THEORETICAL STUDIES OF TRANSMEMBRANE HELIX ASSOCIATION

by

JINMING ZHANG

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2008

UMI Number: 3296938

UMI Microform 3296938

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Biochemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Themis Lazaridis

_____ _____

Date Chair of Examining Committee

Dr. Lesley Davenport

_____ _____

Date Executive Officer

Dr. Marilyn Gunner

_____

Dr. Michael Green

Dr. Fred Naider

_____

Dr. Iban Ubarretxena-Belandia

Dr. Diana Murray

_____

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Theoretical Studies of Transmembrane Helix Association

by

Jinming Zhang

Advisor: Professor Themis Lazaridis

Association of transmembrane (TM) helices plays a key role in the structure and function of membrane proteins. A large number of experimental studies have been devoted to quantifying the interaction between TM helices in detergent micelles and, more recently, in bilayers. However, despite the abundance of experimental data, our understanding of the structural determinants of TM helix association affinity is still limited. In this work a theoretical methodology is first developed for calculating the standard association free energy of TM helices in both micelles and bilayers based on an implicit membrane model (IMM1). This methodology is applied to the TM domain of Glycophorin A (GpA), a human erythrocyte protein. The calculated standard association free energy of GpA in DPC micelles is in good agreement with the experimental value. The translational entropy cost is larger, while the rotational entropy cost is smaller in bilayers than in micelles. The standard association free energy in DMPC bilayers is calculated to be ~1.3 kcal/mol more favorable than in DPC micelles, consistent with available experimental data. Finally we investigated the question why association affinities of GpA and bacteriophage M13 major coat protein (MCP) are dramatically different although they both contain the GxxxG motif. The calculated association free energies follow

experimental observations: the association affinity of MCP-GpA (MCP mutant with all interfacial residues of MCP substituted for those of GpA) falls between those of GpA15p11 (GpA with 15 residues from the wild type sequence plus 11 flanking residues from the ToxCAT construct) and MCP wild type. MCP and MCP-GpA have the same flanking residues used in the ToxCAT assay as those in GpA15p11, but the position of the flanking residues relative to the GxxxG motif is different. MCP exhibits an equally strong inter-helical interaction in the TM domain. A major reason for the weaker association of MCP in the calculations was the non-interfacial residue Lys40, which in the dimer structure is forced to be buried in the membrane interior. To alleviate the desolvation cost, in MCP and MCP-GpA dimers Lys40 gets deprotonated. A second factor that modulates association affinity is the flanking residues. Thanks to them, GpA15p11 exhibits a much stronger association affinity than GpA29 (GpA with 29 residues all from the wild type sequence). The positioning of the flanking residues is also important, as evidenced by the difference in association affinity between MCP and MCP-GpA on one hand and GpA15p11 on the other. Thus, residues outside the contact interface can exert a significant influence on transmembrane helix association affinity. These theoretical studies enhance our understanding of the molecular mechanisms of TM helix association and could be useful for structure prediction and rational design of TM proteins.

ACKNOWLEDGMENTS

First, I would like to send my earnest thanks to my mentor, Dr. Themis Lazaridis, for introducing me to the exciting field of theoretical and computational biophysics. He has given me considerable knowledge, training, supervision and help in a professional way.

I would like to thank Dr. Marilyn Gunner, Dr. Michael Green, Dr. Fred Naider, Dr. Iban Ubarretxena-Belandia, and Dr. Diana Murray for their guidance and helpful discussions. I also thank Dr. M. Mottamal, Dr. M. Mihajlovic, Dr. Z. Li, Mr. J. Ramos and other colleagues in the lab for their technical assistance or discussion.

Specially, I have to thank my parents for their tremendous support and endless love to me during my years of education towards a Ph.D. degree. I thank my siblings and my best friends for being supportive.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ABNR	Adopted Basis Newton-Raphson;
ACE	the acetyl group;
CBX	the methyl amide group;
C_8E_5	pentaoxyethylene octyl ether;
DDMAB	dodecyl-*N,N*-dimethyl ammonium butyrate;
DLPC	1,2-dilauroyl-sn-glycero-3-phosphocholine;
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine;
DPC	N-dodecylphosphocholine;
FRET	fluorescence resonance energy transfer;
GpA	Glycophorin A;
HP	hydrophobic phase;
IMM1	implicit membrane model 1;
MCP	bacteriophage M13 major coat protein;
MCP-GpA	MCP mutant with all the same interfacial residues as GpA wild type;
MD	molecular dynamics;
NOE	the nuclear overhauser effect;
PAGE	polyacrylamide gel electrophoresis;
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine;
RMSD	the root-mean-square deviation;
SDS	sodium dodecylsulfate;
TM	transmembrane;

Chapter I

INTRODUCTION

1.1 Biological Membranes and membrane proteins

The biological membranes (also called biomembranes or the plasma membranes) are selectively permeable lipid bilayers containing a wide variety of biological molecules, mainly proteins and lipids. They place a physical boundary to protect the cell from the external environment and regulate the molecular traffic across that boundary; they divide the intercellular space into discrete compartments to segregate processes and components; they organize complex reaction sequences; they are central to both biological energy conservation and cell-to-cell communication; and they serve as the attachment point for both the intracellular cytoskeleton and, if present, the cell wall (Lehninger et al. 1993).

Biological membranes represent very complex and dynamic structures. A lipid bilayer containing two leaflets is the basic structural element for membranes. A large number of membrane proteins are found on the entire surface of membranes. The cytoskeleton is found underlying the cell membrane in the cytoplasm. Membrane proteins are oriented asymmetrically and the distribution of membrane lipids is also asymmetric on the two faces of the bilayer, for example in the plasma membrane, certain lipids are typically found primarily in the outer face of the bilayer, and others in the inner face. Membrane lipids are in constant motion within the plane of the membrane, undergoing lateral diffusion, thermal motion of the acyl chains, and transbilayer diffusion. Many membrane proteins also can diffuse laterally on the membrane.

Membrane proteins can be divided into two classes: integral (intrinsic) and

peripheral (extrinsic). Integral proteins (also called transmembrane proteins) are bound to the membrane very firmly. They have at least one hydrophobic domain which spans the membrane and anchors them within the membrane and the surrounding hydrophilic domains. The hydrophobic domain consists of α-helices or β sheet protein motifs. Peripheral proteins are loosely or reversibly attached to integral membrane proteins, or associated with peripheral regions of the lipid bilayer. They can be released from membranes by relatively mild treatments, and they are generally water soluble once released from the membrane. Since technically peripheral proteins can be studied as soluble proteins, we have more interest in integral proteins.

Transmembrane (TM) proteins are extremely important, although very difficult to study. They perform various crucial cellular functions, such as transportation of ions, metabolites and larger molecules, transduction of chemical signals, propagation of electrical impulses, anchoring enzymes and proteins to a specific location, regulation of intracellular vesicular transport, etc. (von Heijne 2007) . Although TM proteins represent ~25% of human genome encodings and ~70% of current pharmaceutical targets, because of difficulties in purification and crystallization, only a small number of TM protein structures have been determined. So far only two basic architectures have been discovered in the TM proteins, the α-helix bundle and the β-barrel. In this thesis we study on the α-helical TM proteins.

1.2 Importance of transmembrane helix association

The interaction between TM helices plays a key role in the structure and function of membrane proteins. It is the second essential step in forming the tertiary structure of

multiple-spanning TM proteins according to the two-stage model and its recent refinements (Popot and Engelman 1990; Engelman et al. 2003). Moreover, the dynamic interaction between TM helices seems to be involved in crucial cellular processes, such as signal transduction and membrane transport (Matthews et al. 2006; von Heijne 2007). The oligomerization of membrane receptor proteins is thought to be very important for signal transduction, however the extent to which the TM domain contributes to this oligomerization is still unclear or even controversial (Stanley and Fleming 2005). From the theoretical point of view, it would be useful to be able to predict a) to what extent two given TM helices will associate and b) the structure of the dimer or oligomer. The recent computational design of synthetic TM helical peptides to interact with target TM proteins with high selectivity opens new avenues for therepeutic intervention (Yin et al. 2007). Thus, understanding the rules that govern the association between TM helices is of great importance for membrane protein structure prediction and drug design.

1.3 Experimental measurements on the free energy of association of transmembrane helices in micelles and bilayers

Glycophorin A (GpA) is a small bitopic (single-span) and the major integral membrane protein on red blood cell membranes. It is a 131-residue glycoprotein with several carbohydrate chains covalently attached on its extra-cellular domain containing its N-terminal. In humans, there are two polymorphic versions of GpA, which differ only at residues 1 and 5. The difference gives rise to the MN blood groups. The function of GpA is not very clear, although it is the most important attachment site by which the parasite Plasmodium falciparum invades human red blood cells.

The 23-residue hydrophobic TM domain of GpA has been studied extensively as a model system for the association of TM helices. The association of GpA TM domain and its variants have been studied by SDS polyacrylamide gel electrophoresis (PAGE) (Lemmon et al. 1992a; Lemmon et al. 1992b), analytical ultracentrifugation and fluorescence resonance energy transfer (FRET) in micelles (Fleming et al. 1997; Fisher et al. 1999; Fleming 2002; Fisher et al. 2003; Fleming et al. 2004), and by the ToxR/ToxCAT/GALLEX system in biological membranes (Langosch et al. 1996; Brosig and Langosch 1998; Russ and Engelman 1999; Finger et al. 2006).

The association free energies of wild type GpA and GpA mutants in a number of micelle environments have been determined experimentally: in C_8E_5 micelles by analytical ultracentrifugation (Fleming et al. 1997; Fleming 2002); in a series of detergents with different alkyl chain length and headgroups by FRET (Fisher et al. 1999; 2003); and in C14 betaine micelles by analytical ultracentrifugation (Fleming et al. 2004). The relative association free energy of GpA variants in lipid bilayers has also been determined (Langosch et al. 1996; Brosig and Langosch 1998; Russ and Engelman 1999; Finger et al. 2006).

The relationship between association free energies determined in micelles and bilayers is not well understood (DeGrado et al. 2003). Langosch et al. (1996) have shown that GpA TM segments appear to be less sensitive to mutation and thus more strongly associated in a natural lipid bilayer than in micelles, probably due to a slightly altered structure of the dimer and/or to a higher local concentration and preorientation of the TM helices in a lipid bilayer. Furthermore, it was found that the association of TM helices of the M2 protein from influenza A virus is two orders of magnitude stronger in DLPC

bilayers than in detergent DPC micelles and even stronger in DMPC and POPC bilayers (Cristian et al. 2003). The origin of this difference is unknown. Considering that M2 forms a tetrameric structure with 4 predicted helix-helix association interfaces (Kim et al. 2003), the difference in association constant corresponds roughly to 0.7 kcal/mol for each helix-helix interface.

1.4 Theoretical studies on the free energy of association of transmembrane helices

Although protein-protein and protein-ligand binding free energy in aqueous solution have been extensively studied (Horton and Lewis 1992; Froloff et al. 1997; Gilson et al. 1997; Helms and Wade 1998; Lazaridis 2002; Lazaridis et al. 2002), little work has been done on the theoretical prediction of association free energy of TM helices. Due to the inhomogeneity along the membrane normal, binding free energy calculations in a membrane should be different from those in aqueous solution. Recently, an empirical method has been proposed to estimate association free energies of α-helices in nonpolar media (Lomize et al. 2004). However this method has several limitations: 1) it neglects the translational and rotational entropy changes upon association; 2) it does not address the difference between micelles and bilayers; and 3) possible conformational changes in the helices upon association are not accounted for. More recently, Henin et al. (2005) estimated the free energy change along a reaction coordinate for GpA dimerization and free energy changes due to mutations, and obtained the association free energy by integrating the potential of mean force. Many other simulations of associating TM helices have been reported in micelles or bilayers (Petrache et al. 2000; Feller et al. 2003; Braun et al. 2004; Beevers and Kukol 2006; Bond and Sansom 2006) but have not included

calculations of the association free energy.

1.5 Motifs for association of transmembrane helices

Mutagenesis studies on the GpA TM domain have shown the great importance of the sequence motif LIxxGVxxGVxxT, and especially the GxxxG motif, to the stability of the homo-dimer (Lemmon et al. 1992a; Lemmon et al. 1992b; Lemmon et al. 1994; Langosch et al. 1996; Brosig and Langosch 1998; Russ and Engelman 1999; 2000). Statistical analysis found that the GxxxG motif occurs in membrane proteins at a frequency far above expectation (Senes et al. 2000). As revealed in both solution and solid-state NMR structures of GpA (MacKenzie et al. 1997; Smith et al. 2001), the 7 residues on the LIxxGVxxGVxxT motif form a close inter-helical packing interface facilitated by the two small glycine residues at the center. It has been proposed that a Cα-H··O hydrogen bond network along this motif also contributes significantly to dimer stability (Senes et al. 2001), an idea confirmed by experiments (Arbely and Arkin 2004) and computation (Mottamal and Lazaridis 2005). Computational modeling studies have shown that van der Waals interactions contribute the most to inter-helical interaction (Mottamal et al. 2006) and association free energy (Lomize et al. 2004; Zhang and Lazaridis 2006). More recently, a more extended motif called the glycine zipper, such as (G,A,S)xxxGxxxG and GxxxGxxx(G,S,T), has been identified and found to promote right-handed packing of TM helices (Kim et al. 2005).

The main emphasis so far for explaining the driving force for TM helix association has been on the residues that form the binding interface. The non-interfacial residues and the residues flanking the 13-residue motif, LIxxGVxxGVxxT, are widely thought to be

unimportant for association affinity since hydrophobic mutations in them had little effect on association affinity (Lemmon et al. 1992b; Lemmon et al. 1994; Russ and Engelman 1999; Fleming and Engelman 2001; Fleming et al. 2004). Nevertheless, substituting non-interfacial residues for polar residues may induce strikingly different results on association affinity, which are difficult to rationalize (Lemmon et al. 1992a; Lemmon et al. 1992b). To investigate how the affinity of the GxxxG motif is modulated, the association affinities of bacteriophage M13 major coat protein (MCP) and GpA were recently compared on the *Escherichia coli* inner membrane by the ToxCat assay (Melnyk et al. 2004). Although both GpA and MCP contain a GxxxG motif in their TM domains and form a homo-dimer, the association affinity of MCP is dramatically weaker than GpA. Even when each of the interfacial residues on MCP were substituted with those from the LIxxGVxxGVxxT motif on GpA, the association affinity of that MCP mutant (MCP-GpA) was significantly weaker than that of GpA. This observation can not be explained by steric clashes or the loss of favorable contacts on the inter-helical interface – explanations that worked for most earlier mutagenesis studies (Doura and Fleming 2004; Doura et al. 2004). Another issue that needs to be investigated is whether the flanking residues used in the *in vivo* constructs (ToxR/ToxCAT constructs) have an influence on the association affinity.

1.6 Overview of the present work

From the theoretical point of view, to study the association of TM helices, we need a theoretical methodology for calculating the standard association free energy of TM helices. The core work in my Ph.D. research studies was the development of a

methodology for calculating the standard association free energy of TM helices in both micelles and bilayers based on an implicit membrane model (IMM1). In our methodology, the standard association free energy is decomposed into the effective energy, translational, rotational, and conformational entropy terms. The effective energy of association is obtained by molecular dynamics (MD) simulations in IMM1. The translational and rotational entropy of association are calculated according to the probability distribution of three translational and three rotational dimensions obtained from the MD simulations. The side-chain conformational entropy of association is estimated from the probability distribution obtained by rigid rotation of all side-chain dihedral angles.

Using this methodology we conducted three studies which led to three research papers. First we analyzed the energetics of inter-helical interactions of GpA using its solution and solid-state NMR structures (Mottamal et al. 2006). Second, we applied our methodology to calculate the association free energy of GpA and compared our calculations with the experimental values (Zhang and Lazaridis 2006). Third, we investigated the question why association affinities of GpA and bacteriophage M13 major coat protein (MCP) are greatly different although they both contain the GxxxG motif (Zhang and Lazaridis 2007).

Chapter II

THEORY

2.1 Choice of standard state

Because the free energy of a bimolecular reaction depends on the concentration of the reactants, association free energies are reported at given concentrations (the standard state). The standard state is implied in the units used to define the equilibrium constant. For GpA, sedimentation (Fleming et al. 1997; Fleming 2002; Fleming et al. 2004) and FRET (Fisher et al. 1999; 2003) data were used to obtain the relative amounts of dimer and monomer. The apparent association constant is defined as:

$$K_{app} = (n_{Dimer}/V_{Aq})/(n_{Monomer}/V_{Aq})^2 \quad (1)$$

where n_{Dimer} and $n_{Monomer}$ are numbers of moles of dimer and monomer species, respectively, and V_{Aq} is the volume of the aqueous solution. The standard state implied in equation (1) is 1M in bulk solution. The problem with apparent association constants is that they depend on the amount of detergent (Fleming 2002). To eliminate this problem the mole fraction equilibrium constant, K_X, was introduced, which is related to the apparent association constant by

$$K_X = K_{app} \times [\text{Detergent}] \quad (2)$$

where [Detergent] is the molar concentration of detergent in the bulk solution. The mole

fraction standard free energy of GpA upon association in C_8E_5 and C14 betaine was measured as -7.0 kcal/mol and -5.7 kcal/mol, respectively (Fleming 2002; Fleming et al. 2004). The apparent dissociation constants of GpA in DDMAB and DPC detergent micelles were determined as 0.08±0.04 μM and 0.16±0.08 μM, respectively, at 25mM detergent concentration (Fisher et al. 1999), from which the mole fraction standard association free energy of GpA in DDMAB and DPC is calculated as -7.1 kcal/mol and -7.6 kcal/mol, respectively.

Because molarities lend themselves more naturally to theoretical calculations, we adopt 1M as the standard state for our calculations, but it is defined only within the hydrophobic phase, micelles or bilayers, to eliminate the dependence of association constants on the amount of detergent or lipid. To convert the standard association free energy on the mole fraction scale to our own standard state, we start from the definition of association constants. K_X is defined as (Fleming 2002),

$$K_X = (X_{Dimer})/(X_{Monomer})^2 = (n_{Dimer}/n_{Total})/(n_{Monomer}/n_{Total})^2 \quad (3)$$

where X_i and n_i are the mole fraction and number of moles for protein species i in the hydrophobic phase, respectively, and n_{Total} is the total number of moles of all protein species and detergent/lipid molecules in the hydrophobic phase. At dilute conditions, one can make the approximation that $n_{Total} \approx n_{Detergent}$, where $n_{Detergent}$ is the number of moles of detergents/lipids in the hydrophobic phase. Then

$$K_X = (n_{Dimer}/n_{Detergent})/(n_{Monomer}/n_{Detergent})^2 \quad (4)$$

The molar association constant can be defined as,

$$K_C = [\text{Dimer}]/[\text{Monomer}]^2 = (n_{\text{Dimer}}/V_{HP})/(n_{\text{Monomer}}/V_{HP})^2 \quad (5)$$

where [i] is the molar concentration of species i in the hydrophobic phase and V_{HP} is the total volume of the hydrophobic phase including proteins and detergent/lipid molecules. A similar approximation that $V_{HP} \approx V_{\text{Detergent}}$, where $V_{\text{Detergent}}$ is the volume of the detergent/lipid molecules in the hydrophobic phase, can be made if the proteins are dilute in the hydrophobic phase; then

$$\begin{aligned} K_C &= (n_{\text{Dimer}}/V_{\text{Detergent}})/(n_{\text{Monomer}}/V_{\text{Detergent}})^2 \\ &= [n_{\text{Dimer}}/(n_{\text{Detergent}} \times \nu_{\text{Detergent}})]/[n_{\text{Monomer}}/(n_{\text{Detergent}} \times \nu_{\text{Detergent}})]^2 \\ &= [(n_{\text{Dimer}}/n_{\text{Detergent}})/(n_{\text{Monomer}}/n_{\text{Detergent}})^2] \times \nu_{\text{Detergent}} \end{aligned} \quad (6)$$

where $\nu_{\text{Detergent}}$ is the molar volume of hydrophobic tails of pure detergents/lipids. Combining equations (4) and (6) together, we obtain

$$K_C = K_X \times \nu_{\text{Detergent}} \quad (7)$$

The standard association free energy at 1M (in HP) standard state is:

$$\Delta G^0_{C,HP} = -RT \times \ln K_C = -RT \times \ln(Kx \times \nu_{\text{Detergent}}) = \Delta G^0_X - RT \times \ln(\nu_{\text{Detergent}}) \quad (8)$$

It was estimated that the molar volume of the –CH2-CH2- group is 32.2 cm^3/mol (Sesta 1989), thus the molar volume of hydrophobic tails is about 0.193 L/mol for DPC detergent micelles. Equation (8) then gives -6.1 kcal/mol for the 1M (in HP) standard association free energy of GpA in DPC micelles.

Since a membrane is a two-dimensional medium, the concentration of TM helices in a membrane bilayer would be more naturally expressed in units of mole/area. 1 mole/liter can be easily converted to mole/area if the thickness of hydrophobic core of membrane (T) is specified. 1M (in HP) corresponds to 1660 $Å^3$ per molecule. Therefore, the average area for each molecule is:

$$A = \frac{1660 Å^3}{T} \tag{9}$$

The surface concentration of TM helices ($CON_{surface}$, in units of mole/area) is:

$$CON_{surface} = 1 / (A \times N_a) \tag{10}$$

where N_a is Avogadro's number. For example, if the thickness of hydrophobic core of membrane is 23 Å, the corresponding area for each molecule of each species in membrane is 72.2 $Å^2$/molecule (note that this is similar to the area per lipid in common membranes); the surface concentration is 2.3×10^{-6} mole/m^2. Although surface concentration is more natural for the two-dimensional membrane, the molar concentration is more frequently used in thermodynamics, so 1 molar standard state in hydrophobic

phase is kept in our calculations.

2.2 Association in bilayers

Most previous theoretical calculations of association or binding free energy (Horton and Lewis 1992; Froloff et al. 1997; Gilson et al. 1997; Helms and Wade 1998; Lazaridis 2002; Lazaridis et al. 2002) were performed in bulk solution, which is isotropic in six dimensions (three rotational and three translational). However, in the IMM1 model, the x and y axes on the membrane plane are isotropic, while the z dimension is anisotropic. Among three rotational dimensions (three Euler angles) only one (about the z axis) is isotropic. Different formulas should be used to calculate the rotational and translational entropy lost on isotropic and anisotropic dimensions.

Since rotations and translations on the anisotropic dimensions are restricted to a relatively narrow range, we assume that a MD simulation (for example 1 ns) is sufficient for sampling distribution functions for rotation and translation on the anisotropic dimensions. This assumption was validated by a series of 1ns MD simulations of a GpA monomer which started from different initial positions on the z axis, different initial orientation about y axis and helical axis, respectively, but obtained very similar distributions on these dimensions (data not shown). Therefore, the following basis is proposed for our entropy calculations:

(i) Monomer and dimer translate and rotate freely on the isotropic dimensions and their translational and rotational entropy on these dimensions can be calculated analytically.

(ii) The entropy loss upon association on the isotropic dimensions is contributed by

only one helix. Upon association, one helix still samples the available space specified by the standard state but the second helix has limited freedom to move relative to the other.

(iii) On the anisotropic dimensions the entropy changes for each helix from free to bound state are calculated separately and added.

The translational and rotational entropy loss on the isotropic dimensions can be calculated based on the relative distances or relative Euler angles of one helix with respect to the other as described for the binding process in solution (Lazaridis et al. 2002). For the entropic contribution from anisotropic dimensions, the calculation method should be modified.

As in the calculation of binding free energies in solution (Lazaridis et al. 2002), the standard association free energy ($\Delta G^{0}_{C,HP}$) can be decomposed into the effective energy change ($\Delta W_{association}$), the translational entropy loss ($\Delta S^{translational}_{association}$), the rotational entropy loss ($\Delta S^{rotational}_{association}$), and the conformational entropy loss ($\Delta S^{conformational}_{association}$).

$$\Delta G^{0}_{C,HP} = \Delta W_{association} - T\Delta S^{translational}_{association} - T\Delta S^{rotational}_{association} - T\Delta S^{conformational}_{association} \quad (11)$$

Calculations of these components are discussed in the following sections.

2.2.1 Effective Energy Change upon Association ($\Delta W_{association}$)

Effective energy change upon association is the average effective energy of the dimer ($\overline{W}_{\text{dimer}}$) minus the average effective energy of the two monomers.

$$\Delta W_{association} = \overline{W}_{\text{dimer}} - \overline{W}_{\text{monomer A}} - \overline{W}_{\text{monomer B}} \quad (12)$$

Another expression for $\Delta W_{association}$ can be derived from equation (12) (Lazaridis et al. 2002):

$$\begin{aligned} \Delta W_{association} &= \Delta E^{A} + \Delta E^{B} + E^{\text{inte}} + \Delta\Delta G^{slvA} + \Delta\Delta G^{slvB} \\ &= \Delta W^{A} + \Delta W^{B} + W^{\text{inte}} \end{aligned} \quad (13)$$

where E^{inte} is the inter-helical interaction energy, W^{inte} is the inter-helical effective interaction energy, ΔE^{A} and ΔE^{B} are the changes in intramolecular energy, $\Delta\Delta G^{slvA}$ and $\Delta\Delta G^{slvB}$ are solvation free energy changes, and ΔW^{A} and ΔW^{B} are the reorganization energy of A and B, respectively (Lazaridis et al. 2002). The difference between energies and effective energies is that the latter include solvation effects. That is, W^{inte} is equal to E^{inte} plus the loss of solvation of each helix due to the other helix. ΔW^{A} is equal to ΔE^{A} plus the change in solvation of each atom in helix A due to other atoms in helix A.

2.2.2 Translational Entropy Loss upon Association ($\Delta S^{translational}_{association}$)

The translational entropy loss upon association in lipid bilayers is mostly due to restriction of motion on the plane of the membrane, i.e., isotropic dimensions x and y. For dimerization it is more natural to use polar coordinates (r, θ), where r is the distance between the centers of the helices and θ is the angle between the vector connecting the

centers of mass of the two helices and the vector from the center of reference helix to one of the Cα atoms on the reference helix (figure 1).

Preliminary tests showed that distributions of r and θ are highly coupled, therefore in the calculations we use the coupled distribution of r and θ (equation 14). The loss of translational entropy on the membrane plane is due to: a) reduction in amplitude of r and θ, $\Delta S_1^{r,\theta}$; b) the uneven distribution of r and θ within the allowed range, $\Delta S_2^{r,\theta}$; c) the change of "communal entropy", shown by the first, second and third terms on the right hand side of equation (14), respectively.

$$\Delta S_{trans}^{r,\theta} = \Delta S_1^{r,\theta} + \Delta S_2^{r,\theta} + \Delta S_{communal}^{r,\theta}$$

$$= R\ln\frac{\Delta A}{\pi \cdot R_A^2} - R\{\int_{\Delta A} p(r,\theta)\cdot \ln p(r,\theta)\cdot r\cdot dr\cdot d\theta - \int_{\Delta A}(1/\Delta A)\cdot \ln(1/\Delta A)\cdot r\cdot dr\cdot d\theta\} - R$$

$$= R\ln\frac{\Delta A}{\pi \cdot R_A^2} - R\{\int_{\Delta A} p(r,\theta)\cdot \ln p(r,\theta)\cdot r\cdot dr\cdot d\theta + \ln\Delta A\} - R \qquad (14)$$

where R is the gas constant, p(r, θ) is the normalized probability distribution of r and θ ($\int_{\Delta A} p(r,\theta)\cdot r\cdot dr\cdot d\theta = 1$), R_A is the radius of the average area A for each monomer at 1 molar standard state ($R_A = \sqrt{\frac{A}{\pi}}$) and ΔA is the surface area ($\Delta A = \int r\cdot dr\cdot d\theta$) in which one helix is observed to move relative to the other helix. p(r, θ)=1/ΔA corresponds to a flat distribution. The first and second terms in equation 14 are essentially from a "cell theory" approach; each molecule is allowed to move within a "cell", the size of which is determined by the standard state. It neglects the possibility of multiple occupancy of the cells. The latter contributes a term referred to as "communal entropy" and is equal to R

(Henchman 2003). When 2 moles of monomers form one mole of dimer, this term contributes –R to the translational entropy loss upon association. This term was incorrectly omitted in previous work in bulk solution (Lazaridis et al. 2002).

Translational entropy loss upon association in lipid bilayers may also occur along the z axis, an anisotropic dimension. Preliminary tests also showed coupling of absolute z coordinates of the two helices, therefore the joint distribution was used to calculate the translational entropy lost on the z axis. The translational entropy change upon association on the z axis is the difference in translational entropy of two coupled helices after association and translational entropy of two independent monomers on z axis,

$$\Delta S_{trans}{}^{z} = S_{dimer}{}^{z} - S_{monomer\ A}{}^{z} - S_{monomer\ B}{}^{z}$$

$$= -R \{ \int_{\Delta Z_A^d, \Delta Z_B^d} p(z_A^d, z_B^d) \cdot \ln p(z_A^d, z_B^d) \cdot dz_A^d \cdot dz_B^d \}$$

$$+ R \{ \int_{\Delta Z_A^m} p(z_A^m) \cdot \ln p(z_A^m) \cdot dz_A^m \} + R \{ \int_{\Delta Z_B^m} p(z_B^m) \cdot \ln p(z_B^m) \cdot dz_B^m \} \qquad (15)$$

where all probability distributions are normalized ($\int_{\Delta z_A^m} p(z_A^m) dz_A^m = \int_{\Delta z_B^m} p(z_B^m) dz_B^m = \int p(z_A^d, z_B^d) \cdot dz_A^d \cdot dz_B^d = 1$); A, and B represent helices A and B; d and m denote the helix in a dimer or as a monomer, respectively. In equation (15) Δz_A^d, Δz_A^m, Δz_B^d, and Δz_B^m are the amplitudes of z_A^d, z_A^m, z_B^d, and z_B^m, respectively, obtained from MD simulations, z_A^d, z_A^m, z_B^d, and z_B^m, are absolute z coordinates of helix A in a dimer, helix A as a monomer, helix B in a dimer, and helix B as a monomer, respectively. For a homodimer, the last two terms are identical.

The total translational entropy loss upon association in bilayers is:

$$\Delta S_{bilayers}^{translational} = \Delta S_{trans}^{r,\theta} + \Delta S_{trans}^{z} \quad (16)$$

2.2.3 Rotational Entropy Loss upon Association ($\Delta S_{association}^{rotational}$)

Any orientation of a rod-like TM helix can be defined by a unique combination of three Euler angles. In this work we define the three angles as follows. Starting from a reference where the helix lies along the x axis (see caption of Fig. 2), any orientation can be obtained by a three-step operation (shown in figure 2): a) rotate about the x axis (the helical axis) by an angle α (i.e., α is the rotation angle); b) rotate about the y axis by an angle 90°-β (i.e., β is the tilt angle); c) rotate about the z axis by an angle γ. Thus, to determine the Euler angles of any given orientation of a helix during MD simulations one needs to find the reverse operations to return the orientation back to the reference state.

In aqueous solution it was assumed that the probability distributions of all three Euler angles are independent (i.e., $p(\alpha, \beta, \gamma) = p(\alpha)p(\beta)p(\gamma)$) (Lazaridis et al. 2002). This assumption is likely to fail in membranes; for example, for β=0, all α angles are equally probable, but for β≠0 they are not. Indeed, preliminary tests showed that the distributions of α and β of a helix are coupled with each other. Thus in the membrane it can be assumed that $p(\alpha, \beta, \gamma) = p(\alpha, \beta)\cdot p_\gamma$ for a monomer (p_γ is constant since γ is isotropic). The dimer can rotate isotropically around the z axis. We can think of one helix maintaining its full freedom in the γ angle, but the γ angle of the second helix is now correlated with that of the first helix. The dimer is then described approximately by the probability distribution:

$$p(\alpha_A^d, \beta_A^d, \alpha_B^d, \beta_B^d, \gamma) = p(\alpha_A^d, \beta_A^d, \alpha_B^d, \beta_B^d) \cdot p(\gamma) \quad (17)$$

where $\alpha_A^d, \beta_A^d, \alpha_B^d, \beta_B^d, \gamma$ are the rotational angles of each helix in a dimer from the reference state and $\gamma = \gamma_B - \gamma_A$. Therefore the rotational entropy change is decomposed into two contributions: one from the γ angle and the other from the α and β angles.

Rotational entropy loss in bilayers from the γ angle is given by

$$\Delta S_{rot}^{\gamma} = - R \{ \int_0^{360} p(\gamma) \cdot \ln p(\gamma) \cdot d\gamma \; - \; \int_0^{360} p_\gamma \cdot \ln p_\gamma \cdot d\gamma \} \quad (18)$$

where $\int_0^{360} p(\gamma) d\gamma = 1$, and $\int_0^{360} p_\gamma \cdot \ln p_\gamma \cdot d\gamma$ is the rotational entropy corresponding to free rotation of a monomer (p_γ is a constant value determined from normalization $\int_0^{360} p_\gamma d\gamma = 1$).

Rotational entropy loss in bilayers from angles α and β, anisotropic dimensions, occurs for both helices, shown by equations (19)-(22):

$$\Delta S_{rot}^{\alpha,\beta} = S_{dimer}^{\alpha,\beta} - S_{monomer\ A}^{\alpha,\beta} - S_{monomer\ B}^{\alpha,\beta} \quad (19)$$

$$S_{dimer}^{\alpha,\beta} = - R \{ \int p(\alpha_A^d, \beta_A^d, \alpha_B^d, \beta_B^d) \cdot \ln p(\alpha_A^d, \beta_A^d, \alpha_B^d, \beta_B^d) \cdot \sin\beta_A^d \cdot \sin\beta_B^d \cdot d\alpha_A^d \cdot d\beta_A^d \cdot d\alpha_B^d \cdot d\beta_B^d \} \quad (20)$$

$$S_{monomer\ A}^{\alpha,\beta} = - R \{ \int p(\alpha_A^m, \beta_A^m) \cdot \ln p(\alpha_A^m, \beta_A^m) \cdot \sin\beta_A^m \cdot d\alpha_A^m \cdot d\beta_A^m \} \quad (21)$$

$$S_{monomer\ B}^{\alpha,\beta} = - R \{ \int p(\alpha_B^m, \beta_B^m) \cdot \ln p(\alpha_B^m, \beta_B^m) \cdot \sin\beta_B^m \cdot d\alpha_B^m \cdot d\beta_B^m \} \quad (22)$$

The total rotational entropy loss upon association in lipid bilayers is:

$$\Delta S_{bilayers}^{rotational} = \Delta S_{rot}^{\gamma} + \Delta S_{rot}^{\alpha,\beta} \quad (23)$$

2.2.4 Conformational entropy loss upon association ($\Delta S_{association}^{conformational}$)

The entropy loss due to restriction of side-chain dihedral angles upon association is calculated by exhaustive enumeration, i.e., sampling each side-chain dihedral angle separately keeping the backbone atoms and all the other dihedral angles fixed to calculate the effective energy of different conformations, and then calculating the entropy loss of each dihedral angle according to the probability distribution determined from the effective energies. The total side-chain entropy of a TM helix is assumed to be the sum of contributions from each dihedral angle,

$$S_{helix}^{conformational} = \sum_j \sum_i p_i \ln p_i \quad \left(p_i = \frac{Exp(-W_i / kT)}{\sum_i Exp(-W_i / kT)} \right) \quad (24)$$

where p_i is the probability of conformation i of dihedral angle j and Wi is the effective energy for that conformation. Thus the side-chain conformational entropy change upon association is:

$$\Delta S_{association}^{conformational} = S_{Dimer}^{conformational} - S_{MonomerA}^{conformational} - S_{MonomerB}^{conformational} \quad (25)$$

This method is approximate because it neglects correlations between dihedral angles

(the computational cost prevents full enumeration of all dihedral angle combinations). In test calculations where we allowed simultaneous variation of two dihedral angles from neighboring sidechains in the dimer, we found that the error due to neglect of correlations is very small (less than 1%). Secondly, there is an inconsistency in the fact that the average effective energy and the sidechain entropy are obtained based on different conformational ensembles (one from MD, the other by enumeration). However, this is not expected to have a large effect since the energies of different rotamers are rather similar.

2.3 Association in micelles

Detergent micelles and lipid bilayers are different in two aspects: a) the shape of the hydrophobic environment is different: the lipid bilayers are roughly a flat slab, while the micelles are spherical or elliptical. Given that the helices are fully immersed in the hydrophobic phase, we assume that the effective energy change upon association of TM helices is the same in bilayers and micelles. b) The association entropy of a TM helix in a micelle and a lipid bilayer is different. Before association the orientation of two monomers is already constrained in the lipid bilayer (figure 3A), whereas any orientation is allowed in a micelle (figure 3B). The translational entropy is also different. Bilayers provide for movement in a continuous, two-dimensional medium, while micelles are essentially zero-dimensional media. The movement of the micelle in solution does not contribute to the entropy of the peptide embedded in it. Only the movement of the peptide with respect to the micelle contributes. The translational entropy loss upon association is the sum of two terms: one arising from the overall distribution of helices in micelles and one from the local "vibrations" of the helices within a micelle. The rotational entropy loss

occurs because a monomer can rotate freely within a micelle but two helices in a dimer rotate together. Obviously, association in micelles is very complex and may contain the contributions from the detergent itself (if, for example, the aggregation number changes upon association). Here we will perform a very basic calculation under the following assumptions: a) micelles are ideal, spherical objects b) each micelle cannot contain more than one monomer or dimer.

The translational entropy loss upon association due to the distribution in micelles is calculated in the following way. Let C_P be the standard concentration (molarity in the hydrophobic phase). The number of micelles per monomer or dimer at the standard concentration is:

$$N_m = \frac{1}{\nu_{\text{Detergent}} \times N_{\text{aggregation}} \times C_P} \tag{26}$$

where $\nu_{\text{Detergent}}$ is the molar volume of hydrophobic tails of pure detergents, which is about 0.193 L/mol for DPC, and $N_{\text{aggregation}}$ is the aggregation number of detergent micelles. Although the aggregation numbers of detergent micelles and protein-detergent micelles can be dramatically different (Sanders and Sonnichsen 2006), Nm mainly depends on the aggregation number of detergent micelles ($N_{\text{aggregation}}$) because usually the proteins are highly diluted in experiments and thus the number of protein-detergent micelles is much smaller than that of detergent micelles. The aggregation number of DPC micelles is about 50~60 (le Maire et al. 2000); we used 55. For standard concentration 1M (in HP), N_m is less than 1, which means 1M (in HP) standard state is not accessible in practice if each micelle can contain only one monomer or dimer. Therefore, 1mM (in HP)

state is used and the result is corrected by RTln(1000). At 1mM (in HP) standard state Nm is about 94. The number of states (Ω) for n indistinguishable protein molecules of the same kind in nNm micelles is:

$$\Omega = \frac{nNm \cdot (nNm-1)\cdots(nNm-n+1)}{n\cdot(n-1)\cdots 1} = \frac{(nNm)!}{(nNm-n)!\cdot n!} \qquad (27)$$

If n is a very large number, using Stirling's approximation ($\ln n! \cong n \ln n - n$) gives

$$\ln\Omega \cong n\left\{\ln\left[\frac{Nm^{Nm}}{(Nm-1)^{Nm-1}}\right] + \ln Nm\right\}$$

$$\cong n(\ln Nm + 1) \quad \text{(the limit of } \ln\left[\frac{Nm^{Nm}}{(Nm-1)^{Nm-1}}\right] \text{ for large Nm is 1)} \qquad (28)$$

The 1 in Eq. 28 corresponds to the "communal entropy" discussed above. For the non-covalent association process, 2Monomers ↔ Dimer, the standard free energy change upon association can be defined as the free energy change for 2 moles of monomer at 1 mM to convert to 1 mole of dimer at 1mM. Accordingly, the translational entropy loss upon association due to the distribution in micelles should be

$$\Delta S_{trans}^{state} = S_{dimer}^{1\ mole} - S_{monomer}^{2\ mole}$$

$$= N_A k_B \ln\Omega - 2 N_A k_B \ln\Omega$$

$$= -R(\ln Nm + 1) \qquad (29)$$

where N_A is the Avogadro's number, and k_B is Boltzmann's constant. Thus the translational entropy loss due to the decrease of the number of translation states contributes ~3.3 kcal/mol to the association free energy at 1mM (in HP) standard state, which corresponds to ~ -0.8 kcal/mol, at 1M (in HP) standard state.

The translational entropy change due to the constraints inside the micelles includes contributions from the x-y plane ($\Delta S_{trans}^{r,\theta}$) and the z axis ($\Delta S_{trans}^{z}$, which is assumed to be the same as in bilayers, see equation 15). $\Delta S_{trans}^{r,\theta}$ is equal to the entropy due to the translation of a dimer inside the micelle on the x-y plane ($S_{dimer}^{r,\theta}$), plus the entropy due to the relative translation of one helix with respect to the other in the dimer on the x-y plane ($S_{relative}^{r,\theta}$), less the entropy due to the translation of monomers A and B inside the micelle on the x-y plane ($S_{monomer\ A}^{r,\theta}$ and $S_{monomer\ B}^{r,\theta}$).

$$\begin{aligned}\Delta S_{trans}^{r,\theta} &= S_{dimer}^{r,\theta} + S_{relative}^{r,\theta} - S_{monomer\ A}^{r,\theta} - S_{monomer\ B}^{r,\theta} \\ &= -R\ \ 2\pi \cdot \int_{\Delta r^d} p(r^d) \cdot \ln p(r^d) \cdot r^d \cdot dr^d \ \ - R \int_{\Delta A} p(r,\theta) \cdot \ln p(r,\theta) \cdot r \cdot dr \cdot d\theta \\ &\quad + R\ \ 2\pi \cdot \int_{\Delta r_A^m} p(r_A^m) \cdot \ln p(r_A^m) \cdot r_A^m \cdot dr_A^m \\ &\quad + R\, 2\pi \cdot \int_{\Delta r_B^m} p(r_B^m) \cdot \ln p(r_B^m) \cdot r_B^m \cdot dr_B^m \end{aligned} \qquad (30)$$

where $\Delta r^d, \Delta r_A^m$, and Δr_B^m is the amplitude of r^d, r_A^m, and r_B^m, respectively. r^d, r_A^m, and r_B^m is the distance of the dimer or monomer from the center of the micelle. $p(r^d)$, $p(r_A^m)$, and $p(r_B^m)$ are calculated theoretically according to the energy distribution. ΔA, r, θ, and $p(r,\theta)$ are the same as those in equation (14). As shown in figure 4, the hydrophobic burial (T) of a monomer or dimer changes with the distance from the center

of the micelle (r^m). T can be calculated by T=2$\cdot\sqrt{(11.5)^2-(r^m)^2}$, assuming the helix is an ideal cylinder, the hydrophobic core of detergent micelle is an ideal sphere, and the helical axis is parallel to the z-axis. So we calculate the effective energy (W) as a function of T using IMM1 and obtain p(r^m) as,

$$p(r^m)=\frac{e^{-W(r^m)/kT}}{\int_0^{11.5} e^{-W(r^m)/kT}} \tag{31}$$

Since a 23 Å thickness hydrophobic slab of IMM1 model is used to mimic the hydrophobic environment of micelles when the protein is at the center of the micelle, r^m varies from 0 to 11.5 Å and Δr^m is 11.5 Å. $p(r^d)$, the distribution of a dimer inside the micelle can be calculated similarly.

Thus the total translational entropy change upon association in micelles is:

$$\Delta S_{micelles}^{translational} = \Delta S_{trans}^{state} + \Delta S_{trans}^{r,\theta} + \Delta S_{trans}^{z} \tag{32}$$

The method to calculate rotational entropy loss upon GpA association in micelles is the same as that upon binding in water (Lazaridis et al. 2002).

$$\Delta S_{micelles}^{rotational} = -R\left\{\int_0^{360} p(\alpha)\cdot\ln p(\alpha)\cdot d\alpha - \int_0^{360} p_\alpha\cdot\ln p_\alpha\cdot d\alpha\right\}$$

$$-R\left\{\int_0^{180} p(\beta)\cdot\ln p(\beta)\cdot\sin\beta\cdot d\beta - \int_0^{180} p_\beta\cdot\ln p_\beta\cdot\sin\beta\cdot d\beta\right\}$$

$$- R \{ \int_0^{360} p(\gamma) \cdot \ln p(\gamma) \cdot d\gamma - \int_0^{360} p_\gamma \cdot \ln p_\gamma \cdot d\gamma \} \quad (33)$$

where α, β, and γ are relative Euler angles of one helix with respect to the other helix (reference helix) in the dimer and $\int_0^{360} p_\alpha \cdot \ln p_\alpha \cdot d\alpha$, $\int_0^{180} p_\beta \cdot \ln p_\beta \cdot \sin(\beta) \cdot d\beta$, and $\int_0^{360} p_\gamma \cdot \ln p_\gamma \cdot d\gamma$ are the rotational entropy on α, β, and γ dimensions, respectively, corresponding to the free rotation state of a monomer (p_α, p_β, and p_γ are constants determined from normalization $\int_0^{360} p_\alpha d\alpha = \int_0^{180} p_\beta \sin\beta d\beta = \int_0^{360} p_\gamma d\gamma =$ 1). p(α), p(β), and p(γ) are calculated from the bilayer simulations using IMM1.

Since the same effective energy function was used for both bilayers and micelles, we assume that side-chain conformational entropy change upon association is the same in bilayers and micelles.

Chapter III

ENERGETIC ANALYSIS OF GLYCOPHORIN A DIMER

3.1 Methods

3.1.1 IMM1 model

IMM1 is an implicit membrane model derived from the implicit aqueous model EEF1 (Lazaridis and Karplus 1999). It gives the effective energy (W_{IMM1}) of a protein in a heterogeneous membrane-water system as the sum of the intramolecular energy (E) of the protein and the solvation free energy (ΔG^{slv}) (Lazaridis 2003).

$$W_{IMM1} = E + \Delta G^{slv} \tag{34}$$

The intramolecular energy is calculated by the CHARMM 19 polar hydrogen energy function and the solvation energy is the sum of contributions from each atom or group i.

$$\Delta G^{slv} = \sum_i \Delta G_i^{slv} = \sum_i \Delta G_i^{ref} - \sum_i \sum_{j \neq i} f_i(r_{ij}) V_j \tag{35}$$

where ΔG_i^{ref} is the solvation free energy of group i in a small, model compound and the last term is the solvation free energy lost due to exclusion of solvent by surrounding atoms. One essential difference between EEF1 and IMM1 model is the definition of reference solvation free energy. In IMM1 the reference solvation free energy depends on the position of each atom with respect to the membrane. The values in the interior of the membrane are obtained from solvation data in cyclohexane.

$$\Delta G_i^{ref}(z') = f(z')\,\Delta G_i^{ref.water} + (1- f(z'))\,\Delta G_i^{ref.cyclohexane} \quad (36)$$

where $z'=|z|/(T/2)$ (T denotes the thickness of hydrophobic core of the membrane) and f(z') is defined by:

$$f(z') = \frac{z'^n}{1+z'^n} \quad (37)$$

n is set equal to 10, which gives the appropriate steepness of transition between nonpolar and polar environments (Lazaridis 2003). f=0.5 corresponds to the hydrocarbon-polar headgroup interface.

A second difference between EEF1 and IMM1 model is that a modified dielectric screening function is used to calculate the electrostatic interaction in the membrane.

$$\varepsilon = r^{fij} \quad (38)$$

where f_{ij} reflects the positions of the two interacting groups i and j with respect to the membrane and can be calculated by the empirical model

$$f_{ij} = 0.85 + 0.15\sqrt{f_i f_j} \quad (39)$$

where f_i and f_j are defined by equation (37).

3.1.2 Dimer structures and energetic analysis

All calculations were carried out using the program CHARMM (Brooks et al. 1983). The main effective energy function we use is IMM1 (Lazaridis 2003). Two dimer structures of GpA were used in our energetic analysis: one obtained by solution NMR in detergent (pdb code: 1AFO) (MacKenzie et al. 1997) and the other by solid-state NMR in lipid bilayers (Smith et al. 2001). The backbone root-mean-square deviation (RMSD) of the 20 solution state NMR models from the solid-state NMR structure ranges from 0.87 to 1.24 Å, with an average of 1.1 Å. For the solution NMR structure, we selected the first of 20 models. The backbone RMSD of this model from the others ranges from 0.48 to 1.20 Å, with an average of 0.74 Å. For both structures, residues 73 through 95 were included in the calculations. The initial structures were oriented with their principal axis perpendicular to the membrane surface, and then energy minimized for 800 steps by the Adopted Basis Newton-Raphson (ABNR) method. Subsequently, 1-ns MD simulations starting from the minimized structures were performed with harmonic constraints on all the backbone atoms (force constant 1.0 kcal/mol/$Å^2$) at the temperature of 298.15 K, to examine the effect of thermal motion while remaining close to the experimental structure.

3.2 Results

The minimized and average total and inter-helical interaction energies of the solid-state and solution NMR GpA structures are presented in Table 1. The solid-state NMR structure has a stronger inter-helical interaction probably due to two reasons. First, the OH group of Thr87 makes an intrahelical hydrogen bond with the backbone in the

solution NMR structure but an inter-helical hydrogen bond in the solid-state NMR structure (during the MD simulation, Thr87 rapidly reverts to an intrahelical hydrogen bond). Second, the solid-state NMR structure has a smaller crossing angle. The largest overall contribution to the inter-helical energy is the van der Waals interaction. The desolvation energy contribution is positive, reflecting the loss of protein – lipid interactions upon dimer formation. The electrostatic contribution is very small.

For a more detailed attribution of association energy, the inter-helical interactions were decomposed into contributions between each pair of inter-helical residues. A similar decomposition has been performed previously but without the desolvation contribution (Petrache et al. 2000). The matrices of interaction pairs are shown in Tables 2 and 3 for the solution and solid-state NMR structures, respectively. In most residue pairs, van der Waals interactions make the largest contribution and electrostatic interactions the smallest. The desolvation energies are always positive, except in one case where the pair is just outside the membrane. There are some differences in these two matrices. (1) the largest distance in sequence for two inter-helical residues interacting strongly is three residues in the solution NMR structure and four residues in the solid-state NMR structure. (2) In the solution NMR structure, the first residue with strong interaction with a residue on the other helix is Leu75 and the last one Ile88, but in the solid-state NMR structure, they are Ile73 and Ile91, respectively. That is because the solid-state NMR structure has a longer interface between two helices and a smaller crossing angle than the solution NMR structure. (3) The solid-state NMR structure has a positive interaction between two Thr87 residues due to an unfavorable dipole – dipole interaction between the OH groups.

The inter-helical residue-residue interaction energies shown in Tables 2 and 3

account for about 85.0% of the total inter-helical interaction energy. Inter-helical interactions from the GXXXG motif are -6.5 kcal/mol in both minimized structures, which corresponds to about 30% of the total inter-helical interaction. The van der Waals, electrostatic, and solvation contributions are -8.2, -1.0, and +2.7 kcal/mol, respectively. Thus, there are some important interactions outside this motif, such as Ile76: Leu75 and Val84: Gly83. The 6 Cα-H···O hydrogen bonds in GpA contribute -6.0 kcal/mol to stability, largely due to favorable electrostatic interactions (Mottamal and Lazaridis 2005). However, the overall electrostatic interaction in GpA is close to zero, which means that there must exist repulsive interactions that cancel out the hydrogen bond contribution. In both NMR structures, the electrostatic interactions between the inter-helical residues Gly79 and Gly79 and Gly83 and Gly83 are repulsive (1.5 kcal/mol). In the solid-state structure, the electrostatic interaction between the inter-helical residues Thr87 & Thr87 is also positive (1.3 kcal/mol) mainly due to the interactions between the hydroxyl group atoms. Other weak repulsive electrostatic interactions are between Val80 and Val80 and Ile76 and Ile76. The results in Tables 2 and 3 are consistent with recent mutational studies that show that interactions outside the GXXXG motif are important for helix association (Doura et al. 2004; Schneider and Engelman 2004). For example, a mutant that lacks the GXXXG motif can still dimerize, while a mutant containing the GXXXG motif failed to dimerize (Doura and Fleming 2004).

3.3 Discussion

Energetic analysis of two NMR structures of GpA dimer showed stronger interaction energy for the solid state structure due to differences in T87 hydrogen bonding and the

crossing angles. However, the total energies were similar. The largest contribution to the inter-helical energy was the van der Waals interaction. A large contribution to the inter-helical energy came from the GXXXG motif. However, two-thirds of the interaction energy came from outside the GXXXG motif. Thus, non-GXXXG residues are also important for the stabilization of the helix dimer, in agreement with recent mutational studies on GpA (Doura et al. 2004; Schneider and Engelman 2004).

The folding of GpA is believed to be driven by packing interactions at the dimer interface and to be largely independent of electrostatic forces (Lemmon et al. 1992b; Kim et al. 2003). Indeed, the overall contribution of electrostatic interactions to the inter-helical energy in the native state is very small. However, this small value contains larger compensating favorable and unfavorable interactions. The favorable interactions come from the Cα-H···O hydrogen bonds and the unfavorable ones from symmetric L76-L76, G79-G79, V80-V80, and G83-G83 interactions. In addition, in other systems polar interactions are known to be the major driving force for TM helix association (Choma et al. 2000; Zhou et al. 2001; Zhang and Lazaridis 2006). Therefore, neglect of electrostatic interactions cannot be a generally viable strategy for structure prediction.

Chapter IV

CALCULATING THE FREE ENERGY OF ASSOCIATION OF GLYCOPHORIN A

4.1 Methods

4.1.1 MD simulations

Two structures of Glycophorin A are available: one was determined by solution NMR in DPC micelles (PDB code: 1AFO) (MacKenzie et al. 1997); the other was obtained by solid-state NMR in DMPC bilayers (Smith et al. 2001). Here, the initial structures were obtained from the solid-state NMR structure of GpA. 29 residues (from GLU70 to LEU98) were modeled. The N and C termini were blocked by the acetyl (ACE) and the methyl amide (CBX) groups, respectively. Before MD simulations, the structures were energy-minimized by the Adopted Basis Newton-Raphson (ABNR) method. The nuclear overhauser effect (NOE) constraints between hydrogen bonded O and N atoms were used to reduce statistical fluctuations and to prevent the changes in the extramembranous portions observed in previous work (Lazaridis 2003) (A 5-ns simulation of GpA without NOE constraints produced a ~0.5 Å RMSD in the inter-helical interface, from Leu75 to Thr87, with all crucial inter-helical contacts maintained). 23.0 Å was used as the hydrophobic thickness in IMM1 because that is close to the hydrophobic thickness of a DMPC bilayer (de Planque and Killian 2003). All simulations were conducted with the CHARMM package and the Verlet integrator. The temperature of all simulations was set to 298.15 K at which the experimental measurements were normally conducted. During simulations, the average temperature after 0.1 ns was 298.0 K and 299.2 K for the

monomer and the dimer, respectively.

4.1.2 Association free energy calculations

The effective energies during the last 0.9 ns of several 1 ns MD simulations of GpA monomer and dimer in the membrane were averaged and used to calculate the effective energy change upon association. Since NOE constraints generate an artificial additional energy term, this energy term was removed in our effective energy calculations. The magnitude of this term for the first run was 10 kcal/mol, 4.3 kcal/mol, and 1.4 kcal/mol for the dimer, the monomer, and the effective energy change upon association, respectively.

The following steps were followed for the translational entropy calculations:

(i) Run MD simulations of the dimer and the monomers, and save 1000 coordinate frames in the each trajectory.

(ii) Calculate the coordinates of the center of 21 α carbon atoms on residues from THR74 to GLY94 with CHARMM command RGYR.

(iii) Compute a histogram for each center of mass coordinate using 0.4 Å intervals, normalize it, and calculate the translational entropy loss according to equations (14)-(16) in bilayers or equations (26)-(32) in micelles.

The detailed protocol for the rotational entropy calculations is the following:

(i) For each coordinate frame calculate the coordinates of Point 1 (the center of 21 α carbon atoms on residues from THR74 to GLY94), Point 2 (the center of 7 α carbon atoms on residues from THR74 to VAL80) and Point 3 (the α carbon atom on residue VAL84) on the helix of interest with CHARMM command RGYR.

(ii) Determine the angles for the helix rotated about the z (γ) and y axis (β) from the vector of Point 1 (x1, y1, z1) → Point 2 (x2, y2, z2).

$\gamma = \tan^{-1}((y2-y1)/(x2-x1))$

$90^{o}-\beta = \sin^{-1}((z2-z1)/$ distance between Point 1 and 2)

(iii) There are several steps to find α. First, move the whole helix to bring Point 1 to the origin; second, rotate Point 3 about the z-axis by an angle $-\gamma$; third, rotate Point 3 about the y-axis by an angle $-(90°-\beta)$; finally, α can be calculated from the new coordinates of Point 3 (x3', y3', z3').

$\alpha = \tan^{-1}(z3'/ y3')$

(iv) Calculate the range of each Euler angle; count the number of structures in each 5° (β) or 18° (α or γ) interval in that range; normalize the probability distribution in that range; calculate the rotational entropy loss according to equations (18)-(23) in bilayers or equation (33) in micelles.

The side-chain conformational entropy was calculated by systematic sampling of each dihedral angle. Since the effective energy is sensitive to the translational and rotational configuration of a TM helix (translation on z axis, tilt angle to membrane normal, and rotational angle about the helical axis), side-chain entropies were averaged over 100 frames (after energy minimization) from MD trajectories. For each frame, the side-chain conformational entropy of the monomer or the dimer was calculated according to equations (24). The side-chain conformational entropy loss was calculated by equation (25) using the average side-chain entropies.

4.2 Results

4.2.1 Structural stability of GpA during the MD simulations

Seven 1-ns MD simulations were performed starting with a different random number for the assignment of velocities. The RMSD of the backbone atoms of monomers and dimer with respect to their initial structures for run #1 is shown in figure 5. The RMSD of the dimer for most of the simulation time of run #1 is less than 1.6 Å, which demonstrates that the dimer is sufficiently stable during the simulations. A 5-ns simulation of GpA without NOE constraints was also performed, which produced a ~0.5 Å RMSD in the crucial part for association, from Leu75 to Thr87, although the flanking residues fluctuated considerably.

4.2.2 Effective energy changes of GpA upon association in bilayers

MD simulations of 29-residue GpA monomers, and the dimer were performed separately to calculate the effective energy change (ΔW) upon association. The results are presented in Table 4. The average effective energies of monomer A and monomer B over several runs are close to each other, as they should. The average effective energy change is -14.9±1.8 kcal/mol, which is much smaller than the average inter-helical effective interaction energy of ~-23 kcal/mol. The difference is accounted for by a change in intrahelical energy (reorganization energy) (Lazaridis et al. 2002). For run #1, the average reorganization energy for the two helices is 7.9 kcal/mol and can be decomposed into 8.9 kcal/mol from van der Waals, 0.3 kcal/mol from electrostatics, -1.6 kcal/mol from solvation, and 0.3 kcal/mol from bonded terms. The contribution of each residue to the reorganization energy and its van der Waals, electrostatics, solvation and bonded energy components were calculated; the largest contributions are shown in Table 5. Residues

73-76 and 91-92 contribute large solvation terms to the reorganization energy probably because they are around the hydrophobic-hydrophilic boundary and solvation terms are very sensitive to environment changes that take place upon association. Figure 6 shows the configuration of the monomer and dimer in the membrane. The tilt of each helix remains the same upon association, but the position on the z-axis of each helix shifts 1.1 Å towards the C-terminal direction and its rotation is different, which causes effective energy changes for residues 73-76 and 89-92 after association. The large contribution of the van der Waals term to the reorganization energy is probably caused by changes in side-chain conformations. Indeed, the average side-chain conformation of all major contributors to the van der Waals term except LEU89 changes upon association.

4.2.3 Translational entropy loss of GpA upon association in bilayers

The translational and rotational entropy losses of GpA upon association in bilayers were calculated from MD run #1. The contribution of translation entropy change along the x-y plane to the standard free energy is 1.9 kcal/mol. The entropy changes from the x-y plane due to the change of amplitude, uneven distribution and the change of "communal entropy" are -0.9 kcal/mol, -0.4 kcal/mol, and -0.6 kcal/mol, respectively. The joint probability distribution of relative distance r and angle θ is shown in Figure 7A. The contribution from the z dimension is -0.2 kcal/mol. The entropy changes from the z-axis due to the change of amplitude, and uneven distribution are -0.1 kcal/mol and 0.3 kcal/mol, respectively. The joint probability distribution of z coordinates of helix A and helix B in a dimer is presented in figure 7B. The small positive entropy change from the

z-axis is mainly due to a flatter probability distribution of the helices in a dimer compared to the monomers.

4.2.4 Rotational entropy loss of GpA upon association in bilayers

The rotational entropy change upon association in bilayers is -1.4 kcal/mol. The contributions from γ angle and from α and β angles are both -0.7 kcal/mol. As expected, the rotational entropy on the γ rotational dimension decreases since the monomers are free to rotate on that dimension but the relative rotation of one helix to the other in the dimer is restricted by inter-helical interactions. Figure 8 shows the distribution of the γ angle in the dimer ($\gamma = \gamma_B^d - \gamma_A^d$, γ_A^d and γ_B^d are Euler angles of helix A and helix B in the dimer with respect to the reference state, respectively). The joint probability distribution of α and β angles of helix A as a monomer and in a dimer is shown in figure 9. A -0.7 kcal/mol entropy loss was found on these two dimensions, although the coupled distribution in the monomer is sharper than that in the dimer. This indicates that the coupling between the angles of one and the other helix (Equation 20) is responsible for the entropy loss. The distribution peak is located about 27.5º on the β dimension before and after association, but the location of the distribution peak on the α dimension changes from ~63° to ~279º upon association. This reflects what was mentioned above, i.e., the tilt angle remains the same but the rotation angle changes upon association.

4.2.5 Side-chain conformational Entropy loss of GpA upon association in bilayers

The largest contributions to the side-chain conformational entropy loss are shown in Table 6, and they are from Glu72 (-0.8 kcal/mol), Leu75 (-0.7 kcal/mol), Ile76 (-0.8

kcal/mol), Phe78 (+0.4 kcal/mol), Ile91 (-0.5 kcal/mol), Arg96 (-0.4 kcal/mol) and Arg97 (-0.9 kcal/mol). The other residues' contributions are less than 0.3 kcal/mol. Leu75 and Ile76 are at the inter-helical interface. Glu72 and Ile91 are also found at the inter-helical interface in the solid-state NMR structure. Therefore it is reasonable that these residues contribute significantly to the conformational entropy loss. Interestingly, Val80 and Val84 in the LIxxGVxxGVxxT motif do not contribute significantly to the side-chain conformational entropy. This is consistent with the fact that the deuterium NMR spectra of Val84 in the monomeric and dimeric GpA peptides are remarkably similar, which indicated there is no conformational entropy loss of these valines upon association due to restriction of the side chain by intrahelical packing interactions involving the β-methyl group of valines (Liu et al. 2003). Unexpectedly, Phe78, Arg96, and Arg97, which are not at the inter-helical interface, contribute considerably to the side-chain conformational entropy. The positive side-chain entropy change from Phe78 is probably because Phe78 in the dimer is closer to the center of the hydrophobic core due to the shift of the dimer toward the C terminus (see figure 6), and its nonpolar side-chain gains conformational freedom in the dimer. Despite the shift of the dimer, the side-chains of Arg96 and Arg97 in the dimer are even closer to the interface between polar headgroups and the hydrophobic core due to the change in α angle upon association (see figure 6), which probably leads to stronger interactions between them and Tyr93, therefore Arg96 and Arg97 in the dimer appear to have less side-chain conformational freedom than they do in the monomer.

4.2.6 Standard free energy of GpA upon association in bilayers

Table 7 summarizes the results obtained in the previous sections. The favorable change in effective energy (-14.9±1.8 kcal/mol) is partially compensated by the unfavorable translational, rotational, and conformational entropy change to give a value of -7.7 kcal/mol for the standard association free energy of GpA in DMPC bilayers.

4.2.7 Standard free energy of GpA upon association in micelles

The standard association free energy of GpA in DPC micelles and its components are summarized in Table 8. The effective energy change and side-chain entropy loss are assumed to be the same as in bilayers. The translational entropy change upon association in micelles is less negative because the entropy change due to the change in number of translation states is positive at the (unrealistic) 1M HP standard state. The rotational entropy loss upon association in micelles is much larger than that in bilayers because before association monomers in micelles are free to rotate but monomers in bilayers are already rotationally constrained. The standard association free energy of GpA in DPC micelles calculated at 1M (in HP) standard state is -6.4±1.8 kcal/mol, which is close to -6.1 kcal/mol converted to the same standard state from the experimental value.

4.3 Discussion

In this work the standard association free energy was decomposed into the effective energy change, translational entropy loss, rotational entropy loss, and side-chain entropy loss upon association. The average effective energy change upon association was calculated as -14.9±1.8 kcal/mol, a result of compensation betweeen an inter-helical interaction of -23.0 kcal/mol and a helix reorganization energy of 7.9 kcal/mol. After the

entropic terms were included, the standard association free energy of GpA in DPC detergent micelles was theoretically estimated to be -6.4±1.8 kcal/mol, which is in excellent agreement with the value of -6.1 kcal/mol converted from free energy of association at mole fraction standard state determined experimentally (Fisher et al. 1999). This agreement must be somewhat fortuitous, given the drastic approximations made in the micelle calculation (the effective energy change and side-chain entropy loss in bilayers and micelles were assumed to be the same; the micelle was treated as an ideal sphere; contributions from the detergent were neglected).

Because micelles are convenient for quantitative studies, it is important to understand the difference in association in micelles and bilayers. Our calculations showed that the translational and rotational entropy changes upon association are different in bilayers and in micelles. The contribution of translational entropy to the association free energy in bilayers and micelles is 1.7 kcal/mol and 0.4 kcal/mol, respectively. The reason for the smaller translational entropy loss in micelles could be that, because of the spherical nature of micelles, and the requirement of only one monomer per micelle, much micelle volume remains inaccessible to the monomer. That is, a monomer in a certain micellar hydrophobic volume has less entropy than in an equal amount of bilayer hydrophobic volume, because bilayers are continuous. In contrast, the rotational entropy change in micelles is 2.6 kcal/mol less favorable than in bilayers, because the monomers are preoriented in bilayers but are free to rotate in micelles. Assuming that these are the only factors that are different between the two media, the calculated standard association free energy in DMPC bilayers is about 1.3 kcal/mol more favorable than in DPC micelles. This is consistent with the experimental observations that the association of GpA appears

stronger in membrane bilayers *in vivo* than in detergent micelles (Langosch et al. 1996) and that the difference in association constant of M2 protein in DMPC bilayers and DPC micelles corresponds to more than 0.7 kcal/mol for each helix-helix interface (Cristian et al. 2003). Another relevant observation is that certain epidermal growth factor receptors (erbB) helices have been found to associate *in vivo*, but not in micelles (Stanley and Fleming 2005).

In terms of structural changes upon association, we found that the rotation angle changes upon association but the tilt angle remains the same. This differs from the results of Henin et al. (2005) who proposed that change of tilt is coupled with helix-helix recognition. A possible reason for the discrepancy is that the tilt of the monomers observed by Henin et al. was smaller than in our study, perhaps due to the increased thickness of their membrane, while the tilt of the helices in the dimer by their and our calculations is almost the same due to inter-helical interactions. Thus, the change of tilt in their calculations is larger than ours.

Henin et al. (2005) estimated the association free energy of GpA by integrating the potential of mean force as a function of the distance between the helices. They obained the value -11.5 kcal/mol. The standard state implicit in their calculation is 1 molecule/Å^2 (Gilson et al. 1997), which, for 26 Å hydrophobic thickness, corresponds to 63.8M (in HP) in our standard state. The conversion of their value to our standard state can be done by $\Delta G^0 = \Delta G + \mathrm{RT}\cdot\ln\mathrm{C}$ (C= 63.8M) and gives ~ -9.0 kcal/mol, not very different from ours. Henin et al. find that van der Waals and solvation make about equal contributions to the association free energy. We, however, find that van der Waals is the only favorable force and solvation is unfavorable to association (see Table 4). The origin of this discrepancy is

unclear. It seems more sensible that the removal of lipids from contact with the associating helices should be unfavorable. Our analysis would be aided if the solvation contribution computed by Henin et al. could be resolved into water and organic solvent components.

It is useful to also compare the present results with those of Lomize et al. (2004). The effective energy change upon association was decomposed into -27.4 kcal/mol from van der Waals, 2.0 kcal/mol from electrostatics, 9.9 kcal/mol from solvation, and 0.3 kcal/mol from bonded terms. These values are larger than those of Lomize's. However similar energetic aspects are shown in both calculations: 1) van der Waals is the major driving force for association; 2) electrostatics and solvation energies are not favorable for association. The reorganization energy was neglected in their calculations, as well as the translational and rotational entropy changes. The side-chain conformational entropy loss we obtained is much larger than the 0.9 kcal/mol that Lomize et al. (2004) calculated, because they do not consider contribution from residues that are not at the inter-helical interface.

Grasberger et al. (1986) investigated the effect of restricted mobility on protein association in membrane and their results showed that restriction of translation and rotation in membrane can enhance protein dimerization by 4300 and 132 times respectively, compared to dimerization in aqueous solution. These two enhancements correspond to 4.8 kcal/mol and 2.8 kcal/mol. The enhancement from translational restriction refers to the difference between bulk concentration and local concentration in the membrane, including excluded volume effects, and is not relevant to our study. The rotational restriction enhancement of 2.8 kcal/mol is very similar to our calculation of 2.6

kcal/mol difference in rotational entropy loss between bilayers and micelles. They assumed a maximum tilt of 10°, although from our MD simulations the tilt angle of monomers could be more than 30°. Also the tilt angle distribution from our simulation is much wider than what they estimated.

It has been demonstrated that MD simulations based on IMM1 model can estimate the standard association free energy of TM helices in bilayers and micelles and illustrate possible translational, rotational, and conformational changes upon association. This work opens the way to quantitative investigations of the driving forces of TM helix association and de novo predictions of the propensity of TM helices to associate.

Chapter V

MODULATION OF TRANSMEMBRANE HELIX ASSOCIATION BY FLANKING AND NON-INTERFACIAL RESIDUES

5.1 Methods

5.1.1 IMM1-GC model

This work uses the implicit membrane models, IMM1 and IMM1-GC (Lazaridis 2003; 2005). IMM1 has been described in Chapter III. For negatively charged membranes, such as the inner membrane of *Escherichia coli*, a Gouy-Chapman (GC) term is added to W_{IMM1}. The resulting model is called IMM1-GC model, in which the effective energy is named as $W_{IMM1\text{-}GC}$ (Lazaridis 2005).

$$W_{IMM1\text{-}GC} = W_{IMM1} + E_{GC} \tag{40}$$

E_{GC} is defined as

$$E_{GC} = \sum_i \psi(z_i) q_i \tag{41}$$

where $\psi(z_i)$ is the electrostatic potential generated by a uniformly charged surface at position z_i according to GC theory, and q_i is the partial charge on atom i.

5.1.2 MD simulations

The sequences of GpA29, GpA15p11, MCP and MCP-GpA shown in Table 9 were

subjected to MD simulations. GpA29 and GpA15p11 only differ in the flanking residues surrounding the TM domain (residues 75-89 adopted from GpA wild type sequence) and they were selected in order to investigate whether these flanking residues affect association affinity. GpA29 has the original flanking residues as wild type GpA, whereas GpA15p11, MCP, and MCP-GpA all have the same flanking residues generated from restriction sites or ToxR' domains in a ToxCAT assay (Dawson et al. 2003). The solid-state NMR structure of GpA determined in lipid bilayers (Smith et al. 2001) was used as the initial structure for MD simulations. Since both molecular modeling and systematic substitution studies have shown that GpA and MCP TM helices have a similar GxxxG-mediated inter-helical interface and structural features (Melnyk et al. 2004), the initial structures of MCP and MCP-GpA were modeled from the solid-state NMR structure of GpA by structure threading. All the sequences were blocked by the acetyl group (ACE) and the methyl amide group (CBX) at the N and C termini, respectively. The initial structures were energy-minimized with 800 steps using the Adopted Basis Newton-Raphson (ABNR) method before MD simulations were performed. Because the TM segments of GpA and MCP are mainly α-helical (MacKenzie et al. 1997; Smith et al. 2001; Melnyk et al. 2002; Melnyk et al. 2004), the backbone dihedral angles, phi and psi were constrained (force constant 100.0) at the ideal values, -57º and -47º, respectively, to reduce energy fluctuations and thus the error bars. Backbone dihedral angle constraints were found to give lower average effective energies than NOE constraints. The thickness of the membrane hydrophobic core and the smeared charge offset from the hydrocarbon boundary were set to be 26 Å and 3 Å, respectively. All simulations were conducted at 298.15 K with the CHARMM package and the NOSE integrator. For a dimer, the

simulation was first performed under BESTFIT constraints, which are similar to the absolute positional constraints except that the reference structure is allowed to rotate and translate so as to best fit the selected atoms and minimize the restraint energy, on the backbone atoms (force constant 1.0) for 0.25 ns for equilibration before the production simulation stage. The monomers were directly subjected to simulations for the production stage. The average structural, configurational and energetic properties, including the root-mean-square deviation (RMSD) and the crossing angle of dimers, the translation on z-axis, the tilt angle relative to the membrane normal and the orientation about the helical axis of the GxxxG motif, effective energies, and inter-helical interactions, were calculated from the last 0.9 ns of the simulations as in previous work (Zhang and Lazaridis 2006).

5.1.3 Determination of lysine ionization state

Lys40 on the TM domain of MCP and MCP-GpA dimers points toward the membrane interior. This could shift its pKa sufficiently to change its ionization state. Thus, the ionization state of Lys40 in the monomer and the dimer should be determined. Taking protonated Lys40 as the reference state, the deprotonation energy, $\Delta G_{deprotonation}$, in a monomer and a dimer can be calculated by the equation (Mihajlovic and Lazaridis 2006):

$$\Delta G_{deprotonation} = -2.303RT\,(pH-pKa) + [(W_{Protein\text{-}Membrane\text{-}Lys} - W_{Protein\text{-}Membrane\text{-}Lys+}) - (W_{Model\text{-}Water\text{-}Lys} - W_{Model\text{-}Water\text{-}Lys+})] \quad (42)$$

where the pH and the pKa of lysine in water are taken as 7.0 and 10.53, respectively. $W_{Protein\text{-}Membrane\text{-}Lys}$ and $W_{Protein\text{-}Membrane\text{-}Lys+}$ are average effective energies of a protein

(either a monomer or a dimer) in the membrane with Lys40 deprotonated and protonated, respectively. Simulations of deprotonated and protonated forms of monomers and dimers were performed separately using the IMM1 and IMM1-GC models to calculate $W_{Protein\text{-}Membrane\text{-}Lys}$ and $W_{Protein\text{-}Membrane\text{-}Lys+}$. $W_{Model\text{-}Water\text{-}Lys}$ and $W_{Model\text{-}Water\text{-}Lys+}$ are average effective energies of model compounds where the lysine is deprotonated and protonated, respectively, and fully exposed to aqueous solution. A short peptide, ACE-ALA-ALA-LYS-ALA-ALA-CBX, in extended conformation was used as the model compound. $W_{Model\text{-}Water\text{-}Lys}$ and $W_{Model\text{-}Water\text{-}Lys+}$ were calculated from 10-ns simulations in EEF1 with all residues except lysine fixed. If $\Delta G_{deprotonation} < 0$, then the deprotonated state is more favorable.

5.1.4 Association free energy calculations

The standard association free energy was calculated at 1M in hydrophobic phase (HP) standard state as previously (Zhang and Lazaridis 2006). The standard association free energy (ΔG^0_{HP}) can be decomposed into the free energy upon change in ionization state (ΔG_{ion}), the effective energy change upon association (ΔW_{ass}), the translational entropy loss (ΔS^{trans}_{ass}), the rotational entropy loss (ΔS^{rot}_{ass}), and the conformational entropy loss (ΔS^{conf}_{ass}):

$$\Delta G^0_{HP} = \Delta G_{ion} + \Delta W_{ass} - T\Delta S^{trans}_{ass} - T\Delta S^{rot}_{ass} - T\Delta S^{conf}_{ass} \quad (43)$$

The first term was added due to Lys40 in the TM domain of MCP and MCP-GpA, which can have different ionization states in the monomer and the dimer. For example, if Lys40

prefers the protonated state in the monomer and the deprotonated state in the dimer, the overall dimerization process can be described as a thermodynamic cycle:

$$\begin{array}{ccc} & \xrightarrow{\text{Association II}} & \\ \text{2 MCP monomers (Lys+)} & \rightarrow & \text{MCP dimer (Lys+)} \\ \downarrow \text{Deprotonation I} & & \downarrow \text{Deprotonation II} \\ \text{2 MCP monomers (Lys)} & \rightarrow & \text{MCP dimer (Lys)} \\ & \xrightarrow{\text{Association I}} & \end{array} \qquad (44)$$

The overall dimerization energy can be calculated by pathway I or pathway II. The effective energies during the last 0.9 ns of several 1-ns MD simulations of monomers and dimers in the IMM1-GC model were averaged and used to calculate ΔG_{ion} and ΔW_{ass}. Since the constraints on backbone dihedral angles generate an artificial additional energy term, this energy term was removed in our effective energy calculations. The magnitude of this term was ~28 kcal/mol and ~15 kcal/mol for the dimer and the monomer, respectively. The calculation of the entropic terms is the same as before (Zhang and Lazaridis 2006), except the membrane hydrophobic core thickness was changed from 23 Å to 26 Å.

5.2 Results

5.2.1 Initial dimer structures

The solid-state NMR structure of GpA (Smith et al. 2001) was used as the initial structure of GpA29 and GpA15p11 dimers. The initial structures of MCP and MCP-GpA dimers

were modeled from the solid-state NMR structure of GpA (Smith et al. 2001) by structure threading, based on the modeling results of Melnyk et al. (2004). Figure 10 shows the initial structures of GpA15p11 and MCP dimers. Although GpA15p11 and MCP have the same type of flanking residues from the ToxCAT constructs, the 13-residue GxxxG association core is closer to the N-flank in GpA15p11 while it is near the middle in MCP.

5.2.2 Ionization states of Lys40 in the monomer and dimer of MCP and MCP-GpA

In the model of the MCP and MCP-GpA dimers Lys40 points toward the membrane interior. This environment can induce large pKa shifts. Thus, the preferred ionization states of Lys40 in the monomer and dimer should be determined first. Taking the protonated state as the reference state, deprotonation energies of MCP and MCP-GpA monomers and dimers in zwitterionic (IMM1) and anionic (IMM1-GC) membranes were calculated from separate simulations of the monomer and the dimer with protonated/deprotonated Lys40. The results are shown in Table 10.

The deprotonation energy of MCP and MCP-GpA dimers is highly favorable in both membrane types (The deprotonation energy of the MCP dimer in IMM1 could not be calculated because three MD runs of the protonated MCP dimer could not maintain the dimer structure, while the deprotonated MCP dimer was stable. This suggests that the deprotonation energy for MCP is also highly favorable.) The deprotonation energies of MCP and MCP-GpA monomers are slightly negative in IMM1 and positive in IMM1-GC, apparently due to the negative charge of the membrane in the latter. This is consistent with the experimental determination by FRET that Lys40 is located near the hydrophobic and hydrophilic interface and seems to be protonated when the MCP monomer is

imbedded in an anionic bilayer (Vos et al. 2005). Therefore, the dimerization process in IMM1-GC includes two steps and it can be represented by the thermodynamic cycle in equation (44), compared with only one step in IMM1.

5.2.3 Configuration of monomers and dimers in the anionic membrane

For each dimer of the GpA29, GpA15p11, MCP and MCP-GpA sequences, three 1-ns MD simulations were performed in IMM1-GC, starting with a different random number for the assignment of velocities. For the MCP and MCP-GpA dimers only the deprotonated state of Lys40 was considered. As shown in Table 11, the average backbone RMSDs of GpA29, GpA15p11, MCP and MCP-GpA dimers with respect to their initial structures are 1.1 Å, 1.2 Å, 2.0 Å and 1.5 Å, respectively, which indicates that dimers are sufficiently stable during the simulations. Table 11 also shows the average crossing angles of the dimers during the simulations, which are 39º, 42º, 46º and 42º, respectively. Even though the flanking residues of GpA29 and GpA15p11 sequences are dramatically different, there is almost no difference between their average dimeric structures. The average crossing angle of MCP in our simulations is somewhat larger than the value of ~40º predicted (Melnyk et al. 2004), perhaps because a longer MCP sequence and different energy function were used in our study. As expected, MCP-GpA has a structure more similar structure to GpA15p11 than MCP does, in terms of RMSD and average crossing angles.

To describe the configuration relative to membrane, we calculated the translation of the center of the 13-residue association core on the z-axis (which coincides with the membrane normal), the tilt angle of the helix relative to the membrane normal, and the

orientation about the helical axis of the 13-residue association core. The configurations of different TM helices were compared and the change in their configurations upon association was examined (shown in Table 11 and Figure 11). GpA29 and GpA15p11 are similar in the tilt and orientation of helices in the dimer but significantly different in the translation of the monomer and dimer, the tilt angle of the monomer, and the orientation of the monomer. The translation, tilt angle and orientation of MCP and MCP-GpA are dramatically different for the protonated and deprotonated monomers. For all four sequences, the change of the translation and tilt angle from the uncharged monomer to the dimer was relatively small; however the change of orientation was considerable. Thus different flanking residues and ionizable residues in the TM domain could induce large configurational changes of TM helices in the membrane.

5.2.4 Inter-helical interactions in the anionic membrane

Average inter-helical interactions of the four sequences in the IMM1-GC model are shown in Table 12. Surprisingly, the average inter-helical interaction of GpA15p11 is 6.5 kcal/mol stronger than that of GpA29. The two sequences differ only in the flanking residues. The effect could be direct or indirect, i.e., the flanking residues could increase the inter-helical interaction between the LIxxGVxxGVxxT motifs because of slight structural and configurational changes induced, or they could directly participate in inter-helical interactions. To distinguish between the two possibilities, inter-helical interactions between the 13-residue motifs on different helices were calculated. This interaction was found to be only 1.3 kcal/mol weaker in GpA29 than in GpA15p11. The remainder must be due to direct interactions between flanking residues. The GpA15p11

flanking residues on the N-terminus, especially the first two residues Asn1 and Arg2, could be the major contributors, because 1) they are separated from the LIxxxGxxxGxxxT motif by only 3 and 2 residues, respectively, and this places them near the inter-helical interface; and 2) they are strong H-bond donors or acceptors. In fact, a strong interaction between Asn1 from one helix and Arg2 from the other helix was observed in GpA15p11 and the average interaction between them during the simulations was -2.7 kcal/mol. When two pairs of Asn1 and Arg2 are considered, their contribution to inter-helical interaction is doubled to -5.4 kcal/mol. Hence, flanking residues could participate in inter-helical interactions and affect association affinities.

Consistent with observations from prior experiments (Melnyk et al. 2004), the inter-helical interaction of GpA15p11 is much stronger than that of MCP, and the inter-helical interaction of MCP-GpA is between those of GpA15p11 and MCP. However, when we only include the 13 residues corresponding to the LIxxxGxxxGxxxT motif, the inter-helical interaction of MCP is close to that of GpA15p11, and the inter-helical interaction of MCP-GpA is 2.3 kcal/mol stronger. Therefore, MCP is not a weak scaffold in terms of inter-helical interactions from the 13-residue interfacial core compared to GpA. The stonger inter-helical interactions in GpA15p11 must be due to the flanking residues. Although both MCP and MCP-GpA have the same flanking residues as GpA15p11, Asn1 and Arg2 in MCP are 9 residues further away from the 13-residue interfacial core compared with those in GpA15p11, so these two residues in MCP are not close enough to form H-bonds (see Figure 12). This is supported by the negligible interaction between Asn1 and Arg2 from different helices in MCP and MCP-GpA (see Table 12).

5.2.5 Standard free energies upon association in the anionic membrane

Table 13 summarizes the results for the calculation of standard free energies of the four sequences upon association in the IMM1-GC model. Since for MCP and MCP-GpA Lys40 prefers to be protonated in a monomer and deprotonated in a dimer, the association free energy includes the free energy of deprotonation of the monomers and the association free energy of deprotonated monomers (see two steps of pathway I in equation 44). The effective energy change upon deprotonation is +2.1±1.1 kcal/mol and +2.2±1.6 kcal/mol for MCP and MCP-GpA, respectively. The standard association free energies of GpA29, GpA15p11, MCP, and MCP-GpA are -6.5±0.8 kcal/mol, -13.0±1.5 kcal/mol, -4.9±1.9 kcal/mol, and -11.4±2.5 kcal/mol, respectively, so the association affinity of MCP-GpA is between those of GpA15p11 and MCP. In experiments, the CAT activity was about 10-fold decreased from GpA15p11 to MCP construct; however MCP-GpA had about 60% CAT activity relative to the GpA15p11 construct. Since the CAT activity is directly proportional to the amount of dimer and total expression of each constructs was similar but not known, the CAT activity can be used as a qualitative indicator to compare association affinities, but the difference in the CAT activities can not be simply converted to the difference in association free energies. Thus our calculations are consistent with the experimental observations.

5.3 Discussion

The main results of these calculations are the following: a) a residue can influence association affinity without being at the binding interface b) flanking polar residues can

affect association affinity via direct interactions.

Non-interfacial residues in the putative TM domain can affect the association affinity because the extent of their burial in the membrane could be different in monomers and dimers. The effect is particularly strong for charged residues. In the case of MCP and MCP-GpA, the tilt and orientation of helices in the dimer are highly constrained due to the inter-helical interactions, and thus two Lys40 side-chains in the dimeric structure are forced to point towards the membrane's hydrophobic core. Translation of the protonated dimer downward to expose Lys40 to the headgroup region is prevented by the unfavorable burial of Tyr24 on the other end. However, the protonated monomer can adopt an orientation in which Tyr24 side-chain points up and Lys40 side-chain points down (see Figure 11). This allows the helix to translate ~3.9 Å down towards the C-terminus compared with the deprotonated dimer (see Table 11) and expose Lys40 to the headgroup region (see Figure 11). Deprotonation of Lys40 makes dimerization possible, but at a significant free energy cost. The coupling of protonation/deprotonation and oligomerization has been observed in artificially designed pLeu peptides (Lew et al. 2000; Lew et al. 2003), M2 protein (Salom et al. 2000), and bacterioopsin (Valluru et al. 2006). The energy cost of deprotonation for MCP and MCP-GpA is consistent with these experimental studies and could be the major reason why the association affinity of MCP and MCP-GPA is weaker than that of GpA15p11. Consideration of the heterogeneous membrane environment and how the monomers or oligomers fit into it is essential for a full understanding of binding affinity in transmembrane systems.

The influence of flanking residues surrounding the TM domain on the association

affinity has not been studied extensively. After the TM 13-residue LIxxGVxxGVxxT motif was first identified, few efforts were made to study the residues outside this motif. Moreover, different flanking residues were systematically introduced in ToxR' and ToxCAT assays, but their role has not been discussed. Although information about flanking residues in ToxR' and ToxCAT is not always provided, most studies employed an asparagine consecutive to an arginine at the N-terminal flank of the TM domain. Our study demonstrated that these two residues (Asn1 and Arg2) could produce strong inter-helical interaction by forming H-bonds when they are 3 and 2 residues away from the 13-residue interfacial core. The observation that GpA TM helices seem to associate more strongly in the membrane than in micelles (Langosch et al. 1996; Russ and Engelman 1999) could potentially be due to the flanking residues, rather than differences in the environment. SDS-PAGE experiments used the wild type sequence (like GpA29 in our study) whereas *in vivo* experiments used the non-native flanking residues like GpA15p11 in our study.

Besides the type of flanking residues, their position with respect to the GxxxG motif also affects association affinity. This effect can explain several experimental observations. Langosch et al. (1996) found that when the full 13-residue motif was inserted between the N-terminal and C-terminal flanking sequences, NRAS and ILINP respectively, it gave the strongest dimerization signal, compared with insertion of 12 and 14 residues. The reason for this could be that both the addition of one residue at the N-end and the deletion of one residue from the C-end of the 13-residue motif shift the position of the polar residues at the N- and C- flanks, and affect the formation of hydrogen bonds between the helices. Johnson et al. (2006) studied the effect of GxxxG position on the dimer affinity and have

shown that the association in ToxCAT tests is stronger when the GxxxG motif is 8 residues away from the N-terminal flank compared with that when it is 12 residues away. Besides the position of GxxxG in the membrane, its position relative to the flanking residues is shifted, which may cause a change of inter-helical interactions between the flanking residues.

While the qualitative picture that emerges from this study is physically realistic, the approximate nature of the IMM1 model forces us to view the quantitative aspects with some reservations. Specifically, the deprotonation energies we calculated depend on the solvation parameters of Lys in water and the membrane interior, which for protonated Lys are somewhat arbitrary. In our model the transfer free energies of the protonated and deprotonated Lys side-chains from water to the membrane interior are 14.49 kcal/mol and -0.06 kcal/mol, respectively. The value for protonated Lys is much smaller than one would get using the Born model. However, this value is not far from reality. Recent studies showed that the burial of charged side-chains in membranes is less costly than previously thought due to membrane adjustments or transfer of water molecules (Freites et al. 2005). Recent calculations of the potential of mean force needed to transfer a charged Arg side-chain from water to the membrane core by MD simulations with an explicit membrane yielded the value ~17 kcal/mol (Dorairaj and Allen 2007). The transfer free energy of the protonated Arg side-chain from water to the membrane interior in our model is 20.67 kcal/mol, quite close to the above value.

There is also some uncertainty with the Arg parameterization at the membrane interface. We have found evidence suggesting that the Arg-Arg interaction should be more repulsive than predicted by IMM1 (Mottamal et al. 2006). In our calculations there

were no Arg pairs interacting, but a pair of Asn and Arg in GpA15p11 were found to contribute 2.7 kcal/mol to the inter-helical interactions. This value seems reasonable, but should be treated with some caution.

One final source of uncertainty is that the structure of MCP is putative. The stability of these putative structures under MD simulations and the good agreement with the relative affinities obtained by TOXCAT experiments lends some support to the structures proposed by Melnyk et al. (2004).

Our findings could easily be tested by experiments: 1) insertion of one or a few residues between the non-native flanking residues and the 13-residue association core in the sequence of GpA15p11 should decrease the association affinity; 2) mutation of Lys40 on MCP and MCP-GpA to a hydrophobic residue or even a less polar residue should increase the association affinity; and 3) high pH should also increase the association affinity by lessening the cost of deprotonating Lys.

Chapter VI

CONCLUSIONS

In summary, from our theoretical studies, we can draw the following conclusions:

1) In GpA, the largest contribution to the inter-helical energy was the van der Waals interaction.

2) In GpA, a large contribution to the inter-helical energy came from the GXXXG motif; however non-GXXXG residues are also important for the stabilization of the helix dimer.

3) The overall contribution of electrostatic interactions to the inter-helical energy in GpA dimer is very small. However, this small value contains larger compensating favorable and unfavorable interactions. Therefore, neglect of electrostatic interactions cannot be a generally viable strategy for structure prediction as thought before.

4) With a given dimer structure, our theoretical methodology based on the IMM1 model can successfully estimate the standard association free energy of TM helices in both micelles and bilayers and illustrate possible translational, rotational, and conformational changes upon association. The calculated standard association free energies are in good agreement with the experimental measurements.

5) The translational and rotational entropy changes upon association are different in bilayers and in micelles. The translational entropy cost is larger, while the rotational entropy cost is smaller in bilayers than in micelles. Although these differences were qualitatively predicted before, we gave quantitative descriptions on these differences using GpA as an example.

6) Non-interfacial residues in the putative TM domain can affect the association affinity because the extent of their burial in the membrane could be different in monomers and dimers. The possible influence of non-interfacial residues to TM helix association has not been noticed before and this conclusion will evoke more attention on non-interfacial residues especially when they are charged or polar.

7) Flanking polar residues can affect association affinity via direct interactions and their position with respect to the GxxxG motif also affects association affinity. This conclusion is surprising since previous studies using the *in vivo* constructs haven't noticed the possible role of flanking residues in TM helix association.

We should also be aware of the strengths and weaknesses of our approach:

1) Treating the water and lipid molecules implicitly has a good and a bad side: the good side is that it saves tremendous computing time; while the bad side is that it lacks a lot of detail in interactions between proteins and water or lipid molecules. Given the fact that the structurally or functionally meaningful TM helix association is always specific and sequence-dependent, the interactions between TM helices themselves are dominant, and thus treating the water and lipid molecules implicitly seems not very harmful. However in some particular cases when water or lipid molecules mediate the interactions between TM helices, neglecting the chemical structure of water and lipid molecules may cause inaccuracies.

2) A few physical and physiological aspects of biological membranes are missing in our model. It is questionable to treat the headgroup region as water. Although in IMM1-GC model, smeared charge surface potential is added to represent the electrostatic interactions between TM helix and headgroups, new solvation parameters particularly in

the headgroup environment are still needed. In a biological membrane, the charge distribution of headgroups is horizontally uneven on membrane surfaces when a TM protein is present, but in our IMM1-GC model charges are assumed to be evenly distributed. Moreover, the TM potential is not included in our studies (although it could have been). Despite these problems, using the IMM1 and IMM-GC models we can still reach reasonable conclusions. In chapter III, the conclusions are based on the inter-helical interactions between TM domains. These TM domains are mostly buried in the hydrophobic core of a bilayer and hardly interact with headgroups, therefore imprecise representation of the headgroup region doesn't hurt our conclusions. In chapter IV, our calculation of association free energy captures the main aspect given that TM domains are the determinant for association. The translational and rotational entropy difference in a bilayer and micelles is mainly due to their different geometric structures. In chapter V, flanking residues play a role in association through direct interaction between themselves. When charges on headgroups were explicitly represented, the interactions between protonated Lys40 and headgroups could be stronger and then the deprotonation energy of Lys40 in a monomer could be larger. Thus the influence of Lys40 in association could be more obvious.

3) In this thesis, the influence of membrane thickness to TM helix association has not been addressed. For example, the red blood cell membrane where GpA locates varies from 50nm to 90nm in thickness under different impingement forces (Heinrich et al. 2001). The red cell membrane has a complex architecture containing various superficial glycoproteins and a subsurface network of spectrin besides a lipid bilayer, and it is too complicated to study by our model. More simply in a bilayer system alone, changing the

thickness of its hydrophobic core can also change the TM helix association affinity, whereas the influence of the bilayer thickness is highly sequence dependent.

Our studies have implications for membrane protein structure prediction.

1) We can predict the structure of TM helix dimer (homomeric or heteromeric) using association free energy as a guideline. Although effective energy is the dominant term that drives the folding of a dimer of TM helices, entropic terms, especially the side-chain conformational entropy, could help to differentiate conformations with close effective energies. First, we artificially build a series of dimer structure of two ideal α helices according to combination of different orientation and distance of two helices described by Mottamal et al. (2006). Second, we minimize the energy of those dimer structures and run MD simulation of each minimized dimer structure. MD simulations of a monomer will also be performed. Third, we class all structure frames from different MD trajectories of dimers into a number of groups. Structure frames in each group are similar to each other. Fourth, we calculate the standard association energy for each group and the structure group with the most favorable association free energy could be the predicted structure.

2) For a homo-oligomer of TM helices, our method could probably be used to predict the oligomerization state with reasonable symmetric constraints. Compared with previous studies (Kim et al. 2003; Bu et al. 2007), our method will be expected to give more accurate calculation of the entropic terms. These entropic terms can be important for predicting the correct oligomerization state. First we build an ideal helix as a monomer and then use CHARMM symmetry facilities to generate the structures of a trimer, a tetramer and so on. When the orientation of the monomer and the distance between helices vary, the structure of oligomers changes so that we can build various

oligomer structures. Second, we run MD simulations of the monomer and oligomers. Third, we predict the structure for each oligomerization state according to oligomerization free energy. Fourth, we compare the average free energy change of a single helix from a monomer to an oligomer. The oligomer with the most favorable average free energy change of a single helix will be the most stable oligomerization state.

3) It is difficult to predict the structure of a hetero-oligomer TM helical bundle. Since there is no symmetric constraint, the number of combinations of all conformational variables is tremendously large in a hetero-oligomer. One possible strategy is the following. To reduce the space of conformational search, the inter-helical interface of each two adjacent helices should be determined first. In other words, we first predict several energetically favorable dimer structures of each two adjacent helices. Then we combine these favorable dimer structures to build a hetero-oligomer. Later we run MD simulations of all constructed oligomers. The structure with the most favorable association free energy will be the predicted structure.

4) Predicting the structure of a multiple-span TM protein is even more difficult. The difficulties are a) the conformation of turns connecting different TM domains is very dynamic and hard to predict; b) one TM helical domain could have multiple contacts with other domains, which dramatically increases conformational space. A possible strategy is the following. First, define the TM helical domains according to sequence hydrophobicity. Second, predict several favorable structures of TM helix bundle described previously assuming that TM helices are all disconnected. Third, add turn structures and use connectivity information to evaluate these structures.

Table 1. **Total energies and inter-helical interactions of the two NMR GpA structures.** All energies are in kcal/mol. The values in parentheses are van der Waals, electrostatic, and solvation energies, in this order.

	Solution NMR Structure		Solid-state NMR Structure	
	Total Effective Energy	Inter-helical Interaction	Total Effective Energy	Inter-helical Interaction
After 800 steps ABNR Minimization	-1108.1 (-247.0, -748.4,-170.2)	-23.6 (-36.2, 1.7,10.8)	-1106.5 (-254.0, -737.4,-169.1)	- 29.3 (-42.9, 1.5,12.1)
Average over the last 0.9 ns	-760.8 (-220.3, -729.1,-172.8)	-21.8 (-34.4, 1.7,10.9)	-764.9 (-224.0, -729.2,-171.4)	-24.9 (-39.5, 2.2,12.4)

Table 2. **Residue-residue interactions in the energy-minimized solution NMR GpA structure.** All energies are in kcal/mol. The values in parentheses are van der Waals, electrostatic, and solvation energies, in this order. Only the residue-residue interactions with absolute values greater than 0.5 kcal/mol are shown. The small differences observed in some symmetric positions are due to a slight asymmetry in the minimized structure.

	A75LEU	A76ILE	A79GLY	A80VAL	A82ALA	A83GLY	A84VAL	A87THR	A88ILE
B75LEU		-1.8 (-2.3,0.2,0.3)							
B76ILE	-1.8 (-2.3,0.1,0.3)	-1.4 (-2.6,0.4,0.8)	-1.1 (-0.9,-0.4,0.2)						
B79GLY		-0.9 (-0.8,-0.3,+0.2)		-0.9 (-1.5,0.0,0.6)					
B80VAL			-0.7 (-1.4,0.0,0.6)	-0.5 (-1.3,0.4,0.4)	-0.7 (-0.9,0.0,0.2)	-1.6 (-1.4,-0.6,0.5)			
B82ALA				-0.6 (-0.8,0.0,0.2)					
B83GLY				-1.5 (-1.4,-0.6,0.4)			-1.0 (-1.5,0.0,0.4)		
B84VAL						-1.0 (-1.5,0.0,0.4)		-1.5 (-2.0,-0.2,0.7)	
B87THR							-1.5 (-2.0,-0.2,0.7)	-0.6 (-1.0,-0.2,0.6)	-0.6 (-0.8,-0.1,0.2)
B88ILE								-0.6 (-0.7,-0.1,0.2)	

Table 3. **Residue-residue interactions in the energy-minimized solid-state NMR GpA structure.** All energies are in kcal/mol. The values in parentheses are van der Waals, electrostatic, and solvation energies, in this order. Only the residue-residue interactions with absolute values greater than 0.5 kcal/mol are shown. The small differences observed in some symmetric positions are due to a slight asymmetry in the minimized structure.

	A73ILE	A75LEU	A76ILE	A79GLY	A80VAL	A82ALA	A83GLY	A84VAL	A87THR	A88ILE	A91ILE
B73ILE	-1.2 (-1.1,0.0,-0.1)	-0.6 (-0.8,0.1,0.1)									
B75LEU	-0.6 (-0.8,0.0,0.1)		-1.9 (-2.4,0.1,0.4)								
B76ILE		-1.9 (-2.4,0.1,0.4)	-0.9 (-2.0,0.4,0.7)	-1.2 (-1.1,-0.4,0.3)							
B79GLY			-1.1 (-1.0,-0.4,0.3)		-0.8 (-1.4,0.0,0.6)						
B80VAL				-0.8 (-1.4,0.0,0.6)		-0.7 (-0.9,0.0,0.2)	-1.6 (-1.5,-0.6,0.5)				
B82ALA					-0.7 (-0.9,0.0,0.2)						
B83GLY					-1.6 (-1.5,-0.6,0.5)			-1.0 (-1.5,0.1,0.4)	-0.6 (-0.3,-0.4,0.1)		
B84VAL							-1.0 (-1.5,0.1,0.5)		-2.5 (-2.0,-1.3,0.8)		
B87THR							-0.5 (-0.3,-0.3,0.1)	-1.8 (-2.0,-0.6,0.8)	0.9 (-1.1,1.3,0.7)	-0.5 (-0.9,0.1,0.3)	
B88ILE									-0.6 (-1.1,0.1,0.4)		
B91ILE											-1.4 (-1.9,0.0,0.4)

Table 4. **Average effective energy change upon GpA association in a 23-Å lipid bilayer calculated from 1ns MD simulations (kcal/mol).** Numbers in parentheses are van der Waals, electrostatics, and solvation contributions to ΔW, respectively.

	$W_{\text{Monomer A}}$	$W_{\text{Monomer B}}$	W_{Dimer}
Run #1	-516.5	-516.1	-1047.1
Run #2	-518.9	-516.6	-1049.0
Run #3	-516.5	-515.9	-1048.7
Run #4	-515.2	-516.1	-1046.7
Run #5	-517.5	-516.9	-1049.4
Run #6	-516.8	-516.7	-1046.7
Run #7	-516.2	-516.9	-1049.6
Average	-516.8±1.2	-516.5±0.4	-1048.2±1.3
ΔW	-14.9±1.8 (-27.4, 2.1, 9.9)		

Table 5. **Residue contributions to the reorganization energy upon GpA association in a 23-Å lipid bilayer (run #1) (kcal/mol).** VDW, ELEC, SOLV, and BOND are van der Waals, electrostatics, solvation, and bonded components of the reorganization energy, respectively. Only the residues with the largest absolute contributions (≥1 kcal/mol) are shown.

Residue	ALL	VDW	ELEC	SOLV	BOND
GLU70	1.0	0.3	0.3	0.6	-0.3
ILE73	-2.3	1.4	-1.1	-1.8	-0.7
THR74	2.1	0.4	-0.5	2.0	0.3
LEU75	2.7	-0.2	0.2	3.1	-0.3
ILE76	-2.3	0.0	-0.2	-2.2	0.0
ILE77	-2.0	-1.6	0.2	-0.8	0.2
ILE88	1.8	1.6	0.2	0.1	-0.1
ILE91	4.6	0.2	0.8	3.1	0.5
SER92	-2.3	0.3	0.1	-2.5	-0.2
TYR93	2.1	1.0	0.1	0.5	0.5
ILE95	1.5	1.4	-0.1	0.3	0.0
ARG96	1.0	0.4	-0.1	0.3	0.5
TOTAL	7.9	8.9	0.3	-1.6	0.3

Table 6. **Side-chain entropy changes of some residues upon GpA association in bilayers (kcal/mol).** The temperature T is 298.15 K. Side-chain entropy changes of the other residues are not shown in this table and their absolute values are less than 0.3 kcal/mol.

residue	$-T\cdot\Delta S^{\text{side-chain}}$	dihedral	$-T\cdot\Delta S^{\text{side-chain}}$
GLU72	-0.8	1	-0.4
		2	-0.2
		3	-0.2
LEU75	-0.7	1	-0.5
		2	-0.2
ILE76	-0.8	1	-0.1
		2	-0.7
PHE78	0.4	1	0.3
		2	0.1
ILE91	-0.5	1	-0.0
		2	-0.5
ARG96	-0.4	1	-0.2
		2	-0.1
		3	-0.1
		4	-0.1
		5	-0.0
ARG97	-0.9	1	-0.3
		2	-0.2
		3	-0.4
		4	-0.0
		5	-0.0
All 29 residues	-4.1		

Table 7. **Standard free energy of GpA upon association in membrane bilayers at 1M (in HP) standard state (kcal/mol).** The temperature T is 298.15 K.

ΔW	-14.9±1.8
$\Delta S^{trans} \bullet T$	-1.7
$\Delta S^{rot} \bullet T$	-1.4
$\Delta S^{side\text{-}chain} \bullet T$	-4.1
ΔG^0 at 1M (in HP) standard state	-7.7±1.8

Table 8. **Standard free energy (kcal/mol) of GpA upon association in DPC micelles at 1M (in HP) standard state.** The temperature T is 298.15 K. Numbers in parentheses are entropy changes due to two terms: one arising from the distribution of helices in micelles and one from the local "vibrations" of the helices within a micelle.

ΔW	-14.9±1.8
$\Delta S^{trans} \bullet T$	-0.4 (0.8, -1.2)
$\Delta S^{rot} \bullet T$	-4.0
$\Delta S^{side\text{-}chain} \bullet T$	-4.1
ΔG^0 at 1M (in HP) standard state	-6.4±1.8

Table 9. **The sequences studied in chapter V.** Residues in red are those on the putative TM domain adopted from GpA or MCP wild type sequence. Flanking residues (Dawson et al. 2003) are underlined. Mutated residues are framed.

Name	Sequence
GpA29	ACE GLU PRO GLU ILE THR LEU ILE ILE PHE GLY VAL MET ALA GLY VAL ILE GLY THR ILE LEU LEU ILE SER TYR GLY ILE ARG ARG LEU CBX
GpA15p11	ACE ASN ARG ALA ARG LEU ILE ILE PHE GLY VAL MET ALA GLY VAL ILE GLY THR ILE LEU LEU ILE LEU ILE ASN PRO SER CBX
MCP	ACE ASN ARG ALA ARG TYR ILE GLY TYR ALA TRP ALA MET VAL VAL VAL ILE VAL GLY ALA THR ILE GLY ILE LYS LEU PHE LEU ILE LEU ILE ASN PRO SER CBX
MCP-GpA	ACE ASN ARG ALA ARG TYR ILE GLY TYR ALA TRP ALA MET VAL LEU ILE ILE VAL GLY VAL THR ILE GLY VAL LYS LEU THR LEU ILE LEU ILE ASN PRO SER CBX

Table 10. **Deprotonation energies (kcal/mol) of MCP and MCP-GpA monomers and dimers calculated from 1-ns MD simulations using the IMM1 and IMM1-GC models.**

		IMM1	IMM1-GC
MCP	Monomer	-0.3	2.1
	Dimer	NA	-9.5
MCP-GpA	Monomer	-1.1	2.2
	Dimer	-18.0	-13.3

*Note: The average energy of the protonated MCP dimer in IMM1 was not calculated because none of three runs was able to maintain the dimer structure. The average energies of protonated dimers in IMM1-GC were calculated from one run because two out of three runs were not stable. The other values were calculated from three runs. Equation 9 was used to calculate ionization/deprotonation energy for the monomer and dimer.

Table 11. **Structures and configurations in the IMM1-GC membrane model.** The average RMSD, crossing angle, translation on z-axis, tilt angle, and orientation about the helix axis were calculated from the last 0.9 ns of 1-ns MD simulations. Helix A and B are two helices that form a dimer. The protonated monomer is denoted as "Monomer+".

Structural properties of dimers									
Sequence		GpA29		GpA15p11		MCP		MCP-GpA	
RMSD (Å)		1.1		1.2		2.0		1.5	
Crossing Angle (º)		39		42		46		42	
Configuration of a helix as a monomer or in a dimer									
Sequence		GpA29		GpA15p11		MCP		MCP-GpA	
		Helix A	Helix B	Helix A	Helix B	Helix A	Helix B	Helix A	Helix B
Translation (Å)	Monomer+	/	/	/	/	-4.2	-4.3	-4.1	-4.0
	Monomer	1.3	1.2	3.0	2.9	-0.2	-0.2	-0.4	-0.3
	Dimer	0.4	0.5	2.7	2.8	-0.1	0.2	-0.2	-0.3
Tilt Angle (º)	Monomer+	/	/	/	/	41.2	41.5	40.6	39.7
	Monomer	16.5	15.3	21.0	21.2	23.0	24.6	25.8	23.3
	Dimer	22.5	23.9	25.7	24.2	27.9	27.4	25.6	25.6
Orientation (º)	Monomer+	/	/	/	/	317.3	317.3	315.1	313.0
	Monomer	170.7	160.7	127.7	143.0	231.8	239.7	249.2	238.5
	Dimer	221.7	226.1	216.1	214.7	214.8	214.8	216.0	213.1

Table 12. **Average inter-helical interactions (kcal/mol) in the IMM1-GC model calculated from three independent runs of 1-ns MD simulations.** GpA15p11 is used as the reference to calculate Δinter-helical interactions.

	Inter-helical Interactions	ΔInter-helical Interactions
GpA29	-21.2±0.3	+6.5±0.6
GpA15p11	-27.7±0.5	/
MCP	-22.3±0.4	+5.4±0.6
MCP-GpA	-25.7±0.3	+2.0±0.6
GpA29 (13 residues)	-16.8±0.2	+1.3±0.2
GpA15p11 (13 residues)	-18.1±0.1	/
MCP (13 residues)	-17.4±0.9	+0.7±0.9
MCP-GpA (13 residues)	-20.4±0.3	-2.3±0.3
GpA15p11 (Asn1-Arg2)	-2.7±0.7	/
MCP (Asn1-Arg2)	0.0	+2.7±0.7
MCP-GpA (Asn1-Arg2)	0.0	+2.7±0.7

*Note: Error bars are standard deviations of the mean from three separate runs.

Table 13. **Standard free energies (kcal/mol) upon association in the IMM1-GC model at 1M (in HP) standard state.** The temperature T is 298.15 K. $\Delta G_{deprotonation}$ is twice that in Table 10 on account of two monomers.

	GpA29	GpA15p11	MCP	MCP-GpA
$\Delta G_{deprotonation}$	/	/	4.2±1.1	4.4±1.6
ΔW_{ass}	-17.1±0.8	-22.4±1.5	-17.4±1.9	-24.6±2.5
$\Delta S_{ass}^{trans} \bullet T$	-2.2	-2.0	-2.1	-2.5
$\Delta S_{ass}^{rot} \bullet T$	-2.6	-2.5	-3.0	-2.8
$\Delta S_{ass}^{conf} \bullet T$	-5.9	-4.9	-3.2	-3.4
ΔG^0 at 1M (in HP) standard state	-6.5±0.8	-13.0±1.5	-4.9±2.2	-11.4±3.0

*Note: Error bars are standard deviations of the mean from three separate runs.

Figure 1. The polar coordinate system on the x-y plane of the membrane. The two helices in a dimer are presented as two cylinders. Points A and B are the center of the reference helix A and the center of the moving helix B, respectively. Point C represents a specific Cα atom on the reference helix. r is the distance between points A and B and θ is the angle between AB and AC.

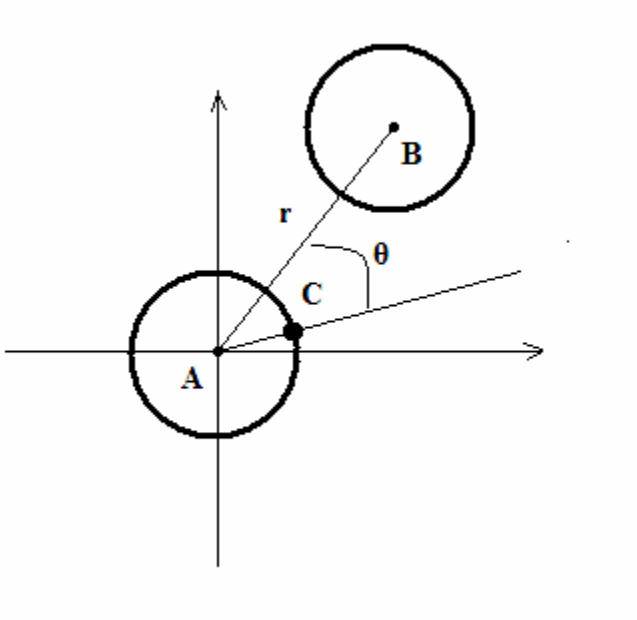


Figure 1

Figure 2. Three Euler angles (A) and the reference state (B) in the rotational entropy calculations. The reference state is defined as follows: a) the center of 21 α carbon atoms on residues from THR74 to GLY94 (from CA 5 to 25) is at the origin (0, 0, 0); b) the center of 7 α carbon atoms on residues from THR74 to VAL80 (from CA 5 to 11) is on the +x axis; c) the α carbon atom on residue VAL84 (CA 15) is on the x-y plane and its coordinate on the positive y axis.

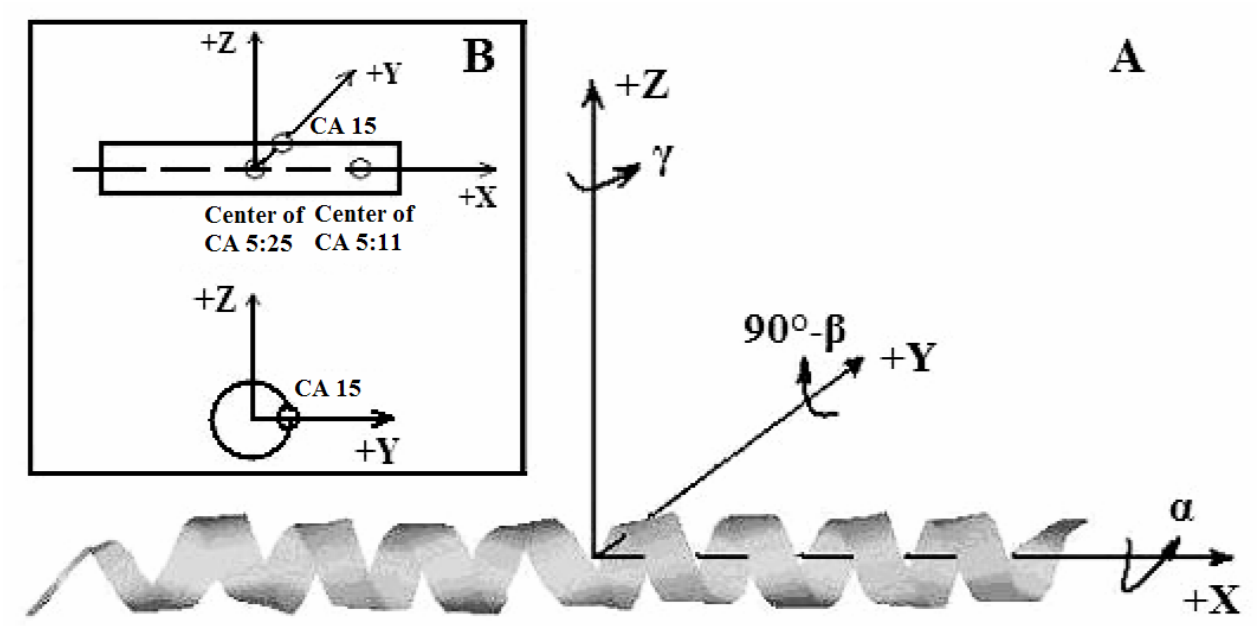


Figure 2

Figure 3. Difference in GpA association in a lipid bilayer (A) and detergent micelles (B).

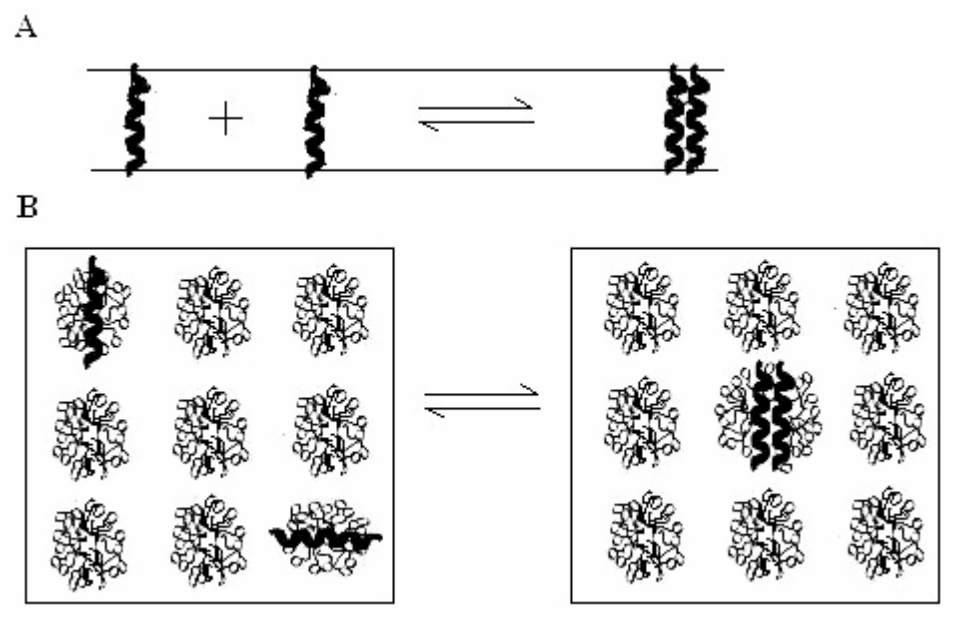


Figure 3

Figure 4. Translation of GpA monomer inside a micelle. To simplify the problem, a GpA helix and the hydrophobic core of a detergent micelle are presented as a cylinder and a sphere, respectively, and the helical axis is parallel to the z-axis (only the cross section is shown). Points O and C are the origin of the coordinate system, and the center of the helix, respectively. Points B1 and B2 are the crossing points of the helical axis and the micelle surface. Points D, A1, and A2 are projections of points C, B1, and B2 on the z-axis, respectively. CD equals the distance of the center of a GpA monomer to the origin on the x-y plane, r^m. OB1 and OB2 denote the radius of the micelle sphere. The radius of the hydrophobic core of the spherical micelle is arbitrarily defined as 11.5 Å since we are using the IMM1 with thickness of 23 Å to mimic the hydrophobic environment of micelles. The thickness of the hydrophobic phase (T) at a distance of r^m is A1A2 (A1A2=OA1+OA2=2·OA1=2·OA2), thus $T=2\cdot\sqrt{(11.5)^2-(r^m)^2}$.

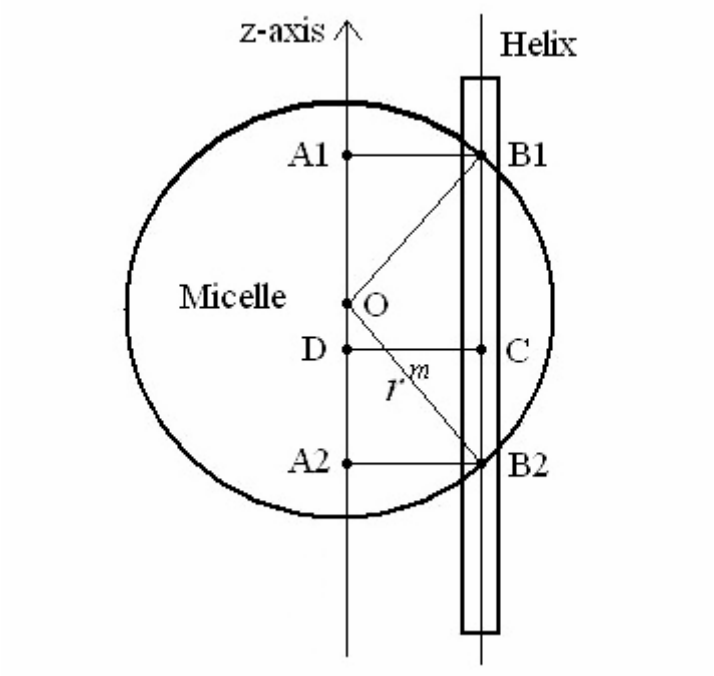


Figure 4

Figure 5. RMSD of GpA monomers (the grey and solid line for monomer A and the grey and dashed line for monomer B) and dimer (the black and solid line) during 1ns MD simulations for run #1.

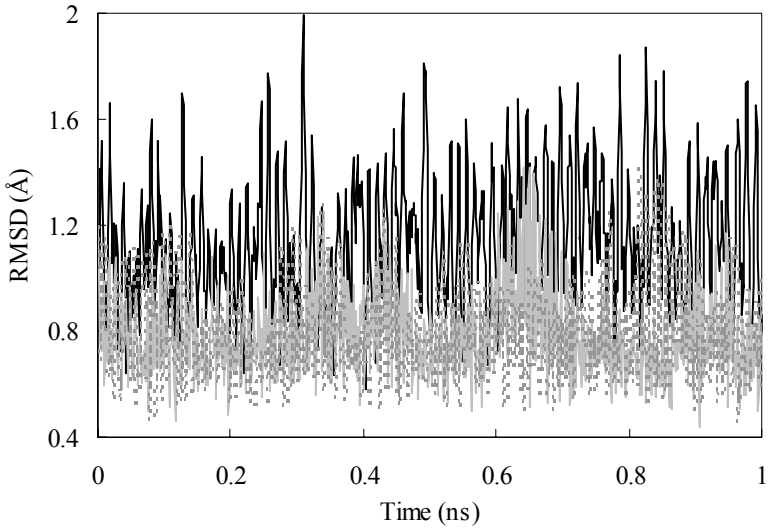


Figure 5

Figure 6. Change in configuration of GpA upon association.

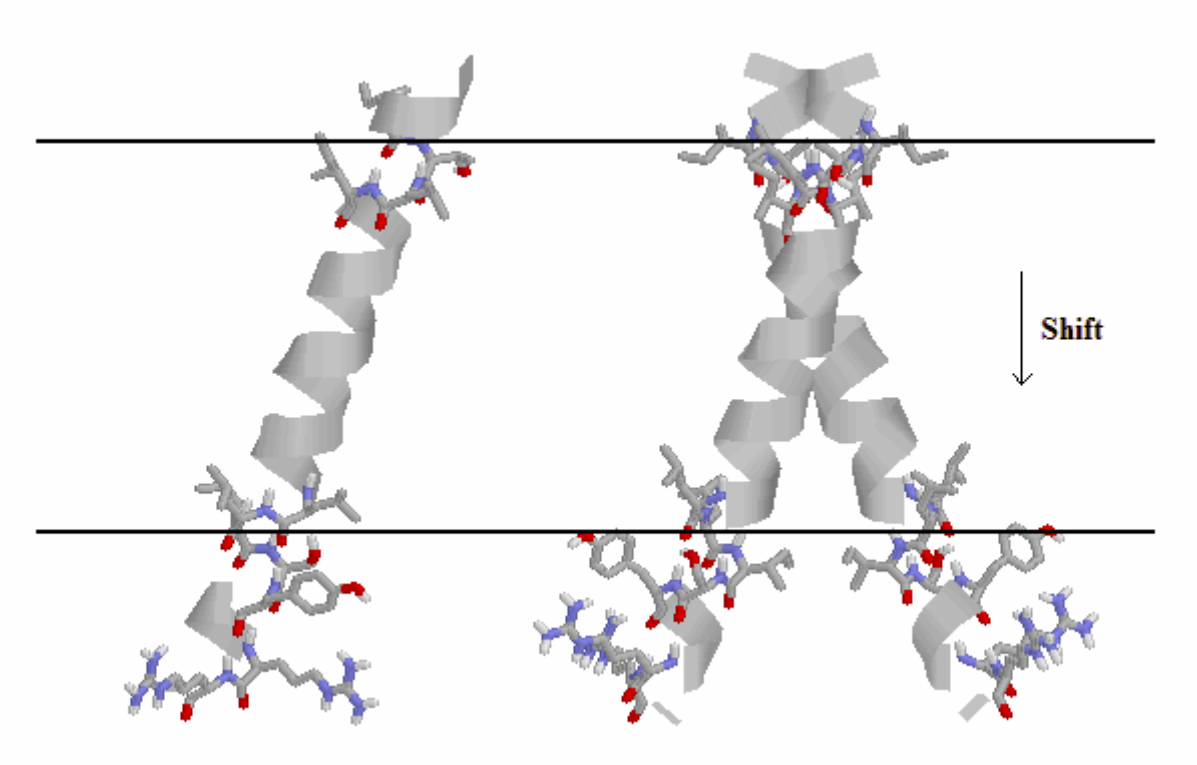
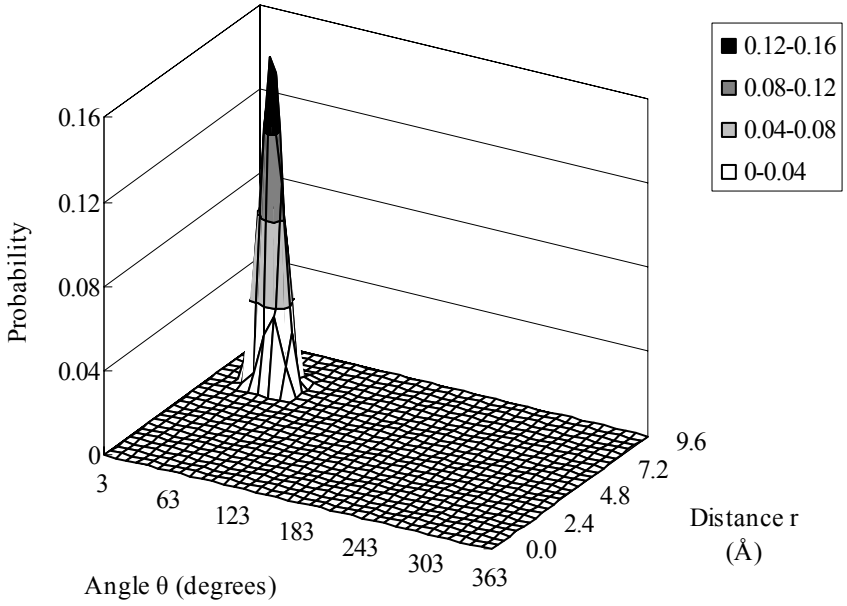


Figure 6

Figure 7. Joint probability distribution of the relative distance r between two helices and the angle θ in a dimer on the x-y plane (A) and joint probability distribution of the z coordinates of helix A and helix B in a dimer (B). The bin size for distance or z is 0.4 Å and the bin size for θ angle is 10º.

A



B

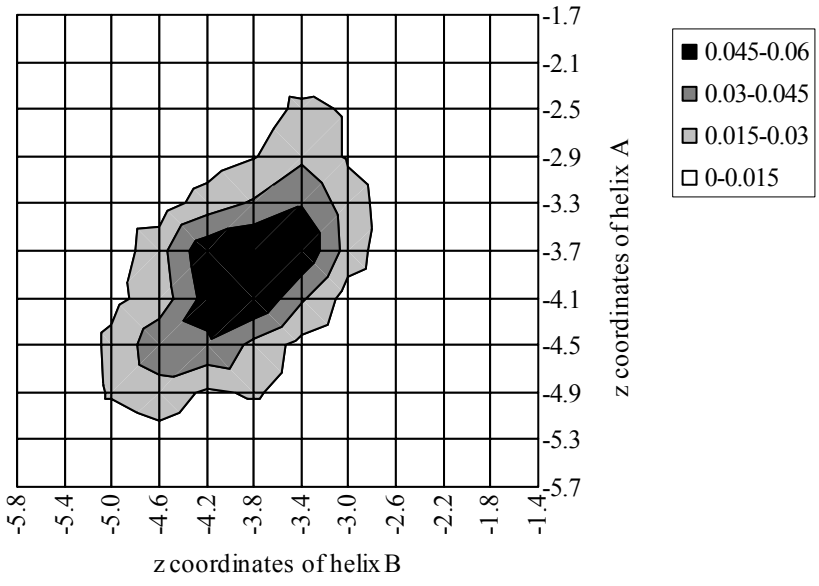


Figure 7

Figure 8. The probability distribution of the γ angle (the difference between helix A and B in a dimer). The bin size is 18°.

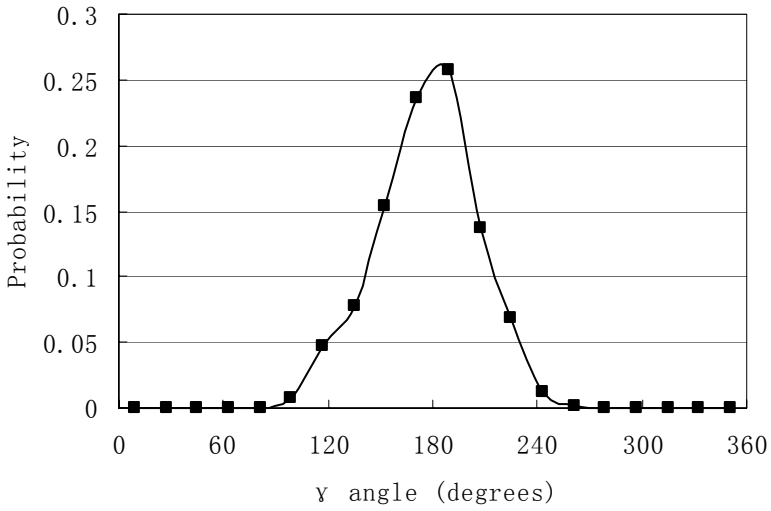
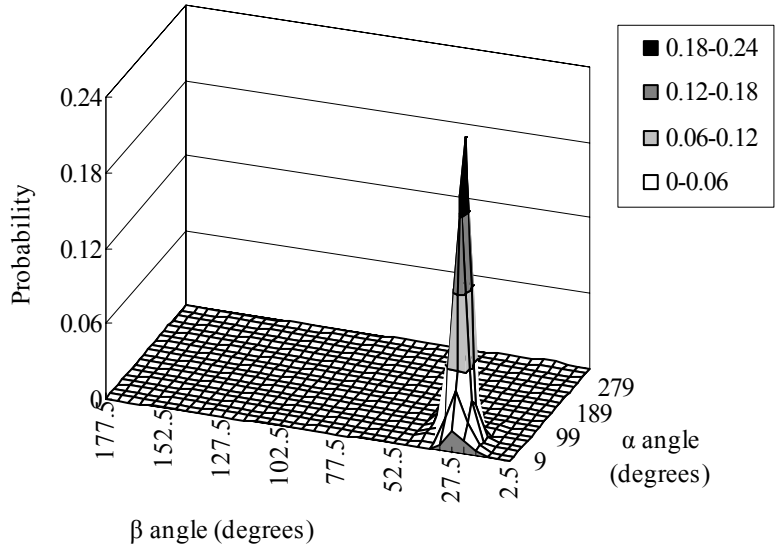


Figure 8

Figure 9. The joint probability distribution of α angle and β angle of helix A as a monomer (figure A) and in a dimer (figure B). The bin size is 18° and 5° for α angle and β angle, respectively.

A



B

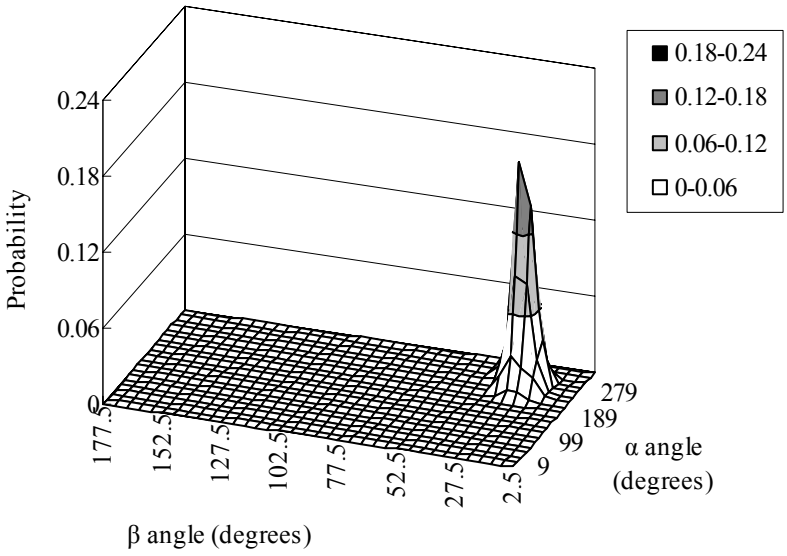


Figure 9

Figure 10. The initial dimer structures of GpA15p11 and MCP. Two glycines in GxxxG motif are shown in 'Balls'.

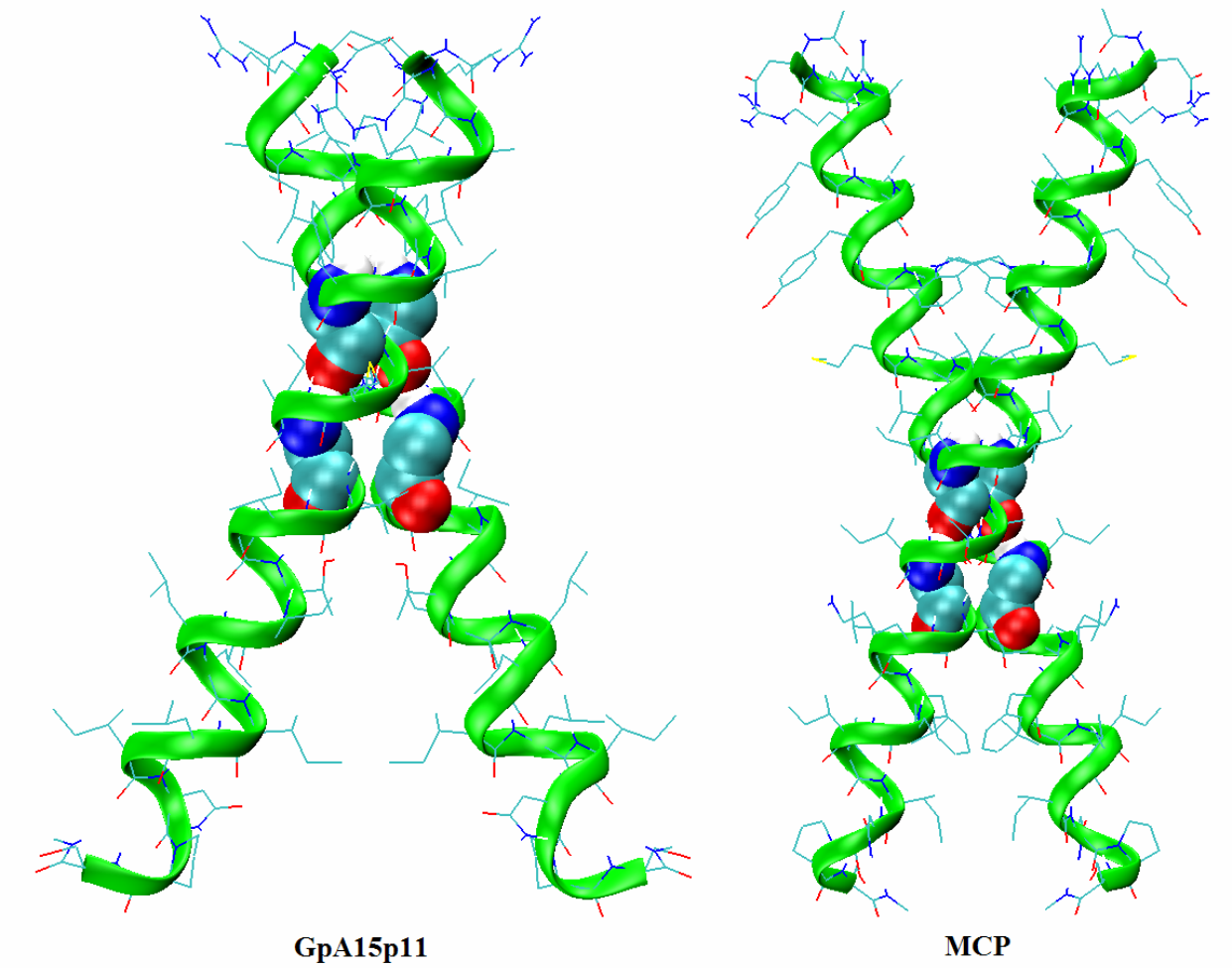


Figure 10

Figure 11. Change of configuration of MCP wild type upon association. Residues Tyr24 and Lys40 are shown in 'Sticks'.

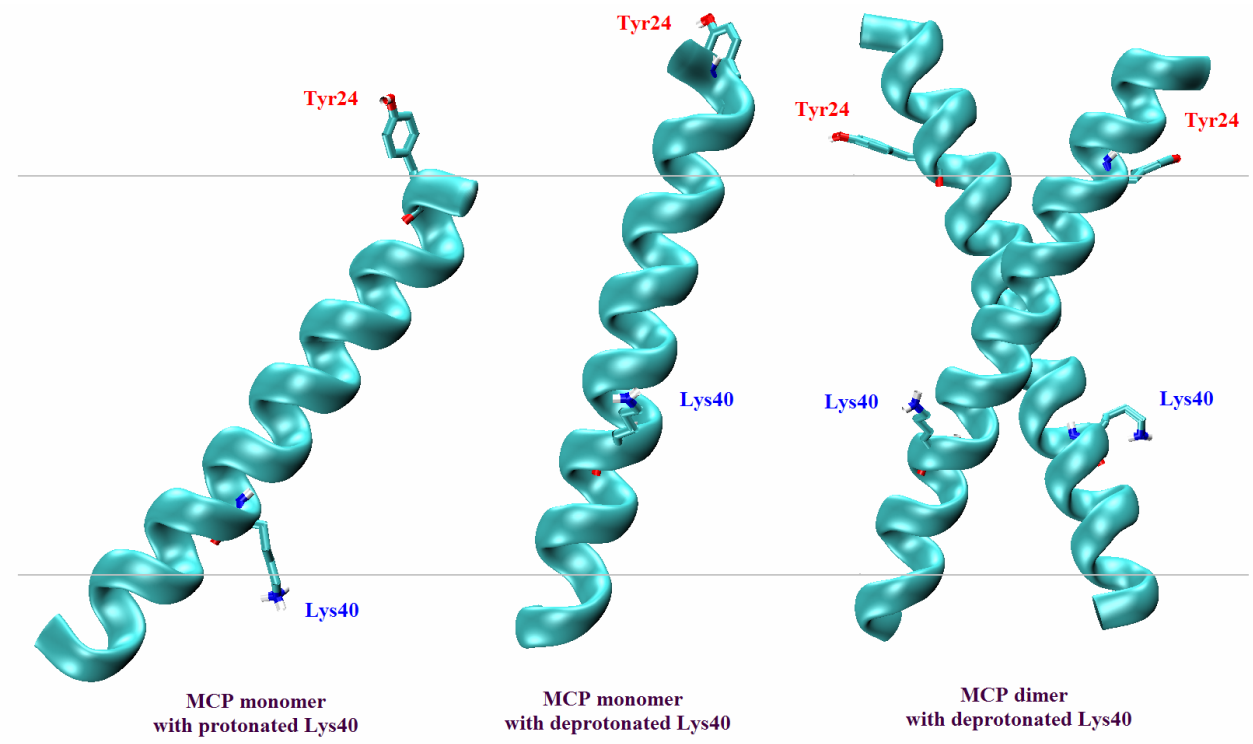


Figure 11

Figure 12. Distance between Asn and Arg from the N-flanks of GpA15p11 and MCP dimers. Residues Asn and Arg from the N-flanks are shown in 'Sticks'.

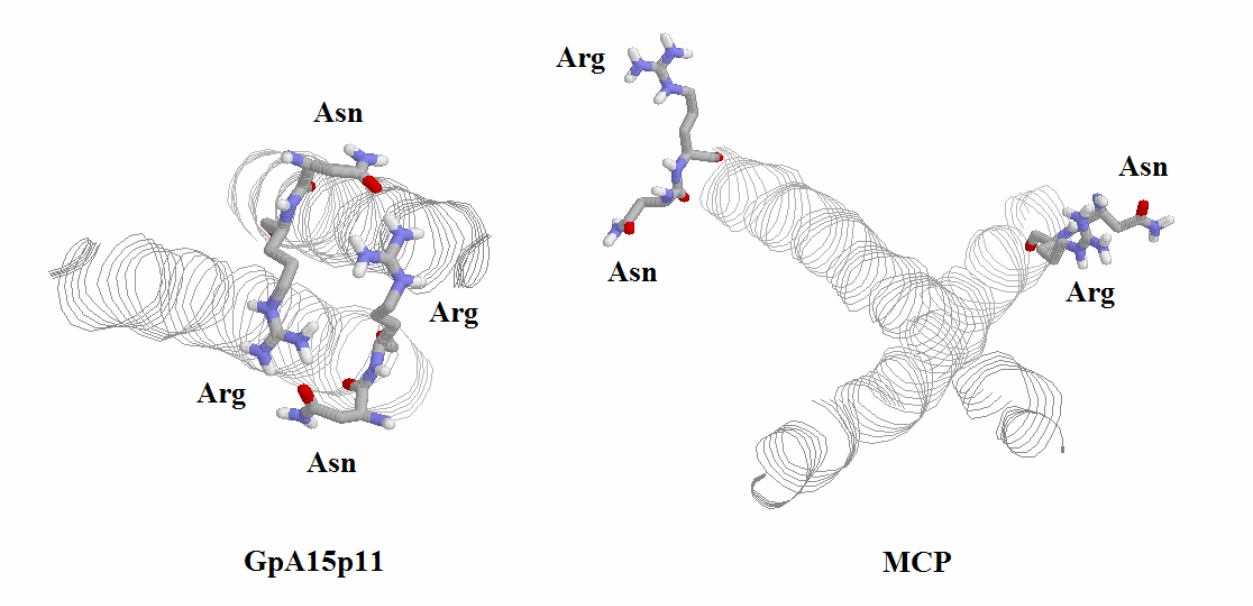


Figure 12

REFERENCES

Arbely, E., and Arkin, I.T. 2004. Experimental measurement of the strength of a C alpha-H...O bond in a lipid bilayer. *J Am Chem Soc* **126:** 5362-5363.

Beevers, A.J., and Kukol, A. 2006. Systematic molecular dynamics searching in a lipid bilayer: application to the glycophorin A and oncogenic ErbB-2 transmembrane domains. *J Mol Graph Model* **25:** 226-233.

Bond, P.J., and Sansom, M.S. 2006. Insertion and assembly of membrane proteins via simulation. *J Am Chem Soc* **128:** 2697-2704.

Braun, R., Engelman, D.M., and Schulten, K. 2004. Molecular Dynamics Simulations of Micelle Formation around Dimeric Glycophorin A Transmembrane Helices. *Biophys J* **87:** 754-763.

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 Comput 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.

Bu, L., Im, W., and Brooks, C.L., 3rd. 2007. Membrane assembly of simple helix homo-oligomers studied via molecular dynamics simulations. *Biophys J* **92:** 854-863.

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.

Cristian, L., Lear, J.D., and DeGrado, W.F. 2003. Use of thiol-disulfide equilibria to measure the energetics of assembly of transmembrane helices in phospholipid bilayers. *Proc. Natl. Acad. Sci. U S A.* **100:** 14772-14777.

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.

de Planque, M.R., and Killian, J.A. 2003. Protein-lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring. *Mol. Membr. Biol.* **20:** 271-284.

DeGrado, W.F., Gratkowski, H., and Lear, J.D. 2003. How do helix-helix interactions

help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles. *Protein Sci.* **12:** 647-665.

Dorairaj, S., and Allen, T.W. 2007. On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *Proc Natl Acad Sci U S A* **104:** 4943-4948.

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.

Engelman, D.M., Chen, Y., Chin, C.N., Curran, A.R., Dixon, A.M., Dupuy, A.D., Lee, A.S., Lehnert, U., Matthews, E.E., Reshetnyak, Y.K., et al. 2003. Membrane protein folding: beyond the two stage model. *FEBS Lett* **555:** 122-125.

Feller, S.E., Gawrisch, K., and Woolf, T.B. 2003. Rhodopsin exhibits a preference for solvation by polyunsaturated docosohexaenoic acid. *J Am Chem Soc* **125:** 4434-4435.

Finger, C., Volkmer, T., Prodohl, A., Otzen, D.E., Engelman, D.M., and Schneider, D. 2006. The stability of transmembrane helix interactions measured in a biological membrane. *J Mol Biol* **358:** 1221-1228.

Fisher, L.E., Engelman, D.M., and Sturgis, J.N. 1999. Detergents modulate dimerization, but not helicity, of the glycophorin A transmembrane domain. *J Mol Biol* **293:** 639-651.

Fisher, L.E., Engelman, D.M., and Sturgis, J.N. 2003. Effect of detergents on the association of the glycophorin a transmembrane helix. *Biophys J* **85:** 3097-3105.

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 alpha-helix dimerization. *J Mol Biol* **272:** 266-275.

Fleming, K.G., and Engelman, D.M. 2001. 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.

Fleming, K.G., Ren, C.C., Doura, A.K., Eisley, M.E., Kobus, F.J., and Stanley, A.M. 2004. Thermodynamics of glycophorin A transmembrane helix dimerization in C14 betaine micelles. *Biophys Chem* **108:** 43-49.

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.

Froloff, N., Windemuth, A., and Honig, B. 1997. On the calculation of binding free energies using continuum methods: application to MHC class I protein-peptide interactions. *Protein Sci.* **6:** 1293-1301.

Gilson, M.K., Given, J.A., Bush, B.L., and McCammon, J.A. 1997. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **72:** 1047-1069.

Grasberger, B., Minton, A.P., DeLisi, C., and Metzger, H. 1986. Interaction between proteins localized in membranes. *Proc. Natl. Acad. Sci. U S A.* **83:** 6258-6262.

Heinrich, V., Ritchie, K., Mohandas, N., and Evans, E. 2001. Elastic thickness compressibilty of the red cell membrane. *Biophys J* **81:** 1452-1463.

Helms, V., and Wade, R.C. 1998. Computational alchemy to calculate absolute protein-ligand binding free energy. *J. Am. Chem. Soc.* **120:** 2710-2713.

Henchman, R.H. 2003. Partition function for a simple liquid using cell theory parametrized by computer simulation. *J. Chem. Phys.* **119:** 400-406.

Henin, J., Pohorille, A., and Chipot, C. 2005. Insights into the recognition and association of transmembrane alpha-helices. The free energy of alpha-helix dimerization in glycophorin A. *J. Am. Chem. Soc.* **127:** 8478-8484.

Horton, N., and Lewis, M. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci.* **1:** 169-181.

Johnson, R.M., Rath, A., and Deber, C.M. 2006. The position of the Gly-xxx-Gly motif in transmembrane segments modulates dimer affinity. *Biochem Cell Biol* **84:** 1006-1012.

Kim, S., Chamberlain, A.K., and Bowie, J.U. 2003. A simple method for modeling transmembrane helix oligomers. *J. Mol. Biol.* **329:** 831-840.

Kim, S., Jeon, T.J., Oberai, A., Yang, D., Schmidt, J.J., and Bowie, J.U. 2005. Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proc Natl Acad Sci U S A* **102:** 14278-14283.

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.

Lazaridis, T. 2002. Binding affinity and specificity from computational studies. *Current Organic Chemistry.* **6:** 1319-1332.

Lazaridis, T. 2003. Effective energy function for proteins in lipid membranes. *Proteins* **52:** 176-192.

Lazaridis, T. 2005. Implicit solvent simulations of peptide interactions with anionic lipid membranes. *Proteins* **58:** 518-527.

Lazaridis, T., and Karplus, M. 1999. Effective energy function for proteins in solution. *Proteins* **35:** 133-152.

Lazaridis, T., Masunov, A., and Gandolfo, F. 2002. Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins.* **47:** 194-208.

le Maire, M., Champeil, P., and Moller, J.V. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim. Biophys. Acta.* **1508:** 86-111.

Lehninger, A.L., Nelson, D.L., and Cox, M.M. 1993. Principle of Biochemistry, Second Edition, Worth Publishers.

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. 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.

Lew, S., Caputo, G.A., and London, E. 2003. The effect of interactions involving ionizable residues flanking membrane-inserted hydrophobic helices upon helix-helix interaction. *Biochemistry* **42:** 10833-10842.

Lew, S., Ren, J., and London, E. 2000. The effects of polar and/or ionizable residues in the core and flanking regions of hydrophobic helices on transmembrane conformation and oligomerization. *Biochemistry* **39:** 9632-9640.

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.

Lomize, A.L., Pogozheva, I.D., and Mosberg, H.I. 2004. Quantification of helix-helix

binding affinities in micelles and lipid bilayers. *Protein Sci* **13:** 2600-2612.

MacKenzie, K.R., Prestegard, J.H., and Engelman, D.M. 1997. A transmembrane helix dimer: structure and implications. *Science* **276:** 131-133.

Matthews, E.E., Zoonens, M., and Engelman, D.M. 2006. Dynamic helix interactions in transmembrane signaling. *Cell* **127:** 447-450.

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.

Melnyk, R.A., Partridge, A.W., and Deber, C.M. 2002. Transmembrane domain mediated self-assembly of major coat protein subunits from Ff bacteriophage. *J Mol Biol* **315:** 63-72.

Mihajlovic, M., and Lazaridis, T. 2006. Calculations of pH-dependent binding of proteins to biological membranes. *J Phys Chem B* **110:** 3375-3384.

Mottamal, M., and Lazaridis, T. 2005. The contribution of C alpha-H...O hydrogen bonds to membrane protein stability depends on the position of the amide. *Biochemistry* **44:** 1607-1613.

Mottamal, M., Zhang, J., and Lazaridis, T. 2006. Energetics of the native and non-native states of the glycophorin transmembrane helix dimer. *Proteins* **62:** 996-1009.

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.

Popot, J.L., and Engelman, D.M. 1990. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29:** 4031-4037.

Russ, W.P., and Engelman, D.M. 1999. TOXCAT: a measure of transmembrane helix association in a biological membrane. *Proc Natl Acad Sci U S A* **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.

Salom, D., Hill, B.R., Lear, J.D., and DeGrado, W.F. 2000. pH-dependent tetramerization and amantadine binding of the transmembrane helix of M2 from the influenza A virus. *Biochemistry* **39:** 14160-14170.

Sanders, C.R., and Sonnichsen, F. 2006. Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* **44 Spec No:** S24-40.

Schneider, D., and Engelman, D.M. 2004. Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. *J Mol Biol* **343:** 799-804.

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. *Proc Natl Acad Sci U S A* **98:** 9056-9061.

Sesta, B. 1989. Physicochemical properties of decyldimethylammonium propanesulfonate and its homologous compounds in aqueous medium. *J. Phys. Chem.* **93:** 7677-7680.

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.

Valluru, N., Silva, F., Dhage, M., Rodriguez, G., Alloor, S.R., and Renthal, R. 2006. Transmembrane helix-helix association: relative stabilities at low pH. *Biochemistry* **45:** 4371-4377.

von Heijne, G. 2007. The membrane protein universe: what's out there and why bother? *J Intern Med* **261:** 543-557.

Vos, W.L., Koehorst, R.B., Spruijt, R.B., and Hemminga, M.A. 2005. Membrane-bound conformation of M13 major coat protein: a structure validation through FRET-derived constraints. *J Biol Chem* **280:** 38522-38527.

Yin, H., Slusky, J.S., Berger, B.W., Walters, R.S., Vilaire, G., Litvinov, R.I., Lear, J.D., Caputo, G.A., Bennett, J.S., and DeGrado, W.F. 2007. Computational design of peptides that target transmembrane helices. *Science* **315:** 1817-1822.

Zhang, J., and Lazaridis, T. 2006. Calculating the free energy of association of transmembrane helices. *Biophys J* **91:** 1710-1723.

Zhang, J., and Lazaridis, T. 2007. Transmembrane Helix Association Affinity Can Be Modulated by Flanking and Non-interfacial Residues. *Proteins* **Submitted**.

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.