J. Membrane Biol. 132, 1–11 (1993)

The Journal of
Membrane Biology

Topical Review

The Topological Analysis of Integral Cytoplasmic Membrane Proteins

Beth Traxler*, Dana Boyd, and Jon Beckwith
Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115

Summary

We review three general approaches to determining the topology of integral cytoplasmic membrane proteins. (i) Inspection of the amino acid sequence and use of algorithms to predict membrane spanning segments allows the construction of topological models. For many proteins, the mere identification of such segments and an analysis of the distribution of basic amino acids in hydrophilic domains leads to correct structure predictions. For others, additional factors must come into play in determining topology. (ii) Gene fusion analysis of membrane proteins, in many cases, leads to complete topological models. Such analyses have been carried out in both bacteria and in the yeast *Saccharomyces cerevisiae*. Conflicts between results from gene fusion analysis and other approaches can be used to explore details of the process of membrane protein assembly. For instance, anomalies in gene fusion studies contributed evidence for the important role of basic amino acids in determining topology. (iii) Biochemical probes and the site of natural biochemical modifications of membrane proteins give information on their topology. Chemical modifiers, proteases and antibodies made to different domains of a membrane protein can identify which segments of the protein are in the cytoplasm and which are on the extracytoplasmic side of the membrane. Sites of such modifications as glycosylation and phosphorylation help to specify the location of particular hydrophilic domains. The advantages and limitations of these methods are discussed.

Key Words gene fusion · topology · membrane protein · membrane spanning segment · protein structure

* *Present address:* Department of Microbiology, University of Washington, Seattle, Washington 98195.

The Topology of a Membrane Protein Provides Important Structural Information

Integral proteins of the bacterial inner membrane and eukaryotic plasma membrane are defined as those that span the membrane one or more times. Topological models for such proteins are presented as two-dimensional diagrams, specifying which residues lie within the lipid bilayer and which residues constitute extramembrane loops or domains. These models usually also present an orientation for the protein such that the location within the cell (e.g., cytosolic or extramembraneous) of each hydrophilic domain is specified. Thus, these descriptions contain a great deal of useful information about the arrangement of the protein in the membrane, even though they do not represent a full three-dimensional structure. Since proteins of this class are generally difficult to crystallize, the construction of accurate topological models by means other than X-ray crystallography can provide information that is useful in the design of experiments to understand membrane protein function.

The nature of the membrane itself simplifies determination of membrane protein topology. The membrane represents a barrier which is impermeable to hydrophilic molecules, in general. Consequently, the parts of a membrane protein that lie on opposite sides of a membrane will be differentially accessible to various agents. This property of membrane proteins has allowed biochemical and immunological approaches to experimental determination of topology. In biochemical experiments, covalent modification of the protein by agents that have access to only one side of the membrane makes possible analysis of topology. Similarly, binding of antibodies to one or the other face of the membrane gives topological information when the immunological sites in a protein are known.

The special properties of membranes also facili-

tate the genetic analysis of membrane protein topology. In vivo, membranes separate compartments that represent different intracellular environments or separate the interior of the cell from the external environment. This property of membranes has permitted the use of gene fusion experiments to generate topological information. In such experiments, reporter genes are employed whose products confer different cellular phenotypes depending on the compartment to which they are localized.

Finally, since the membrane and the aqueous compartments on either side of it constitute such different environments, the parts of a membrane protein that can be stably assembled in one or the other milieu have strikingly different features. For instance, regions of the protein that lie within the membrane can often be identified by their hydrophobic nature. This identification, of course, is essential to the prediction of membrane protein topology.

Theoretical Approaches Can Be Used to Predict Membrane Protein Topology

Inspection of Amino Acid Sequence Can Provide Topological Models

It has become apparent that the information that determines membrane protein topology is distributed throughout the amino acid sequence. There are topological determinants which act locally to determine the membrane insertion and orientation of the segments in which they lie as originally proposed by Blobel [6]. Specifically, each membrane spanning stretch, together with its flanking hydrophilic domains, constitutes a topological determinant that can insert independently into the membrane [9, 10, 13, 51, 66]. By recognizing these topological determinants within membrane protein sequences, we can predict topology.

A first step in constructing topological models of a membrane protein is a definition of which parts of the protein span the membrane and which are extramembraneous. Membrane spanning segments in integral membrane proteins of known structure generally consist of regions of high average hydrophobicity that are long enough (about 20 residues) to span the membrane in α-helical conformation. Often, the existence of such striking hydrophobic stretches is enough to permit formulation of models of membrane protein topology by simple inspection of its sequence. A more formal approach [42] is to use a set of values for the relative hydrophobicity of each amino acyl residue and calculate a running average hydrophobicity over a window of appropriate length. A plot of the average hydrophobicity flanking each residue *vs.* the position in the sequence yields a curve with peaks corresponding to the membrane spanning segments for many membrane proteins. Recently, these methods have been critically reviewed [28]. A similar approach uses the frequency of residues in known membrane spanning segments instead of hydrophobicity and may give slightly better results [17, 61].

Construction of topological models using such plots is often straightforward. Hydrophobicity or probability peaks of the appropriate length correspond to putative membrane spanning segments. Alternating hydrophilic segments are thus placed on opposite sides of the membrane. Several algorithms that attempt to set objective criteria for identification of membrane spanning stretches have been published [17, 27, 40]. In some cases, however, a unique model is not suggested by such analysis, but two or more alternative models may seem plausible. For instance, the existence of unusually long hydrophobic stretches or of putative membrane spanning stretches with charged amino acyl residues within them may lead to models for the same protein that contain different numbers of spanning segments. Distinguishing between such models is sometimes facilitated by consideration of the role of charged residues in determining membrane protein orientation, as discussed below. Hydrophilic domains predicted to be on the same side of the membrane must, of course, be connected by an even number of membrane spanning segments. Discrimination between plausible alternate models may require experiments using the methods outlined in subsequent sections. In such cases topological models derived from sequence analysis are valuable aids in design of experiments.

The Distribution of Basic Amino Acids Allows Prediction of the Orientation of Membrane Proteins

The orientation of membrane proteins is at least in part determined by the distribution of charged residues near the ends of membrane spanning segments (9, 19, 76). von Heijne (73, 75) has pointed out that short cytoplasmic loops of both eukaryotic and prokaryotic membrane proteins have more basic residues than would be expected at random. He suggested that these residues might help to determine the orientation of the membrane spanning segments by directing their own cytoplasmic localization.

Experiments in model bacterial systems using fusion proteins [8], and proteolysis (3, 54, 74) have begun to make possible predictive rules for orienta-

tion. The results suggest that charged basic residues within about thirty residues from the carboxy-terminal end of a membrane spanning segment are primary determinants of cytoplasmic localization. Arginyl and lysyl residues have a much stronger effect than histidinyl residues. It may be that acidic residues can partly neutralize the effects of basic residues when positioned so that they are adjacent in α-helical conformation. Acidic residues also have their own, much weaker effect as cytoplasmic localization determinants [3, 43].

Several reasons why positively charged residues might determine cytoplasmic localization have been proposed. First, burying arginyl or lysyl residues in the membrane during export would be energetically expensive, as their pKa's are farther from neutrality than those of acidic residues. This explanation is consistent with the finding that a group of several acidic residues has an effect similar to a single basic residue in determining localization [3, 43] and that histidinyl residues have an effect only at low pH [3]. On the other hand, the effect of arginyl residues is not always greater than that of lysyl ones, as would be expected from the difference in pka [3]. Second, the membrane potential in bacteria would energetically favor cytoplasmic localization of basic residues. Finally, interaction of basic residues with negatively charged lipid head groups might stabilize cytoplasmic localization by hindering further translocation across the membrane. Recently it has been suggested that the α carbon of arginyl and lysyl residues may, in fact, lie within the membrane in some cases while the basic moiety interacts with the lipid head groups at the surface of the membrane [4]. There is not sufficient evidence at this time to indicate which explanation(s), if any, is the correct one.

Other Features of Membrane Proteins May Be Important in Determining Topology

It seems likely that factors other than the charge distribution in hydrophilic domains are important in determining proper orientation in the membrane. Multiple determinants would allow for greater stability of the final structure. For instance, it is possible that the cytoplasmic domains of membrane proteins have evolved so as to fold rapidly into their native structures in the cytoplasm. At least for secreted proteins, stably folded proteins or domains are usually unexportable (44, 45, 54, 59). If rapid, stable folding of a membrane protein domain can also interfere with its export, then such a domain would remain cytoplasmic and contribute to determining the topology of the protein.

In addition, interactions between hydrophilic domains or between membrane spanning segments are known to be important in the assembly of certain membrane proteins [36, 58, 72, 80]. Such interactions may also contribute to the stability of the final topological structure [57]. At present, it is not possible to predict the existence of rapidly folding domains nor, except in a few cases, can we identify those domains of membrane proteins that interact.

Despite the absence of full information on the nature of topological determinants, predictions from inspection of amino acid sequence together with theoretical considerations have proved remarkably successful in many cases. That is, evaluation of potential membrane spanning segments and of the distribution of charged amino acyl residues in hydrophilic domains has provided models that have been confirmed by other approaches. While it may be that more than one factor is involved in determining the topology of particular regions of a membrane protein, charge distribution and hydrophobicity may be the predominant factors for many proteins. Exceptions to these rules are discussed below.

In contrast to cytoplasmic domains, exported domains of membrane proteins do not seem to have information that determines which side of the membrane they lie on. The topology of the *Escherichia coli* membrane proteins MalF [51] and leader peptidase [54] have been altered so that a normally exported hydrophilic domain is retained in the cytoplasm. This can be achieved by in vitro constructions that remove an odd number of hydrophobic membrane spanning segments preceding the hydrophilic domain under study. The deletion of these segments results in the inversion in orientation of the following portions of the protein. For the normally exported domain, this occurs because it is now separated from a strong cytoplasmic localization signal by an even number of membrane spanning segments.

Finally, manipulations of the MalF system have also suggested that the rules for the effects of basic amino acyl residues on topology are not completely worked out. When the *malF* gene was altered so that two cytoplasmic localization signals that ordinarily were separated by an even number of membrane spanning segments were now separated by an odd number, one cytoplasmic domain remained cytoplasmic and the other was exported [51]. Presumably the stronger of the two signals, the C-terminal one in this case, dominates in this artificial situation. Thus, the assembly in the membrane of this altered protein is not simply following a sequence of events initiated by topological signals early in the protein. Rather, a strong downstream cytoplasmic localization signal can effectively compete with other signals that occur amino-terminal to it. Curiously, these two

"competing" hydrophilic regions both have three basic residues and one acidic residue. The reason for the difference in their strengths as cytoplasmic localization signals may be related to the potential to form stable ion pairs between charged amino acyl residues, thus making the domain more readily exportable. A variety of alternative explanations exist such as differences in the kinetics of folding of the domains.

In summary, we can recognize two classes of topological determinants in membrane protein sequences: hydrophobic segments about 20 to 30 residues long and positively charged hydrophilic segments of about the same length or shorter near the ends of the hydrophobic segments. The hydrophobic sequences usually span the membrane and the basic regions tend to be cytoplasmic. This information permits construction of one or more possible topological models for a membrane protein to aid in the design of experiments to determine topology.

Ultimately, it may be possible to accurately predict from the amino acid sequence the three-dimensional structure of membrane proteins. An example of the way in which topological information can be used in conjunction with biochemical and biophysical data and computer methods to predict a structure is provided by a recent study of the β_2-adrenergic receptor [46].

Gene Fusion Experiments Are Widely Used to Formulate Topological Models

The assumption behind gene fusion analysis of topology is that a reporter protein fused to a hydrophilic domain of a membrane protein can give an indication of the cellular location of that domain. Only reporter proteins with certain features can be used in such studies. That is, the properties of cells in which the reporter protein is localized to the cytoplasm, or of the protein itself, must be distinguishable from those in which the protein has been translocated across a membrane. In bacteria, this membrane would be the cytoplasmic membrane; in eukaryotic cells, it would be the membrane of the endoplasmic reticulum or a subsequent cellular compartment in a protein localization process.

Several Gene Fusion Systems Have Been Used in Prokaryotes

The first protein to be used for such studies was the alkaline phosphatase coded for by the *E. coli phoA* gene [49]. The cellular location of this protein can be easily distinguished since it is enzymatically inactive in the cytoplasm and enzymatically active when translocated across the cytoplasmic membrane [52]. The basis for this difference in activity is the failure to form essential disulfide bonds in the cytoplasm and their formation after translocation into the periplasmic space [21]. Thus, in general, fusion of alkaline phosphatase to periplasmic domains of a membrane protein exhibit enzymatic activity while fusions to cytoplasmic domains exhibit only very low activity. The amount of export of alkaline phosphatase can be precisely determined by the use of a sensitive enzyme assay. Both transposon and plasmid vectors exist for the construction of alkaline phosphatase fusions (24, 26, 35, 48, 78). In addition, beginning with a fusion late in a particular gene or beyond that gene, a set of fusions with specified endpoints can be constructed using oligonucleotide mutagenesis to generate specific deletions [10]. The same or similar starting strain can also be used to create a set of nested fusion endpoints using exonuclease digestion and appropriate restriction enzymes [69, 81]. Polymerase Chain Reaction methods can also be used to create fusions (Boyd, Traxler and Beckwith, *J. Bacteriol., in press*) and have the advantage that the two segments of DNA to be joined need not be in the same parent molecule.

The *E. coli* β-galactosidase has properties that allow it to be used in a fashion complementary to that used with alkaline phosphatase fusions. β-galactosidase is a cytoplasmic enzyme. When it is fused to a protein export signal, it becomes embedded in the membrane during the translocation process, rendering it enzymatically inactive. As a result, fusions of β-galactosidase to a cytoplasmic domain of a membrane protein are active, while those to a periplasmic domain are inactive [30]. In addition to plasmid and transposon systems for constructing fusions, transposons have been described [47, 78] that allow the conversion of an alkaline phosphatase fusion into a β-galactosidase fusion. The two classes of fusions obtained in this way can provide strengthened evidence for particular topological models.

The TEM β-lactamase, encoded by the *bla* gene, can act as a reporter of subcellular location by its property of conferring resistance to antibiotics such as ampicillin. This resistance is only expressed when the enzyme is translocated across the cytoplasmic membrane, since the site of action of ampicillin is on the enzymatic formation of the cell wall. Only fusions of β-lactamase to periplasmic domains of a membrane protein will give ampicillin-resistance [11]. The degree of export can be measured by the degree of antibiotic resistance. As with *phoA*, both a transposon, Tn*blaM* [70] and plasmid vectors [11] have been constructed for gene fusion purposes. The nested fusion approach (*see above*) has also been

used for the isolation of a collection of β-lactamase fusions [7].

A fourth fusion approach to membrane protein topology involves the use of a vector coding for a fragment of the enzyme acetylCoA-carboxylase (Jander and Beckwith, *unpublished results*). This enzyme is the sole protein of *E. coli* that is biotinated. The gene fragment used for this approach encodes a 70 amino acid polypeptide from this protein that includes the site of biotination [18]. Fusion of this fragment to the carboxy-terminus of other proteins results in fusion proteins that are now themselves biotinated. However, since the biotination takes place in the cytoplasm, labeling with biotin might be expected to be restricted to those fusion proteins that are retained in the cytoplasm. This is not entirely true, since fusion proteins that are exported slowly retain that domain in the cytoplasm long enough for biotination to take place. The biotinated portion of the protein is subsequently transported across the membrane. This was found to be the case for secreted proteins that are slowly exported in a *sec* mutant background [62]. Fusions to one periplasmic domain of a membrane protein (MalF) are not biotinated, while those to a cytoplasmic domain are. Yet, a fusion to another periplasmic domain was biotinated, indicating that there is a difference in the rate of export of different periplasmic domains of this protein. Thus, this fusion approach may provide a sensitive test for the kinetics of export of hydrophilic domains of a membrane protein.

Gene Fusions Have Also Been Used in Eukaryotes to Study Membrane Protein Topology

One of the first gene fusion studies to analyze membrane protein topology in yeast was done with a vector encoding the enzyme histidinol dehydrogenase [65]. This enzyme carries out the last step in histidine biosynthesis. When the protein is fused to an export signal, it is translocated into the rough endoplasmic reticulum, where it is no longer accessible to its substrate, histidinol. Thus, fusions of histidinol dehydrogenase to a cytoplasmic protein will complement a HIS4 deletion for growth on histidinol, while those to exported proteins will not. The ability of membrane protein fusions to this enzyme to complement the mutation give an indication of whether the fusion junction is in a cytoplasmic domain or a periplasmic domain.

Glycosylation of secreted proteins has provided the basis for a different gene fusion approach to analyzing topology. When the enzymes galactokinase [33], acid phosphatase [1] or invertase [22] are fused to protein export signals, they are glycosylated, whereas when they are retained in the cytoplasm, the modification does not take place. This difference furnishes a means of distinguishing cytoplasmic and exported domains of membrane proteins. Since glycosylated proteins are easily separated by their molecular weights and further characterized by the use of tunicamycin or Endoglycosidase H, the modified proteins are readily characterized. Treatment of cells with tunicamycin or of extracts with Endoglycosidase H yields proteins missing their normal glycosylation and exhibiting faster mobility on gel electrophoresis.

Gene Fusion Analysis Has Been Successful in Predicting Topology for a Number of Proteins

There are certainly reasons to expect that attempts to determine membrane protein topology using gene fusions might give misleading answers. First, for the most part, the construction of the fusions leads to hybrid proteins in which the reporter protein replaces a carboxy-terminal portion of the membrane protein. If the final topology of a protein is dependent on the interaction of both amino- and carboxy-terminal sequences, the fusion approach could yield a false conclusion. Second, the reporter proteins themselves may have properties that interfere with the topological analysis. For instance, if mere proximity of alkaline phosphatase to the cytoplasmic membrane leads to export of the protein, fusions to some cytoplasmic domains of membrane proteins might exhibit alkaline phosphatase activity. Finally, the enzymatic activity of the reporter proteins might vary according to the amino acid sequences that immediately precede them, resulting in some false negatives.

While each of these concerns is valid, the experience with those proteins of known topology suggests that they are not major problems for the approach. That is, in a number of cases where other techniques have been used to study topology, the fusion results have conformed with the other approaches. In certain of those cases where they do not conform, resolution of the conflicts has led to new information about membrane protein assembly.

Fusion analysis of the proteins Tsr [49], leader peptidase [63] and penicillin-binding protein 3 [7] gave topological structures that were the same as those obtained by studies with proteolysis of the proteins or studies of the relationship between structure and function. In the case of MotB [16], the fusion analysis led to a revision of the topological model, a revision that was confirmed by proteolysis studies. For several complex membrane proteins,

portions of the topological structure determined by structural studies or by proteolysis were the same as those obtained by study of gene fusions [12, 81].

Anomalies in Gene Fusion Analysis Lead to New Insights into Membrane Protein Assembly

In certain cases, the absence of carboxy-terminal sequences of a membrane protein in a gene fusion strain *can* lead to anomalous findings on topology. The first well-established case involved fusions of alkaline phosphatase to different sites within cytoplasmic domains of membrane proteins [8]. The anomalies result from the fact that basic amino acyl residues found in cytoplasmic domains play a role as topological determinants. When alkaline phosphatase is fused to a position in the cytoplasmic domain that precedes the basic amino acyl residues, it is less stably localized to the cytoplasm than if it follows these positively charged amino acyl residues. Then, the former class of fusions would exhibit much higher alkaline phosphatase activities than expected.

In fact, the finding of the apparently anomalous behavior of certain cytoplasmic fusions of alkaline phosphatase led to mutational evidence for the role of basic amino acids as topological determinants. Once these results were understood, it became possible to choose the sites of fusion joints so as to optimize the topological analysis. That is, fusions to proposed cytoplasmic domains are constructed so that the joint is at a position following basic amino acyl residues, the fusions always show low alkaline phosphatase activity (10, 63, Boyd, Traxler and Beckwith, *J. Bacteriol., in press,* and *see cover* of [5]). Choosing the sites of fusion joints in this way gives a clear-cut distinction between cytoplasmic and periplasmic domains of membrane proteins in the cases examined.

Another anomaly was found in the analysis of the galactoside permease of *E. coli,* product of the *lacY* gene [13]. In this case, the periplasmic localization of a particular hydrophilic domain had been established by chemical modification techniques [56]. Yet, an alkaline phosphatase fusion to this domain showed much lower activity than was expected. The low activity was attributed to the fact that the hydrophobic membrane spanning segment that preceded the periplasmic domain contained an arginine that rendered that segment a poor export signal on its own. When the arginine was replaced with an alanine, the high expected levels of alkaline phosphatase export were observed. The question then arises, if the membrane spanning segment is a poor export signal for alkaline phosphatase, how can it act as an export signal for the hydrophilic domain that follows it. One possible explanation is that this export signal does not act on its own, but must interact with carboxy-terminal sequences of the protein (e.g., the next membrane spanning segment) to function properly in membrane assembly. Further genetic studies on this system should contribute to an understanding of how this particular segment of the protein achieves its topology.

The existence of membrane spanning segments that, according to the topology of a protein, should act as export signals, but when isolated from carboxy-terminal sequences do not so act, raises problems for the fusion approach in the analysis of certain proteins. Many of the proteins studied so far have membrane spanning segments that are devoid of charged amino acyl residues and are highly hydrophobic. Nevertheless, some membrane proteins, perhaps because of their function, have membrane spanning segments that are more hydrophilic. It then becomes a challenge to determine whether the fusion approach can be adapted to correctly analyze these cases. One possible solution is the use of sandwich fusions, in which the reporter protein is *inserted* into the membrane protein rather than replacing the carboxy-terminus. In such sandwich fusions, the presence of the complete sequence of the membrane protein may confer proper topology on the hybrid protein. A sandwich fusion vector for alkaline phosphatase has been constructed and tested on the MalF protein and does eliminate the anomalies found with fusions to cytoplasmic domains [26]. However, it is not clear that it will solve all of the other problems described here.

In the case of the ProW protein, involved in proline transport in *E. coli,* alkaline phosphatase fusions at the amino-terminal end of the protein gave results that conflicted with the known topology of the protein (E. Bremer, *personal communication*). The ProW protein, in contrast to most of those studied so far, has a substantial amino-terminal domain that precedes the first membrane spanning segment and is localized to the periplasm. It seems likely that the localization of this domain to the periplasm is dependent on downstream sequences of the protein. Alkaline phosphatase fusions early in the protein would be missing these downstream sequences and, therefore, the reporter protein would not be translocated across the membrane. β-galactosidase fusions also gave anomalous results. Again, it is possible that the sandwich fusion approach will overcome some of the problems associated with topological analysis of proteins such as ProW. Finally, while β-galactosidase fusions yielded the likely correct topology for LacY, there were anomalies in a set of fusion to the cytochrome d terminal oxidase [32].

In each of these cases, unusual features of the protein being analyzed and its assembly in the membrane may be responsible for anomalous results. Further exploration of these anomalies is likely to lead to new insights into the steps by which topological structures are generated.

It is also possible that some of the problems will be seen with one fusion approach and not another. There has not been enough systematic analysis of the different approaches to determine whether this might be the case.

Rules for Gene Fusion Analysis of Topology Are Evolving

The considerations discussed above lead to suggestions for optimizing the utility of gene fusion analysis of topology. First, if there are any apparent complexities in the results, a set of fusions should be constructed with defined junctions. For instance, in the case of the *E. coli* MalG protein, we constructed alkaline phosphatase fusions where the reporter protein was fused to each of the proposed periplasmic domains and to *carboxy-terminal* ends of each proposed cytoplasmic domain (Boyd, Traxler and Beckwith, *J. Bacteriol., in press*). Choosing the fusion junctions in this way eliminated the anomalous results found with fusions in which the junction precedes the basic amino acyl residues needed for stable cytoplasmic anchoring. Second, the enzymatic activity or phenotypic properties should be quantitatively related to the rate of synthesis of the hybrid proteins encoded by gene fusions. The measurements of rates of synthesis and alkaline phosphatase activities have been important in evaluating results with that system [63].

Finally, gene fusion analysis can yield a complete model for the topology of certain membrane proteins, a feat that, to this point, can only be duplicated by structural approaches such as X-ray crystallography, or, in some cases, by use of antibodies (*see below*). The approach has proved, so far, to be practically quite useful in analyzing this aspect of protein structure. The definition of large periplasmic domains of proteins such as MalF [10, 30], SecD and SecF [31], has been an important step in studies of the function and assembly of these proteins. Nevertheless, it is important to consider that the results only allow the formulation of models for topology that should be tested further by other means.

The Topology of Some Eukaryotic Proteins Can Be Studied in Bacteria

Over the years, a striking similarity between the rules for protein secretion and membrane protein assembly in eukaryotes and prokaryotes has been shown. The genes for a number of eukaryotic membrane proteins have been cloned into and expressed in *E. coli*. In some cases, the proteins exhibit the same activity in the bacteria as they do in the eukaryotic cell, indicating that at least a fraction of the protein has been inserted in its correct conformation in the membrane. The human red cell glucose transporter can transport glucose in *E. coli* and the β_2-adrenergic receptor can bind ligand when incorporated into the bacterial cytoplasmic membrane [29, 64]. The only published study on the topology of a eukaryotic membrane protein in bacteria involved the β-lactamase fusion analysis of the β-subunit of sheep-kidney Na, K-ATPase [82]. This analysis indicated that this simple membrane protein (it contains only one membrane spanning segment) had assumed the same conformation in *E. coli* as it did in its original mammalian cell.

This finding raises the possibility that topological analysis of eukaryotic proteins can be done in bacteria. Determining whether this is the case depends on the study of many more examples. Even those cases where proper topology is not generated in bacteria may be useful for understanding any differences that might exist in the mechanism of membrane protein assembly between eukaryotes and prokaryotes.

Another protein analyzed is the human β_2-adrenergic receptor (Lacatena and Tocchini-Valentini, *personal communication*). Here, alkaline phosphatase fusions were used to determine the topology when the protein was expressed in *E. coli*. The data, for the most part, fit the topological model proposed for the protein. However, amino-terminal fusions presented anomalies analogous to those seen with the ProW protein. As with ProW, the amino-terminus of the β_2-adrenergic receptor is exported. As discussed above in the case of ProW, proper assembly of the amino-terminal portions of membrane proteins of this type may require interactions with downstream sequences. A further complication with this protein is that the amount of protein made in *E. coli* far exceeds that expected from the amount of ligand binding seen. Whether this discrepancy represents a mislocalization of the protein, degradation, or simple failure to function effectively in the bacterial membrane, is not clear.

Biochemical Probes Are Important in the Determination of Protein Topology

Preceding sections have summarized strategies for the prediction of topology and molecular genetic approaches for the analysis of topology. However,

the models suggested by the examination of amino acid sequences and of fusion proteins still are considered preliminary in the absence of supporting biochemical data.

Some of the most convincing biochemical topological data are provided by normal protein modifications. Protein phosphorylation and N-linked glycosylation both result from enzymatic reactions carried out on only one side of the membrane. The addition of phosphate groups to a membrane protein (often at His, Tyr, Thr, or Ser residues) occurs in domains of the protein retained on the cytoplasmic side of the membrane [25, 37, 68]. Oligosaccharides are added to the amino group of Asn residues within the consensus sequence Asn-X-Thr/Ser only on the luminal side of the ER for eukaryotic membrane proteins [41]. Following the exit of such proteins from the ER, the modification is maintained on the extracytoplasmic face of the membrane, regardless of the protein's final cellular location. Identification of the location of these modifications in a membrane protein of interest provides strong topological information for particular regions of the polypeptide. The creation of new consensus N-linked glycosylation sites within a protein's primary sequence has recently been used as a probe for protein topology [15]. Site-directed mutagenesis can be used to engineer the amino acid changes in the protein necessary to specify the consensus sequence where it previously did not exist. The presence or absence of the sugar modification indicates the localization of the region containing this sequence. These changes are made in regions of the protein of particular interest and have been used to determine topology between different possible models.

In addition, there are a wide variety of biochemical techniques for the study of protein topology, which are discussed in part below. We have grouped different methodologies into three categories: protein accessibility to chemical modifiers, to antibodies, and to proteases.

For the examination of a protein's topology using biochemical techniques, the uniformity and integrity of the assayed membrane preparation is critical. Most methodologies rely on the exposure of particular domains of proteins in preparations in which one face of the membrane is uniquely accessible. The labeling or modification of the protein at a particular site within the protein's primary sequence can then be correlated with the side of the membrane which was accessible during treatment for the determination of topology. Optimally, one would compare the reactivity of the protein of interest in membrane vesicles of both possible orientations. Many procedures have been described for the isolation of such preparations, e.g., for the cytoplasmic membrane of *E. coli* [34, 39]; the human erythrocyte plasma membrane [67]; and others, reviewed in [38]. In the absence of available preparations in both orientations, one can compare the accessibility of the protein to biochemical manipulations in permeabilized *vs.* nonpermeabilized membranes as was done for the erythrocyte glucose transporter [53], and the β_2-adrenergic receptor [77].

Implicit in this discussion is the assumption that the protein of interest is present in the membrane in one topological arrangement. Under normal physiological conditions, this is almost certainly correct. However, under abnormal conditions such as those often associated with protein overproduction, there may be a heterogenous population of protein in or associated with the membrane. Experiments assaying the structure of overproduced proteins should be pursued only after the demonstration of the similarity of proteins expressed at physiological and elevated levels.

Topology Can Be Examined with Chemical Modifiers

There have been a large number of chemical agents developed for probing membrane protein structure. While no effort will be made to discuss these in detail, we will mention different categories into which these compounds fall. A more extensive survey of these agents can be found in Jennings [38]. For a specific example of the use of a wide variety of different compounds, *see* [56].

First, many chemical procedures for studying topology are based on the reactivity of surface regions of the protein to hydrophilic labeling agents. In implementing these experiments, it is important to consider the reactivity of the labeling agent for the exposed amino acids in the protein of interest as many have considerable specificity (e.g., sulfhydryl reagents such as glutathionemaleimide reacting with Cys and the iodination of Tyr by lactoperoxidase).

With the use of biochemical techniques to identify surface-exposed residues, one must be aware that at least some channel-forming proteins have residues lying within the plane of the membrane which may be reactive to these labeling techniques [60]. Such residues are thought to lie in a solvent accessible pocket. While these data may be useful for identifying important structural or functional features of the protein, their topological indications can be misleading.

Recently, the limited reactivity of many chemical probes has been coupled with the use of site-specific mutagenesis to study protein topology. Specific changes in the protein's primary sequence,

which do not disrupt protein function, can be made such that a reactive site for a particular probe is added. For example, several different amino acyl residues were changed, one at a time, to Cys at specific locations in bacteriorhodopsin, which normally does not contain any Cys residues [2]. The engineered Cys mutant proteins were then tested for reactivity with a specific spin labeling compound and the results were interpreted in terms of which positions within the bacteriorhodopsin sequence were solvent (surface)-exposed and which were membrane-embedded.

Topological information is also available through the use of agents that label hydrophobic regions of membrane proteins to identify membrane bound residues. These compounds are commonly photo-activatable and are available for the labeling of a variety of amino acyl residues [38, 56]. One caution to the use of these agents is that they can label not only membrane-embedded regions of the protein but also hydrophobic sites buried in the interior of soluble domains of the protein. A refinement to this approach is to use a hydrophobic photolabel coupled to either the fatty acid or the head group of phospholipids. These compounds can be introduced into the assayed preparation and activated for labeling residues with which they have contact.

Proteolytic Sensitivity of Proteins is Another Probe for Topology

One of the most extensively used techniques for the demonstration of topology is the in situ proteolysis of the protein of interest with degradative enzymes. These experiments are assumed to demonstrate the surface exposure of sensitive regions of the protein. In practice, it is sometimes unclear what the proteolytic sensitivity or resistance of a particular region means. Often exposed regions of a membrane protein are quite protease resistant, and membrane-embedded domains can be degraded during proteolytic treatment of a membrane preparation. A large number of different proteases are used in these studies, and generally, endopeptidases with high cleavage site specificity (e.g., trypsin) are the most useful. After proteolysis, it is usually necessary to purify the peptides and sequence the point of cleavage. It has also been useful to study topology with exopeptidases such as carboxypeptidases to demonstrate surface accessibility of the C-terminal end of the protein on the exposed face of the membrane (e.g., β_2-adrenergic receptor [23]).

It is often necessary to use high concentrations of proteolytic enzymes to cleave membrane proteins compared to concentrations used in proteolysis of soluble proteins. This is presumably because the proteases have limited access to hydrophilic regions of the protein and, in some cases, a small target size. Recent work has shown that the proteolytic sensitivity of MalF (part of the hetero-oligometric maltose transport complex in *E. coli*) changes as a result of its assembly into the complex [71]. The protein is initially incorporated into the membrane in a protease-sensitive form. Upon assembly into the complex, MalF becomes largely protease resistant. In the absence of the other subunits of the complex, MalF never becomes protease resistant. This difference apparently is not due to changes in the protein's topology but rather to changes in the protein's accessibility to degradation. These results suggest that a useful approach for studying some proteins by proteolysis may be to examine them in the presence and absence of interacting subunits.

Antibodies Can Be Used in Several Ways for Topology Studies

The biochemical techniques for studying membrane proteins are dependent upon having highly enriched membrane fractions containing the protein of interest or a way to selectively purify or detect the protein of interest. Antibodies are a convenient tool for retrieving the reacted protein after treatment.

In addition, antibodies can be used directly in topological studies by determining to which side of a membrane preparation a particular antibody will bind [e.g., 14, 55, 77]. As with other methods, the use of antibodies for studying membrane protein topology relies on the reactivity of specific antibodies with surface-exposed regions of the protein of interest. Antibody binding to a domain of the protein is detected by techniques such as ELISA analysis, radioimmunoassay, or by immunomicroscopy. Either monoclonal or polyclonal antipeptide antibodies can be used for these studies. Monoclonal antibodies against the protein are prepared and characterized for the recognition of small contiguous domains within the protein. "Site-directed" antipeptide antibodies are elicited in response to synthetic peptides corresponding to regions within the protein's primary sequence which are predicted from hydropathy analysis to be solvent exposed.

With a panel of antibodies specific to different domains of the protein, one can identify all solvent-accessible regions of the protein as was done for the β_2-adrenergic receptor and α_1 connexin [77, 79]. However, many antibodies, which recognize a protein in a denatured form, do not bind to the intact protein within the membrane as was observed with both LacY and the erythrocyte glucose transporter,

and, therefore, are not helpful in topology studies [14, 20]. While this problem can limit the utility of the approach, at least some topological information is frequently provided by antibody binding studies.

Conclusion

In this review, we have discussed various methods for the analysis of membrane protein topology, ranging from theoretical to experimental (genetic, immunological, and biochemical) approaches. Each fills a role in the task of defining a protein's arrangement in the membrane.

A noteworthy example where a variety of techniques have been utilized to study protein topology is the body of work on the acetylcholine receptor (AChR) summarized in part in [15, 50]. Each subunit of the AChR originally was proposed to have four membrane spanning segments, based on hydropathy analysis of primary sequences; later analyses suggested an alternative five membrane spanning segment model. Experiments to identify the protein's topology have produced mixed results. Several antibody binding studies were inconsistent with the four membrane spanning segment model and supported a topological structure quite unlike that suggested by any of the hydropathy analyses. However, subsequent biochemical and immunological studies have supported the four membrane spanning segment model, with both the amino- and carboxy-termini exported to the extracytoplasmic face of the membrane. While the topology of the protein has not been unambiguously solved, much data from a variety of labs using a number of different techniques support the four membrane spanning segment model.

The issues raised in the study of the AChR are common to studies of all membrane proteins. When are the models proposed by theoretical analyses considered proven or disproved? How does one discriminate between alternative models or conflicting results? As more membrane proteins are analyzed, the rules governing their arrangement in the membrane should become clear.

This work was supported by a fellowship from the National Institute of General Medical Sciences to B.T., by a grant from the National Science Foundation to D.B. and by a grant from the National Institutes of Health to J.B.. J.B. is an American Cancer Society Research Professor.

References

1. Ahmad, M., Bussey, H. 1988. *Mol. Microbiol.* **2:**627–635
2. Altenbach, C., Marti, T., Khorana, H.G., Hubbell, W.L. 1990. *Science* **248:**1088–1092
3. Andersson, H., Bakker, E., von Heijne, G. 1992. *J. Biol. Chem.* **267:**1491–1495
4. Ballesteros, J.A., Weinstein, H. 1992. *Biophys. J.* **62:**127–128
5. Beckwith, J., Silhavy, T.J. 1992. *The Power of Bacterial Genetics: A Literature-Based Course.* Cold Spring Harbor Laboratory Plainview
6. Blobel, G. 1980. *Proc. Natl. Acad. Sci. USA* **77:**1496–1500
7. Bowler, L.D., Spratt, B.G. 1989. *Mol. Microbiol.* **3:**1277–1286
8. Boyd, D., Beckwith, J. 1989. *Proc. Natl. Acad. Sci. USA* **86:**9446–9450
9. Boyd, D., Beckwith, J. 1990. *Cell* **62:**1031–1033
10. Boyd, D., Manoil, C., Beckwith, J. 1987. *Proc. Natl. Acad. Sci. USA* **84:**8525–8529
11. Broome-Smith, J.K., Tadayyon, M., Zhang, Y. 1990. *Mol. Microbiol.* **4:**1637–1644
12. Calamia, J., Manoil, C. 1990. *Proc. Natl. Acad. Sci. USA* **87:**4937–4941
13. Calamia, J., Manoil, C. 1992. *J. Mol. Biol.* **224:**539–543
14. Carrasco, N., Herzlinger, D., Danho, W., Kaback, H.R. 1986. *Meth. Enzymol.* **125:**453–467
15. Chavez, R.A., Hall, Z.W. 1991. *J. Biol. Chem.* **266:**15532–15538
16. Chun, S., Parkinson, J.S. 1988. *Science* **239:**276–278
17. Crimi, M., Degli Esposti, M. 1991. *TIBS* **16:**119–119
18. Cronan, J.E., Jr. 1990. *J. Biol. Chem.* **265:**10327–10333
19. Dalbey, R.E. 1990. *TIBS* **15:**253–257
20. Davies, A., Ciardelli, T.L., Lienhard, G.E., Boyle, J.M., Whetton, A.D., Baldwin, S.A. 1990. *Biochem. J.* **266:**799–808
21. Derman, A.I., Beckwith, J. 1991. *J. Bacteriol.* **173s:**7719–7722
22. Deshaies, R.J., Schekman, R. 1990. *Mol. Cell. Biol.,* **10:**6024–6035
23. Dohlman, H.G., Bouvier, M., Benovic, J.L., Caron, M.G., Lefkowitz, R.J. 1987. *J. Biol. Chem.* **262:**14282–14288
24. Duchêne, A.M., Patte, J., Gutierrez, C., Chandler, M. 1992. *Gene* **114:**103–107
25. Edelman, A.M., Blumenthal, D.K., Krebs, E.G. 1987. *Annu. Rev. Biochem.* **56:**567–613
26. Ehrmann, M., Boyd, D., Beckwith, J. 1990. *Proc. Natl. Acad. Sci. USA* **87:**7574–7578
27. Eisenberg, D. 1984. *Annu. Rev. Biochem.* **53:**595–623
28. Fasman, G.D., Gilbert, W.A. 1990. *TIBS* **15:**89–92
29. Freissmuth, M., Selzer, E., Marullo, S., Schültz, W., Strosberg, A.D. 1991. *Proc. Natl. Acad. Sci. USA* **88:**8548–8552
30. Froshauer, S., Green, G.N., Boyd, D., McGovern, K., Beckwith, J. 1988. *J. Mol. Biol.* **200:**501–511
31. Gardel, C., Johnson, K., Jacq, A., Beckwith, J. 1990. *EMBO J.* **9:**3209–3216
32. Georgiou, C.D., Dueweke, T.J., Gennis, R.B. 1988. *J. Biol. Chem.* **263:**13130-13137
33. Green, G.N., Hansen, W., Walter, P. 1989. *J. Cell Sci. Suppl.* **11:**109–113
34. Hertzberg, E.L., Hinkle, P.C. 1974. *Biochem. Biophys. Res. Comm.* **58:**178–184
35. Hoffman, C., Wright, A. 1985. *Proc. Natl. Acad. Sci. USA* **82:**5107–5111
36. Huang, K.-S., Bayley, H., Liao, M.-J., London, E., Khorana, H.G. 1981. *J. Biol. Chem.* **256:**3802–3809
37. Hunter, T., Cooper, J.A. 1985. *Annu. Rev. Biochem.* **54:**897–930
38. Jennings, M.L. 1989. *Annu. Rev. Biochem.* **58:**999–1027
39. Kaback, H.R. 1971. *Meth. Enzymol.* **22:**99–120

40. Klein, P., Kanehisa, M., DeLisi, C. 1985. *Biochim. Biophys. Acta* **815**:468–476
41. Kornfeld, R., Kornfeld, S. 1985. *Annu. Rev. Biochem.* **54**:631–664
42. Kyte, J., Doolittle, R.F., 1982. *J. Mol. Biol.* **157**:105–132
43. Laws, J.K., Dalbey, R.E. 1989. *EMBO J.* **8**:2095–2099
44. Lee, C., Li, P., Inouye, H., Beckwith, J. 1989. *J. Bacteriol.* **171**:4609–4616
45. Maher, P.A., Singer, S.J. 1986. *Proc. Natl. Acad. Sci. USA* **83**:9001–9005
46. Maloney Huss, K., Lybrand, T. 1992. *J. Mol. Biol.* **225**:859–871
47. Manoil, C. 1990. *J. Bacteriol.* **172**:1035–1042
48. Manoil, C., Beckwith, J. 1985. *Proc. Natl. Acad. Sci. USA* **82**:8129–8133
49. Manoil, C., Beckwith, J. 1986. *Science* **233**:1403–1408
50. McCrea, P.D., Engleman, D.M., Popot, J.-L. 1988. *TIBS* **13**:289–290
51. McGovern, K., Ehrmann, M., Beckwith, J. 1991. *EMBO J.* **10**:2773–2782
52. Michaelis, S., Inouye, I., Oliver, D., Beckwith, J. 1983. *J. Bacteriol.* **154**:366–374
53. Mueckler, M., Caruso, C., Baldwin, S.A., Panico, M., Blench, I., Morris, H.R., Allard, W.J., Lienhard, G.E., Lodish, H.F. 1985. *Science* **230**:941–945
54. Nilsson, I., von Heijne, G. 1990. *Cell* **62**:1135–1141
55. Ovchinnikov, Y.A., Abdulaev, N.G., Vasilov, R.G., Vturina, I.Y., Kuryatov, A.B., Kiselev, A.V. 1985. *FEBS Lett.* **179**:343–350
56. Page, M.G.P., Rosenbusch, J.P. 1988. *J. Biol. Chem.* **263**:15906–15914
57. Popot, J.-L., Engelman, D.M. 1990. *Biochemistry* **29**:4031–4037
58. Popot, J.-L., Gerchman, S.-E., Engelman, D.M. 1987. *J. Mol. Biol.* **198**:655–676
59. Randall, L.L., Hardy, S.J.S. 1986. *Cell* **46**:921–928
60. Rao, A., Martin, P., Reithmeier, R.A.F., Cantley, L.C. 1979. *Biochemistry* **18**:4505–4516
61. Rao, J.K.M., Argos, P. 1986. *Biochim. Biophys. Acta* **869**:197–214
62. Reed, K.E., Cronan, J.E., Jr. 1991. *J. Biol. Chem.* **266**:11425–11428
63. San Millan, J.L., Boyd, D., Dalbey, R., Wickner, W., Beckwith, J. 1989. *J. Bacteriol.* **171**:5536–5541
64. Sarkar, H.K., Thorens, B., Lodish, H.F., Kaback, H.R. 1988. *Proc. Natl. Acad. Sci USA* **85**:5463–5467
65. Senstag, C., Stirling, C., Schekman, R., Rine, J. 1990. *Mol. Cell Biol.* **10**:672–680
66. Singer, S.J. 1990. *Annu. Rev. Cell Biol.* **6**:247–296
67. Steck, T.L. 1974. *J. Cell Biol.* **62**:1–5
68. Stock, J.B., Ninfa, A.J., Stock, A.M. 1989. *Microbiol. Rev.* **53**:450–490
69. Sugiyama, J.E., Mahmoodian, S., Jacobson, G.R. 1991. *Proc. Natl. Acad. Sci. USA* **88**:9603–9607
70. Tadayyon, M., Broome-Smith, J.K. 1992. *Gene* **111**:21–26
71. Traxler, B., Beckwith, J. 1992. *Proc. Natl. Acad. Sci. USA* **89**:10852–10856.
72. Verrall, S., Hall, Z.W. 1992. *Cell* **68**:23–31
73. von Heijne, G. 1986. *EMBO J.* **5**:3021–3027
74. von Heijne, G. 1989. *Nature* **341**:456–458
75. von Heijne, G., Gavel, Y. 1988. *Eur. J. Biochem.* **174**:671–678
76. von Heijne, G., Manoil, C. 1990. *Protein Engineering* **4**:109–112
77. Wang, H.-Y., Lipfert, L., Malbon, C.C., Bahouth, S. 1989. *J. Biol. Chem.* **264**:14424–14431
78. Wilmes-Riesenberg, M.R., Wanner, B.L. 1992. *J. Bacteriol.* **174**:4558–4575
79. Yaeger, M., Gilula, N.B. 1992. *J. Mol. Biol.* **223**:929–948
80. Yu, X.-M., Hall, Z.W. 1991. *Nature* **352**:64–67
81. Yun, C.-H., Van Doren, S.R., Crofts, A.R., Gennis, R.B. 1991. *J. Biol.Chem.* **266**:10967–10973
82. Zhang, Y., Broome-Smith, J.K. 1990. *Gene* **96**:51–57

Received 10 July 1992; revised 7 October 1992