JOURNAL OF BACTERIOLOGY, Jan. 1993, p. 553–556
0021-9193/93/020553-04$02.00/0

Vol. 175, No. 2

Analysis of the Topology of a Membrane Protein by Using a Minimum Number of Alkaline Phosphatase Fusions

DANA BOYD,* BETH TRAXLER,† AND JON BECKWITH

Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115

Received 19 August 1992/Accepted 2 November 1992

An approach to analyzing the topology of membrane proteins with alkaline phosphatase fusions is described. Precise fusions were constructed by using polymerase chain reaction at the C terminus of each hydrophilic region of the membrane protein. The disruption of topogenic signals is thereby minimized, and predictable anomalous results are avoided. The *Escherichia coli* MalG protein has been analyzed.

The use of gene fusions to study membrane protein topology is now a well-established technique (3, 5, 20). Hydropathy plots and the positive inside rule (11, 22) can be used to generate one or more topological models for a membrane protein from its predicted amino acid sequence. Gene fusions can be used to confirm or discriminate between or among such models. Random methods for generating gene fusions have been widely used for this purpose but are generally inefficient in the sense that many fusions must be constructed to obtain a complete analysis (1, 4, 6, 15, 16, 19, 23). In addition, anomalous results are often obtained when topogenic determinants are interrupted or deleted in the construction of fusions (3, 4, 6, 19, 20). Here we describe a strategy to simplify this analysis. We construct fusions only at the C termini of hydrophilic regions using polymerase chain reaction (PCR). This approach avoids the disruption of the two known classes of topological determinants (membrane-spanning segments and short hydrophilic regions with net positive charge) and requires the construction of a minimal number of fusions to discriminate among topological models.

To illustrate this approach, we have studied the *Escherichia coli* MalG protein. MalG is part of a complex that also contains the inner membrane protein MalF and the membrane-associated cytoplasmic protein MalK (8). This complex works in concert with the periplasmic maltose-binding protein to transport maltose across the inner membrane. This complex is a member of the ABC transporter family (2, 12, 13). A number of these transporters, including the mammalian multidrug transporter, are proposed to have a conserved 12-membrane-spanning segment structure (13). Fusion analysis of the MalF protein offers support for the 8-transmembrane-segment model (4) and the hydropathy analysis of MalG suggests that there are 6 segments (see below), bringing the total number of membrane-spanning segments to 14. Confirmation of the predicted topological structure of MalG will facilitate genetic and biochemical studies of the function of the complex. By constructing a set of alkaline phosphatase fusions to MalG that includes only those with fusion joints at the carboxy terminus of each of the hydrophilic regions (Fig. 1), we have obtained data that point unambiguously to a particular topological model.

Hydropathy analysis (Fig. 2) (9, 14) of the predicted MalG amino acid sequence (7) suggests that it may have six membrane-spanning segments and one large hydrophilic, extramembrane domain. Although a model with six membrane-spanning segments seems most likely, other topologies are conceivable. The clustering of basic residues in the first and third hydrophilic regions suggests that these regions are cytoplasmic.

To test this model, we have constructed fusions of the *malG* gene to the gene encoding alkaline phosphatase. Our choice of sites for fusion joints was governed by the rationale outlined above. Using site-directed mutagenesis, we have constructed fusions at the carboxyl end of each hydrophilic region. A single fusion obtained earlier by transposition of Tn*phoA*, which fortuitously was near the C terminus of the large hydrophilic region, was included in the set. Other fusions were constructed in vitro. First, the parent plasmid pDHB5752, shown in Fig. 3, was constructed. This plasmid complements the *malF* deletion of DHB4 (4) and the *malG* amber mutation of strain BT45 and has alkaline phosphatase activity that is encoded by a tripartite fusion of *xylE*, *malF*, and *phoA* and that is expressed from a promoter regulated by xylose.

Six fusions were constructed in vitro by using oligonucleotides (shown in Fig. 1) that introduce a restriction site, *Bsp*EI, at the fusion joint. This procedure results in conversion of the amino acyl residue at the fusion joint to proline in each case, and although in some cases, the third base of the preceding codon is changed, the residue itself is not changed. Two of the six fusions, GA1 and GA4, were constructed by oligonucleotide-directed deletion mutagenesis of plasmid pDHB5752, using T4 DNA polymerase methods described previously (4). Four additional fusions were constructed by a PCR method. First, a recipient plasmid, pGA4ΔBsp was constructed by deleting the material between the *Bsp*EI site in *malF* and that at the fusion joint of pGA4. This deletion removes a part of *malF* and all of *malG* to fuse *phoA* to *malF*. This fusion is to a cytoplasmic domain and has little alkaline phosphatase activity. The plasmid is similar to pDHB5752 (Fig. 3), except that the fusion joint is at the *Bsp*EI site in *malF*. Next, a fragment containing the desired segment of *malG* and part of *malF* was synthesized by using the oligonucleotides of Fig. 1 and another oligonucleotide (the common *malF* oligonucleotide, which corresponds to a sequence near the amino terminus of MalF) as primers for PCR. Each PCR product was digested with *Bsp*EI, gel purified, and ligated with pGA4ΔBsp that had been cut with the same enzyme. The resultant plasmids have the GA3, GA5, GA6, and GA7 fusion joints, each with its

* Corresponding author.
† Present address: Department of Microbiology, University of Washington, Seattle, WA 98195.

```
GA1                                17
CG CAA AAA GCT CGT TTA TTT ATT ACT CAT Ccg gac tct tat aca caa gta gcg tcc
   gln lys ala arg leu phe ile thr his pro asp ser tyr thr gln val ala ser


GA3                        124
GGC AAA GCG ACG CTG CTG AAA GGT Ccg gac tct tat aca caa gta gcg tcc
gly lys ala thr leu leu lys gly pro asp ser tyr thr gln val ala ser


GA4                        159
CCA TTC ATT GGC CTG AAT ACT CAT Ccg gac tct tat aca caa gta gcg tcc
pro phe ile gly leu asn thr his pro asp ser tyr thr gln val ala ser


GA5                        204
GCG ACA CCG TGG CAG GCC TTC CGT Ccg gac tct tat aca caa gta gcg tcc
ala thr pro trp gln ala phe arg pro asp ser tyr thr gln val ala ser


GA6                        262
CCG CAA AAC TAC CTG TGG GGT GAT Ccg gac tct tat aca caa gta gcg tcc
pro gln asn tyr leu trp gly asp pro asp ser tyr thr gln val ala ser


GA7                        296
CTG ACG GCA GGT GGT GTG AAA GGT Ccg gac tct tat aca caa gta gcg tcc
leu thr ala gly gly val lys gly pro asp ser tyr thr gln val ala ser
```

FIG. 1. MalG-PhoA fusion oligonucleotides. The sequences shown are the complement sequences of the oligonucleotides synthesized. Bases shown in uppercase type code for MalG residues, while those in lowercase type code for Tn*phoA* residues. Underlined bases are different from those in the sequence of MalG or Tn*phoA*. These changes introduce a *Bsp*EI restriction site but do not change the amino acid sequence of MalG or Tn*phoA*, except to convert the residue at the fusion joint to a proline in all cases. The possibility of introducing this site was discovered by using the computer program Degenerate DNA Matcher, which is available as Macintosh-compatible software from Dana Boyd. The number above the last uppercase (MalG) residue is the number of that residue in the MalG sequence (9). The sequence at the GA2 fusion joint, a Tn*phoA* transposition after residue 84 of MalG, is TGG Cct gac, which codes for Trp Pro Asp. The oligonucleotides were designed for deletion mutagenesis with T4 DNA polymerase, so are longer than required for the PCR method.

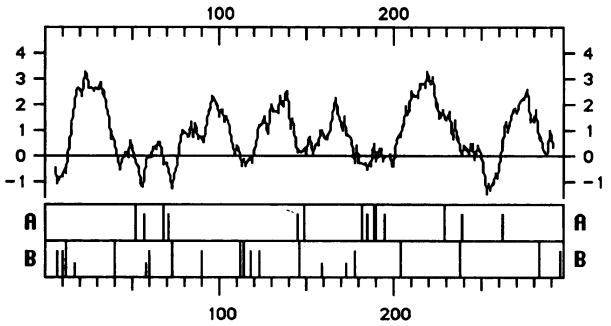

FIG. 2. Analysis of the MalG sequence. Hydropathy plot (above) and distribution of acidic (A) and basic (B) residues (below) in MalG. The average local hydrophobicity at each residue calculated by the method of Kyte and Doolittle (14) is plotted on the vertical axis versus the residue number on the horizontal axis. Higher values represent greater hydrophobicity. At the bottom of the figure the positions of charged residues are represented by vertical lines. In strip A, longer vertical lines represent glutamyl and the shorter ones aspartyl residues. In strip B, arginyl, lysyl, and histidinyl residues are represented by the long, medium, and short lines, respectively. This figure is adapted from output of the program DNA Strider.

*Bsp*EI site in *malG*. The DNA sequence at the fusion joint was confirmed by sequence analysis of each fusion.

To determine relative alkaline phosphatase activities of the fusion proteins, both alkaline phosphatase activity and the rate of fusion protein synthesis were determined for the same culture of each strain. The CLG190 strains with the fusion plasmids (Table 1) were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 40 min. Alkaline phosphatase activity of a fraction was measured as described elsewhere (15). The rate of fusion protein synthesis was determined by pulse-labeling with [^{35}S]methionine (Amersham) immunoprecipitation and sodium dodecyl sulfate-polyacrylamide gel electrophoresis, followed by quantitation with a Phosphorimager as described previously (21). The normalized activity is calculated by dividing the amount of activity by the amount of pulse-label in the fusion protein-sized band for each fusion protein. This gives a units per counts per minute value for each fusion protein. Each of these values is then divided by 0.01 times the value for fusion GA4, the highest value. This calculation method sets the normalized activity of the most active fusion at 100 and gives values for the other fusions that correspond to a relative percentage of the highest value. This important aspect of the gene fusion approach eliminates problems that can arise if different fusions are expressed at different levels (18).

Our results (Fig. 4 and Table 2) show that the fusions to putative periplasmic regions of MalG have much higher (20- to 80-fold) activities than the fusions to the putative cytoplasmic regions. The activities of the periplasmic fusions were comparable to that of wild-type alkaline phosphatase.

TABLE 1. *E. coli* strains

Strain	Genotype	Reference or source
DHB4	F′ *lac pro lacI*^{q}/Δ(*ara leu*)*7697 araD139* Δ(*lac*)*X74 galE galK rpsL phoR* Δ(*phoA*) *Pvu*II Δ(*malF*)*3 thi*	6
CLG190	Same as DHB4, but *pcnB zad*::Tn*10 recA1*	Claude Gutierrez
DHB5752	Same as DHB4, but carries pDHB5752	This study
DHB5730	Same as CLG190, but carries pGA4ΔBsp	This study
DHB5728	Same as CLG190, but carries pGA1	This study
DHB5727	Same as CLG190, but carries pGA2	This study
DHB5733	Same as CLG190, but carries pGA3	This study
DHB5729	Same as CLG190, but carries pGA4	This study
DHB5735	Same as CLG190, but carries pGA5	This study
DHB5737	Same as CLG190, but carries pGA6	This study
DHB5739	Same as CLG190, but carries pGA7	This study
BT45	F′ *lac pro lacI*^{q}/Δ(*lac pro*)*X111 araD rpsL rpoB malG*(Am) *malT*(Con) *zhe*::Tn*10 thi*	This study

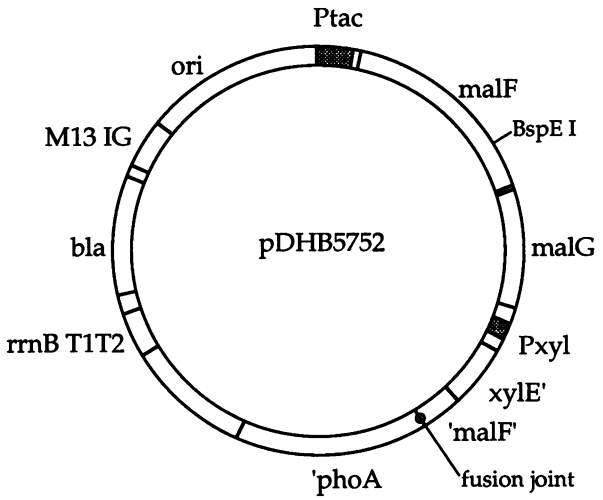

FIG. 3. pDHB5752 expresses wild-type copies of *malF* and *malG* from the *tac* promoter and a tripartite fusion composed of *xylE*, *malF*, and *phoA* from the *xylE* promoter. pDHB5752 was constructed by cloning a *Sac*I-*Pvu*II fragment of *E. coli* DNA from plasmid pHS2 into the M fusion plasmid derivative of pDHB32 (4) after the M fusion plasmid was digested with *Pfl*MI, trimmed with T4 DNA polymerase, and digested with *Sac*I. The intergenic region of phage M13 (M13 IG), the transcriptional terminator region of the *E. coli rrnB* gene (rrnB T1T2), the gene coding for β-lactamase (bla), and the pBR322 replicative origin (ori) are indicated. Transcription from both promoters shown and of *bla* is clockwise. The orientation of the M13 origin is such that the coding strand is packaged during phage growth. The pGA1 and pGA4 fusion plasmids were constructed using single-stranded DNA of pDHB5952 as a template and the corresponding deletion oligonucleotides (Fig. 1) as primers for in vitro second-strand synthesis. Plasmid pGA4ΔBsp, the recipient of the *Bsp*EI-cleaved PCR fragments used to construct PGA3, pGA5, pGA6, and pGA7, is a deletion derivative of pGA4 in which the fusion joint is at the *Bsp*EI site of MalF.

Therefore, the data support the topological model presented in Fig. 4 and make other possibilities less likely.

As was the case with *phoA* fusions to *malF* (4), only the C-terminal fusion GA7, in which the entire *malG* coding sequence is present, is capable of transporting maltose. The pGA7 plasmid confers a weak Mal$^+$ phenotype BT45 on MacConkey maltose indicator medium. Since the fusion protein is unstable (only 34% of it is intact after pulse-labeling and a 30-min chase [Table 2]), it is possible that a breakdown product and not the fusion protein itself is responsible for the phenotype.

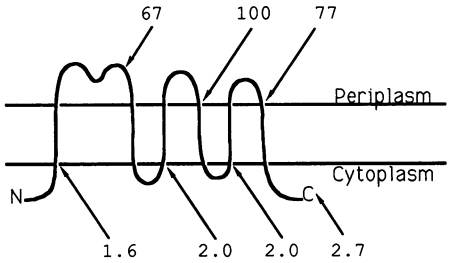

FIG. 4. Topological model for the MalG protein. The proposed topology of the MalG protein in the membrane is shown approximately to scale. The six proposed membrane-spanning segments are as follows: (i) Leu-18–Ile-37, (ii) Val-91–Ala-111, (iii) Gly-124–Phe-144, (iv) Gly-160–Il3-177, (v) Leu-205–Ile-227, and (vi) Phe-263–Ala-281. The normalized activities of each of the fusion proteins from Table 2 are shown with the arrows that point to the positions of the fusion joints in the topological model.

TABLE 2. Activities and stabilities of fusion proteins

Fusion protein	Alkaline phosphatase activity (U[a])		Stability[b]
	Not normalized	Normalized	
GA1	1.5	1.6	1.02
GA2	63	67	0.77
GA3	2.7	2.0	0.17
GA4	128	100	1.08
GA5	2.9	2.0	0.27
GA6	47	77	0.93
GA7	3.2	2.7	0.34

[a] Units of alkaline phosphatase activity and activity normalized to synthesis were measured as described in the text.
[b] Stability is the fraction of the amount synthesized in a pulse that remains intact after a 30-min chase.

The properties of the set of MalG-alkaline phosphatase fusions provide evidence for a particular topological model of the MalG protein of *E. coli*. The model was derived from hydropathy analysis of the MalG sequence and the observation that there are two clusters of basic residues in short hydrophilic regions. It is the most complex model that is suggested by the hydropathy analysis. In testing this model, we have constructed the minimal number of fusions necessary. The models with fewer membrane-spanning stretches are not consistent with the data. More-complex models that show an odd number of membrane-spanning segments inserted between the sites of two fusions are also ruled out, as the C terminus would then be expected to have the orientation opposite to that indicated. Thus, we believe that if a hydropathy analysis tends to strongly provide evidence for one particular model for topology of a membrane protein, that model can be readily tested by the approach we describe here. If the analysis gives two or more equally plausible models, then more-detailed fusion studies or other techniques may be required.

Although the approach we take here simplifies the analysis, it will not resolve all anomalous situations in gene fusion topology studies. In the LacY protein, for example, an alkaline phosphatase fusion to a known periplasmic domain has low activity, because it follows a poor export signal (6). Fusions to the pBR322 Tet protein exhibit a similar phenomenon (1). It has not yet been determined whether the sandwich fusion (10) approach will indicate topology correctly in cases such as these. Similarly, a β-lactamase fusion that fails to indicate topology correctly might not be avoided by our approach (23).

We wish to point out one other potential problem with fusion analysis, which is best illustrated by considering the large periplasmic domain of MalG. The fusion we obtained exhibited high activity, leading to the conclusion that the domain was periplasmic. However, it is possible that this domain contains an additional two (or a larger even number) membrane-spanning segments. While such putative membrane-spanning segments would necessarily be hydrophilic, adjacent pairs of hydrophilic membrane-spanning stretches have been proposed for some channel proteins (17). Presumably, such membrane-spanning segments would require interaction with other parts of the protein or with other proteins to be stably inserted. Replacement of the C terminus of the membrane protein by the large reporter moiety would be expected to disrupt such interactions.

This work was supported by grants from the National Science Foundation to D.B. and from the National Institutes of Health to J.B. J.B. is an American Cancer Society Research Professor.

We thank Howie Shuman for providing pHS2 and Claude Gutierrez for constructing CLG190.

REFERENCES

1. **Allard, J. D., and K. P. Bertrand.** 1992. Membrane topology of the pBR322 tetracycline resistance protein. J. Biol. Chem. **267:**17809–17819.
2. **Ames, G. F.-L., C. S. Mimura, and V. Shyamala.** 1990. Bacterial periplasmic permeases belong to a family of transport proteins operating from *Escherichia coli* to humans: traffic ATPases. FEMS Microbiol. Rev. **75:**429–446.
3. **Boyd, D.** The use of gene fusions to study membrane protein topology. *In* S. H. White (ed.), Membrane protein structure: experimental approaches, in press. Oxford University Press, Oxford.
4. **Boyd, D., C. Manoil, and J. Beckwith.** 1987. Determinants of membrane protein topology. Proc. Natl. Acad. Sci. USA **84:** 8525–8529.
5. **Broome-Smith, J. K., M. Tadayyon, and Y. Zhang.** 1990. Beta-lactamase as a probe of membrane protein assembly and protein export. Mol. Microbiol. **4:**1637–1644.
6. **Calamia, J., and C. Manoil.** 1990. *lac* permease of Escherichia coli: topology and sequence elements promoting membrane insertion. Proc. Natl. Acad. Sci. USA **87:**4937–4941.
7. **Dassa, E., and M. Hofnung.** 1985. Sequence of gene malG in E. coli K12: homologies between integral membrane components from binding protein-dependent transport systems. EMBO J. **4:**2287–2293.
8. **Davidson, A. L., and H. Nikaido.** 1991. Purification and characterization of the membrane-associated components of the maltose transport system of Escherichia coli. J. Biol. Chem. **266:**8946–8951.
9. **Degli Esposti, M., M. Crimi, and G. Venturoli.** 1990. A critical evaluation of the hydropathy profile of membrane proteins. Eur. J. Biochem. **190:**207–219.
10. **Ehrmann, M., D. Boyd, and J. Beckwith.** 1990. Genetic analysis of membrane protein topology by a sandwich gene fusion approach. Proc. Natl. Acad. Sci. USA **87:**7574–7578.
11. **Fasman, G. D., and W. A. Gilbert.** 1990. The prediction of transmembrane protein sequences and their conformation: an evaluation. Trends Biochem. Sci. **15:**89–92.
12. **Higgins, C. F., S. C. Hyde, M. M. Mimmack, U. Gileadi, D. R. Gill, and M. P. Gallagher.** 1990. Binding protein dependent transport systems. J. Bioenerg. Biomembr. **22:**571–593.
13. **Kane, S. E., I. Pastan, and M. M. Gottesman.** 1990. Genetic basis of multidrug resistance of tumor cells. J. Bioenerg. Biomembr. **22:**593–617.
14. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
15. **Manoil, C.** 1991. Analysis of membrane protein topology using alkaline phosphatase and β-galactosidase gene fusions. Methods Cell Biol. **34:**61–75.
16. **Manoil, C., and J. Beckwith.** 1986. A genetic approach to analyzing membrane protein topology. Science **233:**1403–1408.
17. **Miller, V. L., R. K. Taylor, and J. J. Mekalanos.** 1987. Cholera toxin transcriptional activator toxR is a transmembrane DNA binding protein. Cell **48:**271–279.
18. **San Millan, J. L., D. Boyd, R. Dalbey, W. Wickner, and J. Beckwith.** 1989. Use of *phoA* fusions to study the topology of the *Escherichia coli* inner membrane protein leader peptidase. J. Bacteriol. **171:**5536–5541.
19. **Sugiyama, J. E., S. Mahmoodian, and G. Jacobson.** 1991. Membrane topology analysis of Escherichia coli mannitol permease by using a nested-deletion method to create mtlA-phoA fusions. Proc. Natl. Acad. Sci. USA **88:**9603–9607.
20. **Traxler, B., D. Boyd, and J. Beckwith.** Membrane protein topology. J. Membr. Biol., in press.
21. **Traxler, B., C. Lee, D. Boyd, and J. Beckwith.** 1992. The dynamics of assembly of a cytoplasmic membrane protein in Escherichia coli. J. Biol. Chem. **267:**5339–5345.
22. **von Heijne, G.** 1992. Membrane protein structure prediction, hydrophobicity analysis and the positive inside rule. J. Mol. Biol. **225:**487–494.
23. **Wang, R., S. J. Seror, M. Blight, J. M. Pratt, J. K. Broome-Smith, and I. B. Holland.** 1991. Analysis of the membrane organization of an Escherichia coli protein translocator, HlyB, a member of a large family of prokaryote and eukaryote surface transport proteins. J. Mol. Biol. **217:**441–454.