

Genetic Selection for and Molecular Dynamic Modeling of a Protein Transmembrane Domain Multimerization Motif from a Random *Escherichia coli* Genomic Library

Jennifer A. Leeds¹, Dana Boyd¹, Damon R. Huber², G. Koji Sonoda³
Hieu T. Luu¹, Donald M. Engelman³ and Jon Beckwith^{1*}

¹*Department of Microbiology and Molecular Genetics Harvard Medical School Boston, MA 02115, USA*

²*Program in Biological and Biomedical Sciences, Harvard Medical School, Boston MA 02115, USA*

³*Department of Molecular Biophysics and Biochemistry Yale University, New Haven CT 06520, USA*

In order to identify new transmembrane helix packing motifs in naturally occurring proteins, we have selected transmembrane domains from a library of random *Escherichia coli* genomic DNA fragments and screened them for homomultimerization *via* their abilities to dimerize the bacteriophage λ cI repressor DNA-binding domain. Sequences were isolated using a modified λ cI headpiece dimerization assay system, which was shown previously to measure transmembrane helix-helix association in the *E. coli* inner membrane. Screening resulted in the identification of several novel sequences that appear to mediate helix-helix interactions. One sequence, representing the predicted sixth transmembrane domain (TM6) of the *E. coli* protein YjiO, was chosen for further analysis. Using site-directed mutagenesis and molecular dynamics, a small set of models for YjiO TM6 multimerization interface interactions were generated. This work demonstrates the utility of combining *in vivo* genetic tools with computational systems for understanding membrane protein structure and assembly.

© 2001 Academic Press

Keywords: library; transmembrane; *Escherichia coli*; multimerization; lambda repressor

*Corresponding author

Introduction

Integral membrane proteins play essential roles in numerous cellular functions, including cell division, intra and inter-cellular signaling, and transport of macromolecules.^{1–4} Mutations affecting structure and assembly of membrane proteins often lead to cell toxicity in microbes and pathologies such as cystic fibrosis and Alzheimer's disease in humans.^{5,6} With the advent of modern genomics, integral membrane proteins have gained

much attention. For example, they are often identified from genome-wide viability screens for new drug targets. Yet compared to soluble proteins, sufficient tools generally do not exist to map the interactions that are required for tertiary structure and oligomeric assembly of integral membrane proteins.

In depth structural studies of several model integral membrane proteins such as the human erythrocyte protein glycophorin A (GpA) and phospholamban have defined some basic rules of membrane protein structure and assembly. Integral membrane proteins often depend on their transmembrane (TM) segments for assembly into functional structures (for reviews, see^{7,8}). TM domains are generally thought to adopt an alpha-helical secondary structure, with tertiary structure being determined by interhelical interactions between these preformed domains.⁹ The types of helix-helix interactions that can occur among integral membrane proteins are thought to consist primarily of van der Waals interactions,^{10–12} with polar interactions sometimes helping to drive the

Present addresses: G. K. Sonoda, DoubleTwist, 2001 Broadway, Oakland, CA 94612, USA; H. T. Luu, Ursinus College, Collegeville, PA 19426, USA.

Abbreviations used: TM, transmembrane; λ , bacteriophage lambda; GpA, glycophorin A; pfu, plaque-forming units; ES, export signal; E.O.P, efficiency of plating; CTD, C-terminal domain; MBP, maltose-binding protein; ss, signal sequence; ORF, open reading frame.

E-mail address of the corresponding author: jbeckwith@hms.harvard.edu

association.^{13,14} Examples of integral membrane proteins whose structure and assembly depend upon helix-helix interactions include the homodimeric glycoporphin A, heterodimeric T cell receptor, homodimeric *Escherichia coli* chemotactic aspartate receptor Tar and human CFTR.^{4,6,8}

In response to the paucity of proteomic and direct biophysical methods for studying membrane proteins, multidisciplinary approaches to study membrane protein structure that combine *in vivo* genetic techniques with computational modeling are being developed.^{12,15–18} Within this arena, we have developed a technique for detecting homodimerization of transmembrane segments from integral membrane proteins.¹⁵ Here, we expanded this assay system to be able to identify novel interacting TM sequences and combine mutagenesis with computational modeling to propose a set of models for the structure of a TM domain complex. By creating a large catalog of TM domain interaction interfaces, ultimately the results could allow predictions of the tertiary and quaternary structures of TM domains from naturally occurring proteins.

To accomplish our goal of identifying novel TM interacting domains, we utilized our *in vivo* genetic TM interaction technique to monitor homodimerization between transmembrane segments that were selected from a random library of short *E. coli* chromosomal fragments. Our approach was to generate gene fusions that place the candidate DNA segments between the bacteriophage λ cI repressor DNA-binding domain (“headpiece”) and the maltose-binding protein (MBP) lacking its signal sequence. The λ cI repressor is responsible for maintaining lysogeny of the lambda prophage as well as preventing superinfection by homoimmune phage.¹⁹ In order to bind the lambda operator DNA, the repressor must dimerize in the cytoplasm.²⁰ We showed previously that the native C-terminal dimerization domain of the lambda repressor can be replaced by a dimerizing transmembrane domain to reconstitute a functional repressor.¹⁵ By fusing the DNA encoding the mature MBP at the C terminus of the vector system, we were able to select for those DNA fragments that specifically encoded a transmembrane domain. If DNA encoding a TM segment is cloned into the random cloning site, the TM segment will facilitate transport of the MBP domain across the inner membrane. In the absence of an endogenous source of MBP, only those clones expressing a periplasmically localized MBP will be able to grow on maltose as a sole carbon source. Although a dimer is the minimum complex required to confer repression by lambda DNA-binding domains, higher-order TM domain complexes may exist *in vivo* and cannot be distinguished by our assay. Therefore, we refer to the TM domain interactions identified by *in vivo* screens as being “homomultimeric”, which we define as complexes containing two or more identical TM segments.

Our results demonstrate that our assay system can identify novel helix-helix interactions, and that

a combination of genetic and computational approaches result in useful tools to understand integral membrane protein structure and assembly.

Results

Generating an export signal domain library from the *E. coli* genome

We have developed a genetic system that allows us to select for export signals (ES) (transmembrane domains (TM) and cleavable signal sequences (SS)) from a library of random genomic DNA fragments, and then screen the export signals for those encoding transmembrane domains capable of homomultimerizing. The general strategy is to insert short, random DNA fragments in between sequences coding for the bacteriophage λ cI repressor DNA-binding domain and the signal-sequenceless maltose-binding protein. We then select for those fragments that encode export signals, which are capable of transporting the mature MBP across the *E. coli* inner membrane. This is accomplished by selecting for the ability of the hybrid protein to confer growth on maltose as a sole carbon source, of a strain of *E. coli* missing its endogenous source of MBP. We then individually assay the ability of the selected inserts to drive dimerization of the λ cI repressor headpiece by measuring the ability of the clones expressing the hybrid proteins to prevent infection by a λ phage (cI) missing its repressor. This is done by screening for plaque formation on spots of mal+ clones grown on agar containing λ cI phage. The λ immunity assay demands that, at a minimum, a λ cI repressor headpiece dimer has formed, however this *in vivo* method cannot predict or rule out higher order TM complexes. Therefore, we will refer to the TM domain associations identified in this work as “multimeric”, meaning at least two helices with at least one interaction interface are involved.

To distinguish between cleavable signal sequences and transmembrane helices, the λ -immune clones are then screened by Western blot analysis to identify those that produce a full-length tri-hybrid protein (cI headpiece-TM-MBP). Since insertion of a cleavable signal sequence may result in a translational restart and expression of a soluble MBP domain, typically only those inserts encoding TM domains will produce full-length tri-hybrids.

Insert generation and vector construction

In order to construct a library of export signals, chromosomal DNA from *E. coli* W3110 was digested with two sets of enzymes. In both cases, the goal was to generate a large number of fragments in the 48-150 nt size range. This size range corresponds to the minimal size necessary to encode a transmembrane domain up to the maximum sized insert (TM + “extramembraneous material”) that we estimated would not be likely to

contain a potential extramembraneous "dimerizing" or "interfering" domain. For one portion of the library, we completely digested the *E. coli* chromosome with the blunt ending, four-base cutters *AluI* and *RsaI*. This combination of enzymes is predicted to result in a total of 20,000 different fragments (>30 nt), 7168 of which are between 40 - 150 nt. For another portion of the library, we partially digested the *E. coli* chromosome with the enzyme *CviJI**.²¹ Under the conditions chosen, this very frequent cutter (recognition site RGCN) generates an estimated one million different overlapping blunt-ended fragments in the 40 - 150 nt range.

The inserts were cloned into a series of nine vectors (Figure 1 and Table 1). In the vector series, DNA encoding the mature portion of the MBP (Δ ss *malE*) is preceded by a unique *HpaI* cloning site (GTTAAC), shown within the expanded sequence in Figure 1. In each vector, the *HpaI* site is present in one of three reading frames (RF1-RF3) relative to the *malE* DNA, determined by the sequences to the right of AAC. This design allows for the selection of inserted DNA encoding domains capable of exporting the signal-sequenceless MBP. To screen the inserted exporting domains for their capacity to dimerize, the fusion constructs also contain the λ cI headpiece at their N termini (*cl* 1-132). Again, to accommodate all random insert DNA in each vector the λ cI headpiece is present in one of three reading frames relative to the following *HpaI* site, determined by the sequences to the left of GTT. The combination of three reading frames leading from the DNA encod-

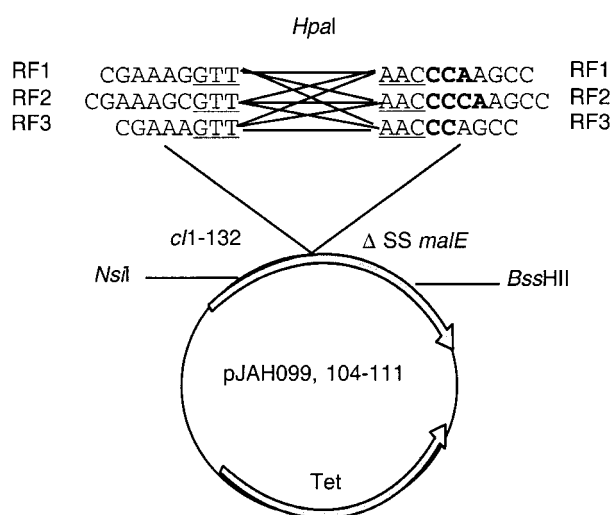


Figure 1. Design of random TM domain cloning vectors. pJAH099, and 104-111 are derived from the backbone plasmid pACYC184. Each vector has sequences encoding the λ cI DNA-binding domain and mature MBP, separated by a unique *HpaI* site. The *HpaI* site is in one of nine different combinations of reading frames in each vector relative to the upstream and downstream coding regions. The DNA sequences creating each of the three possible 3' and 5' reading frames are shown at the top of the figure.

ing the λ headpiece into the *HpaI* site, and out of the *HpaI* site into the DNA encoding the MBP domain results in a total of nine vectors. The vectors are derived from the backbone plasmid pACYC184 and contain the IPTG inducible *lacUV5* promoter driving transcription of the fusion constructs. Cloning details for constructing the vectors are described in the Material and Methods.

Selection of export domains from a random *E. coli* DNA library

Figure 2 depicts the scheme for selection of the export domain libraries. Following digestion of the *E. coli* chromosome and ligation of the small random DNA fragments into a pool of *HpaI*-digested plasmids containing equimolar amounts of each of the nine vectors, the ligation mixes were transformed into a highly electrocompetent strain of *E. coli* (TG1), selecting only for the presence of the plasmid TetR marker. This library represented all possible random DNA fragments. Those plasmids were then extracted and a portion was transformed into JAH143, a strain of *E. coli* lacking an endogenous source of maltose-binding protein (see Table 1 and Materials and Methods for strain construction). These transformants were selected on minimal medium containing tetracycline and IPTG to induce the fusion construct, with maltose as a sole carbon source. Since this selection demanded that MBP be exported to the periplasmic side of the *E. coli* inner membrane, only those inserts encoding transmembrane domains or cleavable signal sequences that were in frame with MBP would confer growth on maltose. Typically, there were

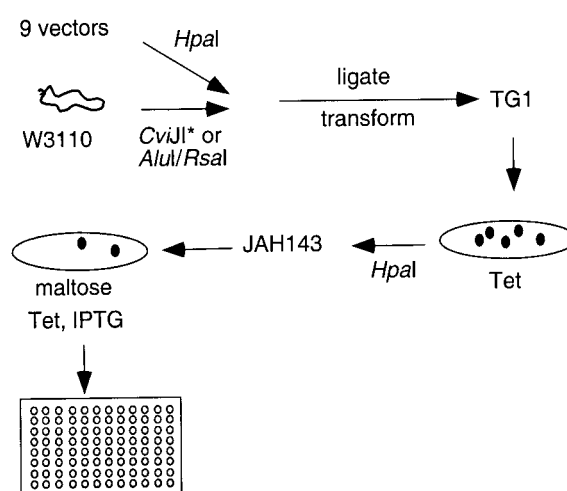


Figure 2. Generation of random export signal libraries. *E. coli* W3110 chromosomal DNA was digested to small fragments, cloned into the pool of nine library vectors, transformed into *E. coli* TG1, and selected on tetracycline. Tetracycline resistant plasmids were extracted and transformed into JAH143 (Δ *malE444*), selecting for growth on maltose as a sole carbon source in the presence of the fusion construct inducer IPTG.

Table 1. Bacterial strains, plasmids, and phages

Strain/plasmid/phage	Description	Source/Reference
<i>E. coli</i>		
W3110	Source of library DNA	Lesley Pratt
JAH013	MC1061/F' <i>lacI</i> ^q , KanR	¹⁵
JAH143	JAH013 Δ <i>malE444</i> , F' <i>lacI</i> ^q , KanR	This work
Plasmids		¹⁵
pJAH01	DNA encoding lambda cI headpiece and linker under the control of the <i>lacUV5</i> promoter in pACYC184	This work
pJAH099	DNA encoding lambda cI headpiece and signal sequenceless MBP separated by an in-frame <i>HpaI</i> site (RF1-1), under the control of the <i>lacUV5</i> promoter in pACYC184	This work
pJAH104	Same as pJAH099 except RF1-2	This work
pJAH105	Same as pJAH099 except RF1-3	This work
pJAH106	Same as pJAH099 except RF2-1	This work
pJAH107	Same as pJAH099 except RF2-2	This work
pJAH108	Same as pJAH099 except RF2-3	This work
pJAH109	Same as pJAH099 except RF3-1	This work
pJAH110	Same as pJAH099 except RF3-2	This work
pJAH111	Same as pJAH099 except RF3-3	This work
pJAH206	YjiO TM6 cloned into pJAH01	This work
pJAH231	pJAH206 V1A	This work
pJAH235	pJAH206 L2A	This work
pJAH261	pJAH206 I5A	This work
pJAH239	pJAH206 I6A	This work
pJAH240	pJAH206 V8A	This work
pJAH242	pJAH206 M9A	This work
pJAH245	pJAH206 G10A	This work
pJAH246	pJAH206 F11A	This work
pJAH262	pJAH206 I12A	This work
pJAH249	pJAH206 S13A	This work
pJAH250	pJAH206 F14A	This work
pJAH264	pJAH206 V15A	This work
pJAH222	pJAH206 G16A	This work
pJAH251	pJAH206 L17A	This work
pJAH252	pJAH206 L18A	This work
pJAH254	pJAH206 L19A	This work
Phages		
λ cI	cI	Laboratory collection
λ vir	virulent	Laboratory collection

two sizes of colonies visible after 48 hours of growth on the minimal maltose plates. The smallest clones (borderline macroscopic) did not restreak on minimal maltose plates. Based on sequence analysis of a subset of inserts, these clones did not encode export domains, whereas the larger colonies did. The frequency of large, *mal*⁺ colonies within the total population of tetracycline resistant transformants was 10⁻³. The *mal*⁺ colonies were sub-streaked and frozen in microtitre dishes.

Expected transmembrane domain library size versus actual number of *mal*⁺ clones analyzed

A priori, we did not know how many homomultimerizing transmembrane domains we should expect to detect by our assay method. However, we could estimate the total number of assayable transmembrane domains by analyzing the expected restriction fragments generated from each restriction digest. This was done by determining the hydrophobicities of the segments predicted to be encoded by random inserts generated from the two restriction digest libraries. From a complete *E. coli* W3110 *AluI/RsaI* digest, 4500 possible

clones were predicted to contain inserts encoding TM domains that were in-frame with λ cI headpiece and MBP. This represents an approximation of the total number of TM segments in the *E. coli* genome. Because of the large number of fragments generated by a *CviJI*^{*} partial digest, estimates for the *CviJI*^{*} library were extrapolated from analysis of the *E. coli* W3110 2 minute region of the chromosome. From these estimates, a *CviJI*^{*} partial digest library was expected to encode 225,000 clones containing inserts encoding TM domains that were in-frame with λ cI headpiece and MBP. Since this estimate was from a partial digest, individual sequences were represented multiple times in the total library. When we performed the actual library selection for *mal*⁺ clones, we retained a total of 477 *mal*⁺ clones, 373 *mal*⁺ clones containing *AluI/RsaI* inserts (a little over 8% of the total possible inserts) and 104 *mal*⁺ clones containing *CviJI*^{*} inserts (0.05% of total possible inserts), for further analysis. This represented about half of the total *mal*⁺ colonies that were obtained from six independent transformations of plasmids containing inserts.

Identification of λ cI headpiece -TM-MBP chimeras that confer immunity to λ cI phage

The random TM segment multimerization assay system is a three-step process. In the first stage, described above and shown in Figure 2, inserts encoding only export signals are selected from a library of all random inserts. In the second stage (Figure 3), the inserts are screened for their ability to dimerize the λ cI headpiece. This is done by screening for their ability to confer immunity to λ cI phage. In this immunity screening assay, a strain is considered to be "immune" if it has an efficiency of plating λ cI phage of 2×10^{-2} . That is, the total number of phage particles applied the plate (10^5 plaque-forming units (pfu)) was chosen so that each bacterial spot was infected with about 25-50 phage. If no plaques are observed then the efficiency of plating (E.O.P) for that strain is $< 2 \times 10^{-2}$. This level of stringency was selected because this is the E.O.P. λ cI observed using an intermediate dimerizing mutant of GpA TM (L75A) when induced with 10 μ M IPTG.¹⁵ We wished to identify TM segments that multimerized in the range of or stronger than the GpA TM (L75A) mutant. The stringency of the screen could

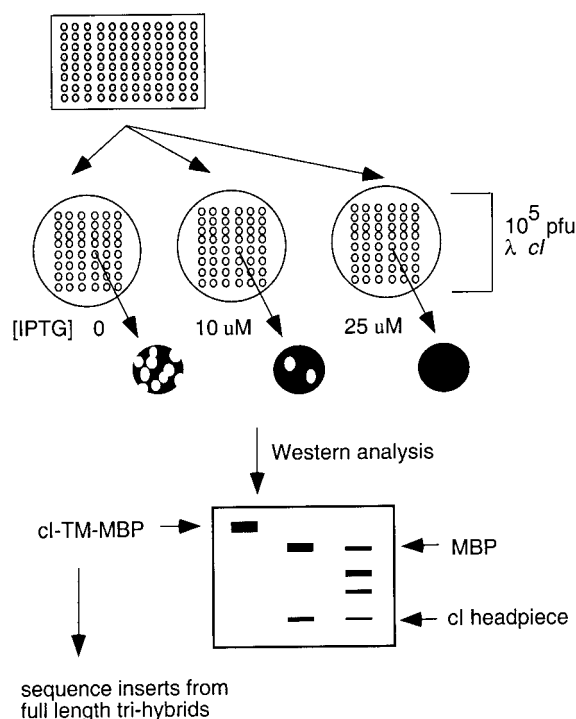


Figure 3. Identification of sequences conferring immunity to λ cI phage. Mal+ clones were induced with 0, 10, or 25 μ M IPTG and patched onto agar seeded with 10^5 pfu of λ cI phage. After overnight incubation, patches were screened for the presence of plaques. Immune clones were subjected to Western analysis and clones producing single, full length cI headpiece-ES-MBP chimerae were sequenced to reveal the identity of the inserted domain.

be increased or decreased to adjust the profile of the multimerizing TM library.

Library cultures were grown in the presence or absence of the inducer IPTG. Mid-logarithmically growing cultures were spotted with a serial dilution block onto media seeded with 10^5 λ cI phage in the presence or absence of IPTG. After overnight growth, bacterial patches were observed for the presence of plaques. To identify potential dimerizing inserts, we retained clones that had the profile of being sensitive to λ cI infection (many plaques observed) when the fusion protein was uninduced and immune to λ cI infection (no plaques observed) when induced with 10 μ M IPTG and/or 25 μ M IPTG. Of the 477 mal+ clones assayed, about 40% were immune to λ cI infection (E.O.P. $\leq 2 \times 10^{-2}$) and not λ vir infection. This group was composed of three subsets, 5% of the total were immune to λ cI without induction, 15% of the total were immune when induced with 10 μ M IPTG and the remaining 20% of the total were immune only when induced with 25 μ M IPTG.

In order to determine whether immunity to λ cI infection was conferred by a single, full-length λ cI headpiece-ES-MBP chimera, Western analyses were performed (data not shown). Total cellular protein was screened for reactivity to anti- λ cI headpiece antibody and anti-MBP antibody. Of the clones conferring immunity to λ cI phage, those producing a single, full-length protein without visible breakdown were candidates for further study. This screening step eliminates clones with inserts encoding cleavable signal sequences that also generate stably expressed λ cI headpiece. Other potential false positives also eliminated at this screening step will be addressed in Discussion. From a Western analysis of 110 independent clones that conferred immunity to λ cI phage in the presence of 10 or 25 μ M IPTG, 12 produced full-length chimeras with no visible breakdown products (Table 2). The underlined portions of the sequences represent the probable TM segments. Sequences with (...) represent inserts consisting of a fusion of two random fragments.

Subcloning of selected TMs into pJAH01 multimerization assay vector

Our assay system is designed to detect inserts that contain both an export signal and at least one dimerization interface. To enrich for TM interaction domains, chromosomal digestions were optimized to give random DNA fragments in the size range corresponding to a TM domain or slightly larger. This was done to limit the chances of picking up DNA fragments encoding "extramembraneous" protein domains that were capable of driving the dimerization of the λ cI headpiece. To confirm which inserts from the pool identified in Table 2 encoded TM domains that were solely responsible for dimerizing the λ cI headpiece, the portion of the random sequence corresponding to the pre-

Table 2. Random TM inserts conferring immunity to λ *cI* infection when fused in-frame between λ *cI* headpiece and Δ SS MBP

Inserted sequence	<i>E. coli</i> chromosomal locus
<u>LLLTCIVAGVMIAILVSC</u> QLFLVAWHKHEVKYDTLITD	<i>rtn</i>
PTPENHEEKNDGEKPKQSITSIKISLTLVLVIGYALGKIAML <u>F</u>	<i>ybeV</i>
PKTDAPVEKEKKGIGVLFLSVAALCYILGQLGFISWVPEYAK	<i>yhfC</i>
TGIGIGLFLLLIAANGVGMVKNPIEGL...PVALGAFTSFPVMM	<i>ygfQ</i>
SFKLLTNHRSCFAFNPFRRFAQYGNQCFTTCARASKRGGSDRIPGFRHGAENTRQDRTKTGSDKTNNFFRNS... <u>LMLTLCSSVMGVLGALQGT</u>	<i>yejE</i>
LFCTVPMLFISAALVLPVIM <u>PVI</u>	<i>ydcP</i> (out of frame)
WHWVFIVTGGIGIIWSLIWFKVYQPPRLTKGISK	<i>dgoT</i>
MHFMHWKVLFAIIAVMGFISFVGLLLAMPETVCRGAVPFSAKSVLRDRNFVFCNRLFLFGAATISL	<i>yjiO</i>
LLLIRLRRLLGLCCFAWATLRLAPYLF <u>LVT</u>	Noncoding DNA
LIVVIMVHIYAALWVKGTITAMVEGWVTTAWAKKHHPRWYREVRKTTEKK	<i>fdol</i>
LLFALAQTRHKQWLSLAKKVLVGLVMGVVFGALHTIYGS <u>DSQ</u>	<i>ydiN</i>
<u>LKIFLPLVGATVFKIAL</u>	<i>ymcA</i> (out of frame)

dicted TM domain (underlined sequences in Table 2) was subcloned into pJAH01 (see Materials and Methods and Supplementary Material, Table 1, for details on plasmid construction). This vector was chosen for subcloning because it contains only the λ cI headpiece and linker as well as a C-terminal cysteine residue. The elimination of MBP sequences reduced the interference of the periplasmic domain with TM dimerization that was observed for other, well characterized dimerizing TM domains such as GpA TM.^{15,16} Interchain disulfide bonding of the terminal cysteine residues, which is dependent upon TM domain dimerization,¹⁵ allowed for identification of chimera dimers on SDS gels (data not shown). The pJAH01 subclones were assayed for their ability to confer immunity to λ cI phage by the quantitative λ cI plaque assay (see Materials and Methods).

Table 3 shows the results of phage immunity assays for clones expressing the subsequences. Six out of the 11 subsequences surveyed did not continue to confer λ cI immunity when fused to the λ cI headpiece. There might be several reasons to account for the failure of the fusions to confer immunity in the new context. Three of the new fusions produced unstable proteins. These included the fusions containing the out-of-frame inserts from *ydcP* and *ymcA* and the fusion containing the non-coding DNA. These protein sequences were stable in the λ cI-TM-MBP tri-hybrid, however, probably because the MBP portion of the fusion protein stabilizes the preceding sequences (J.A.L. & J.B., unpublished results; W. Russ, personal communication). The reason for failing to confer immunity in the other three cases (*ybeV*, *yhfC*, and *fdoI*) might be either that the dimerizing region had been removed or because this new fusion protein has a weaker ability to dimerize than the former.²²

Characterization of the multimerization capability of YjiO TM6

The data in Table 3 demonstrate that several of the TM subsequences continued to show evidence

of dimerization upon subcloning. In order to gain further information about the mechanisms of helix-helix interactions, we have initially chosen to characterize potential TM interaction domain(s) isolated from the *E. coli* protein YjiO in more detail. This TM segment provided the highest level of immunity to λ cI phage of the ones that we examined upon subcloning. The TM domain we selected is the sixth TM segment out of 12 predicted TM domains in YjiO. The protein encoded by the *E. coli* open reading frame (ORF) *yjiO* is a member of the major facilitator superfamily of multi-drug resistance proteins.²³ YjiO has a high degree of homology to *E. coli* proteins MdfA and Cmr and has an overexpression phenotype of multi-drug resistance.²³ Recent 2D electron diffraction experiments suggest that one member of the major facilitator superfamily of multi-drug resistance transporters, the TetA protein, may form a trimeric structure in the bacterial membrane.²⁴ In addition, it has been suggested that TMs 3, 6, 9, and 12 of these family members may face the membrane rather than the water lined channel due to their high hydrophobicities.²⁵ Other than these cursory studies, little is known about the contacts that may be important for multi-drug resistance protein structure and assembly. As an initial step in understanding the roles of TM domains in protein structure and assembly, we mapped the interaction interface(s) of YjiO TM6.

Comparison of the level of immunity to λ cI conferred by multimers of YjiO TM6 to known TM interaction domains

YjiO TM6 was identified as having at least one dimerization domain from a library of random inserts that were selected for their ability to export MBP and screened for their ability to confer dimerization of the λ cI headpiece. In order to estimate the multimerizing capability of YjiO TM6, we compared its ability to confer immunity to λ cI phage to that of known strong, weakly, and non-homointeracting TM domains (Figure 4). Lawns of *E. coli*

Table 3. Effect of subcloning random TM domains into pJAH01 on their ability to confer immunity to λ cI phage

Chromosomal locus	TM inserted	I/S ^a
<i>rtn</i>	LLLT C IVAGVMIAILV S CL	I
<i>ybeV</i>	ISLTVLVIGYALGKI A MLF	S
<i>yhfC</i>	VLFLSVAALCYILGQLGFIS W V	S
<i>ygfQ</i>	TGIGIGL F LLLIAANGVGM V	ND
<i>yejE</i>	LMLTLCS S VMGVL A LAGALQ G T	I
<i>ydcP</i> (out of frame)	LFCTV P MLFIS A ALV L VPVIMP V I	S
<i>dgoT</i>	WHWV F IVTGGIGI I W S LI W F	I
<i>yjiO</i>	VL F AI I AVMG F IS F VGL L L A M	I
Noncoding DNA	LLGLCC F AWATLRL A PY L FL	S
<i>fdoI</i>	LIVVIMV H IYAALW V KGTIT A M V	S
<i>ydjN</i>	VLVGLVMG V V F GL A L H TI Y GS	I
<i>ymcA</i> (out of frame)	LK I FL P LVG A T V FK I AL	S

^a Immunity is defined as a decrease in the efficiency of plating λ cI phage on clones expressing the cI headpiece-TM chimeras, when induced with 25 μ M IPTG. I, immune; S, sensitive.

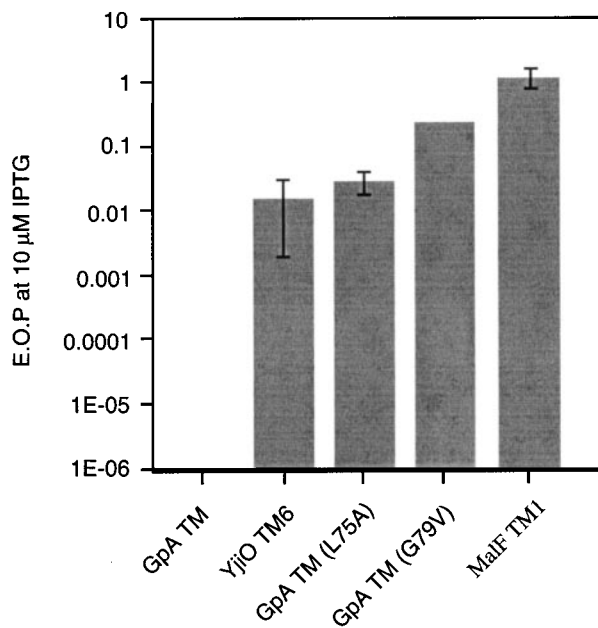


Figure 4. Comparison of efficiency of plating λ *cI* phage on λ *cI* headpiece-YjiO TM6 to that of *cI* headpiece-TM chimerae containing strongly, weakly, or non-dimerizing TM segments. Quantitative assays of E.O.P. λ *cI* phage on various *cI* headpiece-TM chimerae induced with 10 μ M IPTG were performed as described in Materials and Methods. A graphical display of the E.O.P. for each chimera is depicted in the figure. Values denote the mean E.O.P. and standard deviations from five to ten independent λ phage assays.

expressing a fusion of λ *cI* headpiece to YjiO TM6, the strongly dimerizing GpA TM, the weakly dimerizing GpA TM mutants L75A and G79V, or the non-dimerizing MalF TM1 were spotted with dilutions of λ *cI* phage by the standard phage assay. Previous work demonstrated that the L75A mutation is moderately disruptive *in vitro* and *in vivo*, whereas the G79V mutation is strongly disruptive *in vivo* and completely disruptive *in vitro*.^{11,15} Figure 4 shows the efficiencies of plating λ *cI* on *E. coli* expressing those constructs when induced with 10 μ M IPTG, which allowed for intermediate phenotypes. The data suggest that the strength of YjiO TM6 interaction(s) is/are weaker than wild-type GpA TM dimerization, but stronger than those conferred by the GpA TM (L75A) and (G79V) mutants.

Amino acid replacement mutagenesis of YjiO

Lemmon *et al.*¹⁰ demonstrated the utility of amino acid mutational analysis for identifying TM domain residues that are important for GpA TM dimerization. In that report, even fairly conservative changes at dimer interface positions resulted in dramatic decreases in dimerization capacity of the helices. Elegant and extensive analyses of the GpA TM dimer by this and other methods

demonstrated that the interface positions G79 and G83 were especially critical for helix-helix interactions.^{11,12}

Russ & Engelman²⁶ randomized the GpA TM dimer interface positions, while fixing non-interface positions as either all alanine or all leucine, and selected the randomized sequences for strong dimerizing capability. From this survey, the pattern G₈₃LLLA₈₇ emerged in four of the sequences selected from the leucine-backbone library. In a companion study, Senes *et al.*¹⁷ found that the motif GxxxA was overrepresented by 21% ($p > 3.6 \times 10^{-21}$) above that expected at random in all TM sequences identified to date. These studies suggested that, like the GxxxG motif, there is a structural and/or functional role for the GxxxA motif in TM domains. We identified this pattern in the YjiO TM6 sequence, and in a similar location within the hydrophobic TM stretch as it resided in the GpA TM-like domains selected from the leucine-backbone library. Therefore, we initially hypothesized that this motif (GxxxA) was critical for YjiO TM6 dimerization.

In order to experimentally determine which residues were important for YjiO TM6 dimerization, we individually changed most residues in the TM domain to alanine, with additional individual substitutions of G16S and A20G, L, or S (see Supplementary Material, Table 2, for details of mutagenesis). Alanine was chosen as the scanning replacement residue because while being relatively non-disruptive to secondary structure, it is also small and has a relatively hydrophobic side-chain. We measured the ability for the *cI* headpiece-mutated TM chimeras to confer immunity to λ *cI* phage by the standard phage assay (see Materials and Methods). The results of the amino acid substitution experiments are described in Figure 5.

At the G16 position of YjiO TM6, the substitutions to alanine or serine, a residue with a somewhat polar side-chain, had no effect on λ *cI* infection. This was in sharp contrast to the moderately interfering effect of the GpA TM G79A and dramatically interfering effect of the GpA TM G83A or G83S substitutions on GpA TM dimerization.¹⁰ At the A20 position, a change to the small, less hydrophobic flexible glycine residue or the bulkier leucine residue had no effect on λ *cI* immunity (not shown, for clarity). The change of A20 to serine, a more polar substitution, had a modest effect on λ *cI* immunity. In contrast, changes of L17, L18, and L19 to alanine, the "xxx" positions within the GxxxA motif, as well as V15A had dramatic effects on immunity to λ *cI* phage. Changes at most other positions had little or no effect on the ability of YjiO TM6 to multimerize and thus confer repression of λ *cI* phage. These data caused us to reject the hypothesis that the G16 and A20 residues were most critical for multimerization, and instead focus our attention on the entire V15-A20 region.

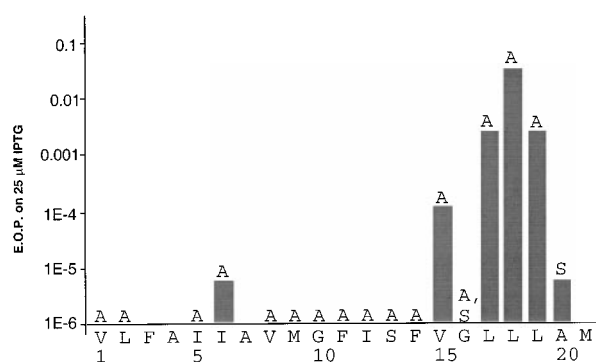


Figure 5. Effect of alanine replacement scanning mutagenesis of YjiO TM6 on repression of λ cI phage by cI headpiece-TM chimeras. The sequence and residue numbering system of the YjiO TM6 is displayed on the X-axis. The efficiency of plating λ cI phage is displayed on the Y-axis. Amino acid replacements are indicated above each position where a change was made. For changes that increase the efficiencies of plating λ cI phage, the new E.O.P. are indicated by the bars. For amino acid replacements that do not increase the E. O. P., the value remains $\leq 10^{-6}$.

Molecular modeling of the YjiO TM6 interaction interface(s)

In order to identify potential structures for YjiO TM6 complex, we used the mutagenesis data to sort through a set of possible homodimeric structures obtained by computational modeling of the wild type YjiO TM6 sequence (see Materials and Methods). The modeling program uses a combination of a grid of mutual rotation starting points followed by a molecular dynamics simulation that explores the local energy minima of the entire two-helix rotational interaction space for both left-handed and right-handed crossing angles.²⁷ After each molecular dynamics simulation, the helices are annealed and structures with minimum energies for those parameters are output. Similar structures from each run are clustered according to a maximum limit set on the average deviation of their α -carbon backbone atomic coordinates. Each cluster has an average structure. We performed three independent molecular dynamics simulations and generated thirty three average structures. These possible models were narrowed down based on the following criteria.

Each simulation generated three clusters that were more highly populated with similar structures than the other clusters from that simulation. In addition, the three most highly populated clusters from the first run appeared to represent the same models as the three most highly populated clusters in the two subsequent runs. Because of the persistence and highly populated nature of these clusters, we reasoned that these clusters represented the three best potential models for TM6 helix-helix interactions. In support of our exper-

imental data, the models represented by the three most highly populated clusters had the lowest potential energy values for the V15-A20 region. Further analyses of all 33 potential models supported our claims.

Since the V15-A20 region was most sensitive to amino acid changes, we hypothesized that residues in this region constituted part of the helix-helix interactions interface(s). Residues at a dimerization interface have close interatomic distances. Therefore, each intermolecular interatomic distance was calculated for each model structure. The same three highly populated clusters that were selected from the previous analysis represented the models that had atoms with the closest contacts residing in the region V15-A20. All three models described asymmetrical interfaces, the homodimeric interactions lacking pseudo 2-fold symmetry, and all three models predicted helix pairs with left handed crossings.

Figure 6 depicts views of the three models for YjiO TM6 interactions. The left-hand side of the Figure shows the relationship of the two helices for each model. The first helix is fixed in the same position for each structure so that one can see the different faces of interactions proposed by the models. The right-hand side of the Figure demonstrates an example of how a mutation in the proposed dimer interface region would affect the van der Waals interactions of the two helices at that position. All three models compare the wild type sequence at L18 (red) to the alanine substitution at that position, and its effect on van der Waals interactions with residue A20 (blue). Coordinates for the three models are available as PDB files at http://beck2.med.harvard.edu/protein_structure/interface.htm.

Because the mutagenic analyses demonstrated that sequential residues (L17L18L19) were important participants in helix-helix interactions, and the computational analyses resulted in multiple likely models of asymmetric homodimeric interactions, it is reasonable to hypothesize that the YjiO TM6 segment contains more than one interaction interface, and is part of a higher-order complex. Continued mutagenic analysis, combined with biophysical characterization of the homomultimer will be required to further establish the nature of the interactions.

Discussion

We have described a gene fusion assay system that permits selection of export signals out of a library of random sequences from whole genomes and through a series of screening steps identifies those which contain TM domains that are capable of forming homomultimeric helix-helix interactions. The reporters for this system consist of two domains flanking a random DNA insertion site. The N-terminal domain of the chimera that

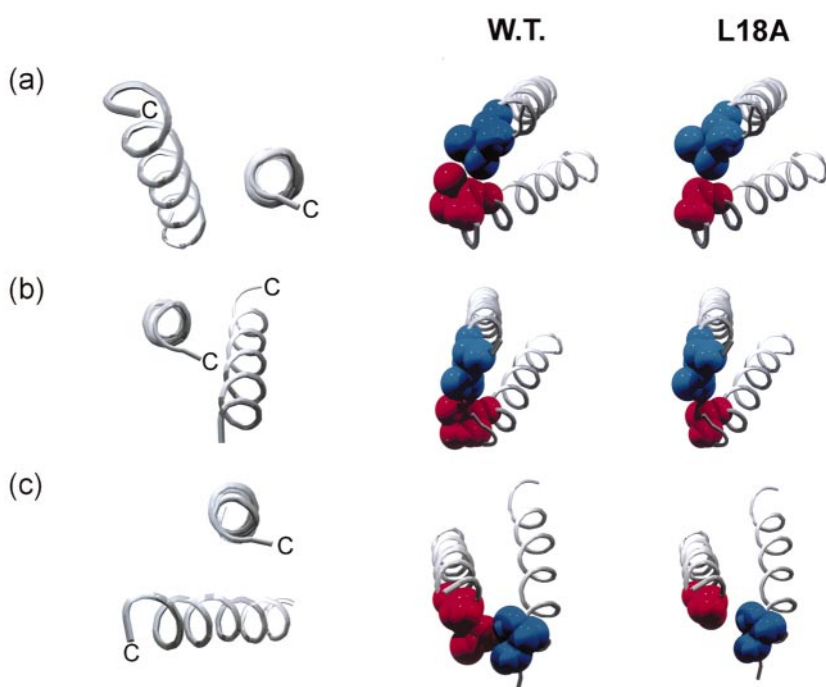


Figure 6. Models for proposed structures of YjiO TM6 homodimeric interactions. (a), (b) and (c) Three proposed models for YjiO TM6 helix-helix interactions. The left-hand panels depict the relationship between the two helices when helix one is held in a similar position for each model. C indicates the position of the C termini for each helix. Note the differing helix-helix interfaces described by each model. The right-hand panels demonstrate the van der Waals surfaces of residues L18 (red) and A20 (blue) (W.T.) compared to the same interactions when L18 is mutated to alanine (L18A).

reports multimerization consists of the DNA-binding domain of the bacteriophage λ repressor, which requires dimerization for binding to its cognate operator DNA.²⁰ The C-terminal domain of the chimera which reports insertion of an export signal consists of the mature portion of the *E. coli* maltose-binding protein. MBP requires a signal sequence or a transmembrane domain for proper localization to the periplasmic side of the *E. coli* inner membrane. The MBP domain, when localized to the periplasm, participates in the uptake of maltose and allows for selection for growth on maltose as a sole carbon source. Using this system, we have identified several amino acid sequences that, when fused to the λ cI headpiece, confer immunity to λ cI phage infection. We further characterized one of those novel TM interaction domains, YjiO TM6, using a combination of mutagenic and computational analyses. Our data suggest that the C-terminal end of the YjiO TM6 contains one or more helix-helix interaction interfaces that are unlike that reported for the well characterized glycoporphin A TM dimerization domain.⁸

One of the most extensively studied TM homomultimerization domains is the dimerization domain from the human erythrocyte protein glycoporphin A (GpA TM). Biochemical and biophysical approaches, *in vivo* genetic analyses, and molecular dynamics all support a critical role for the G₇₉xxxG₈₃ motif in GpA TM dimerization.^{8,12,15,28,30} One hypothesis states that the presence of a GxxxG motif in a transmembrane domain indicates that that segment is likely to participate in helix-helix interactions within that region.^{17,20}

Is it really possible to predict possible helix-helix interaction domains? Russ *et al.*²⁶ and Senes *et al.*¹⁷

describe companion studies that recognize, both genetically and computationally, other TM sequences that confer dimerization and are found to be overrepresented in the population of all TM domains identified to date. One of those sequences is comprised of the amino acid pair GxxxA¹⁷ and more specifically GLLLA.²⁸ Russ & Engelman²⁶ identified the GLLLA motif in sequences selected for their strong propensities to multimerize. They acknowledge that, although the backbone for the random library of sequences is the GpA TM domain, where the randomization occurred in the positions corresponding to the GpA TM interface residues, some of the sequences selected as strongly multimerizing may in fact now behave as left handed or contain asymmetrical interfaces. We identified this sequence within YjiO TM6 and hypothesized that it contained dimer interface residues. What we found, however, is that in contrast to the critical nature of the 1 and 5 positions of the GpA TM GxxxG motif, in YjiO TM6 the three intervening leucines (GLLLA) occupied the positions most sensitive to alanine replacement mutagenesis.

When the YjiO TM6 segment was subjected to molecular modeling, several possible models for YjiO TM6 helix-helix interactions prevailed, all of which depict the α -helices in left-handed crossings with asymmetrical interfaces. These two features are entirely distinct from those of the widely studied GpA TM interface, which consists of a right handed crossing of symmetrical TM segments.¹¹ These particular models were first selected because they represent average structures of highly populated clusters of similar, and therefore suggestive of chemically good, structures. In support of the cluster popularity argument, computational ana-

lyses demonstrated that these models all contained interface residues residing within the V15-A20 region whose van der Waals interactions could be altered by mutations that proved to be highly disruptive in the alanine replacement experiments. None of the likely models represent symmetrical dimers. In addition, the likely models suggested three unique, nonsimultaneous interfaces. This combined with the preliminary structural analyses of the multidrug resistance homolog TetA, which suggest a homotrimeric complex, raise the possibility that the YjiO TM6 helix is involved in a multimeric rather than a dimeric helix-helix interaction. Further mutational analysis and modeling coupled with biophysical studies will be required to establish the size of the YjiO TM6 homomultimeric complex and to test the proposed models.

The goal of our study was to identify helix-helix packing motifs within naturally occurring protein TM domains. However, the types of interactions that we screened for and the models that we generated for the structures of the YjiO TM6 interactions only address the homomultimeric nature of the TM domains. These segments may also be involved in hetero-interactions, perhaps with an antiparallel partner TM in the native protein. By continuing to screen for more homomultimerizing segments, we can construct a catalog of motifs that participate in close helix-helix associations. Ultimately we wish to identify these sequence motifs in naturally occurring proteins and determine, both computationally and experimentally, whether our catalog has any predictive value on the role of these motifs in whole membrane protein structure and assembly.

Technical considerations of the multimerizing TM domain assay

The construction of a TM domain library demands that relatively small DNA fragments be generated. Typical TM domains are in the range of 17-20 amino acid residues. In order to bias the library towards clones containing small inserts, we chose two sets of digestion conditions: one consisting of a full digest with two four-base cutters *AluI* and *RsaI*, and one consisting of a partial digestion with *CviJI**, a mutant four-base cutter used in a permissive buffer with the recognition site RGC.²¹ Computational analyses were performed on the *E. coli* W3110 genome to predict the number of clones generated from each digest that would constitute a complete TM domain library. Our partial library of 477 TM domains constitutes only a small sampling of the total number of TM domains that we estimated would be generated by this approach. Given the successes that we found in screening this small sample, we anticipate that a large amount of information regarding helix-helix interactions will be made available by further screening of such libraries.

Of the 12 sequences that were identified as being potentially dimerizing, 50% came from known integral membrane protein transporters (DgoT, YjiO) or putative proteins with homology to integral membrane transporters (YhfC, YgfQ, YejE, YdjN). That this class predominates may be a result of the number of proteins in the transporter classes (125 out of an approximately 900 *E. coli* predicted integral membrane proteins in the *E. coli* genome (SwissProt database²⁹) and the large number of transmembrane domains found in these categories of proteins (typically 6-14 TMs).

Other TM domains may be underreported in the assay system due to several reasons. First, there are TM sequences that may not be available for selection on maltose as a sole carbon source. This could occur for several reasons, the most simple being that the DNA encoding a particular transmembrane domain is restricted by the enzymes chosen for fragment generation. This situation is minimized, however, by combining the complete *AluI/RsaI* and partial *CviJI** libraries. It is also possible that an insert would encode a sequence that is a target for proteolysis, which would prevent a stable fusion protein containing that TM from forming. Second there are potential interacting TM sequences that are selected on maltose but fail to confer immunity to λ cI at the plaque screening stage. This could occur because the inserts contain "extramembraneous" sequences that interfere with dimerization of either the TM domain or the λ cI headpiece domain. Finally, there is a large class of inserts that are selected on maltose and confer immunity, but upon Western analysis of the chimera using antibodies to both the cI and MBP domains, exhibit multiple products rather than a single, full length cI-ES-MBP chimera. This may be due to protein degradation *in vivo* or upon extraction, or to the presence of translational start signals within the insert, leading to synthesis of two stable proteins. In either case, the clones were not considered further because of the possibility that the stability of the soluble cI headpiece entity would allow a sufficient accumulation of cytoplasmic cI headpiece such that its concentration approximates that which is capable of dimerizing and conferring immunity in the absence of a specific C-terminal dimerization domain. We found instances of this occurring when multiple (three or more) positive charges were applied immediately C-terminal to the cI linker region with no TM domain attached.

In addition to potentially missing some TM domain interactions, the multimerization screen also has the potential to report false positives. As described above, some inserted sequences may contain non-dimerizing or weakly dimerizing TM domains that are flanked by "extramembraneous" sequences which contribute to dimerization. These clones are not eliminated at the Western screening stage, and are only confirmed upon subcloning of the isolated TM fragment.

Once the clones containing inserts that conferred immunity to λ cI phage and encoded full length

fusion proteins were identified, the apparent TM domains were subcloned into an alternative vector. This subsequent analysis was performed to determine whether the apparent TM domains were necessary and sufficient for the phenotype of λ immunity. There was no information available in any database regarding the tertiary structures of any of the TMs identified as possibly multimerizing by our screen. As we observed, only five out of the 11 subclones continued display that phenotype. There are several explanations for this phenomenon. First, it is possible that the window of apparent TM subsequence chosen for subcloning was not the optimal choice. Mingarro *et al.*³⁰ and Orzaez *et al.*³¹ have suggested that the choice of the endpoints of the TM segment has a significant effect on the dimerizing capacity of the GpA TM domain. For example, the choice of the YhfC subsequence could move up to four amino acid residues N-terminal to the window chosen, and the YbeV subsequence move up to seven amino acid residues N-terminal to the window chosen, and they would still maintain a highly hydrophobic nature. The windows that were chosen were done so with an effort to maximize hydrophobicity, and to not include what appeared to be the natural N-terminal stop transfer sequences (usually two or more positively charged residues).

Three of the sequences identified by our multimerization screen constituted segments that were either out of the normal coding frame for that locus (*ymcA*, *ycdP*) or were from a locus that does not normally encode a protein (noncoding DNA). We and others have shown that when amino acid sequences are fused N-terminal to the mature MBP domain, those fusion proteins are stably expressed. However, when we subcloned those sequences into a vector that did not have the MBP domain, the presence of a protein sequence that is not normally expressed in *E. coli* may have resulted in instability of the fusion protein. Western analyses failed to reveal any full-length fusion protein encoded by these subclones, supporting the

provide important sequence and structural information regarding the tertiary and quaternary organization of integral membrane proteins. Here, we demonstrated the potential utility of the *in vivo* random TM multimer assay system to identify novel helix-helix interaction domains as part of a successful route to membrane protein structural information.

Materials and Methods

Bacteria, phage, and plasmids

The bacterial strains and plasmids used in this study are listed in Table 1. For routine culturing, *E. coli* strains were grown on NZ agar plates or in NZ medium.³² When necessary, antibiotics were added to the following final concentrations: kanamycin, 20 $\mu\text{g}/\text{ml}$; tetracycline, 12.5 $\mu\text{g}/\text{ml}$. Phage λ *cl* used in efficiency of plating assays was grown and titered on JAH013³³ by standard methods.³⁴ The low copy number plasmid pACYC184³⁵ was the vector from which the recombinant plasmids used in this study were derived.

Construction of *E. coli* for TM library selection: JAH143 (Table 1) carrying the MBP deletion *malE444* was constructed by first transducing JAH013 (Table 1) to rifampicin resistance with P1 grown on MM105 (*malE*, *metA*, *rpoB*) (laboratory collection). That strain was then transduced to *met* + , and screened for rifampicin sensitivity, with P1 grown on MM39 (*malE*) (laboratory collection) to create JAH143. JAH143 was tested for endogenous sensitivity to λ phage. The frequency of resistant mutants in cultures of both JAH143 and its parent MC1061 (=JAH013) is 10^{-2} .

Library plasmid construction

All fusion constructs initiating with the λ headpiece were expressed from a *lacUV5* promoter on pACYC184 derived plasmids. Library plasmids pJAH099 and 104-111 (Table 1 and Figure 1(a)) which have all nine reading frame combinations were constructed as follows: a PCR product coding for the periplasmic linker from the *Vibrio cholera* ToxR protein followed by the mature *E. coli* MBP was amplified from the template pKKgpa4¹⁶ using one of three upstream primers:

CCGGTTAAC**CCA**AGCCAATCC
CCGGTTAAC**CCA**AGCCAATCC
CCGGTTAAC**CC**AGCCAATCC

*Hpa*I-ToxR-MBP (L) rf1
*Hpa*I-ToxR-MBP (L) rf2
*Hpa*I-ToxR-MBP (L) rf3

hypothesis that these sequences are unstable. Finally, for all of the sequences that failed to confer immunity to λ upon subcloning, it is possible that the TM alone was not sufficient for dimerization of the λ headpiece. By biasing the library towards small inserts, we attempted to minimize this class of clones. However, we do anticipate that they will arise at some frequency.

In summary, multidisciplinary approaches to membrane protein structure and assembly will

which create a *Hpa*I site (underlined) in one of three reading frames (boldface), and the downstream primer (TCCCCCCGGGTTAGTCTGCGCGTCTTTCAGGGCTTC mbp(*Sma*I)right) which creates a stop codon (boldfaced) and a *Sma*I site (underlined). PCR products coding for the λ *cl* N-terminal DNA binding domain and λ linker were amplified from the template pJAH014 (this laboratory) using the upstream primer (CCGGAT CCACAAAAAGAAACCATTAACACAAG lambda 5'(*Bam*HI)) and one of three downstream primers:

CCGGTTAAC CTTT CGTTTGCTTACCCATCTCTCCGC	lambda right (<i>HpaI</i>) rf1
CCGGTTAAC GCTTT CGTTTGCTTACCCATCTCTCCGC	lambda right (<i>HpaI</i>) rf2
CCGGTTAAC TTT CGTTTGCTTACCCATCTCTCCGC	lambda right (<i>HpaI</i>) rf3

which create three or four positively charged residues (KRK, KRKR, and KRK respectively) followed by a *HpaI* site (underlined), in one of three reading frames (bold-faced). The resulting fragments were digested with *HpaI* and ligated together in all nine possible combinations of a λ *cl* headpiece containing fragment and an MBP containing fragment. The joined fragments were then reamplified with the λ 5' (*Bam*HI) upstream primer and the *mbp* (*Sma*I)right downstream primer, digested with *Nsi*I and *Bss*HII, which cut in the *cl* and *malE* coding regions, respectively, and ligated into pJAH087¹⁵ digested with *Nsi*I and *Bss*HII to create pJAH099, 104-111.

Production of *E. coli* export domain library

Chromosomal DNA from *E. coli* W3110 was extracted by the guanadinium HCL method of Pitcher *et al.*³⁶ A 2 μ g sample of DNA was digested either to completion with *Alu*I and *Rsa*I, or partially with *Cvi*JI*; 600 ng of digested DNA was ligated to 100 ng *Hpa*I digested, dephosphorylated, and pooled pJAH099, 104-111 in 1 \times NEB T4 polynucleotide kinase buffer plus 1 μ l of 1 mM ATP (final vol 25 μ l) and NEB high concentration T4 ligase at 4°C overnight. Ligated DNA (100 ng) was transformed into electrocompetent TG1 cells (Stratagene) and plated on NZ containing 12.5 μ g/ml tetracycline. About 1 million tetracycline resistant transformants arose, which were pooled and plasmid DNA was extracted by the Qiagen Midi Prep method. Plasmid DNA (2 μ g) was digested with *Hpa*I to remove any religated vectors, and then 100 ng of the purified *Hpa*I digested plasmid prep was transformed into electrocompetent JAH143, selecting for growth on M63 minimal medium containing methionine, leucine, 12.5 μ g/ml tetracycline, 10 μ M IPTG and maltose as a sole carbon source. After 48 hours, about 150 large colonies were visible in each transformation. A total of 477 of these were picked from six independent transformations (about 50% of the total), subcultured onto the same medium, and picked into NZ broth + glycerol in microtiter dishes for storage at -80°C.

TM library bacteriophage λ *cl* immunity assay

JAH143 expressing the pJAH099, 104-111 *mal+* clones were grown overnight in 96-well microtiter dishes containing 100 μ l NZ broth and 12.5 μ g/ml tetracycline for plasmid selection. Cells were back diluted 1:20 into microtiter dishes containing the same medium, in the presence or absence of IPTG, as indicated. Cells were grown to A_{600} 0.65 and 10 μ l of cultures were spotted onto NZ plates containing 12.5 μ g/ml tetracycline, IPTG at the same concentrations as in the growth medium, and 10⁵ plaque forming units of λ *cl* phage or no phage. After overnight incubation at 37°C, the *E. coli* spots were observed for the presence of plaques. Sensitive strains typically had 25-50 plaques/bacterial spot. Therefore, in this assay, immunity is defined as exhibiting an effi-

ciency of plating, expressed as number of plaques/number of pfu λ *cl* phage, of $\leq 2 \times 10^{-2}$.

Quantitative bacteriophage λ *cl* immunity assay

JAH143 derivatives expressing the vector pJAH01 derived λ *cl* headpiece-TM fusion plasmids were grown overnight in NZ broth containing appropriate antibiotics for plasmid selection. Cells were back diluted 1:100 into NZ broth plus antibiotics in the presence or absence of IPTG, as indicated. Cells were grown to A_{600} 0.65-0.70 and 100 μ l of culture were plated with 3 ml NZ top agar. After approximately 30 minutes, 10 μ l aliquots of λ *cl* phage containing 10-10⁶ plaque forming units (pfu) were spotted on the top agar. After overnight incubation at 37°C, the *E. coli* lawns were observed for plaques. In this assay the level of immunity is assessed by efficiency of plating (expressed as number of plaques/number of pfu), which was calculated as the highest dilution spot containing well isolated plaques (1-20 plaques). As a control for phage infection, lawns were also spotted with the virulent phage λ *vir*.³⁶

Subcloning TM domains into pJAH01

TM domains were amplified from the random library clones using the oligos described in Supplementary Material Table 1. The upstream oligonucleotides add an *Afl*III site (underlined) and positively charged residues to the N-terminal end of the TM, and the downstream oligonucleotides add an *Nco*I site (underlined) to the C-terminal end of the TM. The PCR products were digested with *Afl*III and *Nco*I and cloned into *Afl*III digested and dephosphorylated pJAH01.

Site directed mutagenesis for amino acid replacements in YjiO TM6

Amino acid substitutions were created by incorporation of mutagenic oligonucleotides during PCR amplification of the YjiO TM6 fragment. The oligos used to introduce the site-directed mutations are described in Supplementary material Table 2. PCR products were digested with *Afl*III and *Nco*I and cloned into pJAH01 as described in the previous section.

Molecular modeling

A global molecular dynamics simulation strategy was used to identify structures that settle into local potential energy minima from a search of all possible rotational interactions of restrained α -helices. All calculations were performed using the crystallography and NMR System (CNS).²⁹ Helix modeling was conducted using the CNS Searching of Helix Interactions (CHI, version 1.1).^{12,37,38} A description of the CHI program can be viewed at <http://www.csb.yale.edu/userguides/datamanip/chi/html/chi.html>. All searches conducted were specified to be parallel homo-oligomeric two-helix searches.

Acknowledgments

The authors thank the members of the Beckwith and Engelman Labs for many helpful discussions and careful reading of the manuscript. We also thank Daniel Ritz for help with graphics, Terri Luna for medium preparation and Ann McIntosh for administrative assistance. This work was supported by the National Institutes of Health (grant GM18569-01 to J.A.L and grant GM54160 to J.B and D.M.E.). J.B. is an American Cancer Society Research Professor. D.M.E. is Eugene Higgins Professor of Molecular Biophysics and Biochemistry.

References

- Guzman, L. M., Weiss, D. S. & Beckwith, J. (1997). Domain-swapping analysis of FtsI, FtsL, and FtsQ, bitopic membrane proteins essential for cell division in *Escherichia coli*. *J. Bacteriol.* **179**, 5094-5103.
- Ehrmann, M., Ehrle, R., Hofman, E., Boos, W. & Schlosser, A. (1998). The ABC maltose transporter. *Mol. Microbiol.* **29**, 685-694.
- Weiss, A. & Schlessinger, J. (1998). Switching signals on or off by receptor dimerization. *Cell*, **94**, 277-280.
- Chervitz, S. A., Lin, C. M. & Falke, J. J. (1995). Transmembrane signaling by the aspartate receptor: engineered disulfides reveal static regions of the subunit interface. *Biochemistry*, **34**, 9722-9733.
- Hegde, R. S., Mastrianni, J. A., Scott, M. R., DeFea, K. A., Tremblay, P., Torchia, M. *et al.* (1998). A transmembrane form of the prion protein in neurodegenerative disease. *Science*, **279**, 827-834.
- Wigley, W. C., Vijayakumar, S., Jones, J. D., Slaughter, C. & Thomas, P. J. (1998). Transmembrane domain of cystic fibrosis transmembrane-conductance regulator: design, characterization, and secondary structure of synthetic peptides m1-m6. *Biochemistry*, **37**, 844-853.
- Lemmon, M. A. & Engelman, D. M. (1994). Specificity and promiscuity in membrane helix interactions. *Quart. Rev. Biophys.* **27**, 157-218.
- 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 α -helices. *J. Biol. Chem.* **267**, 7683-7689.
- Popot, J. L. & Engelman, D. M. (1990). Membrane protein folding and oligomerization: the two-stage model. *Biochemistry*, **29**, 4031-4037.
- Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J. & Engelman, D. M. (1992). Sequence specificity in the dimerization of transmembrane α -helices. *Biochemistry*, **31**, 12719-12725.
- MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. (1997). A transmembrane helix dimer: structure and implications. *Science*, **276**, 131-133.
- Adams, P. D., Engelman, D. M. & Brunger, A. T. (1996). Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins: Struct. Funct. Genet.* **26**, 257-261.
- 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.
- Zhou, F. X., Merianos, H. J., Brunger, A. T. & Engelman, D. M. (2001). Polar residues drive association of poly-leucine transmembrane helices. *Proc. Natl Acad. Sci. USA*, **98**, 2250-2255.
- Leeds, J. A. & Beckwith, J. (1998). Lambda repressor N-terminal DNA-binding domain as an assay for protein transmembrane segment interactions *in vivo*. *J. Mol. Biol.* **280**, 799-810.
- 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.
- 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 β -branched residues at neighboring positions. *J. Mol. Biol.* **296**, 921-936.
- Bowie, J. U. (1997). Helix packing in membrane proteins. *J. Mol. Biol.* **272**, 780-789.
- Ptashne, M. (1986). *A Genetic Switch*, Blackwell Scientific Publications & Cell Press, Cambridge, MA.
- Weiss, M. A., Pabo, C. O., Karplus, M. & Sauer, R. T. (1987). Dimerization of the operator binding domain of phage λ repressor. *Biochemistry*, **26**, 897-904.
- Skowron, P. M., Swaminathan, M., McMaster, K., George, D., Van Etten, J. L. & Mead, D. A. (1995). Cloning and applications of the two/three-base restriction endonuclease R-CviJI from IL-3A virus-infected *Chlorella*. *Gene*, **157**, 37-41.
- Langosch, D., Brosig, B., Kolmar, H. & Fritz, H. (1996). Dimerization of the glycophorin-A transmembrane segment in membranes probed with the ToxR transcription activator. *J. Mol. Biol.* **263**, 525-530.
- Edgar, R. & Bibi, E. (1997). MdfA, an *Escherichia coli* multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. *J. Bacteriol.* **179**, 2274-2280.
- Yin, C. C., Aldema-Ramon, M. L., Borges-Walmsley, M. I., Taylor, R. W., Walmsley, A. R. *et al.* (2000). The quaternary molecular architecture of TetA, a secondary tetracycline transporter from *Escherichia coli*. *Mol. Microbiol.* **38**, 482-492.
- Goswitz, V. C. & Brooker, R. J. (1995). Structural features of the uniporter/symporter/antiporter superfamily. *Protein Sci.* **4**, 534-537.
- Russ, W. P. & Engelman, D. M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.* **296**, 911-919.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P. *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905-921.
- Fleming, K. G., Ackerman, A. L. & Engelman, D. M. (1997). The effect of point mutations on the free energy of transmembrane α -helix dimerization. *J. Mol. Biol.* **272**, 266-275.
- Boyd, D., Schierle, C. & Beckwith, J. (1998). How many membrane proteins are there? *Protein Sci.* **7**, 201-205.
- Mingarro, I., Elofsson, A. & von Heijne, G. (1997). Helix-helix packing in a membrane-like environment. *J. Mol. Biol.* **272**, 633-641.
- Orzaez, M., Perez-Paya, R. & Mingarro, I. (2000). Influence of the C terminus of the glycophorin A transmembrane fragment on the dimerization process. *Protein Sci.* **9**, 1246-1253.

32. Guzman, L. M., Baroness, J. J. & Beckwith, J. (1992). FtsL, an essential cytoplasmic membrane protein involved in cell division in *Escherichia coli*. *J. Bacteriol.* **174**, 7716-7728.
33. Leeds, J. A. & Beckwith, J. (2000). A gene fusion method for assaying interactions of protein transmembrane segments *in vivo*. *Methods Enzymol.* **327**, 165-175.
34. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1992). *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
35. Chang, A. C. Y. & Cohen, S. N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. *J. Bacteriol.* **134**, 1141-1156.
36. Pitcher, D. G., Saunders, N. A. & Owen, R. J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Letters Appl. Micro.* **8**, 151-156.
37. 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.
38. Adams, P. D. & Brunger, A. T. (1997). Towards prediction of membrane protein structure. In *Membrane Protein Assembly* (von Heijne, G., ed.), R. G. Landes Co., Austin, TX.

Edited by G. von Heijne

(Received 5 June 2001; received in revised form 10 August 2001; accepted 12 August 2001)



<http://www.academicpress.com/jmb>

Supplementary Material for this paper comprising two Tables are available on IDEAL