J. Membrane Biol. 115, 145–158 (1990)

The Journal of
Membrane Biology
© Springer-Verlag New York Inc. 1990

Observations Concerning Topology and Locations of Helix Ends of Membrane Proteins of Known Structure

Stephen H. White and Russell E. Jacobs
Department of Physiology & Biophysics, University of California, Irvine, Irvine, California 92717

Summary. Hydropathy plots of amino acid sequences reveal the approximate locations of the transbilayer helices of membrane proteins of known structure and are thus used to predict the helices of proteins of unknown structure. Because the three-dimensional structures of membrane proteins are difficult to obtain, it is important to be able to extract as much information as possible from hydropathy plots. We describe an "augmented" hydropathy plot analysis of the three membrane proteins of known structure, which should be useful for the systematic examination and comparison of membrane proteins of unknown structure. The sliding-window analysis utilizes the floating interfacial hydrophobicity scale [IFH(h)] of Jacobs and White (Jacobs, R.E., White, S.H., 1989. *Biochemistry* **28**:3421–3437) and the reverse-turn (RT) frequencies of Levitt (Levitt, M., 1977, *Biochemistry* **17**:4277–4285). The IFH(h) scale allows one to examine the consequences of different assumptions about the average hydrogen bond status (h = 0 to 1) of polar side chains. Hydrophobicity plots of the three proteins show that (i) the intracellular helix-connecting links and chain ends can be distinguished from the extracellular ones and (ii) the main peaks of hydrophobicity are bounded by minor ones which bracket the helix ends. RT frequency plots show that (iii) the centers of helices are usually very close to wide-window minima of average RT frequency and (iv) helices are always bounded by narrow-window maxima of average RT frequency. The analysis suggests that side-chain hydrogen bonding with membrane components during folding may play a key role in insertion.

Key Words membrane protein structure · bacteriorhodopsin · photosynthetic reaction centers · hydrophobicity analysis · transbilyer helix prediction · protein assembly in membranes

Introduction

The accurate prediction of transmembrane protein topology and of helix ends is an important first step for the prediction of membrane protein conformation. The most common approach is to use hydropathy plots (Kyte & Doolittle, 1982) to identify the general locations of transbilayer helices and biochemical analyses to identify residues accessible from the aqueous phases, which must generally constitute helix-connecting links external to the bilayer (*see* review by Engelman et al., 1986). The hydropathy plot method is derived from fundamental thermodynamic considerations which suggest that the helix is likely to be the most common transmembrane conformation of hydrophobic peptide sequences because of the high energetic cost of burying unbonded backbone >NH and >C = 0 groups in a nonpolar phase (Henderson, 1975; Von Heijne & Blomberg, 1979; Engelman & Steitz, 1981; Jähnig, 1983; Jacobs & White, 1989). The predictive ability of polarity plots was tested by Michel et al. (1986) for the light and medium chains of the photosynthetic reaction center (PSRC) of *Rhodospseudomonas viridis* whose high-resolution structure had been determined by Deisenhofer et al. (1985). They found that the most nonpolar domains of the peptide chains corresponded approximately with the observed positions of the transbilayer helices, but that the helix ends were not accurately predicted because of their high polarity which makes them indistinguishable from polar helix-connecting links. Except for this problem, the hydropathy plot method is successful for the one (but only one) class of membrane proteins for which a high resolution structure exists. The use of hydrophobicity plots is now obligatory whenever a membrane protein is sequenced, and many remarkably detailed molecular models of membrane proteins, particularly channels and pumps, have been proposed as a result.

A number of problems remain unresolved regarding the interpretation of hydropathy plots (Lodish, 1988) and their reconciliation with immunological and chemical studies of the peptide sequences exposed to the aqueous phases (McCrea, Engleman & Popot, 1988). Of particular concern is the possibility that very large membrane proteins such as the Na channel or the ACh receptor may have transbilayer peptide conformations that do not adhere rigorously to the motif of bundles of 20AA α-helices. Lodish (1988) suggested, for example, that ion

channels could easily be lined with β-strands or short helices that are relatively polar and which would elude detection by hydropathy plots as currently implemented. McCrea et al. (1988), however, suggested that protein channels may not need to be as polar as assumed in many channel protein models, which is consistent with studies of model peptides (Lear, Wasserman & DeGrado, 1988). Still, it would be surprising to find that the α-helix was the only transmembrane structural element for the very large membrane proteins, considering the rich structural motifs of globular proteins (Richardson, 1981). Wallace, Cascio and Mielke (1986), however, have presented evidence showing that secondary structure prediction schemes based upon soluble proteins are likely to be inappropriate for membrane proteins.

Despite these uncertainties, the α-helical bundle is likely to be the predominant structural scheme of the smaller membrane proteins with 300 or fewer amino acids. It is therefore important to have a systematic hydropathy plot method for analyzing proteins of this type and for seeking anomalies in larger membrane proteins which might be indicative of other structural themes. We present one such method in this paper which focuses on the topology of membrane proteins and the accurate prediction of the ends of transbilayer helices. Edelman and White (1989) considered the possibility that the accurate determination of the full length of transbilayer helices by polarity plots might be limited by nonoptimal amino acid weights (i.e., hydrophobicity scale) and by the use of sliding windows of nonoptimal shape. They used linear optimization methods to derive optimal amino acid weights and window shapes for the PSRC. The resulting so-called linear convolutional recognizer (LCR) improved prediction accuracy compared to standard hydropathy plots but was still inexact. Edelman and White (1989) also examined the general limitations of LCRs and concluded that no simple linear scheme based upon a single set of weights was likely to be completely successful.

One method of circumventing some of the above problems, which retains the inherent simplicity of the hydrophobicity plot, is to combine sliding-window averages of several amino acid parameters. Vogel and Jähnig (1986), for example, have used plots of polarity and Chou and Fasman (1974, 1978) P_α and P_β probabilities to predict the structure of the outer-membrane proteins of *Escherichia coli* while Paul and Rosenbusch (1985) used reverse-turn preferences for predicting the folding pattern of bacteriorhodopsin. Blanck and Oesterhelt (1987) have augmented polarity plots with so-called acrophilicity plots (Hopp, 1985) to make structural predictions about halorhodopsin. Acrophilicity values are derived from the frequency of appearance of amino acids on the outer surfaces of proteins and thus must correlate negatively with the buried hydrophobic segments of proteins and positively with segments forming reverse turns. This is the mirror image of the approach of Rose (1978) and Rose and Roy (1980) who correlated the appearance of amino acids in chain turns at the protein surface with minima in hydrophobicity.

We present here an analysis of the amino acid sequences of the PSRCs of *Rps. viridis* and *Rhodobacter sphaeroides* whose structures are known to high resolution (Deisenhofer et al., 1985; Allen et al., 1987) and of bacteriorhodopsin (BR) whose structure is known to low resolution (Henderson & Unwin, 1975; Leifer & Henderson, 1983). The analysis involves hydropathy plots using the so-called floating interfacial hydrophobicity [IFH(h)] scale of Jacobs and White (1989) and sliding-window averages of the amino acid reverse-turn (RT) frequencies of Levitt (1977). The IFH(h) scale uses the bilayer interface as the reference phase and takes into account the variable hydrogen bonding possibilities of polar side chains by means of the h parameter ($h = 0$ implies no hydrogen bonding and $h = 1$ maximum hydrogen bonding). Hydrophobicity plots of the three proteins show that (i) the intracellular chain ends and helix-connecting links can be distinguished from the extracellular ones and (ii) the main peaks of hydrophobicity are bounded by minor ones which bracket the helix ends. Reverse-turn (RT) frequency plots show that (iii) the centers of helices are usually very close to wide-window minima of average RT frequency and (iv) helix ends are always bounded by narrow-window maxima of average RT frequency. Changes in the appearance of IFH(h) hydrophobicity plots as the hydrogen bond parameter h is varied reveal the topology of the proteins and suggests that changes in side-chain hydrogen bonding patterns during folding may be crucial to insertion.

The analytical method described may be generally useful as a method for the systematic analysis of membrane proteins of unknown three-dimensional structure using PSRC and BR as references. We cannot, of course, make strong claims about the general accuracy of the method because of the limited size of the data base. A broader usefulness for the method is suggested by the conclusion that PSRC and BR are likely to represent two distinctly different classes of membrane proteins (Michel et al., 1985). PSRC consists of three subunits (L, M, and H), which form a structure with a nonpolar

interior. BR, on the other hand, is a functional monomer and is likely to have a very polar interior (Engelman & Zaccai, 1980).

Materials and Methods

Protein Structures

Photosynthetic Reaction Centers[1]

The analysis was performed using the known high-resolution structures of the light (*L*), medium (*M*), and heavy (*H*) subunits of the photosynthetic reaction centers (PSRC) of *Rps. viridis* (Deisenhofer et al., 1985) and *Rb. sphaeroides* (Allen et al., 1987). The helix ends designated in the present paper as "observed" are those published by Deisenhofer et al. (1985) and Allen et al. (1987). The amino acid sequences used were those determined from nucleotide sequences by Williams et al. (1983, 1984, 1986) for *Rb. sphaeroides* and Michel et al. (1985, 1986) for *Rps. viridis*.

Bacteriorhodopsin

This is the only other membrane protein for which significant structural information is available, albeit at low resolution (7 Å). The amino acid sequence of BR has been determined by Ovchinnikov et al. (1979) using peptide sequencing and by Khorana et al. (1979) using nucleotide sequencing. The latter is more accurate and we use it in the analysis. In addition, we include the 13AA leader sequence in the analysis (Seehra & Khorana, 1984) and number the amino acids in the complete sequence from −12 to 248. We note that the PSRC subunits do not have leader sequences.

While there has been considerable discussion about the secondary structure of BR (Jap et al., 1983; Jap & Kong, 1986), we believe that the preponderance of evidence from spectroscopy (Vogel & Gärtner, 1986) and from electron, x-ray, and neutron diffraction studies support a structure consisting of seven transbilayer helices (Henderson & Unwin, 1975; Hayward & Stroud, 1981; Agard & Stroud, 1982; Wallace & Henderson, 1982; Leifer & Henderson, 1983; Popot, Trewhella & Engelman, 1986; Trewhella et al., 1986; Baldwin et al., 1988). The location of the retinal projected onto the bilayer plane is known from neutron diffraction (King et al., 1980; Jubb et al., 1984; Seiff et al., 1985, 1986*a*,*b*; Heyn et al., 1988).

The locations of the helix-connecting links are not known with complete certainty, although several attempts have been made to obtain this information by structural methods (Agard & Stroud, 1982; Wallace & Henderson, 1982; Katre et al., 1984; Jaffe & Glaeser, 1987). There have, however, been extensive studies of this question using combinations of low resolution diffraction data, hydropathy plots, and biochemical methods to assess connecting link accessibility, which give a reasonably consistent assignment of the helices. For the purposes of our analysis, we use helix assignments representing consensus values derived from the studies of Ovchinnikov et al. (1979, 1985), Engelman et al. (1980), Agard and Stroud (1982), Engelman, Steitz and Goldman (1986), and Khorana (1988). We averaged the helix end assignments to arrive at the following consensus locations for the helices (Table 1): A, 8-32; B, 41-64; C, 76-101; D, 106-130; E, 134-156; F, 169-194; G, 200-226. The mean of the standard deviations among all the ends is 2.5, ranging from 0.75 to 6.0. We note that Rosenbusch (1985) and Fimmel et al. (1989) dispute some of the assignments.

[1] *Abbreviations:* BR, bacteriorhodopsin; HR, halorhodopsin; PSRC, photosynthetic reaction center. IFH, interfacial hydrophobicity; RT, reverse-turn; *Rps.*, *Rhodospseudomonas*; *Rb.*, *Rhodobacter*.

Table 1. Summary of helical assignments for BR from various laboratories[a]

Helix letter	This work	Average (Refs. b–g ±SD)	References					
			b	c	d	e	f	g
A [25]								
NT:	8	8 ± 2	7	10	7	4	9	8
CT:	32	32 ± 2	31	29	34	32	35	31
B [25]								
NT:	37	41 ± 2	41	44	39	37	41	42
CT:	61	64 ± 1	65	63	64	65	64	64
C [29]								
NT:	74	76 ± 4	77	82	73	73	73	79
CT:	102	101 ± 1	101	101	100	99	102	101
D [24]								
NT:	106	106 ± 1	107	108	106	105	105	107
CT:	129	130 ± 2	130	127	130	131	131	130
E [22]								
NT:	133	134 ± 1	133	135	134	134	134	133
CT:	154	156 ± 3	156	154	159	160	156	153
F [24]								
NT:	168	169 ± 6	166	178	164	167	164	175
CT:	191	194 ± 3	190	197	191	194	193	198
G [28]								
NT:	198	200 ± 4	197	204	196	199	205	202
CT:	226	226 ± 3	223	223	224	230	230	225

[a] The number in brackets following the helix letter is the length based upon the predicted helix ends in this paper.
[b] Engelman et al., 1980; [c] Engelman et al., 1986; [d] Agard & Stroud, 1982; [e] Ovchinnikov et al., 1979; [f] Ovchinnikov et al., 1985; [g] Khorana, 1988.

Analysis

The basic analysis, described in detail below, consists of sliding-window averages of hydrophobicity and reverse-turn frequency using broad windows to examine gross behavior and narrow windows to examine finer details. The locations of various maxima and minima in the plots of averaged amino acid weights *vs.* amino acid position are correlated with the known positions of helices in PSRC and the consensus positions for BR.

Hydrophobicity Scales

We use the "floating" interfacial hydrophobicity [IFH(*h*)] scale of Jacobs and White (1989), which is based upon a thermody-

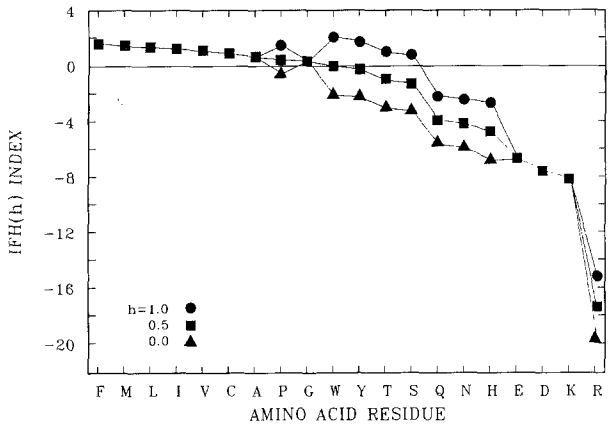

Fig. 1. Plot of the floating interfacial hydrophobicity scale, IFH(h). The parameter h specifies the fraction of designated side group hydrogen bonds made. (*See* text and Jacobs & White, 1989)

namic analysis of helix insertion using the bilayer interface as the reference phase. The free energy of transfer for an extended chain at the bilayer surface into the interior as a helix is designated $\Delta G_{hlx}(h)$. The assumption is made that all backbone hydrogen bonds are satisfied (except for Pro) and that a fraction h of designated side-chain hydrogen bonds are satisfied in some unspecified manner. This means that the probability of insertion of certain residues (particularly Ser, Thr, Pro, Tyr, and Trp) will be strongly determined by their hydrogen bond status. Ser and Thr, for example, can easily hydrogen bond to the main chain of helices. In order that positive values in hydropathy plots correspond to favorable free energies of transfer, the floating interfacial hydrophobicity index is defined as IFH(h) = $-\Delta G_{hlx}(h)$. In this paper, the plots are made with h = 0, 0.5, and 1.0, corresponding to zero, 50%, and 100% hydrogen bonding. The hydrophobicity scales with these three values of h are shown graphically in Fig. 1. We take h = 0.5 as the best mean scale because Chothia (1975) has determined that about 50% of the possible side chain hydrogen bonds of globular proteins are satisified.

Reverse-Turn Frequency

We examined the usefulness of several secondary structure frequency parameters for helix prediction such as the α-helix (P_α) and β-sheet (P_β) probabilities of Chou and Fasman (1974, 1978). However, we found that the results were about the same using either the α-helix or β-sheet preferences. The most conspicuous and consistent features of sliding-window averages of P_α and P_β were minima that occurred near the ends of helices. These minima result from a preponderance of secondary structure-breaking residues, which are the same residues that favor reverse-turn (RT) formation. We therefore examined sliding-window averages of reverse-turn frequency more carefully. The secondary-structure frequency parameters of Levitt (1977) were found to be more useful than those of Chou and Fasman (1974, 1978). Levitt's analysis, based upon more than 50 proteins, categorizes each of the residues in such a way that the residues at the top of a particular secondary structure preference scale do not include amino acids assigned to the top ranks (formers) of the other secondary structure scales. The RT frequency scale includes in the first rank (formers) Pro, Gly, Asp, Ser, and Asn and in the second rank (indifferent) Tyr, Thr, Glu, Gln, and Lys. Various combinations of these same residues are the sole constituents of the lowest ranks (breakers) for both helix and sheet formation.

Choices of Window Length

The characteristics of the sliding windows at various stages of the analysis is a key issue. Broad windows used early in the analysis will produce relatively smooth curves, whereas the narrow ones used later will produce seemingly "noisy" ones. However, one must distinguish between fluctuations due to the lack of a continuum in the physical properties of the amino acid residues and fluctuations that may in fact be indicative of important "signals." If one is ultimately going to select, for example, a helix end to within, say, three amino acids, then the window must be narrow enough to have meaning at this level.

It is common in polarity plots of membrane proteins to use windows whose lengths (20AA) correspond to the number of amino acids which can span the 30 Å thick hydrocarbon region of the bilayer as a helix. For reasons of symmetry, we use a 19AA window for the wide-window scans and note that a single point on the polarity plot represents the average of 19 residues. Thus, in principle, the minimum requirement for the identification of a possible transbilayer helix is that a single point on the hydropathy plots be above the selected reference level [taken as zero for the IFH(h) scales]. A problem with wide windows is the difficulty of "separating" helices with very short connecting links as in BR where there are seven transbilayer helices in about the same span of amino acids that in the L and M subunits contain only five helices. It is useful in such cases to use a narrower window to improve separation, and for BR we use an 11AA window for this purpose. As will be shown, the IFH(h) scales provide additional visual clues for resolving helices and the RT averages additional ones.

We chose 5AA as the finest window because a single turn of a helix or a β-bend occurs within this number. With a window of this width, the fluctuations due to the lack of a continuum in amino acid characteristics become apparent, and as a result it is somewhat difficult to identify significant features. We therefore smooth the resulting set of averages by averaging them with a 3AA window. The end result of this is a window of 7AA which weights each of the seven residues in the proportions 1:2:3:3:3:2:1. That is, a "trapezoidal" window is produced which has a base length of 7AA and a weighting-height of 3AA. The effective width of the window, taken as its full width at half maximum, remains 5AA. An effective window width of 5AA limits the precision of the selection of a single amino acid as an end to about 2.5 residues or about one half of a helix turn. Multiple scans are nothing more than the convolution of rectangles with one another. Thus, two 3AA window scans would produce a triangular window; a large number of repeats with windows of constant length would approximate a gaussian window. *See* Jansson (1984) for a discussion of convolution. Edelman and White (1989) have considered in detail optimal window shapes for linear convolutional analysis. Another approach to smoothing is Fourier transformation (Britton & Green, 1985). We chose not to use this approach because we wanted to be able to carry out the analysis using microcomputer spreadsheet programs.

Scan Terminology

A standard set of abbreviations for the various types of scans (sliding-window averages) will be used throughout the paper.

Scans of average polarity/hydrophobicity will be named after the IFH(h) scale of Jacobs and White (1989) which is used throughout this paper. Scans with effective window lengths of 19, 11, and 5 amino acids will be designated IFH_{19}, IFH_{11}, and IFH_5. Scans of average reverse-turn frequency factors [the P_{ij} found in Table V of Levitt (1977)] will be called $TURN_{19}$ and $TURN_5$.

Error Analysis

The basic method of analyzing the results of the scans is to compare an identified amino acid (candidate) position of a particular scan feature (e.g. maximum in $TURN_5$) with the position of a helix structural feature (e.g. helix end). We are particularly concerned here with the locations of helix ends and centers. Regarding the location of ends, we define "end zones" within which helix ends have a high probability of being located. In all cases the amino acid sequence is read from the N-terminus toward the C-terminus. Let the N-end and C-end of an observed helix be designated N_O and C_O, respectively, and amino acids selected as end-candidates be designated N_C and C_C. The errors in the selections will be $\Delta N = N_C - N_O$ and $\Delta C = C_C - C_O$. Absolute values are *not* used so that the signs of ΔN and ΔC will indicate if the candidate occurred before (minus sign) or after (plus sign) the correct position. Let the first and last members of the end-zone ranges, again reading from N toward C, be N_1 and N_2 and C_1 and C_2 for the N-end and C-end of a helix, respectively. An error in an end-zone range is considered to have occurred if $N_O > N_2$ or $N_O < N_1$ or if $C_O > C_2$ or $C_O < C_1$. The known helix centers will be designated HC_K and candidate centers HC_C so that the error in center selection will be $HC_K - HC_C = \Delta HC$.

Computations

All of the scans reported in this paper were implemented on a microcomputer (12 MHz 80286 processor and 10 MHz math coprocessor) using the 1-2-3 Spreadsheet Program (v. 2.01) of Lotus Development Corp. (Cambridge, MA). The general spreadsheet method used is that of Vickery (1987) because it is easy to implement without extensive programming. We will provide a copy of the spreadsheet for BR to interested readers if a diskette is sent to S.H.W.

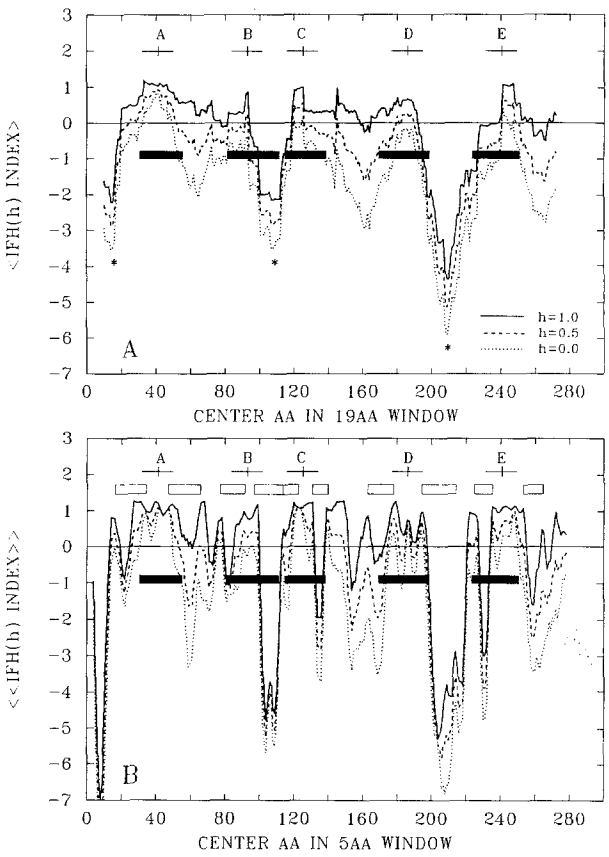

Fig. 2. Sliding-window averages of IFH(h) for the L subunit of *Rb. Sphaeroides*. The known helix locations are indicated by the solid bars and the helix-connecting links located on the cytoplasmic side by *. The lettered crosses indicate the positions of the peak hydrophobicity of each main peak; the horizontal length equals 19AA. The IFH(h) hydrophobicity scale of Fig. 1 is used with the three values of h indicated. $\langle\langle IFH(h)\rangle\rangle$ means that the initial 5AA scan was smoothed by a second scan with a 3AA window to smooth the curve. The resulting window is trapezoidal with a nominal width of 5AA (*see* text). (*A*) Hydropathy plot using 19AA window. (*B*) Plot using the nominal 5AA window. The open bars designate the end zones within which the helix ends are located (*see* text)

Results

Photosynthetic Reaction Centers

The general behavior of the L subunit of *Rb. sphaeroides* is fairly typical of all the subunits and various scans of it are shown in Figs. 2 and 3 for the purpose of illustrating the analytical method. We do not show scans from the other subunits in the interest of conserving space. All scans have superimposed on them the locations of the observed helices (heavy horizontal bars).

The $IFH(h)_{19}$ and $IFH(h)_5$ hydrophobicity plots are shown in Fig. 2*A* and *B*, respectively, for $h = 0$, 0.5, and 1. The five transbilayer helices are easily identified in Fig. 2*A* as the broad peaks with significant regions where the average IFH(0.5) index is greater than zero. The five maxima of hydrophobicity determined from $IFH(0.5)_{19}$ are shown at the tops of the figures by cross-marks whose widths are 19AA wide. The 19 most nonpolar residues (h = 0.5) of each transmembrane helix for each subunit of each PSRC are shown in Tables 2 and 3. These regions and the locations of the maxima for all of the subunits have been listed in Tables 2 and 3 as "IFH_{19} max. ± 9AA" and "IFH_{19} max," respectively. The selected 19 amino acids are always within the known helical domains and the maxima are located on average near the correct helix centers ($\Delta HC = +0.2 \pm 4.8$).

Table 2. Summary of analysis of the primary sequences of the *L*, *M*, and *H* subunits of *Rb. sphaeroides*[a]

Helix letter	Obs. Ends[b]	$TURN_{19}$ min. point	IFH_{19} max. point	IFH_{19} max. ±9AA	IFH_5 end zones	$TURN_5$ max. point
Rb. Sphaeroides			*L* subunit			
A [24]						
NT:	32	—	—	31	16– 34	29
HC:	43	46	40	—	—	—
CT:	55	—	—	49	47– 68	59
B [29]						
NT:	83	—	—	84	78– 91	82
HC:	97	97	93	—	—	—
CT:	111	—	—	102	96–114	115
C [23]						
NT:	116	—	—	116	114–122	115
HC:	127	130	125	—	—	—
CT:	138	—	—	134	131–139	146
D [28]						
NT:	171	—	—	177	159–178	170
HC:	184	181	186	—	—	—
CT:	198	—	—	195	194–209	201
E [26]						
NT:	225	—	—	232	226–235	224
HC:	237	240	241	—	—	—
CT:	250	—	—	250	248–278	252
			M subunit			
A [25]						
NT:	54	—	—	56	40– 58	54
HC:	66	66	65	—	—	—
CT:	78	—	—	74	67– 83	80
B [31]						
NT:	109	—	—	112	105–123	111
HC:	124	123	121	—	—	—
CT:	139	—	—	131	123–140	142
C [22]						
NT:	147	—	—	145	140–150	142
HC:	157	154	154	—	—	—
CT:	168	—	—	163	160–168	171
D [27]						
NT:	200	—	—	206	191–208	200
HC:	213	216	215	—	—	—
CT:	226	—	—	224	224–237	230
E [25]						
NT:	262	—	—	269	257–273	262
HC:	274	269	278	—	—	—
CT:	286	—	—	287	283–296	288
			H subunit			
A [26]						
NT:	12	—	—	12	4– 15	9
HC:	24	22	21	—	—	—
CT:	37	—	—	31	23– 41	40

[a] *Abbreviations:* IFH = average interfacial hydrophobicity index IFH(*h*) per residue for $h = 0.5$ (Jacobs & White, 1989; TURN = average reverse-turn probability per residue (Levitt, 1977; NT = helix N-terminus; CT = helix, C-terminus; HC = helix center. The subscripts refer to the width of the averaging window. For 5AA windows, a second average of 3AA was performed to smooth the curve; this is approximately equivalent to using a trapezoidal window for the scan. The number in brackets next to the helix letter is the helix length. *See* text.

[b] *Reference:* Allen et al., 1987.

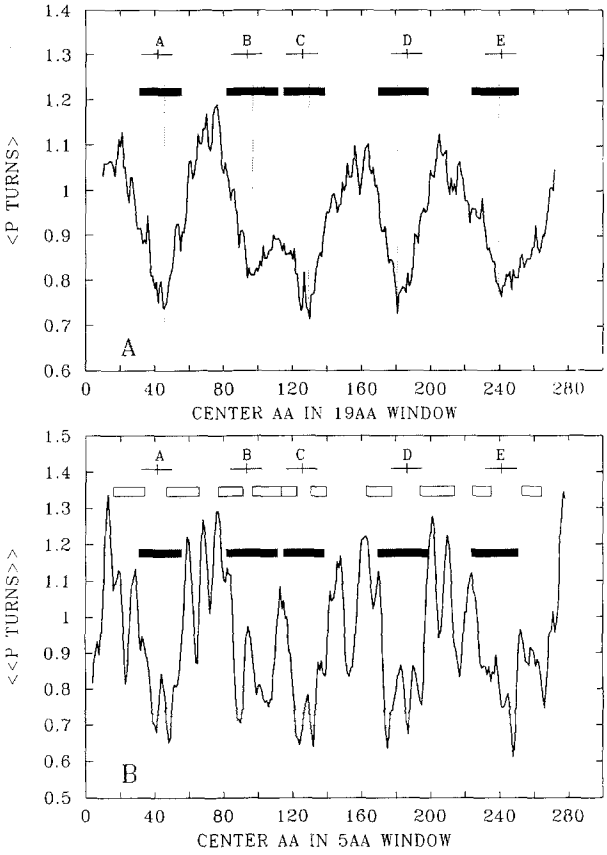

Fig. 3. Sliding-window averages of reverse-turn frequency for the *L* subunit of *Rb. Sphaeroides*. The reverse-turn frequencies are those of Levitt (1977). The meanings of the various symbols are described in the legend of Fig. 2. (*A*) Average with 19AA window. Note that there is a one-to-one correspondence between the major minima, the IFH(*h*) maxima and the known helices. The minima are close the true helix centers. (*B*) Average with a nominal 5AA window. Note that each helix end is bounded by maximum that generally falls within the end zones (open boxes). (*See* text)

We noted elsewhere that the relative values of the IFH(*h*) minima between the helices reflect the topology of the subunits (Jacobs & White, 1989). That is, the minima for the connecting-link and end regions that cross the membrane to the periplasmic surface are not as deep as those that remain on the cytoplasmic surface (indicated by * in Fig. 2*A*). Another very striking feature of these regions is the variations in the plots relative to the zero level as *h* varies from 0 to 1. It appears that the regions that

cross the membranes are relatively more sensitive to the hydrogen bond parameter h than those that do not. These general features can be characterized quantitatively by determining the average fractional change in $IFH(h)_{19}$ as h varies from 0 to 1 for regions outside the helices for which $IFH(1) < 0$. Plots of the fractional change and their averages for the *Rb. Sphaeroides* L subunit are shown in Fig. 5*A*. The fractional change is always less than 0.5 for the cytoplasmic regions. To the extent that h accurately reflects the hydrogen bond-forming potential of the side chains, this suggests the possibility that changes in side-chain hydrogen bonds during folding may be important for helix insertion.

Close examination of the $IFH(0.5)_5$ plots if Fig. 2*B* reveals another important observation. The general pattern for each helix is a more or less broad central peak bounded by narrower secondary peaks; each central peak generally shows several low amplitude oscillations with widths approximately equal to those of the secondary peaks. Comparing these features with the known helix locations reveals that the helix ends are invariably located between the peak of a weak low oscillation on the edge of the main peak and the maximum of the adjacent secondary peak. The region between the central low oscillation maxima and secondary maxima we refer to as "end zones." These are indicated by boxes in Fig. 2*B* and are listed in Tables 2 and 3. Only one helix end is outside of an end zone and that one by only one amino acid.

Figure 3*A* shows the 19AA reverse-turn frequency scan ($TURN_{19}$) of the *L* subunit with the $IFH(0.5)_{19}$ maxima of Fig. 2*A* superimposed on the plot. The pattern of the peaks and valleys of the scan is the reverse of the hydrophobicity scan; deep minima coincide with each helix while the interhelix domains are characterized by large maxima. Note too that there is a one-to-one correspondence between $(IFH(0.5)_{19}$ maxima and $TURN_{19}$ minima. The lowest point (marked with vertical dotted lines) of each minimum for all the subunits are recorded in Tables 2 and 3. These minima correspond very closely to the true centers of the helices with $\Delta HC = -0.1 \pm 3.1AA$.

The 5AA scan ($TURN_5$) of the average reverse-turn probability is shown in Fig. 3*B*. Superimposed on the plot are the $IFH(0.5)_5$ end-zones. The most striking feature of the plot is that every helix is bounded by sharp and relatively high maxima in the average reverse-turn probability which generally fall within the defined end-zones. Note, however, that there are usually smaller maxima within the helix region but that these do not fall within the defined end-zones. The locations of the nearest maxima which bound the helices are listed in Tables 2 and 3. The $TURN_5$ maxima which bound the heli-

Table 3. Summary of the analysis of the primary sequences of the *L*, *M*, and *H* subunits of *Rps. viridis*[a]

Helix letter	Obs. Ends[b]	$TURN_{19}$ min. point	IFH_{19} max. point	IFH_{19} max. ±9AA	IFH_5 end zones	$TURN_5$ max. point
Rps. Viridis			*L* subunit			
A [24]						
NT:	32	—	—	30	16– 33	28
HC:	44	39	39	—	—	—
CT:	55	—	—	48	41– 64	58
B [29]						
NT:	84	—	—	84	78– 95	83
HC:	98	96	93	—	—	—
CT:	112	—	—	112	95–114	113
C [26]						
NT:	115	—	—	113	114–128	113
HC:	127	127	122	—	—	—
CT:	140	—	—	131	128–139	142
D [30]						
NT:	170	—	—	178	159–179	171
HC:	184	186	187	—	—	—
CT:	199	—	—	196	194–208	201
E [27]						
NT:	225	—	—	232	226–235	223
HC:	238	233	241	—	—	—
CT:	251	—	—	250	247–265	252
			M subunit			
A [27]						
NT:	52	—	—	53	35– 59	53
HC:	65	67	62	—	—	—
CT:	78	—	—	71	68– 82	81
B [30]						
NT:	110	—	—	111	104–117	111
HC:	124	122	120	—	—	—
CT:	139	—	—	129	117–140	140
C [26]						
NT:	142	—	—	148	140–154	140
HC:	154	157	157	—	—	—
CT:	167	—	—	166	166–175	171
D [29]						
NT:	197	—	—	206	186–205	199
HC:	211	217	215	—	—	—
CT:	225	—	—	224	222–235	229
E [27]						
NT:	259	—	—	266	255–273	256
HC:	272	275	275	—	—	—
CT:	285	—	—	284	283–294	287
			H subunit			
A [26]						
NT:	12	—	—	12	5– 16	12
HC:	24	22	21	—	—	—
CT:	37	—	—	31	29– 43	41

[a] *See* Table 2 legend for abbreviations.
[b] *Reference:* Deisenhofer et al., 1985.

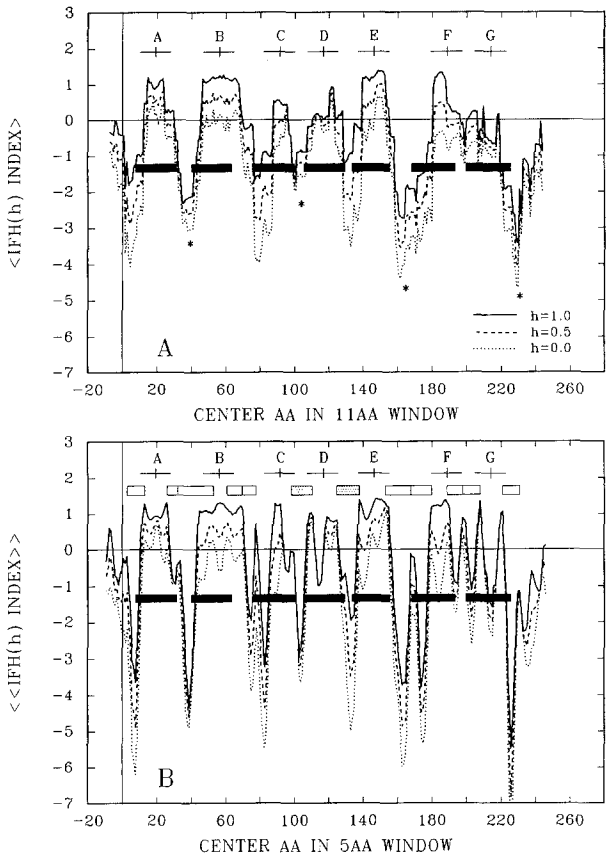

Fig. 4. Sliding-window averages of IFH(h) for bacteriorhodopsin. *See* legend of Fig. 2 for explanation of symbols. The shaded boxes indicate overlap of end zones. The "known" helix positions indicated by the solid bars are the consensus positions described in Table 1. (*A*) Average with 19AA window. (*B*) Average with a nominal 5AA window

ces do so with impressive statistics with $\langle \Delta N \rangle = -1.0 \pm 1.9$ and $\langle \Delta C \rangle = +3.0 \pm 1.5$. Thus, on average, the maxima occur one residue before the helix begins and three residues after it ends. Only five of the maxima of the 44 helix ends (11%) fall outside of the end-zones with the average miss being 3 ± 2 (SD) amino acids. Recalling that helix ends invariably fall within end-zones, a potentially useful stratagem for assigning helix ends of a membrane protein of uncertain structure is to take the helix end as the nearest end-zone terminus should the RT maximum fall outside the end-zone.

Bacteriorhodopsin

Because of the uncertainties in the helix assignments of BR, we do not perform a quantitative analysis. Rather, we broadly compare and contrast the behavior of BR with the PSRC. The main question is whether or not the putative helix ends are consistent with the PSRC observations.

As discussed in Methods, BR has seven helices in a peptide chain of about 250 amino acids while the PSRC *L* and *M* subunits have only five helices in a 300-amino acid chain. The helix-connecting links are therefore shorter and the helices closer together and more difficult to resolve using wide-window IFH(h) plots. We therefore used 11AA windows as discussed in Methods to resolve the seven peaks as shown in Fig. 4*A*. Even then, helices F and G are difficult to resolve but there is a clear "notch" separating them. In addition, the variations in the plot as h varies from 0 to 1 helps one visualize the peaks. Having located the seven peaks, we used 19AA scans (*data not shown*) to locate the $\text{IFH}(0.5)_{19}$ maxima located nearest to the identified IFH_{11} peaks. These maxima are indicated in Fig. 4*A*.

Another result of the close spacing of the helices is that the secondary peaks in the $\text{IFH}(h)_5$ plots (Fig. 4*B*) tend to overlap (shaded end-zone boxes). In addition, there are charged residues buried in the protein interior which cause some of the main peaks in the $\text{IFH}(h)_5$ plot to appear as a series of narrower peaks not easily distinguishable from secondary peaks. Helices C and G are good examples of this behavior. We have selected the end-zones in Fig. 4*B* so that they bracket the putative helix ends. Note that the helix B peak, which is very broad, appears to consist of a merged central and secondary peak on the C end. Despite the resolution problem, it is nevertheless possible to select patterns of peaks and valleys similar to those of the PSRC. The resulting end-zones are shown in Table 4 and Fig. 4*B*.

We have indicated by asterisks in Fig. 4*A* the cytoplasmic helix-connecting links and chain end. The alternating IFH(h) levels of the connecting links seen in the PSRC are absent; all of the minima have approximately the same depth. However, consistent with the PSRC, the noncytoplasmic connecting links tend to be more sensitive to variations in h as illustrated by means of the fractional change in IFH(h) in Fig. 5*B*. These plots are noisier than for PSRC but the pattern is generally consistent. We have included in the IFH(h) plots the 13AA presequence which behaves similarly to the C-terminus of the PSRC and one would conclude by comparison that the N-terminus must cross the membrane. Without the inclusion of the presequence, it would not be possible to decide whether the N-terminus or C-terminus was on the cytoplasmic side. This observation is consistent with the notion of Von Heijne and Segrest (1987) that the presequence of BR acts as a temporary eighth helix so that BR inserts in the membrane as four helical hairpins (Engelman

Table 4. Summary of analysis of the primary sequence of bacteriorhodopsin[a]

Helix letter	Obs. Ends[b]	$TURN_{19}$ min. point	IFH_{19} max. point	IFH_{19} max. ±9AA	IFH_5 end zones	$TURN_5$ max. point
A [25]						
NT:	8	—	—	11	3– 13	7
HC:	20	20	20	—	—	—
CT:	32	—	—	29	26– 32	36
B [24]						
NT:	41	—	—	48	32– 53	36
HC:	53	53	57	—	—	—
CT:	64	—	—	66	61– 70	64
C [26]						
NT:	76	—	—	83	70– 78	73
HC:	88	92	92	—	—	—
CT:	101	—	—	101	98–111	105
D [25]						
NT:	106	—	—	109	98–111	105
HC:	118	117	118	—	—	—
CT:	130	—	—	127	125–138	132
E [23]						
NT:	134	—	—	138	125–138	132
HC:	145	144	147	—	—	—
CT:	156	—	—	156	153–168	157
F [26]						
NT:	169	—	—	180	168–180	164
HC:	181	182	189	—	—	—
CT:	194	—	—	198	189–198	194
G [27]						
NT:	200	—	—	206	198–208	194
HC:	213	215	215	—	—	—
CT:	226	—	—	224	221–231	239

[a] *See* Table 2 legend for abbreviations.

[b] Observed ends are the average values of Table 1. The numbers in brackets next to helix letters are the lengths of consensus helices.

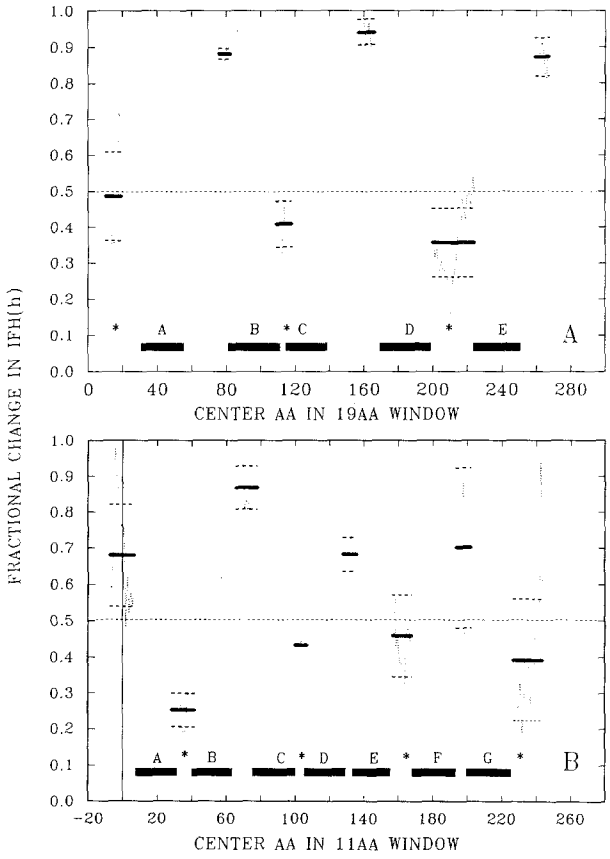

Fig. 5. Plots of the fractional changes in the averaged IFH(*h*) index as *h* varies from 0 to 1. The light dotted curves are plots of $\{\langle IFH(0)\rangle - \langle IFH(1)\rangle\}/IFH(1)$ for the interhelical region where $IFH(1) < 0$. The heavy lines indicate the fractional value averaged over the interval; the dashed lines show the means ± SD. Solid bars show the locations of helices and the asterisks (*) the helix-connecting links on the cytoplasmic surfaces. Note that the cytoplasmic links have fractional values of less than 0.5. (*A*) Fractional changes for the *L* subunit of *Rb. sphaeroides*. (*B*) Fractional changes for bacteriorhodopsin

& Steitz, 1981). Similarly, the C-terminus of PSRC would thus appear to act as a temporary sixth helix during insertion (*vide ut infra*).

The scans of reverse-turn frequency are shown in Fig. 6. The wide-window scan ($TURN_{19}$, Fig. 6*A*) shows a behavior similar to that of the PSRC except for helix D which grossly appears as a major maximum. Closer examination, however, reveals a minor minimum at the center of the maximum which corresponds to the putative center of the helix. The locations of the $TURN_{19}$ minima are listed in Table 4. The behavior of helix D identifies it as unusual compared to all of the PSRC helices and the other six BR helices. The reason for the behavior is that helix D has a very high proportion of Gly residues and an Asp near its center (Fig. 7). We cannot know if this means that this helix has a special structural or functional role until a high-resolution structure is available. As we discuss later, closely related halorhodopsin also shows this behavior but to a greater extent. The $TURN_5$ scan shown in Fig. 6*B* shows that each helix is bounded by reverse-turn maxima which correlate well with the selected end zones (Table 4).

Discussion

The principal observations from our analysis of the PSRC which may be useful for examining mem-

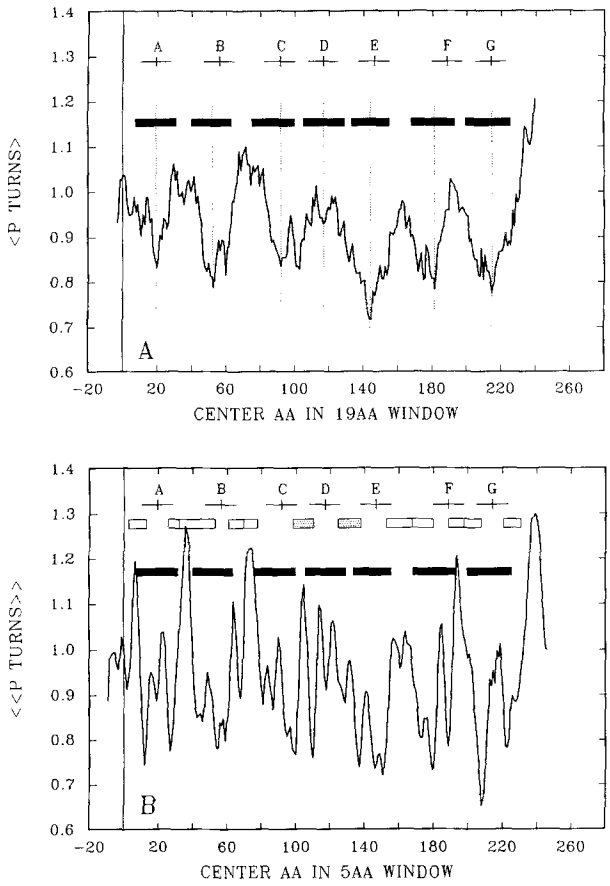

Fig. 6. Sliding-window averages of reverse-turn frequency for bacteriorhodopsin. *See* legends of Figs. 2 and 4 for explanation of symbols. (*A*) Average with 19AA window. Note that the behavior of helix D is anomalous compared to the other helices and those of the PSRC (Fig. 3). (*See* text.) (*B*) Average with nominal 5AA window

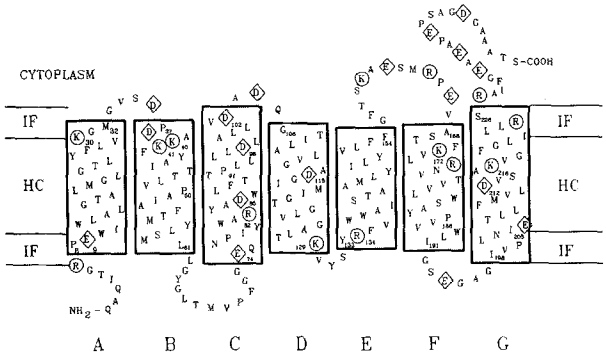

Fig. 7. Helical net representation of the helices of bacteriorhodopsin chosen using the search protocol established from the analysis of the PSRC. (*See* text and Table 1.) The helical nets have been accurately drawn based upon the work of Dunnill (1968). The bilayer is divided into hydrocarbon core (*HC*) and interface (*IF*) based upon the work of Jacobs and White (1989). The HC thickness corresponds to the length of a 20AA helix (30 Å). The amino acid sequence is that of Khorana et al. (1979)

brane proteins of unknown structures are the following: (i) The major maxima of wide-window averages of the IFH(*h*) index and major minima of reverse-turn frequency have a one-to-one correspondence and are very close to the centers of the helices, but the RT minima are somewhat more accurate. (ii) There are secondary maxima in narrow-window averages of IFH(*h*) index which can be used to define helix end zones within which the helix ends are located. (iii) In 90% of the cases there is a local $TURN_5$ maximum within the end zone region which correlates strongly with the location of the helix end. In general, a local maximum will occur about one amino acid before the N-end of the helix and three amino acids after the C-end with a precision of ±2AA. (iv) The variations in the average IFH(*h*) index with changes in *h* in the helix-connecting link and chain ends regions seem to reveal the topology of the protein. The links and ends on the cytoplasmic side tend to have averaged fractional IFH(*h*) changes of 0.5 or less, while those on periplasmic side tend to be larger.

Except for the wide-window RT maximum of helix D, the behavior of BR is generally consistent with that of the PSRC. Of course, the precise ends of the BR helices remain unknown so we cannot be absolutely certain. We note, however, that none of the models shown in Table 1 are inconsistent with the analysis; most of the proposed helix ends are within the selected end zones. In fact, we can use observation (iii) above to "predict" the helix ends of BR. Our assignments based on the PSRC observations are shown in Table 1 and Fig. 7 and are in good agreement with the consensus values.

Khorana and his colleagues have prepared a large number of BR mutants that have single amino acid substitutions in regions thought to be important in BR function. In particular, they have made extensive changes in the F helix (Hackett et al., 1987); substitutions for Asp in the C, D, and F helices and A–B and C–D connectors (Mogi et al., 1988): and cysteine replacements of several different amino acids in the B–C and E–F loops and in helix C (Flitsch & Khorana, 1989). All of these replacements appear to result in properly refolded protein, and it is thus of interest to look at the effects of these substitutions on the analysis of BR. An extensive examination of the effects revealed no significant changes in the augmented hydrophobicity analysis. This is not surprising because most of the substitutions such as Asp → Asn were conservative both in terms of hydrophobicity and RT preference. In any case, the sliding-window averaging minimizes the effects of *single* substitutions. Because the mutants could apparently refold and because

the analysis of the mutant sequences showed no significant changes, we conclude that the our general analysis is reasonable.

Fimmel et al. (1989) have questioned the correctness of some of the consensus values for BR based upon proteinase K digestions of accessible residues and micro-sequencing of released fragments. In particular, they propose that helix D is considerably shorter (16AA) with termination at G120 rather than K129 because a fragment G122-Y133 is produced. They argue that proteinase K is unlikely to have access to the interior portions of a helix terminating at K129. We find it interesting in our analysis that helix D is exceptional because of its high RT frequency arising from the Gly and Asp residues. A possible interpretation is that helix D is inherently unstable because of these residues; once its end is attacked by the proteinase it might partially "unwind" and become accessible for further digestion. We suggest that the Fimmel et al. (1989) proposal is not consistent with the analysis of Engelman et al. (1980) based upon the electron density projection map of BR of Henderson and Unwin (1975). The relative scattering strengths of the seven helices vary from 0.955 to 1.036. If the average length of helices A–C and E–F were 25AA, then one would expect at least one of the helices to have a relative scattering strength of only about 0.64 (=16/25).

The apparent anomalous behavior of helix D in our analysis thus seems to be consistent with its controversial experimental character, which may mean that the analysis has identified a helix with particular structural or functional properties presently unknown to us. The analysis of *Halobacterium halobium* proteins closely related to BR are interesting in this regard. Halorhodopsin (HR) is a protein with about 36% sequence identity with BR (Blanck & Oesterheld, 1987) but which pumps chloride rather than hydrogen ions. HR is generally assumed to be very similar to BR in secondary and tertiary structure. However, while the hydropathy plot of HR shown in Fig. 8*A* is similar to that of BR (Fig. 4*A*), the wide-window RT scan is drastically different as shown in Fig. 8*B* (*compare* to Fig. 6*A*). In this case, hydrophobic regions B, C, and D are anomalous compared to PSRC. This suggests that these regions may be structurally quite different and particularly that key functional differences between BR and HR might reside with these presumptive helices.

An important matter to consider is the meaning of the four principal results of our analysis of the PSRC. It is not at all surprising, of course, that helix-connecting links have high RT frequency averages. It is interesting, however, that the helix ends fall consistently within end-zones defined by

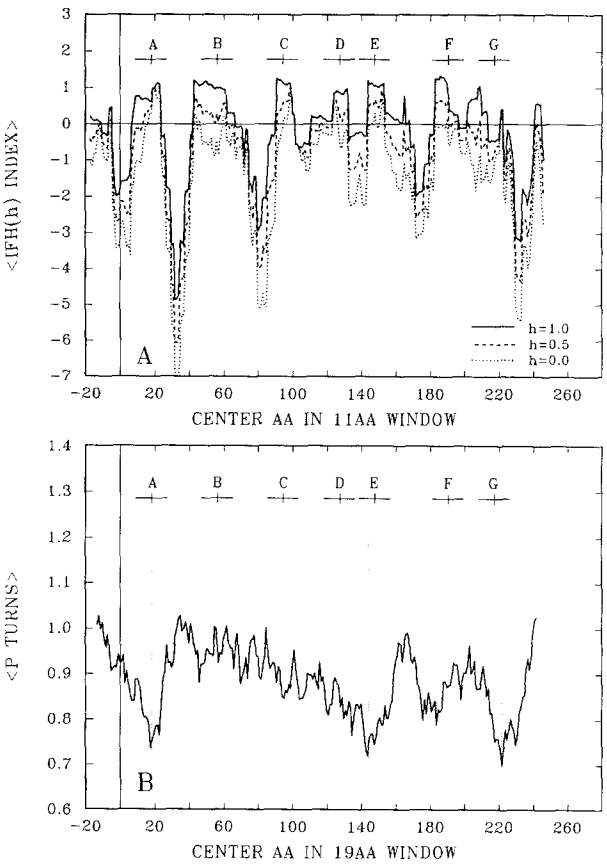

Fig. 8. Sliding-window averages of IFH(h) index and reverse-turn frequency for halorhodopsin. *See* Figs. 2–4 and 6 for explanation of symbols. (*A*) Average of IFH(h) using 19AA window. *Compare* to Fig. 4*A*. (*B*) Average of reverse-turn frequency using 19AA window. *Compare* to Fig. 6*A*. Note the behavior of region corresponding to hydrophobic maxima B, C, and D. (*See* text)

secondary maxima in hydrophobicity. Examination of the residues which contribute to the secondary IFH(h)$_5$ peaks reveals that 45% are drawn from the top ranks of RT frequency residues (Gly, Ser, Asp, Asn, and Pro) with Gly (25%) and Pro (10%) dominating. Except for Asp, these residues are located in the mid-range of the IFH(h) scale (Fig. 1) and thus lead to moderately high hydrophobicity peaks in the IFH(h)$_5$ plots. In the case of BR, which tends to lack obvious secondary peaks, the connecting links are short with a relatively high fraction of charged residues, especially Asp. It is probably significant that HR contains very few Asp or Lys residues compared to BR but a relatively large number of Arg residues. This explains in part the TURN$_{19}$ plot of HR because Arg is neutral with respect RT formation compared to Asp or Lys (Levitt, 1977).

The fact that RT peaks occur just before and after the helix ends is particularly interesting in the context of recent proposals by Presta and Rose

(1988) and Richardson and Richardson (1988) that side-chain/main-chain hydrogen bond formation at helix ends is important to helix formation and stability. Presta and Rose (1988) exhaustively examined the geometry of helix ends from many proteins and identified specific hydrogen bond favoring geometries between side chain polar groups and the four >NH and >CO groups exposed at the N-end and C-end of the helices. Interestingly, however, Presta and Rose (1988) did not find these geometries at the ends of the PSRC transbilayer helices.

We have suggested that the bilayer interface plays an important role in helix formation (Jacobs & White, 1989), and this may preclude the necessity for side-chain/helix-end hydrogen bond interactions for the initiation of helix formation. Our consideration of the functional consequences of the IFH(h) scale led us to suggest that the hydrogen bonding of the side chains could differ before, during, and after helix insertion and thereby play a role in the insertion process. The energetic cost of transferring a >CO or >NH group from water to a nonpolar region is 2–3 kcal/mol (Engelman & Steitz, 1981). An analysis by Roseman (1988) indicates an unbounded >NH and >CO *pair* requires about 6 kcal/mol to be buried in a nonpolar environment. Thus, the four pairs at the ends of a helical hairpin would require 24 kcal/mol to insert into the bilayer interior. It is reasonable to suppose that side chain/main chain hydrogen bonding at these helix ends could transiently lower the free energy costs of insertion. Presta and Rose's (1988) finding of a lack of such interactions in PSRC could be a result of the loss of favorable geometries in the final folded state.

The idea of at least transient side-chain hydrogen bond formation during insertion is supported by our observation that the helix-connecting links which must cross the membrane are characterized by large fractional changes in average hydrophobicity (relative to zero) as h varies from 0 to 1 in IFH(h) plots. One can reasonably hypothesize that during the insertion step the helix-connecting links are best described by $h = 1$ while afterward $h = 0$ is more appropriate. This idea suggests that the early steps of folding lead to side-chain hydrogen bond formation with protein or membrane components other than water which is favorable to insertion, while the later steps lead to a dissolution of these bonds and thereby stabilization of the helices across the bilayer. The participants in these transient nonaqueous phase hydrogen bonds could arise from specific lipid/peptide interactions, interactions with other membrane proteins, or from side-chain/main-chain and side-chain/side-chain interactions.

These suggestions for the role of transient hydrogen bond formation are quite consistent with the proposal of Von Heijne and Segrest (1987) regarding the role of the BR and HR presequences in the insertion processes. They have noted that the presequences of both proteins may form short amphipathic helices, which they suggest interact with polar regions on the other helices during insertion. As we noted earlier, the actual insertion complex is suggested by them to consist of eight helices with the presequence helix being subsequently expelled from the membrane and cleaved in a manner proposed earlier by Engelman and Steitz (1981). Our analysis of BR which includes the presequence is certainly consistent with this. Of further importance in this regard are the helices of the PSRC which are external to the membrane (Deisenhofer et al., 1985; Allen et al., 1987). There are three helices for both the L and M subunits on the periplasmic surface forming a colinear array; two are connectors for helices A-B and C-D and the other follows helix E. Importantly, Yeates et al. (1987) noted that all of these helices are amphipathic. It is thus possible that these helices are important to the insertion process in the ways suggested by Von Heijne and Segrest (1987) and Jacobs and White (1989). One should note, however, that interhelix hydrogen bonds occur only occasionally in the PSRC (Yeates et al., 1987).

It thus seems reasonable that the presequence of BR and the C-termini of the L and M subunits form amphipathic helices which aid insertion and determine in part the topology of the proteins. The fact that the L and M subunits have no presequences suggests that the primary purpose of the BR presequence is insertion. The BR presequence and the L and M subunit "post" sequences do not fit the standard pattern of the leader sequences of exported proteins cataloged by Von Heijne (1985), and it may be useful to refer to them as "insertion sequences." These sequences may appear at either end of a sequence or possibly between helices. These ideas are consistent with the broad conclusions of Randall and Hardy (1989) and the observations of Shaw, Rottier and Rose (1988).

We are pleased to acknowledge many pleasant conversations with Drs. Larry Vickery and Janos Lanyi. We thank Dr. Jay Edelman for his comments on several versions of the manuscript. This work was supported by grants from the National Science Foundation (DMB-8412754, DMB-8807431) and the American Heart Association—California Affiliate with funds contributed by the Orange County, CA. Chapter. REJ is an Established Investigator of the American Heart Association.

References

Agard, D.A., Stroud, R.M. 1982. Linking regions between helices in bacteriorhodopsin revealed. *Biophys. J.* **37**:589–602

Allen, J.P., Feher, G., Yeates, T.O., Komiya, H., Rees, D.C. 1987. Structure of the reaction center from *Rhodobacter*

sphaeroides R-26: The protein subunits. *Proc. Natl. Acad. Sci. USA* **84**:6162–6166

Baldwin, J.M., Henderson, R., Beckman, E., Zemlin, F. 1988. Images of purple membrane at 2.8 Å resolution obtained by cryo-electron microscopy. *J. Mol. Biol.* **202**:585–591

Blanck, A., Oesterhelt, D. 1987. The halo-opsin gene: II. Sequence, primary structure of halorhodopsin and comparison with bacteriorhodopsin. *EMBO J.* **6**:265–273

Britton, D., Green, A. 1985. Application of fourier techniques to hydrophilic analysis. *In:* Synthetic Peptides in Biology and Medicine. K. Alitalo, P. Partanaen, and A. Vaheri, editors. pp. 21–26. Elsevier Science, Amsterdam

Chothia, C. 1975. Structural invariants in protein folding. *Nature (London)* **254**:304–308

Chou, P.Y., Fasman, G.D. 1974. Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. *Biochemistry* **13**:211–221

Chou, P.Y., Fasman, G.D. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**:45–148

Deisenhofer, J., Epp, O., Miki, K., Huber, R., Michel, H. 1985. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3 Å resolution. *Nature (London)* **318**:618–624

Dunnill, P. 1968. The use of helical net-diagrams to represent protein structures. *Biophys. J.* **8**:865–875

Edelman, J., White, S.H. 1989. Linear optimization of predictors for secondary structure. Application to transbilayer segments of membrane proteins. *J. Mol. Biol.* **210**:195–209

Engelman, D.M., Henderson, R., McLachlan, A.D., Wallace, B.A. 1980. Path of the polypeptide in bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA* **77**:2023–2027

Engelman, D.M., Steitz, T.A. 1981. The spontaneous insertion of proteins into and across membranes: The helical hairpin hypothesis. *Cell* **23**:411–422

Engelman, D.M., Steitz, T.A. 1982. On the folding and insertion of globular membrane proteins. *In:* The Protein Folding Problem. D.B. Wetlaufer, editor. pp 87–113. Westview, Boulder, Colo.

Engelman, D.M., Steitz, T.A., Goldman, A. 1982. The identification of helical segments in the polypeptide chain of bacteriorhodopsin. *Methods Enzymol.* **88**:81–88

Engelman, D.M., Steitz, T.A., Goldman, A. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**:321–353

Engelman, D.M., Zaccai, G. 1980. Bacteriorhodopsin is an inside-out protein. *Proc. Natl. Acad. Sci. USA.* **77**:5894–5898

Fimmel, S., Choli, T., Dencher, N.A., Büldt, G., Wittman-Liebold, B. 1989. Topography of surface-exposed amino acids in the membrane protein bacteriorhodopsin determined by proteolysis and micro-sequencing. *Biochim. Biophys. Acta* **978**:231–240

Flitsch, S.L., Khorana, H.G. 1989. Structural studies on transmembrane proteins: 1. Model study using bacteriorhodopsin mutants containing single cysteine residues. *Biochemistry* **28**:7800–7805

Hackett, N.R., Stern, L.J., Chao, B.J., Kronis, K.A., Khorana, H.G. 1987. Structure-function studies on bacteriorhodopsin: V. Effects of amino acid substitutions in the putative F helix. *J. Biol. Chem.* **262**:9277–9284

Hayward, S.B., Stroud, R.M. 1981. Projected structure of purple membrane determined to 3.7 Å resolution by low temperature electron microscopy. *J. Mol. Biol.* **151**:491–517

Henderson, R. 1975. The structure of the purple membrane from *Halobacterium halobium:* Analysis of the x-ray diffraction pattern. *J. Mol. Biol.* **93**:123–138

Henderson, R., Unwin, P.N.T. 1975. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature (London)* **257**:28–32

Heyn, M.P., Westerhausen, J., Wallat, I., Seiff, F. 1988. High-sensitivity neutron diffraction of membranes: Location of the Schiff base end of the chromophore of bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA* **85**:2146–2150

Hopp, T.P. 1985. Prediction of protein surfaces and interaction sites from amino acid sequences. *In:* Synthetic Peptides in Biology and Medicine. K. Alitalo, P. Partanaen, and A. Vaheri, editors. pp. 3–12. Elsevier Science, Amsterdam

Jacobs, R.E., White, S.H. 1989. The nature of the hydrophobic binding of small peptides at the bilayer interface: Implications for the insertion of transbilayer helices. *Biochemistry* **28**:3421–3437

Jaffe, J.S. Glaeser, R.M. 1987. Structure of purple membrane from *Halobacterium halobium:* Recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. *Ultramicroscopy* **23**:17–28

Jähnig, F. 1983. Thermodynamics and kinetics of protein incorporation into membranes. *Proc. Natl. Acad. Sci. USA* **80**:3691–3695

Jansson, P.A. 1984. Deconvolution with Applications in Spectroscopy. Academic, New York

Jap, B.K., Kong, S.-H. 1986. Secondary structure of halorhodopsin. *Biochemistry* **25**:502–505

Jap, B.K., Maestre, M.F., Hayward, S.B., Glaeser, R.M. 1983. Peptide-chain secondary structure of bacteriorhodopsin. *Biophys. J.* **43**:81–89

Jubb, J.S., Worcester, D.L., Crespi, H.L., Zaccai, G. 1984. Retinal location in purple membrane of *Halobacterium halobium:* A neutron diffraction study of membranes labeled *in vitro* with deuterated retinal. *EMBO J.* **3**:1455–1461

Katre, N.V., Finer-Moore, J., Stroud, R.M., Hayward, S.B. 1984. Location of an extrinsic label in the primary and tertiary structure of bacteriorhodopsin. *Biophys. J.* **46**:195–204

Khorana, H.G. 1988. Bacteriorhodopsin, a membrane protein that uses light to translocate protons. *J. Biol. Chem.* **263**:7439–7442

Khorana, H.G., Gerber, G.E., Herlihy, W.C., Gray, C.P., Anderegg, R.J., Nihei, K., Biemann, K. 1979. Amino acid sequence of bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA* **76**:5046–5050

King, G.I., Mowery, P.C, Stoeckenius, W., Crespi, H.L., Schoenborn, B.P. 1980. Location of the chromophore in bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA* **77**:4726–4730

Kyte, J., Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of protein. *J. Mol. Biol.* **157**:105–132

Lear, J.D., Wasserman, Z.R., DeGrado, W.F. 1988. Synthetic amphiphilic peptide models for protein ion channels. *Science* **240**:1177–1181

Leifer, D., Henderson, R. 1983. Three-dimensional structure of orthorhombic purple membrane at 6.5 Å resolution. *J. Mol. Biol.* **163**:451–466

Levitt, M. 1977. Conformational preferences of amino acids in globular proteins. *Biochemistry* **17**:4277–4285

Lodish, H.F. 1988. Multi-spanning membrane proteins: How accurate are the models? *Trends Biochem. Sci.* **13**:332–334

McCrea, P.D., Engelman, D.M., Popot, J.-L. 1988. Topography of integral membrane proteins: Hydrophobicity analysis *vs.* immunolocalization. *Trends Biochem. Sci.* **13**:289–290

Michel, H., Weyer, K.A., Gruenberg, H., Dunger, I., Lott-

speich, F. 1985. The 'heavy' subunit of the photosynthetic reaction centre from *Rhodospseudomonas viridis:* Isolation of the genes, nucleotide and amino acid sequence. *EMBO J.* **4:**1667–1672

Michel, H., Weyer, K.A., Gruenberg, H., Dunger, I., Oesterhelt, D., Lottspeich, F. 1986. The 'light' and 'medium' subunits of the photosynthetic reaction centre from *Rhodospeudomonas viridis:* Isolation of the genes, nucleotide and amino acid sequence. *EMBO J.* **5:**1149–1158

Mogi, T., Stern, L.J., Marti, T., Chao, B.H., Khorana, H.G. 1988. Aspartic acid substitutions affect proton translocation by bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA* **85:**4148–4152

Ovchinnikov, Y., Adbulaev, N., Feigira, M., Kiselev, A., Lobanov, N. 1979. The structural basis of the functioning of bacteriorhodopsin: An overview. *FEBS Lett.* **100:**219–224

Ovchinnikov, Y., Adbulaev, N., Vasilov, R.G., Vturina, I.Y., Kuryatov, A.B., Kiselev, A.V. 1985. The antigenic structure and topography of bacteriorhodopsin in purple membranes as determined by interaction with monoclonal antibodies. *FEBS Lett.* **179:**343–350

Paul, C., Rosenbusch, J.P. 1985. Folding patterns of porin and bacteriorhodopsin. *EMBO J.* **4:**1593–1597

Popot, J.-L., Trewhella, J., Engelman, D.M. 1986. Reformation of crystalline purple membrane from purified bacteriorhodopsin fragments. *EMBO J.* **5:**3039–3044

Presta, L.G., Rose, G.D. 1988. Helix signals in proteins. *Science* **240:**1632–1641

Ptitsyn, O.B. 1974. Invariant features of globin primary structure and coding of their secondary structure. *J. Mol. Biol.* **88:**2878–300

Randall, L.L., Hardy, S.J.S. 1989. Unity in function in the absence of consensus in sequence: Role of leader peptides in export. *Science* **243:**1156–1159

Richardson, J.S. 1981. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34:**167–339

Richardson, J.S., Richardson, D.C. 1988. Amino acid preferences for specific locations at the ends of α helices. *Science* **40:**1648–1652

Rose, G.D. 1978. Prediction of chain turns in globular proteins on a hydrophobic basis. *Nature (London)* **272:**586–590

Rose, G.D., Roy, S. 1980. Hydrophobic basis of packing in globular proteins. *Proc. Natl. Acad. Sci. USA* **77:**4643–4647

Roseman, M.A. 1988. Hydrophobicity of the peptide C=O ··· H—N hydrogen bonded group. *J. Mol. Biol.* **201:**621–623

Rosenbusch, J.P. 1985. Hydrogen-bonding networks in bacteriorhodopsin may afford a proton pathway and structural stability. *Bull. Inst. Pasteur* **83:**20–220

Seehra, J.S., Khorana, H.G. 1984. Bacteriorhodopsin precursor. Characterization and its integration into the purple membrane. *J. Biol. Chem.* **259:**4187–4193

Seiff, F., Wallat, I., Ermann, P., Heyn, M.P. 1985. A neutron diffraction study on the location of the polyene chain of retinal in bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA* **82:**3227–3231

Seiff, F., Wallat, I., Westerhausen, J., Heyn, M.P. 1986*a*. Location of chemically modified lysine 41 in the structure of bacteriorhodopsin by neutron diffraction. *Biophys. J.* **50:**629–635

Seiff, F., Westerhausen, J., Wallat, I., Heyn, M.P. 1986*b*. Location of the cyclohexane ring of the chromophore of bacteriorhodopsin by neutron diffraction with selectively deuterated retinal. *Proc. Natl. Acad. Sci. USA* **83:**7746–7750

Shaw, A.S., Rottier, J.M., Rose, J.K. 1988. Evidence for the loop model of signal-sequence insertion into the endoplasmic reticulum. *Proc. Natl. Acad. Sci. USA* **85:**7592–7596

Trewhella, J., Popot, J.-L., Zaccai, G., Engelman, D.M. 1986. Location of two chymotryptic fragments in the structure of renatured bacteriorhodopsin by neutron diffraction. *EMBO J.* **5:**3045–3049

Vickery, L.E. 1987. Interactive analysis of protein structure using a microcomputer spreadsheet. *Trends Biochem. Sci.* **12:**37–39

Vogel, H., Gärtner, W. 1986. The secondary structure of bacteriorhodopsin determined by Raman and circular dichroism spectroscopy. *J. Biol. Chem.* **262:**11464–11469

Vogel, H., Jähnig, F. 1986. Models for the structure of outer-membrane proteins of *Escherichia coli* derived from Raman spectroscopy and prediction methods. *J. Mol. Biol.* **190:**191–199

Von Heijne, G. 1985. Signal sequences: The limits of variation. *J. Mol. Biol.* **184:**99–105

Von Heijne, G., Blomberg, C. 1979. Trans-membrane translocation of proteins. *Eur. J. Biochem.* **97:**175–181

Von Heijne, G., Segrest, J.P. 1987. The leader peptides from bacteriorhodopsin and halorhodopsin are potential membrane-spanning amphipathic helices. *FEBS Lett.* **213:**238–240

Wallace, B.A., Cascio, M., Mielke, D.L. 1986. Evaluation of methods for the prediction of membrane protein secondary structures. *Proc. Natl. Acad. Sci. USA* **83:**9423–9427

Wallace, B.A., Henderson, R. 1982. Location of the carboxyl terminus of bacteriorhodopsin in purple membrane. *Biophys. J.* **39:**233–239

Williams, J.C., Steiner, L.A., Feher, G. 1986. Primary structure of the reaction center from *Rhodopseudomonas sphaeroides. Proteins* **1:**312–325

Williams, J.C., Steiner, L.A., Feher, G., Simon, M.I. 1984. Primary structure of the L subunit of the reaction center from *Rhodopseudomonas sphaeroides. Proc. Natl. Acad. Sci. USA* **81:**7303–7307

Williams, J.C., Steiner, L.A., Ogden, R.C., Simon, M.I., Feher, G. 1983. Primary structure of the M subunit of the reaction center from *Rhodopseudomonas sphaeroids. Proc. Natl. Acad. Sci. USA* **80:**6505–6509

Yeates, T.O., Komiya, H., Rees, D.C., Allen, J.P., Feher, G. 1987. Structure of the reaction center from *Rhodobacter sphaeroides* R-26: Membrane-protein interactions. *Proc. Natl. Acad. Sci. USA* **84:**6438–6442

Received 19 July 1989; revised 22 November 1989