

Motifs of Serine and Threonine can Drive Association of Transmembrane Helices

Jessica P. Dawson, Joshua S. Weinger and Donald M. Engelman*

Department of Molecular
Biophysics and Biochemistry
Yale University, New Haven
CT 06520-8114, USA

Known sequence motifs containing key glycine residues can drive the homo-oligomerization of transmembrane helices. To find other motifs, a randomized library of transmembrane interfaces was generated in which glycine was omitted. The TOXCAT system, which measures transmembrane helix association in the *Escherichia coli* inner membrane, was used to select high-affinity homo-oligomerizing sequences in this library. The two most frequently occurring motifs were SxxSSxxT and SxxxSSxxT. Isosteric mutations of any one of the serine and threonine residues to non-polar residues abolished oligomerization, indicating that the interaction between these positions is specific and requires an extended motif of serine and threonine hydroxyl groups. Computational modeling of these sequences produced several chemically plausible structures that contain multiple hydrogen bonds between the serine and threonine residues. While single serine or threonine side-chains do not appear to promote helix association, motifs can drive strong and specific association through a cooperative network of interhelical hydrogen bonds.

© 2002 Elsevier Science Ltd.

Keywords: membrane protein; protein folding; TOXCAT; hydrogen bond; association

*Corresponding author

Introduction

The thermodynamics of α -helical membrane protein folding can be conceptualized as a process that occurs in two distinct steps.¹ First, independently stable transmembrane (TM) helices are established across the lipid bilayer and second, the helices interact laterally to form higher-order structures. Previously, as a step towards identifying residues important in mediating TM helix interactions, a library of randomized TM domain interfaces was generated² and then searched for strongly associating TM helices with the TOXCAT system.³ This system can be used to screen for high-affinity homo-oligomerization of TM domains in the *Escherichia coli* inner membrane. The screen identified the GxxxG (GG4) motif in over 80% of the high-affinity isolates. This well-characterized motif is involved in a number of helix oligomers^{4–7} and can drive oligomerization of a poly-methionine

helix.⁸ Its general relevance in helix oligomerization has been supported by its frequent occurrence in transmembrane helices.⁹ The success of the previous library in identifying a biologically relevant oligomerization motif prompted the creation of a second library to screen for additional high-affinity motifs. To prevent selection of the GG4 motif, glycine was omitted from the library design. We find that motifs of four to five serine and threonine residues can drive strong helix interactions.

This result provides additional evidence for the importance of polar residues in mediating and stabilizing the interactions between TM helices. Polar amino acids can be broken into two groups, the strongly polar residues that have at least two polar groups in their side-chain (e.g. Gln, Arg) and the less polar residues that have only a single polar group (e.g. Ser, Thr). In some cases the interactions between strongly polar residues is important for function^{10–12} or for the binding of prosthetic groups.¹³ Recently, a single strong polar residue (Gln, Asn, Glu or Asp) was found sufficient to induce strong helix association in model TM helices,^{14–17} suggesting that this type of interaction also might play a role in the stabilization of helical membrane proteins. In contrast to the strongly polar residues, neither a single serine nor a single threonine residue was able to promote significant

Abbreviations used: TM, transmembrane; cAM, chloramphenicol; CAT, chloramphenicol acetyltransferase; GpA, glycoporphin A; CHI, CNS searching of helix interactions; CNS, crystallography and NMR system.

E-mail addresses of the corresponding author:
donald.engelman@yale.edu

helix association.^{15,16} Although not so well conserved as the strongly polar residues, serine and threonine are the most frequently occurring polar residues in transmembrane helices.⁹ In part this may be explained by serine and threonine's ability to form an intrahelical hydrogen bond with their backbone carbonyl group,¹⁸ partially satisfying the side-chain's hydrogen bonding capacity. When involved in a backbone hydrogen bond, serine and threonine have been shown to promote bends in the helix backbone, possibly facilitating conformational changes in membrane proteins.¹⁹ The results from our study show that serine and threonine can have an important role in membrane protein folding, stabilizing the formation of helical oligomers.

Results

To identify new motifs that drive TM helix-helix interactions, a randomized library of right-handed transmembrane helix interfaces was created and selected for high-affinity homo-oligomerization with the TOXCAT system (Figure 1(a)).³ The library construct encodes a 17 amino acid residue hydrophobic segment with a pattern of seven variable positions that is based on the interfacial surface of the right-handed structure of glycophorin A (Figure 1(b)). The amino acid residues at the variable positions represent the most frequently occurring hydrophobic residues in TM helices,⁹ with the notable omission of glycine to prevent selection of the previously identified GG4 motif. The eight amino acid residues allowed were alanine, valine, leucine, isoleucine, proline, phenylalanine, serine, and threonine with leucine at the intervening positions. This generated a library of 2.1×10^6 possible combinations.

The transmembrane sequences were cloned into the TOXCAT construct and selected for oligomerization at varying concentrations of chloramphenicol (CAM). Transmembrane domains from colonies that grew at the highest CAM concentration (300-400 $\mu\text{g}/\text{ml}$) were sequenced. The membrane insertion and correct orientation of several library isolates was verified by protease sensitivity in a spheroplast assay, as described³ (data not shown). As TOXCAT constructs containing the GpA wild-type TM domain and the disruptive mutant G83I were used as controls, the spheroplast assay also indicated that the levels of expression and insertion are equivalent for all TM domains analyzed.

The most striking result from this library is the overwhelming selection for serine and threonine residues (Figure 1(c)). In these sequences it is apparent that the polar residues are not randomly distributed but show a position-dependent selection. The strongest selection is observed for the threonine at position 13 and the serine at positions 9 and 10. Greater variety is present at positions 5 and 6 where the most common combinations of residues are SS, TS, SP, PS, and AP. The only selec-

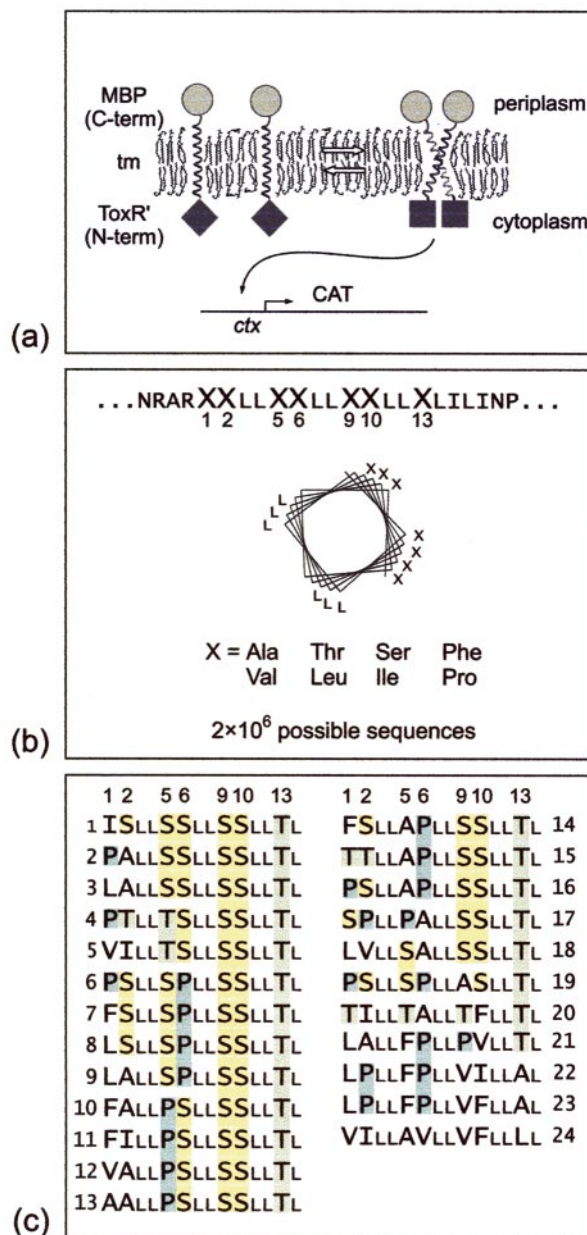


Figure 1. (a) TOXCAT assay. Oligomerization of the TM domains activates the cytoplasmic *toxR* domains, which are then able to activate transcription at the *ctx* promoter that controls expression of chloramphenicol acetyl-transferase (CAT). (b) Library design. The transmembrane sequence is shown with each of the seven variable positions marked by an X. The helical wheel diagram illustrates that the variable positions map to the interface of helices interacting at a right-handed crossing angle. The eight possible amino acid residues at the variable positions are shown. (c) Library results. TM helix sequences of colonies surviving on 300-400 $\mu\text{g}/\text{ml}$ CAM were sorted by polarity and grouped by sequence similarity. The most commonly occurring residues (proline, serine and threonine) were highlighted to emphasize the position-dependent selection.

tion apparent in the first two positions is for serine at position 2, especially in those cases where either SP or AP is present at 5 and 6.

A series of point mutations at positions 5, 6, 9, 10 and 13 were introduced in two library isolates (3 and 8) to explore the specificity of the interaction. Chloramphenicol acetyl-transferase (CAT) assays were used to determine the degree of association in the original library sequences and each of the mutants. Isolates 3 and 8 have a CAT activity comparable to the strongly dimerizing glycoporphin A (GpA), which suggests that the polar motif drives an interaction similar in strength to the GG4 motif found in GpA (Figure 2). This is further supported by the presence of a serine and threonine-rich motif in the previous library of GG4 containing sequences.² The serine to alanine mutations at positions 5, 6, 9, and 10 of isolate 3 were designed to test whether the serine residues function by hydrogen bonding and whether that interaction is cumulative or cooperative. These four mutations each produced a 70-90% drop in CAT activity, suggesting that the interaction is cooperative. In the sequences with proline, serine was often present at a neighboring position, suggesting that serine functions with proline to promote association. The role of these residues was tested by making mutations of serine to alanine, and proline to both alanine and glycine. Each of these three mutations reduced CAT activity by approximately 70%. The most strongly selected residue in this library was threonine at the final variable position. Mutations of threonine to both valine and serine were strongly disruptive, indicating that threonine may have both a packing and a hydrogen bonding role. Remarkably, each of the nine minimal mutations made in these sequences was strongly disruptive (Figure 2), indicating that the interaction is specific and cooperative. Although packing may play a role in the interaction, the serine to alanine mutations indicate that hydrogen bonding is probably the major factor stabilizing the oligomer. It should be noted that both serine and threonine are twofold over-represented in the selected *versus* unselected library. Proline, although

not enriched in the selected library, plays an important role in promoting helix association, as indicated from the mutagenesis results.

To develop chemically reasonable models for the potential interaction interfaces for the sequences, isolates 3 and 8 were modeled using the CNS searching of helix interactions (CHI) from the crystallography and NMR system (CNS).²⁰⁻²² In the simulations, CHI modeled the interaction between two parallel α -helices using the primary sequence of selected library isolates. Since the TOXCAT assay cannot distinguish between dimers and higher-order oligomers, the exact order of the oligomer for the library sequences is not known. Dimers were modeled because they are computationally the least intensive, and higher-order oligomers can be found by propagation from asymmetric dimers.²¹ The computational search produced several models with serine and threonine residues at the interface. Representative structures of the models generated are shown in Figure 3(b) through (d). The models show a series of hydrogen bonds between the polar residues at positions 9, 10 and 13 (SSxxT) that provide a chemically plausible explanation for their sensitivity to non-polar mutations. The role of positions 5 and 6 in the helix interface is more complex. When both residues are polar they can form interhelical hydrogen bonds similar to those between positions 9, 10 and 13. Although not shown here, several models had serine 5 in an intrahelical hydrogen bond. However, if proline is present at one of these positions the models do not provide a clear explanation for their sensitivity to glycine and alanine mutations. It should be noted that although the mutagenesis suggests that each polar residue is involved in a hydrogen bond, the models show that only a limited number of hydrogen bonds can be present at one time. These seemingly contradictory results might be explained by an interface of polar residues that can change between several hydrogen bonding patterns.

Although the design strategy for the library sequences was aimed at promoting helix interactions with right-handed crossing angles, a left-

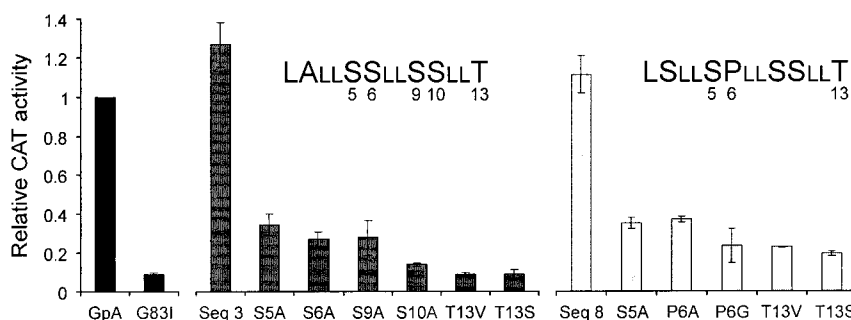


Figure 2. Library mutants. CAT activity measurements were used to assess the degree of association of library isolates 3 and 8 and their mutants. The bars indicate the amount of CAT activity normalized to wild-type GpA. Both wild-type GpA and its disruptive mutant G83I are shown for comparison. The mutated positions are numbered. Error bars represent the standard deviation of at least two measurements for each construct.

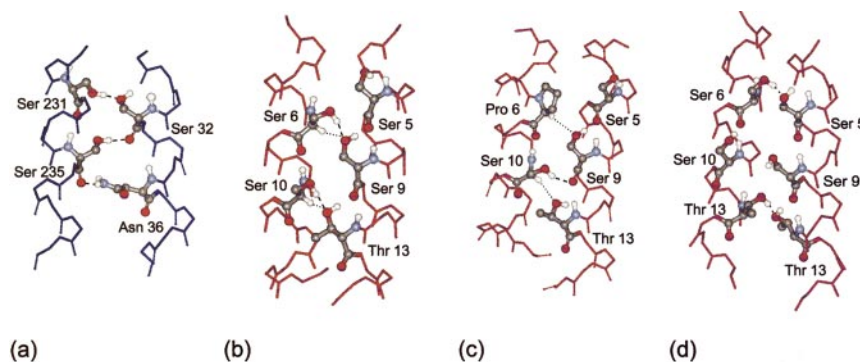


Figure 3. (a) The serine-serine hydrogen bonds between helices A and G of halorhodopsin are shown with broken lines. Serine 231 is in the less frequent *trans* rotamer where it cannot form a backbone hydrogen bond. It is, however, in a position where it can form a cross-strand hydrogen bond with serine 32. The oxygen atom of serine 235 is forming a cross-strand hydrogen bond with the carbonyl group of serine 32. Also noteworthy is the side-chain amino group of asparagine 36 that is close enough to the carbonyl group of serine 235 to form a hydrogen bond. (b) Model of sequence 3 at a right-handed crossing angle. For clarity only the front half of the sequence is shown. The other side of the helix has an identical hydrogen bonding pattern. The hydrogen bonds between serine residues are marked with a broken line. The putative C α hydrogen bonds between the backbone and serine side-chain are shown as a dotted line. Thr13 and Ser9 are in an *i*-4 hydrogen bond while Ser6 and Ser10 are in a cross-strand hydrogen bond. (c) Model of sequence 8 at a right-handed crossing angle. In this model Thr13 is in the less common *i*-3 backbone hydrogen bond and Ser10 is hydrogen bonding with the carbonyl group of Ser9 on the opposite helix. (d) Model of sequence 3 at a left-handed crossing angle. The two possible hydrogen bonds in this sequence are shown as a broken line.

handed interaction could still arise. Therefore, models with both types of crossing angles were examined. Of particular significance was the nature and number of possible stabilization energies between the left and right-handed models. In the left-handed structures, only one or two putative hydrogen bonds across the interface was observed. However, in the right-handed structures, there were four putative hydrogen bonds (Figure 3(b) and (c), broken line) as well as another four possible C α hydrogen bonds (Figure 3(b) and (c), dotted line). Although weaker than a conventional hydrogen bond, C α hydrogen bonds have recently been implicated as a source of stabilization energy for helix-helix interactions, often in the presence of serine and threonine.²³ These factors suggest that the left-handed models are less likely to produce strongly interacting oligomers, since they possess far fewer possible hydrogen bonds, as compared to the right-handed models. Therefore, the right-handed models may be a more accurate representation of the structure of the library isolates.

Since the mutagenesis results suggest that hydrogen bonding is critical to this interaction, helical membrane protein structures from the RCSB PDB were searched for evidence of serine-serine side-chain hydrogen bonding. An example was found between helices A and G in the 1.8 Å resolution structure of halorhodopsin²⁴ (Figure 3(a)). This structure supports the hypothesis that hydrogen bonds can stabilize the association between TM helices.

The SWISS-PROT database was searched for examples of the polar motifs. Twelve TM domains containing either the SxxSSxxT or SxxxSSxxT motifs were identified (Table 1). Eight of these proteins have another transmembrane domain

containing at least three polar residues that could be interacting with the polar motif. With the exception of the two RODA proteins, the set of proteins identified are unique. As this motif is not conserved in any protein family, it is unlikely to be involved in protein function.

Although this complete interface may not occur frequently in TM helices, combinations of two or three serine residues could contribute to stability. It should be noted that, as dimers, these helices have a more stringent requirement for detailed interactions than multi-spanning membrane proteins where each helix has at least two interaction surfaces. Therefore, although multiple polar residues in a specific pattern are required to drive the association of the library isolates, in multi-spanning membrane proteins a smaller number of polar residues could be used to stabilize helix interactions. For example, breaking the full motif into two amino acid pairs revealed that the SS4 motif, found in most of the library isolates, also occurred frequently in helical TMs⁹ with an odds ratio of 1.13 and a significance of 2.2×10^{-4} . This motif was found in the structure of halorhodopsin (Figure 3(a)),²⁴ where serine 231 and serine 235 (SS4) hydrogen bond with serine 32, potentially stabilizing the folded structure.

Discussion

Several recent studies have explored the role of hydrogen bonds in mediating TM helix oligomerization and found that certain polar groups (Gln, Asn, Glu and Asp) can promote helix associations in the absence of specific packing interactions.^{14–17} However, not all polar groups have this capacity, as a single serine or threonine residue is unable to

Table 1. Polar motifs in SWISS-PROT

GAP1_YEAST	LLALSGLSSLFTWGGICICHI
OL1B_HUMAN	FSTCSSHLTVVGFIFYGTGVFSYTR
OLF6_RAT	AFSTCSSHLTVVLIWYGSTIFLHV
PANF_HAEIN	VSYFSSIIITLILTALLIFAAL
QOXM_SULAC	LSGLSSTLTGVNFVMTIT
YMB8_YEAST	TILSTTSSFFTLFIGAICHVE
CYA2_RAT	AIIASILTSSSHTIVLSVYLS
RODA_HELPY	LPLFSYGGSSFITFMILFGIL
RODA_TREPA	ITGIPLLLLSYGGSSLWTAMI
TRG5_ECOLI	VGLVVPTLLSWGASSVITDL
YAT5_SCHPO	GYLFWLWILSALLSSAYTFLW
YMX8_YEAST	SFCLSSAVTVISSFVVFGEIF

TM domains with the SxxSSxxT or SxxxSSxxT motifs are shown. The SWISS-PROT ID is on the left and the TM domain with the motif highlighted is on the right. Although all TM domains from the SWISS-PROT database were searched, only TM domains from polytopic membrane proteins contained the serine-rich motif.

promote significant oligomerization. The latter result is perhaps not surprising because both serine and threonine are capable of forming an intrahelical hydrogen bond with their *i*-3 or *i*-4 carbonyl oxygen atom.^{18,19} Formation of an interhelical hydrogen bond would require breaking the backbone hydrogen bond to form an interhelical bond, a change that produces no additional hydrogen bonds per pair of helices but has the added entropic penalty of constraining the motion of the two helices. However, multiple serine and threonine residues at the helix interface can drive oligomer formation through a series of side-chain hydrogen bonds. The finding that any mutation of a polar to a non-polar residue at the helix interface abolished dimer formation suggests that the interacting hydrogen bonds function cooperatively to create a strong helical oligomer.

A combination of modeling and mutagenesis was used to determine the role of each residue in the helix interface. From the models it appears that the major role of threonine 13 is to accept a hydrogen bond from serine 9 on the opposite helix; however, serine at position 13 could perform that task equally well, suggesting that threonine may promote packing interactions as well as hydrogen bonding. Mutagenesis supports this conclusion, as both serine and valine were disruptive at this position (Figure 2). In the models the γ -methyl groups of the two threonine residues contact each other across the interface. This packing interaction may fix the position of the two helices and promote the hydrogen bonding network.

The other amino acid that was strongly conserved in this library was proline, at either position 5 or 6. The models do not provide an obvious explanation for its selection, but this is not surprising as the modeling program includes a helical structure conformational constraint. If proline creates a significant bend in the helix, the program would not be able to model it. The finding that

glycine, which also permits helix bending, could not substitute for proline indicates that the role of proline may not be to create a distortion in the helix backbone, but rather to free the *i*-4 carbonyl oxygen atom. In any of the polar sequences with proline there is serine or threonine either next to it or at the *i*-4 position. The requirement for a polar residue at this position suggests that it is hydrogen bonding with the exposed carbonyl group. This interhelical hydrogen bond could provide stabilization energy for helix association.

Although this study provides possible explanations for these polar interactions, it was not known if these motifs are present in membrane proteins. Searching the SWISS-PROT database identified 12 TM segments containing serine-rich motifs (Table 1). The small number of identified sequences suggests that the complete motif is not a common theme in natural membrane protein association. However, portions of this motif, like SxxxS (SS4), could help stabilize helix interactions in polytopic proteins. Networks of serine and threonine residues may also be an ideal candidate for creating interactions between subunits of membrane proteins because they are strong, specific and have a low cost of inserting into membranes as compared with other polar residues. There are several pieces of evidence that support the role of serine and the SS4 motif in helix-helix interactions. The SS4 motif is over-represented in a statistical survey of TM helices⁹ and recently serine was shown to have a high propensity for self-interaction in membrane proteins.²⁵ In addition to the statistical analysis a specific example of the SS4 motif's involvement in a TM helix association was found in the structure of halorhodopsin (Figure 3(a)).²⁴ These results indicate that combinations of serine may have a general role in stabilizing TM helix interactions. Previously, it was reported that serine and threonine occur on the lipid-exposed surface of membrane proteins

3.8 times more frequently than Gln, Asn, Glu and Asp combined.¹⁶ Presumably the lipid-exposed serine and threonine residues are in geometries or proximities that limit non-specific aggregation.

Growing evidence indicates that mutations leading to hydrogen bonds between TM helices are the cause of some genetic diseases.^{26,27} As this study shows that only specific serine and threonine-containing motifs create strong helix interactions, this suggests that spontaneous mutations to serine or threonine would not lead to disease-causing hydrogen bonds with another serine or threonine. However, a mutation to serine or threonine could provide a hydrogen bonding partner to a strongly polar residue already present in the sequence, thereby allowing an incorrect hydrogen bond to form.

Described here is a novel interaction motif that creates an association similar in strength to the GG4 motif of GpA. These findings add force to the emerging picture of the role of polar groups in membrane protein folding. Unlike other hydrogen bond driven associations, where only a single position is responsible for oligomerization, the sensitivity to even conservative substitutions at any of five positions along the helix interface show that this interaction is cooperative and specific.

Materials and Methods

Library construction

The oligonucleotide encoding the library (aacacaccg-cagcctaggQZtQZtctgctaQZtQZtctgctaQZtctgatcgccctaacggatatac) was synthesized, where Q and Z refer to mixtures of (a,g,c,t) and (t,c), respectively. To correct for a previously identified nucleotide bias during synthesis the following nucleotide ratios were used: A = 29%, C = 28%, T = 23% and G = 20%. The *StyI* and *DpnII* sites used for cloning into the pccKAN vector are underlined. The resulting codons (QZt) specify the amino acids Phe, Ser, Leu, Pro, Ile, Thr, Val and Ala. The remainder of the cloning process was identical with that of the previous library.² Based on the number of colonies obtained after ligation, we have approximately 70% of the complete library.

Library selection

Between 0.5 and 1 μ l of the titered library glycerol stock (a number of bacteria that could contain two to four libraries of sequences, respectively) was plated on LB/AMP plates of varying CAM concentrations. Bacteria that grew on the 300-400 μ g/ml of CAM plates (less than 1% of the library) were used to inoculate PCR reactions for sequencing. Isolates were screened for membrane insertion by plating on M9-maltose minimal medium plates. Several isolates were also screened for membrane insertion using the protease sensitivity in the spheroplast assay described,³ except that the LexA antibody (Invitrogen) was used to verify the intact state of the spheroplasts. Briefly, spheroplasts were created by digesting bacteria with lysozyme to remove the peptidoglycan layer. Spheroplasts were then digested with proteinase K to remove those proteins present in the periplasm. Only if the TOXCAT chimera is in the mem-

brane and correctly oriented will the maltose binding protein (MBP) portion of the fusion be digested away from the remaining protein. Following digestion the samples were run on SDS-PAGE and then Western blot analysis was carried out using antibodies against MBP (NEB) and LexA (1:10,000 and 1:5000 dilution, respectively).

CAT assays

Cell-free extracts were made by pelleting 200 μ l of cells at an A_{600} of 0.6, resuspending in 500 μ l of 0.1 M Tris (pH 8.0), then lysing with 20 μ l of 100 mM EDTA, 100 mM DTT and 50 mM Tris (pH 8.0) and one drop of toluene from a drawn-out Pasteur pipette, at 30 °C for 30 minutes. The cell-free extract was then diluted 1:200 before being used in the CAT assays. The CAT assays were performed using the Quant-T-CAT kit (Amersham).

CHI simulations

A detailed description of the CHI simulations is provided elsewhere.²⁰⁻²² All simulations were conducted *in vacuo*, a partial mimetic of the low dielectric of the membrane. In the first step of the simulation two α -helices were built with both a right and left-handed crossing angle. Helical modeling was initiated with a series of independent rotations about each helix's central axis. Molecular dynamics simulations, simulated annealing and energy minimizations were carried out at each pair of rotations to yield a total of 512 final structures. Groups of ten or more final structures with a backbone RMSD of less than 1 Å were placed into clusters. The average structure of each cluster was then calculated and taken through energy minimization. As the force field parameters used in CHI do not explicitly describe hydrogen bonds they were defined in any instance where a hydrogen bond acceptor and donor were within 3 Å of each other. None of the well-populated models could be expanded to higher-order oligomers, suggesting that dimers are being observed in the TOXCAT selection.

Searching of SWISS-PROT

The transmembrane domains were extracted from the SWISS-PROT, version 39 updated on 12 July 2001²⁸ and searched for both the SxxSSxxT and SxxxSSxxT patterns.

Acknowledgments

We thank W. Russ, K. Sonoda and A. Senes for intellectual contribution and many helpful discussions, the Engelman laboratory for critical reading of the manuscript, and the NSF and NIH for funding. J.P.D. is supported by a grant from the NSF.

References

1. Popot, J. L. & Engelman, D. M. (1990). Membrane protein folding and oligomerization: the two-stage model. *Biochemistry*, **29**, 4031-4037.
2. Russ, W. P. & Engelman, D. M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.* **296**, 911-919.

3. 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.
4. Lemmon, M. A., Treutlein, H. R., Adams, P. D., Brunger, A. T. & Engelman, D. M. (1994). A dimerization motif for transmembrane alpha-helices. *Nature Struct. Biol.* **1**, 157-163.
5. 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.
6. Op De Beeck, A., Montserret, R., Duvet, S., Cocquerel, L., Cacan, R., Barberot, B. *et al.* (2000). The transmembrane domains of hepatitis C virus envelope glycoproteins E1 and E2 play a major role in heterodimerization. *J. Biol. Chem.* **275**, 31428-31437.
7. Li, Z., Glibowicka, M., Joensson, C. & Deber, C. M. (1993). Conformational states of mutant M13 coat proteins are regulated by transmembrane residues. *J. Biol. Chem.* **268**, 4584-4587.
8. Brosig, B. & Langosch, D. (1998). The dimerization motif of the glycoporphin A transmembrane segment in membranes: importance of glycine residues. *Protein Sci.* **7**, 1052-1056.
9. Senes, A., Gerstein, M. & 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.
10. Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899-929.
11. Luecke, H., Schobert, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. (1999). Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291**, 899-911.
12. Belrhali, H., Nollert, P., Royant, A., Menzel, C., Rosenbusch, J. P., Landau, E. M. & Pebay-Peyroula, E. (1999). Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Struct. Fold. Des.* **7**, 909-917.
13. Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. (1995). Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature*, **376**, 660-669.
14. Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger, A. T. & Engelman, D. M. (2000). Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nature Struct. Biol.* **7**, 154-160.
15. Zhou, F. X., Merianos, H. J., Brunger, A. T. & Engelman, D. M. (2001). Polar residues drive association of polyleucine transmembrane helices. *Proc. Natl Acad. Sci. USA*, **98**, 2250-2255.
16. Gratkowski, H., Lear, J. D. & DeGrado, W. F. (2001). Polar side-chains drive the association of model transmembrane peptides. *Proc. Natl Acad. Sci. USA*, **98**, 880-885.
17. Choma, C., Gratkowski, H., Lear, J. D. & DeGrado, W. F. (2000). Asparagine-mediated self-association of a model transmembrane helix. *Nature Struct. Biol.* **7**, 161-166.
18. Gray, T. M. & Matthews, B. W. (1984). Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. *J. Mol. Biol.* **175**, 75-81.
19. Ballesteros, J. A., Deupi, X., Olivella, M., Haaksma, E. E. & Pardo, L. (2000). Serine and threonine residues bend alpha-helices in the chi(1) = g(-) conformation. *Biophys. J.* **79**, 2754-2760.
20. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905-921.
21. Adams, P. D., Arkin, I. T., Engelman, D. M. & Brunger, A. T. (1995). Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. *Nature Struct. Biol.* **2**, 154-162.
22. Adams, P. D., Engelman, D. M. & Brunger, A. T. (1996). Improved prediction for the structure of the dimeric transmembrane domain of glycoporphin A obtained through global searching. *Proteins: Struct. Funct. Genet.* **26**, 257-261.
23. Senes, A., Ubarretxena-Belandia, I. & Engelman, D. M. (2001). The α -H...O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl Acad. Sci. USA*, **98**, 9056-9061.
24. Kolbe, M., Besir, H., Essen, L. O. & Oesterhelt, D. (2000). Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science*, **288**, 1390-1396.
25. Adamian, L. & Liang, J. (2001). Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. *J. Mol. Biol.* **311**, 891-907.
26. Therien, A. G., Grant, F. E. & Deber, C. M. (2001). Interhelical hydrogen bonds in the CFTR membrane domain. *Nature Struct. Biol.* **8**, 597-601.
27. Smith, S. O., Smith, C. S. & Bormann, B. J. (1996). Strong hydrogen bonding interactions involving a buried glutamic acid in the transmembrane sequence of the neu/erbB-2 receptor. *Nature Struct. Biol.* **3**, 252-258.
28. Bairoch, A. & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl. Acids Res.* **28**, 45-48.

Edited by G. von Heijne

(Received 5 November 2001; received in revised form 12 December 2001; accepted 13 December 2001)