Ann. Rev. Biophys. Biophys. Chem. 1986. 15:321–53

IDENTIFYING NONPOLAR TRANSBILAYER HELICES IN AMINO ACID SEQUENCES OF MEMBRANE PROTEINS

D. M. Engelman, T. A. Steitz, and A. Goldman

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

CONTENTS

PERSPECTIVES AND OVERVIEW

In recent years amino acid sequences for many integral membrane proteins have been determined. At the same time theoretical arguments and experimental evidence have accumulated to indicate that transbilayer

0883–9182/86/0610–0321$02.00

helices are a major motif in integral membrane protein structure. It may be possible to determine the location of such transmembrane helices directly from amino acid sequences using scales of polarity and sequential search protocols. The purpose of this review is to examine the approaches that have been used and to assess their utility. It is an opportune time to consider the issues, since the structure of three integral membrane proteins that are contained in the photosynthetic reaction center of *Rhodopseudomonas viridis* have been determined at high resolution. This new structural information, combined with increasing evidence concerning the structure of bacteriorhodopsin, permits a critical test of the main ideas involved in searching sequences for transbilayer structural elements.

We find that search procedures based on a moving window that scans a sequence twenty residues at a time are suitable for finding the transbilayer helices that are known to exist. For very nonpolar helices separated by polar polypeptides, many of the proposed polarity scales succeed equally well; where the helices contain more polar groups, however, the choice of scales becomes critically important.

In our discussion we consider the arguments that support the notion that helical structure will be a dominant motif in integral membrane protein organization. We introduce and discuss the problem of suitable scaling of amino acids in terms of their polar and nonpolar characteristics, and discuss further the use of such scales in prediction of protein structure. Finally, we examine the cases in which the validity of predictions can be assessed. It is our contention that a suitable scale and protocol can lead to the successful identification of transmembrane helical structures in integral membrane proteins.

INTRODUCTION

Since the determination of the first protein structures, it has been generally observed that the interiors of proteins tend to contain fewer charged and polar residues and more nonpolar residues than the surfaces in contact with water (3, 44, 45, 67, 103). The role of the hydrophobic effect in protein folding has received constant and detailed attention since Kautzmann's influential discussion (43). The notion that hydrophobicity is an energetic determinant in protein folding has led to attempts to characterize the surfaces in contact with water (9, 10, 51), to document the hydrophobic components of interior regions of proteins (10, 12, 28, 39, 79), and to develop quantitative scales of the relative polarity of each amino acid in a polypeptide (12, 19, 21, 22, 29, 40, 49, 53, 57, 60, 62, 63, 74, 77, 88, 89, 91, 93–95, 99–101, 105). The nature of polarity scales and their formulation are the subject of an ongoing discussion (e.g. 12, 29, 77, 78). The main theme of this

discussion is how amino acids partition from water into the interiors of globular proteins. In our treatment a different focus is taken: We examine segments of amino acid sequences as they interact with the nonpolar region of a lipid bilayer. Clearly such interactions are dominated by the hydrophobic effect and the set of interactions involved is different from that involved in the complex interior of a globular protein. Our discussion of scales focuses on the peculiarities of the lipid-protein interface.

HELICES IN LIPID BILAYER ENVIRONMENTS

Theoretical Considerations

Helical structure is known to be induced in polypeptides in nonaqueous environments (85, 86). The large free energy cost of transferring an unsatisfied hydrogen bond donor or acceptor from an aqueous to a nonpolar environment or of breaking such a bond in a nonpolar environment suggests that hydrogen bonds must be systematically satisfied as proteins are inserted into a membrane environment. In the nonaqueous interior of a lipid bilayer where the alternative of hydrogen bonding to water is absent, the energy of each hydrogen bonded pair compared to the unpaired state is approximately 6 kcal/mol (1a), so the lipid environment is extremely unfavorable for unfolding a polypeptide.

Typical energetics for the formation of helices in nonaqueous and aqueous environments and the transfer of a polypeptide between them are summarized in Figure 1. In order to evaluate the relative stability of a transmembrane helix and an unfolded polypeptide in the lipid environment, one must consider at least three factors: hydrogen bonding, conformation entropy, and van der Waals interactions. Although the conformational entropy term favors the unfolded state, it is considerably smaller than the energy term owing to hydrogen bonding in the lipid environment as mentioned above. Since an unfolded polypeptide chain can have many conformations and a folded helix has a well-defined structure, the entropy of the folded structure is much lower. The approximate magnitude of this term is about 1.25 kcal/mol per peptide bond (69). For a 20–amino acid helix about 24 kcal/mol favoring the unfolded state would result from entropy. In a 20–amino acid transmembrane helix, 16 hydrogen bonds would form in the nonaqueous region, contributing -96 kcal/mol favoring the helical conformation. The energy changes due to van der Waals interactions, while important in dictating some details of the final structure, would be small on the scale of energies being considered here since the unfolded chain would have interactions with solvent that would be replaced by interactions with itself as it folded. We conclude that a 20–amino acid α-helix would have a total difference free energy of stabiliza-

tion in the lipid bilayer of approximately 70 kcal/mol compared with the unfolded state (Figure 1).

If we assume that the 20-residue peptide forms a helix at relatively low energy cost in an aqueous environment (7, 47) and that the side chains of the helix are nonpolar, then the chain would be more stable by tens of kcal/mol as a transbilayer helix traversing the nonpolar region of the lipids than it would be either as a helix or as an unfolded chain in the aqueous environment (by about 30 kcal/mol in the example shown in Figure 1). Spontaneous helix formation in water results from a balance of favorable hydrophobic and hydrogen-bonding energies with unfavorable entropy contributions (7, 47).

If the alternative, insertion of a hydrophobic chain of amino acids into a bilayer followed by folding, is considered, the role of hydrogen bonds becomes immediately apparent. Completion of the thermodynamic cycle in Figure 1 results in the conclusion that the insertion of the unfolded chain is extremely unfavored (+42 kcal/mol). The total free energy includes unfavorable hydrogen bond contributions and favorable hydrophobic

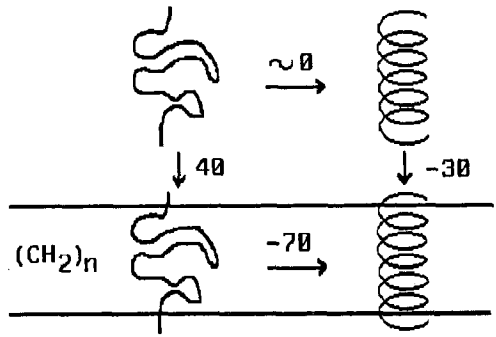


Figure 1 The formation and insertion of a polyalanine helix 20 residues long (21). We assume that the formation of the helix in solution will be at least marginally stable and thus require about 0 kcal/mol. The equilibrium free energy for transferring the helix to a position spanning the nonpolar region of a bilayer includes −32 kcal/mol from the hydrophobic effect; +5 kcal/mol have been included as the entropic term.

The alternative pathway from the unfolded state may be considered, in which the chain first inserts and then forms a helix. To obtain free energies for the process, we consider the combination of chain entropy effects and hydrogen bonding, and obtain an approximate value of −70 kcal/mol for the folding of the chain in the nonaqueous environment. Completing the cycle then gives a value +40 kcal/mol for the process of moving the chain from the aqueous to the nonaqueous environment without folding it. It is clear that the process for insertion of a random polypeptide chain is highly unfavorable, and that some folding that results in the formation of hydrogen bonds must occur prior to the entry of the polypeptide into the nonaqueous environment.

energies, so the unfavorable hydrogen bond energies are even larger than the total. Thus, we conclude that a polypeptide coil cannot be inserted into the bilayer and then fold, but rather a secondary structure must form before insertion into the bilayer (21, 22). This is the major argument favoring the existence of helical structures in membranes: Partially assembled, hydrophobic helices are energetically favored to insert into the bilayer whereas random coils or partial β-sheets (e.g. a beta hairpin) are not.

Although in the nonpolar interior of soluble proteins peptide backbone hydrogen bonds are satisfied by the formation of either α-helix or β-sheet structures, we expect that the helix will be found to be the dominant secondary structure in lipid bilayers (21, 31, 85, 86). Obviously, a single crossing of the lipid bilayer can only be achieved by a helix if all H-bonds are to be satisfied. We have previously argued (21) that the requirement of cotranslational insertion (70, 81) and folding of globular membrane proteins into the lipid bilayer limits the possible secondary structures that can be inserted to a helical hairpin in most cases. One can imagine that pairs of amphipathic but hydrophobic helices might be stable both in an aqueous environment where they are synthesized and in the bilayer where they are assembled into protein. In aqueous solution, the more polar faces of the helix pairs can face water, whereas in the bilayer the helices can rotate to face the polar groups inward, away from lipid (22).

If the alternative of β-sheet structure is considered, it is clear that progressive insertion during protein synthesis would be problematical. The beta strands would have to be inserted as hairpins or single strands in which many hydrogen bond donors and acceptors would be left unsatisfied. While a β-barrel can be imagined as a structural alternative to helices (e.g. 31), the entire barrel would have to form in solution prior to its insertion into the bilayer. Folding in solution requires that many hydrophilic residues be outside the barrel whereas stability in the membrane environment requires the reverse. It may be that different conditions of polarity, such as the creation of large aqueous channels, permit alternative structures of this kind (80, 83). Nonetheless, the use of helical structures as an efficient strategy for progressively satisfying hydrogen bond requirements in nonaqueous environments leads to the expectation that helices are major constituents of membrane protein organization.

Experimentally Observed Membrane Protein Secondary Structures

Present structural data on four polypeptides support the existence of transbilayer helices. At moderate resolution the bacteriorhodopsin structure shows the presence of seven transmembrane rods that have the appropriate dimensions and packing to be α-helices (32, 52). Spectroscopic

studies suggest the presence of a large amount of helix (58), and the prevalent interpretation is that the structure contains seven transbilayer helices. The other three polypeptides are subunits of the photosynthetic reaction center. The recent determination of the reaction center structure has led to the conclusion that two of the four proteins contain five transmembrane helices each and a third subunit contains a single transmembrane helix (14, 15). These appear to be the only structures traversing the lipid bilayer, although the position of the bilayer is inferred from the structure in the crystal. In the case of the photosynthetic reaction center the structure is known at high resolution and is unambiguous. Thus, it can be argued that structural data support the presence of 18 helical segments in three globular and one anchored membrane protein. These helices can usefully serve as tests of procedures for defining transmembrane segments (see below).

It is known that other kinds of transmembrane structures exist. Studies of matrix porin from *Escherichia coli* outer membranes suggest very strongly that β-sheet is the dominant secondary structural feature (18, 80, 83). An assembly of porin molecules forms an aqueous channel through the lipid bilayer. As the channel is large and can accommodate many polar groups, there are additional possibilities for a suitable structure that can be assembled into the bilayer. These possibilities do not exist in a globular membrane protein that is surrounded by lipid and that does not contain an aqueous channel (or an anchored protein). We confine our discussion to the prediction of helical structures that can be tested using the set of globular and anchored membrane protein structures.

DETERMINATION OF HELIX LOCATIONS FROM SEQUENCE DATA

If one assumes that transmembrane α-helices are present in the structure of an integral membrane protein, methods to locate them in the protein sequence on theoretical grounds would be very useful. Many methods for evaluating polarity in amino acid sequences have been developed and applied (2, 19, 20, 22, 50, 78, 88, 93, 94) following the original approach of Rose & Roy (76, 79). Each uses a progressive analysis in which successive regions of the polypeptide are evaluated with respect to some scale of polarity as shown in Figure 2. In the following sections we examine the polarity scales to determine which are appropriate, consider the choice of window length for the moving analysis, and present tests of the significance and success of the predictions. All the methods have as their goal the identification of amino acid sequences that are sufficiently hydrophobic

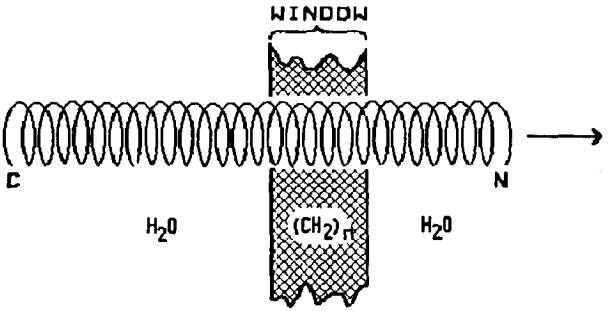


Figure 2 Schematic representation of the free energy computation for locating helices that are stable as transmembrane structures. The amino acid sequence of a protein is arranged as a continuous α-helix and is moved through a nonaqueous window. For each segment of the polypeptide chain the free energy for transferring the segment from the aqueous to the nonaqueous environment is calculated. The free energy transfer is plotted versus the N-terminal amino acid in the segment under consideration. In order to represent favorable insertions as peaks in the graph, the sign for the free energy is reversed, representing the transfer from the nonaqueous to the aqueous environment.

and sufficiently long ($\geq$20 amino acids) to imply the existence of a transmembrane helix.

To frame the discussion of polarity scales we discuss in some detail the scale that we have developed during the past several years (19, 21, 22, 88), partly because we believe it is the most appropriate scale and partly so we can use it as a point of reference in our discussion of other scales and approaches. We assume that various side-chain components may be considered separately, that the details of helical structure are important in establishing an appropriate scale, and that the bilayer interior is a region of dielectric constant 2 containing no hydrogen bond donors or acceptors.

Appropriate Scales of Hydrophobicity for Bilayers and Protein Interiors

There is considerable diversity of opinion concerning the appropriate choice of polarity scale. Scales have been developed on the basis of solubility measurements (13, 29, 48, 63, 90), vapor pressures of side-chain analogs (33, 100–102), and analysis of side-chain distributions with soluble proteins (11, 29, 39, 40, 74, 79, 99). The use of side-chain distributions is complicated by the fact that hydrophobic residues are frequently found on protein aqueous surfaces (72) and by the fact that side chains span regions of different polarity (29, 77). The partition and vapor pressure measurements differ in that different assumptions are made concerning an appropriate analog of the protein dielectric interior. These issues have been extensively

discussed in recent articles (12, 29, 77, 78). It is not surprising that the interior of a protein presents difficulties for modeling. The dielectric environment is extremely nonuniform, being influenced by the presence of many polar groups and hydrogen bond networks; the use of a bulk dielectric constant cannot represent its detailed fluctuations. It may prove necessary, as Guy (29) suggests, to consider a more detailed view of protein structure involving a distinction between the deep interior and the surface regions of a protein.

The hydrophobic environment in a membrane interior is simple compared with a protein interior. The hydrocarbon chains create a comparatively uniform, nonpolar environment. As the environment presents no hydrogen bond donors or acceptors, and as its dielectric constant is lower than that of a protein interior (59), scales developed from the examination of soluble, globular proteins would seem inappropriate for investigating amino acid side chains exposed to the lipid environment. Transfer free energy experiments based on the solubility of compounds in water and a nonpolar solvent analogous to a lipid bilayer are confounded by the very low solubility of even moderately polar compounds in media having dielectric constants of 2. The small number of solved structures creates an inadequate data base for the kind of statistical treatment used in categorizing side chains in globular soluble proteins. A promising approach is examination of the partitioning of compounds between an aqueous phase and the vapor state (33, 102). An alternative approach is the use of theoretical and experimental values for components of each amino acid side chain to derive a polarity scale (19, 21, 22, 93, 95). These alternatives are discussed further below.

A Polarity Scale for Identifying Transmembrane Helices

The arguments that led to the development of the Goldman, Engelman, Steitz (GES) hydrophobicity scale (19, 21, 22, 88) are outlined below. The development is rather similar to that of Von Heijne's early work (93, 94) but differs in some important details that are discussed later.

The major energetic factors favoring the partitioning of an amino acid side chain from aqueous solution into a membrane bilayer are hydrophobic interactions; those factors favoring its solution in the aqueous phase are interactions of polar and charged side chains with water. In order to make a quantitative estimate of the relative energies involved, the free energy of transfer of both the hydrophobic and hydrophilic components of each amino acid from water into oil were assigned. In order to consider the specific case of an α-helical polypeptide in a low-dielectric environment it is important that the scale be specifically adapted to the details of such a structure. This presents a dilemma. Since experimental scales have

previously been based on the properties of individual amino acids in solubility measurements (13, 29, 48, 63, 90) or in transfer to the vapor phase from water (33, 100–102), they do not specifically address the circumstance of amino acids in helices. However, no experimental scale has been developed for helical structures.

We therefore developed a mixed scale (19, 21, 22, 88) in which the nonpolar properties of the amino acids as they exist in a helix were calculated using a semitheoretical approach that combines separate experimental values for the polar and nonpolar characteristics of groups in the amino acid side chains. This procedure, in essence, divides amino acids more finely than a simple consideration of main chain versus side chain characteristics.

Initially (19, 21, 88) we assumed an average hydrophobicity for a 20-residue α-helix based on the surface area of a typical helix. To the favorable baseline hydrophobicity of −30 kcal/mol of helix we added the unfavorable energetic contribution arising from burying the various polar and charged residues. Use of this scale on the sequence of bacteriorhodopsin showed seven plausible hydrophobic regions (88). This scale was modified (22; Goldman, unpublished, 1982) to calculate the hydrophobicities for each of the 20 amino acids as they occur in an α-helix.

The hydrophobic component (Table 1) of the free energy of water-oil transfer can be calculated from the surface area of an amino acid side chain in an α-helix (9, 51, 71). Hydrophobic interactions tend to reduce the nonpolar surface area in contact with water. Their approximate magnitude has been obtained by measuring the partitioning of compounds between water and nonpolar solvents. The hydrophobic free energy thus measured has been shown (9, 71) to correlate linearly with total surface area in contact with water (51). Thus, calculation of the total contact surface area of a polypeptide that can be removed from interaction with water leads to an approximate value for the hydrophobic transfer free energy. We have used the surface area computations of Richmond & Richards (73) to obtain the surface area for each amino acid as it would be exposed in an α-helix of polyalanine. (The solvent-accessible surface varies somewhat in actual cases, depending on the neighboring residues, but this is a second-order effect.) The surface areas could then be converted into hydrophobic free energies (Table 1). In this way the experimental free energies of transfer can be adapted to the specific case of amino acids in an α-helix.

The free energy for inserting charged groups into a bilayer can be considered as having two components: the energy required to produce an uncharged species by protonation or deprotonation, and the energy required to partition the uncharged but polar portions of side chains from water to the nonaqueous phase (21). Our calculations using the Born

approximation (5) showed that the transfer of a formal charge from an aqueous to a nonaqueous phase requires very substantial energy, probably on the order of 40 kcal/mol (21). Recent calculations by Honig & Hubbell (35) have led to a similar conclusion. The alternative of producing the uncharged species and partitioning it requires 10–17 kcal/mol (21, 35). Therefore, we consider that potentially charged amino acids (glutamic acid, lysine, aspartic acid, histidine, and arginine) will be transferred as the uncharged species. If we assume that the process occurs at or near neutrality, we can calculate the energy required for protonation or deprotonation by assuming a standard pK and a requirement for 99% conversion to the uncharged species. The energies obtained are included in the hydrophilic energies listed in Table 1.

There are also energy costs associated with the transfer of uncharged polar groups. These energies arise principally from the participation of side

Table 1 Transfer free energies for amino acid side chains in α-helical polypeptides[a]

	Hydrophobic	Hydrophilic	Water-oil
Phe	−3.7		−3.7
Met	−3.4		−3.4
Ile	−3.1		−3.1
Leu	−2.8		−2.8
Val	−2.6		−2.6
Cys	−2.0		−2.0
Trp	−4.9	3.0	−1.9
Ala	−1.6		−1.6
Thr	−2.2	1.0	−1.2
Gly	−1.0		−1.0
Ser	−1.6	1.0	−0.6
Pro	−1.8	2.0	0.2
Tyr	−3.7	4.0	0.7
His	−3.0	6.0	3.0
Gln	−2.9	7.0	4.1
Asn	−2.2	7.0	4.8
Glu	−2.6	10.8	8.2
Lys	−3.7	12.5	8.8
Asp	−2.1	11.3	9.2
Arg	−4.4	16.7	12.3

[a] Values are given for the hydrophobic and hydrophilic components of the transfer of amino acid side chains from water to a nonaqueous environment of dielectric 2. The hydrophobic term is based on a treatment of the surface area of the groups involved. The hydrophilic term principally involves polar contributions arising from hydrogen bonding interaction. Also included in the hydrophilic term is the energy required to convert the charged side chains to neutral species at pH 7 (19, 21, 22, 88).

Table 2 Approximate water-oil transfer free energies for various groups[a]

Group	G (H_2O-Oil)
—OH	4.0
—NH_2	5.0
—COOH	4.3
C=O	2.0

[a] Values are derived principally on the basis of observations using nonpolar oils. The studies on which they are based are summarized by Davis (7).

chain groups in hydrogen bonds with water. It is difficult to treat the hydrogen bonding potential explicitly; one must rely to a large extent on experimental measurements based on the solubility of various compounds. From extensive reviews of the data (13, 91) we conclude that the energies required for transfer of polar groups from water to oil are approximately as shown in Table 2.

Additional important specific considerations emerge regarding α-helices. Serine and threonine in α-helical segments of proteins are known to participate in shared hydrogen bonds with the backbone carbonyl groups (27). Such sharing reduces the free-energy contribution opposing transfer from the aqueous environment to the nonpolar region of the membrane. A further consideration is the interaction of groups along the helical axis, which is discussed below.

The contributions from different polar interactions were combined for the hydrophilic term in Table 1. The net transfer free energies are the sum of hydrophobic and hydrophilic components for each amino acid. The scale uses a finer division of properties than some other scales have employed, treating the contributions of hydrophobic surfaces and individual side-chain polar groups separately. Its strengths are that it specifically addresses the issue of helical structure and that it is based on a transfer from an aqueous to a low-dielectric hydrocarbon region. In our discussion of other approaches below, the GES scale is taken as a point of comparison.

Comparison with Other Polarity Scales

We now concentrate our attention on other scales that have been developed for the examination of transbilayer helices. Of greatest importance in our discussion are the scales of Von Heijne (94, 95) and Kyte & Doolittle (50) and a scale based on partitioning between water and the vapor phase (33, 100). Other scales have emerged from an examination of partitioning of amino acids into protein interiors or into more polar solvents such as

alcohols, e.g. the scales of Nozaki & Tanford (63), Rose & Roy (79), Guy (29), and Janin (39); these scales are not considered in detail for the reasons presented above, the main point being that partitioning of side chains from water to a protein interior is not equivalent to partitioning of side chains from water to a lipid environment.

Figure 3 shows a comparison of the GES scale (21, 22, 88) with that of Von Heijne (the VH scale). Here we use Von Heijne's revised scale (93, 94) since the original scale (95) had several incorrect chemical assumptions. The GES and VH scales correspond rather closely, with the exceptions of threonine, serine, proline, and lysine. In the cases of serine and threonine the differences are accounted for by the consideration in the GES scale of hydrogen bonds between side chains and main chains within an α-helix.

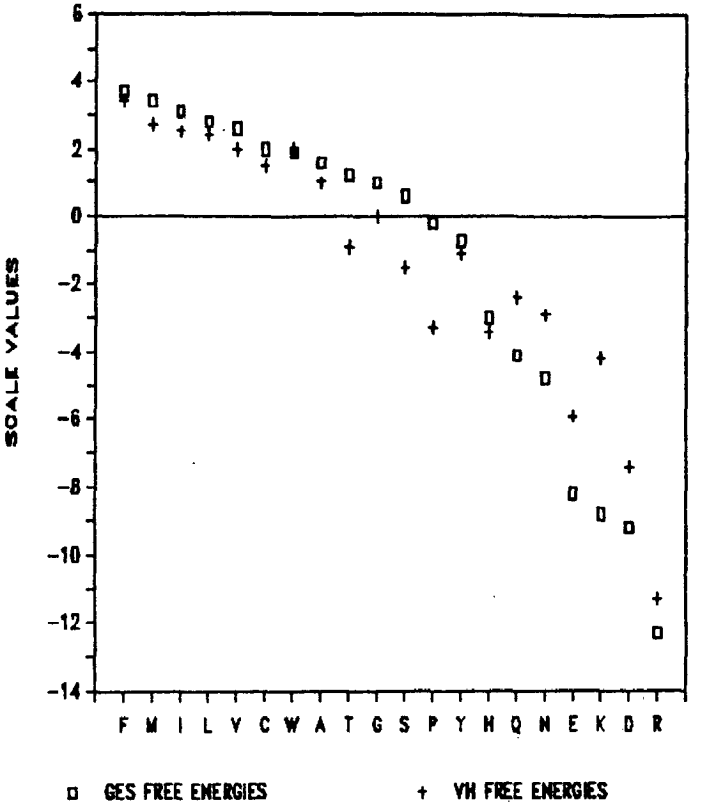


Figure 3 Comparison of the GES (19, 21, 22, 88) and VH (93, 94) scales. □: GES free energies; +: VH free energies. Free energies are represented on the vertical axis in kcal/mol for the transfer of each side chain from a nonaqueous to an aqueous environment. The scales are generally similar with the exception of Ser, Thr, Pro, Lys.

Proline is given a much more polar character in the VH scale, as it is assigned two hydrogen bonds with water. This is incorrect since the only hydrogen bonding group is the main-chain carbonyl group that is not satisfied because of the closure of the imino ring. On the other hand, lysine is given a more nonpolar character by Von Heijne than aspartic or glutamic acid. Although lysine is more hydrophobic (by ~ 1 kcal/mol) than aspartic or glutamic acids, its pK is further removed from neutrality, giving rise to a hydrophilic component that eliminates the difference in our view (at pH 7.0). With these qualifications, the VH scale is, on the whole, rather similar to the GES scale.

A much used scale is that proposed by Kyte & Doolittle (the KD scale) (50). In an extensive and carefully reasoned article they examined a number of alternative polarity scales. A combination of scales based on the observed behavior on partitioning from the aqueous environment to protein interiors and on water-vapor partition gave the best agreement with known cases from soluble, globular proteins. While this scale has many virtues, it is clear that the model under consideration does not address accurately the conformational and environmental aspects we have discussed above.

In Figure 4 the scale derived by Kyte & Doolittle is compared with the GES scale. The most striking difference is that the polarities of aspartic acid, glutamic acid, lysine, and arginine are not as strong in the KD scale as in the GES scale. This is a consequence of the scaling procedures used by Kyte & Doolittle to merge different scales in their analysis. As in the case of the VH scale, the contributions of threonine and serine are considered more polar than we think appropriate. Further, tryptophan is considered a substantially or partially polar amino acid because of the ring nitrogen. While this may explain why tryptophan is predisposed to orient near interfaces, the polarity seems inappropriate in terms of the overall nonpolarity of the side chain. In general, the KD scale is in reasonable although not detailed agreement with the GES scale on the matter of the hydrophobic amino acid side chains. The differences, however, have important consequences in the prediction of transmembrane helices in cases where polar or potentially charged groups are in regions traversing the membrane (see below).

A significant contribution has been provided by measurements of water-vapor partition coefficients for model compounds containing amino acid side chain components. Since the vapor state does not provide hydrogen bonding groups, it would seem to be a good choice as an analog for a nonpolar bilayer interior. Free energies derived from these measurements (33, 100) have been merged and corrected by Kyte & Doolittle (50) to give a vapor-water transfer free energy scale (the VW scale).

Figure 5 shows a comparison of the VW scale and the GES polarity scale.

Some very large differences are evident. In developing their hydropathy scale, Kyte & Doolittle found it reasonable to adjust a number of the VW values based on chemical arguments. For example, because of the nonpolar character of its side chain, phenylalanine would be unlikely to have an equal probability of being found in an aqueous environment and in a membrane or protein interior. Similarly, on this experimental scale methionine is found to have a slightly polar character, in contradiction to its occurrence in the interior of known proteins and the apparently nonpolar character of its side chain. Also, cysteine is given a polar character, as is tryptophan. The cases of threonine and serine are interesting; the magnitudes of polarity are in agreement with the presumed value for the solvation of a hydroxyl group given in Table 2. Of course, as with the scales previously discussed the

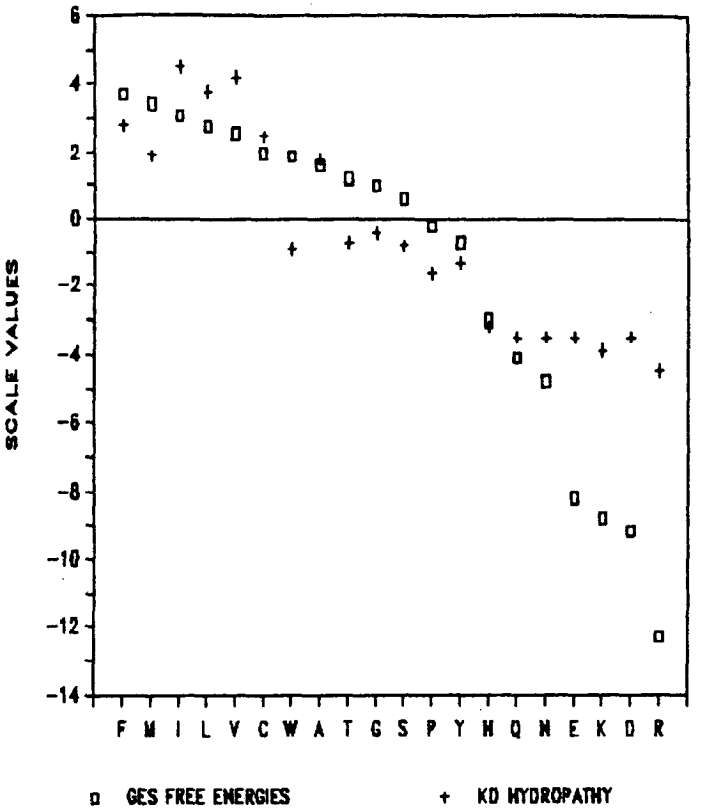


Figure 4 Comparison between the GES scale and the KD hydropathy scale (50). Values for the scales are on the vertical axis (free energy for the GES scale and the hydropathy for the KD scale).

structural assumptions regarding a free amino acid versus an amino acid in a helix alter the view of threonine and serine as polar amino acids. Rather striking is the extreme polar character accorded tyrosine using the transfer energy measurements. It is possible that interactions in the vapor phase, such as the dimerization of carboxylate groups, will distort estimates of the transfer energy. Furthermore, the present data do not give values for glycine, proline, or arginine.

While the VW scale appears useful a priori, the measurements that have so far been made using water-vapor transfer have resulted in somewhat perplexing conclusions concerning polarity. Kyte & Doolittle felt compelled to modify the direct conclusions from the transfer measurements and to reset the point on the scale at which zero transfer free energy is located.

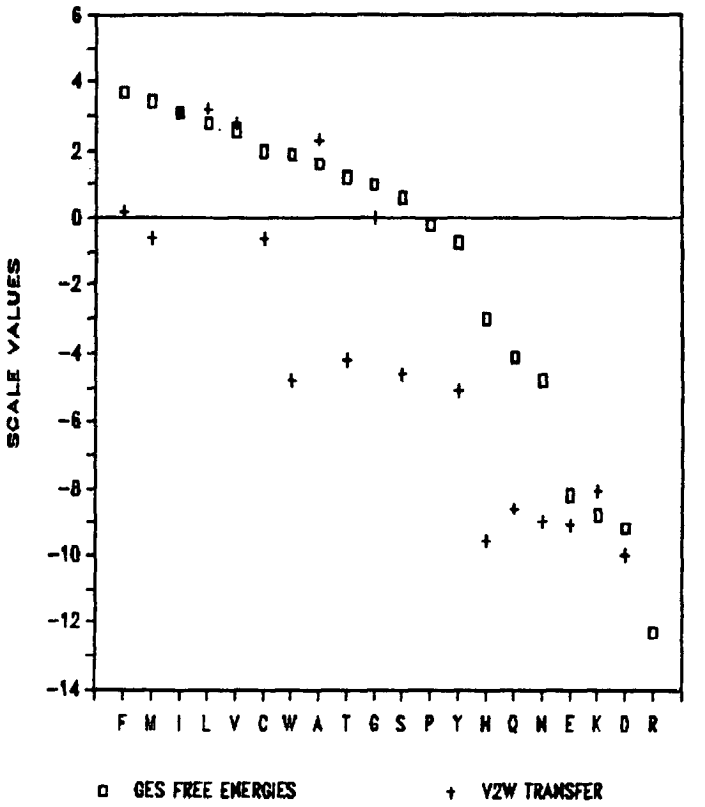


Figure 5 Comparison of GES and VW scales (29). The vertical axis represents transfer free energy in kcal/mol.

Rose et al (77) have noted that these values are not in agreement with other measures of polarity or with the observed occurrence of amino acids in protein interiors. We agree that some features of the scale derived from transfer free energies are surprising and that the scale may not be useful in efforts to predict transmembrane segments of polypeptide chains.

Polarity scales based on transfer of amino acids from water to various alcohols have been widely used (33, 63). Guy (29) has summarized the results from a number of experiments of this kind and has put them on a common scale for alcohol transfer. A striking fact is that the polar amino acids are assigned values near zero on this scale. Thus, virtually any polypeptide would be predicted to partition into a nonpolar phase based on this analysis. Clearly a restructuring would be needed for such a polarity scale to be useful in the kinds of scanning procedures under consideration in this article. Moreover, the arguments we have made concerning the suitability of alcohol partitioning measurements suggest that the scale would be inappropriate for this application.

Overview of Polarity Scales

In the foregoing discussion the GES scale was elaborated and compared with other approaches for the specific case of amino acid side chains in nonpolar environments. To apply any of these scales to the identification of transmembrane helices requires additional considerations and a computational approach. By applying the scales to prediction problems, differences in their properties become apparent and arguments concerning suitability are clarified. The application of scales is discussed below.

USE OF SCALES TO IDENTIFY TRANSMEMBRANE HELICES

The general approach to identifying transmembrane helices that has emerged in several publications is to use a scanning procedure by which an amino acid sequence can be progressively evaluated in terms of its polarity and hence its tendency to form transbilayer helices. Progressive analysis was first used in the study of globular, soluble proteins (76). In early versions, such methods invoked a smoothing algorithm applied to the detailed, residue-by-residue values of polarity or hydrophobicity. More recently, the approach taken for studies of membrane proteins has been to use a window scan of amino acid sequences (e.g. 21, 50, 88, 94). In a window scan, the sum of hydrophobicity or polarity values for a number of amino acids is taken progressively through the sequence (Figure 2). A plot is made of the position of a reference amino acid, either the first or the middle amino

acid in the region being summed, versus the value of the summed polarity or average polarity in the region. The use of such scans involves several choices of approach. These include attention to factors that arise as a consequence of the detailed structural model under consideration as well as choice of the window length appropriate for the search.

Energetic Consequences of Helical Conformation

In examination of potential α-helical structures, specific considerations of structural details are important. Two of these are the polar interaction of side chains along the helical structure and the different exposure of side chains to solvent when a helix is compared to isolated amino acids or extended chains. If nearby side chains have the potential to interact in the nonpolar environment, their interaction will modify the energy calculation for a transfer from the aqueous to the nonaqueous environment. For example, if the side chains of aspartic or glutamic acid are located one turn of a helix away from the side chains of arginine or lysine, interactions are possible and expected. Whether such interactions actually involve the formation of an ion pair or the formation of a strong hydrogen bond (1a) is an issue that cannot be addressed with the present information. The reader is referred to the excellent article by Honig & Hubbell in which the issues of group interactions in a nonpolar environment are treated (35) and to the review in this volume (36). These articles conclude that the energy required to transfer polar groups as ion pairs or as strongly hydrogen-bonded structures is certainly less than that required to transfer the groups separately.

At issue is the question of how much less energy is needed to transfer ion pairs. There are examples in protein structures in which it appears that the energy needed to transfer an ion pair from the aqueous environment to a protein interior may be very small. Benzamidine binds strongly to the catalytic pocket in trypsin. In this case it appears that the cost of forming the internal ion pair between the amide and a carboxyl group in the active site is very small (4). On the other hand, the treatment of Honig & Hubbell suggests that 10–15 kcal may be required to move a carboxyl and amino group as an interacting pair from the aqueous to the nonaqueous environment. It is therefore appropriate to include a term in the scanning procedure to allow for the interaction of polar groups. We have suggested (19, 88) that the value of this term might be 5–10 kcal/mol. The exact value is not known at present. However, some reduction of the energy requirement of the groups taken separately is appropriate where the groups are located 1,4 or 1,5 in the amino acid sequence. In the calculations presented below, the GES scale includes 10 kcal/mol as the contribution from paired amino and carboxyl groups along a helix.

Entropy of Immobilization

An additional factor that must be considered in the computation of an energy profile is the entropy of immobilization involved in moving a macromolecule from a solution to a lipid bilayer. Of the six degrees of freedom a molecule has in solution, three are restricted by binding to the lipid bilayer. In the case of a loss of all six degrees of freedom in enzyme substrate interactions, the extreme value for the entropy of immobilization is thought to be about 20 kcal/mol (65a). The loss of one translational and two rotational degrees of freedom would reduce this value to about 10 kcal/mol (38). The fact that the macromolecule is not totally immobilized with the lost degrees of freedom (owing to the fluid character of the lipid bilayer) means that some further reduction is in order. We have adopted the use of 5 kcal/mol as the unfavorable free energy term, which represents the immobilization of a polypeptide chain binding to a lipid bilayer.

Choice of Window Length and Scanning Procedures

In choosing a window length for sequence analysis to locate transmembrane helical structures, two factors are important: the hydrophobic width of the bilayer itself and the orientation of a possible helix with respect to the bilayer plane. Progressive sequential analysis requires some decision as to the length of sequence that will be examined at each step. In early analyses windows as short as 7 (50) and as long as 20 amino acids (88, 19) were used. Others have adopted smoothing procedures or have used model functions to smooth the erratic behavior of small averaging windows.

The hydrophobic thickness of a lipid bilayer may vary considerably depending on the composition of the lipid fatty acyl chains and on the content of cholesterol. It has been shown, for example, that the hydrophobic thickness of the bilayer is proportional to chain length for fluid phosphatidylcholine bilayers formed from a series of phosphatidylcholines with different fatty-acid chain lengths. Thus, the thickness can vary by more than a factor of two (54). In choosing the length of the test window it would be optimal if one knew the hydrophobic thickness of the particular bilayer into which a protein was to be inserted. Nonetheless, a typical value for many lipid bilayers is of the order of 30 Å. For an α-helix to span a 30 Å distance, 21 residues are required because the interval between residues along the helix axial direction is 1.5 Å.

If the helix is tilted with respect to the bilayer plane, a longer helix can be accommodated in the hydrophobic region. It may be that a protein that consists of many helices contains helices of different tilts, and that a series of test window dimensions can reveal the presence of more extensive hydrophobic helices. A final point is that lipid bilayers in the fluid state

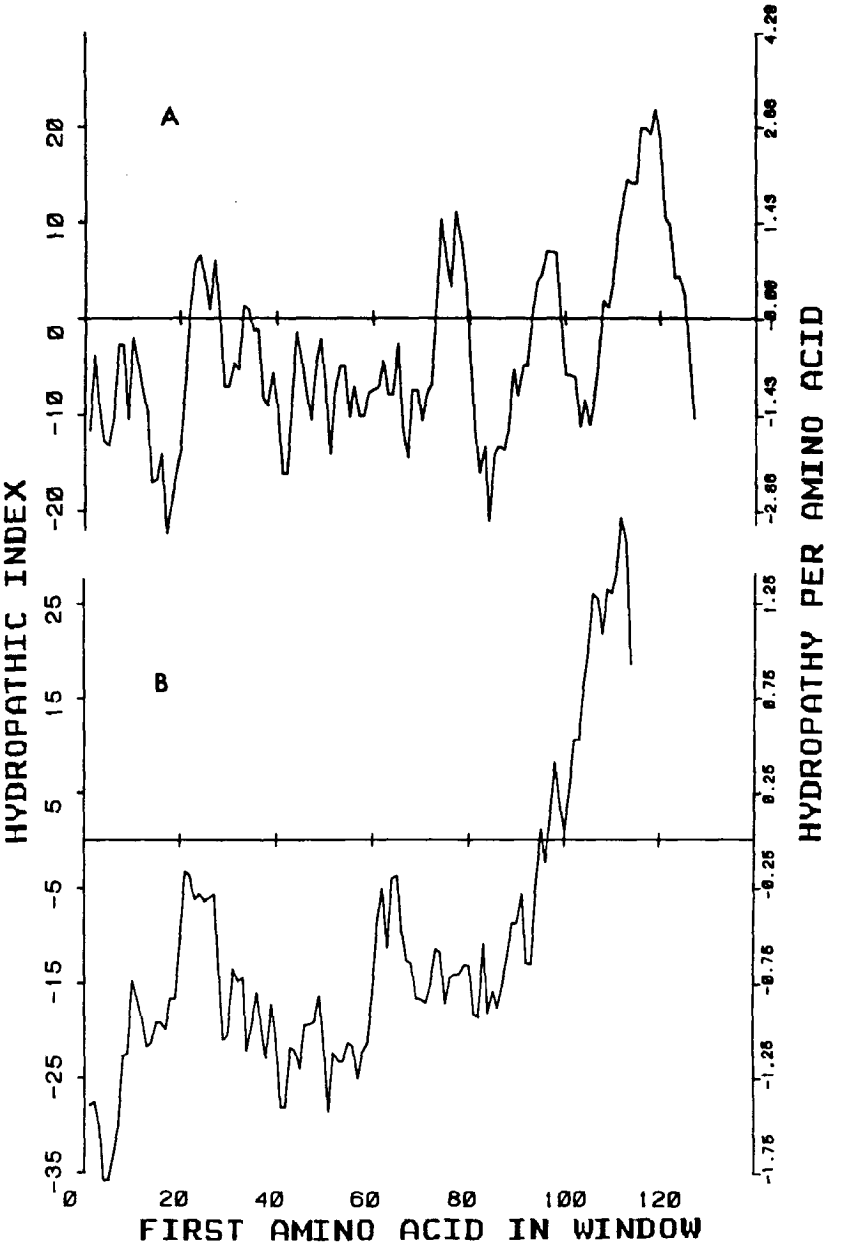


Figure 6 The sequence of cytochrome b5 (87) is analyzed using the KD hydropathy scale and windows of 7 (*A*) and 20 (*B*) amino acids. When a window of 7 is used, several important peaks occur in the profile. A lower background prediction is obtained when a 20-amino acid window is used, clearly contrasting with the experimentally known C-terminal anchoring region of the protein.

appear to be readily distorted (55, 66). Thus, the presence of a helix with a nonpolar dimension that does not match the hydrophobic thickness of the bilayer may be accommodated through distortion of the bilayer thickness.

The above discussion suggests that a reasonable choice for the test window is on the order of 20 amino acids, but that no unique number can be readily assigned. It would appear that short windows (on the order of 10 amino acids) or long windows (on the order of 30 amino acids) are unlikely to be optimal choices. It is possible that an inappropriate choice of window may give misleading results. Figure 6 shows the sequence of cytochrome b5, a protein that is anchored by its hydrophobic carboxyl terminus to the membrane of the endoplasmic reticulum (87). The KD scale was used with a window of 7 amino acids as Kyte & Doolittle originally specified (50). A peak is seen corresponding to the hydrophobic carboxy terminus, but additional peaks also appear. If the window is instead set at 20 amino acids, it becomes clear that the truly significant hydrophobic feature is the carboxy terminus and that the other peaks do not extend above zero.

Comparison of Scales

If a membrane-traversing region is extremely hydrophobic in character, virtually any scale will reveal its presence. Difficulties arise, however, in cases in which a helix contains some polar amino acids. Since the partitioning of helices is so strongly favored by the presence of exclusively nonpolar amino acids, stable structures are possible in which one or more amino acids in the middle of an otherwise nonpolar helix have strongly polar character (21). An example of this kind is provided by bacteriorhodopsin.

Figure 7 shows analyses of the bacteriorhodopsin sequence (46, 64) using the KD scale, the GES polarity scale, and the VW scale. The analyses shown in Figure 7 *B*, *C*, and *D* were each carried out with a window of 20 amino acids. It is evident that the analysis using the KD scale leads to a clear identification of only two helices; the other five expected helices are much less plainly revealed except as broad maxima. The VW scale gives only five peaks, and the values are radically shifted so that the free energies would lead to the prediction that the structures are unstable with the possible exceptions of one helix. Using the GES scale, seven distinct peaks separated by clear minima are observed.

Figure 7 Different analyses of the sequence of bacteriorhodopsin. *A* and *B* show the effect of using the KD scale and windows of 7 and 20 amino acids respectively. The appropriate choice of 20 does not reveal many of the helices nor does the choice of 7. *C* and *D* show the application of the VW and GES scales respectively, each with a window of 20. The VW scale, while revealing many of the helices in profile, appears far too negative in predicting stability. The GES scale, on the other hand, shows seven well-defined maxima which are thought to correspond to the seven helices present in the structure (see Figure 12).

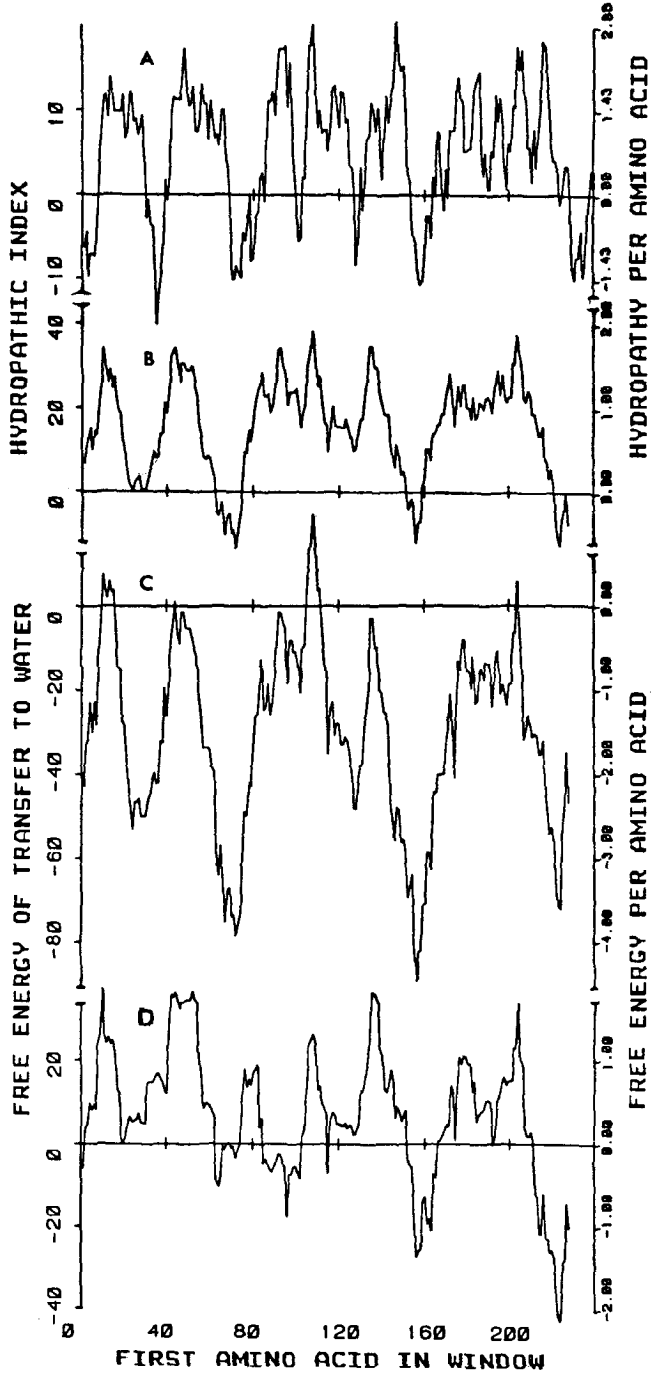


Figure 8 shows analyses of the sequence of glycophorin (92), which is known to span the red cell membrane (6), using the three scales and a window of 20 amino acids. In this case, where a very strongly nonpolar helix appears to be present, each scale gives a clear identification.

All these scales give, as expected, very similar results for the cases in which helices can be identified on the basis of inspection for nonpolar amino acids. The more difficult task of identifying helices that contain polar amino acids or that are not separated by clearly or strongly polar regions appears to be best accomplished using the GES scale.

Significance of Peaks in the Sequence Analysis

Given a choice of scale and window, the question arises of how peaks in the analysis are to be interpreted. An interesting test of the magnitude required for a peak to be biologically significant uses a series of deletion mutations in the anchoring peptide of the vesicular stomatitis viral coat protein (1). This protein appears to be anchored by a single 20–amino acid membrane-spanning sequence near its carboxy terminus (42, 75). Using genetic techniques several altered forms have been produced in which the hydrophobic region of the presumed anchor sequence has been varied in length and the cellular location of the modified protein has been determined.

Figure 9 shows the relative membrane stabilization calculated for the different modified coat proteins using a window of 20 amino acids and the GES scale. Rose and colleagues have determined the disposition of the different modified proteins (1) and have found that reduction of the anchoring sequence to 8 amino acids does not anchor the protein. However, proteins with a hydrophobic sequence of 14 or more amino acids are clearly anchored. With an anchor sequence of 12 amino acids, the protein appears to bind well in cytoplasmic membranes but only sparingly in the plasma membrane (which may be thicker). We can therefore say from inspection of Figure 9 that the peak corresponding to 20 kcal/mol appears to correlate with stable insertion and anchoring of the membrane protein.

It is not the case, however, that all proteins containing hydrophobic sequences identified in this way are membrane-spanning or, indeed, even membrane-associated proteins. An example is the sequence of trypsinogen (30), which is analyzed in Figure 10. Here a clear hydrophobic stretch is identified that, were it known to be a membrane protein, would be suspected as a transmembrane segment. Trypsinogen is, of course, a secreted, soluble protein. It is hazardous to assume that proteins that show peaks of about 20 kcal/mol must be integral membrane proteins.

We have confined our attention to prediction methods using polarity scales to identify nonpolar helices. Additional transmembrane structures may be found where the constraints are different, as in assemblies that form

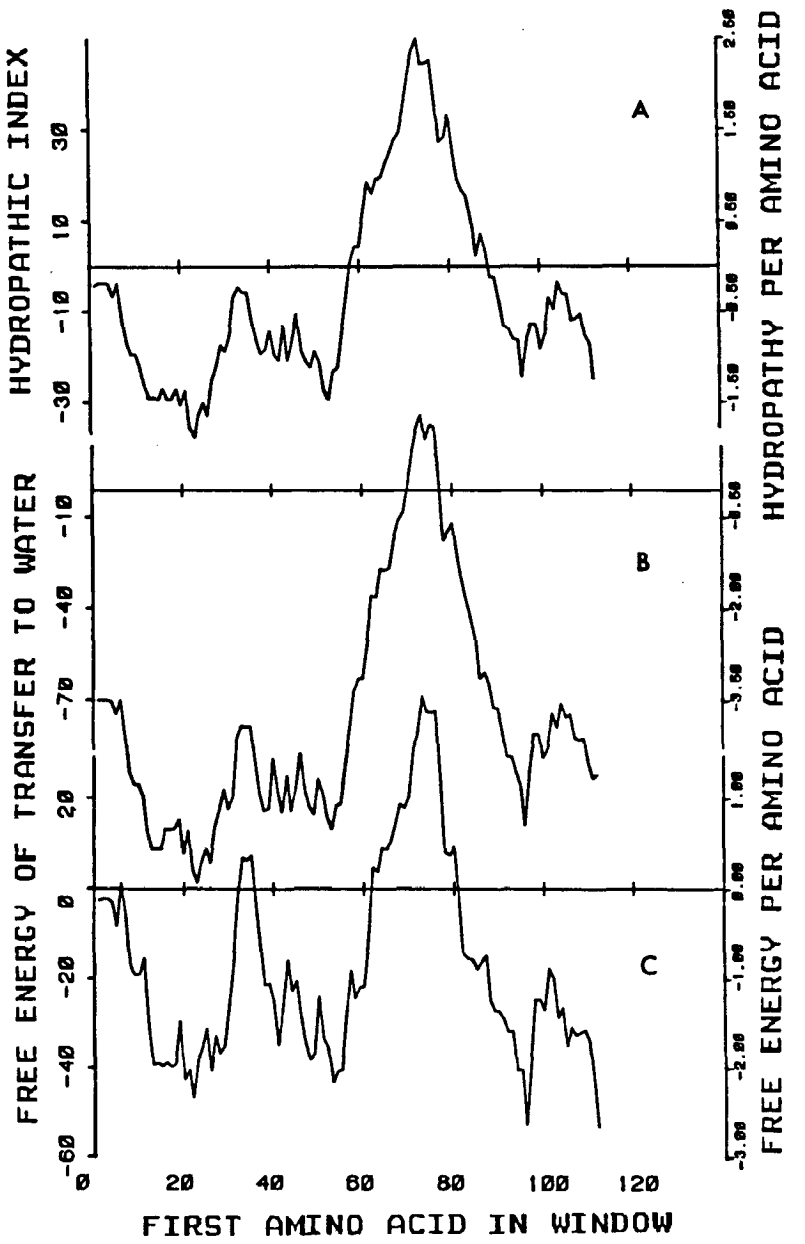


Figure 8 The sequence of glycophorin (92) is examined using the KD, VW, and GES scales. Each scale reveals the transmembrane region of the polypeptide.

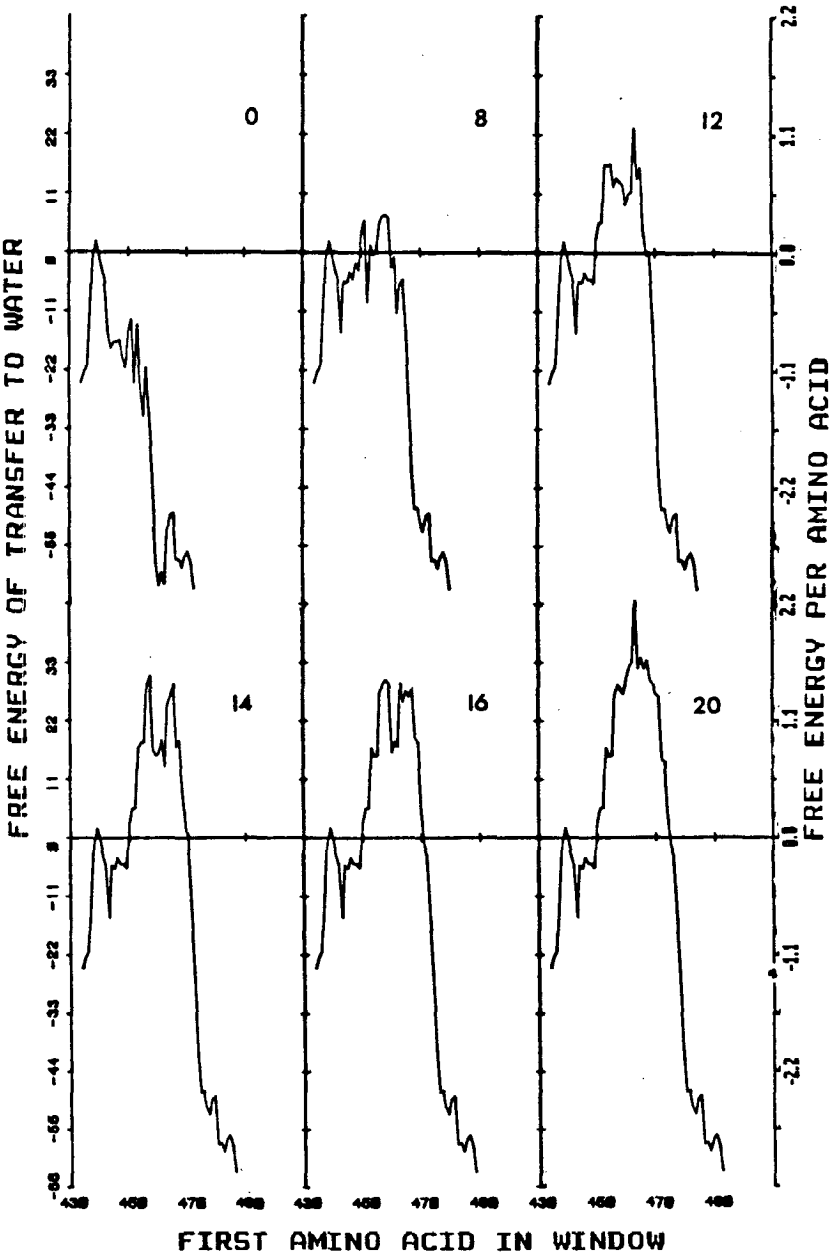


Figure 9 Profiles for the anchoring segment of the VSV G protein in progressively shortened versions (1). The VSV G protein has a membrane-anchoring sequence of 20 amino acids, which has been progressively shortened by genetic modification (1). Shown are the complete deletion of the 20 amino acid region (1), and anchoring segments of 8, 12, 14, 16, and the native 20 amino acids. For successful anchoring in cytoplasmic and plasma membranes 12–14 amino acids are required, so it appears that an anchoring energy of about 20 kcal/mol is sufficient.

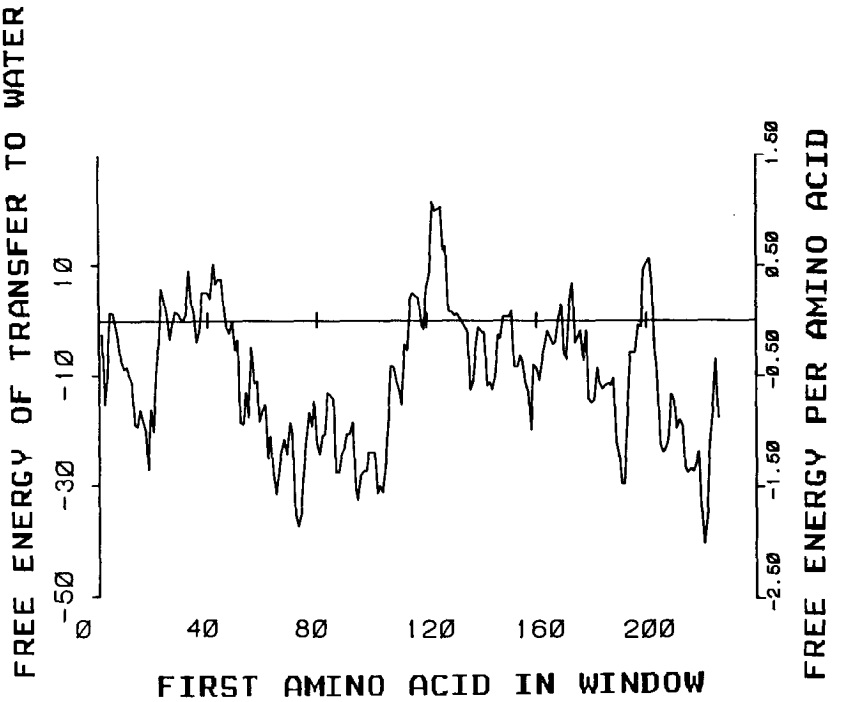


Figure 10 The free energy profile for trypsinogen (4) determined using a window of 20 amino acids and the GES scale. Note that a strongly hydrophobic peak exists near the significance level.

aqueous pores. In recent years there have been a number of efforts to examine the possible presence of amphiphilic helices that may provide polar pores (17, 24, 28). These have led, for example, to detailed models for the disposition of chains in the acetylcholine receptor (24, 26, 68). Some of the predictions have been criticized on statistical grounds (25). A major difficulty is how to distinguish whether a potential amphiphilic helix exists in the bilayer as part of a pore or in a soluble globular domain, since most helices in soluble proteins are amphiphilic (82). While these efforts may reveal additional aspects of membrane protein structure, we cannot test these aspects in the absence of well-established structural observations; consequently, our decision has been to set them aside until their veracity can be tested experimentally.

Segments of an amino acid sequence that form two closely spaced helical regions with a turn between them may not be readily identified as a helix pair by any of the procedures described above. If the region between the helices contains no amino acids of strikingly polar character, the turn may not be revealed. This does not mean that the ends of the helices are nonpolar or that the turn is unstable. It is well known that in the region in which a

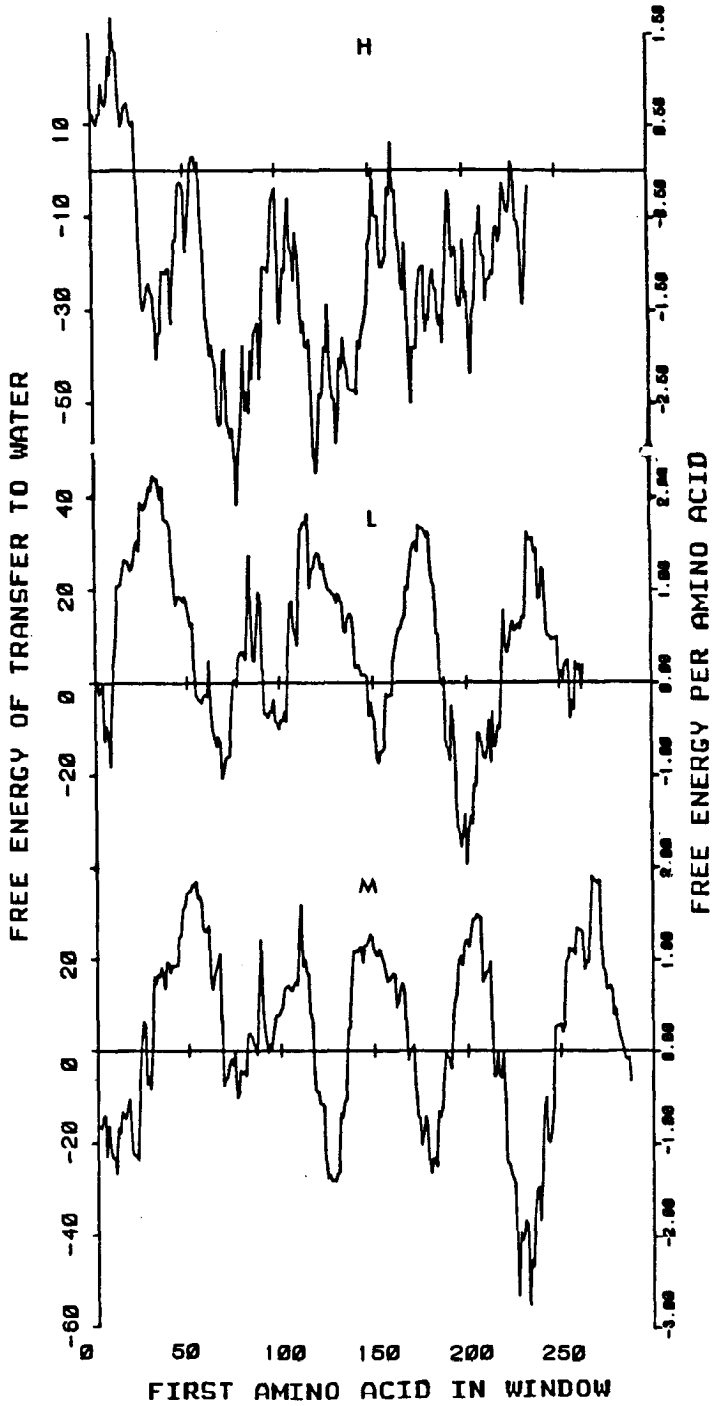
helix ends, a fractional charge exists as a consequence of charge separation in the peptide bonds aligned along the helix (34, 84, 96). Furthermore, the hydrogen bond donors or acceptors are not satisfied by backbone acceptors or donors. Not only does this render the end of the helix strongly polar [and suggest that helix ends will not be found in the interior of membrane bilayers (21, 31)] but the polarity produced in this way would not be revealed by the progressive analysis employed in scanning sequences for the polar characteristics of the amino acid side chains present. The structure of the inserted portion of the cytochrome b5 molecule is not known, but a hairpin of helices would not be excluded merely by the fact that two distinct peaks are not seen in the analysis shown in Figure 7.

TESTS OF PREDICTIONS USING KNOWN STRUCTURES

While there are many proteins, such as the red cell membrane glycophorin (92), for which the transmembrane structure is strongly implied by a range of data (e.g. 6), in only a few cases is helical structure established with a high level of confidence. The best-established examples are found in the structure of the photosynthetic reaction center of *Rhodopseudomonas viridis*, which has recently been determined at high resolution (14, 15). The macromolecular assembly consists of four polypeptide chains; two of these (L and M) are globular integral membrane proteins and one (H) is an anchored membrane protein. While the sequences of the *R. viridis* proteins are not yet published, they are highly homologous (61; H. Michel, personal communication) to the sequences of photosynthetic reaction centers from other organisms such as *R. capsulata* (104). The crystal structure shows a region in which bundles of helices traverse an apparently nonpolar region. Although the structure was crystallized in the presence of detergent and not in the presence of phospholipid, the distribution of polar and nonpolar groups suggests that a defined region containing a number of helices spans the membrane. Using the published sequences of the *R. capsulata* subunits it is therefore possible to construct a test of the prediction methods.

All of the putative membrane-spanning helices observed in the crystal structure are predicted from the hydrophobicity analyses of the sequences. Figure 11 shows the sequence analysis for the L, M, and H subunits made using the GES scale and a window of 20 amino acids. Four helices each are

Figure 11 Sequences for the H, L, and M subunits of the photosynthetic reaction center. Sequences of *R. capsulata* (104), the GES scale, and a window of 20 amino acids were used to examine the structure. One transmembrane helix is predicted for the H subunit, and five for both the L and M subunits.



suggested by broad maxima in both the L and M cases, and a fifth helix is suggested by a relatively sharp maximum between the first two broad maxima. In the L subunit there is an additional peak that is at the margin of significance and is located near the first sharp peak. It does not correspond to a transmembrane helix in the structure. The H-subunit profile shows a single broad maximum suggesting a single transmembrane helix. The maxima from the polarity profiles were used for the predictions shown in Table 3.

Also in Table 3 are the positions of helices in *R. viridis* established from the crystal structure of Deisenhofer et al (14, 15). The agreement is striking. Of the 220 amino acids assigned by the polarity profile to 11 helices, it appears that only 2 amino acids lie outside of the helices that are actually found in the protein structures. The observed helices are actually somewhat longer than the scanning window of 20 residues. This is not surprising, since the actual helices may have hydrophilic extensions beyond the region of the nonpolar lipid bilayer. As these sequences contain very nonpolar regions with few polar amino acids, helix predictions are relatively insensitive to the choice of scale used (see above). The agreement between the predicted and established transmembrane helix location is striking and is highly encouraging for those who wish to apply prediction methods to membrane protein sequences.

Table 3 Comparison of predicted and observed membrane spanning helices in photosynthetic reaction centers[a]

Subunit	Helix	Predicted	Observed
L	A	32–51	32–55
	B	84–103	84–112
	C	116–135	115–140
	D	175–194	170–199
	E	233–252	225–251
M	A	52–71	52–78
	B	111–130	110–139
	C	148–167	142–167
	D	206–225	197–225
	E	267–186	259–285
H	A	12–31	12–37

[a] Predictions are based on the energy plots shown in Figure 11. They are based on the amino acid sequences from *R. capsulata* (104), which are known to be highly homologous to those of *R. viridis* (61). The structures of the subunits are known at high resolution for *R. viridis* (14, 15) and the transmembrane helices are located as shown. It is to be expected that actual helices may be longer than those predicted on the basis of spanning the nonpolar region of the lipid bilayer.

Although it is less well established, the structure of bacteriorhodopsin also provides a test of predictive methods. The structure is known to contain seven transmembrane helices (32, 52), and use of the current GES polarity scale on the sequence showed seven nonpolar regions (19, 22, 88), suggesting the locations of such helices in the amino acid sequence (Figure 7). The first hydrophobicity analysis of the bacteriorhodopsin sequence was performed using the earlier version of the GES scale (88). This analysis prompted a revision in the proposed positions of helices F and G in the sequence from the original model (20). While the application of the initial scale suggested locations for all seven helices, the current scale more convincingly delineates the existence and positions of helices C and G. Subsequent experiments have been consistent with and thus support the use of the computer-generated model. While the exact sequence locations of the helices in the actual structure are not known, a number of recent chemical-modification and protease-digestion studies narrow the possible locations substantially (8, 16, 26, 37, 41, 56, 65, 97, 98). There remains some ambiguity in the precise location of the short loop connecting helices F and G, but the water-accessible portions of the rest of the sequence are well defined. The regions that are predicted and those that are defined by various modification and digestion studies are compared in Figure 12. As with the case of the photosynthetic reaction centers, the agreement is excellent.

CONCLUSIONS

We conclude that the prediction of regions of integral membrane protein polypeptide sequences that span the nonpolar region of the membrane as helical structures is sound and useful.

Where membrane proteins have extremely nonpolar regions as transbilayer elements, virtually any scale of polarity can reveal their presence. On the other hand, if polar groups are present, scales that take into account the details of helical structure and transfer from water to a lipid bilayer interior are more successful in revealing important possible helices. Of the scales developed, the GES scale is most appropriate (see Table 1).

Scanning procedures should employ a window of a length approximating the hydrophobic dimension of a lipid bilayer. While this is a variable, a value of 20 amino acids is a reasonable choice.

Using studies with altered protein sequences, it is possible to establish that a peak of about 20 kcal/mol on the GES scale with a window of 20 amino acids is a significant feature. However, it is important to note that not all polypeptide sequences that contain such features have membrane-spanning helices; some soluble proteins may include features of this kind,

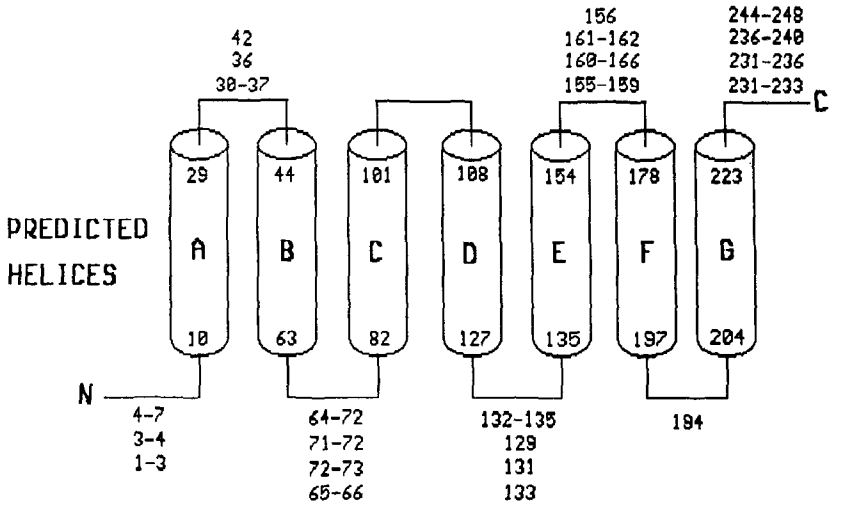


Figure 12 Predicted and observed features of bacteriorhodopsin topology. Seven helices are predicted on the basis of the GES scale (19, 22, 88; Figure 7). The predicted hydrophobic regions are indicated on the presumed transmembrane helices, indicated by the amino acid sequence numbers. Modifications using reagents active in the aqueous phase should reveal the regions between predicted helices. Such reagents include enzyme, lactoperoxidase-catalyzed iodination, and antibody binding. Experimentally observed modifications are shown either as spans of amino acids in the case of enzyme cleavages or as single amino acids in the cases of modification or antigenic identification (8, 16, 26, 37, 41, 56, 65, 97, 98). The prediction of helix B has two possible extremes; that which is preferred on the basis of experimental observation is shown. Some debate concerning the location of helix F continues, and recent antibody experiments suggest that the helix may be located a few amino acids toward the amino terminus from the location shown here. The loop between helix C and D is short, and may not be accessible to the reagents used.

and extramembrane domains of membrane proteins may also include such structures. Caution is therefore recommended in the absence of confirmatory evidence.

The most striking observation, however, is that all of the known transbilayer structural elements in helical membrane proteins are accurately predicted by the polarity analysis we discuss in this article. It appears that the prediction of some secondary structural elements of membrane proteins may be, in this sense, more successful than that of proteins in the aqueous milieu.

ACKNOWLEDGMENTS

We are indebted to a number of people who have helped us with the manuscript, particularly to N. Stewart for her able assistance in its production and to H. Michel and K. Dill for useful discussions. This work was supported by the National Institutes of Health Grant GM-22778.

NOTE

The program written by A. Goldman to examine sequences as we have discussed is available upon request.

Literature Cited

1. Adams, G. A., Rose, J. K. 1985. *Cell* 41: 1007
1a. Allen, L. C. 1975. *Proc. Natl. Acad. Sci. USA* 72: 1401
2. Argos, J. K., Rao, J. K. M., Hargrave, P. A. 1982. *Eur. J. Biochem.* 128: 565
3. Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., Sarma, V. R. 1965. *Nature* 206: 757
4. Bode, W., Schwager, P. 1975. *J. Mol. Biol.* 98: 693
5. Born, M. 1920. *Z. Phys.* 1: 45
6. Bretscher, M. S. 1971. *Nature New Biol.* 231: 229
7. Brown, J. E., Klee, W. A. 1971. *Biochemistry* 10: 470
8. Brunner, J., Franzusoff, A. J., Lüscher, B., Zugliani, C., Semenza, G. 1985. *Biochemistry* 24: 5422
9. Chothia, C. 1974. *Nature* 248: 338
10. Chothia, C. 1975. *Nature* 254: 304
11. Chothia, C. 1976. *J. Mol. Biol.* 105: 1
12. Chothia, C. 1985. *Ann. Rev. Biochem.* 53: 537
13. Davis, S. S., Higuchi, T., Rytting, J. H. 1974. *Adv. Pharm. Sci.* 4: 73
14. Deisenhofer, J., Epp, O., Miki, K., Huber, R., Michel, H. 1984. *J. Mol. Biol.* 180: 385
15. Deisenhofer, J., Epp, O., Miki, K., Huber, R., Michel, H. 1985. *Nature* In press
16. Dumont, M. E., Trewhella, J., Engelman, D. M., Richards, F. M. 1985. *J. Membr. Biol.* In press
17. Eisenberg, D., Weiss, R. M., Terwilliger, T. C. 1982. *Nature* 299: 371
18. Engel, A., Massalzki, A., Schindler, H., Dorset, D. L., Rosenbusch, J. 1985. *Nature* 317: 645
19. Engelman, D. M., Goldman, A., Steitz, T. 1982. *Methods Enzymol.* 88: 81
20. Engelman, D. M., Henderson, R., McLaughlin, A. D., Wallace, B. A. 1980. *Proc. Natl. Acad. Sci. USA* 77: 2023
21. Engelman, D. M., Steitz, T. A. 1981. *Cell* 23: 411
22. Engelman, D. M., Steitz, T. 1984. In *The Protein Folding Problem*, ed. D. Wetlauffer, p. 87. Boulder, Colo: Westview
23. Deleted in proof
24. Finer-Moore, J., Stroud, R. M. 1984. *Proc. Natl. Acad. Sci. USA* 81: 155
25. Flinta, C., Von Heijne, G., Johannson, J. 1983. *J. Mol. Biol.* 168: 193
26. Gerber, G. E., Gray, C. P., Wildenauer, D., Khorana, H. G. 1977. *Proc. Natl. Acad. Sci. USA* 74: 5426
27. Gray, T. M., Mathews, B. W. 1984. *J. Mol. Biol.* 175: 75
28. Guy, H. R. 1984. *Biophys. J.* 45: 249
29. Guy, H. R. 1985. *Biophys. J.* 47: 61
30. Hartley, B. S., Brown, J. R., Kauffman, D. L., Smillie, L. B. 1965. *Nature* 207: 1157
31. Henderson, R. 1979. *Soc. Gen. Physiol. Ser.* 33: 3
32. Henderson, R., Unwin, P. N. T. 1975. *Nature* 257: 28
33. Hine, J., Mookerjee, P. K. 1975. *J. Org. Chem.* 40: 292
34. Hol, W. G. J., van Duijnen, P. T., Berendsen, H. J. C. 1978. *Nature* 294: 532
35. Honig, B. H., Hubbell, W. L. 1984. *Proc. Natl. Acad. Sci. USA* 81: 5412
36. Honig, B. H., Hubbell, W. L., Flewelling, R. F. 1986. *Ann. Rev. Biophys. Biophys. Chem.* 15: 163
37. Huang, K. S., Bayley, H., Liao, M.-J.,

London, E., Khorana, H. G. 1981. *J. Biol. Chem.* 256:3802
38. Jahnig, E. 1983. *Proc. Natl. Acad. Sci. USA* 80:3691
39. Janin, J. 1979. *Nature* 277:491
40. Janin, J. 1979. *Bull. Inst. Pasteur* 77: 337
41. Katre, N. V., Finer-Moore, J., Hayward, S. B., Stroud, R. M. 1984. *Biophys. J.* 46:195
42. Katz, F. N., Rothman, J. E., Lingappa, V. R., Blobel, G., Lodish, H. F. 1977. *Proc. Natl. Acad. Sci. USA* 74:3278
43. Kautzmann, W. 1959. *Adv. Protein Chem.* 14:1
44. Kendrew, J. C. 1962. *Brookhaven Symp. Biol.* 15:216
45. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R. 1960. *Nature* 185:422
46. Khorana, H. G., Gerber, G. E., Herlihy, W. C., Gray, C. P., Anderegg, R. J., et al. 1979. *Proc. Natl. Acad. Sci. USA* 76: 5046
47. Kim, P. S., Baldwin, R. L. 1982. *Ann. Rev. Biochem.* 51:459
48. Klein, R., Moore, M., Smith, M. 1971. *Biochim. Biophys. Acta* 233:420
49. Krigbaum, W. R., Komoriya, A. 1979. *Biochim. Biophys. Acta.* 576:204
50. Kyte, J., Doolittle, R. F. 1982. *J. Mol. Biol.* 157:105
51. Lee, B., Richards, F. M. 1971. *J. Mol. Biol.* 55:379
52. Leifer, D., Henderson, R. 1983. *J. Mol. Biol.* 163:451
53. Levitt, M. 1976. *J. Mol. Biol.* 104:59
54. Lewis, B. A., Engelman, D. M. 1983. *J. Mol. Biol.* 166:203
55. Lewis, B. A., Engelman, D. M. 1983. *J. Mol. Biol.* 166:211
56. Liao, M. J., Huang, K.-S., Khorana, H. G. 1984. *J. Biol. Chem.* 259:4194
57. Manovalan, P., Pomuswamy, P. K. 1978. *Nature* 275:673
58. Mao, D., Wallace, B. A. 1984. *Biochemistry* 23:2667
59. Matthew, J. B. 1985. *Ann. Rev. Biophys. Biophys. Chem.* 14:387
60. Meirovitch, H., Rackovsky, S., Scheraga, H. 1980. *Macromolecules* 13: 1398
61. Michel, H., Weyer, K. A., Gruenberg, H., Oesterhelt, D., Lottspeich, F. 1985. *EMBO J.* 4:1667
62. Nishikawa, K., Ogi, T. 1980. *Int. J. Pept. Protein Res.* 16:19
63. Nozaki, Y., Tanford, C. 1971. *J. Mol. Chem.* 246:2211
64. Ovchinnikov, Y., Adbulaev, N., Feigira, M., Kiselev, A., Lobonov, N. 1979. *FEBS Lett.* 100:219
65. Ovchinnikov, Y. A., Abdulaev, N. G., Vasilov, R. G., Vturina, I. Y., Kuryatov, A. B., Kiselev, A. V. 1985. *FEBS Lett.* 179:343
65a. Page, M. I., Jencks, W. P. 1971. *Proc. Natl. Acad. Sci. USA* 68:1678
66. Parsegian, V. A., Fuller, N., Rand, R. P. 1979. *Proc. Natl. Acad. Sci. USA* 76: 2750
67. Perutz, M. F., Muirhead, H., Cox, J. M., Goaman, L. C. G. 1968. *Nature* 219: 131
68. Popot, J.-L., Changeux, J.-P. 1984. *Physiol. Rev.* 64:1162
69. Privalov, P. L. 1979. *Adv. Protein Chem.* 33:167
70. Redman, C. M., Sabatini, D. D. 1966. *Proc. Natl. Acad. Sci. USA* 56:608
71. Reynolds, J. A., Gilbert, D. B., Tanford, C. 1974. *Proc. Natl. Acad. Sci. USA* 71: 2925
72. Richards, F. M. 1977. *Ann. Rev. Biophys. Bioeng.* 6:151
73. Richmond, T., Richards, F. 1978. *J. Mol. Biol.* 119:537
74. Robson, B., Osguthorpe, D. J. 1979. *J. Mol. Biol.* 132:19
75. Roe, J. K., Welch, W. J., Sefton, B. M., Esch, F. S., Ling, N. C. 1980. *Proc. Natl. Acad. Sci. USA* 77:3884
76. Rose, G. D. 1978. *Nature* 272:586
77. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., Zehfus, M. H. 1985. *Science* 229:834
78. Rose, G. D., Gierasch, L. M., Smith, J. A. 1985. *Adv. Protein Chem.* 37:1
79. Rose, G. D., Roy, S. 1980. *Proc. Natl. Acad. Sci. USA* 77:4643
80. Rosenbusch, J. P. 1974. *J. Biol. Chem.* 249:8019
81. Rothman, J., Lodish, H. 1977. *Nature* 269:775
82. Schiffer, M., Edmundson, A. B. 1967. *Biophys. J.* 7:121
83. Schindler, M., Rosenbusch, J. P. 1984. *FEBS Lett.* 173:85
84. Sheridan, R. P., Levy, R. M., Salemme, F. R. 1982. *Proc. Natl. Acad. Sci. USA* 79:4545
85. Singer, S. J. 1962. *Adv. Protein Chem.* 17:1
86. Singer, S. J. 1971. *Structure and Function of Biological Membranes.* New York: Academic. 145 pp.
87. Spatz, L., Strittmatter, P. 1971. *Proc. Natl. Acad. Sci. USA* 68:1042
88. Steitz, T., Goldman, A., Engelman, D. 1982. *Biophys. J.* 37:124
89. Sweet, R. M., Eisenberg, D. 1983. *J. Mol. Biol.* 171:479
90. Tanford, C. 1962. *J. Am. Chem. Soc.* 84: 4240
91. Tanford, C. 1980. *The Hydrophobic Effect.* New York: Wiley

92. Tomita, M., Furthmayr, H., Marchesi, V. T. 1978. *Biochemistry* 17:4756
93. Von Heijne, G. 1981. *Eur. J. Biochem.* 116:419
94. Von Heijne, G. 1981. *Eur. J. Biochem.* 120:275
95. Von Heijne, G., Blomberg, C. 1979. *Eur. J. Biochem.* 97:175
96. Wada, A. 1976. *Adv. Biophys.* 9:1
97. Walker, J. E., Carne, A. F., Schmitt, H. W. 1979. *Nature* 278:655
98. Wallace, B. A., Henderson, R. 1975. *Biophys. J.* 39:233
99. Wertz, D. H., Scheraga, H. A. 1978. *Macromolecules* 11:9
100. Wolfenden, R. 1983. *Science* 222:1087
101. Wolfenden, R., Andersson, L., Cullis, P. M., Southgate, C. C. B. 1981. *Biochemistry* 20:849
102. Wolfenden, R., Cullis, P. M., Southgate, C. C. F. 1979. *Science* 206:575
103. Wyckoff, H. W., Hardman, K. D., Allewell, N. M., Inagami, T., Johnson, L. N., Richards, F. M. 1967. *J. Biol. Chem.* 242:3984
104. Youvan, D. C., Bylina, E. J., Alberti, M., Begusch, H., Hearst, J. E. 1984. *Cell* 37:949
105. Yunger, L. M., Cramer, R. D. 1981. *Mol. Pharmacol.* 20:602

Annual Review of Biophysics and Biophysical Chemistry
Volume 15, 1986

CONTENTS