximal histidine (corresponding to F7 in vertebrate globins) is of particular interest. In most Mbs, the wild-type residue is a serine. The neutron diffraction structure of sperm whale Mb establishes that the NδH atoms of the proximal histidine are within H-bonding distance of the seryl oxygen at F7

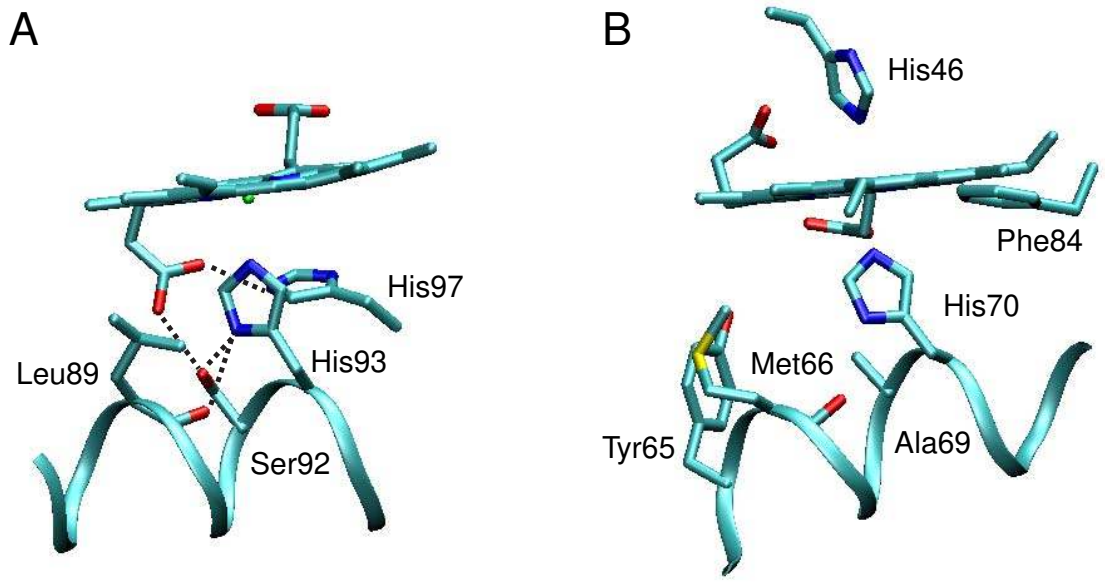


Figure 5.1. The environment of the proximal histidine in sperm whale myoglobin (5mbn, A) and *Synechocystis* rHb-R (1mwb, B). In A, the NδH group of His93 forms a hydrogen bond with Ser92 Oγ and Leu89 O. In B, the NδH group of His93 forms a hydrogen bond with Met66 O. The replacement of Ala69 with a Ser brings an O acceptor in the proximity of His70 NδH.

and the backbone oxygen of leucine F4 (*16*). Furthermore, the F7 hydroxyl group forms a hydrogen bond with the 7-propionate. This region of the Mb structure is shown in Figure 5.1A. Replacement of theserine has diverse outcomes depending on the source organism of the protein. It is capable of altering the geometry of the proximal histidine with respect to the heme group (*25-28*), increasing the affinity for small ligands in a modest *trans* effect (*25*), promoting heme loss (*25, 26*), and facilitating solvent access in the heme pocket (*25*). In vertebrate Hbs, residues at F7 possess side chains that are unable to participate in H-bonding interactions. Such is also the case in leghemoglobin (Lb) A, which contains a valine at that position; introduction of a serine at F7 does not increase the resemblance to Mb and speaks to the correlation of interactions on the proximal side of the heme (*21*).

The hexacoordinate globin of interest to this study is from *Synechocystis* sp. PCC 6803 (or S6803), a cyanobacterium incapable of nitrogen fixation. The protein belongs to the subfamily of "truncated globins" (trHbs). TrHbs are distantly related to the better-known vertebrate globins and can be organized into three groups according to phylogeny (*29, 30*). Thus far, a few structures have been solved for Group I and Group II trHbs. They present unique features: a distinctive helical topology referred to as the 2-on-2 fold (*31*), a heme pocket accessible via an apolar tunnel, and an elaborate network of H-bonds on the distal side of the heme (*32*).

Many Group I trHbs, including that from *Synechocystis*, contain an alanine immediately before the proximal residue (Figure 5.1B). Whether an exogenous ligand is bound or not, the proximal histidine of S6803 Hb assumes a staggered orientation with respect to the heme pyrrole nitrogens (Figure 2) (*33-35*), whereas the orientation of the same residue is eclipsed in Mb. S6803 rHb also has unusual ligand binding properties

(*36-38*), in part because of endogenous hexacoordination. To examine the role of residue F7, a serine was introduced at this position. *In vitro*, S6803 rHb can attach the heme covalently through the formation of a bond between His117 and the heme 2-vinyl (*39*). In the absence of information on the physiologically relevant form(s) of the protein, we first focused on the protein in which this cross-link, extremely rare among Hbs, was not formed (rHb-R). NMR structural characterization, denaturation experiments, and cyanide-binding experiments were performed to establish proximal influences in determining the globin properties. We report on the subtle changes induced by the H–OH substitution and compare these to observations made in selected pentacoordinate globins.

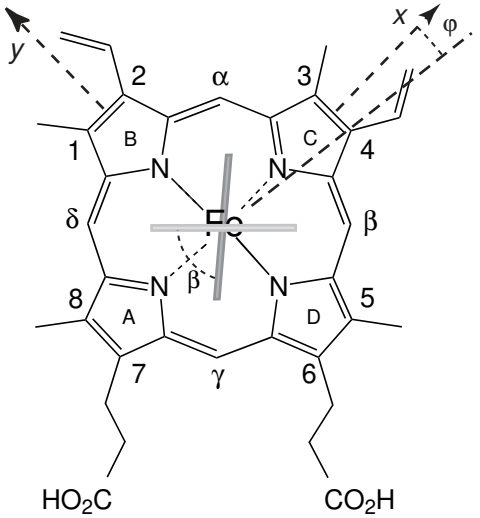


Figure 5.2. Structure of the *b* heme group with the nomenclature used in this work. The light and dark rectangles indicate the orientation of the proximal His70 and distal His46, respectively, in S6803 rHb-R. The φ and β angles characterize the geometry of the axial ligands: β is the acute angle between the two imidazole planes, and φ is the angle between the bisector of the two planes and the x axis. The proximal His93 in myoglobin is oriented along the N_B–N_D axis (not shown).

5.2. Results

5.2.1. Characterization of A69S S6803 rHb-R in the Ferric State

The hemichrome state of the protein (ferric, *bis*-histidine coordination) is a paramagnetic, $S = ½$ complex. The chemical shifts of the heme, axial histidines, and heme pocket residues were therefore expected to be sensitive reporters of structural and electronic perturbations arising from the A69S replacement. Figure 5.3 presents the one-dimensional spectrum of ferric A69S S6803 rHb-R along with that of the wild-type protein. The similarities are evident and support that the A69S replacement caused only minor three-dimensional and electronic changes to rHb-R. Resonance assignments for the heme group and for nearby residues (Table 5.1) were readily derived from an analysis of the homonuclear NOESY and TOCSY data and comparisons to data collected on H117A and wild-type rHb-Rs. The chemical shift differences between wild-type and A69S rHb-Rs were modest and localized on the proximal side of the heme, within a few residues of the replacement. On the distal side of the heme group, only His46 displayed chemical shift deviations larger than 0.1 ppm. Table 5.1 summarizes the chemical shift information for the axial residues and the heme group.

The geometry of the heme cavity is defined by a set of NOEs. The proximal histidine (His70) NδH was found to be in dipolar contact with Met66 CαH, as observed for the wild-type and H117A hemichromes. Met66 assumed the same position with respect to the heme group and the flanking Tyr65, as evidenced by contacts involving Met66 $C\varepsilon H_3$, Tyr65 CεHs, and the heme 8-CH_3. Compared to the wild-type protein, Met66 $C\varepsilon H_3$ was shifted upfield by ~ 0.2 ppm, a displacement that is also observed in H117A rHb-R and, therefore, cannot be attributed to H-bonding of Ser69 to His70 or the 7-propionate with

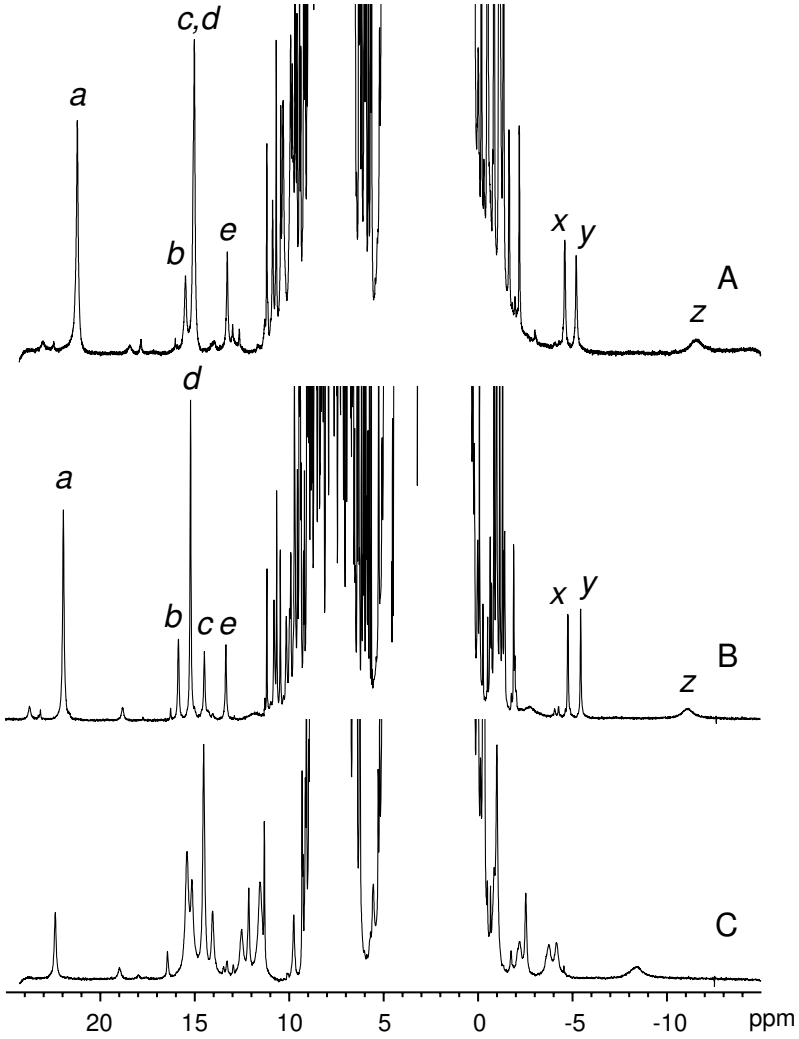


Figure 5.3. ^{1}H spectrum of (A) ferric wild-type S6803 rHb-R, (B) ferric A69S S6803 rHb-R, and (C) metcyano A69S S6803 rHb-R. Data were collected by SAK and BCV at 25 °C, pH 7.2. In traces A and B, the labels are as previously used (*44*): *a*, heme 5-CH_3; *b*, heme 2-α vinyl; *c*, His70 NδH; *d*, heme 1-CH3; *e*, His46 NδH; *x*, *y*, 2-β vinyls; *z*, His46 CεH. The resonance at 22.5 ppm in trace C arises from Tyr22 OηH hydrogen-bonded to the cyanide ligand.

certainty. NOEs between the heme 8-CH_3 and Tyr65, heme 2-vinyl and Phe84, heme 2-vinyl and Tyr53, heme 4-vinyl and Phe34, heme 3-CH_3 and Leu79, and heme 3-CH_3 and Val87 all confirmed similar placement of the heme. The relative proportions of native protein populating two states that differ by a 180° rotation about the heme α-γ meso axis is 95:5 in the wild-type hemichrome (*4*). The ratio shifted slightly in favor of the minor form in the variant. The proximal helix was traced by NH_i-NH_{i+1} NOEs from Arg64 to Asn76, except at the level of Arg67 and Glu68 because of overlap in the ^1H data. Helical conformation was also identified throughout the F helix with a number of resolved $H\beta_i$-NH_{i+1} NOEs. The $C\beta H_2$ group of Ser69, however, was not identified. An absence of NOE is not necessarily meaningful, but since strong effects are observed between Ala69 $C\beta H_3$ and His70 NδH, M66 $C\varepsilon H_3$, and Tyr65 CεHs in the wild-type and H117A proteins, this suggested that the serine side chain was not rigidly held in a single rotameric state and argued against the formation of a side chain hydrogen bond. No changes were observed in the few NOEs of the heme 7-propionate.

5.2.2. Characterization of A69S S6803 rHb-R in the Cyanide-bound State

Binding of cyanide to the wild-type protein results in a broadening of the ^1H spectrum. The proximal-side substitution H117A has the remarkable consequence of sharpening the heme resonances (*36*). The A69S replacement did not have the same effect (Figure 5.3C); the lines for the cyanide complex remained as broad as for the wild-type, but the shifts were affected to some extent. A natural abundance ^1H-^{13}C HMQC spectrum (not shown) identified the signals from His70 CαH and CβHs. In the cyanomet form, the average axial histidine Hβ shift is indicative of imidazolate character. The average values for cyanomet H117A and A69S rHb-Rs were 6.5 ppm and 7.1 ppm, respectively.

For comparison, wild-type sperm whale metMbCN and S92A metMbCN have shifts of 9.1 ppm and 8.5 ppm, respectively (*28*), and peroxidases display values between 14 ppm and 22 ppm (*40*). The moderate change in the average chemical shift observed in Mb and S6803 rHb-R is consistent with some reorganization of the proximal site geometry.

The ^{15}N chemical shift of the bound cyanide can also be used as a probe of hydrogen bonding and imidazolate character of the axial histidine in *trans*. At 32 °C and neutral pH, the difference in chemical shift between free and bound cyanide was determined to be 692 ppm in the wild-type protein (*36*). This difference increases to 701 ppm in the A69S variant. Evidence for H-bonding of the bound cyanide is provided by the 22-ppm exchangeable signal attributed to Tyr22 (B10) Oη^1H in Figure 3C (*36*). Thus, this aspect of the structure appeared conserved. The difference of less than 10 ppm in C^{15}N shift between the two proteins did not indicate a significant change in imidazolate character, as the deviation was a small fraction of the total range of shifts observed for globins and peroxidases (*41*).

5.2.3. Thermal Denaturation of Ferric Wild-type and A69S S6803 rH-Rs

The effect of the replacement on the thermodynamic stability of the protein was inspected with equilibrium unfolding experiments. Thermal denaturation data were collected by measuring absorbance and ellipticity as functions of temperature. The first method yielded the apparent fraction of natively bound heme and the second, the apparent fraction of folded polypeptide. Both fractions were calculated employing a two-state model; they are shown in Figure 5.4.

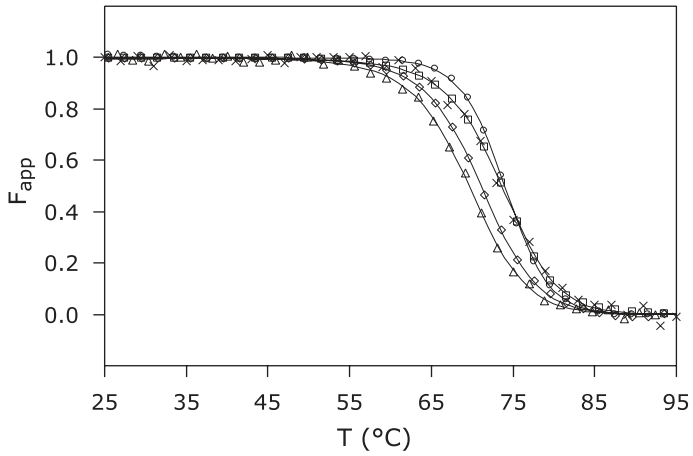


Figure 5.4. Thermal denaturation of S6803 rHb-R. Data were collected by JAK and SAK at pH 7.2 on the wild-type and A69S proteins in the ferric state using CD and visible spectroscopy. ×: ferric wild-type S6803 rHb-R, CD; □: ferric wild-type S6803 rHb-R, visible; △: ferric A69S S6803 rHb-R, CD; ◊: ferric A69S S6803 rHb-R, visible; ○: metcyano A69S rHb-R, visible.

The near coincidence of the two curves for the wild-type protein supported that heme release occurred concomitantly with loss of secondary structure. In the thermal denaturation of *b* hemoproteins, the interpretation of denaturation curves is complicated by the solubility properties of the heme group. When the thermal energy is sufficient, the heme is released from its binding site, but remains associated, presumably non-specifically, with the protein matrix. Thus, the experiment measures the difference in free energy between specific and non-specific binding of heme to the protein. The stability of the apoprotein also factors into the response as a function of external conditions (*22, 42*). When the A69S rHb-R data were compared to those obtained with the wild-type protein, it appeared that the midpoint of the thermal transition occurred at a slightly lower temperature in the variant (Table 5.2). Assuming that the non-specific affinity and the apoprotein stability were unchanged by the replacement, the result implied a lower specific affinity for the heme group, translating into an apparent lower

stability of the holoprotein. More striking than the lowering of T_m, however, was the non-coincidence of the curves obtained by monitoring the CD (backbone) and Soret (heme) signals. The discrepancy was small but reproducible, even though the thermal denaturation process was only partially reversible. The non-coincident curves were indicative of three-state behavior at a minimum.

5.2.4. Chemical Denaturation of Ferric Wild-type and A69S S6803 *rHb*-Rs

To verify that heme affinity was decreased by the substitution, urea-induced denaturation of the wild-type and variant proteins was performed. At neutral pH, the process was fully reversible and lent itself to thermodynamic analysis. Figure 5.5 presents the apparent fraction of folded protein as a function of urea concentration. Non-coincidence of the signals arising from heme release and secondary structure loss was observed for both wild-type and A69S rHb-Rs, with the midpoint of the transition corresponding to heme release occurring at a higher urea concentration. Thermodynamic parameters obtained via two-state analysis are listed in Table 5.3. The difference in the free energy of unfolding for the two proteins was small. In further support that the experimental methods were monitoring slightly different transitions, the *m* values obtained for the CD and absorbance data were not within error of each other. In addition, comparison of the $\Delta G°$ values obtained on the two hemichromes via the same method indicated that the substitution resulted in a change in the balance of free energy associated with heme ligation and secondary structure formation.

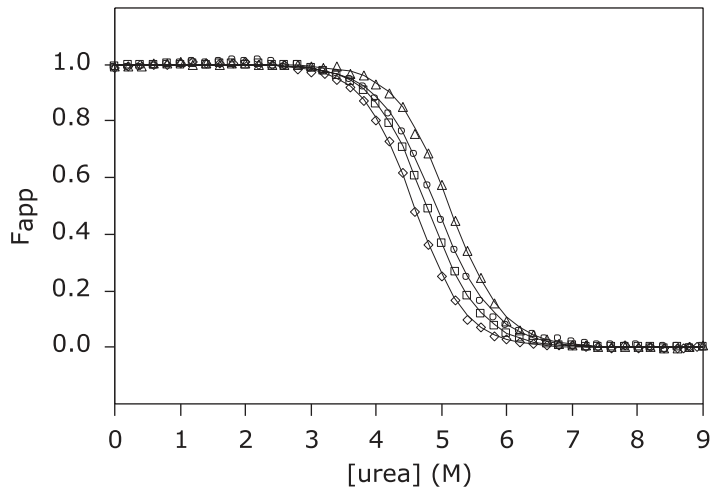


Figure 5.5. Urea-induced denaturation of S6803 rHb-R. Data were collected by JAK and SAK at pH 7.2 and 25 °C on the wild-type and A69S proteins in the ferric state using CD and visible spectroscopy. ○: ferric wild-type S6803 rHb-R, CD; △: ferric wild-type S6803 rHb-R, visible; ◊: ferric A69S S6803 rHb-R, CD; □: ferric A69S S6803 rHb-R, visible.

5.2.5. Urea Gradient Gels

In its apoprotein state, S6803 rHb has limited stability (*4*). Preliminary measurements indicated no difference due to the A69S replacement. This was confirmed by analyzing denaturation profiles obtained with urea gradient gels of wild-type and A69S apoproteins. The transitions observed for the apoglobins run individually lacked a well-defined range of urea concentration in which the native state was fully populated. For all practical purposes, the transition was complete at 2 M urea at pH 7.5. Gels run on mixtures of the two apoproteins (so that they were subjected to identical gradients) yielded indistinguishable traces. Such a gel is shown in Figure 5.6. It is noteworthy that under the chosen conditions, the unfolding and refolding rates were sufficiently fast to yield a detectable, continuous band of protein. This indicated that the half-times of the species co-existing at any urea concentration were shorter than 0.1 times the duration of the electrophoretic separation (*43*).

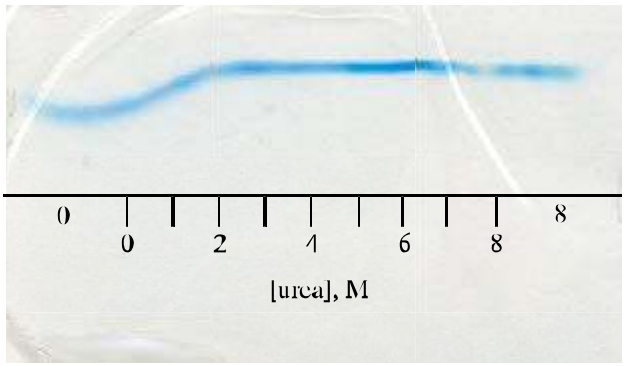


Figure 5.6. Urea-gradient gel electrophoresis of wild-type and A69S apoproteins stained with Coomassie brilliant blue. The scale represents approximate urea concentrations. "0" and "8" indicate regions of the gel with constant urea concentration. The proteins begin to unfold immediately upon exposure to urea; by 2 M urea, the protein is completely unfolded, and the horizontal line corresponds to protein having an electrophoretic mobility consistent with the unfolded state. The two traces, which were observed individually, overlay perfectly in the mixture indicating that the substitution did not affect the apoprotein's resistance to urea denaturation. Data were collected by JAK.

Urea gradient gel eletrophoresis was also performed on the holoproteins (Figure 5.7) in an effort to obtain equilibrium and kinetic information simultaneously. Both Coomassie staining and heme staining were applied to reveal protein and dissociated heme, respectively. The gels were loaded either with native protein (0 M urea; Figure 5.7) or unfolded protein (8 M urea; data not shown) and returned identical results in a demonstration that the process was fully reversible. In the wild-type S6803 rHb-R gels, a native band with a constant mobility was observed up to ~ 2 M urea (Figure 5.7A). At this concentration, a faint band was also detected with mobility consistent with unfolded protein; this band became intense above 2.5 M. Between 2 and 2.5 M urea, there was no detectable amount of protein with intermediate mobility. As the concentration of urea increased, the rate at which the heme was lost from the holoprotein also increased. The interrupted pattern of the stained gel indicated that, when the urea concentration neared 2

M, heme loss became irreversible because released heme migrated through the gel matrix faster than the protein could refold and recapture it. Under those conditions, the globin found itself in the apoprotein state at urea concentrations that were sufficient to denature it completely. Thus, the profile obtained for the holoprotein was not expected to mimic the transition observed for the urea titration, which takes place between 3.2 and 6 M urea (Figure 5.5). Heme staining of gels run for a shorter time confirmed that at ~2.2 M urea, the heme had dissociated irreversibly from the protein (Figure 5.7B). The sensitivity of these experiments was often insufficient to detect the heme when bound to the protein.

When the same experiments were performed on A69S rHb-R, the profile was similar, but not identical (Figure 5.7C). Although a discontinuity was observed at approximately the same urea concentration as for the wild-type rHb-R, detectable amounts of protein had intermediate mobility at the edges of the transition.

5.2.6. Heme Release from A69S S6803 rHb-R upon Solution Acidification

S6803 rHb-R can undergo a chemical modification by which the *b* heme becomes covalently attached to His117 via the 2-vinyl group, yielding "rHb-A" (*39*). When the cross-link is absent, lowering the pH results in the release of the heme group from its pocket in addition to the protonation of His46 and His70. The heme release can be detected optically as the cofactor spin state changes from low to high. pH titrations of the wild-type rHb-R display a sharp transition with a midpoint at pH 3.6 and a Hill coefficient of ~ 3.4 (*44*). A second transition with a higher pK_a (~ 5) is also observed. The titration curve is shown in Figure 5.8, which also contains A69S S6803 rHb-R data.

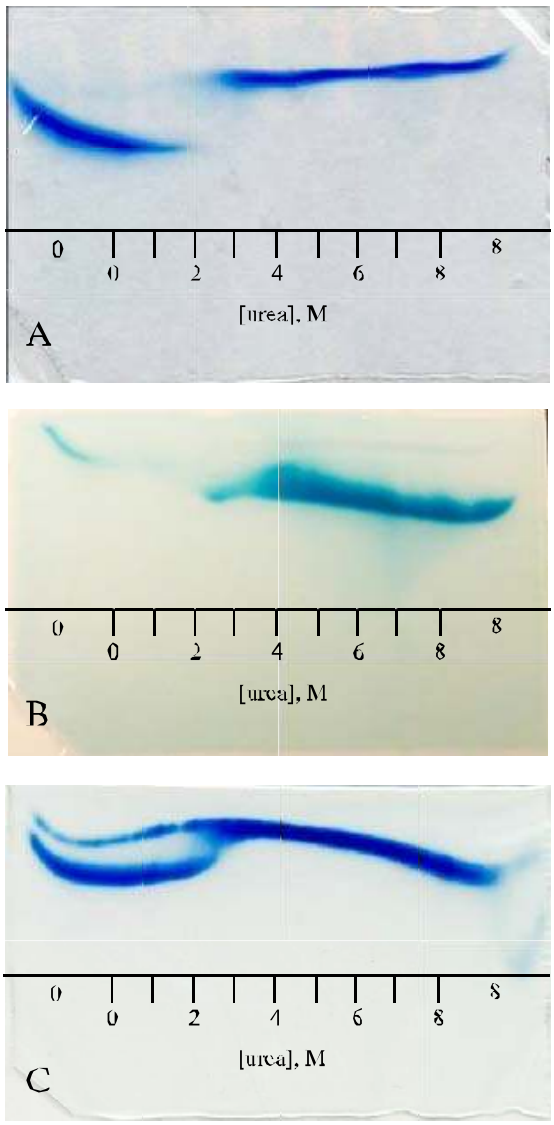


Figure 5.7. Holoprotein urea-gradient gels: A) 4-h electrophoresis of wild-type S6803 ferric rHb-R visualized by Coomassie blue. The two species observed are the natively folded holoprotein (0-2 M) and the unfolded apoprotein (~2.5-8 M). B) 40-min electrophoresis of wild-type and A69S S6803 ferric rHb-R stained for heme. The thin band in the 0-M buffer region corresponds to heme specifically bound to natively folded protein; the thick band seen from 2-8 M urea arises from heme that has escaped irreversibly from the protein. C) 4-h electrophoresis of A69S ferric S6803 rHb-R visualized by Coomassie blue. The sample contained both the apo- and holoprotein. The apoprotein (thin band) mirrors the transition seen in Figure 5.6. The A69S holoprotein (thick band) follows the pattern of the wild-type holoprotein (Figure 5.7A) except for the transition is seen at 2-2.5 M urea. In all cases the proteins were loaded on the gel in the native state. The scale represents approximate urea concentrations. "0" and "8" indicate regions of the gel with constant urea concentration. Data were collected by JAK.

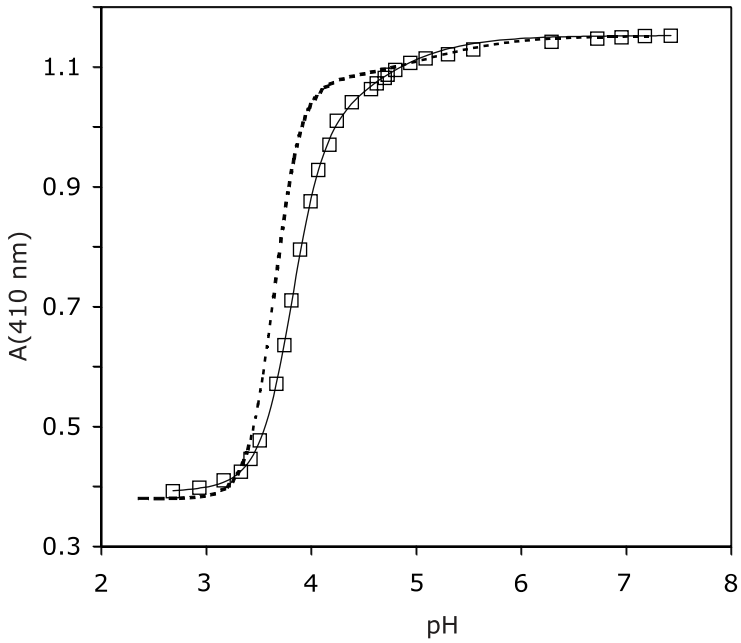


Figure 5.8. pH titration of ferric A69S rHb-R. Data were collected by SAK and BCV at 25 °C. The solid line represents the fit to a two-pK_a model. The main transition has pK_a = 3.8 and Hill coefficient ~ 3. The dashed line represents the results of the titration of ferric wild-type rHb-R (*44*), scaled to attain identical intensity at pH 7.2.

In the variant, the midpoint for heme release occurred at pH 3.8, and the Hill coefficient was ~ 3. The higher pK_a of the main transition obscured the second transition. In both proteins, the Hill coefficient suggested that release was coupled to at least three protonation events, which would arise from His46, His70 and a third group, possibly the heme 7-propionate as a hydrogen bond was noted to Lys42 (*33*).

5.2.7. Cyanide Binding by Ferric A69S S6803 rHb-R.

In the process of binding cyanide, the distal ligand to the heme iron (His46) is ultimately replaced by CN^-, the dissociation product of HCN. Hence, multiple equilibria must be considered, including:

$$\text{rHb-R (hexacoordinate)} \underset{k_H}{\overset{k_{-H}}{\rightleftharpoons}} \text{rHb-R (pentacoordinate)}$$

$$\text{HCN} \rightleftharpoons \text{H}^+ + \text{CN}^-$$

$$\text{rHb-R (pentacoordinate)} + \text{HCN} \overset{k_{HCN}}{\rightleftharpoons} \text{cyanomet rHb-R} + \text{H}^+$$

$$\text{rHb-R (pentacoordinate)} + \text{CN}^- \overset{k_{CN^-}}{\rightleftharpoons} \text{cyanomet rHb-R}$$

Wild-type S6803 rHb-R binds cyanide slowly at neutral pH (*36*): the time required for half-conversion to the cyanomet form in the presence of 5 mM total cyanide (HCN and CN^-) is approximately 10 min, with a corresponding bimolecular rate constant of ~ 3×10^{-1} M^{-1} s^{-1}. This value is three orders of magnitude slower than in vertebrate globins and pentacoordinate 2-on-2 globins (*45*).

The H117A replacement results in a protein that binds cyanide 8 times more slowly than wild type (*36*). This effect is noteworthy as His117 is located on the proximal side of the protein and is not positioned in the ligand access tunnel described in 2-on-2 globins (*29*). To probe a distinct proximal influence, the binding experiment was repeated with A69S S6803 rHb-R. Upon addition of a 200-, 500-, or 1000-fold excess of cyanide to protein solution, the absorption at 423 nm was measured as a function of time. This manual mixing procedure yielded the relative apparent rates of association. To test the hypothesis that HCN dissociation played a significant role, the experiments were repeated in D_2O.

Apparent pseudo-first-order rate constants obtained for the wild-type and A69S ferric proteins are shown in Figure 5.9 as a function of total cyanide concentration. These constants are associated with the major kinetic phase of four phases required to approach randomly distributed residuals in the fit of the time courses. The major phase accounted for approximately 80% of the total protein population. As shown by the slope of the

lines, binding to the A69S variant was slower than to the wild-type protein. The ratio of the apparent second-order rate constants for the two proteins was approximately 2. Thus, the A69S replacement, as the H117A replacement, decelerated cyanide binding. In D_2O, the number of kinetic phases remained the same in the A69S and wild-type proteins; the major phase was decelerated by a factor of 1.5 – 2 in both proteins (data not shown). At pH 8.2, the pseudo-first-order rate constant obtained for a solution containing a 200-fold excess of total cyanide was accelerated by a factor of 5 relative to the same solution at pH 7.2 (data not shown). This increase in pH corresponded to a 9-fold increase in free cyanide concentration. The acceleration did not scale according to the apparent bimolecular rate constant obtained at pH 7.2.

Thermal denaturation of the cyanomet complex of wild-type (data not shown) and A69S rHb-R (Figure 5.4) in the presence of a 1000-fold excess of cyanide showed that heme release occurred at a higher temperature and with a steeper slope than in the absence of exogenous ligand. This is also observed with H117A rHb-R (*36*).

5.3. Discussion

The results obtained on A69S S6803 rHb-R take their significance in comparison with Mb and Lb, proteins for which the role of residue F7 has been analyzed in detail by several methods. In pig Mb (Ser92 at F7), disruption of the F7-F8 and F7-7-propionate H-bonding interactions causes an increase in affinity for small ligands and an enhanced propensity to lose heme. Interactions between the 7-propionate and His97 (FG3) are also affected, and the new geometry increases permeability to solvent (*25*). In human Mb, ligand binding is unaffected by F7 replacement, and the orientation of the proximal histidine rotates from the eclipsed position by a few degrees (*26*). Slight rotation of the

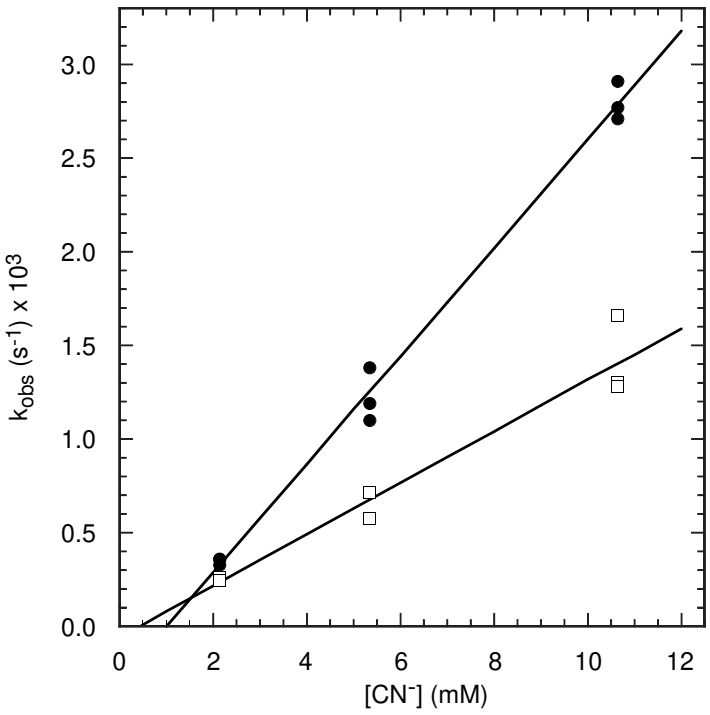


Figure 5.9. Apparent pseudo-first-order rate constants for association of cyanide with wild-type (•) and A69S (□) ferric S6803 rHb-Rs. The protein concentrations were 10.7 μM and 10.8 μM, respectively. The slopes indicated apparent bimolecular rate constants of 3.0×10^{-1} M^{-1} s^{-1} (wild-type) and 1.4×10^{-1} M^{-1} s^{-1} (A69S). The multiple exponentials required to fit the data contributed to errors in the values shown, as illustrated by the non-zero extrapolation in the absence of cyanide. As a result, the assessment was used only in a semi-qualitative fashion. Data were collected by HJN.

proximal histidine is also reported for the S92D variant of horse heart Mb in the cyanomet state (*27*). The role of proximal interactions was reinspected by Peterson and coworkers, who studied variants containing multiple replacements and 8-residue F-helix swaps (*46*). These authors conclude that the hydrogen bond network involving Ser92 serves principally to stabilize the proximal pocket and retain the heme within the protein. The reverse, stabilization upon introduction of a hydrogen-bonding residue at F7, was tested with the V91S variant of Lb; in fact, heme loss is accelerated by the replacement (*21*). In this protein, the orientation of the proximal histidine is staggered with respect to

the pyrrole nitrogens, a property achieved via steric interactions in the absence of hydrogen bonds (*21*). Common effects of F7 replacement are, therefore, small structural changes and holoprotein destabilization.

5.3.1. Structure

In S6803 rHb-R, His70 adopts a staggered conformation that presumably stabilizes distal ligand binding (*20*). The A69S replacement affected the chemical shift of the axial histidines moderately (Table 5.1), with larger proximal than distal deviations. Minor shifts were also observed for the heme group. Peripheral substituents (protons in α to PPIX ring) experienced an upfield shift on pyrrole A and C and a downfield shift on pyrrole B and D. In previous work, the paramagnetic shift of the methyl groups was used to assess the orientation of the axial histidines in the wild-type protein (*33*). The same heuristic approach (*47*) applied to the A69S variant showed that the shifts were consistent with a rotation of a few degrees of the bisector of the two histidine planes and a slight decrease in the angle between the two imidazole planes. The method is approximate and does not provide information on individual ligands; comparison of distal and proximal shifts, however, supported that a change in the orientation of the proximal histidine was most likely responsible for the effect. The shift pattern suggested that His70 rotated to align the projection of its plane closer to the x-axis (Figure 5.2). Such a move could accommodate the Ser69 hydroxyl group and, perhaps, reorient the NδH vector so as to favor hydrogen bonding. However, no direct evidence for formation of this hydrogen bond was found. The 7-propionate, which could potentially form a hydrogen-bond with Ser69 OH as in Mb, interacts with Lys42 in the wild-type protein. NOE perturbations did not suggest rearrangement near the propionates. Other perturbations were minor as well and suggested conserved geometric and electronic properties in the variant.

5.3.2. Stability

Urea-induced denaturation of the A69S and wild-type apoproteins, as observed by urea-gradient gel electrophoresis, indicated that the substitution had no effect on the resistance to denaturant. In both cases, the unfolding transition began at the lowest urea concentration and was complete by 2 M. The observation of a single, narrow protein band in the gel on which a mixture of wild-type and A69S proteins was run suggested that alteration in the free energy of holoprotein folding could be attributed to changes in heme-protein interactions.

The temperature- and urea-induced denaturation results indicated that the substitution had only slight effects on the relative free energies of the folded and unfolded states of the protein in the presence of the heme group. The most notable feature of the denaturation curves was the non-coincidence of the CD and absorption data. Through the thermal experiments, the A69S substitution revealed a decoupling of heme-iron decoordination and loss of secondary structure. This is consistent with the lower helical propensity of serine compared to alanine and with an intrinsically less stable turn of helix encompassing that position. In the urea experiments, neither protein followed a two-state mechanism. In all non-coincident cases, the CD data displayed a lower C_m than the absorbance data; therefore, backbone unraveling began at lower denaturant concentration than heme loss. A possible interpretation is that the serine side chain introduced strain into the F helix, made it more susceptible to denaturation, and facilitated decoordination of His70 once this helix started to fray. In this view, an intermediate state would be populated that contains heme specifically bound to a protein with a partially folded F helix. An extreme case of decoupling has been reported in *Drosophila* cytochrome *c*,

where replacement of Pro30 is thought to disrupt a hydrogen-bond network involving the axial histidine (*48*).

The urea-induced denaturation of the holoproteins in solution yielded dramatically different results than the same experiment performed within the confines of an acrylamide gel. In solution, the system is allowed to come to equilibrium, and the folding and unfolding rates (as well as the rates of specific and non-specific heme association and dissociation) must only be considered in the sense that these quantities determine the time required to reach equilibrium and the equilibrium constant at each urea concentration. In the urea-gradient gel, however, an additional factor must be considered. Because the molecules are migrating, the ability of the protein to remain associated with the heme prosthetic group depends critically on the relative rates of protein refolding, heme association, and protein and heme migration. The urea-gradient gel method offers kinetic insights that are otherwise experimentally difficult to obtain.

The overall gel pattern obtained for the wild-type and A69S holoproteins suggested that, at ~ 2 M urea, the rate of heme migration exceeded the rate of holoprotein refolding, and the two species observed in the gel were the folded holoprotein and the unfolded apoprotein. There was, however, a reproducible difference between the profiles obtained for the wild-type and the variant proteins. Whereas the wild-type gel showed only native and unfolded baselines, the A69S gel displayed hints of a transition. This could be explained in one of two ways: either the variant protein had a higher non-specific heme affinity (slower k_{diss}) or it refolded slightly faster than the wild-type protein. The former effect was not expected because non-specific heme affinity arises from low solubility of the cofactor, a property counteracted by urea. In addition, the two proteins differ by an OH group with limited ability to affect non-specific interactions. Hence, the dissimilarity

between the wild-type and variant likely reflected a change in the refolding rate. The slight decrease in the free energy of folding for the variant relative to wild type (Table 5.3) indicated that the unfolding rate must have been increased also, arguing for a lowering of the transition state.

Kinetics studies performed on cytochrome b_{562} have indicated that holoprotein folding from a partially-folded apo state to which the heme group is associated proceeds significantly faster than refolding from the totally denatured state (*49*). If the same behavior were followed by S6803 rHb-R, the difference between the holoprotein gels for the wild-type and A69S proteins could arise from changes in either stage: apoprotein folding or apo-to-holoprotein transition. This scenario is likely to apply only to hemoproteins that have high-stability apo forms and can remain folded at denaturant concentrations that are sufficient to disrupt the specific heme-protein contacts formed in the holoprotein. By comparison, the rHb-R apoproteins began unfolding immediately upon addition of urea (Figure 5.6). In addition, no evidence for the population of the folded apoprotein was detected in either the holoprotein urea-gradient gels (Figure 5.7) or the titration experiments performed in solution (Figure 5.5). Rather, it is likely that the change in the staining pattern observed in the urea-gradient gel upon the A69S substitution reflects an alteration of the kinetics of folding to the holoprotein state from the fully unfolded state.

The acid-induced denaturation of the two proteins provided another measure of relative heme affinities. Two of the three protonation events linked with heme release are expected to involve the axial ligands. The 0.2-unit increase in the pH midpoint for the decoordination of the heme iron in the A69S protein was consistent with a moderate decrease in the strength of the association with the cofactor. Assuming once more that

non-specific heme affinity remained unchanged, these results indicated a modest $\Delta\Delta G°$ (~ 1 kJ mol^{-1}) relative to the total free energy of the process and were consistent with those for the urea-induced denaturation experiments.

5.3.3. Cyanide Binding

Fitting the time course of cyanide binding by S6803 rHb-R under pseudo-first-order conditions requires several kinetic phases. This is unlike other globins, which exhibit monophasic behavior (*45*). The major phase extracted from wild-type S6803 rHb-R data has an apparent bimolecular rate constant comparable to that observed in *Glycera dibranchiata* Hb component III (*50*). In general, cyanide binding occurs minimally through two processes, one involving HCN (with bimolecular rate constant k_{HCN}) and another CN^- (with bimolecular rate constant k_{CN-}). The observed rate constant contains a contribution from each process, weighted by the HCN and CN^- fractional populations at the given pH. In *G. dibranchiata* Hb, the slow observed rate constant is attributed to the absence of a functional group (e.g, His or H_2O) capable of catalyzing HCN dissociation in the heme cavity. The preference for CN^- binding over HCN binding ($k_{CN-} > k_{HCN}$) is manifested in the pH dependence of the reaction kinetics. It is noteworthy that the 5-fold acceleration caused by a change of pH from 7.0 to 8.0 (*50*) is also comparable to the effect noted in S6803 rHb-R. Distinct isoforms of *G. dibranchiata* Hb display rate constants that vary within an order of magnitude and are likely to reflect the influence of minor structural perturbations on the balance of HCN and CN^- binding.

The similarity of S6803 rHb-R and *G. dibranchiata* Hb with respect to cyanide binding is intriguing. In S6803 rHb, not only can HCN dissociation contribute to slow binding, but also decoordination of His46. The rate constants for His46 ligation (k_H) and

deligation (k_{-H}) in the ferrous and presumably cross-linked state have been reported (*37*); they are rapid (k_H = 4200 s^{-1} and k_{-H} = 930 s^{-1}, resulting in an equilibrium constant K_H of 4.5) and suggest that the dissociation of HCN may dominate the reaction. The k_{-H} value, however, depends on whether exogenous ligand binding is initiated by flash photolysis (*37*) or rapid mixing (*38*); decoordination occurs almost two orders of magnitude more slowly in the rapid mixing case (k_{-H} = 14 s^{-1}, dictating an equilibrium constant K_H of 300). Rapid mixing conditions resemble more closely our experimental method, though the oxidation and heme cross-link states of the protein still differ. NMR data on the ferric S6803 rHb-R do not reveal any spectral contribution from a pentacoordinate species at room temperature, in support of an equilibrium favoring the hexacoordinate state ($K_H >> 5$).

Previous studies on the H117A variant of S6803 rHb indicate that the presence of His117 increases the rate of cyanide binding (*36*) relative to an alanine at this position. The solution structure of the wild-type globin (*33*) suggests that His117, the residue that cross-links the heme, samples rapidly multiple conformations, both in and out of the heme pocket. A role for His117 in the dissociation of HCN is therefore conceivable. A similar role for Ala69 is not as readily envisioned. The side chain is located at the edge of the heme pocket, and polarity is increased by the serine substitution. An alternative interpretation is that the residues at position 69 and 117 affect the association and dissociation kinetics of His46. In its dissociated state, His46 can serve as a proton acceptor (as HisE7 does in vertebrate globins) and is a good candidate for catalyzing HCN dissociation. In this view, proton transfer efficiency would depend on the residence time of His46 in the decoordinated state. It is possible that the A69S replacement consolidates the His46-Fe bond, thereby slowing down HCN dissociation. The

observation of larger-than-average chemical shift deviations for His46 may reflect this subtle alteration. A similar effect accompanied by reseating of the heme group can be invoked for the deceleration caused by the H117A replacement. A confirmation of the role of His117 and connection to His46 kinetics was found in the behavior of the cross-linked protein, which bound cyanide ~ 3 times faster than S6803 rHb-R (wild-type protein). The small solvent isotope effect (k_{H_2O}/k_{D_2O} = 1.5 – 2) was inconclusive as it could arise from a HCN bending mode or betray indirect influences associated with the replacement of labile hydrogens on the protein.

Assuming that the deceleration in A69S rHb-R was related to a reinforcement of the distal coordination bond, the difference in rate constants compared to wild type indicated a *trans* effect being exerted by the proximal substitution. Because the acid-induced heme release and the denaturation monitored in the visible supported a lowering of the heme affinity via axial bond weakening, the stabilization of the His46-Fe bond would appear smaller than the destabilization of the His70-Fe bond, with a net effect of holoprotein destabilization.

A recent study of neuroglobin and cytoglobin, both *bis*-histidine globins, relates the hyperthermostability of these proteins to their high K_H in the ferrous state (*13*). In addition, exogenous ligand binding and distal histidine substitution causes a reduction of this stability. S6803 rHb-R, with T_m of 75 °C and K_H of 300, does not cluster with the neuroglobin/cytoglobin set. In fact, the thermal denaturation of wild-type, A69S (current study), and H117A (*36*) proteins show that the cyanide-bound state is more stable than the hemichrome state. In wild-type S6803 rHb-R, cyanide binding results in helical rearrangement and formation of a distal hydrogen-bonding network between the ligand and the protein (*35, 36*). These conformational changes are more extensive than

observed in neuroglobin upon ligand binding (*51*) and lead to a repacked S6803 rHb-R interior with additional favorable interactions in the native state. Thus, the outcomes of hexacoordination and endogenous-exogenous ligand substitution differ in the cyanobacterial 2-on-2 globins and other types of globins.

5.4. Conclusion

The 2-on-2 globin fold has structural features that distinguish it from the canonical globin fold, including unique pathways to the heme pocket and ligand stabilization mechanism. Additional data will be necessary to achieve a full description of these traits and their effects. Available information, however, already indicates that the residues near the heme group are essential to controlling interactions with heme and small ligands, as in other globins. The A69S replacement in S6803 rHb revealed that structural stability and ligand binding properties could be finely adjusted by proximal influences.

The amino acid preceding the proximal histidine in globins has been recognized as a partial determinant of properties such as the azimuthal angle of the proximal histidine and heme affinity. The introduction of a residue capable of hydrogen bonding at F7 in the hexacoordinate S6803 rHb-R resulted in modest decreases in resistance to temperature-, urea-, and acid-induced denaturation as well as changes in exogenous ligand-binding kinetics. A decrease in helical propensity triggering *trans* effects can explain the observations. A behavior common to globins emerges by which alteration of F7 alone, regardless of the nature of the change, deteriorates heme affinity. Proximal networks therefore appear optimized for each specific combination of residues. In contrast to cytochrome b_5, which remains endogenously hexacoordinate even in the presence of high concentration of potential ligands, and other *bis*-histidine complexes that have the ability to bind ligands but become destabilized in the process, the S6803

rHb-R protein seems to maintain relative proximal and distal bond strengths that allow for increased stability in the exogenous-ligand-bound state. This could speak to a physiological role, particularly if covalent attachment of the heme to His117 is not a permanent characteristic of the protein *in vivo*.

Table 5.1: Selected ^{1}H NMR chemical shifts for S6803 rHb [a]

	Assignment [b]	WT [c]	A69S	Δ
heme	1-methyl	15.03	15.21	+0.18
	3-methyl	9.99	9.44	–0.55
	5-methyl	21.28	21.95	+0.67
	8-methyl	10.33	9.68	–0.65
	2-α-vinyl	15.56	15.86	+0.30
	trans-2-β-vinyl	–4.57	–4.77	–0.20
	cis-2-β-vinyl	–5.22	–5.45	–0.23
	4-α-vinyl	6.80	6.34	–0.46
	trans-4-β-vinyl	–1.53	–1.40	+0.13
	cis-4-β-vinyl	–2.05	–1.91	+0.14
	6-α-propionate	8.32	8.64	+0.32
	6-α'-propionate	9.67	9.97	+0.30
	6-β-propionate	1.41	1.51	+0.10
	6-β'-propionate	0.67	0.67	0.00
	7-α-propionate	3.76	3.66	–0.10
	7-α'-propionate	1.74	1.18	–0.51
	7-β-propionate	–0.44	–0.64	–0.20
	7-β'-propionate	–0.81	–0.95	–0.14
	α-meso	1.61	1.56	–0.05
	β-meso	0.21	0.64	+0.43
	γ-meso	–1.1	–1.03	+0.07

	δ-meso	0.43	0.67	+0.24
His70	NH	9.90	9.90	0.00
	CαH	6.75	6.90	+0.15
	CβH	9.62	10.15	+0.53
	CβH'	8.92	8.98	+0.06
	NδH	15.0	14.5	–0.5
His46	NH	10.71	10.65	–0.06
	CαH	7.70	7.60	–0.10
	CβH	10.82	10.79	–0.03
	CβH'	9.20	9.08	–0.12
	NδH	13.2	13.4	+0.2
	CεH	–11.6	–11.3	+0.3
axial His	m	11.7 [d]	11.8	
	n	5.4 [d]		
	o	–1.7 [d]	–2.8	

[a] In 95:5 $^{1}H_2O$:$^{2}H_2O$, at 25 °C and pH 7.3. [b] Refer to Figure 2 for heme nomenclature. [c] As in (*4*) in 95:5 $^{1}H_2O$:$^{2}H_2O$, at 25 °C and pH 6.9-7.5. [d] In $^{2}H_2O$, pH*7.2 at 35 °C (*4*).

Data acquired by SAK and BCV, analyzed and interpreted by JTL.

Table 5.2: Thermal denaturation of ferric wild-type and A69S rHb-R

Protein	T_M (°C) [a]	$\Delta H°$ (kJ mol^{-1}) [b]	ΔC_p (kJ mol^{-1} K^{-1})
wt (vis and CD)	73.6 ± 0.2	288 ± 10	7.1 [c]
A69S (vis)	71.15 ± 0.04	284 ± 3	7.1 [c]
A69S (CD)	69.62 ± 0.13	271 ± 6	7.1 [c]
CN-A69S (vis)	73.99 ± 0.05	368 ± 4	5.8 ± 1.8

[a] Midpoint of thermal denaturation in pH 7.2, 20 mM phosphate buffer. Errors reported are standard deviations from global fitting of multiple data sets. [b] Change in enthalpy evaluated at the midpoint of the thermal unfolding transition. [c] Value fixed during fitting.

Data acquired by JAK and SAK, analyzed, and interpreted by JAK.

Table 5.3: Urea denaturation of ferric wild-type and A69S rHb-R

	$\Delta G°$ (kJ mol^{-1}) [a]	*m* (kJ mol^{-1} M^{-1}) [b]	C_m (M) [c]
wt (vis)	30.6 ± 0.3	5.99 ± 0.05	5.11
wt (CD)	27.4 ± 0.3	5.57 ± 0. 06	4.9
A69S (vis)	28.38 ± 0.17	5.94 ± 0.04	4.78
A69S (CD)	28.6 ± 0.4	6.25 ± 0.08	4.6

[a] Free energy of unfolding at 25 °C in pH 7.2, 20 mM phosphate buffer. Errors reported are standard deviations from global fitting of multiple data sets. [b] Dependence of free energy of unfolding on denaturant concentration. [c] Midpoint of urea denaturation, calculated from $\Delta G°$ and *m*.

Data acquired by JAK and SAK, analyzed, and interpreted by JAK.

References

(1) Walker, F. A. (2004) Models of the bis-histidine-ligated electron-transferring cytochromes. Comparative geometric and electronic structure of low-spin ferro- and ferrihemes. *Chem Rev 104*, 589-615.

(2) de Sanctis, D., Dewilde, S., Vonrhein, C., Pesce, A., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T., Ponassi, M., Nardini, M., and Bolognesi, M. (2005) Bishistidyl heme hexacoordination, a key structural property in Drosophila melanogaster hemoglobin. *J Biol Chem 280*, 27222-9.

(3) Couture, M., Das, T. K., Savard, P. Y., Ouellet, Y., Wittenberg, J. B., Wittenberg, B. A., Rousseau, D. L., and Guertin, M. (2000) Structural investigations of the hemoglobin of the cyanobacterium *Synechocystis* PCC 6803 reveal a unique distal heme pocket. *Eur. J. Biochem. 267*, 4770-4780.

(4) Lecomte, J. T. J., Scott, N. L., Vu, B. C., and Falzone, C. J. (2001) Binding of ferric heme by the recombinant globin from the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochemistry 40*, 6541-6552.

(5) Scott, N. L., Falzone, C. J., Vuletich, D. A., Zhao, J., Bryant, D. A., and Lecomte, J. T. J. (2002) The hemoglobin of the cyanobacterium *Synechococcus* sp. PCC 7002: evidence for hexacoordination and covalent adduct formation in the ferric recombinant protein. *Biochemistry 41*, 6902-6910.

(6) Dewilde, S., Kiger, L., Burmester, T., Hankeln, T., Baudin-Creuza, V., Aerts, T., Marden, M. C., Caubergs, R., and Moens, L. (2001) Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family. *J Biol Chem 276*, 38949-38955.

(7) Trent, J. T., 3rd, and Hargrove, M. S. (2002) A ubiquitously expressed human hexacoordinate hemoglobin. *J. Biol. Chem. 277*, 19538-19545.

(8) Burmester, T., Ebner, B., Weich, B., and Hankeln, T. (2002) Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. *Mol Biol Evol 19*, 416-21.

(9) Sawai, H., Kawada, N., Yoshizato, K., Nakajima, H., Aono, S., and Shiro, Y. (2003) Characterization of the heme environmental structure of cytoglobin, a fourth globin in humans. *Biochemistry 42*, 5133-42.

(10) Feng, L., Gell, D. A., Zhou, S., Gu, L., Kong, Y., Li, J., Hu, M., Yan, N., Lee, C., Rich, A. M., Armstrong, R. S., Lay, P. A., Gow, A. J., Weiss, M. J., Mackay, J. P., and Shi, Y. (2004) Molecular mechanism of AHSP-mediated stabilization of alpha-hemoglobin. *Cell 119*, 629-40.

(11) Feng, L., Zhou, S., Gu, L., Gell, D. A., Mackay, J. P., Weiss, M. J., Gow, A. J., and Shi, Y. (2005) Structure of oxidized alpha-haemoglobin bound to AHSP reveals a protective mechanism for haem. *Nature 435*, 697-701.

(12) Weiland, T. R., Kundu, S., Trent, J. T., 3rd, Hoy, J. A., and Hargrove, M. S. (2004) Bis-histidyl hexacoordination in hemoglobins facilitates heme reduction kinetics. *J. Am. Chem. Soc. 126*, 11930-11935.

(13) Hamdane, D., Kiger, L., Dewilde, S., Uzan, J., Burmester, T., Hankeln, T., Moens, L., and Marden, M. C. (2005) Hyperthermal stability of neuroglobin and cytoglobin. *Febs J 272*, 2076-84.

(14) Ptitsyn, O. B., and Ting, K. L. (1999) Non-functional conserved residues in globins and their possible role as a folding nucleus. *J. Mol. Biol. 291*, 671-682.

(15) Knapp, J. E., Bonham, M. A., Gibson, Q. H., Nichols, J. C., and Royer, W. E., Jr. (2005) Residue F4 plays a key role in modulating oxygen affinity and cooperativity in Scapharca dimeric hemoglobin. *Biochemistry 44*, 14419-30.

(16) Cheng, X. D., and Schoenborn, B. P. (1991) Neutron diffraction study of carbonmonoxymyoglobin. *J Mol Biol 220*, 381-99.

(17) Sinclair, R., Hallam, S., Chen, M., Chance, B., and Powers, L. (1996) Active site structure in cytochrome c peroxidase and myoglobin mutants: effects of altered hydrogen bonding to the proximal histidine. *Biochemistry 35*, 15120-8.

(18) Roncone, R., Monzani, E., Murtas, M., Battaini, G., Pennati, A., Sanangelantoni, A. M., Zuccotti, S., Bolognesi, M., and Casella, L. (2004) Engineering peroxidase activity in myoglobin: the haem cavity structure and peroxide activation in the T67R/S92D mutant and its derivative reconstituted with protohaemin-l-histidine. *Biochem J 377*, 717-24.

(19) Liong, E. C., Dou, Y., Scott, E. E., Olson, J. S., and Phillips, G. N., Jr. (2001) Waterproofing the heme pocket. Role of proximal amino acid side chains in preventing hemin loss from myoglobin. *J Biol Chem 276*, 9093-100.

(20) Samuni, U., Ouellet, Y., Guertin, M., Friedman, J. M., and Yeh, S. R. (2004) The absence of proximal strain in the truncated hemoglobins from *Mycobacterium tuberculosis*. *J. Am. Chem. Soc. 126*, 2682-2683.

(21) Kundu, S., Snyder, B., Das, K., Chowdhury, P., Park, J., Petrich, J. W., and Hargrove, M. S. (2002) The leghemoglobin proximal heme pocket directs oxygen dissociation and stabilizes bound heme. *Proteins 46*, 268-77.

(22) Hargrove, M. S., and Olson, J. S. (1996) The stability of holomyoglobin is determined by heme affinity. *Biochemistry 35*, 11310-11318.

(23) Inaba, K., Ishimori, K., Imai, K., and Morishima, I. (2000) Substitution of the heme binding module in hemoglobin alpha- and beta-subunits. Implication for different regulation mechanisms of the heme proximal structure between hemoglobin and myoglobin. *J Biol Chem 275*, 12438-45.

(24) Inaba, K., Ishimori, K., and Morishima, I. (1998) Structural and functional roles of heme binding module in globin proteins: identification of the segment regulating the heme binding structure. *J Mol Biol 283*, 311-27.

(25) Smerdon, S. J., Krzywda, S., Wilkinson, A. J., Brantley, R. E., Jr., Carver, T. E., Hargrove, M. S., and Olson, J. S. (1993) Serine92 (F7) contributes to the control of heme reactivity and stability in myoglobin. *Biochemistry 32*, 5132-5138.

(26) Shiro, Y., Iizuka, T., Marubayashi, K., Ogura, T., Kitagawa, T., Balasubramanian, S., and Boxer, S. G. (1994) Spectroscopic study of Ser92 mutants of human myoglobin: hydrogen bonding effect of Ser92 to proximal His93 on structure and property of myoglobin. *Biochemistry 33*, 14986-14992.

(27) Lloyd, E., Burk, D. L., Ferrer, J. C., Maurus, R., Doran, J., Carey, P. R., Brayer, G. D., and Mauk, A. G. (1996) Electrostatic modification of the active site of myoglobin: characterization of the proximal Ser92Asp variant. *Biochemistry 35*, 11901-11912.

(28) Wu, Y., Chien, E. Y., Sligar, S. G., and La Mar, G. N. (1998) Influence of proximal side mutations on the molecular and electronic structure of cyanomet myoglobin: an 1H NMR study. *Biochemistry 37*, 6979-6990.

(29) Wittenberg, J. B., Bolognesi, M., Wittenberg, B. A., and Guertin, M. (2002) Truncated hemoglobins: A new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes and plants. *J. Biol. Chem. 277*, 871-874.

(30) Vuletich, D. A., and Lecomte, J. T. J. (2006) A phylogenetic and structural analysis of truncated hemoglobins. *J. Mol. Evol. in press*.

(31) Milani, M., Pesce, A., Nardini, M., Ouellet, H., Ouellet, Y., Dewilde, S., Bocedi, A., Ascenzi, P., Guertin, M., Moens, L., Friedman, J. M., Wittenberg, J. B., and Bolognesi, M. (2005) Structural bases for heme binding and diatomic ligand recognition in truncated hemoglobins. *J Inorg Biochem 99*, 97-109.

(32) Dantsker, D., Samuni, U., Ouellet, Y., Wittenberg, B. A., Wittenberg, J. B., Milani, M., Bolognesi, M., Guertin, M., and Friedman, J. M. (2004) Viscosity-dependent relaxation significantly modulates the kinetics of CO recombination in the truncated hemoglobin TrHbN from Mycobacterium tuberculosis. *J Biol Chem 279*, 38844-53.

(33) Falzone, C. J., Vu, B. C., Scott, N. L., and Lecomte, J. T. J. (2002) The solution structure of the recombinant hemoglobin from the cyanobacterium *Synechocystis* sp. PCC 6803 in its hemichrome state. *J. Mol. Biol. 324*, 1015-1029.

(34) Hoy, J. A., Kundu, S., Trent, J. T., 3rd, Ramaswamy, S., and Hargrove, M. S. (2004) The crystal structure of *Synechocystis* hemoglobin with a covalent heme linkage. *J. Biol. Chem. 279*, 16535-16542.

(35) Trent, J. T., 3rd, Kundu, S., Hoy, J. A., and Hargrove, M. S. (2004) Crystallographic analysis of synechocystis cyanoglobin reveals the structural changes accompanying ligand binding in a hexacoordinate hemoglobin. *J. Mol. Biol. 341*, 1097-1108.

(36) Vu, B. C., Nothnagel, H. J., Vuletich, D. A., Falzone, C. J., and Lecomte, J. T. (2004) Cyanide binding to hexacoordinate cyanobacterial hemoglobins: Hydrogen bonding network and heme pocket rearrangement in ferric H117A *Synechocystis* Hb. *Biochemistry 43*, 12622-12633.

(37) Hvitved, A. N., Trent, J. T., 3rd, Premer, S. A., and Hargrove, M. S. (2001) Ligand binding and hexacoordination in *Synechocystis* hemoglobin. *J. Biol. Chem. 276*, 34714-34721.

(38) Smagghe, B. J., Sarath, G., Ross, E., Hilbert, J. L., and Hargrove, M. S. (2006) Slow Ligand Binding Kinetics Dominate Ferrous Hexacoordinate Hemoglobin Reactivities and Reveal Differences between Plants and Other Species. *Biochemistry 45*, 561-570.

(39) Vu, B. C., Jones, A. D., and Lecomte, J. T. J. (2002) Novel histidine-heme covalent linkage in a hemoglobin. *J. Am. Chem. Soc. 124*, 8544-8545.

(40) La Mar, G. N., Satterlee, J. D., and de Ropp, J. S. (1999) Nuclear magnetic resonance of hemoproteins, in *The Porphyrin Handbook* (Smith, K. M., Kadish, K., and Guilard, R., Eds.) pp 185-298, Academic Press, Burlington, MA.

(41) Shiro, Y., Iizuka, T., Makino, R., Ishimura, Y., and Morishima, I. (1989) ^{15}N NMR study of cyanide ($C^{15}N$) complex of cytochrome P-450_{cam}. Effects of *d*-camphor and putidaredoxin on iron-ligand structure. *J. Am. Chem. Soc. 111*, 7707-7711.

(42) Mukhopadhyay, K., and Lecomte, J. T. (2004) A relationship between heme binding and protein stability in cytochrome b5. *Biochemistry 43*, 12227-36.

(43) Creighton, T. E., and Shortle, D. (1994) Electrophoretic characterization of the denatured states of staphylococcal nuclease. *J Mol Biol 242*, 670-82.

(44) Vu, B. C., Vuletich, D. A., Kuriakose, S. A., Falzone, C. J., and Lecomte, J. T. (2004) Characterization of the heme-histidine cross-link in cyanobacterial hemoglobins from *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7002. *J. Biol. Inorg. Chem. 9*, 183-194.

(45) Milani, M., Ouellet, Y., Ouellet, H., Guertin, M., Boffi, A., Antonini, G., Bocedi, A., Mattu, M., Bolognesi, M., and Ascenzi, P. (2004) Cyanide binding to truncated hemoglobins: a crystallographic and kinetic study. *Biochemistry 43*, 5213-5221.

(46) Peterson, E. S., Friedman, J. M., Chien, E. Y., and Sligar, S. G. (1998) Functional implications of the proximal hydrogen-bonding network in myoglobin: a resonance Raman and kinetic study of Leu89, Ser92, His97, and F-helix swap mutants. *Biochemistry 37*, 12301-12319.

(47) Bertini, I., Luchinat, C., Parigi, G., and Walker, F. A. (1999) Heme methyl ^{1}H chemical shifts as structural parameters in some low-spin ferriheme proteins. *J. Biol. Inorg. Chem. 4*, 515-519.

(48) Koshy, T. I., Luntz, T. L., Schejter, A., and Margoliash, E. (1990) Changing the invariant proline-30 of rat and Drosophila melanogaster cytochromes c to alanine or valine destabilizes the heme crevice more than the overall conformation. *Proc Natl Acad Sci U S A 87*, 8697-701.

(49) Garcia, P., Bruix, M., Rico, M., Ciofi-Baffoni, S., Banci, L., Ramachandra Shastry, M. C., Roder, H., de Lumley Woodyear, T., Johnson, C. M., Fersht, A. R., and Barker, P. D. (2005) Effects of heme on the structure of the denatured state and folding kinetics of cytochrome b562. *J Mol Biol 346*, 331-44.

(50) Mintorovitch, J., van Pelt, D., and Satterlee, J. D. (1989) Kinetic study of the slow cyanide binding to Glycera dibranchiata monomer hemoglobin components III and IV. *Biochemistry 28*, 6099-104.

(51) Vallone, B., Nienhaus, K., Matthes, A., Brunori, M., and Nienhaus, G. U. (2004) The structure of carbonmonoxy neuroglobin reveals a heme-sliding mechanism for control of ligand affinity. *Proc. Natl. Acad. Sci. U. S. A. 101*, 17351-6.

Chapter 6

Conclusions and Future Directions

b Hemoproteins represent a class of proteins with diverse functions that are controlled chiefly by the protein environment. We have examined two members of this family, rat microsomal cytochrome b_5 (cyt b_5) and the truncated hemoglobin from the cyanobacterium *Synechocystis* sp. PCC 6803 (S6803 rHb-R), in an attempt to discern the role that both structured and unstructured elements play in determining thermodynamic and kinetic properties. The heme-binding region of cyt b_5 was grafted into an alternate scaffold to assess its effect on apoprotein characteristics as well as its ability to bind heme in a non-native environment. Additionally, a potential hydrogen-bond donor was introduced into S6803 rHb-R to examine the function of interactions in the proximal heme pocket in determining holoprotein properties in both the hemichrome and ligand-bound states.

The results obtained with the EbE proteins, which contained the heme-binding region of cyt b_5 inserted into the SH3 fold of PsaE, provided a better understanding of the properties of the parent proteins and proteins with large loops in general. The ability of PsaE to accommodate as many as 40 extra residues without extreme thermodynamic or structural consequences indicated that this particular scaffold was resilient. The available data also suggested that the cytochrome's binding region may not destabilize significantly the apoprotein core-2 fold; rather, it appears to have a low intrinsic stability. In addition, it was determined that the subset of residues chosen from the cytochrome b_5 protein were unable to bind heme in the background of the PsaE scaffold. The possibility exists that some communication between the two hydrophobic cores of the protein is

required for function, and core 2 has a dual role rather than serving merely a structural purpose. Also, it may be that the PsaE scaffold prevented the heme-binding region from folding correctly around the heme. Finally, the EbE4 results provided some clues about the importance of the loop-scaffold interface in proteins that contain long stretches of unstructured amino acids. We have shown here that the nature of the boundary between structured and disordered protein segments has the ability to dictate several biophysical properties including free energy of folding, protein aggregation propensity, population of multiple native states, and folding kinetics. This may indicate that, as expected, evolution of the joining region is critical in maintaining proper protein function.

Several questions regarding the cytochrome b_5 heme-binding region and long loops in general remain unanswered. It is unknown what portion of the cytochrome is necessary for specific heme-protein interactions to occur and whether or not a shortened version of the protein would be able to perform the physiologically relevant electron-transfer chemistry. Small *b* heme-binding model systems have been created in an attempt to determine the minimum requirements for stable complex formation (*1, 2*), but the ability of these holoproteins to perform chemistry has not been explored extensively. Studies on a proteolytic fragment of Mb have shown that ability to bind the heme group and the physiologically relevant ligand are not the only considerations (*3*). The conformational and thermodynamic stabilities of the complex are also important. Hence, an investigation of the minimal functional portion of cytochrome b_5 would add to the available database of heme-binding model systems and aid in the development of protein engineering principles.

In terms of proteins containing large flexible regions, a more extensive set of proteins could be examined in order to determine the range of consequences of loop insertion. It is clear that the effect of a disordered region depends on the natures of both the structured and unstructured portions of the protein. A side-by-side investigation into the average number of residues on either side of the interface between the loop and scaffold that are able to dictate a protein's biophysical properties and the conservation of those residues among an evolutionarily-related family of proteins would be instructive. Further analysis of the ability or inability of certain loops to form residual structure at various insertion points within the same scaffold would take the opposite approach as that taken in the EbE proteins and inform on the effects of the scaffold on the loop. Each of these studies could contribute to the overall knowledge of how proteins fold into their native structures (assuming that the proteins are folded) and how the roughness of the protein-folding landscape depends on the particular residues present.

In addition to flexible apoprotein regions, ordered holoprotein components also play a large role in modulating function. We have determined the effects of a conservative proximal modification on the specific heme affinity and ligand-binding properties of S6803 rHb-R. In contrast to the Mb case, no evidence for formation of a side-chain side-chain hydrogen bond was found upon introduction of a serine immediately before the proximal histidine. In addition, the data do not support a Mb-like orientation of the proximal histidine with respect to the heme plane. Rather, the A69S substitution resulted in modest structural changes, enhanced decoupling of the losses of secondary structure and heme iron ligation, and moderate decreases in resistance to temperature-, urea-, and acid-induced denaturation as well as changes in exogenous ligand-binding kinetics. It

appeared that the proximal network present in wild-type S6803 rHb-R was optimized for the specific combination of residues found in this particular protein. This may be a general property of this class of proteins, because substitution at the analogous position in other systems has led to decreased heme affinity regardless of whether hydrogen-bonding interactions were being introduced or removed. Finally, the increased thermal stability of S6803 rHb-R in the ligand-bound state relative to the hemichrome state indicated that the relative proximal and distal pocket interactions and coordination bond strengths had been optimized for this form. Depending on the physiological role ultimately determined for S6803 rHb-R, this may become an important characteristic. Further investigation of the function of this and other truncated hemoglobins (many of which have yet to be structurally characterized or even overexpressed) could aid in developing a more complete understanding of this newly-discovered class of proteins and complement the available sequence alignment studies (*4, 5*).

References

(1) Discher, B. M., Noy, D., Strzalka, J., Ye, S., Moser, C. C., Lear, J. D., Blasie, J. K., and Dutton, P. L. (2005) Design of amphiphilic protein maquettes: controlling assembly, membrane insertion, and cofactor interactions. *Biochemistry 44*, 12329-43.

(2) Gibney, B. R., Isogai, Y., Rabanal, F., Reddy, K. S., Grosset, A. M., Moser, C. C., and Dutton, P. L. (2000) Self-assembly of heme A and heme B in a designed four-helix bundle: implications for a cytochrome c oxidase maquette. *Biochemistry 39*, 11041-9.

(3) Di Iorio, E. E., Yu, W., Calonder, C., Winterhalter, K. H., De Sanctis, G., Falcioni, G., Ascoli, F., Giardina, B., and Brunori, M. (1993) Protein dynamics in minimyoglobin: is the central core of myoglobin the conformational domain? *Proc Natl Acad Sci U S A 90*, 2025-9.

(4) Vuletich, D. A., and Lecomte, J. T. (2006) A phylogenetic and structural analysis of truncated hemoglobins. *J Mol Evol 62*, 196-210.

(5) Wittenberg, J. B., Bolognesi, M., Wittenberg, B. A., and Guertin, M. (2002) Truncated hemoglobins: a new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants. *J Biol Chem 277*, 871-4.

Appendix A

EbE1-4 Gene Construction and Protein Sequences

A.1. EbE Gene Construction

A.1.1. Oligonucleotides Used to Create EbE1 Gene

Uppercase letters indicate restoration of the NcoI and BamHI restriction sites

Oligo	Coding Strand (5' → 3')
1	CATGGctatcgaacgcggatccaaagttaaaattctccggaaagaatcc
2	tactggtacggcgatgtcggcacagtagcaagcattgataaaagcggtatt
3	atctatcccgtcattgttcgctttaacctaactaaattcctcgaagagcac
4	cccgggggcgaagaagtcctgagagaacaggccggcggtgacgcgactgaa
5	aacttcgaagacgttggccatagtaccgacgctcgagaactgtcgaaaacg
6	tacaacaccaacaacttcgctgagcatgaactcgaagtcgttggctgataG

Oligo	Noncoding Strand (5' → 3')
7	GATCCtatcagccaacgacttcgagttcatgctcagcgaagtt
8	gttggtgttgtacgttttcgacagttctcgagcgtcggtactatggccaac
9	gtcttcgaagttttcagtcgcgtcaccgccggcctgttctctcaggacttc
10	ttcgcccccggggtgctcttcgaggaatttagttaggttaaagcgaacaat
11	gacgggatagataataccgcttttatcaatgcttgctactgtgccgacatc
12	gccgtaccagtaggattctttccggagaattttaactttggatccgcgttcgatagC

A.1.2. Primers Used to Create EbE2,3 Genes

Primers used in the creation of EbE2, EbE3 genes from the EbE1 gene using Stratagene QuikChange Mutagenesis protocol. Highlighted regions indicate insertion of Gly codon; the EbE1 oligo into which the Gly was inserted is parenthetically noted. All primers are written 5'→ 3'.

EbE2:
ccgtcattgttcgctttaacgggctaactaaattcc (3)

ggaatttagttagcccgttaaagcgaacaatgacgg (10)

EbE3:
ccgtcattgttcgctttaacgggctaactaaattcc (3)
ctgtcgaaaacgtacgggaacaccaacaacttcgc (6)

gcgaagttgttggtgttcccgtacgttttcgacag (10)
ggaatttagttagcccgttaaagcgaacaatgacgg (8)

A.1.3. Oligonucleotides Used to Create EbE4 Gene

For the coding strand, EbE1 oligonucleotides 1,2,4,5 were reused. Alternate oligos 3 and 6 are shown below. Highlighted regions depict changes from the EbE1 sequence. All primers are written 5'→ 3'.

Oligo 3:
atctatcccgtcattgttcgctttaaagtatacgacctaactaaattcctcgaagagcac

Oligo 6:
tacatcatcaacttcgctgagcatgaactcgaagtcgttggctgataG

For the noncoding strand, EbE1 oligonucleotides 7,9,11,12 were reused. Alternate oligos 8 and 10 are shown below. Highlighted regions depict changes from the EbE1 sequence. All primers are written 5'→ 3'.

Oligo 8:
gatgatgtacgttttcgacagttctcgagcgtcggtactatggccaac

Oligo 10:
ttcgcccccggggtgctcttcgaggaatttagttaggtcgtatactttaaagcgaacaat

A.2. EbE Protein Sequences

```
Cyt b5 ---------A EQSDKDVKYY TLEEIQKHKD SKSTWVILHH KVYD-LTKFL EEHPGGEEVL
PsaE   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF N----KVNYN GFSGSAGGL-
EbE1   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF N----LTKFL EEHPGGEEVL
EbE2   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF N---GLTKFL EEHPGGEEVL
EbE3   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF N---GLTKFL EEHPGGEEVL
EbE4   AIERGSKVKI LRKESYWYGD VGTVASIDKS GIIYPVIVRF KVYD-LTKFL EEHPGGEEVL

Cyt b5 REQAGGDATE NFEDVGHSTD ARELSKTY-I IGELHPDDRS KIAKPSETL
PsaE   ---------- ---------- ---------- -NTNNFAEHE LEVVG----
EbE1   REQAGGDATE NFEDVGHSTD ARELSKTY-- -NTNNFAEHE LEVVG----
EbE2   REQAGGDATE NFEDVGHSTD ARELSKTY-- -NTNNFAEHE LEVVG----
EbE3   REQAGGDATE NFEDVGHSTD ARELSKTYG- -NTNNFAEHE LEVVG----
EbE4   REQAGGDATE NFEDVGHSTD ARELSKTY-I I---NFAEHE LEVVG----
```

Red letters indicate the portions of the cytochrome b_5 heme-binding loop used in the construction of the EbE proteins. Blue letters indicate the SH3 scaffold of PsaE used for the EbE proteins.

Protein	PsaE N term	Cytochrome b_5	PsaE C term
EbE1	1 – 41	33 – 75	56 – 69
EbE2	1 – 41 + Gly	33 – 75	56 – 69
EbE3	1 – 41 + Gly	33 – 75	Gly + 56 – 69
EbE4	1 – 40	28 – 76	59 – 69

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	^{1}H chemical shifts for the EbE1 protein																		
2	pH	7.2																	
3	T	292																	
4																			
5	Residue	NH	aH1	aH2	bH1	bH2	g1H1	g1H2	g1H3	g2H3	d1H1	d1H2	d1H3	d2H3	eH	eNH1	zH	z3H	hH
6	A1E																		
7	I2E																		
8	E3E	8.32	4.55																
9	R4E	8.89	3.62																
10	G5E	9.10	4.45	3.42															
11	S6E	8.25	4.39																
12	K7E	8.80	5.08																
13	V8E	8.73	5.31		1.70				0.60	0.49									
14	K9E	9.45	5.07		1.47	1.67													
15	I10E	8.19	3.83		2.58		1.04	1.78		0.93			-0.08						
16	L11E	8.94	4.62		1.80	1.53	1.52						0.73						
17	R12E	6.62	4.48								3.14					8.66			
18	K13E	8.63	2.97																
19	E14E																		
20	S15E																		
21	Y16E	8.66	3.84		1.64	2.49					5.81				6.31				
22	W17E	8.02	4.36		2.83	2.07					6.89				7.21	9.70	7.16	6.89	6.98
23	Y18E	7.25	3.90			3.77					6.88				6.76				
24	G19E	8.66	3.86	3.14															
25	D20E																		
26	V21E	8.62	5.21		1.87				0.77	0.87									
27	G22E	8.79	4.50	2.79															
28	T23E	8.21	5.03		3.60					0.98									
29	V24E	9.14	3.74		2.29				0.62	0.76									
30	A25E	9.58	4.35		1.26														
31	S26E	7.63	4.35																
32	I27E	8.32	4.77																
33	D28E	9.43	4.61																
34	K29E	8.85	4.51																
35	S30E	8.95	4.35																
36	G31E	8.64	3.69	4.20															

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
37	I32E	7.21	4.31		2.04														
38	I33E	7.95	3.76		1.51		0.91	1.09		0.39			0.70						
39	Y34E	7.61	4.80		2.31	2.82					7.08				6.80				
40	P35E		4.42		2.30		2.19	2.01			3.76	3.88							
41	V36E	8.80	4.27		2.09				0.72	0.85									
42	I37E	8.62	4.54																
43	V38E	8.75	4.19		1.62				-0.24	0.63									
44	R39E	8.73	4.57		1.46	1.63													
45	F40E	8.67	4.62								6.40				6.46		6.59		
46	N41E	8.94	4.80		2.94	2.88													
47	L32b																		
48	T33b																		
49	K34b																		
50	F35b																		
51	L36b																		
52	E37b																		
53	E38b																		
54	H39b																		
55	P40b																		
56	G41b																		
57	G42b																		
58	E43b																		
59	E44b																		
60	V45b																		
61	L46b																		
62	R47b																		
63	E48b																		
64	Q49b																		
65	A50b																		
66	G51b																		
67	G52b																		
68	D53b																		
69	A54b																		
70	T55b																		
71	E56b																		
72	N57b	8.26	4.51																

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
73	F58b	8.32	4.55		2.99	3.14					7.21				7.31		7.25		
74	E59b																		
75	D60b	8.31	4.60		2.72	2.60													
76	V61b	8.10	4.12		2.14				0.90	0.90									
77	G62b	8.50	3.89	3.89															
78	H63b																		
79	S64b																		
80	T65b																		
81	D66b																		
82	A67b																		
83	R68b																		
84	E69b																		
85	L70b																		
86	S71b																		
87	K72b																		
88	T73b																		
89	Y74b										7.04				6.76				
90	N56E																		
91	T57E																		
92	N58E	8.59	4.91																
93	N59E	7.65	5.52		2.77	2.49													
94	F60E	9.52	4.48																
95	A61E	9.59	5.04		1.57														
96	E62E	9.68	3.68		2.14	2.14	2.32												
97	H63E	7.78	4.68		3.07	3.44					7.05				7.82				
98	E64E	7.52	4.33		2.40	2.54													
99	L65E	7.33	5.14		0.90	1.83	1.68						0.60	0.34					
100	E66E	8.52	4.72		1.91	1.98													
101	V67E	9.09	4.28		2.00				1.07	0.97									
102	V68E	8.75	4.50		2.24				0.88	0.69									
103	G69E	7.92	3.72	3.85															

Appendix C
EbE4 chemical shift data

			EbE4-N1	EbE4-N2
1E	ALA	HA	4.07	
1E	ALA	HB	1.46	
2E	ILE	HA	3.76	
3E	GLU	H	8.45	
3E	GLU	HA	4.53	4.54
3E	GLU	HB2	1.77	
3E	GLU	HB3	2.02	
3E	GLU	HG2	2.23	
3E	GLU	HG3	2.23	
3E	GLU	N	125.7	
4E	ARG	H	8.84	8.85
4E	ARG	HA	3.60	3.71
4E	ARG	HB2	1.71	
4E	ARG	HB3	1.66	
4E	ARG	HG2	1.42	
4E	ARG	HG3	1.68	
4E	ARG	N	121.7	122.9
5E	GLY	H	9.08	8.78
5E	GLY	HA2	3.40	3.49
5E	GLY	HA3	4.43	4.35
5E	GLY	N	114.7	114.9
6E	SER	H	8.19	8.07
6E	SER	HA	4.40	4.36
6E	SER	HB2	3.78	3.70
6E	SER	HB3	3.78	3.70
6E	SER	N	117.5	117.4
7E	LYS	H	8.81	8.75
7E	LYS	HA	5.06	5.08
7E	LYS	HB2	1.64	
7E	LYS	HB3	1.47	
7E	LYS	N	125.1	124.6
8E	VAL	H	8.69	8.69
8E	VAL	HA	5.29	5.21
8E	VAL	HB	1.70	
8E	VAL	HG1	0.60	0.56
8E	VAL	HG2	0.50	
8E	VAL	N	114.9	115.3
9E	LYS	H	9.40	9.45
9E	LYS	HA	5.07	4.82
9E	LYS	HB2	1.64	
9E	LYS	HB3	1.50	
9E	LYS	HG2	1.08	
9E	LYS	N	124.1	124.8

10E	ILE	H	8.22	8.37
10E	ILE	HA	3.84	3.65
10E	ILE	HB	2.61	
10E	ILE	HG12	1.80	1.36
10E	ILE	HG13	1.00	
10E	ILE	HG2	0.92	
10E	ILE	HD1	-0.01	0.27
10E	ILE	N	125.0	125.1
11E	LEU	H	8.93	8.89
11E	LEU	HA	4.63	4.62
11E	LEU	HB2	1.47	1.40
11E	LEU	HB3	1.82	1.85
11E	LEU	HG	1.59	
11E	LEU	HD1	0.75	
11E	LEU	HD2	0.90	
11E	LEU	N	125.5	125.7
12E	ARG	H	6.61	6.54
12E	ARG	HA	4.49	4.47
12E	ARG	HB3	1.89	1.89
12E	ARG	N	119.5	119.3
13E	LYS	H	8.59	8.58
13E	LYS	HA	3.08	3.10
13E	LYS	HB2	1.50	
13E	LYS	HB3	1.50	
13E	LYS	HG2	1.28	
13E	LYS	HG3	1.01	
13E	LYS	N	132.5	132.0
14E	GLU	H	8.71	8.80
14E	GLU	HA	4.07	4.10
14E	GLU	HB2	1.50	
14E	GLU	HB3	1.92	
14E	GLU	HG2	2.26	
14E	GLU	HG3	2.05	
14E	GLU	N	115.40	115.80
15E	SER	H	7.50	7.55
15E	SER	HA	4.62	4.66
15E	SER	HB2	3.93	
15E	SER	HB3	3.93	
15E	SER	N	112.40	112.40
16E	TYR	HD1	5.63	
16E	TYR	HE1	6.37	
17E	TRP	H	8.06	
17E	TRP	HA	4.59	
17E	TRP	HB2	2.65	
17E	TRP	HB3	3.54	
17E	TRP	HD1	7.00	
17E	TRP	HE1	9.57	
17E	TRP	HE3	6.98	7.70
17E	TRP	HZ2	6.54	7.50
17E	TRP	HZ3	6.60	7.16
17E	TRP	HH2	6.31	7.25

17E	TRP	N	115.5	
17E	TRP	NE1	127.0	
18E	TYR	H	7.32	
18E	TYR	HA	3.88	3.90
18E	TYR	HB2	3.68	
18E	TYR	HB3	2.71	
18E	TYR	HD1	6.87	
18E	TYR	HE1	6.75	
18E	TYR	N	119.2	
19E	GLY	H	8.72	8.89
19E	GLY	HA2	3.12	3.04
19E	GLY	HA3	3.89	3.89
19E	GLY	N	117.0	117.8
20E	ASP	H	8.61	
20E	ASP	HA	4.89	4.80
20E	ASP	HB2	2.96	
20E	ASP	HB3	2.96	2.84
20E	ASP	N	122.9	
21E	VAL	H	8.71	8.46
21E	VAL	HA	5.30	5.10
21E	VAL	HB	1.92	1.84
21E	VAL	HG1	0.82	0.82
21E	VAL	HG2	0.90	
21E	VAL	N	117.3	117.0
22E	GLY	H	8.78	9.02
22E	GLY	HA2	4.48	4.77
22E	GLY	HA3	2.71	3.37
22E	GLY	N	113.9	113.9
23E	THR	H	7.94	8.03
23E	THR	HA	5.08	5.22
23E	THR	HB	3.64	3.76
23E	THR	HG2	0.99	1.06
23E	THR	N	113.6	114.1
24E	VAL	H	9.13	
24E	VAL	HA	3.78	4.07
24E	VAL	HB	2.27	
24E	VAL	HG1	0.63	
24E	VAL	HG2	0.78	
24E	VAL	N	126.6	
25E	ALA	H	9.72	9.34
25E	ALA	HA	4.35	4.32
25E	ALA	HB	1.24	1.40
25E	ALA	N	136.0	134.2
26E	SER	H	7.66	7.64
26E	SER	HA	4.28	4.30
26E	SER	HB2	3.76	3.86
26E	SER	HB3	3.76	3.86
26E	SER	N	108.3	109.2
27E	ILE	H	8.16	
27E	ILE	HA	4.80	
27E	ILE	HB	1.72	

27E	ILE	HG12	1.36		
27E	ILE	HG13	1.05		
27E	ILE	HG2	0.79		
27E	ILE	HD1	0.75		
27E	ILE	N	120.5		
28E	ASP	H	9.60		
28E	ASP	HA	4.60		
28E	ASP	HB2	2.54		
28E	ASP	HB3	2.83		
28E	ASP	N	130.0		
29E	LYS	H	8.75		
29E	LYS	HA	4.50		
29E	LYS	HB2	1.69		
29E	LYS	HB3	2.12		
29E	LYS	HG2			
29E	LYS	HG3	1.45		
29E	LYS	N	125.1		
30E	SER	H	8.91		
30E	SER	HA	4.33		
30E	SER	HB2	4.02		
30E	SER	HB3	4.02		
30E	SER	N	117.0		
31E	GLY	H	8.71		
31E	GLY	HA2	4.21		
31E	GLY	HA3	3.69		
31E	GLY	N	111.9		
32E	ILE	H	7.18		
32E	ILE	HA	4.30		
32E	ILE	HB	2.03		
32E	ILE	HG12			
32E	ILE	HG13			
32E	ILE	HG2			
32E	ILE	HD1		0.67	
32E	ILE	N	115.3		
33E	ILE	H	7.89		
33E	ILE	HA	3.80		
33E	ILE	HB	1.54		
33E	ILE	HG12	1.07		
33E	ILE	HG13	0.94		
33E	ILE	HG2	0.41		
33E	ILE	HD1	0.70		
33E	ILE	N	119.2		
34E	TYR	H	7.47		
34E	TYR	HA	4.85		
34E	TYR	HB2	2.88		
34E	TYR	HB3	2.30		
34E	TYR	HD1	7.09		
34E	TYR	HE1	6.83		
34E	TYR	N	117.2		
35E	PRO	HA			
35E	PRO	HB2			158

35E	PRO	HB3		
35E	PRO	HG2		
35E	PRO	HG3		
35E	PRO	HD2	3.87	
35E	PRO	HD3	3.75	
36E	VAL	H	8.75	
36E	VAL	HA	4.16	
36E	VAL	HB	2.03	
36E	VAL	HG1	0.82	
36E	VAL	HG2		
36E	VAL	N	120.0	
37E	ILE	H	8.5	
37E	ILE	HA	4.11	
37E	ILE	HB	1.85	
37E	ILE	HG12	1.48	
37E	ILE	HG13	1.03	1.03
37E	ILE	HG2	0.65	0.84
37E	ILE	HD1	0.75	0.67
37E	ILE	N	129.5	
38E	VAL	H	9.08	9.19
38E	VAL	HA	4.16	4.46
38E	VAL	HB	1.96	2.30
38E	VAL	HG1	-0.12	0.56
38E	VAL	HG2	0.74	1.03
38E	VAL	N	130.5	127.8
39E	ARG	H	8.77	9.27
39E	ARG	HA	5.13	4.58
39E	ARG	HB2		
39E	ARG	HB3	1.59	1.72
39E	ARG	N	124.4	126.5
40E	PHE	H	9.30	9.13
40E	PHE	HA	4.80	4.66
40E	PHE	HB2	2.16	
40E	PHE	HB3	2.89	
40E	PHE	HD1	6.50	
40E	PHE	HE1	6.58	
40E	PHE	HZ	6.73	
40E	PHE	N	126.2	125.0
28b	LYS	H	8.35	
28b	LYS	HA	4.96	
28b	LYS	HG2	1.56	
28b	LYS	HG3	1.56	
28b	LYS	N	121.50	
29b	VAL	H	8.82	
29b	VAL	HA	4.68	
29b	VAL	HB	1.89	
29b	VAL	HG1	0.65	
29b	VAL	HG2	0.56	
29b	VAL	N	125.00	
30b	TYR	H	8.68	
30b	TYR	HA	4.54	

30b	TYR	HB2	3.13	
30b	TYR	HB3	2.89	2.96
30b	TYR	HD1	7.18	7.18
30b	TYR	HE1		
30b	TYR	N	125.70	
32b	LEU	H	7.72	7.67
32b	LEU	HA	4.04	
32b	LEU	HB2		
32b	LEU	HB3		
32b	LEU	HG		
32b	LEU	HD1		
32b	LEU	HD2		
32b	LEU	N	122.20	121.40
33b	THR	H	8.14	8.24
33b	THR	HA	3.90	
33b	THR	HB	4.28	
33b	THR	HG2	1.23	
33b	THR	N	114.50	114.80
34b	LYS	H	7.97	
35b	PHE	H	8.14	
35b	PHE	HA	4.31	
35b	PHE	HB2	3.16	
35b	PHE	HB3		
35b	PHE	HD1	6.95	
35b	PHE	HE1		
35b	PHE	HZ		
35b	PHE	N	119.80	
36b	LEU	H	8.59	
36b	LEU	HA	4.05	
36b	LEU	HB2		
36b	LEU	HB3		
36b	LEU	HG		
36b	LEU	HD1		
36b	LEU	HD2		
36b	LEU	N	120.30	
37b	GLU	H	7.70	
37b	GLU	HA	4.01	
37b	GLU	HB2	2.04	
37b	GLU	HB3	2.19	
37b	GLU	HG2		
37b	GLU	HG3	2.44	
37b	GLU	N	119.20	
38b	GLU	H	7.48	
38b	GLU	HA	4.10	
38b	GLU	HB2	1.73	
38b	GLU	N	115.40	
54b	ALA	H	8.30	8.24
54b	ALA	HA	4.33	
54b	ALA	HB	1.38	1.40
54b	ALA	N	124.50	124.70
55b	THR	H	8.11	8.14

55b	THR	HA	4.20		
55b	THR	HB	4.23		
55b	THR	HG2	1.16		
55b	THR	N	112.70	113.00	
58b	PHE	H	8.20		
58b	PHE	HA	4.55		
58b	PHE	HB2	2.96		
58b	PHE	HB3	3.15		
58b	PHE	HD1			
58b	PHE	HE1			
58b	PHE	N	120.70		
59b	GLU	H	8.29		
59b	GLU	HA	4.20		
59b	GLU	HB2	1.89		
59b	GLU	N	121.40		
60b	ASP	H	8.27		
60b	ASP	HA	4.59		
60b	ASP	HB2	2.70		
60b	ASP	HB3	2.56		
60b	ASP	N	121.20		
61b	VAL	H	8.03		
61b	VAL	HA	4.09		
61b	VAL	HB	2.12		
61b	VAL	HG1			
61b	VAL	HG2			
61b	VAL	N	119.60		
62b	GLY	H	8.45	8.44	
62b	GLY	HA2	3.88		
62b	GLY	HA3			
62b	GLY	N	111.30	111.10	
72b	LYS	HA	4.48		
73b	THR	H	8.06		
73b	THR	HA	5.11		
73b	THR	HB	3.81		
73b	THR	HG2	0.99		
73b	THR	N	116.30		
74b	TYR	H	8.72		
74b	TYR	HA	4.86		
74b	TYR	HB2	2.51		
74b	TYR	HB3	2.91		
74b	TYR	HD1	6.90		
74b	TYR	HE1			
74b	TYR	N	124.00		
75b	ILE	H	8.49		
75b	ILE	HA	5.21	4.88	
75b	ILE	HB	1.62		
75b	ILE	HG12			
75b	ILE	HG13			
75b	ILE	HG2	0.65		
75b	ILE	HD1			
75b	ILE	N	119.90		161

76b	ILE	H	8.90	8.94
76b	ILE	HA	4.31	4.11
76b	ILE	HB	1.78	1.36
76b	ILE	HG12		
76b	ILE	HG13		
76b	ILE	HG2	0.50	0.25
76b	ILE	HD1		
76b	ILE	N	126.50	128.70
59E	ASN	H	7.99	8.24
59E	ASN	HA	5.40	5.21
59E	ASN	HB2	2.46	2.14
59E	ASN	HB3	2.91	2.54
59E	ASN	HD21	7.57	7.19
59E	ASN	HD22	6.77	6.44
59E	ASN	N	123.80	120.90
59E	ASN	ND2	112.14	110.02
60E	PHE	H	9.53	8.78
60E	PHE	HA	4.50	4.71
60E	PHE	HB2	2.38	2.60
60E	PHE	HB3	3.35	3.22
60E	PHE	HD1	7.09	7.11
60E	PHE	HE1	7.07	
60E	PHE	HZ		
60E	PHE	N	119.00	120.90
61E	ALA	H	9.53	9.29
61E	ALA	HA	5.01	4.99
61E	ALA	HB	1.54	1.35
61E	ALA	N	122.50	123.50
62E	GLU	H	9.58	8.80
62E	GLU	HA	3.67	3.63
62E	GLU	HB2	2.19	
62E	GLU	HB3	2.10	2.26
62E	GLU	HG2	2.26	2.41
62E	GLU	N	120.70	120.30
63E	HIS	H	7.74	7.58
63E	HIS	HA	4.67	4.68
63E	HIS	HB2	3.34	3.39
63E	HIS	HB3	3.07	3.07
63E	HIS	HD2	7.04	7.02
63E	HIS	HE1	7.80	7.73
63E	HIS	N	110.50	109.20
64E	GLU	H	7.41	7.68
64E	GLU	HA	4.31	4.22
64E	GLU	HB2	2.47	
64E	GLU	HB3	2.47	
64E	GLU	HG2	1.97	
64E	GLU	HG3	1.65	
64E	GLU	N	119.20	119.30
65E	LEU	H	7.24	7.23
65E	LEU	HA	5.12	5.11
65E	LEU	HB2	1.78	

65E	LEU	HB3	0.92	0.81
65E	LEU	HG	1.66	1.60
65E	LEU	HD1	0.60	
65E	LEU	HD2	0.28	
65E	LEU	N	118.20	118.20
66E	GLU	H	8.52	8.39
66E	GLU	HA	4.70	4.66
66E	GLU	HB2	1.91	
66E	GLU	HB3	1.91	
66E	GLU	HG2		
66E	GLU	HG3	2.18	
66E	GLU	N	121.70	121.30
67E	VAL	H	9.01	8.98
67E	VAL	HA	4.29	4.17
67E	VAL	HB	2.00	1.97
67E	VAL	HG1	1.11	1.02
67E	VAL	HG2	0.96	0.92
67E	VAL	N	128.10	128.00
68E	VAL	H	8.73	8.68
68E	VAL	HA	4.50	4.43
68E	VAL	HB	2.25	2.17
68E	VAL	HG1	0.89	0.88
68E	VAL	HG2	0.70	0.70
68E	VAL	N	123.30	123.70
69E	GLY	H	7.88	7.89
69E	GLY	HA2	3.86	3.82
69E	GLY	HA3	3.72	3.82
69E	GLY	N	117.70	117.90

VITA
Jane A. Knappenberger

GRADUATE EDUCATION
The Pennsylvania State University
Fall 2001 – Spring 2006
Advisor: Juliette T. J. Lecomte
Awards: Geiger Graduate Fellowship 2005-2006
Roberts Graduate Fellowship. 2001-2003.
Travel Award: The Pennsylvania State University Center for Biomolecular Structure and Function NSF Research and Training Group. 2001, 2003.

UNDERGRADUATE EDUCATION
Immaculata College
Fall 1997- Fall 2000
B.A. Chemistry, B.A. Mathematics
Summa Cum Laude, Honors Program
Awards: The Wendall P. MacIntosh Medal for Chemistry
The Anderson Medal for Mathematics
American Chemical Society, Philadelphia Section Annual Award, 2001.
American Chemical Society, Polymer Education Committee Award for Outstanding Performance in Organic Chemistry, 1999.
Who's Who Among College Students in American Colleges and Universities
Pennsylvania Athletic Conference All-Conference Academic Team
Immaculata College Presidential Scholarship

RESEARCH EXPERIENCE
July 2001 – April 2006
The Pennsylvania State University; Advisor: Juliette T. J. Lecomte
January 2001 - June 2001
The Pennsylvania State University; Advisor: C. Robert Matthews
May 2000 - August 2000
NSF Research Experience for Undergraduates (REU) Program
The Pennsylvania State University; Advisor: C. Robert Matthews
August 1999 - May 2000
Independent Study funded by NIH through Immaculata College's office of Sponsored Research
Immaculata College; Advisor: Kathryn A. Lysko
May 1999 - August 1999
NSF REU Program
Bowling Green State University; Advisor: John R. Cable

TEACHING EXPERIENCE
August 2001 - December 2001
The Pennsylvania State University
Teaching Assistant; Principles of Chemistry

PUBLICATIONS
Knappenberger, J.A. and Lecomte, J.T.J. Control of thermodynamic and kinetic properties in a family of PsaE-cytochrome b_5 chimeras via minor loop modifications. (Manuscript in preparation)
Knappenberger, J.A., Kuriakose, S.A., Nothnagel, H.J., Vu, B.C., Vuletich, D.A., and Lecomte, J.T.J. Proximal influences in two-on-two globins: Effect of the Ala69Ser replacement on *Synechocystis* sp. PCC 6803 hemoglobin. *Biochemistry*. (submitted)
Knappenberger, J.A., Kraemer-Pecore, C.M., and Lecomte, J.T.J. Insertion of the cytochrome b_5 heme-binding loop into an SH3 domain. Effects on structure and stability, and clues about the cytochrome's architecture. (2004) *Prot. Sci.* 13: 2899-2908.
Knappenberger, J.A., Smith, J.E., Thorpe, S.H., Zitzewitz, J.A., and Matthews, C.R. A Buried Polar Residue in the Hydrophobic Interface of the Coiled-coil Peptide, GCN4-p1, Plays a Thermodynamic, not a Kinetic Role in Folding. (2002) *J. Mol. Biol.* 321(1): 1-6.

AFFILIATIONS
Protein Society, Sigma Zeta Honor Society, Kappa Gamma Pi Honor Society, Delta Epsilon Sigma Honor Society