doi:10.1016/j.jmb.2004.06.042 *J. Mol. Biol.* (2004) **341**, 991–998

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Sequence Context Modulates the Stability of a GxxxG-mediated Transmembrane Helix–Helix Dimer

Abigail K. Doura†, Felix J. Kobus†, Leonid Dubrovsky, Ellen Hibbard and Karen G. Fleming*

T. C. Jenkins Department of Biophysics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218 USA

**Corresponding author*

To quantify the relationship between sequence and transmembrane dimer stability, a systematic mutagenesis and thermodynamic study of the protein–protein interaction residues in the glycophorin A transmembrane helix–helix dimer was carried out. The results demonstrate that the glycophorin A transmembrane sequence dimerizes when its GxxxG motif is abolished by mutation to large aliphatic residues, suggesting that the sequence encodes an intrinsic propensity to self-associate independent of a GxxxG motif. In the presence of an intact GxxxG motif, the glycophorin A dimer stability can be modulated over a span of -0.5 kcal mol^{-1} to $+3.2$ kcal mol^{-1} by mutating the surrounding sequence context. Thus, these flanking residues play an active role in determining the transmembrane dimer stability. To assess the structural consequences of the thermodynamic effects of mutations, molecular models of mutant transmembrane domains were constructed, and a structure-based parameterization of the free energy change due to mutation was carried out. The changes in association free energy for glycophorin A mutants can be explained primarily by changes in packing interactions at the protein–protein interface. The energy cost of removing favorable van der Waals interactions was found to be 0.039 kcal mol^{-1} per A^2 of favorable occluded surface area. The value corresponds well with estimates for mutations in bacteriorhodopsin as well as for those mutations in the interiors of soluble proteins that create packing defects.

Keywords: glycophorin A; GxxxG; ultracentrifugation; thermodynamics; membrane protein

Introduction

The relationships between sequence, structure and stability for membrane proteins are not well understood. While hydrophobicity can be used to predict the occurrence of transmembrane α-helices in open reading frames,[1,2] little is known about the principles that govern the subsequent interactions between helices leading to the formation of native membrane protein structures. In recent years, a pattern of two glycine residues separated by three intervening residues, known as a GxxxG (or GG4) motif, has emerged as a characteristic signature of transmembrane α-helix dimerization. This motif was identified by a statistical analysis of pair-wise amino acid patterns in putative transmembrane helix sequences from SwissProt, revealing that the GG4 motif occurs at a frequency ~32% higher than the randomly expected value.[3] In an accompanying study, a genetic screen designed to select transmembrane domains with high-affinity homo-oligomerization properties from a randomized sequence library contained a high occurrence of the GG4 motif.[4] Indeed, the importance of the GxxxG motif in the dimerization of the transmembrane protein, glycophorin A, had been recognized for several years.[5–8] In addition, the introduction of a GxxxG motif in a background of a polyvaline or polymethionine transmembrane sequence has been shown to enhance oligomerization.[6] Thus, the functional role for GxxxG was postulated as a

† A.K.D. and F.J.K. authors contributed equally to this work.

Abbreviations used: GpATM, glycophorin A transmembrane.

E-mail address of the corresponding author: karen.fleming@jhu.edu

motif to drive transmembrane helix–helix dimerization. A GG4 motif has subsequently been shown to be involved in the oligomerization of the erbB tyrosine kinase receptors,[9] the yeast ATP synthase,[10] the *Helicobacter pylori* vacuolating toxin,[11,12] the yeast α-factor receptor,[13] and assembly of the γ-secretase complex.[14]

Insight into the role that the GxxxG motif plays in transmembrane helix–helix dimerization is derived mainly from studies on the glycophorin A transmembrane (GpATM) dimer. From biochemical data as well as structural studies, the GxxxG motif in glycophorin A is part of a larger protein–protein interaction motif: L75I76xxG79V80xxG83V84xxT87,[5,7,8,15] where the GxxxG motif glycine residues are Gly79 and Gly83. Inspection of the NMR structure for the GpATM dimer suggests that the glycine residues allow a close approach of the helices providing a smooth surface for packing interactions.[15] It has been suggested that the close approach of the glycine residues stabilizes the helix dimer by allowing the formation of an interhelical $C^{\alpha}H\cdots O$ hydrogen bond.[16] In contrast, molecular dynamics calculations suggest that the interaction energy at the glycine residues in the GpATM dimer is an unfavorable electrostatic contact.[17] Nevertheless, the pronounced destabilizing effects of mutations at the glycine residues[5–8] provide evidence that the two glycine residues are extremely important for strong dimerization. Surprisingly, previous sedimentation equilibrium experiments demonstrated that sequences containing mutations to alanine at the GxxxG motif still retain a significant propensity to dimerize.[5] To address this apparent paradox, we hypothesized that the sequence context surrounding the GxxxG motif must play an active role in specifying and stabilizing the GpATM helix–helix interaction. A systematic thermodynamic study to determine the free energies of association of GpATM sequences containing point mutations to other hydrophobic residues was conducted to address this question.

Results

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

A systematic mutagenesis and thermodynamic study was carried out to determine how the sequence context surrounding the GxxxG motif influences the stability of the glycophorin A transmembrane dimer. Single-point mutations to Gly, Ile, Leu, and Val were introduced at each of the protein–protein contact positions in the glycophorin A transmembrane domain. In addition, to address the potential for interhelical hydrogen bonding at position Thr87 by thermodynamic methods, a Ser mutation was constructed at this site. The effect of each of these mutations on the free energy of association was determined using sedimentation equilibrium analytical ultracentrifugation under conditions that measure reversible interactions between membrane proteins in micellar solutions.[18,19] Previously, the dimer stabilities measured by sedimentation equilibrium on a series of alanine mutants of the GpATM have been shown to scale linearly with the apparent stabilities obtained using an *in vivo* bacterial genetic assay.[5]

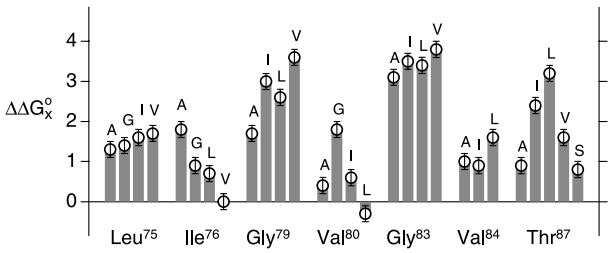

Figure 1. The free energy perturbations due to mutations at the protein–protein interface of the GpATM. Shown are the $\Delta\Delta G^{\circ}_{Mut}$ values for single-point mutations at the protein–protein interaction surface of the glycophorin A transmembrane dimer. The interaction residues are given in three-letter code along the abscissa, and the identities of the single-point mutations are given in one-letter code above each energy value. $\Delta\Delta G^{\circ}_{Mut}$ values were calculated as described in Experimental Procedures. The error bars represent the standard deviations of the values and were propagated as described by Bevington.[37] For comparison the $\Delta\Delta G^{\circ}_{Mut}$ values of the previously published alanine mutants[5] are included.

The free energy perturbation due to mutation was quantified by subtracting the free energy of association of the wild-type from that of the mutant. These $\Delta\Delta G^{\circ}_{Mut}$ values are shown in Figure 1. Overwhelmingly, mutations at the glycophorin A dimer interface lead to a loss of dimer stability. Only Ile76Val and Val80Leu dimerize as strongly as the wild-type sequence. In contrast to the high frequency of stabilizing mutations observed for other transmembrane proteins,[20–22] no single-point mutation was found to significantly stabilize the dimer. As expected, all mutations at the two glycine residues in the GxxxG motif show strong destabilizing effects, yet these sequences retain a significant propensity to dimerize. One of the flanking residues, Val80, is relatively insensitive when mutated to another nonpolar residue, whereas the other flanking residue, Val84, shows greater sensitivity. Mutations at distant sites on both ends of the helix from the GxxxG motif exhibit a wide variation of perturbations to the free energy of association. Accordingly, the free energy perturbation is not equal for all residue types. For example, the cost of a single-point mutation to valine can range from 0 (at Ile76) to ~1.6 (at Leu75 and Thr87) to ~3.8 kcal mol^{-1} (at Gly79 and Gly83) (1 cal = 4.184 J).

All mutant sequences dimerize, but to different extents

The propensity for all mutants to form dimers at

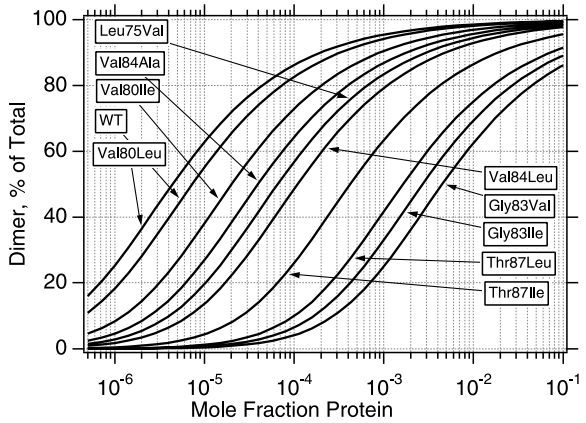

Figure 2. Equilibrium distributions of glycophorin A transmembrane dimer mutants. The percentage dimer is given as a function of the total mole fraction of protein in the micellar detergent phase. The distributions were calculated using the experimentally determined ΔG°_{x} values.

experimentally detectable concentrations enabled the determination of association free energy values for all sequence permutations in this study. The experimental free energy values were used to calculate the oligomeric species populations over a wide range of protein concentrations. The conditions under which proteins can be characterized as "monomeric" or "dimeric" can then be determined. This result is visualized in Figure 2, which shows the population distribution of dimeric species for a subset of the mutants as a function of protein:detergent mole fraction. At the high mole fraction of 10^{-2}, the thermodynamic data show that even the most destabilizing mutants, Gly83Ile, Gly83Val and Thr87Leu, are >60% dimeric. While not shown in the Figure, this is true for all mutants with $\Delta\Delta G^{\circ}_{Mut}$ values >2.5 kcal mol^{-1}. Many of these extremely disruptive mutations abolish the GxxxG motif, demonstrating that the GpATM sequence can dimerize considerably in the absence of an intact GG4. The population of these destabilizing mutants becomes >90% monomeric only when the concentration of protein in the micellar detergent is decreased 100-fold (e.g. 10^{-4}). At this lower mole fraction, only the wild-type and the Val80Leu sequences are >90% dimeric; all other sequences have mixed populations of monomer and dimer at this concentration. The free energy data demonstrate that the populations of monomer and dimer can be altered significantly by mutations at any of the protein–protein interaction sites along the glycophorin A transmembrane sequence.

Structural consequences of mutations

To rationalize the modulation of dimer stability by sequence context, structural models were generated for each mutant, and a model-based parameterization of the free energy perturbation was carried out. The modeling procedure used was designed to minimize local steric clash and maximize local van der Waals interactions by allowing changes in side-chain conformations throughout the structure. No backbone rearrangements were considered in this procedure. The computational models were used to calculate the change in favorable occluded surface area (Δ*FOS*), unfavorable occluded surface area (Δ*UOS*) and side-chain conformational entropy ($\Delta\Delta S_{SC}$) for all residues for each of the mutants relative to the wild-type structure. Favorable occluded surface area is interpreted as an indication of favorable van der Waals packing interactions. Unfavorable occluded surface area is interpreted as the presence of a steric clash. Occluded surface area was scored as unfavorable when the molecular surfaces of two atoms overlapped. The conformational heterogeneity present in the family of NMR structures was accommodated by calculating mean and standard deviation values for each of the parameters for each mutant–wild-type pair of the 20 NMR structures.

Coefficients for mean Δ*FOS* and Δ*UOS* parameters were determined by least-squares linear regression analysis against the experimental $\Delta\Delta G^{\circ}_{Mut}$ values. After the initial parameterization, eight mutants with Δ*UOS* values of ≥50 Å^2 were disproportionately influencing the fit. These were

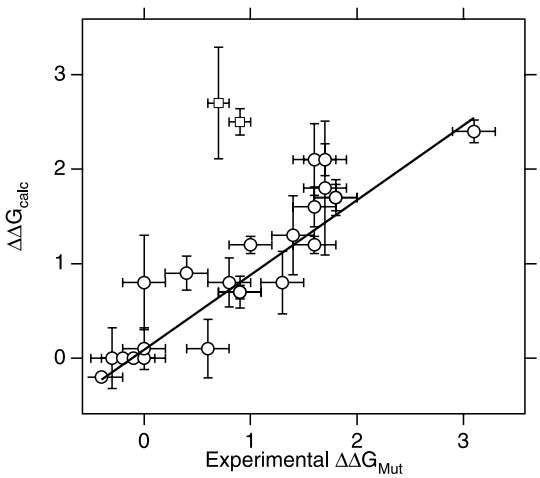

Figure 3. A linear correlation between parameterized structural features and free energy perturbation. The parameterized structure-based free energy perturbation is plotted *versus* the experimentally measured $\Delta\Delta G^{\circ}_{Mut}$. The linear correlation ($R = 0.92$, $R_{jackknife} = 0.89$) has a slope of 0.79. Open circles show the 23 mutants that were used in this parameterization. Open squares represent the application of the coefficients derived from the linear regression to the Ile76Gly and Ile76Leu mutants. These two mutants could not be described by this analysis, even though their modeled structures appear to be chemically reasonable. The error bars in the vertical direction represent the standard deviation of the structural features amongst the 20 models of the glycophorin A NMR structure. The error bars in the horizontal direction represent the experimental standard deviation of the free energy measurement.

all leucine and isoleucine mutants at Gly79, Gly83 and Thr87 as well as valine mutations at Gly79 and Gly83. Inspection of these mutant structures suggested that the perturbations generated in the automatic modeling protocol resulted in models containing extremely large steric clashes. Since the thermodynamic data demonstrate that all of these mutants do in fact dimerize, such steric clashes would presumably be relieved *in vivo*. We hypothesize that structural relaxation would require backbone rearrangements and/or helix rotation or bending to form an alternative interface. However, computational modeling of such structures is beyond the scope of this study, and these mutants were eliminated from further consideration. Figure 3 shows the linear correlation between the coefficient-weighted structure-based parameters and the experimentally measured free energy perturbation for the remaining mutants with the exception of two outliers. These outliers (Gly and Leu mutations at Ile76) are described poorly by this automated modeling procedure, even though their structural models appear to be chemically reasonable and lack large steric clashes. The coefficient value for the (ΔFOS parameter was found to be 0.039, which suggests that the free energy cost to remove 1 $Å^2$ of favorable occluded surface area is 39 cal mol^{-1}.

In contrast to the buried surface area calculation, the use of the occluded surface algorithm facilitates a comparison of the nature of the structural perturbations at each residue. The energetic contributions due to changes in favorable and unfavorable van der Waals interactions and side-chain conformational entropy can be calculated from each parameter using the fitted coefficient values. Figure 4 shows the energetic contributions for each of the mutants used in the linear parameterization. Changes in the favorable occluded surface area dominate the energetic effects. In general, the loss of favorable van der Waals interactions occurs at positions where large to small mutations were introduced (Leu75Ala, Leu75Gly, Ile76Ala, Ile76Gly, Ile76Val Val80Ala, Val80Gly, Val84Ala, Val84Ile, Val84Leu, Thr87Ala and Thr87Ser). At sites where small to large mutations were introduced, the results are varied. At Gly79Ala, Gly83Ala and Thr87Val the destabilization of the dimer is primarily due to the introduction of large steric clashes. For the Val84Ile and Val84Leu mutants a loss of favorable occluded surface area is observed. A number of small to large mutant sites show a mixed contribution from the loss of favorable OS and the introduction of steric clash (Leu75Ile, Leu75Val, Ile76Leu, Thr87Val). The contribution from side-chain conformational entropy appears to be less important than changes in van der Waals interactions. Nevertheless, a small net stabilization due to changes in side-chain conformational entropy is observed at several sites (Leu75Ala, Leu75Gly, Ile76Ala, Ile76Val, Thr87Ala, and Thr87Val).

The structural parameterization reveals insight into the detailed nature of the optimized wild-type interface. Changes in the same structural parameters do not explain the effects of different mutations at a particular site. For instance, mutations at Leu75 result in a free energy perturbation of $\sim$1.5–1.8 kcal mol^{-1} regardless of the residue type. Structurally, these effects can be rationalized by a loss of favorable occluded surface area (Leu75Ala and Leu75Gly) or alternatively by the combination of a loss of favorable OS and the introduction of a modest steric clash (Leu75Ile and Leu75Val). Of particular interest is a consideration of the effects of point mutations at Thr87. The destabilizing effects of valine, serine and alanine are well explained by the combination of an introduction of steric clash (in the case of Val) and a loss of favorable OS (for Val, Ser and Ala). This correlation suggests that Thr87 in the wild-type sequence contributes favorable van der Waals

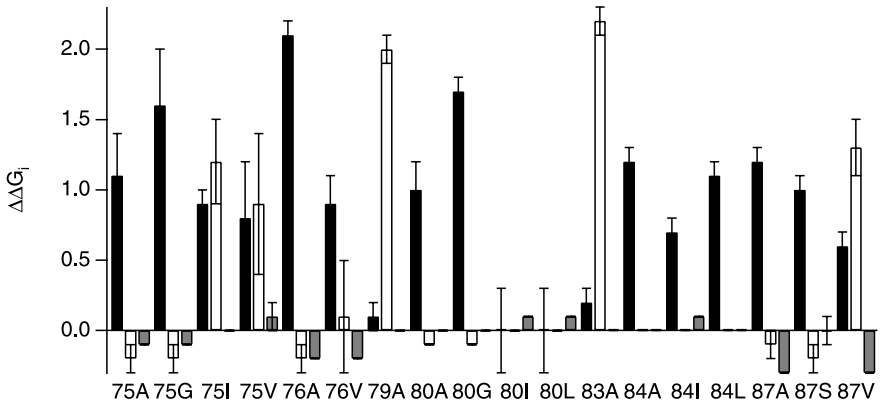

Figure 4. Contributions of each of the structure-based energy terms to the of $\Delta\Delta G_{calc}$ sum. Three bars are shown for each mutant. The black bars represent energy values due to changes in favorable occluded surface area. The white bars show energy contributions due to changes in unfavorable occluded surface area. The grey scale bars show contributions due to changes in the side-chain entropy term. The error bars are propagated from the variation in the family of NMR structures.

interactions rather than forms an interhelical hydrogen bond in C_8E_5 micelles. This finding is not inconsistent with solid-state NMR measurements, where an interhelical hydrogen bond was observed,[23,24] since the detergent micelle environment may be a more hydrated environment than the lipid bilayer in which the solid state NMR experiments were carried out.

Discussion

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

The energetic consequences of single-point mutations at Gly79 and Gly83 are consistent with the importance of closely packed glycine residues as one defining structural feature of the GpA transmembrane dimer. Mutations at the glycine residues were previously shown to result in complete disruption of the dimer as measured by SDS-PAGE.[8] In addition, the Gly83Ile sequence has a low signal and is normally considered to be a "monomeric" standard in the TOXCAT assay for helix–helix interaction in bacterial membranes.[25] Indeed, the glycine residues at these sites in the glycophorin A transmembrane domain are required for the strongest dimerization of the helices to occur. However, this thermodynamic study demonstrates that the glycophorin A transmembrane domain still dimerizes when the GxxxG motif is disrupted by mutation to alanine, valine, leucine or isoleucine. Such self-interactions must be encoded by the remaining residues at the protein–protein interface. Mutations in the GpATM GxxxG motif do not result in uncontrolled protein aggregation, but rather in a defined oligomeric species. This suggests that these non-GxxxG mediated transmembrane helix–helix interactions are still specific, albeit less stable. Additional mutagenesis studies could be used to test this idea and to determine how the interface of the helix is responsible for the dimerization.

The dimerization potential of these weaker mutants was revealed only by these sedimentation equilibrium studies. The access to association free energy values for the weaker mutants reflects the ability of sedimentation equilibrium to access a wider range of protein concentrations than other techniques used to assess oligomeric states of membrane proteins. Due to the existence of a conserved hierarchy of stability for the GpATM alanine sequence variants between C_8E_5 micelles, SDS micelles, and bacterial membranes,[5] we hypothesize that mutant sequences with lesser propensities to self-associate should still populate dimeric forms in biological membranes. The reduced dimensionality constraints in membranes compared to micelles is expected to further stabilize interactions of transmembrane sequences.[26]

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

Mutations in the protein–protein interaction sequence flanking the GxxxG motif in glycophorin A demonstrate that the free energy of association can be modulated over a $\Delta\Delta G_{Mut}$ span of -0.5 kcal mol^{-1} to $+3.2$ kcal mol^{-1} in the presence of a GxxxG motif. At 25 °C, this free energy range corresponds to a shift of over 1000-fold in the species population distribution. Mutations at all of the GpATM protein–protein interaction residues can significantly affect the dimer stability. At Ile76, Val80 and Thr87 mutations exhibit a variety of free energies of association that are stabilizing, energetically neutral, or significantly destabilizing. In contrast, all mutations at Leu75 and Val84 appear to be energetically equivalent and destabilizing regardless of the residue type.

In the sequence context of the GpATM, Thr87 appears to be particularly sensitive. Mutations to leucine or isoleucine at Thr87 have energetic consequences similar to that of mutating either of the glycine residues in the GxxxG motif to large aliphatic residues. This thermodynamic result indicates that this threonine residue is particularly well suited to allow strong helix–helix interactions along with the GpATM GxxxG motif. However, the sequences of several unrelated transmembrane proteins known to interact using GxxxG lack threonine at this position, but instead have alanine, valine or leucine.[9–14] If the GpATM serves as a general model for the dimerization propensity of a GxxxG-mediated dimer, these proteins may dimerize to a lesser extent than GpATM. Alternatively, the ability of threonine to strengthen a GxxxG-mediated dimer may be dependent on the remaining sequence context.

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

The over-representation of pair-wise patterns in the statistical analysis does not always correlate with increased thermodynamic stability of helix–helix dimers. The Leu75Gly mutation introduces a second GxxxG motif in tandem to the first, creating a GG4G4 triplet motif, which is among the most over-represented triplets.[3] The introduction of this motif leads to a lower rather than a higher stability for the transmembrane dimer. As discussed above, mutations in Thr87 disrupt a GxxxGxxxT triplet motif in the GpATM. This triplet has a *p*-value consistent with a modestly significant over-representation in the statistical analysis of open reading frames.[3] Mutations at Ile76 and Val80, which constitute an over-represented IV4 motif, show opposing phenotypes, since mutations at these sites were found to be either stabilizing or destabilizing. Thus, in the absence of a defined context, statistical motifs do not necessarily

indicate a propensity for transmembrane domains to dimerize. These results suggest that statistical motifs should be studied in greater detail to determine their functional significance.

The van der Waals interactions are modulated by the sequence context

The molecular basis for the change in stability due to mutation was probed by assessing the structural consequences of mutation using computational models based on the wild-type NMR structure. The computational modeling protocol tests the hypothesis that changes in side-chain conformations can account for the loss of dimer stability. Overall, a loss of local favorable inter-monomer contacts appears to play a dominant role, suggesting that optimized van der Waals interactions specify the wild-type dimer interface. Since side-chain conformational freedom is already restricted by the initial helix formation, a small contribution from side-chain conformational entropy is observed. Although threonine pays the highest side-chain conformational entropy cost upon dimerization, changes in packing interactions dominate the energetic effects at Thr87.

The best-fit coefficient for loss of favorable occluded surface area was found to be 0.039 kcal mol^{-1} $Å^{-2}$. Since the occluded surface area calculation is based on a molecular surface, which is smaller than an accessible surface area, the coefficient 0.039 corresponds to the value of 0.018 kcal mol^{-1} $Å^{-2}$ of buried accessible surface area. This value agrees well with our earlier estimate derived from only three sequences[27] as well as with the value of 0.026 kcal mol^{-1} $Å^{-2}$ of buried surface area determined for a series of mutations in the polytopic membrane protein, bacteriorhodopsin.[22] In addition, by comparison of thermodynamic stability and crystal structures, it has been suggested that the reduction of van der Waals interactions by the introduction of cavities within the interiors of several soluble proteins has an energy cost of 0.020 kcal mol^{-1} $Å^{-2}$ of accessible surface area.[27]

Dimers containing mutations at Gly79 and Gly83 create large steric clashes that are not well explained by this computational modeling protocol. These mutations appear to require backbone rearrangements to accommodate dimer formation. Due to the close inter-helical proximity of Gly79 and Gly83, we hypothesize that mutation at these sites will lead to a disruption of interactions in the remaining protein–protein interface as a consequence of a global separation of the helices. In contrast, the structural effects caused by mutations at the non-GxxxG sites can be explained largely by local side-chain rearrangements. By analogy to the role of glycine residues in collagen, glycine residues in the GxxxG motif may simply facilitate a close approach of two helices.[29,30] The other dimer interface residues are then allowed to interact *via* significantly stronger van der Waals interactions than would be possible otherwise. The wide variation in the free energy perturbation at these sites suggests that modulation of van der Waals packing at the dimer interface by non-GxxxG residues can serve to tune the equilibrium stability to a value optimized for biological function.

Conclusions

Mutants of the GpA transmembrane helix form dimers over a wide range of stabilities. Sequences containing mutations that abolish the GxxxG motif in the glycophorin A transmembrane domain are predominantly dimeric at high protein/detergent mole fractions where structural studies are performed.[15,31] This result suggests that the sequence context surrounding the glycophorin A GxxxG motif has an independent propensity to dimerize. Those mutants containing an intact GxxxG motif show free energies of association that can vary by several kcal mol^{-1}. Thus, the GxxxG motif is necessary but not sufficient to achieve strong transmembrane helix–helix association. The context of flanking residues determines the stability of the dimer that forms. Structural parameterization of the association energetics suggests that helix–helix association is modulated largely by side-chain packing interactions at the dimer interface. A free energy cost of 0.039 kcal mol^{-1} was observed for each 1 $Å^2$ reduction in favorable inter-monomer occluded surface area. This value agrees with free energy estimates for the creation of a packing void in soluble proteins,[28] suggesting that similar principles for protein stability may be acting in both transmembrane and soluble proteins.

Experimental Procedures

Sample preparation and analytical ultracentrifugation

Single-point mutants were generated using the Stratagene Quickchange protocol with appropriate primers. All plasmids were confirmed by DNA sequencing. All mutant proteins were purified using the published protocol.[5] Immediately before sedimentation equilibrium analysis, samples were exchanged by ion-exchange chromatography into buffer containing C_8E_5 as described.[5]

Sedimentation equilibrium experiments were performed at 25 °C using a Beckman XL-A analytical ultracentrifuge as described.[5,18] The samples were centrifuged for lengths of time sufficient to achieve equilibrium. Data obtained from absorbance at 230 nm were analyzed by non-linear least-squares curve fitting of radial concentration profiles using the Windows version of NONLIN[32] using the equations describing the reversible association in sedimentation equilibrium. For each global fit, nine equilibrium data sets were collected. These consisted of three different initial protein concentrations analyzed at three rotor speeds (20,000, 24,500, 30,000 rpm) (i.e. such that the speed factor ratios were minimally 1.0, 1.5 and 2.25). The monomeric molecular

masses and partial specific volumes were calculated using the program SEDNTERP,[33] and these parameters were held constant in fitting the absorbance *versus* radius profiles.

The experimental free energy cost of mutation, $\Delta\Delta G^{\circ}_{Mut}$ equals $\Delta G^{\circ}_{Mut} - \Delta G^{\circ}_{WT}$ where $\Delta G^{\circ}_{x} = -RT \ln K_x$, and $K_x = K_{Assc,App}[\text{micellar } C_8E_5]_w$, and $K_{Assc,App}$ and $[\text{micellar } C_8E_5]_w$ are the experimentally determined monomer–dimer association constant and micellar C_8E_5 concentrations expressed on the molar aqueous scale.[18]

Computational modeling of glycophorin A mutants

The PDB coordinates (pdb1AFO.ent) containing 20 models for the solution NMR structure of the wild-type glycophorin A transmembrane dimer were used as a basis for modeling. Each model was truncated to focus on the transmembrane domain residues, 74–91. The truncated forms were minimized using CNS employing the CHARMM22 parameter set with full van der Waals radii. The minimization included the publicly available NMR constraint tables. To automatically generate point mutations, each of the minimized models was mutated using the "mutate" and "deball" commands available in WHAT IF. As a control, the minimized WT structures were subjected to the "deball" procedure. which allows rearrangement of the dimer interface at the level of changes in side-chain conformation. No backbone rearrangements or rigid body helical rotations are modeled by this procedure. In all, 660 pdb files were created and analyzed (33 mutants, 20 models each).

Structure-based calculations

Version 7.2.2 of the occluded surface algorithm, OS,[34] was used to quantify favorable (ΔFOS) and unfavorable (ΔUOS) inter-monomer contacts as well as to determine the exposed molecular surface area used in the side-chain conformational entropy calculations (ΔS_{SC}).[35,36] The change in side-chain conformational entropy of dimerization for each sequence, $\Delta\Delta S_{SC}$, was calculated as the difference between the side-chain conformational entropy of the monomers and the dimer.

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

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

$$\Delta\Delta G^{\circ}_{Calc} = \sigma\Delta FOS + \alpha\Delta UOS + (-T\Delta\Delta S_{SC}) \equiv \Delta\Delta G^{\circ}_{Mut} \quad (1)$$

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

Acknowledgements

This work was supported by a grant from the NIH (GM57534) and by a Career Award from the Department of Defense (DAMD17-02-1-0427).

References

1. Wimley, W. C., Creamer, T. P. & White, S. H. (1996). Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry*, **35**, 5109–5124.
2. Wimley, W. C. & White, S. H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* **3**, 842–848.
3. Senes, A., Gerstein, M. & Engelman, D. M. (2000). Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in associated with beta-branched residues at neighboring positions. *J. Mol. Biol.* **296**, 921–936.
4. Russ, W. P. & Engelman, D. M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.* **296**, 911–919.
5. Fleming, K. G. & Engelman, D. M. (2001). Specificity in transmembrane helix-helix interactions defines a hierarchy of stability for sequence variants. *Proc. Natl Acad. Sci. USA*, **98**, 14340–14344.
6. Brosig, B. & Langosch, D. (1998). The dimerization of the glycophorin A transmembrane segment in membranes: importance of glycine residues. *Protein Sci.* **7**, 1052–1056.
7. Lemmon, M. A., Flanagan, J. M., Hunt, J. F., Adair, B. D., Bormann, B. J., Dempsey, C. E. & Engelman, D. M. (1992). Glycophorin A dimerization is driven by specific interactions between transmembrane a-helices. *J. Biol. Chem.* **267**, 7683–7689.
8. Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J. & Engelman, D. M. (1992). Sequence specificity in the dimerization of transmembrane a-helices. *Biochemistry*, **31**, 12719–12725.
9. Mendrola, J. M., Berger, M. B., King, M. C. & Lemmon, M. A. (2002). The single transmembrane domains of ErbB receptors self-associate in cell membranes. *J. Biol. Chem.* **277**, 4704–4712.
10. Arselin, G., Giraud, M. F., Dautant, A., Vaillier, J., Brethes, D., Coulary-Salin, B. *et al.* (2003). The GxxxG motif of the transmembrane domain of subunit e is involved in the dimerization/oligomerization of the yeast ATP synthase complex in the mitochondrial membrane. *Eur. J. Biochem.* **270**, 1875–1884.
11. McClain, M. S., Cao, P. & Cover, T. L. (2001). Amino-

terminal hydrophobic region of Helicobacter pylori vacuolating cytotoxin (VacA) mediates transmembrane protein dimerization. *Infect. Immun.* **69**, 1181–1184.
12. McClain, M. S., Iwamoto, H., Cao, P., Vinion-Dubiel, A. D., Li, Y., Szabo, G. *et al.* (2003). Essential role of a GxxxG motif for membrane channel formation by Helicobacter pylori vacuolating toxin. *J. Biol. Chem.* **278**, 12101–12108.
13. Overton, M. C., Chinault, S. L. & Blumer, K. J. (2003). Oligomerization, biogenesis, and signaling is promoted by a glycophorin A-like dimerization motif in transmembrane domain 1 of a yeast G protein-coupled receptor. *J. Biol. Chem.* **278**, 49369–49377.
14. Lee, S. F., Shah, S., Yu, C., Wigley, C., Li, H. & Lim, M. *et al.* (2004). A conserved GxxxG motif in APH-1 is critical for assembly and activity of the gamma-secretase complex. *J. Biol. Chem.* **279**, 4144–4152.
15. MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. (1997). A transmembrane helix dimer: structure and implications. *Science*, **276**, 131–133.
16. Senes, A., Ubarretxena-Belandia, I. & Engelman, D. M. (2001). The Cα–H···O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl Acad. Sci. USA*, **98**, 9056–9061.
17. Petrache, H. I., Grossfield, A., MacKenzie, K. R., Engelman, D. M. & Woolf, T. B. (2000). Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: molecular dynamics calculations. *J. Mol. Biol.* **302**, 727–746.
18. Fleming, K. G. (2002). Standardizing the free energy change of transmembrane helix–helix interactions. *J. Mol. Biol.* **323**, 563–571.
19. Fleming, K. G. (1998). *Chemtracts: Biological Applications of the Analytical Ultracentrifuge* (Hanson, J. C., ed.), pp. 985–990, Springer-Verlag, New York.
20. Zhou, Y. & Bowie, J. U. (2000). Building a thermostable membrane protein. *J. Biol. Chem.* **275**, 6975–6979.
21. Bowie, J. U. (2001). Stabilizing membrane proteins. *Curr. Opin. Struct. Biol.* **11**, 397–402.
22. Faham, S., Yang, D., Bare, E., Yohannan, S., Whitelegge, J. P. & Bowie, J. U. (2004). Side-chain contributions to membrane protein structure and stability. *J. Mol. Biol.* **335**, 297–305.
23. Smith, S. O., Song, D., Shekar, S., Groesbeek, M., Ziliox, M. & Aimoto, S. (2001). Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. *Biochemistry,* **40**, 6553–6558.
24. Smith, S. O., Eilers, M., Song, D., Crocker, E., Ying, W., Groesbeek, M. *et al.* (2002). Implications of threonine hydrogen bonding in the glycophorin A transmembrane helix dimer. *Biophys. J.* **82**, 2476–2486.
25. Russ, W. P. & Engelman, D. M. (1999). TOXCAT: A measure of transmembrane helix association in a biological membrane. *Proc. Natl Acad. Sci. USA*, **96**, 863–868.
26. Grasberger, B., Minton, A. P., DeLisi, C. & Metzger, H. (1986). Interaction between proteins localized in membranes. *Proc. Natl Acad. Sci. USA*, **83**, 6258–6262.
27. Fleming, K. G., Ackerman, A. L. & Engelman, D. M. (1997). The effect of point mutations on the free energy of transmembrane a-helix dimerization. *J. Mol. Biol.* **272**, 266–275.
28. Eriksson, A. E., Baase, W. A., Xhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*, **255**, 178–183.
29. Yonath, A. & Traub, W. (1969). Polymers of tripeptides as collagen models. IV. Structure analysis of poly(L-prolyglycyl-L-prolin). *J. Mol. Biol.* **43**, 461–477.
30. Traub, W., Yonath, A. & Segal, D. M. (1969). On the molecular structure of collagen. *Nature*, **221**, 914–917.
31. Liu, W., Crocker, E., Siminovitch, D. J. & Smith, S. O. (2003). Role of side-chain conformational entropy in transmembrane helix dimerization of glycophorin A. *Biophys. J.* **84**, 1263–1271.
32. Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halvorson, H. R. (1981). Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* **36**, 575–588.
33. Laue, T. M., Shah, B., Ridgeway, T. M. & Pelletier, S. L. (1992). *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J. & Horton, J. C., eds), pp. 90–125, Royal Society of Chemistry, Cambridge.
34. Pattabiraman, N., Ward, K. B. & Fleming, P. J. (1995). Occluded molecular surface: analysis of protein packing. *J. Mol. Recogn.* **8**, 334–344.
35. Lee, K. H., Xie, D., Freire, E. & Amzel, L. M. (1994). Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins: Struct Funct. Genet.* **20**, 68–84.
36. Baker, B. M. & Murphy, K. P. (1998). Predication of binding energetics from structure using empirical parameterization. *Methods Enzymol.* **295**, 294–315.
37. Bevington, P. R. (1969). *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill Book Company, New York.

Edited by G. von Heijne

(Received 8 April 2004; received in revised form 1 June 2004; accepted 1 June 2004)