MEMBRANE PROTEIN FOLDING:
THE ROLE OF AMINO ACID SEQUENCE IN SPECIFYING THE STABILITY AND CONFORMATION OF TRANSMEMBRANE OLIGOMERS

by

Abigail E. Kroch

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
February 2006

ABSTRACT

Proteins carry out essential processes of the cell, and to function properly they must adopt and maintain their native fold. Both thermodynamic and structural data for soluble protein have provided insight into the forces that drive folding in an aqueous environment. However, structural and energetic characterization for membrane proteins has lagged behind that of soluble proteins. Therefore, membrane protein model systems have been extremely valuable to understand the interactions that stabilize and specify membrane protein conformation. Early thermodynamic studies on membrane proteins demonstrated that interactions between helices promote and specify the native fold; and a large scale mutagenesis of a transmembrane dimer, Glycophorin A (GpA), showed that the amino acid sequence on a single face of the helix conferred stability to the dimer. When the NMR structure was solved, the motif highlighted by mutagenesis was found at the dimer interface. Therefore, preferential interactions at the dimer interface may stabilize and specify the native fold of a membrane protein.

When this thesis study was commenced, interactions in a single sequence motif were thought to provide the driving force for association in the GpA dimer. However, in our studies, we have demonstrated that this motif, albeit critical for strong dimerization, is neither necessary nor sufficient to drive association. This work highlights that the entire sequence context provides the framework for the stability and specificity of transmembrane helix-helix interactions. The correlation of structural models to energetic measurements strongly suggests that local

packing defects may be responsible for the perturbation upon amino acid substitution. However, upon experimental examination of structural changes, packing defects may also initiate global conformation change in the dimer and therefore these mutations would perturb interactions between the helices and between the helices and solvating lipids. The information derived from these studies on a model membrane protein dimer was also applied to transmembrane segments that are involved in vesicle fusion. The association of these proteins, syntaxin and synaptobrevin, may be essential to the process of fusion. Both homo-dimerization and hetero-dimerization of these proteins is weak but is still modulated by the amino acid sequence. In this case, a dynamic equilibrium for protein-protein interactions may be critical for biological function. Therefore, amino acid sequences can encode both stable and transient protein interactions that are biologically relevant. Thus, when we study membrane protein structure and interactions, it is necessary to consider the forces that both drive and repel folding, since it is the equilibrium of these forces that establishes biologically functional proteins.

Thesis Advisor: Dr. Karen G. Fleming

Secondary Reader: Dr. Bertrand Garcia-Moreno

Thesis Committee: Dr. George Rose

Dr. Joel Schildbach

To my Opa,

Morris Adolph Kroch,

Whose courage and strength allowed my very existence.

ACKNOWLEDGEMENTS

This work would not have been possible without the help and support of my friends and family. I thank my mother and father for their encouragement and endorsement of everything I have done, and for reminding me that the most important thing is the work, and especially for my father's advice and assistance in the completion of this thesis.

I thank my sisters, Miriam and Deborah, who are always available for support, encouragement, and a distracting conversation. And, my nieces Katie and Sarah, and my nephew Jake, who provide endless entertainment and remind me of the things that are truly important. My grandfather and my late grandmother, Sam and Jane Burgess, who have always demonstrated their pride and support through the years.

To all my friends, for without them, I could not have made it through, and especially could not have had so much fun in graduate school. I especially thank, Eric Davison, who is always willing to feed me, listen to me, and support me, more than I could truly deserve. I thank Alex Ebie, my labmate and confidant, who has helped me through the most trying times. I also thank my sister moon, Kittiya Lee; Michelle Beaucher, Prashanth Rangan, Naveen Michaud, Katherine Tripp, Mark Zweifel, Naomi Courtemanche, Guru Thuduppathy, Timo Street, Judy Gerring, Kate Murphy, Scott Jackson, Susu Hunniecutt, Youjin Lee, Alana McGill, Kedy'ky Sherrill, and her dear grandmother Dessie Mae Tyler, who I will always miss; and, all the members of the Biology class that entered in 2000. I also thank the Midtown Yoga Studio, and all the

teachers there for their support and help in the advancement in my yoga practice, which undoubtedly helped in the completion of this work.

I would like to thank the faculty of the Biology and Jenkins departments for their mentorship and support over the years. I thank Allen Shearn, who acted as my advocate; Doug Barrick, a wonderful mentor, karaoke singer, and an inspiration as a scientist. I especially thank the members of my committee, Bertrand Garcia-Moreno, Neil Clarke, and George Rose for being supportive, and providing me with sound advice and creative ideas. I thank all the support staff at Hopkins, especially Jerry Levin, Joan Miller, and Ken Rutledge, and all the staff of the Biology and Jenkins Departments, the guys in the Mudd Stockroom, the delivery guys, and the housekeeping staff. I especially thank Deborah DiLazzero, who has been instrumental in all my recent achievements.

I must thank all the members of the Fleming Lab, who have made coming to work stimulating and fun, Ann Marie Stanley, who has been my partner in crime over the years and always up for a stimulating scientific conversation; Felix Kobus, our former post doc; and the newer members of our lab, Alex Ebie, Nancy Burgess, and Preston Moon, who keep the lab a fun place to work. I also thank our former technicians, Matt Eisley and Maridel Lares; and all former undergraduates, especially Yasuko Takeda, and all former rotation students. Finally, I must thank my advisor, Dr. Karen G. Fleming, who taught me an analytical approach to biological problems, got me excited about studying biophysics and gave me the confidence to pursue it. Karen has given me the support and the freedom to become a confident and independent scientist.

TABLE OF CONTENTS

Abstract		ii
Acknowledgments		v
Table of Contents		vii
Abbreviations		x
List of Tables		xi
List of Figures		xii
CHAPTER 1. Introduction		**1**
1.1	Overview and Perspectives	1
1.2	Forces that drive membrane protein folding	4
1.3	A thermodynamic approach to membrane protein folding	8
1.4	Relationship between sequence and stability	18
1.5	The relevance to membrane protein stability in the cell	21
1.6	Amino acid sequence specifies stable helix-helix interactions	27
1.7	Thermodynamic measurements for membrane protein stability	32
1.8	Thesis Overview	35
CHAPTER 2. Sequence context modulates the stability of a GxxxG mediated transmembrane helix-helix dimer		**39**
2.1	Summary	39

2.2 Introduction 40

2.3 Materials and Methods 42

2.4 Results 46

2.5 Discussion 56

2.6 Conclusions 61

CHAPTER 3. Complex interactions at the helix-helix interface stabilize the Glycophorin A transmembrane dimer 63

3.1 Summary 63

3.2 Introduction 64

3.3 Materials and Methods 67

3.4 Results 71

3.5 Discussion 86

3.6 Conclusions 92

CHAPTER 4. Packing defects in the Glycophorin A transmembrane dimer induce global rearrangements 93

4.1 Summary 93

4.2 Introduction 94

4.3 Materials and Methods 98

4.4 Results 101

4.5 Discussion 120

4.6 Conclusions 125

CHAPTER 5. Alternate interfaces may mediate homomeric and heteromeric assembly in the transmembrane domains of SNARE proteins **126**

5.1 Summary 126

5.2 Introduction 128

5.3 Materials and Methods 132

5.4 Results 135

5.5 Discussion 150

5.6 Conclusions 154

CHAPTER 6. Conclusions and Remarks **156**

REFERENCES. **159**

VITA. **171**

ABBREVIATIONS

GpA	(Glycophorin A)
BR	(Bacteriorhodopsin)
TM	(transmembrane domain)
C_8E_5	(Polyoxyethylene 5 octyl ether)
C14	(3-(N,N-Dimethylmyristylammino)propanesulfonate)
SDS-PAGE	(sodium dodecyl sulfate polyacrylamide gel electrophoresis)
CuOP	(copper phenanthroline)
DTNB	(5,5'-dithio-bis(2-nitrobenzoic acid))
AUC	(Analytical Ultracentrifugation)
SNARE	(soluble NSF attachment protein receptors)
SN	(*Staphylococcal aureus* Nuclease)
Syx	(Syntaxin 1A)
Syb	(Synaptobrevin 2 aka VAMP2)

LIST OF TABLES

CHAPTER 3.

Table 3.1	Double Mutant Analysis of GpA TM	72
Table 3.2	Alignment of experimental constructs that test the role of the GxxxG motif	81

CHAPTER 4.

Table 4.1	Analysis of structural models generated by CHI.	116

CHAPTER 5.

Table 5.1	Free energy of association for syntaxin and synaptobrevin constructs.	136

LIST OF FIGURES

CHAPTER 1.

Figure 1.1 The crystal structure of Bacteriorhodopsin. 10

Figure 1.2 The two-stage model for membrane protein folding. 12

Figure 1.3 The thermodynamic cycle for single point substitutions in a transmembrane α-helical dimer. 17

Figure 1.4 The GpA TM dimer NMR structure. 22

Figure 1.5 The ToxR assay measures dimer affinity *in vivo* 25

CHAPTER 2.

Figure 2.1 The free energy perturbations due to mutations at the protein-protein interface of the GpA TM. 47

Figure 2.2 Equilibrium distributions of GpA TM dimer mutants. 49

Figure 2.3: A linear correlation between parameterized structural features and free energy perturbation. 52

Figure 2.4: Contributions of each of the structure-based energy terms to the of $\Delta\Delta G_{calc}$ sum. 54

CHAPTER 3.

Figure 3.1 Relative dimer population of selected GpA TM mutants. 74

Figure 3.2 The free energy of coupling for alanine double mutants. 76

Figure 3.3 The coupling of sites along the helix-helix interface. 77

Figure 3.4 A GxxxG motif is neither necessary nor sufficient for dimerization. 82

Figure 3.5 The free energy of association for double mutants can be predicted using structure-based calculations. 85

Figure 3.6 Long range coupling can be mediated through a change in the helix-helix crossing angle. 90

CHAPTER 4.

Figure 4.1 Possible structural rearrangements in the GpA symmetrical homodimer. 97

Figure 4.2 Crosslinking of interface and diagnostic cysteine Substitutions. 103

Figure 4.3 Crosslinking of double mutants at WT interface and lipid sites. 106

Figure 4.4 Rate of crosslinking for double mutants with diagnostic cysteine substitutions. 109

Figure 4.5 Crosslinking of GpA single alanine variants with diagnostic cysteine substitutions. 111

Figure 4.6 Forced choice experiments demonstrate preferential interfaces**.** 113

Figure 4.7 CHI generated models of strongly coupled double mutants. 117

Figure 4.8 Density plot of inter-subunit occluded surface area per residue in WT and mutant models. 119

CHAPTER 5.

Figure 5.1	Sedimentation equilibrium data for synaptobrevin and syntaxin constructs.	138
Figure 5.2	Relative dimer population of syntaxin and synaptobrevin constructs.	139
Figure 5.3	Homodimer crosslinking for syntaxin and synaptobrevin constructs.	142
Figure 5.4	Heterodimer crosslinking for syntaxin and synaptobrevin constructs.	145
Figure 5.5	Computational modeling for syntaxin and syntaxin-	147

CHAPTER 1

Introduction

1.1 Overview and Perspectives

A central quest of modern biology has been to understand how cellular function is ensured at the molecular level. A major step in understanding biological macromolecules was the development of X-ray crystallography to observe structures at the atomic level (Hodgkin 1972). The subsequent discovery of the DNA double-helix (Watson and Crick 1953) provided an obvious structural basis for transcription and replication. Early models of protein structure, such as the α-helix modeled by Linus Pauling (Pauling and Corey 1951a), suggested that protein structures could also be a simple assembly that dictates function. However, when Kendrew solved first structure for myoglobin (Kendrew et al. 1958), it was found that protein structures were intricate and complex. In contrast to the structure of DNA, the structure of a protein seemed to provide little insight into the structural basis for the protein function.

Since the initial structural studies of proteins, a major question has been how proteins obtain their intricate and complex folds. Over time, both theoretical and experimental studies have revealed general principles of protein folding. In particular, two puzzles have arisen from polymer theory: first, a principle later referred to as the Blind Watchman's paradox states that biological proteins cannot have arisen from random sequences (Dawkins 1996; Dill 1999); second, Levinthal's paradox (Levinthal 1968) states that protein folding cannot be a

random process. Together, these principles imply that the amino acid sequence of a protein must encode its proper fold through predetermined folding pathways. The particular folding pathway will be unique for different proteins, but if the native fold is at an energetic minimum, folding can be studied using equilibrium thermodynamics, regardless of the particular folding pathway. In fact, in the 1960's, Afinsen's work on ribonuclease showed that the native fold of a protein is indeed at an energetic minimum (Anfinsen 1973). Although, these seminal experiments provide a strong framework for understanding protein folding, the experiments that describe both the structure and the energetics of proteins had been carried out only in aqueous environments. It was not until years later that similar experiments could be carried out for proteins that exist in membraneous environments (London and Khorana 1982).

Membrane proteins comprise 30% of open reading frames in the typical genome (Wallin and von Heijne 1998), and 75 % of therapeutic targets (Baldwin 1993). They serve in the unique role of mediating signals between the cell and its environment by acting as channels and conduits for biological signals. To function properly, these proteins must acquire the native fold. Accordingly, mutations in membrane proteins that affect stability and conformation can lead to pathologies. For instance, cystic fibrosis is caused by mutations that affect the insertion of transmembrane segments into the bilayer (Choi et al. 2005) and the assembly of transmembrane helices (Chen et al. 2004). Furthermore, mutations that affect the association of receptor proteins also lead to pathologies. For instance, the neu oncogene contains mutations in the transmembrane region of

the rat homologue of the human epidermal growth factor receptor (Her-2) that lead to constitutive activity (Bargmann et al. 1986), perhaps due to increased dimerization (Smith et al. 1996). In addition, the maturation of red blood cells depends on the oligomerization of the Epo receptor (EpoR), which may be mediated through interactions in the transmembrane segments (Constantinescu et al. 2001). Therefore, the assembly of transmembrane helices in biological processes must be selected for stability, conformation, and a tightly regulated equilibrium for protein-protein interactions.

Advancement in the study of membrane proteins has been significantly slower than for soluble proteins. For instance, less than 1% of proteins currently deposited in the protein data bank are membrane proteins (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html; http://www.rcsb.org/pdb/). Nevertheless, knowledge of membrane protein structure and folding has progressed considerably since the 1980's. Early theories suggested that membrane protein folding could simply be the inverse of soluble protein folding. It is widely accepted that the native fold of a soluble protein is acquired through the formation of a hydrophobic core and a hydrophilic surface. It was hypothesized that the collapse of membrane proteins would result in a hydrophilic core with a hydrophobic surface (Engelman and Zaccai 1980). However, the first membrane protein structure demonstrated that the core of a membrane protein has equivalent hydrophobicity to soluble protein cores (Deisenhofer et al. 1984; Rees et al. 1989). Therefore, burial of a hydrophilic surface cannot be a driving force for membrane protein folding.

Since the protein core is similar in both membrane and soluble proteins, the forces that drive membrane protein folding may be analogous to those that drive soluble protein folding. However, the amino acid sequence of a membrane protein must be optimized for stability in the apolar membrane. Since transient interactions between transmembrane segments are essential for cellular processes, the amino acid sequence for a membrane protein may also be optimized for a particular equilibrium in the bilayer, and not for a stable interaction. How amino acid sequence specifies both the stability and conformation of a membrane protein remains unclear. To begin to address this question, it becomes necessary to perform thermodynamic analysis of membrane protein stability.

1.2 Forces that drive membrane protein folding

For both membrane and soluble proteins the free energy minimum of the native conformation will be dependent on the interactions between solvent molecules, interactions within the protein itself, and interactions between solvent and protein molecules. The classical model for soluble protein folding suggests that the major force driving folding is the hydrophobic effect (Kauzmann 1959; Dill 1990). At physiological temperatures, the hydrophobic effect is primarily entropic, but is commonly defined by a large change in heat capacity (ΔC_p) (Dill 1990; Murphy et al. 1990). Since membrane proteins do not exist in an aqueous environment, the role of the hydrophobic effect is questionable. Furthermore, the partitioning of nonpolar solutes into a lipid bilayer has been shown to be enthalpy driven, referred to as the “nonclassical” hydrophobic effect (Seelig and Ganz

1991). Therefore, the forces that drive membrane protein folding may be distinct from those that drive soluble protein folding.

A thermodynamic analysis of folding depends in part on the nature of the unfolded state. For membrane proteins, an interpretation of unfolded state is complex. It could be considered as an extended form in an aqueous environment, as with soluble proteins, or as some folded element in the lipid bilayer. Since membrane proteins do not exist in aqueous solutions, it is difficult to choose between these alternatives. At the same time, the unfolded state for soluble proteins is also not populated except under extreme conditions of pH or temperature or with the use of denaturants. Therefore, considering the unfolded state of a membrane protein to be an extended form in an aqueous solution may not be a huge extrapolation from the same assumption in the case of soluble proteins. If so, the partitioning of a membrane protein into an apolar environment during folding may be parallel to hydrophobic collapse in soluble protein folding. Furthermore, although partitioning into a lipid environment is enthalpically driven, it is accompanied by a large negative change in heat capacity, the hallmark of the hydrophobic effect (Wimley and White 1993). Without ionizable groups to interact with water and solubilize the protein, the hydrophobic collapse of a membrane protein would lead to a preferential insertion into the bilayer. Then, the forces that specify the native conformation in soluble proteins and membrane proteins would depend on the differences between the two environments.

As a soluble protein collapses into a native structure, the hydrophobic effect drives secondary structure formation through desolvation of the amide

backbone and tertiary structure formation by the partitioning of hydrophobic sidechains into the protein core. It is unclear whether the backbone or side chain effects dominate folding. Thus, a proper model for soluble protein folding includes both effects (Dill 1999). Although, it is important to note that these forces may not account for specificity in the native fold of a soluble protein. For membrane proteins, backbone desolvation continues to be the driving force for secondary structure formation; however, the forces driving tertiary folding may be different in a hydrophobic bilayer. Although the early structures of membrane proteins showed that they are not “inside-out” soluble proteins, the protein core can be more polar than the protein surface (Rees and Eisenberg 1989). This result implies that the partitioning of polar residues into the core of a membrane protein could be a driving force for folding. If this were true, the core of a membrane protein should always be less hydrophobic than the surface. However, the availability of more membrane protein structures over the years had shown that such is not the case, since hydrophobicity is an inadequate predictor of helix orientation in a membrane protein (Stevens and Arkin 1999). Given that the hydrophobic effect cannot drive tertiary folding events in membrane proteins, and the burial of hydrophilic surface is not a driving force, the forces driving tertiary and quaternary membrane protein folding may be distinct from those that drive soluble protein folding.

As more membrane structures are solved, it becomes possible to compare general differences in structure for membrane and soluble proteins. Statistical analysis of known structures indicates that membrane proteins have similar or

slightly tighter packing than soluble proteins (Chamberlain et al. 2003; Eilers et al. 2002). However, the distribution of amino acids in the core of a membrane protein is different than in soluble proteins. Membrane proteins tend to have smaller residues in the core, whereas soluble proteins tend to have bulkier residues in the core (Adamian and Liang 2001). Statistical studies of membrane protein structures show that small residues, which have high packing values, are more likely to be at a helix-helix interface. Large hydrophobic residues, which have low packing values, are more likely to be on a lipid-exposed surface (Eilers et al. 2000). For membrane proteins the helix-helix interaction surface is smooth whereas the lipid exposed surface is rough (Renthal 1999). An interaction surface with small residues would appear smooth, while bulky residues would make the lipid surface appear rough.

Why would membrane proteins be optimized for smooth interaction surfaces? There are two general explanations. First, the conformational entropy cost of partitioning small residues into the protein core would be less than that of large residues; second, small residues permit better packing at an interface. A smooth interface for packing allows a close approach of helices optimizing favorable van der Waals interactions between both sidechains and backbone atoms and minimizing the loss of entropy upon folding. Therefore, membrane protein folding may be dominated by van der Waals interactions that are modulated by the types of amino acid residues and their arrangement in the tertiary structure.

1.3 A thermodynamic approach to membrane protein folding

A rigorous analysis of the forces driving protein folding can be conducted using equilibrium thermodynamics. The basis for a thermodynamic approach was first proposed by Afinsen in his “thermodynamic hypothesis.” He described this hypothesis in his 1972 Nobel prize speech.

> This hypothesis states that the three-dimensional structure of a native protein in its normal physiological milieu … is the one in which the Gibbs free energy of the whole system is lowest; that is, that the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence

At that time, it was unknown whether the physiological milieu in which this statement held true was only cytosolic, or whether proteins in the cellular membrane also existed at an energetic minimum. Furthermore, the role of the amino acid sequence in determining the native fold for membrane proteins could be different than for soluble proteins. It has only been in the last twenty years that these questions have been addressed. Study of membrane proteins (MPs) has been limited to systems that are experimentally accessible, since hydrophobic polypeptides are difficult to work with experimentally. The first protein to permit such study was bacteriorhodopsin (BR). Structural and energetic work on BR has provided a basic understanding of membrane protein folding.

Folding studies on BR led to the conclusion that thermodynamics can be used to study membrane proteins. In BR, seven transmembrane helices form the native assembly and bind a retinal group (Figure 1.1). The binding of retinal can serve as an indicator that BR has acquired the native fold. Reversible

denaturation of BR was first indicated by the restoration of retinal absorption in detergent (Huang et al. 1981; London and Khorana 1982). A denatured BR can also refold in a lipidic environment into a structure that is indistinguishable from the native structure (Popot et al. 1986). BR can be cleaved into isolated fragments without loops between the helices and can refold and re-associate to bind retinal and regain native visible absorption spectrum (Liao et al. 1984; Popot et al. 1987). These seminal studies support the conclusion that native BR resides at an energetic minimum that is not dependent on the mechanism of folding. Furthermore, the native fold can be achieved in many lipidic environments, and the association of helices to form the native structure does not require peptide linkers between the helices. Therefore, as Afinsen described for soluble proteins, interatomic contacts in membrane proteins determine the native fold. Thus, the thermodynamic hypothesis also holds for membrane proteins, allowing equilibrium studies to be used to investigate the energetics of membrane protein folding.

The two-stage model for membrane protein folding.

The studies on BR lead to a model of membrane protein folding known as the two-stage model (Popot and Engelman 1990). Popot and Engelman proposed that helical membrane protein folding could be separated into distinct

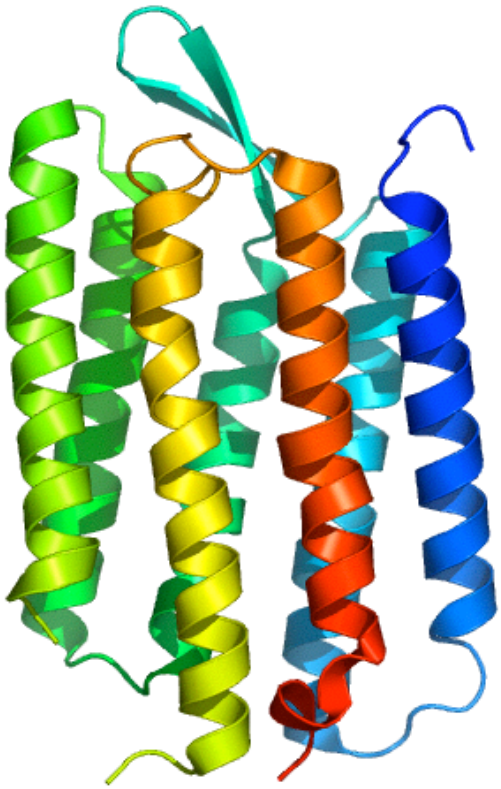

Figure 1.1 The crystal structure of Bacteriorhodopsin (1C3W; Luecke et al. 1999). In the figure, each of the seven helices has a unique color. The association of these helices into the native form allows it to bind retinal and function as a photoreactive center. Folding studies show that native BR can refold and reassociate without the loops that connect the transmembrane helices (Popot et al. 1987), indicating that the interactions between helices may specify the native fold.

thermodynamic stages: insertion of a peptide into the bilayer, and association of helices into a native structure (Figure 1.2). In the first stage of folding, secondary structure is acquired concomitant with the insertion of helices into a lipid bilayer. Unfolded polypeptides would be extremely unfavorable in the hydrophobic environment of a lipid bilayer due to the presence of unsatisfied amide hydrogen bonds. Therefore, when the peptides fold into an α-helix, hydrogen bonds are satisfied, and the helix should insert favorably into the membrane. Early calculations for insertion energetics predicted that inserted helices would be stable with a free energy value of approximately -35 kcal mol^{-1} (Jahnig 1983). Actual experimental values show that free energy of insertion is favorable for hydrophobic polypeptides in the range of -5 → -20 kcal mol^{-1} (White and Wimley 1999). Therefore, after insertion the helices are extremely stable and will not unfold spontaneously in the bilayer. The process of insertion may be more complex than proposed in this model, and some thermodynamic models include multiple individual stages to membrane protein folding (Jacobs and White 1989; White and Wimley 1999). In these models, insertion is divided into three independent steps, and the association of α-helices is a final independent stage. In this final stage, the functional helical bundle is formed.

Since insertion and association are thermodynamically distinct, each stage can be considered independently. In membrane proteins, the secondary structure of a helical bundle will not unfold spontaneously, but the bundle can disassociate to individual stable helix monomers. Therefore, when considering the assembly of membrane proteins embedded within a lipidic environment, the “unfolded”

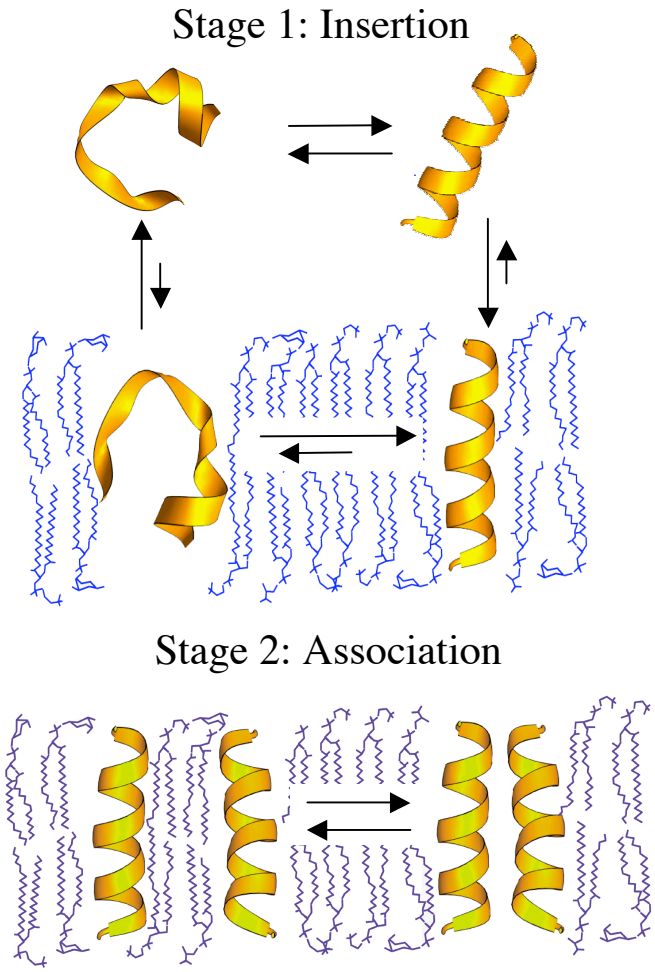

Figure 1.2 The two-stage model for membrane protein folding (Popot and Engelman 1990). Stage one involves the insertion of a hydrophobic peptide into the membrane bilayer. The acquisition of secondary structure occurs during the process of insertion. It is unfavorable for an α-helix to unfold in the bilayer due to the necessity to satisfy hydrogen bonds in an apolar environment. Stage two involves the association of these stable monomers to form functional multimers in the bilayer. Cartoons for helices were provided by Karen Fleming and generated using RIBBONS (Carson 1997). Cartoons for lipids were provided by Kevin MacKenzie.

state is a stable α-helix. This is truly distinct from the unfolded state for soluble proteins. Furthermore, the thermodynamically independent stages of membrane protein folding allow an independent study of secondary structure and tertiary structure. As Afinsen described for the energetic minimum of soluble proteins, for membrane proteins each independent minimum should be encoded by the amino acid sequence. Thus, the effect of sequence on the free energy of insertion and the free energy of association can be studied separately under appropriate experimental conditions.

Hydrophobicity and the free energy of insertion.

The hydrophobicity of an amino acid sequence can be used to predict the location of transmembrane segments in polytopic proteins. Partitioning studies of model residue compounds into apolar solvents have been used to measure of the free energy of insertion for an amino acid residue into a hydrophobic environment (Nozaki and Tanford 1971; Engelman et al. 1986; Fauchere et al. 1988; Radzicka et al. 1988). Octanol is a common model for an apolar environment, such as the interior of a protein or a lipid membrane (Eisenberg and McLachlan 1986; Fauchere et al. 1988; Shirley et al. 1992). A whole residue scale was developed by measuring the free energy of transfer for amino acid sidechains in pentapeptide models from water to octanol (Wimley and White 1996). This scale is particularly useful, because it includes the contribution of the peptide bond to the solvation free energy. The free energy of insertion for a polypeptide should be additive, since the sum of the free energies of insertion for individual amino acids provides an accurate measure for the insertion of the

entire sequence (Wimley and White 1999; Wimley and White 2000). According to hydrophobicity scales, ionizable residues are extremely costly to insert into a hydrophobic environment. Therefore, the presence of ionizable residues in a transmembrane sequence may affect the ability of that sequence to insert into the membrane, or lead to a preferential partitioning of ionizable residues at the membrane protein core. In soluble proteins, the burial of an ionizable residue can be energetically costly, since it will have a favorable interaction with the water solvent. For membrane proteins, an ionizable residue would not have a favorable interaction with the lipid solvent, and therefore, could have a dramatic affect on the fold or energetics of a membrane protein.

The role of ionizable residues in membrane proteins has been the subject of recent controversy. These residues are essential to the paddle-model proposed for regulation of the voltage gated potassium channel, KCSA. MacKinnon proposed that the S4 helix of KCSA has 4 arginines exposed to the lipid face in the native structure (Jiang et al. 2003). Critics of the model state that the arginines would be unfavorable at the lipid face and would partition in the protein core (Cuello et al. 2004). Supporters note that the free energy of insertion for the entire sequence will determine whether a polypeptide would insert (Freites et al. 2005; Hessa et al. 2005b). Therefore, the mere presence of ionizable residues may not be an immediate indicator of the ability of a sequence to insert or of the conformation of the corresponding peptide in the bilayer; rather, the hydrophobicity of the entire amino acid sequence of a protein will determine whether insertion is favorable. The role of a transmembrane sequence containing

ionizable residues may be critical to protein function. In the paddle model, the exposure of arginines to the lipid face may be critical to act as a voltage sensor. In some proteins, ionizable residues may also have an effect on specifying the native conformation. However, the overall hydrophobicity of a transmembrane segment is important when considering membrane protein folding in distinct stages. If a polypeptide can insert stably, the association of the resultant helices can be considered independently, and the role of the amino acid sequence on association can also be studied independently.

Using thermodynamics to study helix-helix interactions.

The second stage of membrane protein folding is the association of transmembrane helices in the bilayer. The hydrophobic effect may be responsible for the insertion of transmembrane α-helices, but should not affect interactions between helices in the bilayer. However, the assembly of α-helices may be driven by a "solvophobic effect" dependent on the lipid (Lemmon and Engelman 1994). The free energy of association in a lipid environment can be broken down into the individual components as follows:

$$\Delta G_{HH} = \Delta G_{PP} + \Delta G_{PL} + \Delta G_{LL} \qquad (1.1)$$

The individual components include the free energy of protein-protein interactions, ΔG_{PP}; free energy of protein-lipid interactions, ΔG_{PL}; and free energy of lipid-lipid interactions, ΔG_{LL}. Each of these components will contribute to the stability of an interaction; however, the dominant force is unknown. Helices are rigid structures, and therefore, lipid entropy may drive association. Also, in the apolar environment, the strength of van der Waals and electrostatic interactions may be

increased and could provide an attractive force between helices. Calculations predict that enthalpy favors assembly of helices over dispersion in the lipid phase (Wang and Pullman 1991). The relative strengths of these forces will be determined by the environment and specified by the amino acid sequence of the protein.

The role of amino acid sequence in the association of helices can be investigated by making substitutions in the peptide sequence, which can be described as a thermodynamic cycle (Figure 1.3). According to the principles of thermodynamics, the change in the free energy of association upon mutation will be due to the effect on both dimer and monomer stability. The cycle is symmetrical and the following equation should hold.

$$\Delta G_1 + \Delta G_2 = \Delta G_3 + \Delta G_4 \qquad (1.2)$$

If the surrounding sequence is adequate to promote insertion, a single substitution should still allow a stable insertion of the peptide in a hydrophobic environment. In fact, for most of our studies, the assumption should hold that the substitution of a hydrophobic residue by another hydrophobic residue will have negligible effect on the free energy of insertion. In this case, we can consider that $\Delta G_3 = 0$, and the following equation applies.

$$\Delta G_1 + \Delta G_2 = \Delta G_4 \equiv \Delta\Delta G_{MUT} \qquad (1.3)$$

Therefore, the change in the free energy of association for a wildtype versus a mutant sequence can be considered as a change in dimer stability only. This rationale is critical to interpretation of energetic studies of membrane proteins and provides a different approach than studies of soluble proteins. If amino acid

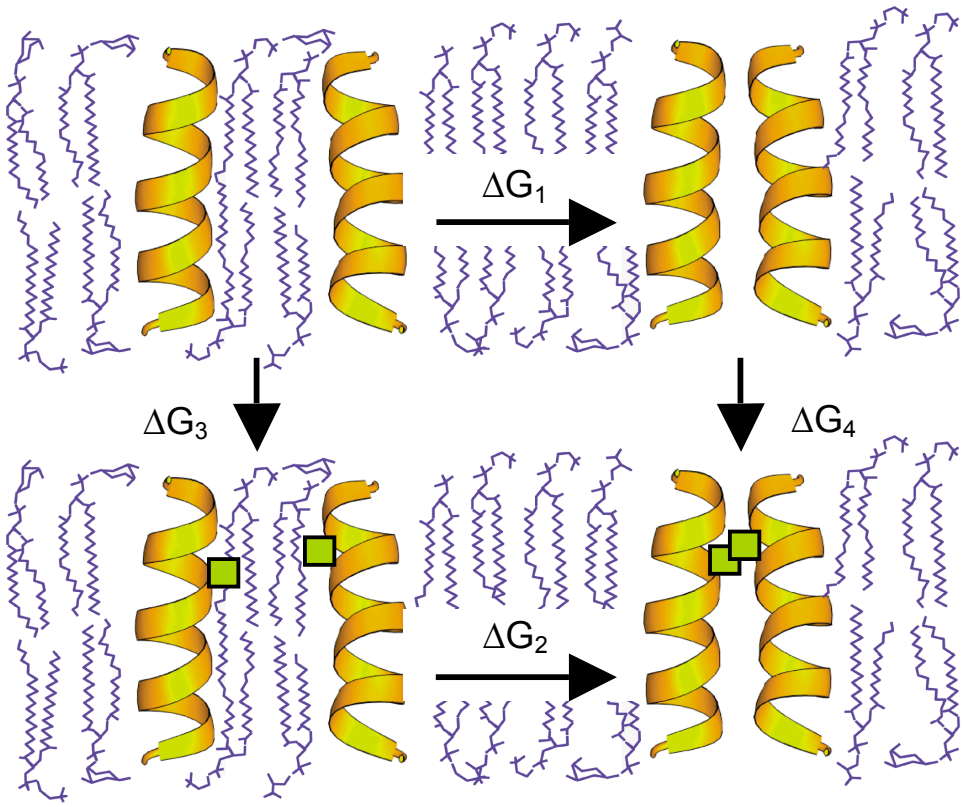

Figure 1.3 The thermodynamic cycle for single point substitutions in a transmembrane α-helical dimer. The association of the WT is described by ΔG_1. A substitution is made in the protein represented by a square, and the association of the mutant is described by ΔG_2. The free energy change upon substitution in the monomer is given by ΔG_3. The free energy change upon substitution in the dimer is given by ΔG_4. The principles of thermodynamics dictate that this cycle sums to zero. A conservative substitution allows the change in free energy upon substitution to be equivalent to the change in dimer stability. Cartoons for helices were provided by Karen Fleming and generated using RIBBONS (Carson 1997). Cartoons for lipids were provided by Kevin MacKenzie.

substitutions do not affect the free energy of insertion, the unfolded state should still be a stable a-helical monomer. Therefore, for membrane proteins, an isolated study of the thermodynamics of tertiary and quaternary interactions can be performed.

1.4 Relationship between sequence and stability

Rigorous thermodynamic analysis of membrane proteins has proved extremely difficult. Most membrane proteins do not denature reversibly with heat or denaturants (Haltia and Freire 1995). Using calorimetry, the major contribution to the heat of unfolding is the cytoplasmic domains (Haltia et al. 1994). Even under extremely denaturing conditions, transmembrane helices do not unfold in a lipidic environment (Arrondo et al. 1994; Hill et al. 1988). Therefore, a large contribution to membrane protein stability is the strength of intramolecular hydrogen bonds that stabilize an α-helix in a hydrophobic environment. These interactions may be the major contributors to the overall stability of a membrane protein, but they may not specify the functional native fold. To study the interactions that contribute to the stability of the native fold, the association of transmembrane helices can be studied in isolation using a simple model system.

Glycophorin A (GpA) has become a model for membrane protein folding to study sequence specific helix-helix interactions. GpA is a transmembrane dimer that was originally purified from human erythrocytes (Furthmayr et al. 1975). Although isolation produced a clearly homogeneous protein, two electrophoretic bands representing different oligomeric states were observed on SDS-PAGE. It was later shown buy AUC that these states are monomer and dimer (Dohnal et

al. 1980). The dimer was found to be extremely stable, and it could only be disassociated in the presence of heat and SDS (Furthmayr and Marchesi 1976). The transmembrane segment was determined by the inhibition of full length protein association by trypsin digests of purified protein and confirmed by synthetic peptides that mimic the assembly of GpA (Bormann et al. 1989; Segrest et al. 1973). These studies showed that association was driven by the transmembrane segment and was dependent on protein concentration and reversible. The stability of GpA and the presence of monomer and dimer on SDS-PAGE made it a model protein for studying role of amino acid sequence in the association in α-helical membrane proteins.

The role of amino acid sequence in stabilizing a transmembrane helical bundle was unclear until the early 1990's. The interactions driving association could be lipophobic, via interactions between side chains and lipids, or sequence specific interactions at interface, via interactions between the side chains of the oligomers (Jahnig 1983; Lemmon and Engelman 1994). By performing a large scale mutagenesis, it was possible to probe sequence specific interactions, and infer the role of sequence in the energetics of association (Lemmon et al. 1992b). To permit a large scale analysis, a chimeric construct with soluble *Staphylococcal* nuclease (SN) N-terminal to the GpA transmembrane sequence was generated (referred to as SN-GpA), which is easily purified and expressed in *E*. *coli*. A qualitative analysis of monomer and dimer concentrations by SDS-PAGE allowed the determination of the relative stabilities of GpA mutants (Lemmon et al. 1992a).

The saturation mutagenesis determined that substitutions at specific residues modulated dimerization (Lemmon et al. 1992b). These residues demonstrated a periodicity consistent with a single face of the α-helix. The sequence of interaction was determined to be $L^{75}I^{76}xxG^{79}V^{80}xxG^{83}V^{84}xxT^{87}$. The sites that showed the greatest destabilization were two glycines, Gly^{79} and Gly^{83}. All substitutions at these sites lead to a dramatic decrease in dimerization. This led to the hypothesis that the presence of glycines at these sites allows a close approach of the helices, which may be stabilized by electrostatic interactions between the backbones. The preferential face for helix-helix interactions, distinct from the helix-lipid face, may also be specified by van der Waals packing interactions.

The motif for dimerization in GpA has been studied in great detail using multiple techniques. SDS-PAGE analysis of chimeric proteins showed that the motif for dimerization was adequate to drive dimerization in other hydrophobic transmembrane segments (Lemmon et al. 1994). Alanine insertion mutagenesis provided further information into the role of the amino acid sequence in specifying the dimer (Mingarro et al. 1996). This study suggested that $G^{79}V^{80}xxG^{83}V^{84}xxT^{87}$ was the most important motif for dimerization in the GpA TM. Although insertions that disrupted Leu^{75} and Ile^{76} led to a decrease in dimerization, it was not comparable to the decrease in dimerization caused by disruptions in the remaining motif. The ability of this sequence to drive dimerization in other hydrophobic transmembrane segments suggested universal

roles for specific interaction motifs in transmembrane helices to promote oligomerization.

When the NMR structure of the GpA TM was determined (MacKenzie et al. 1997) (Figure 1.4), it became one of the only membrane proteins for which there were both structural data and stability measurements. The NMR structure provided further insight into the roles of the amino acid residue in specification of the dimer (MacKenzie et al. 1997). The proposed dimer interface (Lemmon et al. 1992b) was confirmed, demonstrating that mutagenesis data is reflective of structural conformation in membrane proteins. In particular, the glycines residues, shown to be the most destabilizing in mutagenesis, are the sites of closest approach in the dimer. These residues are thought to be “helix-breaking” in soluble α-helices, but in the GpA TM they appear to generate specific interactions between the monomers. The quantity of experimental data on the GpA TM promotes the use of this model system to develop general rules of membrane protein folding and assists in the development of experimental techniques to study membrane proteins.

1.5 The relevance to membrane protein stability in the cell

Membrane protein insertion in vivo.

Although thermodynamic models outlining a spontaneous insertion of helices into the bilayer are useful for understanding membrane protein folding, spontaneous insertion is not the primary mechanism of membrane protein folding *in vivo*. In the cell, membrane protein insertion is mediated by the translocon via direct association with the ribosome throughout the translation of an amino acid

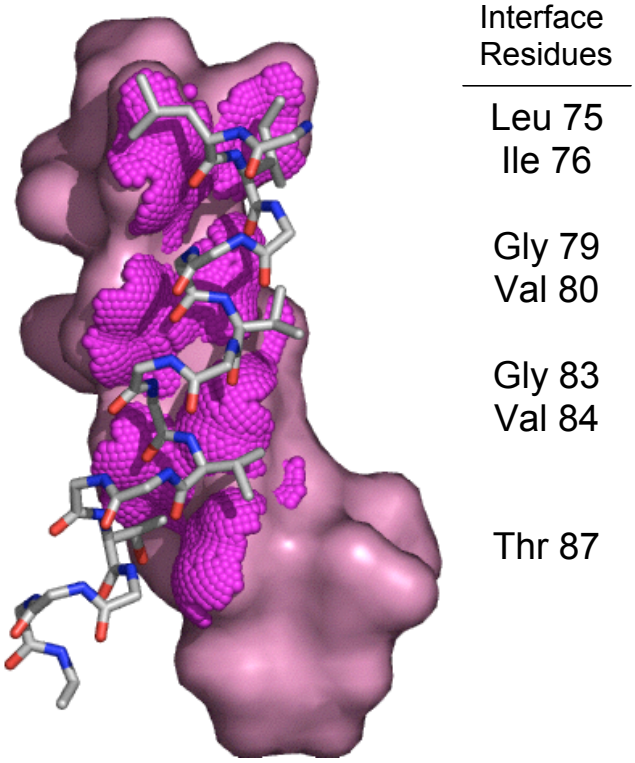

Figure 1.4 The GpA TM dimer NMR structure (1AFO; MacKenzie et al. 1997). Interface residues listed on the right of the model. The monomer B is represented in sticks and monomer A is represented as a surface model. The dot surface is the surface of interaction for monomer A. The dot surface shows that GpA is well packed at the dimer interface, with a smooth surface of interaction and minimal voids. The residues in the center of the helices have the closest approach, which includes the critical glycine residues. This figure was created using MacPymol (DeLano Scientific L.LC.).

sequence. The crystal structure of the translocon suggests that the polypeptide chain is translated into an intramembrane proteinaceous pore and partitions directly into the membrane (Van den Berg et al. 2004). The ability of the translocon to distinguish between proteins that should be secreted or inserted into the membrane is dependent on the hydrophobicity of the amino acid sequence (Chen and Kendall 1995). Hydrophobicity scales determined *in vitro* have been used as adequate predictors of transmembrane segments (Engelman et al. 1986; Kyte and Doolittle 1982; Wimley et al. 1996; Wimley and White 1996; Jayasinghe et al. 2001), but the correlation between the hydrophobicity of residues *in vitro* and *in vivo* was not established until recently. A whole residue *in vivo* biological scale for hydrophobicity was developed to determine the ability of a residue to insert into a membrane bilayer (Hessa et al. 2005a). The scale was created by comparing the relative amount of secreted protein versus inserted protein for all residues types in a host sequence. Importantly, a strong correlation was found between *in vitro* and *in vivo* hydrophobicity scales (Hessa et al. 2005a). Therefore, the free energy minimum for membrane proteins is related whether the mechanism of insertion is biological, such as translocation into the membrane, or experimental, such as solubilized by detergent micelles. This validates the thermodynamic analysis of membrane proteins in membrane mimetic environments, demonstrating that *in vitro* studies do elucidate membrane protein stability *in vivo*.

Studying helix-helix association in vivo.

The experimental conditions to study membrane proteins *in vitro* are drastically different than those in a cellular membrane. For soluble proteins, polypeptides can be solubilized under many conditions that may approximate the physiological conditions in the cytosol. However, due to their inherent hydrophobic nature, it is necessary to use detergents to solubilize membrane proteins *in vitro*. Therefore, there are many inherent differences in the physiological and experimental conditions for membrane proteins. For instance, a detergent micelle can be considered a 3 dimensional environment, whereas a bilayer is inherently 2 dimensional. In addition, the bilayer is composed of phospholipids with 2 acyl chains, whereas detergents have variable headgroups and a single acyl chain. It has been difficult to ascertain the effects of such different environments on sequence specific interactions hydrophobic polypeptides. More recently, an assay was developed to measure dimerization *in vivo*, allowing experimentalists to directly compare membrane protein stabilities in detergent and bilayer environments.

The ToxR assay was first developed to measure the propensity for GpA dimerization in the *E. coli* inner membrane (Langosch et al. 1996; Figure 1.5). A chimeric protein is created that contains a transmembrane domain of interest fused to maltose binding protein (MBP) at the N-terminus and the transcriptional regulator ToxR' at the C-terminus (Langosch et al. 1996). The presence of MBP in the periplasm allows the experimentalist to confirm the orientation of the protein in the cellular bilayer, since MBP must be expressed in the periplasm to allow cells to be grown in the presence of maltose. If the transmembrane

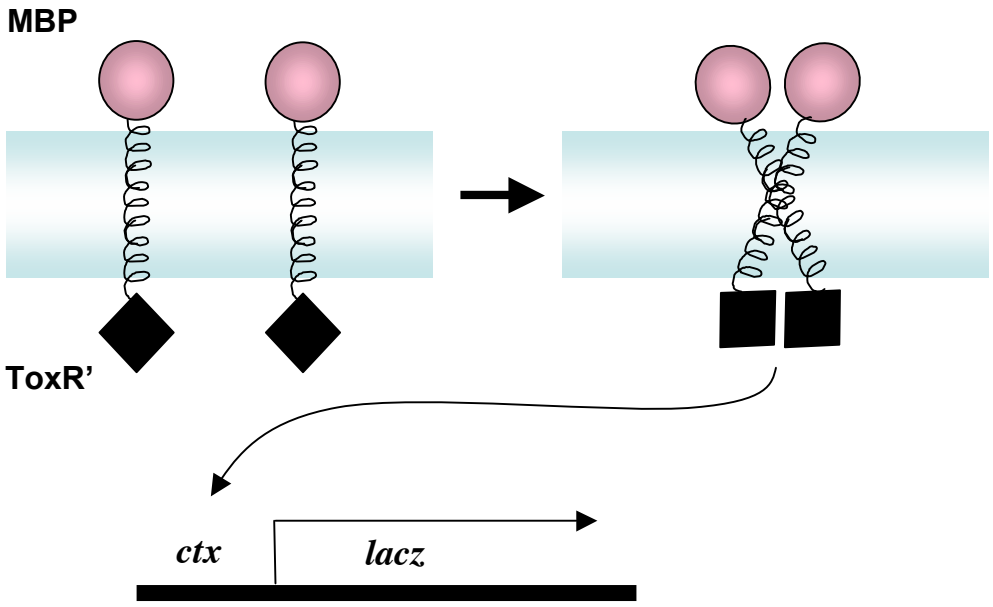

Figure 1.5 The ToxR assay measures dimer affinity *in vivo* (Langosch et al. 1996). A chimeric protein with a transmembrane domain of interest will insert into the bilayer with MBP as an extracellular domain and the ToxR' regulatory domain as an intracellular domain. If the transmembrane domain drives dimerization, the ToxR' domains can activate the transcription of a reporter gene, after binding to the *ctx* promoter. In this assay, *lacz* is the reporter gene and the level of dimerization can be detected by the amount of β-galactosidase activity.

segment drives dimerization, the ToxR' domains are drawn in close proximity, activating the *ctx* promoter and initiating transcription of the *lacZ* reporter gene. The ToxR construct containing the GpA TM induces high activity of β-galactosidase, indicating strong dimerization of the transmembrane segment. Alanine substitutions were made in this construct to reflect alterations in the previously determined motif for dimerization. The importance of the GpA dimerization motif was confirmed *in vivo*. Small differences were found in the effects of mutations compared to the saturation mutagenesis (Lemmon et al. 1992b). Although a significant decrease in dimerization was demonstrated by a reduction in the ToxR signal for I^{76}, G^{79}, G^{83}, T^{87}, the decrease in dimerization for L^{75}, I^{76}, V^{80}, V^{84} was not as significant as in the previous SDS-PAGE study. Nevertheless, the overall motif for dimerization in the cellular membrane was consistent with that determined in detergent micelles.

TOXCAT is a similar assay, which uses chloramphenicol-acetyl transferase as a reporter for dimerization (Russ and Engelman 1999). In a TOXCAT study, mutations were made to multiple residue types to compare the effects of polar mutations in the *in vitro* SDS study to the effects in an *in vivo* bilayer. The results show that the GpA variants follow the general trend of the SDS study, although polar substitutions appear to be more tolerated by the dimer when inserted into the bilayer than when inserted into SDS micelles. Although lipid bilayers and detergent micelles are inherently different, it appears that the protein-protein interactions *in vivo* are conserved in *in vitro* analysis. Evidently,

the interactions that drive helix-helix association can be independent of environment and are primarily dependent on the amino acid sequence.

1.6 Amino acid sequence specifies stable helix-helix interactions

Both *in vivo* and *in vitro* data on the association of GpA demonstrate that the stability of the dimer is dependent on the amino acid sequence and that interactions at the helix-helix interface are critical in stabilizing the dimer. Detailed information on this model system may allow the development of rules that are generally applicable to membrane protein stability. Although, like soluble proteins, membrane protein stability is dependent on amino acid sequence, the interactions between amino acid side chains that stabilize a protein in the hydrophobic milieu may be distinct from those in soluble proteins. The types of residues in the primary sequence will dictate the types of stabilizing interactions in a protein. In particular, the role of polar and ionizable residues may be different in a hydrophobic environment. In this section, the role of sequence in stabilizing transmembrane oligomers will be explored.

The role of glycines in driving association and the GxxxG motif.

Glycine residues, thought to be helix-breaking for soluble α-helices, are overrepresented in membrane proteins (Senes et al 2000). These residues may more abundant in membrane proteins, because they provide favorable interactions that contribute to membrane protein stability. For instance, glycines in the GpA TM interaction motif have been shown to be the most critical sites for strong dimerization. Electrostatic calculations on the GpA NMR structure predict

that the glycines permit Cα--H····O hydrogen bonds across the dimer interface that contribute to dimer stability (Senes et al. 2001). In this study, Cα--H····O hydrogen bonds were also predicted at glycine residues in two other membrane proteins, glycerol facilitator and calcium ATPase. In contrast, molecular dynamics simulations predict that an electrostatic repulsion occurs at the glycines in the GpA dimer interface (Petrache et al. 2000). Nevertheless, a hypothesis has emerged that glycine residues comprise an important motif for helix-helix interactions in many membrane proteins. The validity of this motif, known as the GxxxG or GG4 motif, is supported by two studies from the Engelman group published in tandem.

The first study uses the TOXCAT assay to screen for motifs of strong dimerization in transmembrane helices (Russ and Engelman 2000). In this screen, cells transfected with a randomized library of TM sequences were grown in the presence of chloramphenicol (CAM) to select for sequences that will dimerize strongly in the cellular membrane. In the randomized transmembrane sequence, one face of the helix was allowed to vary to any combination of nine amino acids: glycine, alanine, valine, leucine, serine, threonine, proline, and arginine. The opposing face was held constant as a sequence of leucines or alanines. In fact, the only sites allow varied were those consistent with the interface motif for glycophorin A. An overall pattern for strong dimerization residues emerged with the vast majority of strongly dimerizing sequences containing the GxxxG motif. Few sequences in the library contained other residues at these locations, with the single exception of serine in a small

percentage of sequences. The varied sites were consistent with the location of interface residues for GpA, but the exact sequence motif of GpA was not selected in the screen. Although, an overall motif was found that matched the GpA TM dimerization motif. Accordingly, these results provided strong evidence for a generalized motif of GxxxG driving dimerization in transmembrane α-helices.

In the accompanying study, a statistical analysis of membrane proteins probed the presence of overrepresented sequence motifs (Senes et al. 2000). Approximately 13,000 putative transmembrane domains were isolated from the Swiss-Prot database for statistical analysis. The average amino acid composition of these TMs was calculated. It was found that the hydrophobic residues Leu, Ile, Val, and Ala comprise 50% of all residues in a TM. Gly and Phe contribute 10% each, and all other residues comprise 30%. Ionizable residues are underrepresented, consistent with the unfavorable free energy of transfer for an ionizable residue into an apolar environment. This study investigated the probability of residues in a transmembrane segment and combinations of residues. A particular pair of residues, two glycines separated by three residues (GxxxG or GG4) showed the strongest positive correlation.

GG4 occurs 32% more than expected based on a random distribution of the residue pair according to the probability of glycine occurrence in a transmembrane sequence (Senes et al. 2000). This correlation is much greater than all other pairs found in the analysis and is considered highly significant. In conjunction with the accompanying TOXCAT selection study, the

overrepresentation of GG4 in transmembrane helices suggests that these residues play an important biological role and suggests that these residues are driving helix-helix interactions *in vivo*. However, in statistical studies on soluble protein structures, GG4 occurred 41 % more than expected, while only 50% of those helices containing a GG4 motif participate in helix-helix interactions (Kleiger et al. 2002). Therefore, GxxxG may not have a specific role for membrane proteins or for helix-helix interactions in general. Although GxxxG may be a determinant for strong helical association for some proteins, many membrane proteins that strongly associate may not contain the motif, and proteins that do contain the motif may not strongly associate. Other sequence dependent interactions may be critical in driving association of α-helices in a hydrophobic environment.

The role of ionizable and polar residues in driving association.

Another sequence dependent interaction emerging in the literature is the association of transmembrane helices induced by ionizable and polar residues. Ionizable residues were shown to drive association of polyleucine transmembrane segments by TOXCAT. An unsatisfied hydrogen bond is energetically unfavorable in the apolar lipidic environment, and in order to be satisfied, a hydrogen bond can form across a protein interface (Cosson et al. 1991; Smith et al. 1996). Side chains that can be both a hydrogen bond donor and acceptor provide the greatest driving force for association (Zhou et al. 2001). This is strong evidence that the presence of a ionizable residue can stabilize an oligomer by hydrogen bonding across an interface. Polar residues can also

mediate dimerization. The presence of an asparagine residue was found to mediate oligomerization of hydrophobic peptides in a detergent environment (Choma et al. 2000; Zhou et al. 2000). However, the ability of a polar or ionizable residue to drive association may be dependent on the sequence context.

The position of an asparagine in a polyleucine ToxR construct modulates the strength of association (Ruan et al. 2004). If interactions due to the polar residue were the only force driving association, the position of the asparagine in a transmembrane sequence should not affect its ability to dimerize. The ability of polar residues to drive association in the M13 major coat protein was also shown to be dependent on the sequence context (Dawson et al. 2003). The packing interactions surrounding polar residues were found to modulate the stability of an interface. Therefore, van der Waals interactions can specify and overcome the lipophobic force of polar residues to drive association. Because conservative hydrophobic substitutions have little effect on the free energy of insertion, by making these substitutions, the role of van der Waals interactions on helix association may be studied independently. Although it is important to note that interactions with the solvent may provide a component to the free energy of association for transmembrane helices, this evidence suggests that packing interactions of hydrophobic residues provide a major contribution to the stability and specificity of membrane protein interactions.

1.7 Thermodynamic measurements for membrane protein stability

Since membrane proteins cannot be denatured reversibly (Haltia and Freire 1995), thermodynamic analysis of the association of transmembrane oligomers can elucidate the principles of polytopic membrane protein stability. Sedimentation equilibrium analytical ultracentrifugation has emerged as a dominant means of obtaining energetic measurements of helix-helix stability for transmembrane oligomers in a detergent environment (Lebowitz et al. 2002). Using this method, the free energy of helix-helix association can be determined directly, since the fundamental measurement of AUC is molecular mass. The apparent protein mass will be dependent on the protein concentration, the stoichiometry, and the association constant. Therefore, it can be used to determine the concentration of protein oligomers at equilibrium. To obtain accurate thermodynamic data, the chemical molecular equilibrium must be reached, and transport equilibrium must be established. Transport equilibrium will occur when the gravitational potential is offset by the chemical potential, and there is no longer a net transport of molecules in solution. Under these conditions, AUC provides a thermodynamic measurement for the association of transmembrane oligomers.

AUC was used to perform a quantitative analysis on the GpA TM using the SN-GpA construct in a detergent environment. The fusion protein has a higher molecular weight and a larger molar extinction coefficient than the peptide alone. This allows access to lower concentrations in the centrifuge and a larger distribution of species at accessible speeds. The *Staphylococcal* nuclease

portion of the fusion construct does not drive association, since AUC of the SN portion alone results in a single monomeric sedimenting species (Fleming et al. 1997). Sedimentation equilibrium for membrane proteins presents unique problems. Membrane proteins must be solubilized by detergents, which results in a sedimenting species that is a protein/detergent complex. If the detergents sediments, an additional species that would need to be accommodated in global fitting. Furthermore, without explicit knowledge of the amount of detergent bound, it is impossible to determine oligomeric states from the sedimenting mass. The reversibility of the system will depend on reversible interactions between: protein/protein, protein/detergent, and detergent/detergent. The following equation mathematically represents how the buoyant molecular mass of the complex can be reduced into both detergent and protein components (Casassa and Eisenberg 1964).

$$M_{pr}\left(1-\phi'\rho\right) = M_{pr}\left[\left(1-\bar{\nu}_{pr}\rho\right) + \delta_{\det}\left(1-\bar{\nu}_{\det}\rho\right)\right] \qquad (1.4)$$

The partial specific volume of the protein/detergent complex is ϕ', and the partial specific volumes of the protein and detergent are $\bar{\nu}_{pr}$ and $\bar{\nu}_{\det}$. The amount of detergent bound is $\delta_{\det}$ in grams of detergent per gram of protein, and ρ is the solvent density. The equation can be reduced experimentally by choosing a detergent that is neutrally buoyant, or the buoyancy of the solution can be matched to the detergent using D_2O. This reduces the fitting function to the following equation, in which the buoyant molecular mass of the complex is comprised of the protein component alone (Reynolds and Tanford 1976).

$$M_{pr}\left(1-\phi'\rho\right) = M_{pr}\left(1-\bar{\nu}_{pr}\rho\right) \qquad (1.5)$$

The detergent C_8E_5 was used to determine the free energy of association for SN-GpA and mutants. Both transport and chemical equilibrium were established for SN-GpA in the detergent environment of C_8E_5. Furthermore, the $\bar{\nu}_{\mathrm{det}}\rho$ for C_8E_5 is equal to 0.999, eliminating the detergent component to the buoyant molecular mass. Using C_8E_5 to solubilize the protein, the sedimentation equilibrium data for SN-GpA show that the dimerization reaction is reversible and the apparent free energy of dimerization was determined to be 9.0 ± 0.1 kcal mol^{-1} in 33 mM C_8E_5.

Using AUC, the energetic cost of substitutions in the GpA TM can measured quantitatively (Fleming and Engelman 2001). An alanine scan showed that substitutions at all helix facing residues were destabilizing. The most costly substitution was $Gly^{83}Ala$, with a loss of free energy equal to 3.2 kcal mol^{-1}. Substitutions to alanine at lipid facing residues did not destabilize the GpA TM dimer, demonstrating little or no change in the free energy of association. The hierarchy of mutations was retained in the sedimentation equilibrium data that was determined in the SDS-PAGE saturation mutagenesis and ToxR assays. However, thermodynamic data provide further insight into the importance of sequence context. SDS-PAGE and ToxR/TOXCAT provide compelling evidence for the importance of the GxxxG motif in driving association, but the AUC data show that even when these sites are mutated, there remains a propensity for association. In fact, although dimer formation can be disrupted, no single alanine substitution can eliminate the dimer species. Instead, a range of effects is found, suggesting that the entire sequence specifies the stability. This is a powerful

interpretation of the data, which may influence our understanding of the importance of the role of sequence context *in vivo*. The entire amino acid sequence may determine both stability and a specific equilibrium of oligomers.

1.8 Thesis Overview

Technological advances in the study of membrane protein structure and energetics have allowed researchers to investigate and to understand membrane protein folding. Although these studies still lag behind those in the soluble protein field, advancements have allowed a rigorous comparison of membrane and soluble protein structure and stability. In this thesis, we investigate the role of amino acid sequence in membrane proteins using thermodynamics. The available structural and mutagenesis data for glycophorin A provide a strong framework for a detailed thermodynamic analysis. We later use the techniques and knowledge obtained from on analysis of GpA to study a more relevant biological system. These studies unequivocally demonstrate the importance of the entire sequence context in stabilizing an interface, specifying the native fold, and determining a specific equilibrium of interactions.

Chapter 2

In this chapter, we expand on previous studies on GpA variant association (Fleming et al. 1997; Fleming and Engelman 2001), and we perform a large scale mutagenesis including conservative hydrophobic substitutions. Our results provide further insight into the importance of sequence context in specifying the

stability of the dimer. All GpA variants dimerize, even those that have large aliphatic substitutions in the GxxxG motif. These data demonstrate that a GxxxG motif is not necessary for dimerization of the GpA TM. Furthermore, by obtaining a database of free energy values, it was possible to make a rigorous comparison of sequence and stability. Using molecular models based on the NMR structure and the free energy values obtained in these experiments, a structure-based parameterization was conducted. This parameterization suggests that favorable packing interactions dominate the forces that stabilize the GpA dimer. These results support a model for membrane protein stability through sequence dependent van der Waals interactions.

Chapter 3

Energetic coupling (or cooperativity) is essential to stabilize and specify the native fold of soluble proteins. Chapter three focuses on the coupling between sites at the interface of the GpA TM dimer. In this chapter, we also test the role of the critical sequence motif, GxxxG, in driving dimerization. We find that long range coupling stabilizes the dimer and that the entire sequence context modulates dimerization, not a simple sequence motif. We employ the structure-based relationship determined in Chapter 2 to correlate structural models based on the NMR structure and energetics for double mutants. We find that the energetics of association for double mutants that are strongly coupled do not correlate well with the structural parameters calculated from the corresponding mutant models. We suggest that these models do not represent the mutant

structure and that mutating sites that are strongly coupled may induce global conformational change. These results elucidate the role of amino acid sequence in specifying the structure and stability of a membrane protein.

Chapter 4

In chapter four, we test the relationship between structure and energetics by probing changes in the structure of the GpA TM upon mutagenesis. We experimentally investigate alterations in the dimer interface caused by mutations by site-engineered disulfide crosslinking. Furthermore, a modeling protocol is used to generate models *de novo*, which are used to explore the possible structural rearrangements that occur in GpA double mutants. The results demonstrate that the entire amino acid sequence has an intricate role in specifying the conformation of the interface. We find that no single structural change can explain the energetics of GpA mutants. Mutant structures include changes in helix curvature, distance between helices, and rotation of helices about the interface. In fact, some single mutants seem to explore alternate interfaces. The role of amino acid sequence in specifying the dimer interface appears to be complex and cannot be easily predicted based on energetics or structural models.

Chapter 5

In this thesis, the GpA TM has been the primary model system used to study helix-helix interactions. GpA is an extremely stable dimer that is

constitutively dimeric *in vivo*. An equilibrium of association for weaker oligomers may be critical for cellular processes that require dynamic interactions between proteins. This hypothesis is explored in chapter five, which focuses on two transmembrane proteins involved in vesicle fusion, syntaxin and synaptobrevin. We employ AUC to study homodimerization of syntaxin and synaptobrevin and mutated variants. A conserved dimer interface was proposed for these proteins, with only single site difference at the interface. Although mutation of this site has a dramatic effect on the association of synaptobrevin, the corresponding mutation in syntaxin has no energetic effect.

The results of cysteine crosslinking experiments suggest that the homodimerization interfaces for syntaxin and synaptobrevin are distinct, while the interface for heterodimerization may be more similar to the synaptobrevin interface. Using molecular modeling we calculate chemically reasonable structures that correlate with the experimental data. In conjunction with the AUC studies, these data suggest that amino acid sequence specifies a functional equilibrium for the interactions occurring within and between these proteins. Furthermore, inherent plasiticity in a transmembrane sequence may contribute to a protein's promiscuity in interactions necessary for cellular processes. Further investigation into the role of amino acid sequence in energetics of association for both stable and weakly associating proteins should lead to a better understanding of membrane protein function and the forces that drive protein folding in a membrane.

CHAPTER 2

Sequence context modulates the stability of a GxxxG mediated transmembrane helix-helix dimer

2.1 Summary

To quantify the relationship between sequence and transmembrane dimer stability, a systematic mutagenesis and thermodynamic study of the protein-protein interaction residues in the glycophorin A transmembrane dimer was carried out. This study addresses the ability for a GxxxG motif to drive dimerization and results demonstrate that the glycophorin A transmembrane sequence dimerizes when its GxxxG motif is abolished by mutation to large aliphatic residues. Therefore, the entire amino acid sequence encodes an intrinsic propensity to self-associate independent of a GxxxG motif. In the presence of an intact GxxxG motif, the glycophorin A dimer stability can be modulated over a span of –0.5 to +3.2 kcal mol^{-1} by mutating the surrounding sequence context. Thus, flanking residues play an active role in determining the transmembrane dimer stability. To assess the structural consequences of the thermodynamic effects of mutations, molecular models of mutant transmembrane domains were constructed, and a structure-based parameterization of the free energy change due to mutation was carried out. The changes in association free

energy for glycophorin A mutants can be primarily explained by changes in packing interactions at the protein-protein interface. The energy cost of removing favorable van der Waals interactions was found to be 0.039 kcal/mol per $Å^2$ of favorable occluded surface area. The value corresponds well with estimates for mutations in Bacteriorhodopsin as well as for those mutations in the interiors of soluble proteins that create packing defects.

2.2 Introduction

The relationships between sequence, structure and stability for membrane proteins are not well understood. While hydrophobicity can be used to predict the occurrence of transmembrane α-helices in open reading frames (Wimley et al. 1996; Wimley and White 1996), little is known about the principles that govern the subsequent interactions between helices leading to the formation of native membrane protein structures. In recent years a pattern of two glycines separated by three intervening residues, known as a GxxxG (or GG4) motif, has emerged as a characteristic signature of transmembrane α-helix dimerization. This motif was identified by a statistical analysis of pair-wise amino acid patterns in putative transmembrane helix sequences from SwissProt, revealing that the GG4 motif occurs at a frequency ~32% higher than the randomly expected value (Senes et al. 2000). In an accompanying study, a genetic screen designed to select transmembrane domains with high affinity homo-oligomerization properties from a randomized sequence library contained a high occurrence of the GG4 motif (Russ and Engelman 2000). Indeed, the importance of the GxxxG motif in the

dimerization of the transmembrane protein, glycophorin A, had been recognized for several years (Brosig and Langosch 1998; Fleming and Engelman 2001; Lemmon et al. 1992a; Lemmon et al. 1992b). In addition, the introduction of a GxxxG motif in a background of a polyvaline or polymethionine transmembrane sequence has been shown to enhance oligomerization (Brosig and Langosch 1998). Thus, the functional role for GxxxG was postulated as a motif to drive transmembrane helix-helix dimerization. A GG4 motif has subsequently been implicated in the oligomerization of the erbB tyrosine kinase receptors (Mendrola et al. 2002), the yeast ATP synthase (Arselin et al. 2003), the *H. pylori* vacuolating toxin (McClain et al. 2001; McClain et al. 2003), the yeast α-factor receptor (Overton et al. 2003), and assembly of the γ-secretase complex (Lee et al. 2003).

Insight into the role that the GxxxG motif plays in transmembrane helix-helix dimerization is derived mainly from studies on the glycophorin A transmembrane (GpA TM) dimer. From biochemical data as well as structural studies, the GxxxG motif in glycophorin A is part of a larger protein-protein interaction motif: $L^{75}I^{76}xxG^{79}V^{80}xxG^{83}V^{84}xxT^{87}$ (Lemmon et al. 1992a; Lemmon et al. 1992b; MacKenzie et al. 1997; Fleming and Engelman 2001), where the GxxxG motif glycines are Gly^{79} and Gly^{83}. Inspection of the NMR structure for the GpA TM dimer suggests that the glycine residues allow a close approach of the helices providing a smooth surface for packing interactions (MacKenzie et al. 1997). It has been suggested that the close approach of the glycines stabilizes the helix dimer by allowing the formation of an interhelical CαH••••O hydrogen

bond (Senes et al. 2001). In contrast, molecular dynamics calculations suggest that the interaction energy at the glycines in the GpA TM dimer is an unfavorable electrostatic contact (Petrache et al. 2000). Nevertheless, the pronounced destabilizing effects of mutations at the glycine residues (Lemmon et al. 1992a; Lemmon et al. 1992b; Brosig and Langosch 1998; Fleming and Engelman 2001) provide evidence that these two glycines are extremely important for strong dimerization. Surprisingly, previous sedimentation equilibrium experiments demonstrated that sequences containing mutations to alanine at the GxxxG motif still retain a significant propensity to dimerize (Fleming and Engelman 2001). To address this apparent paradox, we hypothesized that the sequence context surrounding the GxxxG motif must also play an active role in specifying and stabilizing the GpA TM helix-helix interaction. A systematic thermodynamic study to determine the free energies of association of GpA TM sequences containing point mutations to other hydrophobic residues was conducted to address this question.

2.3 Materials and Methods

Sample Preparation and Analytical Ultracentrifugation.

Single point mutants were generated using the Stratagene Quikchange protocol with appropriate primers. All plasmids were confirmed by DNA sequencing. All mutant proteins were purified using the published protocol (Fleming and Engelman 2001). Immediately before sedimentation equilibrium

analysis, samples were exchanged by ion-exchange chromatography into buffer containing C_8E_5 as described previously (Fleming and Engelman 2001).

Sedimentation equilibrium experiments were performed at 25˚C using a Beckman XL-A analytical ultracentrifuge as described previously in detail (Fleming 2002; Fleming and Engelman 2001). The samples were centrifuged for lengths of time sufficient to achieve equilibrium. Data obtained from absorbance at 230 nm were analyzed by nonlinear least-squares curve fitting of radial concentration profiles using the Windows version of NONLIN (Johnson et al. 1981) using the equations describing the reversible association in sedimentation equilibrium. For each global fit, nine equilibrium data sets were collected. These consisted of three different initial protein concentrations analyzed at three rotor speeds (20000, 24500, 30000) (*i.e.* such that the speed factor ratios were minimally 1.0, 1.5 and 2.25). The monomeric molecular masses and partial specific volumes were calculated using the program SEDNTERP (Laue et al. 1992), and these parameters were held constant in fitting the absorbance versus radius profiles.

The experimental free energy cost of mutation, $\Delta\Delta G^o_{Mut}$ equals $\Delta G^o_{Mut} - \Delta G^o_{WT}$ where $\Delta G^o_x = -RT \ln K_x$, and $K_x = K_{Assc,App}[micellar\, C_8E_5]_w$, and $K_{Assc,App}$ and [micellar C_8E_5]$_w$ are the experimentally determined monomer-dimer association constant and micellar C_8E_5 concentrations expressed on the molar aqueous scale (Fleming 2002).

Computational modeling of glycophorin A mutants.

The pdb coordinates (pdb1AFO.ent) containing 20 models for the solution NMR structure of the wild type glycophorin A transmembrane dimer were used as a basis for modeling. Each model was truncated to focus on the transmembrane domain residues, 74-91. The truncated forms were minimized using CNS employing the CHARMM22 parameter set with full van der Waals radii. The minimization included the publicly available NMR constraint tables. To automatically generate point mutations, each of the minimized models was mutated using the *mutate* and *debump* commands available in WHAT IF. As a control, the minimized WT structures were subjected to the *debump* procedure. This procedure allows rearrangement of the dimer interface at the level of changes in side chain conformation. No backbone rearrangements or rigid body helical rotations are modeled by this procedure. Six hundred sixty pdb files were created and analyzed (33 mutants x 20 models each).

Structure-based calculations

Version 7.2.2 of the occluded surface algorithm, OS (Pattabiraman et al. 1995), was used to quantify favorable (ΔFOS) and unfavorable (ΔUOS) inter-monomer contacts as well as to determine the exposed molecular surface area used in the side chain conformational entropy calculations (ΔS_{SC}) (Baker and Murphy 1998; Lee et al. 1994). The change in side-chain conformational entropy of dimerization for each sequence, $\Delta\Delta S_{SC}$, was calculated as the difference between the side chain conformational entropy of the monomers and the dimer.

The inter-monomer occluded surface area represents that portion of the molecular surface area of an atom on one chain that is occluded by any atom on the opposing chain. Occluded surface area calculations are advantageous over buried surface area calculations because they reveal atomic level descriptions of the packing changes that occur upon mutation. In addition, the extent of unfavorable van der Waals interactions can be quantified by summing occluded surface area regions where overlapping van der Waals radii occur. The sum of the inter-monomer occluded surface areas for all atoms is most closely related to the buried molecular surface area. Since the OS calculation is based on a molecular surface, the absolute value of the favorable occluded surface area is smaller than the traditional buried accessible surface area for an oligomeric protein. In the case of the glycophorin A mutants, the favorable occluded surface area scales linearly with the buried accessible surface area with a slope of 0.47 (R=0.88) (data not shown).

Differences in each structure-based parameter were calculated by taking the difference between mutant and wildtype of the calculated parameter for each structural model. This generated a set of 20 measures of each parameter for each mutant from which mean and standard deviation values were calculated. Coefficients for the ΔFOS and ΔUOS parameters were determined by simultaneously fitting 23 point mutants to the following linear model:

$$\Delta\Delta G^{o}_{Calc} = \sigma * \Delta FOS + \alpha * \Delta UOS + (-T\Delta\Delta S_{SC}) \equiv \Delta\Delta G^{o}_{Mut} \qquad (2.1)$$

where σ and α are the best-fit coefficients for the ΔFOS and ΔUOS parameters.

2.4 Results

The free energy costs of single point mutations at interface sites.

A systematic mutagenesis and thermodynamic study was carried out to determine how the sequence context surrounding the GxxxG motif influences the stability of the glycophorin A transmembrane dimer. Single point mutations to Gly, Ile, Leu, and Val were introduced at each of the protein-protein contact positions in the glycophorin A transmembrane domain. In addition, to address the potential for interhelical hydrogen bonding at position Thr^{87} by thermodynamic methods, a Ser mutation was constructed at this site. The effect of each of these mutations on the free energy of association was determined using sedimentation equilibrium analytical ultracentrifugation under conditions that measure reversible interactions between membrane proteins in micellar solutions (Fleming 1998; Fleming 2002). Previously, the dimer stabilities measured by sedimentation equilibrium on a series of alanine mutants of the GpA TM have been shown to scale linearly with the apparent stabilities obtained using an *in vivo* bacterial genetic assay (Fleming and Engelman 2001). The free energy perturbation due to mutation was quantified by subtracting the free energy of association of the wild type from that of the mutant. These $\Delta\Delta G^{o}_{Mut}$ values are shown in Figure 2.1. Overwhelmingly, mutations at the glycophorin A dimer interface lead to a loss of dimer stability. Only Ile^{76}Val and Val^{80}Leu dimerize as strongly as the wild type sequence. In contrast to the high frequency of stabilizing mutations observed for other transmembrane proteins (Bowie 2001; Faham et al. 2004; Zhou and Bowie 2000), no single point mutation was found to significantly stabilize the dimer. As

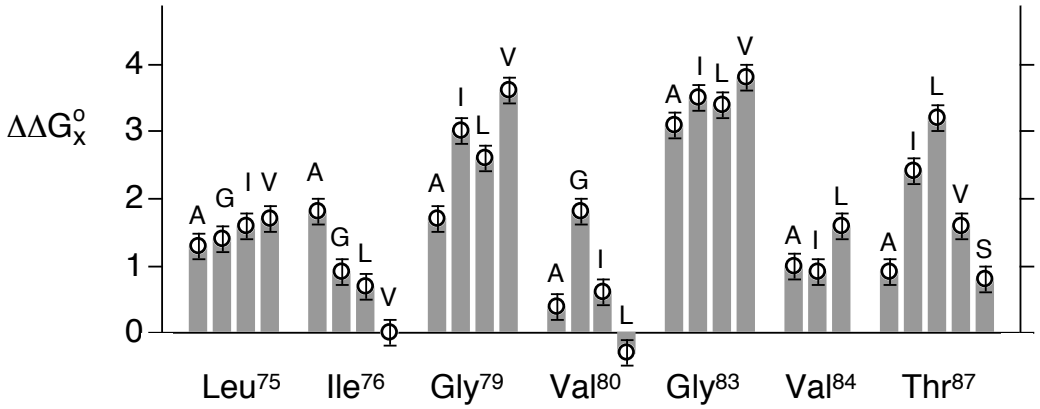

Figure 2.1 The free energy perturbations due to mutations at the protein-protein interface of the GpA TM. Shown are the $\Delta\Delta G^{o}_{Mut}$ values for single point mutations at the protein–protein interaction surface of the glycophorin A transmembrane dimer. The interaction residues are given in three letter code along the abscissa, and the identity of the single point mutations are given in one letter code above each energy value. $\Delta\Delta G^{o}_{Mut}$ values were calculated as described in the methods section. The error bars represent the standard deviations of the values and were propagated according to Bevington (Bevington 1969). For comparison the $\Delta\Delta G^{o}_{Mut}$ values of the previously published alanine mutants (Fleming and Engelman 2001) are included.

expected, all mutations at the two glycines in the GxxxG motif show strong destabilizing effects, yet these sequences retain a significant propensity to dimerize. One of the flanking residues, Val^{80}, is relatively insensitive when mutated to another nonpolar residue whereas the other flanking residue, Val^{84}, shows greater sensitivity. Mutations at distant sites on both ends of the helix from the GxxxG motif exhibit a wide variation of perturbations to the free energy of association. Accordingly, the free energy perturbation is not equal for all residue types. For example, the cost of a single point mutation to valine can range from 0 (at Ile^{76}) to ~1.6 (at Leu^{75} and Thr^{87}) to ~3.8 kcal mol^{-1} (at Gly^{79} and Gly^{83}).

All mutant sequences dimerize, but to different extents

The propensity for all mutants to form dimers at experimentally detectable concentrations enabled the determination of association free energy values for all sequence permutations in this study. The experimental free energy values were used to calculate the oligomeric species populations over a wide protein concentration range. The conditions under which proteins can be characterized as “monomeric” or “dimeric” can then be determined. This result is visualized in Figure 2.2, which shows the population distribution of dimeric species for a subset of the mutants as a function of protein:detergent mole fraction. At the high mole fraction of 10^{-2}, the thermodynamic data show that even the most destabilizing mutants, Gly^{83}Ile, Gly^{83}Val and Thr^{87}Leu, are >60% dimeric. While not shown in the figure, this is also true for all mutants with $\Delta\Delta G^{o}_{Mut}$ values > 2.5kcal mol^{-1}. Many of these extremely disruptive mutations abolish the GxxxG

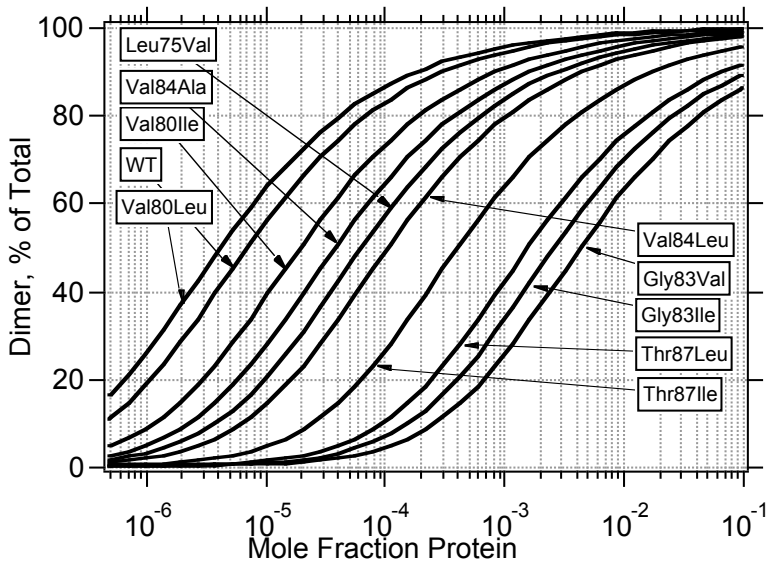

Figure 2.2 Equilibrium distributions of GpA TM dimer mutants. The percent dimer is given as a function of the total mole fraction of protein in the micellar detergent phase. The distributions were calculated using the experimentally determined ΔG_x^o values.

motif, demonstrating that the GpA TM sequence can dimerize considerably in the absence of an intact GG4. The population of these destabilizing mutants becomes >90% monomeric only when the concentration of protein in the micellar detergent is decreased100-fold (e.g. 10^{-4}). At this lower mole fraction, only the wild type and the Val^{80}Leu sequences are >90% dimeric; all other sequences have mixed populations of monomer and dimer at this concentration. The free energy data demonstrate that the populations of monomer and dimer can be significantly altered by mutations at any of the protein-protein interaction sites along the glycophorin A transmembrane sequence.

Structural consequences of mutations

To rationalize the modulation of dimer stability by sequence context, structural models were generated for each mutant, and a model-based parameterization of the free energy perturbation was carried out. The modeling procedure used was designed to minimize local steric clash and maximize local van der Waals interactions by allowing changes in side chain conformations throughout the structure. No backbone rearrangements were considered in this procedure. The computational models were used to calculate the change in favorable occluded surface area (ΔFOS), unfavorable occluded surface area (ΔUOS) and side chain conformational entropy ($\Delta\Delta S_{SC}$) for all residues for each of the mutants relative to the wild type structure. Favorable occluded surface area is interpreted as an indication of favorable van der Waals packing interactions. Unfavorable occluded surface area is interpreted as the presence of

a steric clash. Occluded surface area was scored as unfavorable when the molecular surfaces of two atoms overlapped. The conformational heterogeneity present in the family of NMR structures was accommodated by calculating mean and standard deviation values for each of the parameters for each mutant-wild type pair of the 20 NMR structures.

Coefficients for mean ΔFOS and ΔUOS parameters were determined by least squares linear regression analysis against the experimental $\Delta\Delta G^{o}_{Mut}$ values. After the initial parameterization, eight mutants with ΔUOS values of ≥ 50 A^2 were disproportionately influencing the fit. These were all leucine and isoleucine mutants at Gly^{79}, Gly^{83} and Thr^{87} as well as valine mutations at Gly^{79} and Gly^{83}. Inspection of these mutant structures suggested that the perturbations generated in the automatic modeling protocol resulted in models containing extremely large steric clashes. Since the thermodynamic data demonstrates that all of these mutants do in fact dimerize, such steric clashes would presumably be relieved *in vivo*. We hypothesize that structural relaxation would require backbone rearrangements and/or helix rotation or bending to form an alternative interface. However, computational modeling of such structures is beyond the scope of this study, and these mutants were eliminated from further consideration. Figure 2.3 demonstrates the linear correlation between the coefficient-weighted structure-based parameters and the experimentally measured free energy perturbation for the remaining mutants with the exception of two outliers. These outliers (Gly and Leu mutations at Ile^{76}) are poorly described by this automated modeling

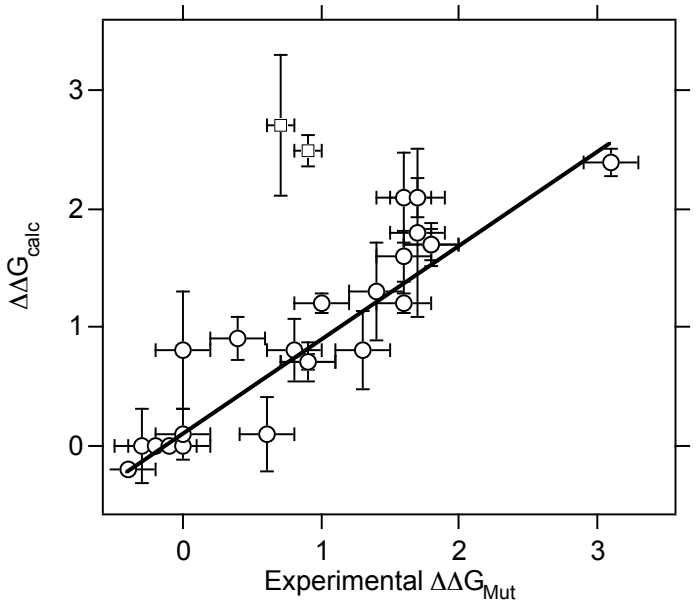

Figure 2.3 A linear correlation between parameterized structural features and free energy perturbation. The parameterized structure-based free energy perturbation is plotted versus the experimentally measured $\Delta\Delta G^{o}_{Mut}$. The linear correlation (R=0.92, $R_{jackknife}$ = 0.89) has a slope of 0.79. Open circles show the 23 mutants that were used in this parameterization. Open squares represent the application of the coefficients derived from the linear regression to the Ile^{76}Gly and Ile^{76}Leu mutants. These two mutants could not be described by this analysis even though their modeled structures appear to be chemically reasonable. The error bars in the vertical direction represent the standard deviation of the structural features amongst the 20 models of the glycophorin A NMR structure. The error bars in the horizontal direction represent the experimental standard deviation of the free energy measurement.

ΔFOS parameter was found to be 0.039, which suggests that the free energy cost to remove one square Angstrom of favorable occluded surface area is 0.039 kcal mol^{-1}.

In contrast to the buried surface area calculation, the use of the occluded surface algorithm facilitates a comparison of the nature of the structural perturbations at each residue. The energetic contributions due to changes in favorable and unfavorable van der Waals interactions and side chain conformational entropy can be calculated from each parameter using the fitted coefficient values. Figure 2.4 shows the energetic contributions for each of the mutants used in the linear parameterization. Changes in the favorable occluded surface area dominate the energetic effects. In general, the loss of favorable van der Waals interactions occurs at positions where large to small mutations were introduced (Leu^{75}Ala, Leu^{75}Gly, Ile^{76}Ala, Ile^{76}Gly, Ile^{76}Val Val^{80}Ala, Val^{80}Gly, Val^{84}Ala, Val^{84}Ile, Val^{84}Leu, Thr^{87}Ala and Thr^{87}Ser). At sites where small to large mutations were introduced, the results are varied. At Gly^{79}Ala, Gly^{83}Ala and Thr^{87}Val the destabilization of the dimer is primarily due to the introduction of large steric clashes. For the Val^{84}Ile and Val^{84}Leu mutants a loss of favorable occluded surface area is observed. A number of small to large mutant sites show a mixed contribution from the loss of favorable OS and the introduction of steric Waals interactions. Nevertheless, a small net stabilization due to changes in side chain conformational entropy is observed at several sites (Leu^{75}Ala, Leu^{75}Gly, Ile^{76}Ala, Ile^{76}Val, Thr^{87}Ala, and Thr^{87}Val).

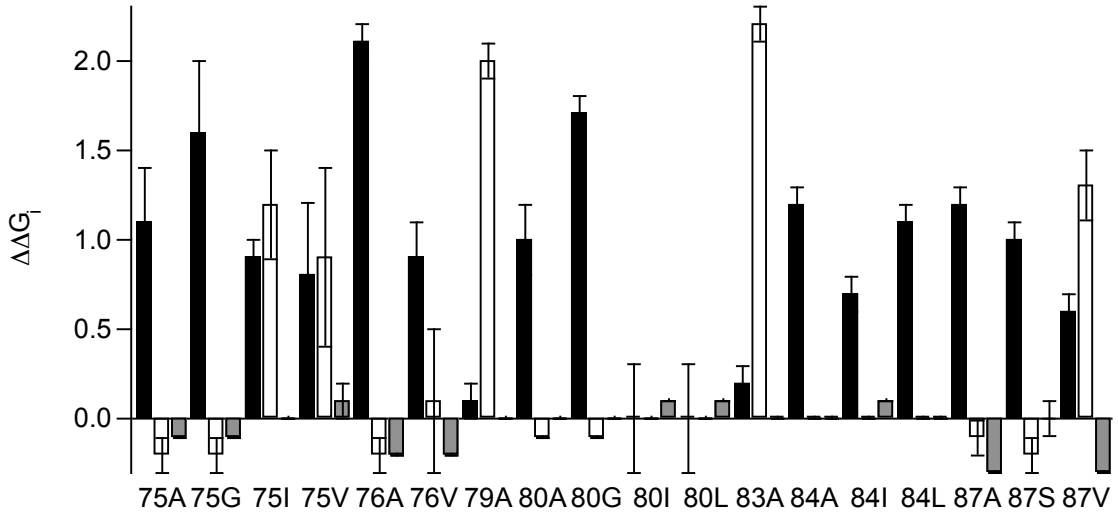

Figure 2.4 Contributions of each of the structure-based energy terms to the of $\Delta\Delta G_{calc}$ sum. Three bars are shown for each mutant. The black bars represent energy values due to changes in favorable occluded surface area. The white bars show energy contributions due to changes in unfavorable occluded surface area. The grayscale bars show contributions due to changes in the side chain entropy term. The error bars are propagated from the variation in the family of NMR structures.

The structural parameterization reveals insight into the detailed nature of the optimized wild-type interface. Changes in the same structural parameters do not explain the effects of different mutations at a particular site. For instance, mutations at Leu^{75} result in a free energy perturbation of ~1.5-1.8 kcal mol^{-1} regardless of the residue type. Structurally these effects can be rationalized by a loss of favorable occluded surface area (Leu^{75}Ala and Leu^{75}Gly) or alternatively by the combination of a loss of favorable OS and the introduction of a modest steric clash (Leu^{75}Ile and Leu^{75}Val). Of particular interest is a consideration of the effects of point mutations at Thr^{87}. The destabilizing effects of valine, serine and alanine are well explained by the combination of an introduction of steric clash (in the case of Val) and a loss of favorable OS (for Val, Ser and Ala). This correlation suggests that Thr^{87} in the wild-type sequence contributes favorable van der Waals interactions rather than forms an interhelical hydrogen bond in C_8E_5 micelles. This finding is not inconsistent with solid state NMR measurements where an interhelical hydrogen bond was observed (Smith et al. 2001; Smith et al. 2002), since the detergent micelle environment may be a more hydrated environment than the lipid bilayer in which the solid state NMR experiments were carried out.

2.4 Discussion

The GxxxG motif is not necessary for dimerization of the glycophorin A transmembrane sequence.

The energetic consequences of single point mutations at Gly^{79} and Gly^{83} are consistent with the importance of closely packed glycine residues as one defining structural feature of the GpA transmembrane dimer. Mutations at the glycine residues were previously shown to result in complete disruption of the dimer as measured by SDS-PAGE (Lemmon et al. 1992b). In addition, the Gly^{83}Ile sequence has a low signal and is normally considered to be a "monomeric" standard in the TOXCAT assay for helix-helix interaction in bacterial membranes (Russ and Engelman 1999). Indeed, the glycine residues at these sites in the glycophorin A transmembrane domain are required for the strongest dimerization of the helices to occur. However, this thermodynamic study demonstrates that the glycophorin A transmembrane domain still dimerizes when the GxxxG motif is disrupted by mutation to alanine, valine, leucine or isoleucine. The remaining residues at the protein-protein interface must encode these interactions in the dimer. Mutations in the GpA TM GxxxG motif do not result in uncontrolled protein aggregation, but rather in a defined oligomeric species. This suggests that these non-GxxxG mediated transmembrane helix-helix interactions are still specific, albeit less stable. Additional mutagenesis studies could be used to test this idea and to determine how the interface of the helix is responsible for the dimerization.

The dimerization potential of these weaker mutants was revealed only by these sedimentation equilibrium studies. The access to association free energy values for the weaker mutants reflects the ability of sedimentation equilibrium to access a wider range of protein concentrations than other techniques used to assess oligomeric states of membrane proteins. Due to the existence of a conserved hierarchy of stability for the GpA TM alanine sequence variants between C_8E_5 micelles, SDS micelles, and bacterial membranes (Fleming and Engelman 2001), we hypothesize that mutant sequences with lesser propensities to self-associate should still populate dimeric forms in biological membranes. The reduced dimensionality constraints in membranes compared to micelles is expected to further stabilize interactions of transmembrane sequences (Grasberger et al. 1986).

The sequence context modulates the stability of a GxxxG containing transmembrane sequence.

Mutations in the protein-protein interaction sequence flanking the GxxxG motif in glycophorin A demonstrate that the free energy of association can be modulated over a $\Delta\Delta G_{Mut}$ span of -0.5 to +3.2 kcal mol^{-1} in the presence of a GxxxG motif. At 25˚C, this free energy range corresponds to a shift of over 1000-fold in the species population distribution. Mutations at all of the GpA TM protein-protein interaction residues can significantly affect the dimer stability. At Ile^{76}, Val^{80} and Thr^{87} mutations exhibit a variety of free energies of association that are stabilizing, energetically neutral, or significantly destabilizing. In contrast, all

mutations at Leu^{75} and Val^{84} appear to be energetically equivalent and destabilizing regardless of the residue type.

In the sequence context of the GpA TM, Thr^{87} appears to be particularly sensitive. Mutations to leucine or isoleucine at Thr^{87} have energetic consequences similar to that of mutating either of the glycine residues in the GxxxG motif to large aliphatic residues. This thermodynamic result indicates that this threonine residue is particularly well suited to allow strong helix-helix interactions along with the GpA TM GxxxG motif. However, the sequences of several unrelated transmembrane proteins known to interact using GxxxG lack threonine at this position, but instead have alanine, valine or leucine (Arselin et al. 2003; Lee et al. 2003; McClain et al. 2001; McClain et al. 2003; Mendrola et al.; Overton et al. 2003). If the GpA TM serves as a general model for the dimerization propensity of a GxxxG mediated dimer, these proteins may dimerize to a lesser extent than GpA TM. Alternatively, the ability of threonine to strengthen a GxxxG mediated dimer may also be dependent on the remaining sequence context.

Statistical patterns of open reading frames do not always correlate with dimerization propensity.

The over-representation of pair-wise patterns in the statistical analysis does not always correlate with increased thermodynamic stability of helix-helix dimers. The Leu^{75}Gly mutation introduces a second GxxxG motif in tandem to the first, creating a GG4G4 triplet motif which is among the most overrepresented

triplets (Senes et al. 2000). The introduction of this motif leads to a lower rather than a higher stability for the transmembrane dimer. As discussed above, mutations in Thr^{87} disrupt a GxxxGxxxT triplet motif in the GpA TM. This triplet has a p-value consistent with a modestly significant over-representation in the statistical analysis of open reading frames (Senes et al. 2000). Mutations at Ile^{76} and Val^{80}, which constitute an over-represented IV4 motif, show opposing phenotypes since mutations at these sites were found to be either stabilizing or destabilizing. Thus, in the absence of a defined context, statistical motifs do not necessarily indicate a propensity for transmembrane domains to dimerize. These results suggest that statistical motifs should be studied in greater detail to determine their functional significance.

Van der Waals interactions are modulated by the sequence context

The molecular basis for the change in stability due to mutation was probed by assessing the structural consequences of mutation using computational models based on the wild-type NMR structure. The computational modeling protocol tests the hypothesis that changes in side chain conformations can account for the loss of dimer stability. Overall, a loss of local favorable inter-monomer contacts appears to play a dominant role suggesting that optimized van der Waals interactions specify the wild-type dimer interface. Since side chain conformational freedom is already restricted by the initial helix formation, a small contribution from side-chain conformational entropy is observed. Although threonine pays the highest side chain conformational entropy cost upon

dimerization, changes in packing interactions dominate the energetic effects at Thr^{87}.

The best-fit coefficient for loss of favorable occluded surface area was found to be 0.039 kcal/mol per $Å^2$. Since the occluded surface area calculation is based on a molecular surface, which is smaller than an accessible surface area, the coefficient 0.039 corresponds to the value of 0.018 kcal/mol per $Å^2$ of buried accessible surface area. This value agrees well with our earlier estimate derived from only three sequences (Fleming et al. 1997) as well as with the value of 0.026 kcal/mol per $Å^2$ of buried surface area determined for a series of mutations in the polytopic membrane protein, Bacteriorhodopsin (Faham et al. 2004). In addition, by comparison of thermodynamic stability and crystal structures, it has been suggested that the reduction of van der Waals interactions by the introduction of cavities within the interiors of several soluble proteins has an energy cost of 0.020 kcal/mol per A^2 of accessible surface area (Eriksson et al. 1992).

Dimers containing mutations at Gly^{79} and Gly^{83} create large steric clashes that are not well explained by this computational modeling protocol. These mutations appear to require backbone rearrangements to accommodate dimer formation. Due to the close inter-helical proximity of Gly^{79} and Gly^{83}, we hypothesize that mutation at these sites will lead to a disruption of interactions in the remaining protein-protein interface as a consequence of a global separation of the helices. In contrast, the structural effects caused by mutations at the non-GxxxG sites can be largely explained by local side change rearrangements. By

analogy to the role of glycines in collagen, glycines in the GxxxG motif may simply facilitate a close approach of two helices (Traub et al. 1969; Yonath and Traub 1969). The other dimer interface residues are then allowed to interact *via* significantly stronger van der Waals interactions than would otherwise be possible. The wide variation in the free energy perturbation at these sites suggests that modulation of van der Waals packing at the dimer interface by non-GxxxG residues can serve to tune the equilibrium stability to a value optimized for biological function.

2.6 Conclusions

Mutants of the GpA transmembrane helix form dimers over a wide range of stabilities. Sequences containing mutations that abolish the GxxxG motif in the glycophorin A transmembrane domain are predominantly dimeric at high protein/detergent mole fractions where structural studies are performed (Liu et al. 2003; MacKenzie et al. 1997). This result suggests that the sequence context surrounding the Glycophorin A GxxxG motif has an independent propensity to dimerize. Those mutants containing an intact GxxxG motif show free energies of association that can vary by several kcal mol^{-1}. Thus, the GxxxG motif is necessary but not sufficient to achieve strong transmembrane helix-helix association. The context of flanking residues determines the stability of the dimer that forms. Structural parameterization of the association energetics suggests that helix-helix association is modulated largely by side chain packing interactions at the dimer interface. A free energy cost of 0.039 kcal/mol was

observed for each square angstrom reduction in favorable inter-monomer occluded surface area. This value agrees with free energy estimates for the creation of a packing void in soluble proteins (Eriksson et al. 1992) suggesting that similar principles for protein stability may be acting in both transmembrane and soluble proteins. These results demonstrate the van der Waals packing interactions are a major force in the stabilization of membrane proteins, and that the stability is encoded in the amino acid sequence.

CHAPTER 3

Complex interactions at the helix-helix interface stabilize the glycophorin A transmembrane dimer

3.1 Summary

To explore the residue interactions in the glycophorin A dimerization motif, an alanine scan double mutant analysis at the helix-helix interface was carried out. These data reveal a combination of additive and coupled effects. The majority of the double mutants are found to be equal to or slightly more stable than would be predicted by the sum of the energetic cost of the single point mutants. The proximity of the mutated sites is not related to the presence of coupling between those sites. The energetic effects of weakly coupled and additive double mutants can be explained by changes in van der Waals interactions at the dimer interface. Previous studies reveal that a single face of the glycophorin A monomer contains a specific glycine-containing motif (GxxxG) that is thought to be a driving force for the association of transmembrane helices. This study provides further evidence that the relationship of the GxxxG motif to the remainder of the helix-helix interface is complex. Sequences containing mutations that abolish the GxxxG motif retain an ability to dimerize, while sequences that contain a GxxxG motif appear unable to form dimers. These results emphasize that the sequence context of the dimer interface modulates

the strength of the glycophorin A GxxxG mediated transmembrane dimerization reaction.

3.2 Introduction

The specificity of protein-protein interactions is critical to the establishment of native protein structure and to the assembly of protein complexes. A common method to study the interactions that specify native protein folds is site specific mutagenesis (Ackers and Smith 1985). The interpretation of structural effects of mutation may be simplified in membrane proteins. As compared to soluble proteins, membrane proteins exist in a constrained environment in which the association of helices can be considered independently from the formation of secondary structure elements (Popot and Engelman 1990). Thus, an advantage to the study of helical membrane protein folding is that the unfolded state may be considered a stable helix monomer, whereas the unfolded state in soluble proteins explores a greater conformational space (Creamer et al. 1995; Shortle 1996). By comparing the structural and thermodynamic effects of mutants in soluble and membrane proteins, it may be possible to elucidate whether similar principles determine the native fold in such different environments.

The human erythrocyte protein glycophorin A (GpA) contains a single transmembrane domain that associates to form a symmetric homodimer. Multiple thermodynamic studies using qualitative (Russ and Engelman 1999) and quantitative techniques (Fleming et al. 1997) have been carried out to probe the GpA TM dimerization reaction. These experiments resulted in a prediction of a

sequence motif involved in dimerization (Lemmon et al. 1992b), which was found at the dimer interface when the NMR structure was solved (MacKenzie et al. 1997). Initial mutational analysis suggested that the glycines within the glycophorin motif were the most critical residues for dimerization (Langosch et al. 1996). Genetic and statistical screens lead to the hypothesis that the interactions in the glycine motif (GxxxG) are the primary force driving the helices to associate and that the presence of this motif could mediate strong association in other transmembrane alpha helices (Russ and Engelman 2000; Senes et al. 2000). In contrast, large scale mutagenesis has shown that an intact GxxxG motif is not necessary for dimerization of the GpA TM (Doura et al. 2004). These results suggest that the adjacent amino acids at a GxxxG dimer interface can be a determinant for the strength of dimerization. Current simple models of helical protein interaction suggest that both packing moments (Liu et al. 2004) and the presence of small side chains (Jiang and Vakser 2004) can be predictors of a helix-helix interface. However, previous thermodynamic studies show that the ability of helices to dimerize is governed by complex principles and involves detailed interactions at the helix-helix interface (Doura et al. 2004).

Site-specific mutagenesis has been used extensively in biochemistry to determine the effect of a residue on the structure and function of a protein. The effect of a single point mutant may be attributed to many interactions between the mutated site and the remaining sequence. Double-mutant cycles were first used to more directly probe the interaction between two residues (Ackers and Smith 1985). Using this method, one can determine whether an interaction

between sites exists and quantify the strength of that interaction. Double-mutant cycles have been used in several soluble proteins (Chen and Stites 2001) to probe cooperativity in folding (Horovitz and Fersht 1992) and enzymatic activity (Mildvan et al. 1992). In simple cases, the presence of coupling can be explained by the proximity of the mutated residues (Wells 1990). However, coupling can also be due to long-range electrostatic interactions (Perry et al. 1989) established through both direct and indirect pathways (LiCata et al. 1993). While there is no direct correlation between the distance between mutated sites in the native structure and the energy of coupling in soluble proteins, coupling is more likely to occur when residues are less that 12 Å apart (Fodor and Aldrich 2004). In addition, some non-additive interactions are believed to be caused by interactions in the denatured state (Green and Shortle 1993). The ensemble of conformations available to an unfolded protein in solution is much greater than that available in a membrane environment. The decreased ambiguity in the denatured state of membrane proteins may lessen the effect of mutation on the monomeric transmembrane helix as compared to a soluble denatured helix. The intrinsic contribution of the monomer stability to the free energy of association could remain the same in a mutated transmembrane domain. Therefore, in membrane proteins the interpretation of the effect of mutation on the free energy of association can be considered as changes in inter-helical interactions only. The intent of this study is to employ double mutant cycles to investigate the role of coupling at the helix-helix interface in stabilizing the transmembrane protein glycophorin A.

3.3 Materials and Methods

Mutagenesis and Protein Purification

Double mutants were generated using the pET11A-SNGpA99 construct (Lemmon et al. 1992a) as a template for site-directed mutagenesis using the Quikchange kit (Stratagene, LaJolla CA) with the appropriate primers. All mutant SNGpA fusion proteins were purified using extractions in the detergent thesit (Fluka, Switzerland) as described previously (Lemmon et al. 1992a). Immediately before sedimentation equilibrium analysis, the SNGpA fusion protein of interest was exchanged into the desired detergent (C_8E_5, Sigma-Aldrich) by ion exchange chromatography as described previously (Fleming et al. 1997).

Sedimentation Equilibrium Analytical Ultracentrifugation

Experiments were carried out at 25ºC in a Beckman XL-A analytical ultracentrifuge. The wavelength of absorbance chosen in each experiment was dependent on the ability to observe an adequate dimer population to determine an accurate equilibrium constant. Mutants with a greater propensity for dimerization were observed at 230 nm. For those mutants with low propensity to dimerize, the wavelength of absorbance chosen for the experiment was 280 nm, which allows a higher protein concentration in the cell. Experiments carried out at both wavelengths were shown to have consistent standard state free energies of association (data not shown). A minimum of nine data sets were used in a global fitting of the data using MacNonlin (Johnson et al. 1981). The data used in

analysis consisted of three significantly different initial protein concentrations run at three or four significantly different speeds. The monomeric sigma was calculated from the amino acid composition using SEDNTERP (Laue 1992) and held constant during global fitting. Each free energy was independently measured a minimum of three times to determine an average and standard deviation.

A standard state free energy value (ΔG_x^o) is calculated by assuming an ideal dilute solution (Fleming 2002). By extrapolating to the standard state, it is possible to directly compare experiments conducted at different detergent concentrations. This facilitates the determination of an accurate free energy value by adjustment of the experimental conditions to populate both monomeric and dimeric species. The equation below is used to calculate the standard state free energy of association (Fleming 2002).

$$\Delta G_x^o = -RT\ln(K_{app}[micellarDet]_w) \qquad (3.1)$$

Calculation of the additivity threshold

The $\Delta\Delta G_{MUT}^o$ due to mutation was determined by taking the difference between the average standard state free energy of association for the WT and mutant protein as shown in the following equation:

$$\Delta\Delta G_{MUT}^o = \Delta G_{x,MUT}^o - \Delta G_{x,WT}^o \qquad (3.2)$$

The free energy of coupling was determined by comparing the free energy perturbation for a double mutant to the sum of the previously determined free energy perturbations for the corresponding single point mutants as follows:

$$\Delta G_{Coup}^o = \Delta\Delta G_{MUT1MUT2}^o - \Delta\Delta G_{MUT1}^o - \Delta\Delta G_{MUT2}^o \qquad (3.3)$$

A $\Delta G^{o}_{Coup} = 0$ ± *threshold* indicates an additive mutation, where the threshold represents the limiting value for additivity based on propagation of the experimental standard deviations. If the ΔG^{o}_{Coup} is outside the threshold, the double mutant is considered coupled. The standard deviation of the $\Delta\Delta G^{o}_{MUT}$ for the single point mutants was previously determined and is given as σ_{MUT} (Fleming and Engelman 2001). The standard deviation in the $\Delta\Delta G^{o}_{MUT}$ of double mutants, $\sigma_{\Delta\Delta G_{\exp erimental}}$ (Equation 3.4), is determined using the standard deviations obtained from multiple centrifugation experiments of each variant as follows.

$$\sigma_{\Delta\Delta G_{\exp erimental}} = \sqrt{\sigma^2_{\Delta G_{WT}} + \sigma^2_{\Delta G_{MUT1\,MUT2}}} \qquad (3.4)$$

Since each $\Delta\Delta G^{o}_{MUT}$ value contains the $\Delta G^{o}_{x,WT}$ term the equation for coupling can be expressed as follows.

$$\Delta G^{o}_{Coup} = (\Delta G^{o}_{x,MUT1MUT2} - \Delta G^{o}_{x,WT}) - (\Delta G^{o}_{x,MUT1} - \Delta G^{o}_{x,WT}) - (\Delta G^{o}_{x,MUT2} - \Delta G^{o}_{x,WT}) \qquad (3.5)$$

Two $\Delta G^{o}_{x,WT}$ terms can be canceled out, and the equation can be simplified by rearrangement of terms into:

$$\Delta G^{o}_{Coup} = \Delta G^{o}_{x,MUT1MUT2} - \Delta G^{o}_{x,MUT1} - \Delta G^{o}_{x,MUT2} + \Delta G^{o}_{x,WT} \qquad (3.6)$$

The threshold for the ΔG^{o}_{Coup} is therefore calculated by the square root of the sum of the standard deviation of each component as follows:

$$threshold = \sqrt{\sigma^2_{\Delta G_{WT}} + \sigma^2_{\Delta G_{MUT1}} + \sigma^2_{\Delta G_{MUT2}} + \sigma^2_{\Delta G_{MUT1\,MUT2}}} \qquad (3.7)$$

Computational Modeling and Structure-Based Parameterization

Computational modeling and structure-based parameterization were carried out as described previously (Doura et al. 2004). The coordinates (1AFO) for the WT glycophorin A NMR structure were used as a basis to model the mutant structures (MacKenzie et al. 1997). Amino acid substitutions were made in the 20 NMR structures using WHAT-IF (Vriend 1990) and mild steric clashes were relieved by using the *deball* function. Structural parameters were calculated using the occluded surface algorithm version 7.2.2 (Pattabiraman et al. 1995). The occluded surface algorithm calculates three values used in the parameterization: favorable interchain occluded surface (FOS), unfavorable interchain occluded surface (UOS), and exposed surface (ES). The $\Delta\Delta G^{o}_{MUT}$ is the basis for the parameterization, therefore each WT value is subtracted from each mutant value. Exposed surface is used as the basis for a calculation of side chain conformational entropy by comparing the conformational freedom of a residue in an extended state to the conformational freedom in monomer and dimer models considering the maximum possible conformational entropy for that side chain (Lee et al. 1994; Baker and Murphy 1998). The difference between the monomer and dimer side chain conformational entropy values results in the ΔS_{sc} for a mutant model. Structure-based parameterization is than carried out using the following parameters: ΔFOS, the change in inter-monomer favorable packing interactions, ΔUOS, the change in inter-monomer unfavorable packing interactions, and $T\Delta\Delta S_{sc}$, the change in side chain conformational entropy. Each parameter is the difference of the average value for the 20 mutant models and the corresponding value for the 20 WT NMR structures. The parameterization

was done by a simultaneous fit of the single point mutant values to the following equation by floating the coefficients in equation 3.8.

$$\Delta\Delta G^{o}_{Calc} = \sigma * \Delta FOS + \alpha * \Delta UOS + (-T\Delta\Delta S_{SC}) \equiv \Delta\Delta G^{o}_{Mut} \qquad (3.8)$$

The previously determined values for σ and α (using only the single point mutant models) were used to predict the $\Delta\Delta G^{o}_{Calc}$ for the double mutants. These values are $\sigma = -0.039$ and $\alpha = 6.44 \times 10^{-2}$. No significant changes in the coefficients and in the correlation between the $\Delta\Delta G^{o}_{MUT}$ and the $\Delta\Delta G^{o}_{Calc}$ are observed when double mutant models were included in the parameterization of σ and α.

3.4 Results

To better understand the role of coupling in specifying the native fold of a membrane protein, a double-mutant analysis of the GpA TM was carried out. The study comprises double alanine mutations at the dimer interface. A total of twenty-one double alanine mutants were created and analyzed using sedimentation equilibrium analytical ultracentrifugation. Table 3.1 shows both the experimentally determined changes in the free energy of association as well as those calculated assuming additivity for each double mutant. Experimentally, all mutations are destabilizing with respect to the wild-type with the exception of Leu^{75}Ala-Val^{84}Ala. By comparing the free energy of association if additive to the experimentally determined free energy of association, it is possible to quantify

Mutant	Sequence	Additive $\Delta\Delta G^{o}_{MUT}$ (a)	Experimental $\Delta\Delta G^{o}_{MUT}$ (b)	ΔG^{o}_{coup} (c)	1.1.1. (d)
WT	ITLIIFGVMAGVIGTILLISTGI		0.0 ± 0.1		
75A76A	ITAAIFGVMAGVIGTILLISTGI	3.1 ± 0.2	2.4 ± 0.1	-0.7	± 0.2
75A79A	ITAIIFAVMAGVIGTILLISTGI	3.0 ± 0.2	2.3 ± 0.1	-0.7	± 0.3
75A80A	ITAIIFGAMAGVIGTILLISTGI	1.7 ± 0.2	1.3 ± 0.3	-0.4	± 0.3
75A83A	ITAIIFGVMAAVIGTILLISTGI	4.4 ± 0.2	3.6 ± 0.2	-0.8	± 0.3
75A84A	ITAIIFGVMAGAIGTILLISTGI	2.3 ± 0.2	0.1 ± 0.1	-2.2	± 0.3
75A87A	ITAIIFGVMAGVIGAILLISTGI	2.2 ± 0.2	2.2 ± 0.1	0.0	± 0.3
76A79A	ITLAIFAVMAGVIGTILLISTGI	3.5 ± 0.2	2.9 ± 0.2	-0.6	± 0.3
76A80A	ITLAIFGAMAGVIGTILLISTGI	2.2 ± 0.2	2.5 ± 0.3	0.3	± 0.4
76A83A	ITLAIFGVMAAVIGTILLISTGI	4.9 ± 0.2	3.4 ± 0.2	-1.6	± 0.3
76A84A	ITLAIFGVMAGAIGTILLISTGI	2.8 ± 0.2	1.2 ± 0.1	-1.6	± 0.3
76A87A	ITLAIFGVMAGVIGAILLISTGI	2.7 ± 0.2	2.8 ± 0.2	0.1	± 0.2
79A80A	ITLIIFAAMAGVIGTILLISTGI	2.1 ± 0.3	2.4 ± 0.1	0.3	± 0.3
79A83A	ITLIIFAVMAAVIGTILLISTGI	4.8 ± 0.3	3.6 ± 0.3	-1.2	± 0.4
79A84A	ITLIIFAVMAGAIGTILLISTGI	2.7 ± 0.3	3.0 ± 0.2	0.3	± 0.4
79A87A	ITLIIFAVMAGVIGAILLISTGI	2.6 ± 0.2	2.8 ± 0.2	0.2	± 0.3
80A83A	ITLIIFGAMAAVIGTILLISTGI	3.5 ± 0.3	3.5 ± 0.1	0.0	± 0.3
80A84A	ITLIIFGAMAGAIGTILLISTGI	1.4 ± 0.3	2.0 ± 0.2	0.6	± 0.4
80A87A	ITLIIFGAMAGVIGAILLISTGI	1.3 ± 0.2	3.7 ± 0.2	2.4	± 0.3
83A84A	ITLIIFGVMAAAIGTILLISTGI	4.1 ± 0.3	3.2 ± 0.1	-0.9	± 0.3
83A87A	ITLIIFGVMAAVIGAILLISTGI	4.0 ± 0.2	3.3 ± 0.1	-0.7	± 0.3
84A87A	ITLIIFGVMAGAIGAILLISTGI	1.9 ± 0.2	1.0 ± 0.1	-0.9	± 0.3
79L83L	ITLIIFLVMALVIGTILLISTGI	5.8 ± 0.3	4.1 ± 0.2	-1.8	± 0.4

Table 3.1 Double mutant analysis of GpA TM. The sequence for each double mutant is shown in column two. The motif for dimerization is colored in red and sites of substitution are colored in blue. (a) The $\Delta\Delta G^{o}_{MUT}$ of the double mutant predicted by the sum of the $\Delta\Delta G^{o}_{MUT}$ for the single point mutants. (b) The $\Delta\Delta G^{o}_{MUT}$ for the double mutant experimentally determined by sedimentation equilibrium analytical ultracentrifugation. (c) ΔG^{o}_{coup} is calculated using equation 3.3. (d) T is the threshold for additivity, which is calculated using equation 3.7.

the energy of interaction between the mutated sites. This value is referred to as the free energy of coupling (ΔG^{o}_{coup}). Although the vast majority of mutations are destabilizing with respect to the WT, the combination of double mutations can have a stabilizing effect as compared to the sum of single point mutant losses in free energy. The most stabilizing ΔG^{o}_{coup} is -2.2 kcal mol^{-1} for the mutant Leu^{75}Ala-Val^{84}Ala (i.e. the double mutant is more stable than the combination of the single point mutants). The most destabilizing ΔG^{o}_{coup} is +2.4 kcal mol^{-1} observed for the mutant Val^{80}Ala-Thr^{87}Ala (i.e. the double mutant is less stable than the combination of the single point mutants). Ten of the twenty-one mutants (48%) have coupling free energies that are stabilizing. Nine of the twenty-one mutants (42%) exhibit an additive ΔG^{o}_{coup}. The remaining two (9.5%) have a destabilizing ΔG^{o}_{coup}.

The free energy of coupling is stabilizing in most cases

All GpA TM double mutants retain the ability to dimerize, and none of the double mutant proteins associate more strongly than the wildtype protein (Figure 3.1, Table 3.1). The most destabilizing mutant that associates under experimental conditions, Val^{80}Ala-Thr^{87}Ala, is almost fifty percent dimeric at a protein: detergent mole fraction of 10^{-3}. The double mutant Gly^{79}Ala and Gly^{83}Ala, which contains the two most destabilizing single point mutants, retains the ability dimerize, even though this mutant does not contain the GxxxG dimerization motif. The free energy of coupling is determined by comparing the free energy of association for the double mutant protein to the sum of the free energies for the

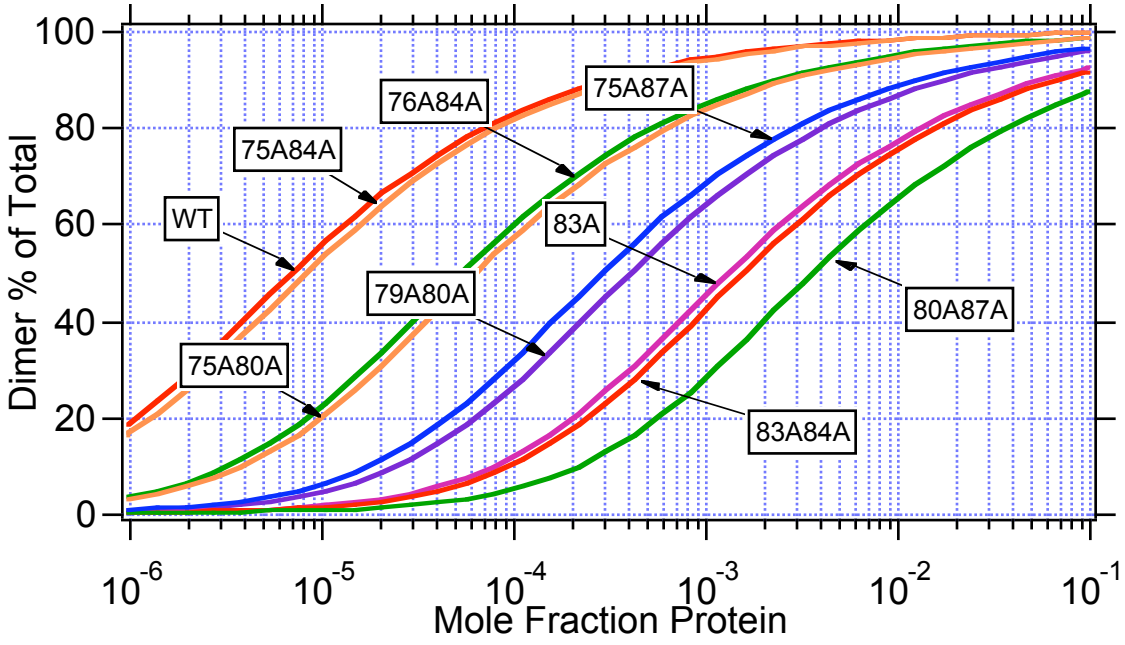

Figure 3.1 Relative dimer population of selected GpA TM mutants. The fraction dimer is plotted versus the mole fraction protein for selected mutants. The populations are calculated data based on the standard state free energy values determined by sedimentation equilibrium analytical ultracentrifugation. For reference the WT and $Gly^{83}Ala$ single point mutant distributions are shown.

corresponding single point mutant proteins (Figure 3.2). A free energy of coupling equal to zero denotes an additive interaction. The majority of mutations exhibit a stabilizing free energy of coupling. Those with the strongest stabilizing free energy of coupling are mutations that occur when the primary site mutated is Leu^{75} or Ile^{76}. The residue pair Leu^{75} and Val^{84} exhibits the most stabilizing interaction. The ΔG^{o}_{coup} between Leu^{75} and Val^{84} is great enough to result in a WT-like propensity to dimerize in the double mutant (Figure 3.1). Residue Ile^{76} exhibits stabilizing coupled effects when doubly mutated with residues Gly^{83} or Val^{84}. These data demonstrate that in many cases the double substitutions to alanine in the GpA TM allow the recovery of favorable interactions, providing for greater overall stability as compared to the sequences that contain single point mutations to alanine.

The pattern of coupling is complex

In soluble proteins, the simplest explanation for thermodynamic coupling is a van der Waals interaction between mutated residues (Wells 1990). This simple pattern of coupling can be observed by comparing the proximity of mutated residues to the presence of coupling. Interestingly, in the GpA TM there appears to be no relationship between the proximity of mutated residues and their coupling, and the coupling of residues in the GpA TM appears to follow no simple pattern. This is visualized in Figure 3.3 where each structure shows the pattern of coupling between the first substitution and all other secondary substitutions created in the double mutant proteins. Each schematic represents the coupling

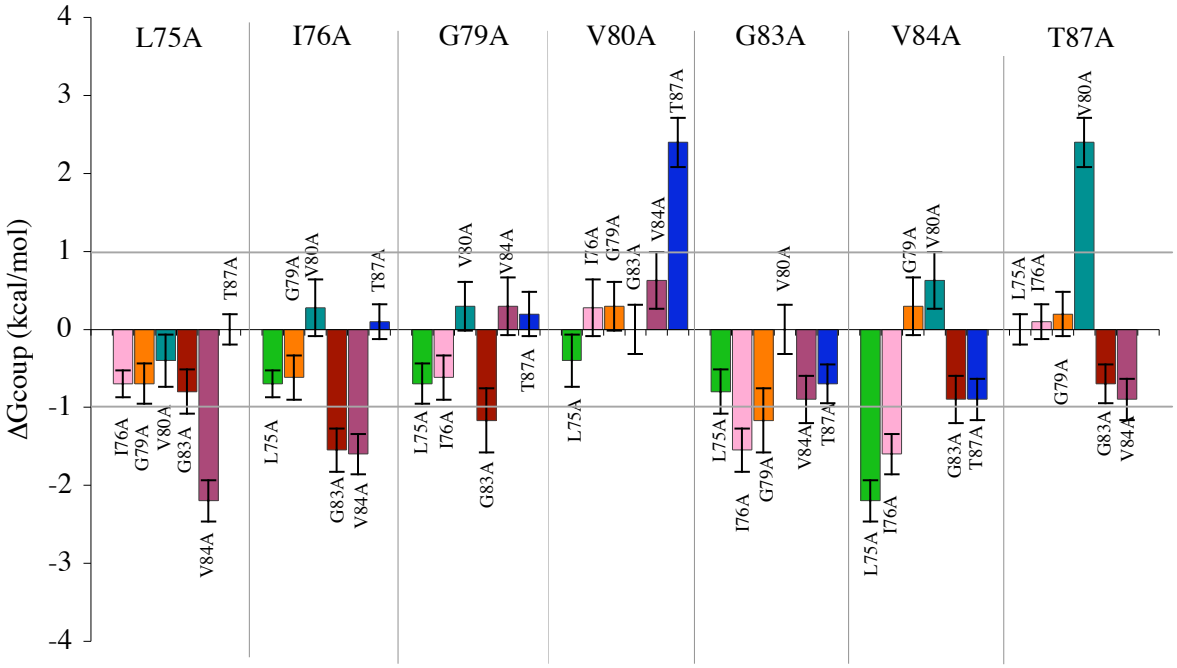

Figure 3.2 The free energy of coupling for alanine double mutants. Each bar represents the free energy of coupling for a double mutant. The first mutation in each double mutant is noted in the header and the second is the label on the bar. Each bar is colored to represent the mutation at the second site, with each site represented by a unique color. The error bars reflect the standard deviation of each measurement. A free energy of zero indicates additivity. The dark gray lines delineate weakly coupled versus strongly coupled double mutants. A $\Delta G_{coup} \leq \pm 1$ kcal mol^{-1} is considered weakly coupled. A $\Delta G_{coup} \geq \pm 1$ kcal mol^{-1} is considered strongly coupled. Note that each double mutant is shown twice.

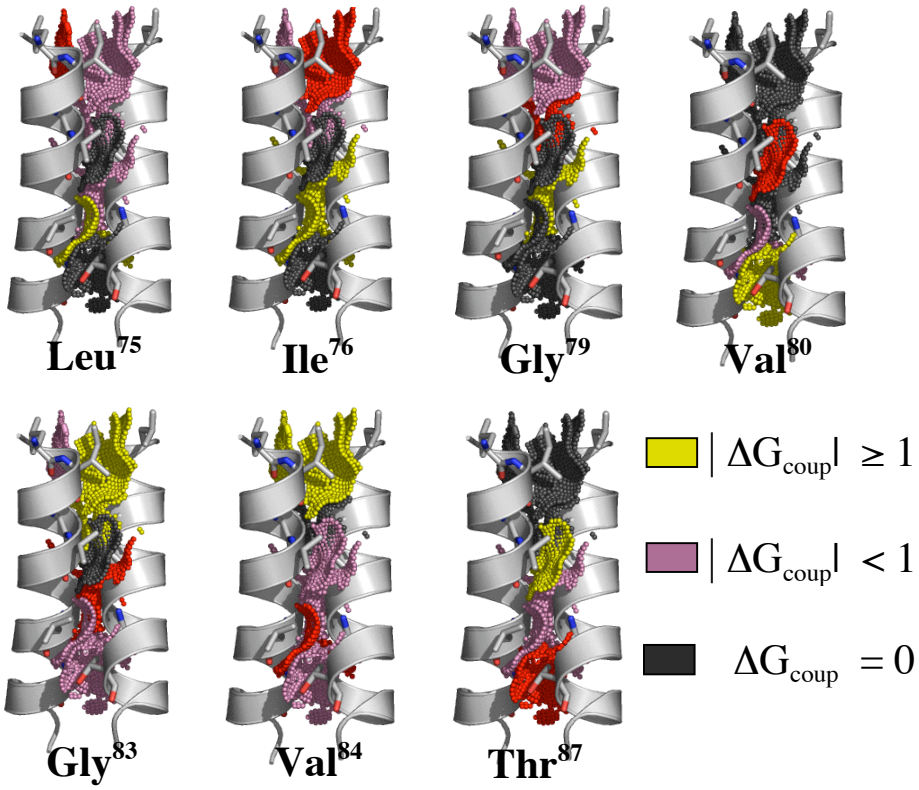

Figure 3.3 The coupling of sites along the helix-helix interface. Each GpA TM schematic structure represents the type of coupling between the indicated primary site and each secondary site along the dimerization interface. These couplings are projected onto the WT structure to compare the pattern of coupling for each site. The dot surface represents the molecular contact surface area of each residue at the dimer interface in the WT structure. The primary site of substitution is colored in red and indicated in the text below the schematic. Strong coupling is represented by a coloring of the dot surface for the secondary mutations in yellow. Weak coupling is represented by a coloring of the dot surface for the secondary mutations in pink. An additive free energy of coupling is represented by a coloring of the dot surface for the secondary mutations in dark gray.

for six double mutants for which the primary substitution is indicated in text below the model. Yellow denotes strong coupling ($|\Delta G_{coup}| \geq 1$), pink denotes weak coupling ($|\Delta G_{coup}| < 1$), and gray denotes additivity ($|\Delta G_{coup}| = 0$). Notably, residues in van der Waals contact do not exhibit strong coupling in any of the double mutants created. In contrast, GpA TM residues within van der Waals contact in the WT structure have a free energy of coupling that is either additive or weakly coupled and slightly stabilizing. Residues that are strongly coupled are distant and on opposite ends of the helix pivot point.

The pattern of coupling for each site along the interface is different from the pattern for all other sites. Leu^{75} exhibits a similar coupling profile to Ile^{76}. Val^{80} exhibits a particularly unique coupling profile. It is additive with the residues to the N-terminal side yet is coupled in a destabilizing fashion to the C-terminus. Val^{84} (the valine adjacent to the second glycine in the GxxxG motif) exhibits a different coupling profile than Val^{80}. Gly^{79} and Gly^{83} also exhibit very different coupling profiles, for instance, Gly^{79} is not coupled to as many sites as Gly^{83}. This distinction may be due to the fact that the helices are packed more tightly at Gly^{83} than they are at Gly^{79}. The difference between the coupling at the N and C-termini with Val^{80} and the distinct profiles of Gly^{79} and Gly^{83} indicate a lack of symmetry in the nature of interactions at the distal ends of the helix and within the GxxxG motif.

The entire GpA transmembrane sequence participates in dimerization

Saturation mutagenesis suggested the important motif for dimerization was $L^{75}I^{76}xxG^{79}V^{80}xxG^{83}V^{84}xxT^{87}$ (Lemmon et al. 1992b). Further studies lead to the conclusion that interactions in the GxxxG motif provide the driving force for association (Russ and Engelman 2000). It has been suggested that this motif may be an interaction motif for many α-helical membrane proteins (Kobus and Fleming 2005). However, recent studies have demonstrated that the presence of a GxxxG is not sufficient to drive dimerization. The epidermal growth factor receptors (erbBs) and CCK4 are single pass transmembrane proteins that encode a GxxxG motif. Using analytical ultracentrifugation, no homomeric preferential interactions were found for CCK4 or the erbBs. (Kobus and Fleming 2005; Stanley and Fleming 2005). The major coat protein of M13 bacteriophage (MCP) contains a GxxxG motif that has little ability to drive association in TOXCAT (Melnyk et al. 2004). Strong dimerization is only found in MCP when substitutions are made in the transmembrane sequence to include the N-terminal leucine and C-terminal threonine of the GpA motif. The C-terminal threonine was also implicated as a major component of the dimerization motif when alanine insertion mutagenesis on GpA found that the addition of an alanine at residue 86 greatly disrupts association (SN-GpA-A^{86}). This insertion causes a 100° displacement of the remaining residues. These results suggest that not only is Thr87 a critical residue for dimerization, but that the remaining sequence may also be involved dimerization.

The implications for the remaining sequence in dimerization can be found by comparison of experimentally. The sequences that flank the GxxxG motif differ from transmembrane sequences that have been used in the constructs used in different experiments (Table 3.2). In this study, a mutant was generated that contains double alanine substitutions at Leu^{75} and Thr^{87} and a truncation of the transmembrane sequence at residue 89 (SN-GpA_mult). Although this sequence is significantly shorter than the SN-GpA construct, it is comparable to the transmembrane segment length in the TOXCAT vector. Despite the truncation, SN-GpA_mult retains adequate hydrophobicity to insert stably into micelles or a lipid bilayer according to the DAS transmembrane segment predictor (Cserzo et al. 1997). In addition to changes in hydrophobicity, the shortened length of the construct may affect its insertion. However, the length of the transmembrane segment in the SN-GpA_mult construct is 17 amino acids, which is within the range for known transmembrane segments (Senes et al. 2000). Furthermore, transmembrane segments as short as 7-10 hydrophobic residues have been shown to stably insert into membranes *in vivo* (Sakaguchi et al. 1992). Therefore, we do not expect the shortened length of the SN-GpA_mult protein to strongly disrupt dimerization. Our results indicate that the SN-GpA_mult sequence has little ability to drive dimerization, even though it retains the GxxxG motif. In fact, at the highest protein/detergent mole ratio to measure preferential interactions, SN-GpA_mult is monomeric (Figure 3.4a). In conjunction with the evidence in the literature, these data indicate that a GxxxG motif alone cannot drive association of transmembrane helices. Furthermore, it

Construct	Amino Acid Sequence	Result
SN-GPA	SN-ITLIIFGVMAGVIGTILLISTGI	Strong dimer
TOXCAT-GpA	MalE–ASLIIFGVMAGVIGTTILI-ToxR′	Strong dimer
SN-GpA-A^{86}	SN-ITLIIFGVMAGVIAGTILLISYGIRRR	Weak dimer
SN-GpA_mult	SN-ITAIIFGVMAGVIGATS	No dimer
TOXCAT-MCP	MalE-AMVVIVGATIGILKFTILI-ToxR′	Weak dimer
SN-CCK4xs	SN-GMIQTIGLSVGAAVAYIIAVLGLMFYSK	No dimer

Table 3.2: Alignment of experimental constructs that test the role of the GxxxG motif. These chimeric proteins have been used to investigate the amino acid sequence contribution to dimerization. The SN-GpA construct was developed in Lemmon et al. 1992a and mutated for alanine insertion mutagenesis (SN-GpA-A^{86}) in Mingarro et al. 1996. This construct was also mutated for the SN-GpA_mult mutant in this study and SN was fused to CCK4 in Kobus and Fleming 2005. The TOXCAT vector was developed by Russ and Engelman 1999 and used to analyze the MCP transmembrane domain in Melnyk et al. 2004. The range of stability of these sequences may involve more than substitutions at the interface, but also the sequence that flanks the interaction motif.

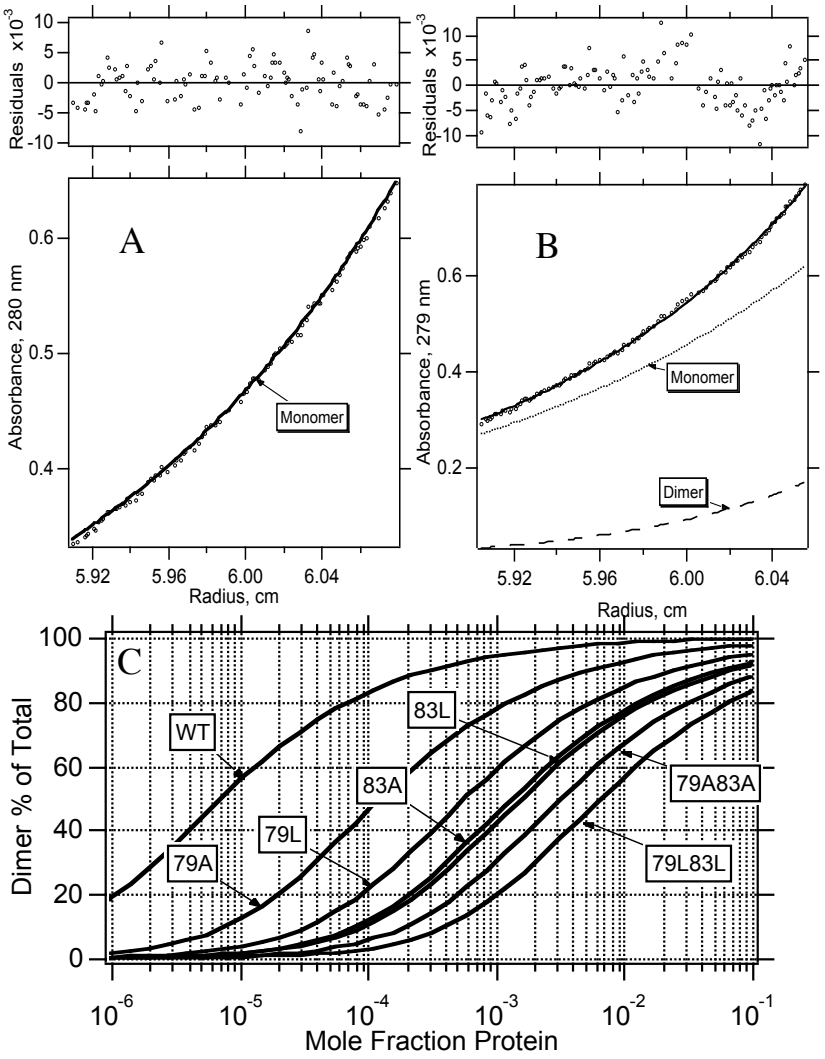

Figure 3.4 A GxxxG motif is neither necessary nor sufficient for dimerization. Panels A and B are representative sedimentation equilibrium data for selected mutants. Panel A shows a purely monomeric population for the GpA_mult mutant and the residuals of the single species fit of AUC data. Panel B shows the monomer dimer fit and the residuals for AUC data of the double mutant $Gly^{79}Leu$-$Gly^{83}Leu$. Panel C shows the relative populations of the single and double mutants affecting the GxxxG motif as a function of mole fraction protein. These distributions are calculated from the experimental $\Delta G°$ values.

is important to note that the sequence that flanks the interaction motif may also be involved in dimerization.

Large aliphatic substitutions at the GxxxG motif are tolerated

The GxxxG motif has been proposed to be the primary determinant for strong interactions in the GpA transmembrane domain (Senes et al. 2004). However, alanine substitutions at Gly^{79} and Gly^{83} still dimerize. When a double mutant to alanine is created at the GxxxG motif, the GpA TM still dimerizes and the ΔG^{o}_{coup} is stabilizing. This result may relate to the proposal that an AxxxA motif can also drive dimerization (Senes et al. 2000), and an AxxxA motif is created by alanine substitution of the glycines in the GxxxG motif. The result that the ΔG^{o}_{coup} is stabilizing for the Gly^{79}Ala-Gly^{83}Ala mutant may support the hypothesis that an AxxxA motif also drives dimerization. This may occur because alanine has a small side chain that can facilitate a close approach of the helices to maintain strong van der Waals interactions. To address this question we created a double mutant to leucine at Gly^{79} and Gly^{83}. The introduction of leucine is expected to create a large steric clash that should not allow backbone interactions at positions 79 and 83. Single point substitutions to leucine at positions 79 and 83 have been shown to dimerize with a weaker affinity than single point substitutions to alanine at those sites (Doura et al. 2004). Figure 3.4b shows that the data obtained for the Gly^{79}Leu-Gly^{83}Leu mutant is best fit to a monomer-dimer model, demonstrating that a sequence replacing a GxxxG motif with leucines also has the ability to dimerize in the GpA TM ($\Delta\Delta G^{o}_{MUT}$ = +4.1 kcal

mol^{-1}). Therefore, the substitution of large aliphatic residues in the GxxxG motif does not eliminate dimerization in the GpA TM.

The energetics of additive and weakly coupled mutants can be predicted by structure-based calculations

Structure-based parameterization using computational models has been used to explain the relationship between structural changes and energetic changes for glycophorin A single point mutants (Doura et al. 2004). To determine whether this relation applies to double mutants, a structure-based calculation was carried out using the WT GpA TM NMR structure as the basis for double mutant models. The coefficients determined in the single-point mutant parameterization were used to predict the $\Delta\Delta G^{o}_{MUT}$ for a double mutant based its computational model. Of the twenty-one double mutants, the association free energies of fifteen can be predicted using structure-based calculations (Figure 3.5) (R=0.89, p=5.364e-05). All fifteen are either additive or weakly coupled. In contrast, strongly coupled double mutants do not have free energies of association that are well predicted by the structure-based calculation. To test whether the predictive value of the parameterization can be improved, an additional parameterization was carried out that included both double and single point mutant data. There was no significant change in the values of the coefficients or in the correlation of predicted and experimental changes in free energy.

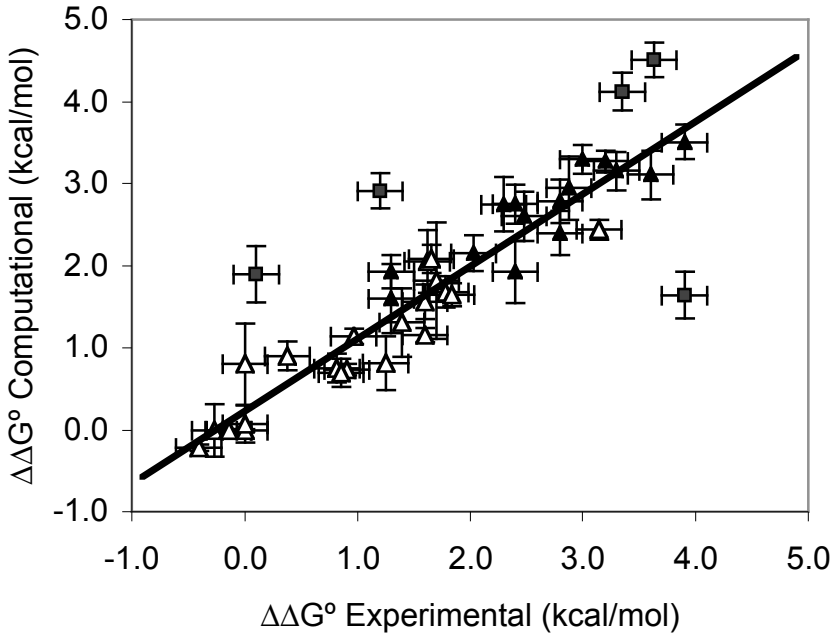

Figure 3.5 The free energy of association for double mutants can be predicted using structure-based calculations. A comparison between the free energies determined through analysis of structural parameters and those determined experimentally is shown. Single-point mutant free energies are indicated by open triangles. Weakly-coupled and additive double-point mutant free energies are indicated by closed triangles. Strongly-coupled double-mutant free energies are indicated by closed squares. The trend-line indicates the correlation between closed and open triangle data (R^2 = 0.9209, slope = 0.88). The X-value error bars are the experimental standard deviation. The Y-value error bars are based on the conformational heterogeneity in the NMR structure.

Structurally, coupling may be due to global structural changes in the dimer, an accumulation of small changes at the level of side chain conformation at the dimer interface, or structural changes in the unfolded state. Under the assumption that small conservative mutations do not affect the monomeric state of a transmembrane helix it can be inferred that coupling in the GpA TM is due to either global or local structural changes in the dimer. The modeling protocol used in this study allows for changes in side chain conformation at the dimer interface but does not allow backbone rearrangements or other large global changes in structure. The inability to use these models to predict strong coupling indicates that the most likely explanation for strong coupling in the GpA TM is global structural rearrangements in the dimer structure.

3.5 Discussion

GxxxG is neither necessary nor sufficient for dimerization in the GpA TM

Double mutant cycles allow the exploration of the sequence context of the glycophorin A motif in more detail. Using this method, we were able to determine the degree of coupling at the dimer interface to reveal how sites interact to specify a transmembrane dimer. Several previous studies attempt to develop simple models for helix-helix interactions based on statistics (Senes et al. 2000) and genetic screens (Russ and Engelman 2000). These studies postulated that the interactions at the GxxxG motif provide a driving force for helix-helix interactions in a membrane environment. Single point mutants at Gly^{79} and Gly^{83} lead to large disruptions in the propensity of dimerization in the GpA TM

(Lemmon et al. 1992b). These results confirm that these glycines are important in stabilizing the interface. However, all single point substitutions at Gly^{79} and Gly^{83} maintain a propensity to dimerize. When a double mutant to alanine is created at the GxxxG motif the protein is still able to dimerize and the ΔG_{coup} is stabilizing. Since the alanine side chain is small, this substitution may still allow a close approach of the helices in the dimer. The introduction of leucines at these sites should create a large steric clash and would not be expected to allow backbone interactions at position 79 and 83. Surprisingly, the mutant Gly^{79}Leu-Gly^{83}Leu dimerizes with significant affinity. These data indicate that a GxxxG-like motif is not necessary for dimerization in the GpA TM. Because the sedimentation equilibrium method allows a direct determination of mass, it is known this LxxxL sequence still forms a defined oligomer (dimer) that must be specified by the remaining amino acid sequence. The GpA_mult construct contains a GxxxG motif but does not dimerize with significant affinity. These results, along with the studies of the CCK4 and erbB transmembrane sequences (Kobus and Fleming 2005; Stanley and Fleming 2005), show that a GxxxG motif is not sufficient to drive association.

Long range coupling specifies native interactions in the GpA TM

Thermodynamic coupling of residues can be easily explained in those situations where the coupling of residues is correlated to the proximity of residues in the secondary and tertiary structure of the protein. Mutations at sites that are in direct contact can affect the total stability of the protein by changing

multiple components. The free energy change can be due to changes in a particular site's inherent contribution as well as alterations in the interactions between that site and other sites in the protein. Previous studies in soluble proteins have shown that residues outside of van der Waals radii are largely energetically independent (Chen and Stites 2001). Since glycophorin A is a very well packed and compact structure, it is an unexpected result that most sites are not strongly coupled although they are proximal. The greatest degree of coupling is between sites 75 and 84, which have an intramonomer distance of 14 Å and an intermonomer distance of 15.5 Å. In fact, all coupled residues are greater than 10 Å apart in the dimer, with the average of 12 Å distance between coupled residues. Therefore, no proximal residues are coupled and the average distance of coupled residues is at the limit for coupling observed in soluble proteins. Corroborating this result, recent studies have also shown that the long-range coupling between Leu^{75} and Thr^{87} are necessary to initiate high affinity dimerization in the membrane protein bacteriophage M13 major coat protein (MCP), a GxxxG containing low affinity dimer (Melnyk et al. 2004). Therefore, in contrast to soluble protein studies, the nature of the coupling along the interface of the GpA TM appears unique due to the predominance of long-range coupling stabilizing the native fold.

The different role of coupling in stabilizing soluble and membrane proteins may be a result of the distinct differences in the structures of membrane and soluble helix bundles. The helices in membrane proteins tend to be longer and have different packing angle preferences than in soluble proteins (Bowie 1997a;

Bowie 1997b). The greater degree of long range coupling in a transmembrane protein as compared to a soluble protein may be due to lesser constraints at the ends of the helices, which allows for greater conformational rearrangements as a function of mutation. A possible model for the structural effects that result in long range coupling in a transmembrane dimer is visualized in Figure 3.6. When the crossing angle is changed, there could be an effect on the interactions on the opposing end through a pivot about the crossing point of the helices. This combination of effects could lead to both positive and negative coupling between the distal ends of the helices. The function of the sequence context at the distal ends in a GxxxG mediated association could be to stabilize a specific crossing angle that allows a close approach of the helices. The presence of glycine residues at the crossing point may maximize favorable inter-monomer interactions. Therefore long range coupling could be essential for the formation of the native-like structure in a GxxxG protein.

Structural rearrangements are necessary to explain strongly coupled mutants.

Structure-based parameterization has been used to probe the relationship between structure and energetics in many soluble proteins (Baker and Murphy 1998) and in the membrane protein bacteriorhodopsin (Faham et al. 2004). Recently, a set of coefficients was generated in a parameterization using structural models and energetic data for GpA TM single point mutants (Doura et al. 2004). Using these coefficients it is possible to calculate an accurate $\Delta\Delta G^{o}_{mut}$

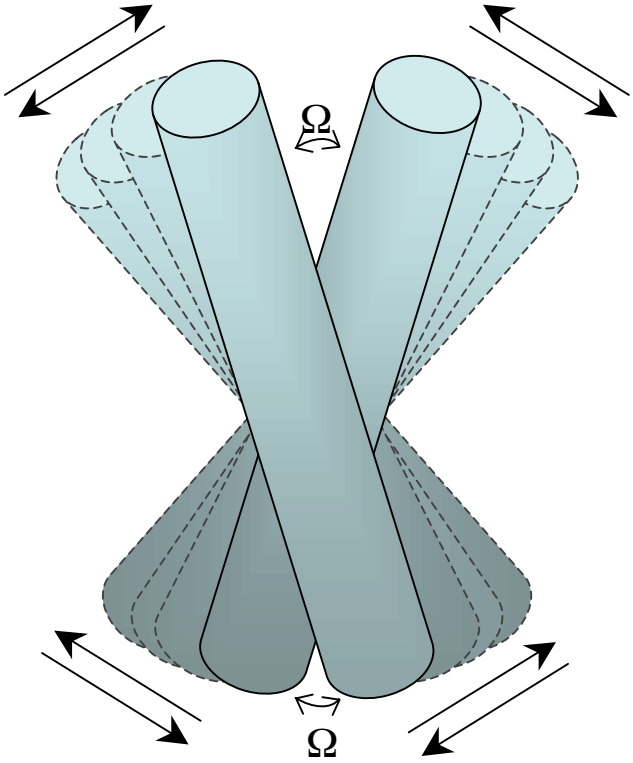

Figure 3.6 Long range coupling can be mediated through a change in the helix-helix crossing angle. The schematic illustrates a possible mechanism of long range coupling in a symmetric dimer. Changes in interactions at a distal end of the helix could have a significant effect on the opposing end mediated by a change in the crossing angle (Ω) through a movement about the pivot point.

for most GpA TM double mutants. In this study, a structure-based parameterization using both single and double mutants was also carried out. Neither the coefficients nor the correlation to the experimental free energies change significantly when double mutant models are included in the parameterization. Therefore, using a limited set of data, a correlation between calculated and experimental free energies can be determined.

The ability to predict the energetics of association for most double mutants using models based on the WT NMR structure indicates that modeling these mutations does not require global rearrangements in the structure of the GpA TM. Although many double mutants are weakly stabilizing, their energetics can still be explained using structural modeling that only allows structural rearrangements at the level of side chain conformation. Since most mutations can be explained using simple models that only deviate slightly from the NMR structure, the GpA TM dimer structure appears to be relatively insensitive to mutation even though these mutations can cause the energetics of association to vary greatly. Additionally, the parameters used to calculate the free energies are based only on changes at the helix-helix interface, validating the assumption that these mutations only affect the dimer stability. Although the molecular modeling in this study is simple, it is still possible to predict the stability of many sequence permutations. In contrast, the free energy of association for strongly coupled mutants cannot be predicted based on this simple model. We hypothesize that it is necessary to introduce global structural rearrangements to explain the interactions that stabilize strongly coupled mutants. The strength of a study

comparing experimentally determined free energies to those calculated based on simple models is that it may be possible to predict which mutations are inducing global structural changes. Conversely, without experimentally determined free energies it is not possible to distinguish the probability that a structural model is valid. In the next chapter we will test whether the relationship between structure and energetics determined by structure-based parameterization can serve as a method to compare more comprehensive models for a particular GpA mutant. If a valid model can be calculated, it may be possible to predict structural rearrangements induced by mutation.

3.6 Conclusions

The affinity of the glycophorin A transmembrane dimer can be greatly affected by the presence of mutations at the helix-helix interface. Double point substitutions to alanine exhibit a range of effects. Although all double mutants are destabilized with respect to the WT, few double mutants are coupled in a destabilizing fashion. Most double mutants are only weakly coupled or additive. Double mutant cycles reveal that long-range coupling is essential for the native structure and association of the GpA TM. Structure-based parameterization reveals that mutations at those sites that are strongly coupled are likely to induce global structural changes. Moreover, we find that a GxxxG motif is neither necessary nor sufficient for dimerization to occur. The interactions along the dimer interface appear complex and are greatly modulated by sequence context.

CHAPTER 4

Packing defects in the glycophorin A transmembrane dimer induce rearrangements of the dimer interface

4.1 Summary

To study the relationship between structure and energetics, it is necessary to have a system that is amenable to both thermodynamic and structural analysis. One particularly useful model system has been the glycophorin A transmembrane dimer (GpA TM). The NMR structure has been solved and extensive energetic measurements of dimer stability have been made. However, the effect of amino acid substitutions on the global structure of GpA has not been investigated experimentally. To test if rearrangements occur in the dimer interface, we employed site-engineered disulfide crosslinking in GpA mutants. Most mutants tested show a crosslinking profile indicative of a WT-like interface. Although, some double mutants appear to create global structural rearrangement of the dimer interface. Surprisingly, we find that single mutants at the N-terminus, which create large packing defects in the dimer interface, can induce these alternate interfaces. Using computational modeling, we predict that global rearrangements in dimer structure alter favorable packing interactions that establish the dimer stability for some mutants. These results demonstrate that mutations in the GpA TM can induce global structural changes in the dimer, and

that role of amino acid sequence in stabilizing and specifying the dimer interface is complex.

4.2 Introduction

Amino acid substitutions have been used extensively to elucidate the relationship between sequence, stability, and structure (Ackers and Smith 1985; Turner et al. 1992). To investigate this relationship, it is useful to have both structural and energetic data for a particular protein (Baker and Murphy 1998). Both energetic and structural characterization of membrane proteins has lagged significantly behind that of soluble protein structures (Lau and Bowie 1997; White 2004). The added constraints of the lipid bilayer imply that the principles of membrane protein folding may be different than soluble protein folding, and that the role of amino acid sequence in specifying the native fold may be dependent on the apolar environment (Popot and Engelman 2000). Although the number of membrane protein structures has increased in recent years, most of these proteins are not characterized energetically due to difficulties in reversible folding conditions for membrane proteins (Haltia and Freire 1995). Therefore, the comparison of sequence, stability, and conformation of membrane proteins is limited to model systems that can be characterized both structurally and energetically.

The most well characterized α-helical membrane proteins are bacteriorhodopsin (BR) and glycophorin A (GpA). BR has been characterized structurally by crystallography (Luecke et al. 1999) and energetically by

reversible denaturation in SDS (Chen and Gouaux 1999). The relationship between structure and energetics has been compared for BR alanine substitutions (Faham et al. 2004) and a large set of GpA mutants (Doura and Fleming 2004; Doura et al. 2004). The studies on both GpA and BR directly correlate the energetic cost of a mutation to changes in the buried surface of the protein, which can be interpreted as changes in favorable packing interactions. In the BR study, the energetic consequences for some mutants could not be explained by structural changes in the native state. Therefore, the energetic consequences must occur through changes in the denatured state of BR. In the study of a simple α-helical dimer, such as GpA, dramatic changes in the structure of the denatured state might be avoided. In this case, the relationship between dimer stability and conformation can be compared directly.

Structural parameters calculated from double mutant models were used to correlate energetics to changes in the GpA TM dimer structure (Chapter 3; Doura and Fleming 2004). Although most double mutants correlate well to structural models, the energetics for strongly coupled double mutants could not be predicted based on simple structural models. The modeling protocol did not allow global structural rearrangements from the WT; therefore, this lack of correlation may arise from inaccurate models. The global structural changes promoted by these mutants may be limited. In a symmetric homodimer, there are a discrete number of geometric movements that can be observed, including: (R) the rotation of helices about the interface; (Ω) a change in the helix-helix crossing angle; (D) a change in the distance between the helices; (C) and a change in helix

curvature (Figure 4.1). Therefore, the structural consequences of mutations in the GpA TM may be more easily ascertained than those in a larger polytopic membrane protein.

The structural changes induced by substitutions in the GpA transmembrane sequence have yet to be investigated experimentally. In this study, we employ a biochemical technique, disulfide-engineered crosslinking, to test for structural rearrangements of the GpA dimer interface. Cysteine crosslinking has been used extensively in both membrane and soluble proteins to determine the proximity of residues in the native fold (Wu and Kaback 1996; Hughson et al. 1997; Nagy et al. 2000; van Montfort et al. 2001; Ermolova et al. 2003). This technique could be used to screen variants for changes in the interface and can provide quick structural information on large numbers of sequences. The resulting data not only provides direct information on the conformation of the protein, but can also be used to predict more accurate structures using computational modeling (Adams et al. 1995).

Modeling is a powerful technique to predict membrane protein structure (Lehnert et al. 2004). The computational model of glycophorin A (Adams et al. 1996) correlated well to the NMR structure that was solved later (MacKenzie et al. 1997). A combination of molecular modeling and disulfide crosslinking should be useful in determining structural rearrangements in GpA mutants. By comparing energetic and structural data it may be possible to understand how a given amino acid sequence specifies both the stability and conformation of a protein.

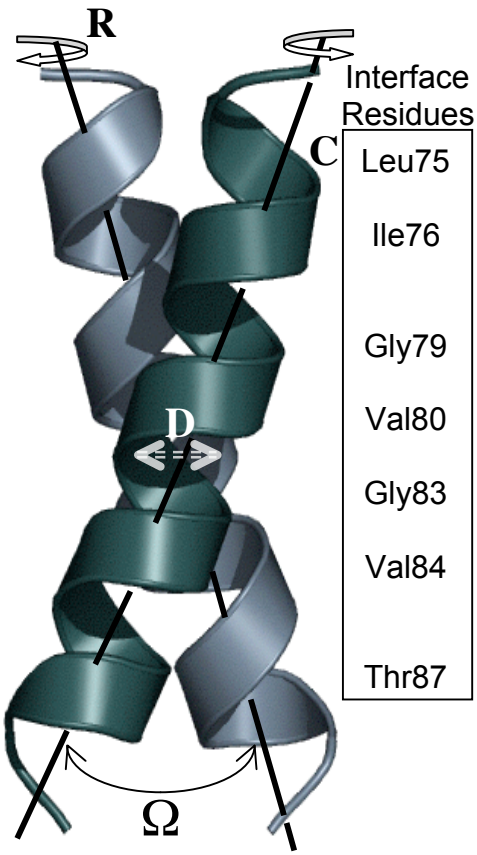

Figure 4.1 Possible structural rearrangements in the GpA symmetrical homodimer. A limited number of geometric rearrangements could explain energetic effects of strongly coupled mutants. (R) Rotation of helices about the interface. (D) Distance of closest approach of helices. (Ω) Crossing angle of helices. (C) Curvature of helix from the normal. The residues listed on the right are the critical residues for interaction in the WT GpA TM, which are substituted with alanine residues in the described experiments.

4.3 Materials and Methods

Cloning, Expression, and Protein Purification

The GpA fusion construct with *Staphylococcal* nuclease was used in this study (Lemmon et al. 1992a). SNGpA sequence variants were generated using the pET11A-SNGpA99 construct as a template for site-directed mutagenesis using the Quikchange kit (Stratagene, LaJolla CA) with appropriate primers. All mutant SNGpA fusion proteins were purified using extractions in the detergent thesit (Fluka, Switzerland) with a single salt extraction in 0.5M NaCl (Lemmon et al. 1992a).

Crosslinking and Gel Densitometry

Crosslinking experiments were carried out using the oxidizing agent $Cu^{II}SO_4$ and 1,10 phenanthroline at a final concentration of 1mM and 3mM. Reactions were carried out with 250mM NaCl and 1% Thesit. Reactions were quenched by the addition of NEM, EDTA, and SDS-PAGE gel loading buffer. Samples were then boiled for 4 minutes at 95ºC and run on a 10% acrylamide gel.

Gel Densitometry was performed using the Kodak 1D analysis software. This program calculates the relative intensity of bands in a given lane. This was used to compare the relative increase in dimer band intensity over time while under oxidizing conditions. To determine the rate of crosslinking it was necessary to determine the timescale over which the reaction rate was linear. This was

determined by graphing the increase in dimer intensity over time. The rate of crosslinking is calculated as the increase in dimer intensity divided by the dimer intensity at last time point in the linear range. For the 91Cys mutants, the crosslinking reaction was linear until 1 minute, whereas the 81Cys mutants displayed a linear rate over 15 minutes.

Computational Modeling

Structural models were generated using the program CHI (Adams et al. 1995) and the CHARMM parameter set (Brooks et al. 1983). CHI uses simulated annealing and energy minimization to generate clusters of low energy structural conformations. An average structure that represents each low energy cluster is calculated. A full search for all possible homodimers was performed a minimum of five times to fully explore all possible models. A representative structure for each variant was determined as follows: first, asymmetrical models are eliminated, because they are likely to form higher order oligomers; second, all structures that did not correlate to the crosslinking data were eliminated.

It is then determined whether these models correlate to the structure-based parameterization carried out on single and double mutants of glycophorin A (Doura and Fleming 2004; Doura et al. 2004). Structural parameters are calculated using the occluded surface algorithm version 7.2.2. (Pattabiraman et al. 1995). Occluded surface is based on molecular surface area and is related to buried surface area. This program allows the calculation of both favorable occluded surface (FOS) and unfavorable occluded surface (UOS) on a per

residue basis. Side chain conformational entropy (S_{sc}) is calculated using the exposed surface per residue value (ES) given by the occluded surface program. The S_{sc} of each monomer is subtracted from the S_{sc} of the dimer to give the change in side chain conformational entropy upon dimerization (ΔS_{sc}). The values for these parameters calculated for the WT NMR structures are subtracted for the values calculated for the mutant models. The linear correlation between the $\Delta\Delta G$ determined experimentally and the $\Delta\Delta G$ calculated based on structural parameters are based on a global fit of 50 mutants to determine a single set of coefficients that describe all mutants. These values $\sigma = -0.039$ and $\alpha = 6.44 \times 10^{-2}$ used to calculate free energies for the CHI generated mutant models in equation 4.1.

$$\Delta\Delta G^{o}_{Calc} = \sigma * \Delta FOS + \alpha * \Delta UOS + (-T\Delta\Delta S_{SC}) \qquad (4.1)$$

δ describes whether the calculated free correlates to the experimentally determined free energy of association. It is calculated by subtracting the calculated free energy from experimentally determined free energy as in equation 4.2.

$$\delta = \Delta\Delta G_{\exp} - \Delta\Delta G_{calc} \qquad (4.2)$$

Models with δ values less than 1 kcal mol^{-1} are considered consistent with the structure-based parameterization.

Crossing angles between the helices of the dimer are calculated using WHATIF (Vriend 1990). Helix curvature is calculated using *helanal* (Bansal et al. 2000). Interface residues are determined using OS v7.2.7 (Pattabiraman et al. 1995), and are considered to be those residues that occlude surface on the

opposing helix. These parameters are used to describe the models, and compare between mutant and WT structural features. To visualize structural rearrangements, mutant models are aligned to a WT model using MacPymol (Delano Scientific L.L.C.). The WT model is the seventh model of the NMR structure deposited in the PDB. This model best represents the average structure of the 20 NMR structures in the PDB. Density plots (Figure 4.8) were generated using *Mathematica* (Wolfram Scientific).

4.4 Results

Site engineered disulfide cross-linking can test interface rearrangements.

Crosslinking can be used to test whether certain sites are proximal in the native structure (Hughson et al. 1997; Nagy et al. 2000). In a symmetrical homodimer, such as glycophorin A (GpA), it is possible to make a single cysteine substitution and test its ability to form an intermonomer disulfide linkage. This method can be useful to determine whether this site is at the dimer interface. A cysteine residue should crosslink the dimer if it is directed towards the opposing helix; it should not crosslink if it points away from the opposing helix toward the lipid face. As a proof of concept, cysteine substitutions were first made at sites that are known to be critical interface residues, Gly79 and Thr87. Mutations at these sites can be extremely disruptive (Lemmon et al. 1992b; Fleming et al. 1997; Doura et al. 2004); although, we expect that these substitutions do not induce a rearrangement of the dimer interface (Doura and Fleming 2004). Purified proteins were exposed to the oxidizing agent, copper-phenanthroline

(CuOP), in the presence of ambient oxygen. It is important to note that for some variants dimer is observed by SDS-PAGE in the absence of oxidizing agents, due to a strong affinity for dimerization. No protein was found to crosslink in ambient oxygen. Thus, this dimer is not a covalent crosslinked dimer, but a representation of the native monomer-dimer equilibrium. Consequently, the ability for a cysteine residue to crosslink the dimer is determined by an increase in dimer concentration after exposure to oxidizing agent, which indicates the formation of covalent crosslinked dimers.

Crosslinking occurs at critical interface sites (Gly79Cys and Thr87Cys) within 5 minutes after exposure to oxidizing agent (Figure 4.2a,b). Gly79Cys crosslinking occurs more rapidly than Thr87Cys, which may be due to a closer proximity of the 79Cys residues in the dimer. These results are consistent with the known structure, which shows that the Gly79 residues in the dimer are approximately 4.7 Å apart, and that the Thr87 residues are 6.7 Å apart. Therefore, the pattern of crosslinking is consistent with the WT structure, and cysteine substitutions at interface residues are able to crosslink the dimer.

To use crosslinking to observe changes in global conformation induced by mutations in the GpA transmembrane domain, it is necessary to choose a site that facilitates an intermonomer disulfide linkage, but also does not contribute to the interface motif. For these reasons, Ile91 was chosen to test interface rearrangements. It is a site with close proximity across the dimer interface. This residue is not considered part of the glycophorin A motif, since mutations at Ile91

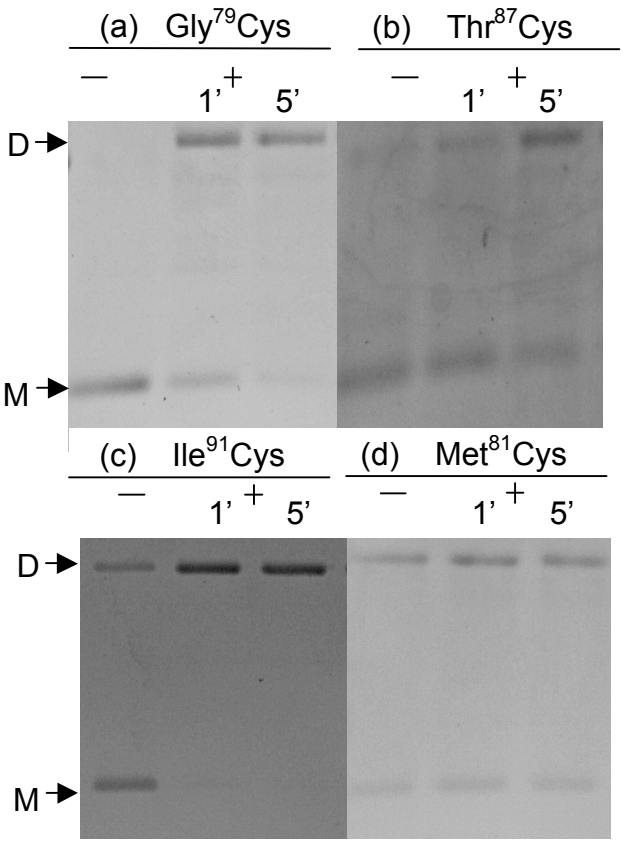

Figure 4.2 Crosslinking of interface and diagnostic cysteine substitutions. SDS-PAGE of purified protein crosslinked with 3mM CuOP. Samples were taken at one and five minutes as indicated. (-) Protein in buffer only; (+) Protein with oxidizing agent. Proteins with a strong inherent affinity for dimerization demonstrate a monomer-dimer equilibrium in the absence of crosslinking agents. Therefore, crosslinking is visualized by an increase in dimer concentration after exposure to oxidizing agent. (a-c) Residues facing the dimer interface form covalent crosslinked dimers; (d) Lipid facing residues do not form crosslinked dimers, since the monomer-dimer equilibrium is not perturbed by the addition of oxidizing agent. D indicates dimer and M indicates monomer species.

do not have a large effect on the ability for the protein to dimerize (Lemmon et al. 1992b). When 91Cys is exposed to CuOP, complete conversion to dimer occurs by 1 minute (Figure 4.2c).

For diagnostic purposes, it is useful to test sites that are both lipid facing and helix facing in the wildtype dimer. Therefore, a substitution to cysteine was made at the lipid facing site Met81. As expected, no crosslinking is observed for Met81Cys (Figure 4.2d). As shown, this mutant is not disruptive and is primarily dimeric in the absence of crosslinking agents. When Met81Cys protein is exposed to crosslinking agent, the concentration of dimer does not increase on the same time course that promotes full conversion to dimer for interface cysteine substitutions. This indicates a lack of ability for cysteines at this site to crosslink the stable dimer conformer. Since Met81Cys does not crosslink in the wildtype, it is a useful diagnostic substitution to test for crosslinking at an alternate interface. We believe that the negligible energetic disruptions caused by Met81Cys and Ile91Cys (Lemmon et al. 1992b) suggest that the addition of these substitutions to other GpA mutants should be minimally disruptive of both their energetic and structure features. Accordingly, these diagnostic cysteine substitutions to identify changes in the dimer interface for GpA mutants.

Crosslinking data probes interface rearrangements in double mutants.

A previous study on double substitutions in the GpA TM suggested that strongly coupled double mutants might undergo such rearrangements (Chapter 3; Doura and Fleming 2004). The energetics of these double mutants was not

well described by structural models based on local side chain rearrangements in the WT NMR structure. These substitutions may induce a rotation of the helices to form dimers with alternate protein-protein interfaces. To test this possibility, strongly coupled double mutants were cloned with diagnostic cysteine mutants. Crosslinking of 91Cys in the WT background promotes conversion to dimer within 1 minute; therefore, the concentration of dimer after 1 minute of crosslinking was chosen as a metric to compare the degree of crosslinking for GpA mutants.

Crosslinking results show that most double mutants retain the wildtype dimer interface. For double mutants with 91Cys, there is a rapid conversion to dimer under oxidizing conditions (Figure 4.3), similar to the results for the 91Cys mutation in the wildtype background (Figure 4.2). To determine whether the wildtype interface is specifically retained, the protein must also not crosslink in the 81Cys variant. Our results show that most mutants do not crosslink with 81Cys. While there may still be subtle structural rearrangements in these mutant dimers that cannot be detected by this insensitive technique, the data suggest that most mutants retain the wildtype interface, with 91 remaining at the helix-helix interface and 81 at the helix-lipid face.

Crosslinking data does show that alternate interfaces are present in two double mutants. These double mutants, Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala, exhibited crosslinking with both 91Cys and 81Cys (Figure 4.3). For these mutants there may be a loss or change of specificity of the dimer interface: the sequence may promote the population of an alternate interface that includes

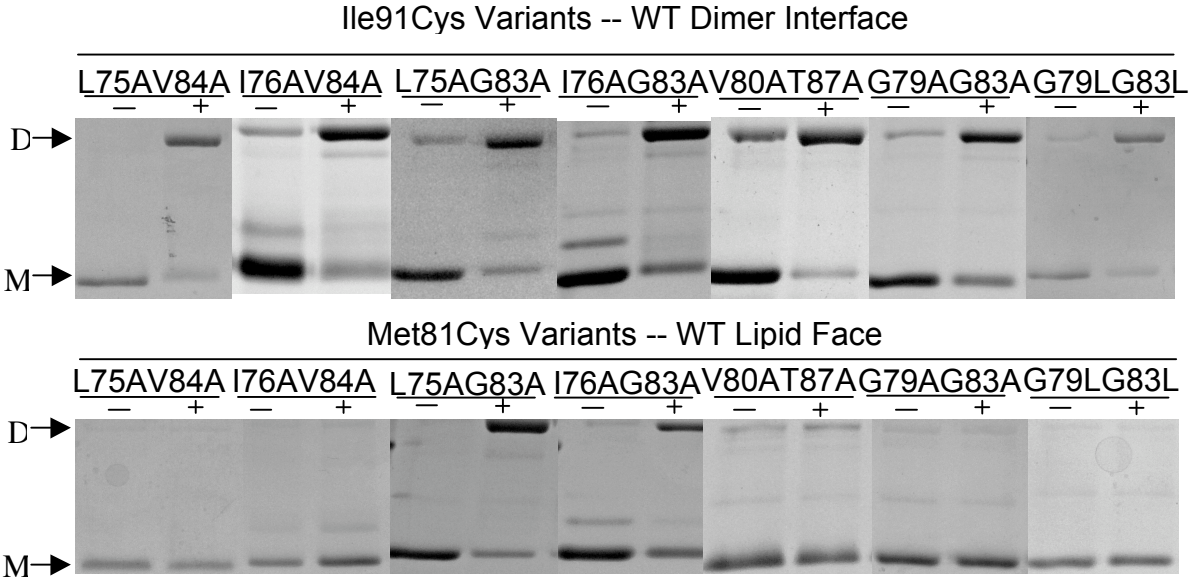

Figure 4.3 Crosslinking of double mutants at WT interface and lipid sites. SDS-PAGE of double mutants with a diagnostic cysteine mutation. (-) Protein in buffer only, (+) Protein with 1mM CuOP. Most double mutants with a cysteine at position 91 crosslink to dimer, whereas a cysteine at position 81 does not rapidly form crosslinked dimer in most proteins. Two double mutants (Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala) exhibit crosslinking with both 91Cys and 81Cys variants. D indicates dimer and M indicates monomer species.

both 91Cys and 81Cys; or the sequence may allow population of two distinct forms of dimers. We hypothesize that the form captured by 91Cys crosslinking reflects interactions at the canonical WT GpA interface, and the form captured by the 81Cys crosslinking suggests interactions involving the residues normally at the helix-lipid interface of WT GpA. Interestingly these double mutants both have a N-terminal alanine substitution and a substitution at Gly83. Furthermore, they have similar dimer stabilities, although their free energies of coupling are quite different. Unifying features of Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala may be responsible for global structural rearrangements of the dimer interface, reflected by a greater degree of crosslinking for WT lipid-facing residues.

We also determined the rate of crosslinking, which can be used as an orthogonal measure of the reactivity of cysteines in the dimer. CuOP is known to be a promiscuous catalyst of disulfide bonds. Since it is not consumed during the crosslinking reaction, it can catalyze disulfide bonds between sites that are not near in the stable conformation, but are only proximal in transient protein conformations resulting from molecular motion (Careaga and Falke 1992; Hughson et al. 1997). Therefore, a change in reactivity may result from an alteration in the conformation of the stable state or in a change in molecular motion. A large change in the rate of crosslinking may be indicative of a change in the stable structure, while small changes in the rate of crosslinking may be indicative of altered motion in the dimer.

The reactivity of cysteine residues in different mutant backgrounds was compared to determine if mutations induce changes in stable structure and in

protein motion. Reactivity of a cysteine is determined by calculating the rate of crosslinking over the initial linear range using gel densitometry (Figure 4.4). The considerable increase in crosslinking at 81Cys indicates that Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala populate conformations in which their Met81Cys residues are proximal in the dimer. The rate of crosslinking for some other double mutants also increases compared to that of 81Cys in the WT background, albeit to a much lesser degree. In particular, the rate of crosslinking for Val80Ala-Thr87Ala-81Cys is increased; after 15 minutes of crosslinking the protein is predominantly dimeric. Leu75Ala-Val84Ala, Gly79Ala-Gly83Ala, and Gly79Leu-Gly83Leu also show an increase in the rate of crosslinking for their 81Cys variants. It is important to recognize that the rates of crosslinking for these 81Cys variants are still significantly less than their 91Cys variants. Therefore, these mutants are not likely to cause the population of a stable alternate interface, but may promote crosslinking by a close approach of 81Cys in the dimer resulting from increased protein motion. Therefore, although these sequences most likely retain the canonical WT interface, they may have increased molecular motion as compared to WT.

Single site substitutions are responsible for population of alternate interfaces.

To determine whether the crosslinking profiles for the Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala double mutants arose from the thermodynamic coupling between the sites or was an underlying unanticipated feature of one of the

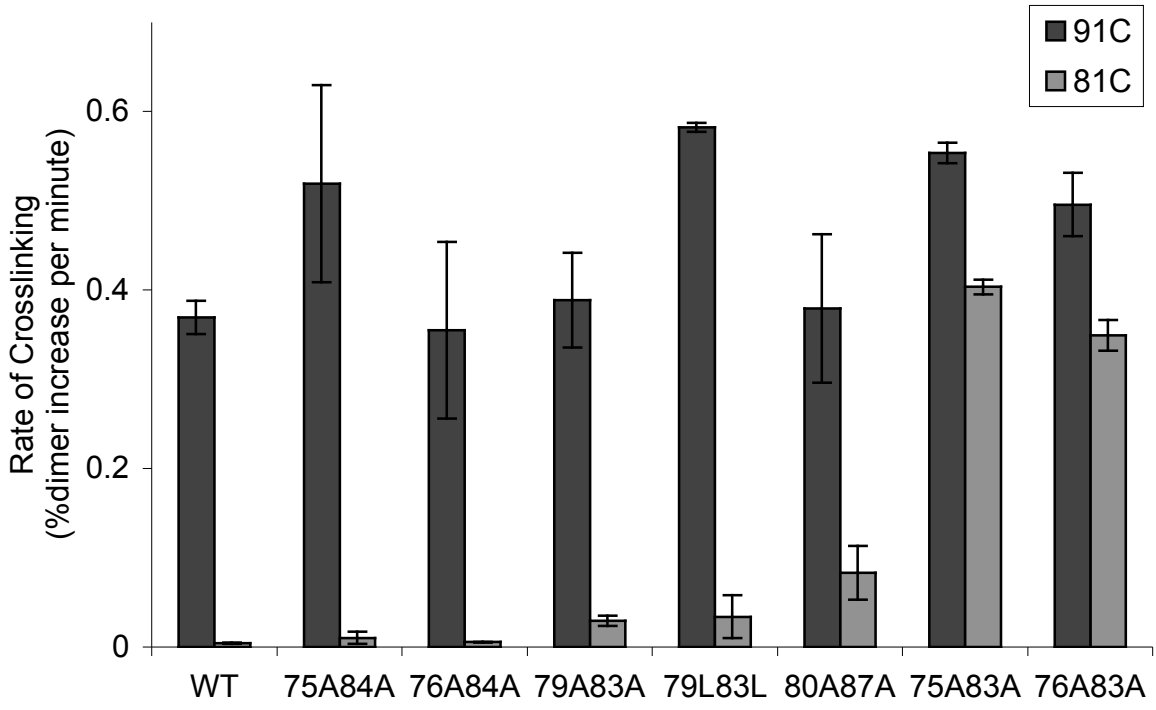

Figure 4.4 Rate of crosslinking for double mutants with diagnostic cysteine substitutions. The rate of crosslinking was determined by gel densitometry. Error bars are estimated based on duplicate experiments for 91Cys and 81Cys in the wildtype background. The rate of crosslinking for the 91Cys mutants (black) are comparable and much more rapid than the rate of crosslinking for the 81Cys mutants (gray). D indicates dimer and M indicates monomer species.

corresponding single mutants, we carried out crosslinking experiments in each of the single point mutants individually. The single mutants tested were N-terminal alanine substitutions (Leu75Ala or Ile76Ala) and Gly83Ala, which is present in both double mutants. This site is a part of the GxxxG motif, and therefore, we reasoned that this substitution might induce the observed structural rearrangements. However, crosslinking of Gly83Ala with 91Cys and 81Cys demonstrate that this mutant retains a wildtype interface (Figure 4.5). In contrast, crosslinking results demonstrate that the Leu75Ala and Ile76Ala dimers crosslink in both the 81Cys and the 91Cys variants. The rate and profile of crosslinking for these two single mutants is similar to that seen in the Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala double mutants. Because of this surprising result, other single alanine mutants were tested for the ability to promote alternate interfaces. No other single sites promote crosslinking at 81Cys. Therefore, we conclude that the alternate interface in the double mutants may be promoted by the N-terminal single site alanine substitutions, Leu75Ala and Ile76Ala, and not the important Gly83 site, or as a consequence of energetic coupling between the mutated sites in the double mutants.

Forced choice experiments demonstrate preferential interfaces.

The crosslinking data show that alternate interfaces can be promoted by alanine substitutions in Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala. To better understand the nature of this alternate interface, forced choice experiments were performed to determine preferential sites of interaction in the mutant dimers. In

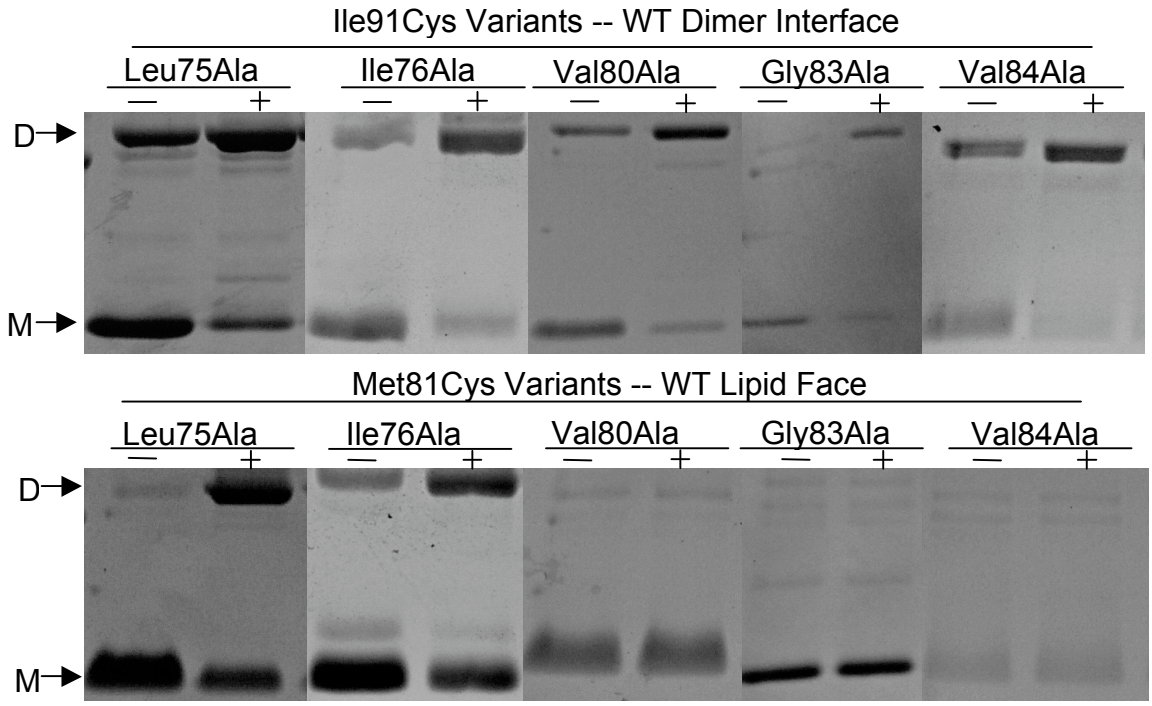

Figure 4.5 Crosslinking of GpA single alanine variants with diagnostic cysteine substitutions. SDS-PAGE of single alanine mutants with a diagnostic cysteine mutation at either 81 or 91 that were crosslinked for 1 minute with 1mM CuOP. (-) Protein in buffer only; (+) Protein with oxidizing agent. All single mutants with a cysteine at 91 crosslink, while most single mutants with a cysteine at 81 do not crosslink. Two mutants (Leu75Ala and ile76Ala) exhibit crosslinking with both 91Cys and 81Cys variants. D indicates dimer and M indicates monomer species.

this experiment, a reactive thiol can interact with a neighboring reactive thiol, or with a reactive compound in solution. If the cysteines in the dimer are not in stable proximity, crosslinking will be inhibited by exogenous DTNB (5,5'-dithio-bis(2-nitrobenzoic acid)) (Faulstich and Heintz 1995; Nagy et al. 2000). Crosslinking will be dependent on the distance between the thiols and the DTNB concentration. These experiments can demonstrate whether mutants that exhibit crosslinking at both 81Cys and 91Cys have a preference for crosslinking at one site or the other.

The 91Cys and 81Cys variants of Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala were crosslinked with increasing concentrations of DTNB (Figure 4.6). Leu75Ala-Gly83Ala-81C crosslinking is completely inhibited with addition of DTNB, whereas the 91Cys variant has only slightly decreased crosslinking in the presence of DTNB (Figure 6a). The Leu75Ala-Gly83Ala double mutant demonstrates a preference for an interface that contains residue 91. Ile76Ala-Gly83Ala demonstrates an inhibition of crosslinking in both the 81Cys and 91Cys variants (Figure 4.6b). Therefore, Ile76Ala-Gly83Ala demonstrates less preference for the canonical WT interface. The different results for the Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala forced choice since similar results were obtained when performing forced choice experiments on the single mutants (data not shown).

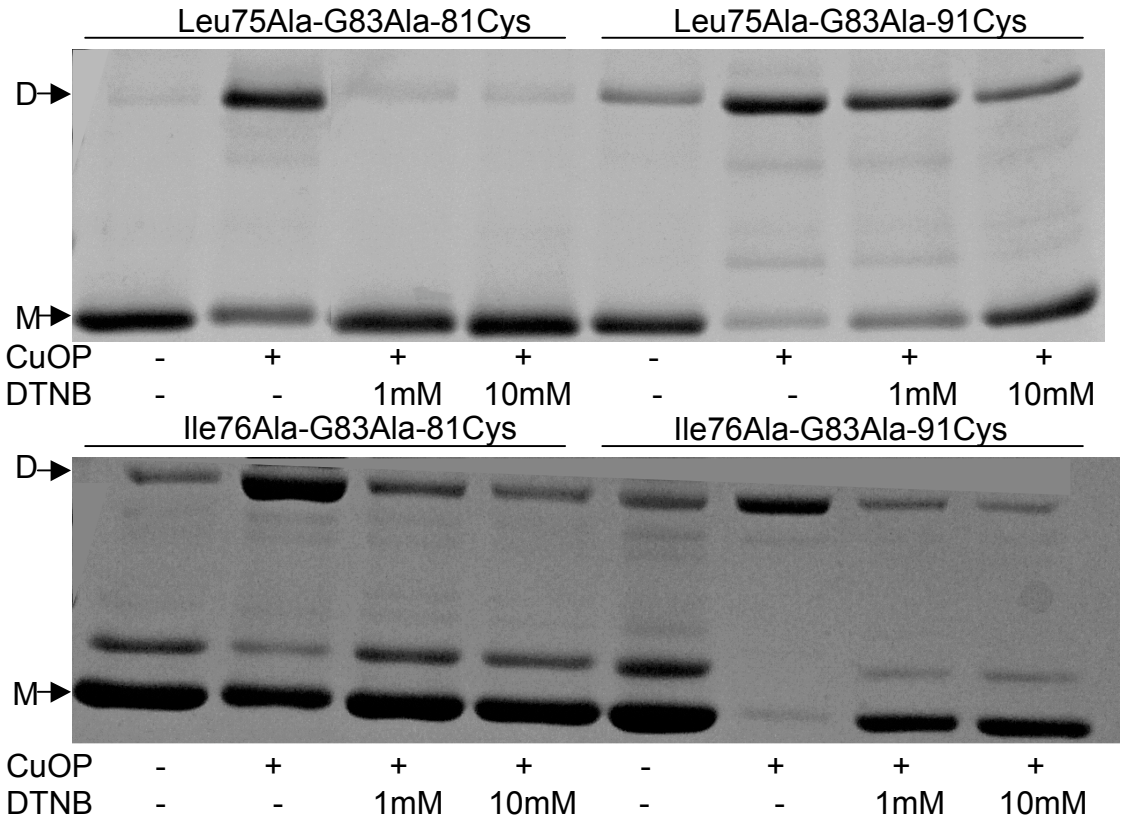

Figure 4.6 Forced choice experiments demonstrate preferential interfaces**.** These experiments compare the reactivity of cysteines at diagnostic sites. Purified proteins were exposed to 1mM CuOP for 15 minutes and increasing concentrations of DTNB. Crosslinking is completely inhibited for the Leu75Ala-Gly83Ala-81Cys variant, whereas crosslinking is favorable even in 10mM DTNB for the Leu75Ala-Gly83Ala-91Cys variant. For the Ile76Ala-Gly83Ala double mutant crosslinking is inhibited in the 91Cys variant and the 81Cys variant. D indicates dimer species; M indicates monomer species.

Modeling predicts subtle structural rearrangements in some mutants.

To test the structural rearrangements in GpA strongly coupled mutants, structural models were calculated using the established modeling program, CHI. This program was used to generate a model of glycophorin A, later found to be consistent with the NMR structure (Adams et al. 1996; MacKenzie et al. 1997), and other membrane protein oligomers (Adams et al. 1995). In contrast to our previous modeling of the GpA double mutants, CHI does not rely on the previously solved structure, and will allow bending, rotation of the helices, and changes in crossing angle to find the low energy structures. To explore all possible conformations, the program was executed a minimum of five times with different random starting velocities (Adams et al. 1995). Experimental data was used to distinguish between potential models: the chosen model must be a symmetrical homodimer, and must be consistent with the crosslinking data. Using these criteria, the number of accepted models for each mutant was very small. Some mutant sequences generated several models that were structurally similar, and the model chosen was a representative average. The previous structure based parameterization was used to calculate a free energy from the chosen model, and this value was compared to the experimentally determined free energy. This computational approach allowed the calculation of structural models for five GpA double mutants that correlated well to crosslinking data; three of those double mutants correlated to the previous structure based parameterization.

The following structural parameters were calculated from the models: helix crossing angle, distance of closer approach, and helix curvature (Table 4.1). Slight differences occur in the residues at the dimer interface and in the distance and site of closest approach between the helices. No mutant model demonstrates a crossing angle outside the range of the WT NMR structures, or a large change in overall helix curvature. To visualize global and local differences between models and the WT, the mutant models that correlated well with experimental data were aligned with a representative WT NMR structure (Figure 4.7). Most models show a global separation of the helices at the N-terminus as compared to WT, with the exception of Ile76Ala-Val84Ala. The alignment demonstrates Ile76Ala-Val84Ala shows the greatest change in helix curvature as compared to WT, allowing a closer approach of the helices at the N-terminus.

Structures also exhibit a slight rotation of the helices, although the rotation is not enough to alter the interface or allow a close approach of Met81, the residue used for outside crosslinking. Models for mutants with substitutions in the GxxxG motif, Gly79Ala-Gly83Ala and Gly79Leu-Gly83Leu are also similar to the WT NMR structure, but relieve the steric clash created by the addition of a larger side chain by allowing a global separation and slight rotation of the helices. These data indicate that many structural rearrangements can be induced by amino acid substitutions at the helix-helix interface. The energetic cost of substitutions may be caused by structural rearrangements that impact the

Model	δ (a)	Interface (b)	Crossing angle (c)	Distance of closest Approach (d)	Helix curvature (e)
WT	0±0.8	tLIxfGVxAGVxGTIxlI	-38.11± 9.37	83→4.23 Å ± 0.07	C/C
L75AV84A	-0.7	xAIxxGVxAGAIxTIxxI	-37.00	84 →3.9 Å	C/C
I76AV84A	-0.0	TLxxFGVxAGxxGTxxLI	-41.94	83 →4.0 Å	C/C
V80AT87A	1.5	xLIxFGAxAGVxxAIxLI	-39.72	80 →4.4 A	C/C
G79AG83A	2.9	xLIxxAVxxAVxxTIxxI	-38.11	80 →4.7 Å	C/C
G79LG83L	4.3	xLIIxLVxxLVxxTIxxI	-35.49	80 →4.6 A	L/L
L75AG83A	3.7	TxxIFxVMxxVIxTIxxI	34.25	81 → 4.7 Å	C/C
I76AG83A	4.4	TxxIFxxMAxVIxTIxxI	36.04	81 → 5.7 Å	C/C

Table 4.1 Analysis of structural models generated by CHI. (a) Difference between experimentally determined $\Delta\Delta G_{\exp}$ (Doura and Fleming 2004) and $\Delta\Delta G_{calc}$ calculated based on the CHI mutant model (b) Residues at the dimer interface in the model (c) Crossing angle between helices of the dimer (d) Residue with the closest approach to the opposing helix and the distance between Cα atoms of that residue on the opposing helix. (e) Curvature of helices a/b in the dimer as calculated by *helanal* (Bansal et al. 2000); qualitatively *C* is curved and *L* is linear.

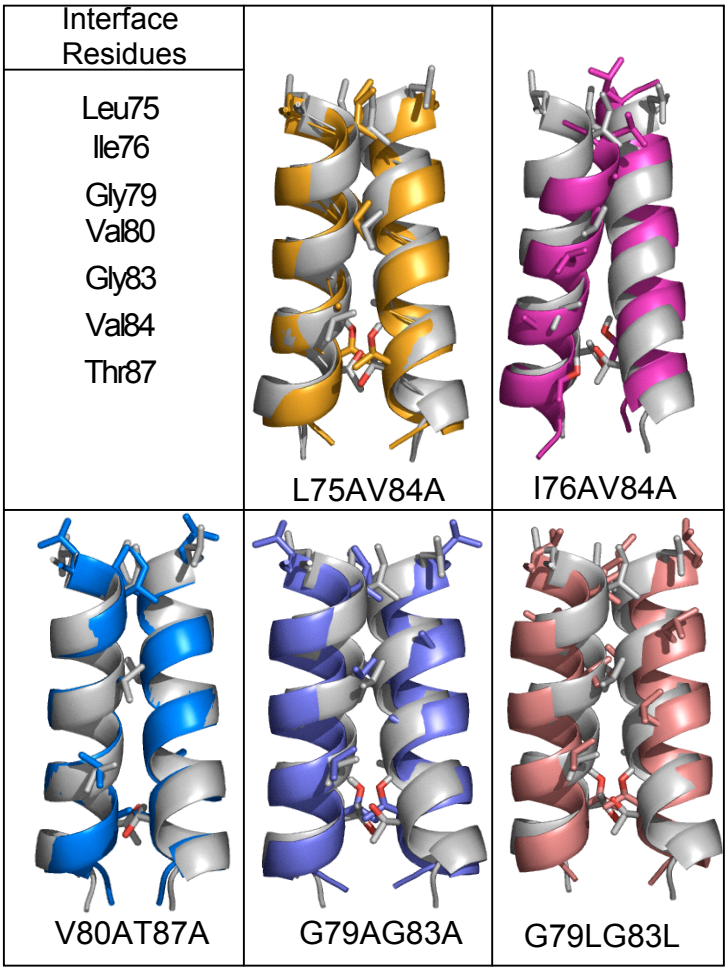

Figure 4.7 CHI generated models of strongly coupled double mutants. Each model is aligned with a WT NMR structure to illustrate the structural rearrangements in these mutants. The WT model is in gray and mutant model is colored. In both WT and mutant models, interface residue side chains are shown in a stick representation and labeled on the left according the their position in the sequence.

favorable packing interactions that stabilize the dimer.

To visualize the change in packing interactions, the average occluded surface per residue was calculated on the mutant models and the pattern of packing interactions was compared for WT and mutant models (Figure 4.8). In general, all mutants show decreased packing at residues 75 and 76, even when the mutated sites are distant. In addition, most mutants show increased packing at the C-terminus, especially at residues 88 and 91. Ile76Ala-Val84Ala shows a slight change in the helix interface, indicated by the shift in interactions from the wildtype. Although the gross structural features of WT and mutant models are similar, the overall packing of mutants appears to be different than WT. These data indicate that the strong coupling free energies observed experimental may be explained by subtle global structural changes in these mutated variants that lead to alterations in favorable packing interactions.

Modeling for lipid face crosslinking mutants is inconsistent with experimental results.

Models were also generated for the mutants Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala using CHI. Crosslinking data shows that these mutants are promiscuous; however, forced choice experiments suggest that there may be a preference for a particular interface. CHI generates models that have a face of interaction including both of these residues at the dimer interface, which is consistent with the initial crosslinking (Table 4.1). However, the details of the models do not agree with the crosslinking data. The models generated for

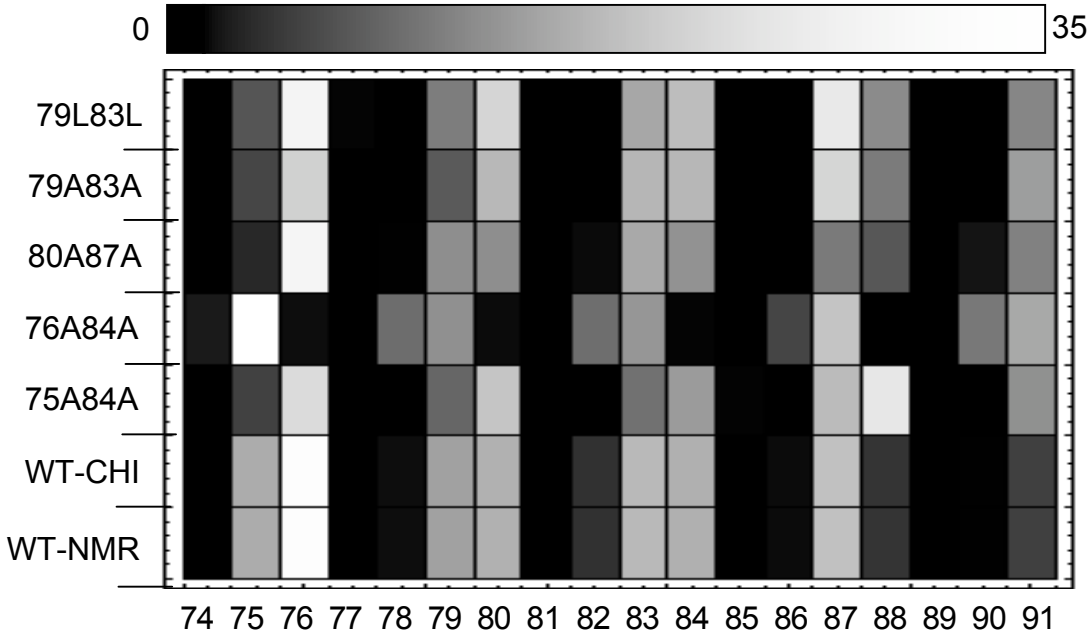

Figure 4.8 Density plot of inter-subunit occluded surface area per residue in WT and mutant models. The average occluded surface area per residue was calculated for each dimer model. The scale is indicated in the gradient box, with zero occluded surface in black and the maximum occluded surface in the model in white. In this case, the maximum occluded surface is 35 $Å^2$. The residue number is indicated on the x-axis, and the model for the corresponding mutant is indicated on the y-axis. For comparison, a WT model was calculated using CHI; it is shown to have the same degree and periodicity of occluded surface as the average for the NMR structures.

Leu75Ala-Gly83Ala and Ile76Ala-Gly83Ala both have a closer approach at residue 81 than 91, which is inconsistent with the observation that the 91Cys variants of these mutants are more reactive than the 81Cys variants. In addition, forced choice experiments demonstrate that Leu75Ala-Gly83Ala has more reactivity at 91 than 81, and that Ile76Ala-Gly83Ala has similar reactivity at 81 and 91. Therefore, it is not possible to confirm a representative model. Possible explanations for this result would be that these mutants exhibit a heterogeneous population of conformations at equilibrium, and a loss of specificity at the WT interface, which could account for the experimental crosslinking data.

4.5 Discussion

Energetic coupling may be explained by many subtle global structural rearrangements in some double mutants.

Using cysteine crosslinking and molecular modeling, the structural rearrangements that occur in glycophorin A mutants were investigated. In this study, we attempt to correlate the peptide sequence, energetics of association, and structure of this transmembrane dimer. Crosslinking data were used to probe structural changes. The resulting data show that most coupled double mutants retain use of the wildtype interface. Computational modeling of these mutants predicts subtle rearrangements in the dimer structure. These rearrangements accommodate many geometric changes, including: a global separation of helices in the dimer, a slight rotation of the helices about one another, and slight

changes in helix curvature. These models not only provide insight into the structural rearrangements in GpA mutants, but also allow for investigation into the structural basis for energetic coupling.

Our previous model for energetic coupling suggested that long-range coupling could result from changes in the crossing angle (Doura and Fleming 2004). The geometry of two crossing helices would create a greater structural and energetic connection between the distal ends than between closer sites. This hypothesis appears to be incorrect for the double mutants tested in this study. No models show a change in crossing angle outside the range of the 20 WT NMR models. Therefore, a change in helix-helix crossing angle is not likely to be the determinant of long range coupling. Studies indicate that bilayer thickness may be the predominant factor in determining helix tilt, and therefore may also be critical in determination of helix crossing angle (Killian 2003). Furthermore, molecular dynamics simulations of the GpA dimer in different lipid environments suggest that crossing angle is dependent on the lipid composition (Petrache et al. 2000). Therefore, the helix-helix crossing angle may be more dependent on the environment and less on the nature of the interactions between the helices. We hypothesize that all other possible geometric changes may be induced by amino acid substitutions. Therefore, long range coupling between sites in GpA cannot only be explained by rearrangements of a single geometric relation in the dimer.

Double mutants at the glycines residues are not fully explained by changes in packing interactions.

In the WT structure the glycines at 79 and 83 are extremely well packed (MacKenzie et al. 1997), hence, substitutions at these sites are extremely destabilizing (Lemmon et al. 1992b; MacKenzie et al. 1997; Doura et al. 2004). These sites compose the GxxxG motif, which is thought to induce strong helix-helix interactions (Senes et al. 2000). The strength of a GxxxG interaction may be due to a proposed C_α hydrogen bond that forms between the backbone of the two helices (Senes et al. 2001). However, substitutions in the GxxxG motif to large aliphatic residues still retain the ability to dimerize (Doura et al. 2004). In this study, crosslinking data show that proteins containing substitutions to alanines and leucines in the GxxxG motif to still retain the wildtype interface. Therefore, even when the GxxxG motif is disrupted, the interactions at these sites may still be important in establishing the dimer interface. These data confirm the result that a GxxxG motif is not necessary for dimerization (Doura and Fleming 2004), and suggests that the GxxxG motif is also not necessary to specify the WT interface.

While the crosslinking results for Gly79Ala-Gly83Ala and Gly79Leu-Gly83Leu suggest that these mutants use the canonical GpA interface, the modeling protocol did not generate structural models consistent with the structure-based parameterization. Models were generated that are structurally similar to wildtype, but have a calculated free energy much greater than the

experimentally determined free energy of association. A possible explanation for this lack of correlation is that the structure-based calculations are based primarily on changes in packing interactions at the dimer interface. The stability of variants that contain double substitutions in the GxxxG motif could arise from a number of effects. Using a correlation that includes only changes in packing interactions would result in an overestimate of the calculated free energy. The structure-based calculation does not include electrostatic interactions and only considers changes in interactions within the dimer. If the substitution to a larger aliphatic residue created a more favorable interaction with the lipid solvent at that interface, the experimental free energy would be less than expected based on only on changes in packing. Although these substitutions are unlikely to cause a structural change in the monomeric state, changes in the interactions of the individual monomers with lipid also may contribute to the energetics of association. Further data would be necessary to distinguish the possible energetic effects of substitution, and confirm the dimer structure in double mutants with substitutions in the GxxxG motif.

Packing defects can cause sampling of alternate interfaces

Crosslinking data suggest that global rearrangements in the dimer interface occur when the N-terminal sites Leu75 and Ile76 are mutated to alanine. Notably, Leu75 and Ile76 occlude the most surface area on the opposing helix in the WT structure. The substantial loss of packing interactions in these mutated variants may promote a global rotation of the helices. Interestingly, most

strongly coupled mutants include a substitution at either Leu75 or Ile76 at the N-terminal end of the helix, and a second mutation at either Val84 or Gly83. The coupling of these sites appears to be complex and due to different energetic and structural interactions. In particular, the global rotation induced by Leu75Ala and Ile76Ala is retained in the double mutants with Gly83, but not in the double mutants with Val84. Therefore, a second mutation at Val84 causes a rearrangement to use the wildtype interface. In fact, Leu75Ala-Val84Ala demonstrates a free energy of association very similar to the wildtype, and the model shows a high level of similarity to the wildtype structure and similar pattern of packing interactions. Therefore, the predominance of strong coupling with Leu75 and Ile76 with other residues may be a result of the influence of these residues on helix rotation.

Structural information obtained by site-engineered disulfide crosslinking provides insight into the structure-based parameterization on single and double mutants (Doura and Fleming 2004; Doura et al. 2004). In the single mutant study, structural parameters calculated from models of Ile76 variants did not correlate well to experimentally measured free energy (Doura et al. 2004). Calculated free energies for models of alanine and valine substitutions at Ile76 correlated to the experimental values, whereas models containing leucine and glycine substitutions at this site were poorly described by the structure-based calculation. We hypothesize that the difficulty in characterizing substitutions at this site may be due to global structural changes, in particular, an induced rotation in the helix-helix interface. Although most single and double mutants are well characterized

by the structure-based parameterization, the limited structural data obtained in this study indicate that some of these mutants undergo larger structural rearrangements than predicted. In particular, it has become apparent that models for variants with substitutions at Ile76 and Leu75 are likely to be inaccurate. The computational modeling for mutants that demonstrate this alteration of the dimer interface did not correlate well to experimental data. This does not exclude the possibility that a unique structure could explain our experimental results. However, it also does not prohibit the possibility that the decreased specificity in of the dimer interface allows these variants to sample multiple conformations.

4.6 Conclusions

The dimer interface of glycophorin A has been well characterized using mutagenesis and structural techniques. Until this study, there was no experimental data on the structural rearrangements induced by interfacial substitutions. The current data suggest that the N-terminal leucine and isoleucine residues are critical to specify the dimer interface, although these residues are not the most critical for dimer stability. Mutations at the GxxxG motif, which is critical for strong dimerization, seem to retain the WT canonical interface. These data highlight that the interactions necessary to specify the native fold may not be the same as those that stabilize it. These results suggest that a complex interplay between residues of the interface motif is critical to stabilize and specify the glycophorin A transmembrane dimer.

CHAPTER 5

Alternate interfaces may mediate homomeric and heteromeric assembly in the transmembrane domains of SNARE proteins

5.1 Summary

Thus far, this thesis has focused on the role of amino acid sequence in stabilizing and specifying the model membrane protein Glycophorin A. GpA is an extremely stable dimer, which is probably constitutively dimeric *in vivo*. In biological systems, intramolecular interactions are not only essential to specify stable structures, but also transient interactions that are necessary for dynamic biological processes. These transient interactions must also be encoded in the amino acid sequence. Vesicle fusion is a dynamic biological process that requires membrane protein participation. The fusion of a vesicle to a target membrane is mediated by temporally and spatially regulated interactions within a set of evolutionarily conserved proteins. Integral to proper fusion is the interaction between proteins originating on both vesicle and target membranes to form a protein bridge between the two membranes, known as the SNARE complex. This protein complex includes two single pass transmembrane helix proteins: syntaxin and synaptobrevin. Experimental data and amino acid sequence analysis suggest that an interface of interaction is conserved between the transmembrane regions of the two proteins. However, conflicting reports have been presented in

the literature on the role of the synaptobrevin transmembrane domain in mediating important protein-protein interactions.

To address this question, a thermodynamic study was carried out to quantitatively determine the self-association propensities of the transmembrane domains of synaptobrevin and syntaxin. Our results show that the transmembrane domain of synaptobrevin has only a modest ability to self-associate, whereas the transmembrane domain of syntaxin is able to form stable homodimers. Nevertheless, by a single amino acid substitution, synaptobrevin can be driven to dimerize with the same affinity as syntaxin. Furthermore, crosslinking studies show that dimerization of synaptobrevin is promoted by oxidizing agents. Despite the presence of a conserved cysteine in the same location as in synaptobrevin, syntaxin dimerization is not promoted by oxidization. This analysis suggests that subtle yet distinct differences are present between the two transmembrane dimer interfaces. A syntaxin/synaptobrevin heterodimer is able to form under oxidizing conditions, and we propose that the interface of interaction for the heterodimer may resemble the homodimer interface formed by the synaptobrevin transmembrane domain. Computational analysis of the transmembrane sequences of syntaxin and synaptobrevin reveal structural models that correlate with the experimental data. These data may provide insight into the role of transmembrane segments in the mechanism of vesicle fusion.

5.2 Introduction

The fusion of a vesicle to a target membrane is essential to many biological functions in eukaryotes. The mechanism of fusion has been dissected into distinct and essential stages. One essential stage is the bridging of the target and vesicle membranes to form a single membrane, which requires a set of proteins termed the SNARE (soluble NSF attachment protein receptors) complex (Szule and Coorssen 2003). The SNARE complex includes SNAP25 and two single pass transmembrane proteins: syntaxin, which is located on the target membrane (t-SNARE), and synaptobrevin, which is located on the vesicle membrane (v-SNARE) (Ungar and Hughson 2003). The crystal structure and electron microscopy data demonstrate that the SNARE complex is a parallel helical bundle with the transmembrane segments emerging as tails inserted into the two merging membranes (Sutton et al. 1998). Evidence suggests that the SNARE complex creates a protein bridge that draws the two membranes together and overcomes repulsive forces to allow membrane mixing (Hanson et al. 1997). A *cis* to *trans* transition occurs in the complex when the membranes merge, and the transmembrane segments emerge inserted into the same membrane (Jahn et al. 2003). The soluble domain complex has enormous stability and the formation of this complex could overcome the energetic barrier to fusion (Fasshauer et al. 2002). However, the stability of the interactions of the transmembrane domains has not been quantified, and these segments may provide additional energy and mediate interactions required for the mechanism of intracellular vesicle fusion.

These transmembrane segments may act as more than just membrane anchors. They could be integral to membrane fusion through a direct role in bilayer mixing and by promotion of protein complex formation and oligomerization (Szule and Coorssen 2003). Synthetic peptides corresponding to the transmembrane domains of syntaxin and synaptobrevin have been shown to drive fusion *in vitro*, in the absence of their respective cytoplasmic domains (Langosch et al. 2001). This fusiogenic activity may involve sequence specific interactions between the SNARE protein transmembrane domains. Biochemical studies suggest that the TM segments of syntaxin and synaptobrevin promote the formation of the SNARE complex (Poirier et al. 1998) and that sequence specific interactions occur between these transmembrane domains (Margittai et al. 1999). In addition, mutations in the transmembrane domains of syntaxin and synaptobrevin have been shown to lead to impaired neurotransmission in *Caenorhabditis elegans* (Nonet et al. 1998; Saifee et al. 1998). Further study into the nature of interactions between transmembrane segments will be necessary to better understand the mechanism of vesicle fusion.

Previous biochemical studies have provided insight into the specificity of interactions for both homo- and hetero-oligomerization of the syntaxin and synaptobrevin TMs. SDS-PAGE analysis of full length synaptobrevin mutants led to a possible motif of interaction for homodimerization (Laage and Langosch 1997). Interestingly, this motif is highly conserved in the syntaxin TM, except for a single amino acid difference at the N-terminus (Laage et al. 2000). This sequence similarity suggested an analogous region of interaction for the

syntaxin/synaptobrevin heterodimer and the syntaxin homodimer. This hypothesis is supported by evidence of heterodimerization of syntaxin and synaptobrevin using SDS-PAGE (Laage et al. 2000). Homodimerization of syntaxin and synaptobrevin have also been measured using the ToxR assay *in vivo* (Laage et al. 2000). The addition of the synaptobrevin and the proposed syntaxin interaction motif to a polyalanine helix supports dimerization in the ToxR assay, which lead to the conclusion that this interaction sequence promotes homodimerization in both proteins.

The ToxR and SDS-PAGE data provided strong evidence for interactions within the transmembrane regions of syntaxin and synaptobrevin. However, these studies have been subjects of controversy in the recent literature. Bowen and coworkers observed association of synaptobrevin using an assay similar to ToxR, the TOXCAT assay. These data showed that the dimerization signal given by the synaptobrevin TM is not significantly greater than their negative control, the glycophorin A mutant GpA-G83I (Bowen et al. 2002). In further contradiction to the previous studies, the authors also found that dimerization of synaptobrevin observed using SDS-PAGE was insignificant and dependent on the purification protocol. Furthermore, strong dimerization was only found with the substitution of a transmembrane residue to an asparagine. The presence of a polar residue is unfavorable in the apolar membrane environment and has been shown to drive association of transmembrane helices (Choma et al. 2000; Dawson et al. 2003). In response, the authors of the SDS-PAGE and ToxR studies demonstrated again that the ToxR signal for synaptobrevin was substantially greater than their

negative control, GpA-G83A, and that dimerization can be observed by SDS-PAGE in the presence of the crosslinking agents (Roy et al. 2004). These authors also demonstrated that the construct used by Bowen and coworkers elicited lower response in ToxR than the construct that was used in the initial study. Although the significance of synaptobrevin dimerization is still unclear, it appears that dimerization propensity depends on the environment, experimental conditions, and exact sequence of the experimental constructs.

In this study we address the role of the transmembrane regions of syntaxin and synaptobrevin and their propensities for homo- and hetero-dimerization using quantitative techniques to analyze oligomerization *in vitro*. Sedimentation equilibrium analytical ultracentrifugation (AUC) can be used to determine the free energy of association for a transmembrane protein in a detergent environment (Fleming et al. 1997). Using this technique, the fundamental measurement is the protein mass. The following properties can be definitively determined: the stoichiometries of any protein complexes; the association propensity; and the detergent concentration dependence on association. If the proteins fail to associate, sedimentation equilibrium will demonstrate the monomer molecular weights, as opposed to a qualitative decrease in dimerization signal as in the TOXCAT and ToxR assays. Sedimentation equilibrium has been used extensively to measure the free energy of association in a detergent environment for glycophorin A and mutants (Fleming et al. 1997; Fleming and Engelman 2001a; Doura et al. 2004; Doura and Fleming 2004). Since this method has been useful in understanding the

sequence context of interactions in glycophorin A, it may bring clarity to the controversy surrounding the association propensities of the syntaxin and synaptobrevin transmembrane sequences.

In addition, the presence of conserved cysteines in the transmembrane domains permit the use of oxidizing conditions to observe weak oligomerization by SDS-PAGE. These crosslinking studies provide insight into the permissible interface of interaction for syntaxin and synaptobrevin homodimers and the syntaxin/synaptobrevin heterodimer. Including these experimental constraints in computational modeling, it is possible to better understand the nature of the interaction in homodimers and heterodimers of syntaxin and synaptobrevin transmembrane domains. Our findings strongly suggest that sequence specific interactions do occur in the transmembrane segments of syntaxin and synaptobrevin, albeit weakly when compared to the stability of the GpA TM. Although the wildtype synaptobrevin transmembrane sequence may have little ability to drive homodimerization, the sequence does encode a sterically permissible interface for association.

5.3 Materials and Methods

Cloning, Expression, and Protein Purification

All proteins used in this study were expressed as fusion constructs with *Staphylococcal* nuclease (Lemmon et al. 1992). The syntaxin1A construct, SN-Syx, was cloned by using the pET11A-SNGpA99 construct as a template for successive rounds of site-directed mutagenesis using the Quikchange kit

(Stratagene, LaJolla, CA) with the appropriate primers. The synaptobrevin2 or VAMP-2 construct, SN-Syb, was cloned by ligation of extended oligonucleotides containing the synaptobrevin transmembrane sequence into the pET11A-SN vector. Substitutions were made using the Quikchange kit with the wildtype constructs. Proteins were purified using extractions in the detergent thesit (Fluka, Switzerland) with a single salt extraction in 1M NH_4OAc (Sulistijo et al. 2003) followed by further purification on a SP column by FPLC.

Sedimentation Equilibrium Analytical Ultracentrifugation

Proteins were exchanged into the detergent C_8E_5 for AUC (Fleming et al. 1997). Sedimentation equilibrium experiments were carried out at 25ºC in a Beckman XL-A analytical ultracentrifuge. Protein distributions were monitored using the absorbance at 280nm. For an accurate and precise measurement of the free energy of association, a minimum of twelve data sets were used in a global fitting of the data using MacNonlin (Johnson et al. 1981). The data used in analysis consisted of three significantly different initial protein concentrations run at four significantly different speeds. The buoyant molecular mass (σ) was calculated from the amino acid composition using SEDNTERP (Laue 1992) and held constant during global fitting. Each free energy was independently measured a minimum of three times with at least two different detergent concentrations. The standard state free energy value (ΔG_x^o) is calculated from the experimentally measured apparent equilibrium constant (K_{app}) and by assuming an ideal dilute solution, using the following equation (Fleming 2002).

$$\Delta G_x^o = -RT\ln(K_{app}[micellarDet]_w) \qquad (5.1)$$

Disulfide Crosslinking under Oxidative Conditions

Crosslinking experiments were carried out using copper-phenanthroline, a final concentration of 1mM $Cu^{II}SO_4$ and 2mM 1,10 phenanthroline, or iodine at a final concentration of 0.1mM (data not shown) with 20mM Tris pH 8.0, 250mM NaCl and 1% Thesit. Both oxidizing agents produced comparable results under similar experimental conditions. Reactions were quenched by the addition of 40mM NEM, 20mM EDTA, and SDS-PAGE gel loading buffer. Samples were immediately boiled for 4 minutes at 95ºC and observed by SDS-PAGE. Although the length of the amino acid sequences of the SN-Syx and SN-Syb clones are identical, the SN-Syb proteins migrates more quickly than the SN-Syx proteins, allowing visualization of the heterodimer by SDS-PAGE.

Computational Modeling

Structural models were generated for the syntaxin homodimer and the syntaxin/synaptobrevin heterodimer for the WT sequences (Table 5.1; syntaxin residues 267-292; synaptobrevin residues 98-113) using the program CHI (Adams et al. 1995; Adams et al. 1996) with the CHARMM parameter set (Brooks et al. 1983). CHI uses simulated annealing and energy minimization to find clusters of low energy structures. The average structure that represents each cluster is then calculated. A full search for all possible homodimers was performed a minimum of five times to fully explore all possible models (Fleming and Engelman 2001a). A representative structure for the syntaxin homodimer and the syntaxin/synaptobrevin heterodimer was determined by consideration of

symmetrical structures that correlate to the experimental data. Structures were analyzed using the occluded surface algorithm version 7.2.2 (Pattabiraman et al. 1995) to quantify packing interactions, and WHATIF (Vriend 1990) to calculate helix crossing angles. Figures were constructed using MacPymol (DeLano Scientific LLC).

5.4 Results

Synaptobrevin and syntaxin transmembrane domains have surprisingly different propensities for association

Previous studies have shown that the transmembrane regions of syntaxin and synaptobrevin may be involved in homodimerization (Laage et al. 2000). To address the ability of these regions to drive dimerization, appropriate fusion constructs containing the transmembrane domain of syntaxin and synaptobrevin were cloned. The sequences are shown in Table 5.1, in which the previously determined conserved motif of dimerization is highlighted. There is a single residue difference in the dimerization face in the two proteins: Leu99 in synaptobrevin, which is aligned to the Met267 in syntaxin. Substitutions were made at these positions to create the syntaxin interface in the synaptobrevin background, called SN-SybLM, and to create the synaptobrevin interface in the

Construct	Sequence	$\Delta G_x°$(kcal/mol)
SN-Syb	EPE-^{98}ILGVICAIILIIIIVYFSTZ	Monomeric
SN-SybLM	EPE-^{98}IMGVICAIILIIIIVYFSTZ	-3.3 ± 0.2
SN-Syx	EPE-^{266}IMIIICCVILGIIIASLLIZ	-3.5 ± 0.3
SN-SyxML	EPE-^{266}ILIIICCVILGIIIASLLIZ	-3.2 ± 0.1

Table 5.1 Free energy of association for syntaxin and synaptobrevin constructs. The fusion proteins were cloned with *Staphylococcal* nuclease N-terminal to the indicated transmembrane sequence. The previously determined dimerization interface for synaptobrevin is highlighted (Laage and Langosch 1997). This sequence is highly conserved in the syntaxin transmembrane region. The experimental free energies were measured by sedimentation equilibrium analytical ultracentrifugation. There was no detectable dimer for the SN-Syb construct, which is listed as monomeric. The error is the standard deviation of at least three independent experiments.

syntaxin background, called SN-SyxML. These constructs were analyzed by sedimentation equilibrium analytical ultracentrifugation, using well developed protocols for SN-GpA (Fleming et al. 1997; Fleming and Engelman 2001b).

Figure 5.1 presents representative data for AUC experiments for the four constructs used in this study. The wildtype syntaxin construct, SN-Syx (Figure 5.1a), is best fit to a monomer-dimer equilibrium, with a standard state mole fraction free energy of association equal to 3.5 ± 0.3 kcal mol^{-1}. In contrast, no dimer is observed for the wildtype synaptobrevin construct, SN-Syb at the highest experimentally accessible protein/detergent mole fraction (Figure 5.1c). Alteration of the single residue difference in the conserved interface, the Met267Leu substitution in syntaxin (SN-SyxML), creates no disruption in the free energy of association (Figure 5.1b, Table 5.1). However, when the corresponding mutation is made in the synaptobrevin dimerization motif, the SN-SybLM construct, dimer is observed at a comparable concentration to both SN-Syx and SN-SyxML (Figure 5.1d, Table 5.1). We therefore find that the substitution to methionine in the synaptobrevin transmembrane domain provides at least an additional 1.3 kcal mol^{-1} in the free energy of association.

It is important to consider the concentration dependence of association, since the conditions in the cell are likely to be vastly different than experimental conditions. Therefore, the populations of oligomers over a range of concentrations were calculated using the experimentally determined thermodynamic values. Figure 5.2 demonstrates the dimer population as a function of the mole fraction protein for all constructs. For reference, these are

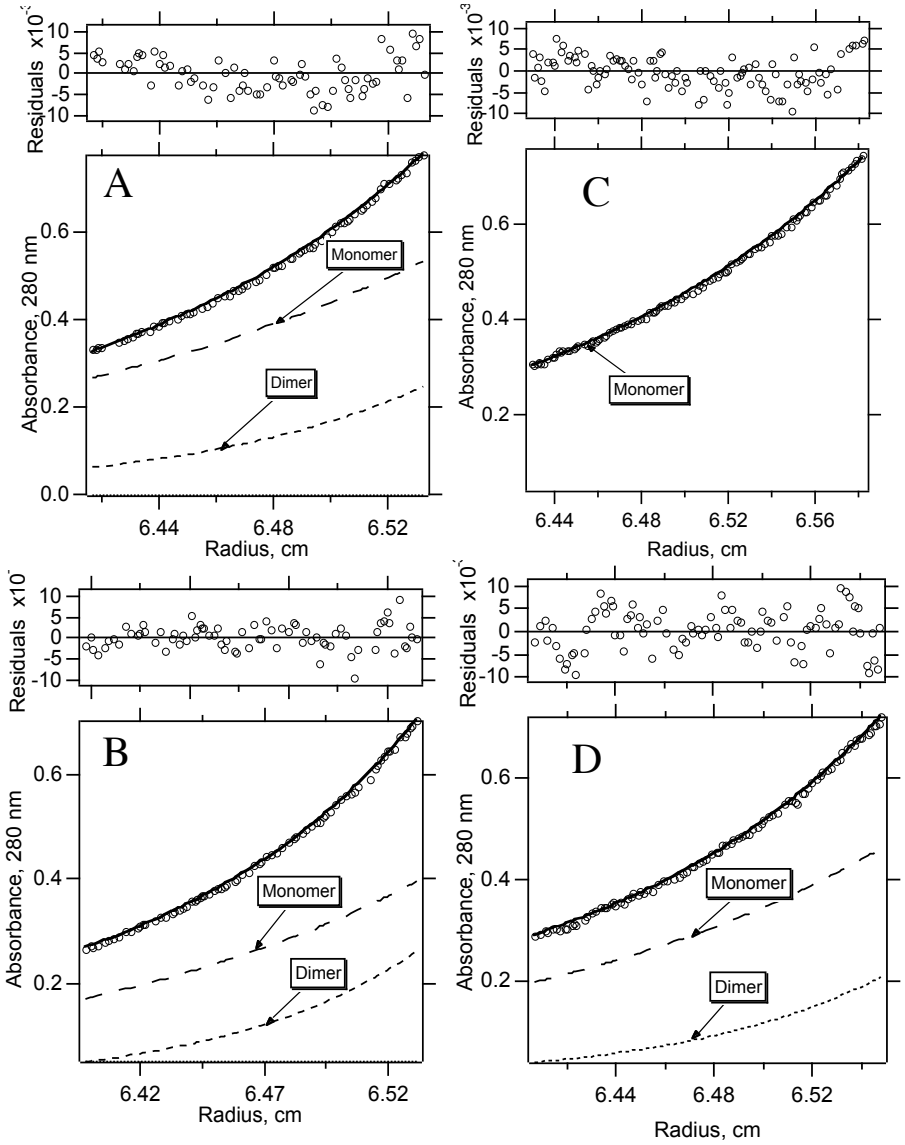

Figure 5.1 Sedimentation equilibrium data for synaptobrevin and syntaxin constructs. The raw data for a representative data set is shown as open circles. The solid line is the fit for which the above residuals appear small and random. The concentration of monomer (dashed line) and dimer (double-dashed line) are calculated based on the globally determined equilibrium constant. (A) SN-Syx and (B) SN-SyxML and (D) SN-SybLM best fit to a monomer-dimer equilibrium. (C) SN-Syb is best fit to a single monomeric species.

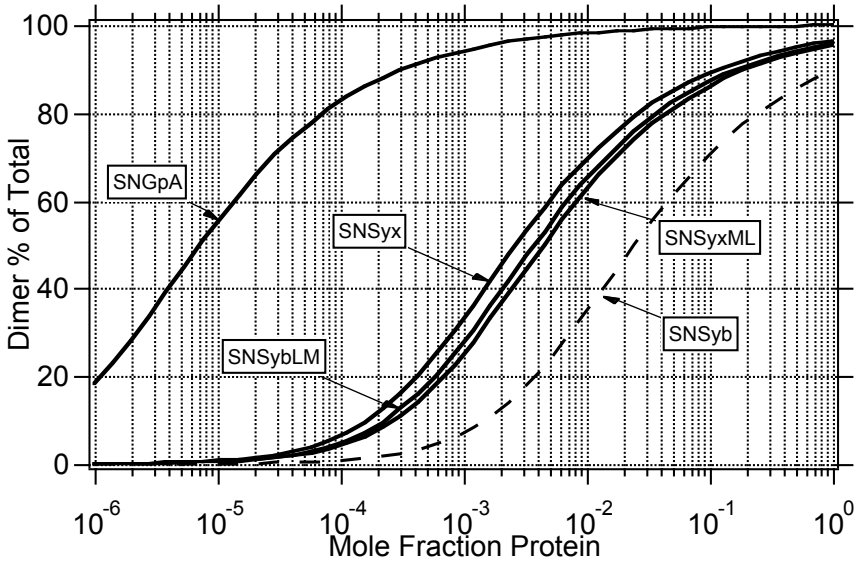

Figure 5.2 Relative dimer population of syntaxin and synaptobrevin constructs. The fraction dimer is plotted versus the mole fraction protein. The population distribution is calculated based on the free energy of association measured by analytical ultracentrifugation. For reference, the calculated distribution for the strong dimer glycophorin A is shown. The broken line shows the maximum possible dimer distribution for the SN-Syb construct, which approximates the random distribution for a dimer in a micellar environment (Kobus and Fleming 2005).

compared to distribution for SN-GpA analyzed in the same micellar environment. SN-Syx, SN-SyxML, and the SN-SybLM are significantly less stable than Glycophorin A. The free energies of association for these SNARE constructs can be considered identical since they are within error of each other (Table 5.1). Since there was no dimer observed experimentally, the preferential interaction in the synaptobrevin transmembrane domain is likely to be too weak to measure under equilibrium conditions using the described method. Measuring the association in detergent micelles is limited, because at high protein-detergent ratios random association becomes the driving force for oligomerization (Kobus and Fleming 2005). The distribution for SN-Syb, shown in Figure 5.2, was calculated based on the maximum possible experimental free energy value, which approximates the random distribution and is shown as a dotted line in Figure 5.2 (Kobus and Fleming 2005). The apparent concentration of these transmembrane proteins during the process of fusion could be much greater than what is considered experimentally, since the interactions between the soluble domains will drive them into close proximity. Therefore, presumably weak interactions could still be influential in the mechanism of fusion in vivo.

Crosslinking reveals differences in the reactivity of cysteine residues in synaptobrevin and syntaxin homodimers.

It has been shown that synaptobrevin will form dimer in the presence of oxidizing agents (Laage and Langosch 1997; Roy et al. 2004). Under these conditions, a single cysteine residue in the transmembrane region reacts to form a disulfide bonded dimer, which can be visualized using SDS-PAGE gel

electrophoresis. Syntaxin has two cysteines in the transmembrane region that may also be important for homo- and hetero-dimerization and that may indicate the interface of interaction (Table 5.1). Furthermore, because cysteine 271 is a conserved residue in the synaptobrevin interface, it may be involved in heterodimerization of syntaxin and synaptobrevin. To test the position of these cysteines in interface of the transmembrane dimers, and to determine if that role is altered in the SN-SyxML and SN-SybLM mutants, crosslinking experiments were carried out. If the same interface were involved in homodimerization for both syntaxin and synaptobrevin, all constructs would be expected to dimerize in the presence of oxidizing agents. Furthermore, since the syntaxin constructs and the SN-SybLM construct demonstrate a similar free energy of association, a similar motif for interaction may exist in these proteins, distinct from the wildtype synaptobrevin construct.

Using copper phenanthroline (CuOP) (Figure 5.3) and iodine (data not shown), mutant and WT fusion proteins were subjected to oxidizing conditions that promote the formation of favorable disulfide bonds. Both oxidizing agents promoted dimer formation for SN-Syb and SN-SybLM proteins. Notably, SN-Syb does not populate dimer at time zero, but dimer formation appears after exposure to CuOP. SN-SybLM populates dimer at time zero and becomes predominately dimeric after exposure to CuOP. Additionally, the rate of crosslinking for SN-SybLM appears more rapid than the rate for SN-Syb. This may be explained in part due to the increased propensity for dimerization in the SN-SybLM protein, but may also be due to a better accommodation for disulfide bond formation in

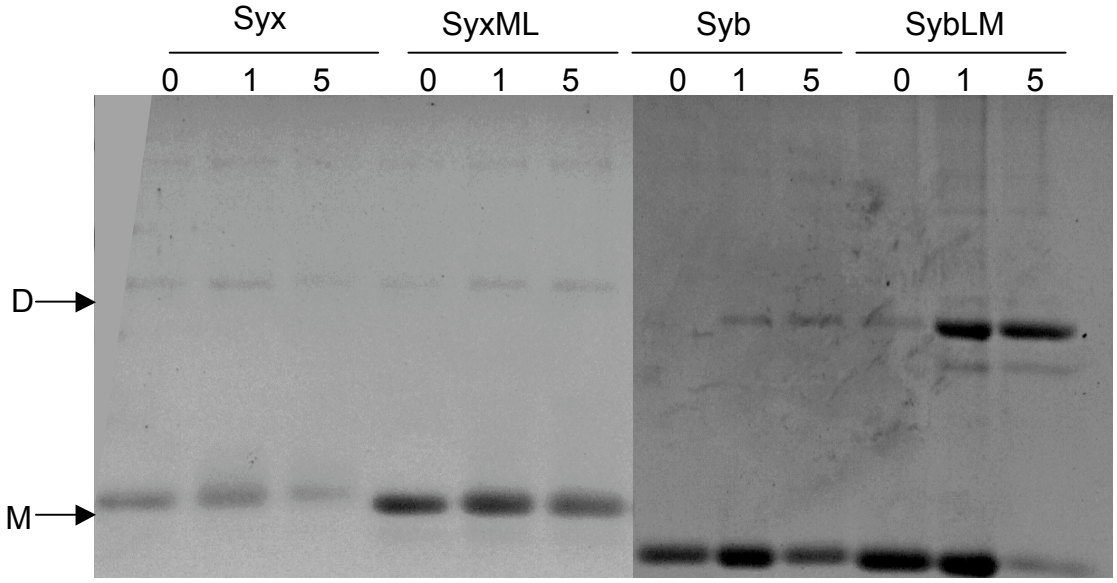

Figure 5.3 Homodimer crosslinking for syntaxin and synaptobrevin constructs. Disulfide crosslinking was carried out on purified protein using copper phenanthroline as an oxidizing agent. Both mutant and wildtype syntaxin and synaptobrevin constructs were exposed to oxidizing agent and samples were taken at 1 minute and 5 minutes and observed by SDS-PAGE. Both SN-Syx and SN-SyxML did not crosslink under these conditions. SN-Syb shows no dimer at time zero and an increase in dimer concentration after exposure to CuOP. SN-SybLM shows dimer at time zero and a larger increase in dimer concentration after exposure to oxidizing agent. An increase in dimer concentration is indicative of the formation of disulfide bonds across a dimer interface. D-Dimer and M-Monomer.

the structure of the SN-SybLM dimer. This data is consistent with previous studies, and demonstrates that Cys103 is at the dimer interface and forms a disulfide bond across the interface in synaptobrevin.

In this study, we also test the ability of cysteine residues in the syntaxin transmembrane domain to promote dimerization. Under conditions that promote dimerization for synaptobrevin, neither SN-Syx nor SN-SyxML formed crosslinked dimer (Figure 5.3). Since syntaxin has a greater affinity than WT synaptobrevin, there is a visible dimer band at time zero for WT and mutant SN-Syx proteins. This dimer band is due to the inherent affinity of the transmembrane sequences, and does not represent a crosslinked dimer. After exposure to oxidizing agent there is no increase in dimer concentration for SN-Syx or SN-SyxML, which is consistent with an absence of crosslinking in these dimers. The oxidizing agent, CuOP, is a promiscuous catalyst of disulfide bond formation, even for cysteines that are only proximal due to molecular motion (Careaga and Falke 1992; Hughson et al. 1997). The lack of reactivity of the cysteines in the syntaxin dimer indicates that these cysteines are in a stable conformation that does not permit disulfide bond formation across the interface. Furthermore, at high protein concentrations, SDS-PAGE reveals higher order oligomers for both SN-Syx and SN-SyxML, which are also not affected by oxidizing agents (data not shown). These higher order oligomers are not unique to our experimental conditions, but have also appeared on immunoblots of full length syntaxin protein in rat brain membranes (Yoshida et al. 1992). In contrast, observations show that the SN-SybLM mutant, which has the putative syntaxin

dimer interface, does not populate higher order oligomers at similar concentrations. These data indicate that although SN-SybLM demonstrates a free energy of association equivalent to the syntaxin constructs, the interface for interaction in this construct is distinct. Therefore, inherent differences exist in the homodimerization interfaces for the syntaxin and synaptobrevin constructs.

Crosslinking promotes heterodimerization of syntaxin and synaptobrevin

Previous data shows that the transmembrane regions of syntaxin and synaptobrevin can form heterodimers under oxidizing conditions (Laage et al. 2000). Using CuOP as an oxidizing agent, it is possible to crosslink all combinations of syntaxin and synaptobrevin constructs (Figure 5.4). Inherent differences in the migration of the constructs allow the visualization of heterodimer formation by SDS-PAGE. The three upper arrows indicate the migration of the dimer species on the gel, with the middle arrow showing the heterodimer. The lower arrows indicate the SN-Syx and SN-Syb monomers. SN-SyxML and SN-Syx do not crosslink to each other, but both variants do crosslink with SN-Syb and SN-SybLM to form disulfide linked heterodimers. It appears that neither mutation in the SN-Syb or SN-Syx transmembrane region affects the rate of heterodimer formation under oxidizing conditions. Although the conserved cysteine residue cannot promote homodimerization in syntaxin, these residues promote heterodimerization of syntaxin and synaptobrevin. In contrast to the syntaxin homodimer, the cysteine residues in the heterodimer are reactive and can form a disulfide bond across the dimer interface. This is similar to the

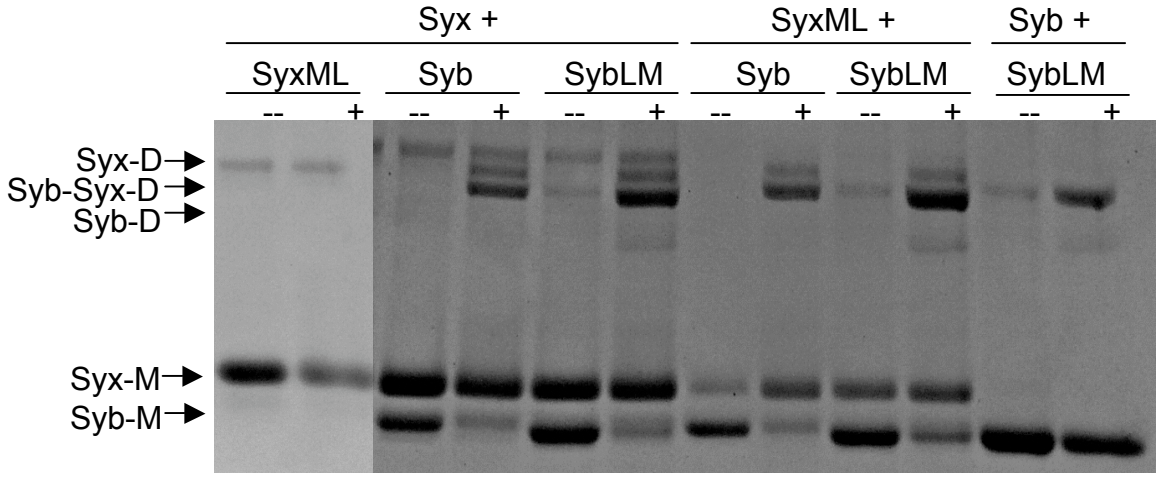

Figure 5.4 Heterodimer crosslinking for syntaxin and synaptobrevin constructs. The inherently different migration properties of the SN-Syb and SN-Syx fusion proteins on SDS-PAGE allow the visualization of a SN-Syx/SN-Syb heterodimer. Proteins were mixed and exposed to copper-phenanthroline for 5 minutes. The effect of mutation on crosslinking is also tested by inclusion of the SN-SyxML and SN-SybLM constructs. Heterodimer formation occurs between SN-Syx and SN-Syb proteins and does not seem affected by the mutations at the interface. D-Dimer and M-Monomer.

result for the synaptobrevin homodimer, which permits a disulfide bonded dimer. Therefore, the crosslinking data suggest that the heterodimer interface may be similar to the synaptobrevin crosslinked homodimer interface, but distinct from the syntaxin homodimer interface.

Alternate interfaces for syntaxin homodimer and syntaxin-synaptobrevin heterodimer can be calculated using computational modeling

Using the previous mutagenesis data, a model for the structure of the synaptobrevin homodimer (Figure 5.5a) was calculated using CHI program (Fleming and Engelman 2001a). which has been successful for prediction of other transmembrane homodimer structures (Adams et al. 1996; MacKenzie et al. 1997). This model is consistent with our experimental data, predicting that Leu99 is located at the interface in close contact with the opposing helix. Furthermore, the model shows that Cys103, the residue necessary for crosslinking, is also located at the interface and could be predicted to form an intermonomer disulfide bond under oxidizing conditions. Using the same modeling protocol, the syntaxin homodimer and syntaxin/synaptobrevin heterodimer structures were calculated and the experimental data was used to distinguish between the models. Although there are conserved residues in syntaxin that can mediate dimerization in the synaptobrevin transmembrane domain, thermodynamic and crosslinking data suggest that these residues are not necessarily important in homodimerization of the syntaxin transmembrane domain. Crosslinking data show that Cys271 and Cys272 do not form a disulfide

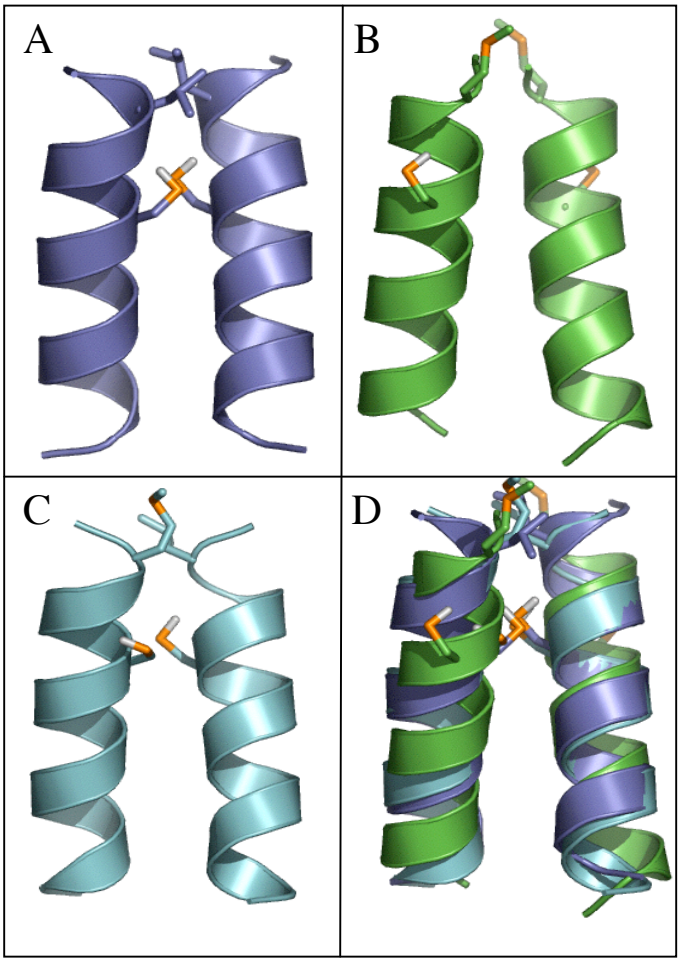

Figure 5.5 Computational modeling for syntaxin and syntaxin-synaptobrevin heterodimer. Models were generated using the program CHI. Each model is represented as a cartoon helix with the important experimental residues shown as sticks. (A) Computational model for synaptobrevin transmembrane domain homodimer (Fleming and Engelman 2001a). Shown in sticks are Leu99 and Cys103 (B) Computational model for syntaxin transmembrane homodimer. Shown in sticks are Met267 and Cys271 (C) Computational model for the syntaxin-synaptobrevin heterodimer, syntaxin is the helix on the left and synaptobrevin is the helix on the right. (D) Alignment of syntaxin, synaptobrevin, and heterodimer models.

linkage between two syntaxin monomers. The lack of reactivity of these cysteine residues in the homodimer under conditions that promote disulfide linkages in the heterodimer and synaptobrevin homodimer can be due to multiple structural features. The cysteine residues may be at the interface, but too distant to form a disulfide linkage. The cysteine residues may be at the interface, but in an inadequate rotamer orientation to promote crosslinking. Finally, the cysteine residues could be located on the lipid face and therefore not reactive at the interface. The final explanation seems to be the most likely. CuOP is a promiscuous catalyst and would be able to promote disulfide linkages in the dimer between cysteines that are distant and in a rotamer conformation that is transient. Additionally, mutagenesis data for syntaxin suggest that Met267 is not a critical interface residue. Therefore a likely interface for the syntaxin dimer would not include Met267, Cys271 or Cys272. The interface for interaction for the synaptobrevin/syntaxin heterodimer should include either Cys271 or Cys272 of syntaxin and Cys103 of synaptobrevin, since crosslinking data shows that these residues promote a disulfide bonded heterodimer, and therefore that the syntaxin-synaptobrevin heterodimer may use a similar interface to the synaptobrevin homodimer.

Models were generated for the syntaxin homodimer and the syntaxin/synaptobrevin heterodimer and subjected to the structural interpretation of the experimental data. Using CHI, an exhaustive search is performed to find low energy structures. The structures obtained in this search include both symmetrical and non- symmetrical dimers. Only symmetrical dimers with the

interface implicated in the experiments are chosen. In this study, there was a single model that fit these criteria for the syntaxin homodimer and synaptobrevin/syntaxin heterodimer (Figure 5.5b,c). The important experimental residues are shown as sticks in Figure 5.5. For syntaxin those residues are Met267 and Cys271; and for synaptobrevin those residues are Leu99 and Cys103. In accordance with the experimental results, the crosslinking residues Cys271 point away from each other and would be unable to form a disulfide bond between the monomers. The symmetrical model for the syntaxin dimer that demonstrated an inability for interaction at the cysteine residues also demonstrated that the Met267 is not at the dimer interface, and in the model these residues are directed away from each other and are perpendicular to the interface. This is consistent with the experimental result that the Met267Leu substitution has no energetic consequence. As discussed, the heterodimer interface is expected to be distinct from the syntaxin homodimer interface. Therefore, in the heterodimer model, the synaptobrevin dimerization interface is used and a disulfide bond could form between Cys271 on the syntaxin monomer and Cys103 on the synaptobrevin monomer. The conformations of the helix backbone in these models are very similar to each other, which can be seen when all three models are aligned (Figure 5.5d). It is important to note that these are not experimentally determined structures, but are chemically reasonable models that explain the experimental data.

A more detailed comparison of the models can be performed by examination of van der Waals packing interactions using the occluded surface

algorithm (Pattabiraman et al. 1995). Buried surface area and occluded surface have been shown to correlate to energetics in membrane proteins using structure-based parameterization (Doura et al. 2004; Faham et al. 2004). The occluded surfaces for the synaptobrevin and syntaxin models are comparable, each homodimer occluding approximately 330 $Å^2$. The heterodimer model occludes approximately 350 $Å^2$. Our results demonstrate that at equilibrium, syntaxin populates dimer more readily than synaptobrevin. This result cannot be explained using the theory that buried surface area is the primary determinant for the propensity of helices to dimerize. Nevertheless, the computational structures provide a structural framework compatible with the existing experimental data. Therefore, these models may provide insight into the possible interface of interaction for these transmembrane proteins.

5.5 Discussion

Synaptobrevin dimerization is not driven by the transmembrane domain packing interactions

Although the significance of synaptobrevin dimerization has been disagreed upon in the literature, the TOXCAT and ToxR assays have both shown that the synaptobrevin transmembrane domain associates with a greater affinity than the weak dimers GpA83A and GpA83I (Bowen et al. 2002; Roy et al. 2004). Using analytical ultracentrifugation the free energy of association measured for the synaptobrevin transmembrane domain is weaker than both GpA controls, which have a standard state free energy of association of approximately 3 kcal

mol^{-1} (Doura et al. 2004). In fact, at limiting protein to detergent ratios, we find that the synaptobrevin transmembrane domain does not drive association. Although ToxR data has been shown to correlate with *in vitro* data for the model transmembrane domain glycophorin A (Brosig and Langosch 1998), recent work has shown that not all transmembrane sequences that produce a positive result in these *in vivo* assays show significant affinity for association *in vitro* (Stanley and Fleming 2005). However, consistent with previous studies, we do find that synaptobrevin dimerization can be driven by disulfide crosslinking. In fact, oxidizing conditions were used in the previous study to promote association and to determine the dimerization interface (Laage and Langosch 1997). The formation of a disulfide bond is integral to this dimerization, and sequence context may facilitate this bond by promoting a sterically permissible interface and allowing proper geometry for disulfide linkage. Interestingly, the substitution Leu99Met (SN-SybLM), which creates the putative syntaxin interface, promotes synaptobrevin dimerization in the absence of oxidizing agents. Furthermore, the rate of disulfide bond formation under oxidizing conditions is increased as compared to the wildtype synaptobrevin sequence. This mutant provides insight into the equilibrium population for the heterodimer since the interface of interaction would have only a single amino acid difference on one monomer. These results may indicate a greater propensity for interaction between the transmembrane domains of synaptobrevin and syntaxin than for two synaptobrevin monomers. The implication for the role of the transmembrane region of synaptobrevin in protein-protein interactions is still unclear. The

cysteine residue that promotes crosslinking in synaptobrevin has also been reported to be palmitoylated in adult rat brains (Veit et al. 2000). However, these authors show that palmitoylation of synaptobrevin 2 does not occur in embryonic rat brains and that the palmitoylation in adult rat brain is substiochiometric *in vivo*, indicating that a significant population will not be palmitoylated in adult synaptic vesicles. Although palmitoylation may disrupt association of a synaptobrevin dimer, the fact that unpalmitoylated species exist indicates that this association may still be significant *in vivo*. Furthermore, in addition to disulfide linkage, other factors may drive the association *in vivo*, which is optimized by a sterically permissible interface that stabilize the structure or specify the conformation.

Much of our understanding about the association of membrane proteins is derived from the study of stable α-helical complexes. For instance, the well studied transmembrane dimer Glycophorin A is predominantly dimeric *in vivo*. This protein has served as a model for transmembrane helix interactions in part due to the fact that an extremely stable dimer allows extensive mutagenesis. Transmembrane helices for which association occurs as a part of a cellular process, such as vesicle fusion, would not be expected to associate as strongly as GpA. An accessible equilibrium of species is necessary to participate in dynamic events. Therefore, a thermodynamically weak propensity for association may be integral to cell function. Although the native synaptobrevin sequence may have little ability to drive association, it may still play a major role in important protein-protein interactions.

Implications for multiple interfaces in the syntaxin transmembrane domain

Our data demonstrate that the syntaxin transmembrane domain has a thermodynamic measurable propensity for self-association. Although syntaxin contains a sequence motif similar to that proposed for synaptobrevin, our results suggest that an alternate interface promotes the dimerization observed *in vitro*. The substitution Met267Leu, which creates the putative synaptobrevin interface, has no effect on the ability to dimerize, although the converse mutation promotes dimerization in the synaptobrevin transmembrane domain. Most importantly, the conserved cysteine residue, which promotes dimerization in the synaptobrevin transmembrane domain under oxidative conditions by formation of an intermonomer disulfide linkage, does not form a disulfide linkage in syntaxin under the same experimental conditions. Although the syntaxin sequence corresponding to the previously determined synaptobrevin dimer interface can drive association in other sequence contexts (Laage et al. 2000), the entire syntaxin transmembrane sequence may use an alternate interface for homodimerization. Interestingly, the conserved cysteines do crosslink in the heterodimer, suggesting an alternate interface for homo- and hetero-dimerization for the syntaxin transmembrane domain. The formation of higher order oligomers visualized by SDS-PAGE also indicates multiple interfaces for interaction. These higher order oligomers may be functionally important, since studies suggest that an arrangement of five to eight syntaxin transmembrane segments may compose a proteinaceous fusion pore (Han et al. 2004).

Syntaxin is known to interact with numerous proteins using both transmembrane and soluble domains in the protein. The membrane proximal domain (H3 domain) is known to promote homo-oligomerization (Misura et al. 2001) and is the region in the solved structure of the SNARE complex (Sutton et al. 1998). It has been predicted that large conformational rearrangements occur in the H3 domain to account for the promiscuity of interactions (Misura et al. 2000). Full length syntaxin is also known to interact with many proteins other than SNARE proteins (Yang et al. 2000; Arien et al. 2003; Ganeshan et al. 2003; Liu et al. 2004). Furthermore, the transmembrane domain is known to promote interactions of the full length protein (Lewis et al. 2001). It is conceivable that in addition to the soluble domains, the transmembrane domain also has an inherent plasticity that accounts for interactions with multiple proteins. At the physiological level, these alternate interfaces could act as a molecular switch that participates in dynamic interactions that facilitate vesicle fusion. Further biochemical and structural data are necessary to verify the existence of multiple interfaces in the syntaxin transmembrane domain and the greater role of the transmembrane domains syntaxin and synaptobrevin in vesicle fusion.

5.6 Conclusions

Much of our understanding about the association of membrane proteins is derived from the study of stable α-helical complexes. For instance, the well studied transmembrane dimer glycophorin A is predominantly dimeric *in vivo*. This protein has served as a model for transmembrane helix interactions in part

due to the fact that an extremely stable dimer allows extensive mutagenesis. Transmembrane helices for which association occurs as a part of a cellular process, such as vesicle fusion, would not be expected to associate as strongly as GpA. An accessible equilibrium of species is necessary to participate in dynamic events. Therefore, a thermodynamically weak propensity for association may be integral to cell function. Although the syntaxin and synaptobrevin sequences may have little ability to drive association, their association may still play a major role in important protein-protein interactions.

CHAPTER 6

Conclusions and Remarks

A profound understanding of protein folding will have an enormous intellectual and technological impact on the biological sciences. Membrane proteins are not well studied structurally or thermodynamically, and neither are the interactions that drive membrane protein folding. In contrast, the hydrophobic effect is considered to be the dominant force to drive folding of soluble proteins, although the amino acid sequence defines the native fold. However, how the native fold is achieved is not well understood in either soluble or membrane proteins. The role of sequence in defining the native fold may be more easily explored in membrane proteins, and may be applicable to all proteins. Therefore, the exploration of membrane protein folding may also have significant impact on the understanding of soluble protein.

In this thesis, we focused on the relationship between amino acid sequence, energetics, and the native fold of membrane proteins. Although, this work does not lead us to a simple model to understand folding, it brings us to a deeper understanding of the role of sequence context. We find that the entire sequence is involved in specifying both stability and conformation. We show that perturbations in local packing interactions can largely explain the energetic effect of mutations. However, we also demonstrate that the strong coupling of distant sites in the transmembrane sequence is also integral to defining the stability.

Therefore, the entire length of the transmembrane sequence is interacting to specify the stability, not just local interactions.

Limited structural data provide insight into the nature of interactions that define stability, but also point out many underlying questions that should be explored. Although changes in packing may explain energetics, we find that large rearrangements may occur in single mutants that have modest effects on stability. This result highlights that the interactions that specify the native fold may not be the same as those that stabilize it. Furthermore, although mutations may not perturb the α-helical structure of the unfolded state of a membrane protein, the interactions between lipid and protein in both native and unfolded states define stability. These interactions are also dependent on peptide sequence, must be examined in more detail to truly understand folding.

There are feasible experiments that could shed light on the questions that arise from this work. In particular, the determination of high resolution structures can better elucidate the structural rearrangements that are induced upon mutations. Furthermore, although it may be experimentally difficult, the determination of the contribution of enthalpy and entropy to stability may provide insight into the forces that drive folding. In particular, this information may distinguish between the contribution of interactions within and between protein and lipid in driving association. Finally, extensive study of another model system will demonstrate the universality of the results revealed in this study. A useful model system may be syntaxin, which we demonstrate may have multiple interfaces for helix-helix interaction. Therefore, in this system, the role of

sequence in defining a lipid and a helix interaction face of the protein may be more complex. Additionally, helix-helix association can be studied in a bilayer using rigorous thermodynamic techniques. This should be more applicable to the environment in the cell, and may also provide insight into the effect of environment on association. These experiments highlight a few of the many avenues that should be explored to better understand membrane protein folding.

REFERENCES

Ackers, G.K., and Smith, F.R. 1985. Effects of site-specific amino acid modification on protein interactions and biological function. *Annu Rev Biochem* **54:** 597-629.

Adams, P.D., Arkin, I.T., Engelman, D.M., and Brunger, A.T. 1995. Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. *Nat Struct Biol* **2:** 154-162.

Adams, P.D., Engelman, D.M., and Brunger, A.T. 1996. Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins* **26:** 257-261.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* **181:** 223-230.

Arien, H., Wiser, O., Arkin, I.T., Leonov, H., and Atlas, D. 2003. Syntaxin 1A modulates the voltage-gated L-type calcium channel (Ca(v)1.2) in a cooperative manner. *J Biol Chem* **278:** 29231-29239.

Arrondo, J.L., Castresana, J., Valpuesta, J.M., and Goni, F.M. 1994. Structure and thermal denaturation of crystalline and noncrystalline cytochrome oxidase as studied by infrared spectroscopy. *Biochemistry* **33:** 11650-11655.

Arselin, G., Giraud, M.F., Dautant, A., Vaillier, J., Brethes, D., Coulary-Salin, B., Schaeffer, J., and Velours, J. 2003. The GxxxG motif of the transmembrane domain of subunit e is involved in the dimerization/oligomerization of the yeast ATP synthase complex in the mitochondrial membrane. *Eur J Biochem* **270:** 1875-1884.

Baker, B.M., and Murphy, K.P. 1998. Prediction of binding energetics from structure using empirical parameterization. *Methods Enzymol* **295:** 294-315.

Baldwin, J.M. 1993. The probable arrangement of the helices in G protein-coupled receptors. *Embo J* **12:** 1693-1703.

Bansal, M., Kumar, S., and Velavan, R. 2000. HELANAL: a program to characterize helix geometry in proteins. *J Biomol Struct Dyn* **17:** 811-819.

Bargmann, C.I., Hung, M.C., and Weinberg, R.A. 1986. Multiple independent activations of the neu oncogene by a point mutation altering the transmembrane domain of p185. *Cell* **45:** 649-657.

Bormann, B.J., Knowles, W.J., and Marchesi, V.T. 1989. Synthetic peptides mimic the assembly of transmembrane glycoproteins. *J Biol Chem* **264:** 4033-4037.

Bowen, M.E., Engelman, D.M., and Brunger, A.T. 2002. Mutational analysis of synaptobrevin transmembrane domain oligomerization. *Biochemistry* **41:** 15861-15866.

Bowie, J.U. 1997a. Helix packing angle preferences. *Nat Struct Biol* **4:** 915-917.

Bowie, J.U. 1997b. Helix packing in membrane proteins. *J Mol Biol* **272:** 780-789.

Bowie, J.U. 2001. Stabilizing membrane proteins. *Curr Opin Struct Biol* **11:** 397-402.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. 1983. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem* **4:** 187-217.

Brosig, B., and Langosch, D. 1998. The dimerization motif of the glycophorin A transmembrane segment in membranes: importance of glycine residues. *Protein Sci* **7:** 1052-1056.

Careaga, C.L., and Falke, J.J. 1992. Thermal motions of surface alpha-helices in the D-galactose chemosensory receptor. Detection by disulfide trapping. *J Mol Biol* **226:** 1219-1235.

Carson, M. 1997. Ribbons. *Methods in Enzymology* **277:** 493-505.

Casassa, E.F., and Eisenberg, H. 1964. Thermodynamic Analysis of Multicomponent Solutions. *Adv Protein Chem* **19:** 287-395.

Chamberlain, A.K., Faham, S., Yohannan, S., and Bowie, J.U. 2003. Construction of helix-bundle membrane proteins. *Adv Protein Chem* **63:** 19-46.

Chen, E.Y., Bartlett, M.C., Loo, T.W., and Clarke, D.M. 2004. The DeltaF508 mutation disrupts packing of the transmembrane segments of the cystic fibrosis transmembrane conductance regulator. *J Biol Chem* **279:** 39620-39627.

Chen, G.Q., and Gouaux, E. 1999. Probing the folding and unfolding of wild-type and mutant forms of bacteriorhodopsin in micellar solutions: evaluation of reversible unfolding conditions. *Biochemistry* **38:** 15380-15387.

Chen, H., and Kendall, D.A. 1995. Artificial transmembrane segments. Requirements for stop transfer and polypeptide orientation. *J Biol Chem* **270:** 14115-14122.

Chen, J., and Stites, W.E. 2001. Energetics of side chain packing in staphylococcal nuclease assessed by systematic double mutant cycles. *Biochemistry* **40:** 14004-14011.

Choi, M.Y., Partridge, A.W., Daniels, C., Du, K., Lukacs, G.L., and Deber, C.M. 2005. Destabilization of the transmembrane domain induces misfolding in a phenotypic mutant of cystic fibrosis transmembrane conductance regulator. *J Biol Chem* **280:** 4968-4974.

Choma, C., Gratkowski, H., Lear, J.D., and DeGrado, W.F. 2000. Asparagine-mediated self-association of a model transmembrane helix. *Nat Struct Biol* **7:** 161-166.

Constantinescu, S.N., Keren, T., Socolovsky, M., Nam, H., Henis, Y.I., and Lodish, H.F. 2001. Ligand-independent oligomerization of cell-surface erythropoietin receptor is mediated by the transmembrane domain. *Proc Natl Acad Sci U S A* **98:** 4379-4384.

Cosson, P., Lankford, S.P., Bonifacino, J.S., and Klausner, R.D. 1991. Membrane protein association by potential intramembrane charge pairs. *Nature* **351:** 414-416.

Creamer, T.P., Srinivasan, R., and Rose, G.D. 1995. Modeling unfolded states of peptides and proteins. *Biochemistry* **34:** 16245-16250.

Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* **10:** 673-676.

Cuello, L.G., Cortes, D.M., and Perozo, E. 2004. Molecular architecture of the KvAP voltage-dependent K+ channel in a lipid bilayer. *Science* **306:** 491-495.

Dawkins, R. 1996. *The blind watchmaker: Why the evidence of evolution reveals a universe without design.* Norton, New York.

Dawson, J.P., Melnyk, R.A., Deber, C.M., and Engelman, D.M. 2003. Sequence context strongly modulates association of polar residues in transmembrane helices. *J Mol Biol* **331:** 255-262.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. 1984. X-ray structure analysis of a membrane protein complex. Electron density map at 3 A resolution and a model of the chromophores of the photosynthetic reaction center from Rhodopseudomonas viridis. *J Mol Biol* **180:** 385-398.

Dill, K.A. 1990. The meaning of hydrophobicity. *Science* **250:** 297-298.

Dill, K.A. 1999. Polymer principles and protein folding. *Protein Sci* **8:** 1166-1180.

Dohnal, J.C., Potempa, L.A., and Garvin, J.E. 1980. The molecular weights of three forms of glycophorin A in sodium dodecyl sulfate solution. *Biochim Biophys Acta* **621:** 255-264.

Doura, A.K., and Fleming, K.G. 2004. Complex interactions at the helix-helix interface stabilize the glycophorin A transmembrane dimer. *J Mol Biol* **343:** 1487-1497.

Doura, A.K., Kobus, F.J., Dubrovsky, L., Hibbard, E., and Fleming, K.G. 2004. Sequence context modulates the stability of a GxxxG-mediated transmembrane helix-helix dimer. *J Mol Biol* **341:** 991-998.

Eilers, M., Patel, A.B., Liu, W., and Smith, S.O. 2002. Comparison of helix interactions in membrane and soluble alpha-bundle proteins. *Biophys J* **82:** 2720-2736.

Eilers, M., Shekar, S.C., Shieh, T., Smith, S.O., and Fleming, P.J. 2000. Internal packing of helical membrane proteins. *Proc Natl Acad Sci U S A* **97:** 5796-5801.

Eisenberg, D., Wilcox, W., and McLachlan, A.D. 1986. Hydrophobicity and amphiphilicity in protein structure. *J Cell Biochem* **31:** 11-17.

Engelman, D.M., Steitz, T.A., and Goldman, A. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu Rev Biophys Biophys Chem* **15:** 321-353.

Engelman, D.M., and Zaccai, G. 1980. Bacteriorhodopsin is an inside-out protein. *Proc Natl Acad Sci U S A* **77:** 5894-5898.

Eriksson, A.E., Baase, W.A., Xhang, X.-J., Heinz, D.W., Blaber, M., P., B.E., and Matthews, B.W. 1992. Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect. *Science* **255:** 178-183.

Ermolova, N., Guan, L., and Kaback, H.R. 2003. Intermolecular thiol cross-linking via loops in the lactose permease of Escherichia coli. *Proc Natl Acad Sci U S A* **100:** 10187-10192.

Faham, S., Yang, D., Bare, E., Yohannan, S., Whitelegge, J.P., and Bowie, J.U. 2004. Side-chain contributions to membrane protein structure and stability. *J Mol Biol* **335:** 297-305.

Fasshauer, D., Antonin, W., Subramaniam, V., and Jahn, R. 2002. SNARE assembly and disassembly exhibit a pronounced hysteresis. *Nat Struct Biol* **9:** 144-151.

Fauchere, J.L., Charton, M., Kier, L.B., Verloop, A., and Pliska, V. 1988. Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int J Pept Protein Res* **32:** 269-278.

Faulstich, H., and Heintz, D. 1995. Reversible introduction of thiol compounds into proteins by use of activated mixed disulfides. *Methods Enzymol* **251:** 357-366.

Fleming, K.G. 1998. Measuring transmembrane a-helix energies using analytical ultracentrifugation. In *Chemtracts: biological applications of the analytical ultracentrifuge*. (ed. J.C. Hanson), pp. 985-990. Springer-Verlag, New York.

Fleming, K.G. 2002. Standardizing the free energy change of transmembrane helix-helix interactions. *J Mol Biol* **323:** 563-571.

Fleming, K.G., Ackerman, A.L., and Engelman, D.M. 1997. The Effect of Point Mutations on the Free Energy of Transmembrane a-Helix Dimerization. *Journal of Molecular Biology* **272:** 266-275.

Fleming, K.G., and Engelman, D.M. 2001a. Computation and mutagenesis suggest a right-handed structure for the synaptobrevin transmembrane dimer. *Proteins* **45:** 313-317.

Fleming, K.G., and Engelman, D.M. 2001b. Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proc Natl Acad Sci U S A* **98:** 14340-14344.

Fodor, A.A., and Aldrich, R.W. 2004. On evolutionary conservation of thermodynamic coupling in proteins. *J Biol Chem* **279:** 19046-19050.

Freites, J.A., Tobias, D.J., von Heijne, G., and White, S.H. 2005. Interface connections of a transmembrane voltage sensor. *Proc Natl Acad Sci U S A* **102:** 15059-15064.

Furthmayr, H., and Marchesi, V.T. 1976. Subunit structure of human erthrocyte glycophorin A. *Biochemistry* **15:** 1137-1144.

Furthmayr, H., Tomita, M., and Marchesi, V.T. 1975. Fractionation of the major sialoglycopeptides of the human red blood cell membrane. *Biochem Biophys Res Commun* **65:** 113-121.

Ganeshan, R., Di, A., Nelson, D.J., Quick, M.W., and Kirk, K.L. 2003. The interaction between syntaxin 1A and cystic fibrosis transmembrane conductance regulator Cl- channels is mechanistically distinct from syntaxin 1A-SNARE interactions. *J Biol Chem* **278:** 2876-2885.

Grasberger, B., Minton, A.P., DeLisi, C., and Metzger, H. 1986. Interaction Between Proteins Localized in Membranes. *Proceedings of the National Academy of Sciences of the United States of America* **83:** 6258-6262.

Green, S.M., Meeker, A.K., and Shortle, D. 1992. Contributions of the polar, uncharged amino acids to the stability of staphylococcal nuclease:

evidence for mutational effects on the free energy of the denatured state. *Biochemistry* **31:** 5717-5728.
Green, S.M., and Shortle, D. 1993. Patterns of nonadditivity between pairs of stability mutations in staphylococcal nuclease. *Biochemistry* **32:** 10131-10139.
Haltia, T., and Freire, E. 1995. Forces and factors that contribute to the structural stability of membrane proteins. *Biochim Biophys Acta* **1241:** 295-322.
Haltia, T., Semo, N., Arrondo, J.L., Goni, F.M., and Freire, E. 1994. Thermodynamic and structural stability of cytochrome c oxidase from Paracoccus denitrificans. *Biochemistry* **33:** 9731-9740.
Han, X., Wang, C.T., Bai, J., Chapman, E.R., and Jackson, M.B. 2004. Transmembrane segments of syntaxin line the fusion pore of Ca2+-triggered exocytosis. *Science* **304:** 289-292.
Hanson, P.I., Heuser, J.E., and Jahn, R. 1997. Neurotransmitter release - four years of SNARE complexes. *Curr Opin Neurobiol* **7:** 310-315.
Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and von Heijne, G. 2005a. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433:** 377-381.
Hessa, T., White, S.H., and von Heijne, G. 2005b. Membrane insertion of a potassium-channel voltage sensor. *Science* **307:** 1427.
Hill, B.C., Cook, K., and Robinson, N.C. 1988. Subunit dissociation and protein unfolding in the bovine heart cytochrome oxidase complex induced by guanidine hydrochloride. *Biochemistry* **27:** 4741-4747.
Hodgkin, D.C. 1972. In *Nobel Lectures, Chemistry 1963-1970*. Elsevier Publishing Company, Amsterdam.
Horovitz, A., and Fersht, A.R. 1992. Co-operative interactions during protein folding. *J Mol Biol* **224:** 733-740.
Huang, K.S., Bayley, H., Liao, M.J., London, E., and Khorana, H.G. 1981. Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. *J Biol Chem* **256:** 3802-3809.
Hughson, A.G., Lee, G.F., and Hazelbauer, G.L. 1997. Analysis of protein structure in intact cells: crosslinking in vivo between introduced cysteines in the transmembrane domain of a bacterial chemoreceptor. *Protein Sci* **6:** 315-322.
Jacobs, R.E., and White, S.H. 1989. The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices. *Biochemistry* **28:** 3421-3437.
Jahn, R., Lang, T., and Sudhof, T.C. 2003. Membrane fusion. *Cell* **112:** 519-533.
Jahnig, F. 1983. Thermodynamics and kinetics of protein incorporation into membranes. *Proc Natl Acad Sci U S A* **80:** 3691-3695.
Jayasinghe, S., Hristova, K., and White, S.H. 2001. MPtopo: A database of membrane protein topology. *Protein Sci* **10:** 455-458.
Jiang, S., and Vakser, I.A. 2004. Shorter side chains optimize helix-helix packing. *Protein Sci* **13:** 1426-1429.

Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B.T., and MacKinnon, R. 2003. X-ray structure of a voltage-dependent K+ channel. *Nature* **423:** 33-41.

Johnson, M.L., Correia, J.J., Yphantis, D.A., and Halvorson, H.R. 1981. Analysis of Data from the Analytical Ultracentrifuge by Nonlinear Least-Squares Techniques. *Biophysical Journal* **36:** 575-588.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181:** 662-666.

Killian, J.A. 2003. Synthetic peptides as models for intrinsic membrane proteins. *FEBS Lett* **555:** 134-138.

Kleiger, G., Grothe, R., Mallick, P., and Eisenberg, D. 2002. GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry* **41:** 5990-5997.

Kobus, F.J., and Fleming, K.G. 2005. The GxxxG-containing transmembrane domain of the CCK4 oncogene does not encode preferential self-interactions. *Biochemistry* **44:** 1464-1470.

Kyte, J., and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157:** 105-132.

Laage, R., and Langosch, D. 1997. Dimerization of the synaptic vesicle protein synaptobrevin (vesicle-associated membrane protein) II depends on specific residues within the transmembrane segment. *Eur J Biochem* **249:** 540-546.

Laage, R., Rohde, J., Brosig, B., and Langosch, D. 2000. A conserved membrane-spanning amino acid motif drives homomeric and supports heteromeric assembly of presynaptic SNARE proteins. *J Biol Chem* **275:** 17481-17487.

Langosch, D., Brosig, B., Kolmar, H., and Fritz, H.J. 1996. Dimerisation of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. *J Mol Biol* **263:** 525-530.

Langosch, D., Crane, J.M., Brosig, B., Hellwig, A., Tamm, L.K., and Reed, J. 2001. Peptide mimics of SNARE transmembrane segments drive membrane fusion depending on their conformational plasticity. *J Mol Biol* **311:** 709-721.

Laue, T.M., Shah, B., Ridgeway, T.M., and Pelletier, S.L. 1992. Computer-aided Interpretation of Analytical Sedimentation Data for Proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. (eds. S.E. Harding, A.J. Rowe, and J.C. Horton), pp. 90-125. Royal Society of Chemistry, Cambridge.

Laue, T.M., Shah, B., Ridgeway, T. M., and Pelletier, S. L. 1992. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. (ed. S.E. (Harding, Rowe, A. J., and Horton, J. C., eds)), pp. 90-125. Royal Society of Chemistry, Cambridge.

Lebowitz, J., Lewis, M.S., and Schuck, P. 2002. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci* **11:** 2067-2079.

Lee, K.H., Xie, D., Freire, E., and Amzel, L.M. 1994. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins* **20:** 68-84.

Lee, S.F., Shah, S., Yu, C., Wigley, C., Li, H., Lim, M., Pedersen, K., Han, W., Thomas, P., Lundkvist, J., et al. 2003. A conserved GXXXG motif in APH-1 is critical for assembly and activity of the gamma-secretase complex. *J Biol Chem*.

Lehnert, U., Xia, Y., Royce, T.E., Goh, C.S., Liu, Y., Senes, A., Yu, H., Zhang, Z.L., Engelman, D.M., and Gerstein, M. 2004. Computational analysis of membrane proteins: genomic occurrence, structure prediction and helix interactions. *Q Rev Biophys* **37:** 121-146.

Lemmon, M.A., and Engelman, D.M. 1994. Specificity and promiscuity in membrane helix interactions. *Q Rev Biophys* **27:** 157-218.

Lemmon, M.A., Flanagan, J.M., Hunt, J.F., Adair, B.D., Bormann, B.J., Dempsey, C.E., and Engelman, D.M. 1992a. Glycophorin A Dimerization is Driven by Specific Interactions between Transmembrane a-Helices. *Journal of Biological Chemistry* **267:** 7683-7689.

Lemmon, M.A., Flanagan, J.M., Hunt, J.F., Adair, B.D., Bormann, B.J., Dempsey, C.E., and Engelman, D.M. 1992b. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *J Biol Chem* **267:** 7683-7689.

Lemmon, M.A., Flanagan, J.M., Hunt, J.F., Adair, B.D., Bormann, B.J., Dempsey, C.E., and Engelman, D.M. 1992a. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *J Biol Chem* **267:** 7683-7689.

Lemmon, M.A., Flanagan, J.M., Treutlein, H.R., Zhang, J., and Engelman, D.M. 1992c. Sequence Specificity in the Dimerization of Transmembrane a-Helices. *Biochemistry* **31:** 12719-12725.

Lemmon, M.A., Flanagan, J.M., Treutlein, H.R., Zhang, J., and Engelman, D.M. 1992b. Sequence specificity in the dimerization of transmembrane alpha-helices. *Biochemistry* **31:** 12719-12725.

Lemmon, M.A., Treutlein, H.R., Adams, P.D., Brunger, A.T., and Engelman, D.M. 1994. A dimerization motif for transmembrane alpha-helices. *Nat Struct Biol* **1:** 157-163.

Levinthal, C. 1968. Are there pathways for protein folding? *J Chim Phys* **65:** 44-45.

Lewis, J.L., Dong, M., Earles, C.A., and Chapman, E.R. 2001. The transmembrane domain of syntaxin 1A is critical for cytoplasmic domain protein-protein interactions. *J Biol Chem* **276:** 15458-15465.

Liao, M.J., Huang, K.S., and Khorana, H.G. 1984. Regeneration of native bacteriorhodopsin structure from fragments. *J Biol Chem* **259:** 4200-4204.

LiCata, V.J., Dalessio, P.M., and Ackers, G.K. 1993. Single-site modifications of half-ligated hemoglobin reveal autonomous dimer cooperativity within a quaternary T tetramer. *Proteins* **17:** 279-296.

Liu, J., Ernst, S.A., Gladycheva, S.E., Lee, Y.Y., Lentz, S.I., Ho, C.S., Li, Q., and Stuenkel, E.L. 2004a. Fluorescence resonance energy transfer reports

properties of syntaxin1a interaction with Munc18-1 in vivo. *J Biol Chem* **279:** 55924-55936.
Liu, W., Crocker, E., Siminovitch, D.J., and Smith, S.O. 2003. Role of side-chain conformational entropy in transmembrane helix dimerization of glycophorin A. *Biophys J* **84:** 1263-1271.
Liu, W., Eilers, M., Patel, A.B., and Smith, S.O. 2004b. Helix packing moments reveal diversity and conservation in membrane protein structure. *J Mol Biol* **337:** 713-729.
London, E., and Khorana, H.G. 1982. Denaturation and renaturation of bacteriorhodopsin in detergents and lipid-detergent mixtures. *J Biol Chem* **257:** 7003-7011.
Luecke, H., Schobert, B., Richter, H.T., Cartailler, J.P., and Lanyi, J.K. 1999. Structure of bacteriorhodopsin at 1.55 A resolution. *J Mol Biol* **291:** 899-911.
MacKenzie, K.R., Prestegard, J.H., and Engelman, D.M. 1997. A Transmembrane Helix Dimer: Structure and Implications. *Science* **276:** 131-133.
Margittai, M., Otto, H., and Jahn, R. 1999. A stable interaction between syntaxin 1a and synaptobrevin 2 mediated by their transmembrane domains. *FEBS Lett* **446:** 40-44.
McClain, M.S., Cao, P., and Cover, T.L. 2001. Amino-terminal hydrophobic region of Helicobacter pylori vacuolating cytotoxin (VacA) mediates transmembrane protein dimerization. *Infect Immun* **69:** 1181-1184.
McClain, M.S., Iwamoto, H., Cao, P., Vinion-Dubiel, A.D., Li, Y., Szabo, G., Shao, Z., and Cover, T.L. 2003. Essential role of a GXXXG motif for membrane channel formation by Helicobacter pylori vacuolating toxin. *J Biol Chem* **278:** 12101-12108.
Melnyk, R.A., Kim, S., Curran, A.R., Engelman, D.M., Bowie, J.U., and Deber, C.M. 2004. The affinity of GXXXG motifs in transmembrane helix-helix interactions is modulated by long-range communication. *J Biol Chem* **279:** 16591-16597.
Mendrola, J.M., Berger, M.B., King, M.C., and Lemmon, M.A. 2002. The single transmembrane domains of ErbB receptors self-associate in cell membranes. *J Biol Chem* **277:** 4704-4712.
Mildvan, A.S., Weber, D.J., and Kuliopulos, A. 1992. Quantitative interpretations of double mutations of enzymes. *Arch Biochem Biophys* **294:** 327-340.
Mingarro, I., Whitley, P., Lemmon, M.A., and von Heijne, G. 1996. Ala-insertion scanning mutagenesis of the glycophorin A transmembrane helix: a rapid way to map helix-helix interactions in integral membrane proteins. *Protein Sci* **5:** 1339-1341.
Misura, K.M., Scheller, R.H., and Weis, W.I. 2000. Three-dimensional structure of the neuronal-Sec1-syntaxin 1a complex. *Nature* **404:** 355-362.
Misura, K.M., Scheller, R.H., and Weis, W.I. 2001. Self-association of the H3 region of syntaxin 1A. Implications for intermediates in SNARE complex assembly. *J Biol Chem* **276:** 13273-13282.

Murphy, K.P. 1999. Predicting binding energetics from structure: looking beyond DeltaG degrees. *Med Res Rev* **19:** 333-339.
Murphy, K.P., Privalov, P.L., and Gill, S.J. 1990. Common features of protein unfolding and dissolution of hydrophobic compounds. *Science* **247:** 559-561.
Nagy, J.K., Lau, F.W., Bowie, J.U., and Sanders, C.R. 2000. Mapping the oligomeric interface of diacylglycerol kinase by engineered thiol cross-linking: homologous sites in the transmembrane domain. *Biochemistry* **39:** 4154-4164.
Nonet, M.L., Saifee, O., Zhao, H., Rand, J.B., and Wei, L. 1998. Synaptic transmission deficits in Caenorhabditis elegans synaptobrevin mutants. *J Neurosci* **18:** 70-80.
Nozaki, Y., and Tanford, C. 1971. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J Biol Chem* **246:** 2211-2217.
Overton, M.C., Chinault, S.L., and Blumer, K.J. 2003. Oligomerization, biogenesis, and signaling is promoted by a glycophorin A-like dimerization motif in transmembrane domain 1 of a yeast G protein-coupled receptor. *J Biol Chem* **278:** 49369-49377.
Pattabiraman, N., Ward, K.B., and Fleming, P.J. 1995. Occluded Molecular Surface: Analysis of Protein Packing. *Journal of Molecular Recognition* **8:** 334-344.
Pauling, L., and Corey, R.B. 1951a. Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc Natl Acad Sci U S A* **37:** 235-240.
Pauling, L., and Corey, R.B. 1951b. The polypeptide-chain configuration in hemoglobin and other globular proteins. *Proc Natl Acad Sci U S A* **37:** 282-285.
Perry, K.M., Onuffer, J.J., Gittelman, M.S., Barmat, L., and Matthews, C.R. 1989. Long-range electrostatic interactions can influence the folding, stability, and cooperativity of dihydrofolate reductase. *Biochemistry* **28:** 7961-7968.
Petrache, H.I., Grossfield, A., MacKenzie, K.R., Engelman, D.M., and Woolf, T.B. 2000. Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: molecular dynamics calculations. *J Mol Biol* **302:** 727-746.
Poirier, M.A., Xiao, W., Macosko, J.C., Chan, C., Shin, Y.K., and Bennett, M.K. 1998. The synaptic SNARE complex is a parallel four-stranded helical bundle. *Nat Struct Biol* **5:** 765-769.
Popot, J.L., and Engelman, D.M. 1990. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29:** 4031-4037.
Popot, J.L., and Engelman, D.M. 2000. Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* **69:** 881-922.
Popot, J.L., Gerchman, S.E., and Engelman, D.M. 1987. Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process. *J Mol Biol* **198:** 655-676.

Popot, J.L., Trewhella, J., and Engelman, D.M. 1986. Reformation of crystalline purple membrane from purified bacteriorhodopsin fragments. *Embo J* **5:** 3039-3044.

Radzicka, A., Pedersen, L., and Wolfenden, R. 1988. Influences of solvent water on protein folding: free energies of solvation of cis and trans peptides are nearly identical. *Biochemistry* **27:** 4538-4541.

Rees, D.C., DeAntonio, L., and Eisenberg, D. 1989. Hydrophobic organization of membrane proteins. *Science* **245:** 510-513.

Renthal, R. 1999. Transmembrane and water-soluble helix bundles display reverse patterns of surface roughness. *Biochem Biophys Res Commun* **263:** 714-717.

Reynolds, J.A., and Tanford, C. 1976. Determination of molecular weight of the protein moiety in protein-detergent complexes without direct knowledge of detergent binding. *Proc Natl Acad Sci U S A* **73:** 4467-4470.

Roy, R., Laage, R., and Langosch, D. 2004. Synaptobrevin transmembrane domain dimerization-revisited. *Biochemistry* **43:** 4964-4970.

Ruan, W., Lindner, E., and Langosch, D. 2004. The interface of a membrane-spanning leucine zipper mapped by asparagine-scanning mutagenesis. *Protein Sci* **13:** 555-559.

Russ, W.P., and Engelman, D.M. 1999. TOXCAT: A measure of transmembrane helix association in a biological membrane. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 863-868.

Russ, W.P., and Engelman, D.M. 2000. The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol* **296:** 911-919.

Saifee, O., Wei, L., and Nonet, M.L. 1998. The Caenorhabditis elegans unc-64 locus encodes a syntaxin that interacts genetically with synaptobrevin. *Mol Biol Cell* **9:** 1235-1252.

Sakaguchi, M., Tomiyoshi, R., Kuroiwa, T., Mihara, K., and Omura, T. 1992. Functions of signal and signal-anchor sequences are determined by the balance between the hydrophobic segment and the N-terminal charge. *Proc Natl Acad Sci U S A* **89:** 16-19.

Seelig, J., and Ganz, P. 1991. Nonclassical hydrophobic effect in membrane binding equilibria. *Biochemistry* **30:** 9354-9359.

Segrest, J.P., Kahane, I., Jackson, R.L., and Marchesi, V.T. 1973. Major glycoprotein of the human erythrocyte membrane: evidence for an amphipathic molecular structure. *Arch Biochem Biophys* **155:** 167-183.

Senes, A., Engel, D.E., and DeGrado, W.F. 2004. Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Curr Opin Struct Biol* **14:** 465-479.

Senes, A., Gerstein, M., and Engelman, D.M. 2000. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol* **296:** 921-936.

Senes, A., Ubarretxena-Belandia, I., and Engelman, D.M. 2001. The Calpha --- H...O hydrogen bond: a determinant of stability and specificity in

transmembrane helix interactions. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 9056-9061.
Shirley, B.A., Stanssens, P., Hahn, U., and Pace, C.N. 1992. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* **31:** 725-732.
Shortle, D. 1996. The denatured state (the other half of the folding equation) and its role in protein stability. *Faseb J* **10:** 27-34.
Smith, S.O., Eilers, M., Song, D., Crocker, E., Ying, W., Groesbeek, M., Metz, G., Ziliox, M., and Aimoto, S. 2002. Implications of threonine hydrogen bonding in the glycophorin A transmembrane helix dimer. *Biophysical Journal* **82:** 2476-2486.
Smith, S.O., Smith, C.S., and Bormann, B.J. 1996. Strong hydrogen bonding interactions involving a buried glutamic acid in the transmembrane sequence of the neu/erbB-2 receptor. *Nat Struct Biol* **3:** 252-258.
Smith, S.O., Song, D., Shekar, S., Groesbeek, M., Ziliox, M., and Aimoto, S. 2001. Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. *Biochemistry* **40:** 6553-6558.
Stanley, A.M., and Fleming, K.G. 2005. The transmembrane domains of ErbB receptors do not dimerize strongly in micelles. *J Mol Biol* **347:** 759-772.
Stevens, T.J., and Arkin, I.T. 1999. Are membrane proteins "inside-out" proteins? *Proteins* **36:** 135-143.
Sulistijo, E.S., Jaszewski, T.M., and MacKenzie, K.R. 2003. Sequence-specific dimerization of the transmembrane domain of the "BH3-only" protein BNIP3 in membranes and detergent. *J Biol Chem* **278:** 51950-51956.
Sutton, R.B., Fasshauer, D., Jahn, R., and Brunger, A.T. 1998. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature* **395:** 347-353.
Szule, J.A., and Coorssen, J.R. 2003. Revisiting the role of SNAREs in exocytosis and membrane fusion. *Biochim Biophys Acta* **1641:** 121-135.
Traub, W., Yonath, A., and Segal, D.M. 1969. On the molecular structure of collagen. *Nature* **221:** 914-917.
Ungar, D., and Hughson, F.M. 2003. SNARE protein structure and function. *Annu Rev Cell Dev Biol* **19:** 493-517.
Van den Berg, B., Clemons, W.M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. 2004. X-ray structure of a protein-conducting channel. *Nature* **427:** 36-44.
van Montfort, B.A., Schuurman-Wolters, G.K., Duurkens, R.H., Mensen, R., Poolman, B., and Robillard, G.T. 2001. Cysteine cross-linking defines part of the dimer and B/C domain interface of the Escherichia coli mannitol permease. *J Biol Chem* **276:** 12756-12763.
Veit, M., Becher, A., and Ahnert-Hilger, G. 2000. Synaptobrevin 2 is palmitoylated in synaptic vesicles prepared from adult, but not from embryonic brain. *Mol Cell Neurosci* **15:** 408-416.
Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8:** 52-56, 29.

Wallin, E., and von Heijne, G. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **7:** 1029-1038.
Wang, J., and Pullman, A. 1991. Do helices in membranes prefer to form bundles or stay dispersed in the lipid phase? *Biochim Biophys Acta* **1070:** 493-496.
Watson, J.D., and Crick, F.H. 1953. The structure of DNA. *Cold Spring Harb Symp Quant Biol* **18:** 123-131.
Wells, J. 1990. Additivity of mutational effects in proteins. *Biochemistry* **29:** 8509-8517.
White, S.H. 2004. The progress of membrane protein structure determination. *Protein Sci* **13:** 1948-1949.
White, S.H., and Wimley, W.C. 1999. Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* **28:** 319-365.
Wimley, W.C., Creamer, T.P., and White, S.H. 1996. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35:** 5109-5124.
Wimley, W.C., and White, S.H. 1993. Membrane partitioning: distinguishing bilayer effects from the hydrophobic effect. *Biochemistry* **32:** 6307-6312.
Wimley, W.C., and White, S.H. 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol* **3:** 842-848.
Wimley, W.C., and White, S.H. 2000. Designing transmembrane alpha-helices that insert spontaneously. *Biochemistry* **39:** 4432-4442.
Wu, J., and Kaback, H.R. 1996. A general method for determining helix packing in membrane proteins in situ: helices I and II are close to helix VII in the lactose permease of Escherichia coli. *Proc Natl Acad Sci U S A* **93:** 14498-14502.
Yang, B., Steegmaier, M., Gonzalez, L.C., Jr., and Scheller, R.H. 2000. nSec1 binds a closed conformation of syntaxin1A. *J Cell Biol* **148:** 247-252.
Yonath, A., and Traub, W. 1969. Polymers of tripeptides as collagen models. IV. Structure analysis of poly(L-proly-glycyl-L-proline). *J Mol Biol* **43:** 461-477.
Yoshida, A., Oho, C., Omori, A., Kuwahara, R., Ito, T., and Takahashi, M. 1992. HPC-1 is associated with synaptotagmin and omega-conotoxin receptor. *J Biol Chem* **267:** 24925-24928.
Zhou, F.X., Cocco, M.J., Russ, W.P., Brunger, A.T., and Engelman, D.M. 2000. Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat Struct Biol* **7:** 154-160.
Zhou, F.X., Merianos, H.J., Brunger, A.T., and Engelman, D.M. 2001. Polar residues drive association of polyleucine transmembrane helices. *Proc Natl Acad Sci U S A* **98:** 2250-2255.
Zhou, Y., and Bowie, J.U. 2000. Building a thermostable membrane protein. *J Biol Chem* **275:** 6975-6979.

Abigail E. Kroch

Formerly Abigail K. Doura
Email: akroch@jhu.edu
Born February 14, 1978

Education:

Johns Hopkins University Baltimore, MD	PhD	February 2006	Biology
University of Chicago Chicago, IL	BA	June 2000	Biology

Publications

Kroch and Fleming, K.G.
(2006) "Packing defects in the Glycophorin A transmembrane dimer induce rearrangements of the dimer interface."
Manuscript in preparation

Kroch and Fleming, K.G.
(2006) "Alternate interfaces may mediate homomeric and heteromeric assembly in transmembrane domains of SNARE proteins"
Journal of Molecular Biology, in press.

Doura, A.K. and Fleming, K.G.
(2004) "Complex Interactions at the helix-helix interface stabilize the transmembrane dimer Glycophorin A"
Journal of Molecular Biology; **343**(5): 1487-97

Doura, A.K., Kobus, F.J., Dubrovsky, L., Hibbard, E., Fleming, K.G.
(2004) "Sequence context modulates the stability of a GxxxG mediated transmembrane helix-helix dimer"
Journal of Molecular Biology; **341**(4): 991-8.

Fleming, K.G., Ren, C.C., Doura, A.K., Eisley, M.E., Kobus, F.J., Stanley, A.M.
(2004) "Thermodynamics of glycophorin A transmembrane helix dimerization in C14 betaine micelles."
Biophysical Chemistry; **108** (1-3):43-9.

Research Experience:

5/2001- **PhD Student**
Karen Fleming, Department of Biophysics,
Johns Hopkins University, Baltimore, MD

3/2001- **Rotation Student**
5/2001 Ian Paulsen, Department of Microbial Ecology and Evolution
The Institute for Genomic Research, Rockville, MD

1/2001-3/2001	**Rotation Student** Marnie Halpern, Carnegie Institute for Embryology Baltimore, MD
11/2000-1/2001	**Rotation Student** Karen Fleming, Department of Biophysics Johns Hopkins University, Baltimore, MD
9/2000-11/2000	**Rotation Student** Allen Shearn, Department of Biology Johns Hopkins University, Baltimore, MD
9/1999-6/2000	**Undergraduate Honors Research Student** Victoria Prince, Department of Organismal Biology and Anatomy University of Chicago, Chicago, IL
6/1999-9/1999	**Summer IRTA Fellow** Tatijana Piotrowski, Laboratory of Molecular Genetics NICHD, NIH, Bethesda, MD
6/1998-9/1998	**Research Assistant** Daniel Kessler, Department of Developmental Biology University of Pennsylvania, Philadelphia, PA

Honors and Awards

2005	Postdoctoral Fellowship UCSF Department of Biochemistry and Biophysics
2005	Student Travel Award Biophysical Society
2004	Invited to talk based on poster FASEB meeting in Molecular Biophysics of Cellular Membranes
2001	NSF Predoctoral Fellowship Honorable Mention
2000	BA with Honors in Biology University of Chicago
2000	Richter Fellowship University of Chicago
1999	Summer IRTA Fellowship National Institutes of Health

Presentations

Feb. 2005	“Complex interactions at the helix-helix interface stabilize the transmembrane dimer Glycophorin A” Poster Presentation at the Biophysical Society Conference
Oct. 2004	“Complex interactions at the helix-helix interface stabilize the transmembrane dimer Glycophorin A” Platform Presentation at the Gibbs Conference of Biothermodynamics

Sept. 2004 "Complex interactions at the helix-helix interface stabilize the transmembrane dimer Glycophorin A"
Platform Presentation at the Johns Hopkins Institute for Biophysical Research Retreat

June 2004 "Complex interactions at the helix-helix interface stabilize the transmembrane dimer Glycophorin A"
Platform and Poster Presentation at the FASEB meeting on the Molecular Biophysics of Cellular Membranes

Oct. 2003 "Double-Mutant Cycles of the Glycophorin A Transmembrane Domain"
Poster Presentation at the Gibbs Conference on Biothermodynamics

Oct. 2002 "Empirical Structure-Based Parameterization of the Effect of Mutations on the Free Energy of GpA Transmembrane Dimerization"
Poster Presentation at the Gibbs Conference on Biothermodynamics