STRUCTURAL AND MECHANISTIC MOTIFS IN MEMBRANE PROTEINS:

The Three-Dimensional Modelling of Rhodopsin, Band 3, and the Nicotinic Acetylcholine Receptor

Thesis by

Nancy Swick Vogelaar

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1989

(Submitted December 7, 1988)

To my family, and especially to my father,

who gave me his love of science.

Listen to this, O Job,

Stand and consider the wonders of God...

Job 37:14

ACKNOWLEDGEMENTS

While here at Caltech, many people have influenced both my life and work. Premier among those affecting my work is, of course, my advisor Sunney Chan. I am grateful for his support and encouragement throughout these years.

For many of the programs used in this thesis, I would like to thank Tim Hunkapiller, who developed them as part of his own research. I would also like to thank Dr. Leroy Hood for the use of his computer facility for the sequence analyses.

For the use of the *Biograf* program and his computational facilities, I am indebted to Dr. Bill Goddard. His personal interest in this work has also been greatly appreciated.

Reid Leonard has played an important role in this work through the many useful discussions we have had on the acetylcholine receptor. I am also grateful to Dr. Henry Lester for making information available prior to publication.

There are some people whom I would like to thank for their help on work not discussed in this thesis. Kathy Kanes, a fellow Chan group member, was a collaborator on much of the experimental work involving band 3. Working with her was both productive and enjoyable.

Gary Pipes was of great assistance in the work involving gel electrophoresis, protein blotting, and the preparation of samples for protein sequencing. John Tomich was instrumental in guiding our work in peptide purification. For teaching me his protocol for protein electroelution and for other needed assistance, I thank Warren Kibbe.

My earliest work here involved a significant amount of electron microscopy. Tim Pearson was a collaborator in this work and taught me a great deal. For the use of his microscopy lab, I would like to thank Dr. Jean-Paul Revel. I would also like to thank Mr. Pat Koen and Mrs. Jean Edens for their expert assistance.

Throughout my work, the Chan group has been a constant source of both inspiration and diversion and to all its members I am indebted.

Last, but certainly not least, I would like to acknowledge a few people who have enriched my life during this time. My husband Bruce has been a constant source of encouragement during these last few years, and going through graduate school together has made this time very special. Gary Kruppa, Joe Rushanan, and Dave James have all been terrific friends through thick and thin. Dr. Edward and Mrs. Margaret Merchant have provided us with a "home away from home." Their love and gracious hospitality will never be forgotten.

Finally, I would like to thank my "quintet buddies" Jo, John, Harry, Ellen, Sarah, and Jeff, and my "horn buddies" Wayne and Dave for their friendship and their music. It's been great, guys!

This work was supported by grant GM-22432 from the National Institute of General Medical Sciences, U.S. Public Health Service.

ABSTRACT

Because so little is known about the structure of membrane proteins, an attempt has been made in this work to develop techniques by which to model them in three dimensions. The procedures devised rely heavily upon the availability of several sequences of a given protein. The modelling procedure is composed of two parts. The first identifies transmembrane regions within the protein sequence on the basis of hydrophobicity, β-turn potential, and the presence of certain amino acid types, specifically, proline and basic residues. The second part of the procedure arranges these transmembrane helices within the bilayer based upon the evolutionary conservation of their residues. Conserved residues are oriented toward other helices and variable residues are positioned to face the surrounding lipids. Available structural information concerning the protein's helical arrangement, including the lengths of interhelical loops, is also taken into account. Rhodopsin, band 3, and the nicotinic acetylcholine receptor have all been modelled using this methodology, and mechanisms of action could be proposed based upon the resulting structures.

Specific residues in the rhodopsin and iodopsin sequences were identified, which may regulate the proteins' wavelength selectivities. A hinge-like motion of helices M3, M4, and M5 with respect to the rest of the protein was proposed to result in the activation of transducin, the G-protein associated with rhodopsin. A similar mechanism is also proposed for signal transduction by the muscarinic acetylcholine and

β-adrenergic receptors.

The nicotinic acetylcholine receptor was modelled with four transmembrane helices per subunit and with the five homologous M2 helices forming the cation channel. Putative channel-lining residues were identified and a mechanism of channel-opening based upon the concerted, tangential rotation of the M2 helices was proposed.

Band 3, the anion exchange protein found in the erythrocyte membrane, was modelled with 14 transmembrane helices. In general the pathway of anion transport can be viewed as a channel composed of six helices that contains a single hydrophobic restriction. This hydrophobic region will not allow the passage of charged species, unless they are part of an ion-pair. An arginine residue located near this restriction is proposed to be responsible for anion transport. When ion-paired with a transportable anion it rotates across the barrier and releases the anion on the other side of the membrane. A similar process returns it to its original position. This proposed mechanism, based on the three-dimensional model, can account for the passive, electroneutral, anion exchange observed for band 3. Dianions can be transported through a similar mechanism with the additional participation of a histidine residue. Both residues are located on M10.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

CHAPTER 1

INTRODUCTION

Band 3, the anion-exchange protein located in the erythrocyte membrane, is being investigated by the Chan laboratory, with an emphasis toward the determination of its structure and membrane topology. Toward this end, both the experimental technique of peptide mapping and the use of predictive algorithms have been used to examine the protein. The application of these predictive algorithms to band 3, and to membrane proteins in general, will be the main focus of the present work.

Proteins located in the membranes of cells are not as well characterized as soluble proteins. One of the main reasons for this disparity is that these proteins are difficult to study using many standard biophysical techniques, due to interference caused by the presence of a lipid bilayer. In addition, their physical properties often make purification, reconstitution, and other procedures difficult. Progress in the structure determination of some of these proteins has been made, however, and will be discussed briefly below. Because band 3 is an ion transporter, the structures of ion transport proteins will be emphasized.

There are three major types of ion transporter, each with a different channel structure. In the first type, the channel through the membrane is formed by a β-sheet like structure. The hydrophobic side chains extend outward into the bilayer and the hydrophilic amide linkages line the channel. Proteins that are believed to have channels of

this type include gramicidin A (1) and porin (2). A second type of ion transporter has been proposed that consists of adjacent, amphiphilic helices without intervening aqueous compartments. The only water present is water of hydration. This type of structure has been proposed for some H^+-transporters including bacteriorhodopsin (3). Most ion transporters, however, are believed to incorporate a third type of channel, which is composed of several amphipathic α-helices. These helices are approximately perpendicular to the plane of the bilayer and are arranged such that their hydrophilic faces are directed toward the channel and their hydrophobic faces are toward the lipid. The peptide chain extending from these membrane helices is envisioned to adopt a structure, which would allow for the continuation and widening of the channel on both sides of the membrane. The vestibules so formed are thought to have a net charge arranged such that a dense layer of countercharges collects at the mouths of the channel (4). Such an arrangement allows the efficient gathering of transportable ions and, in part, determines the charge selectivity of the channel protein. The determination of the channel's selectivity on the basis of size, shape, and perhaps also charge, is presumed to occur at the narrowest region of the channel, which is thought to be located within the transmembrane region.

In addition to their overall channel structure, many ion transport proteins are gated, meaning that they open and close in response to their environment and do not, in general, allow the free exchange of ions across the membrane. The nicotinic acetylcholine receptor is an example of such a protein (5) and will be discussed in more detail

below. The mechanisms responsible for the opening and closing of such channels are currently unknown.

Band 3 is also a gated protein, although it is unusual in that it effects the 1 for 1 exchange of anions across the membrane, rather than the unidirectional transport most common to gated proteins (6). A schematic drawing of the mechanism utilized by the band 3 protein can be seen in figure 1.1. First, anions bind to the outward-facing transport site. The protein is then able to change conformation, resulting in the transport of the ion across the membrane. After translocation, the anion dissociates from the protein and the inner transport site is free to bind an anion from the cytoplasmic compartment. Once bound, the protein may again translocate, effecting the transport of the second anion, and the process continues. The protein cannot change conformation unless an anion is bound to its transport site. This type of mechanism has been termed an alternating site or "ping-pong" mechanism (7). It should be emphasized that although the transport characteristics of many of these proteins have been examined, the molecular details of channel regulation and ion transport are still largely unknown.

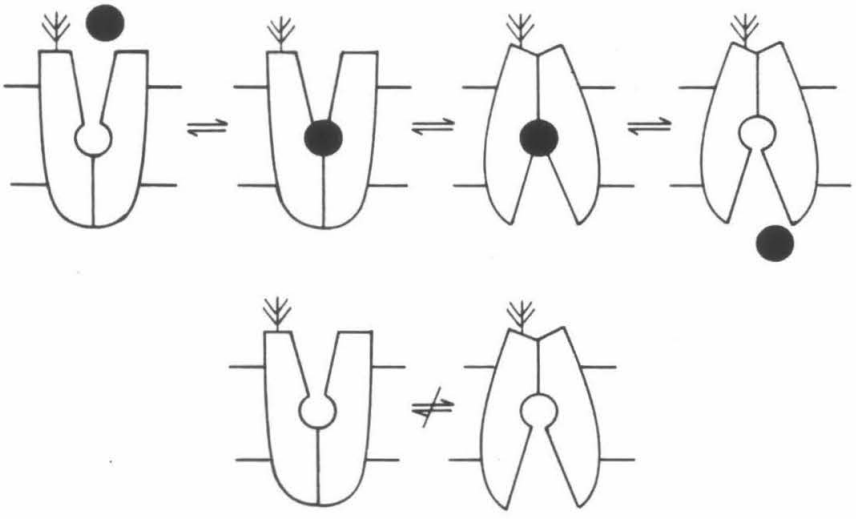
If the structures of these proteins could be obtained, especially in the transmembrane regions, possible mechanisms for gating and ion selectivity could be proposed and tested. Toward this end, a variety of experimental techniques have been used to examine the structures of membrane proteins. The most powerful technique, that of x-ray crystallography, has yielded the high-resolution structures of the reaction centers of *Rhodopseudomonas viridis* (8) and *Rhodobacter*

Figure 1.1

Ping-Pong or Alternating Site Mechanism

This schematic drawing illustrates the individual steps required for a 1 for 1 exchange of substrate across the membrane. It is essential to the effectiveness of this mechanism that the transporter be unable to switch from one conformation to the other in the absence of substrate, or to do so only very slowly.

ALTERNATING SITE MODEL



● = SUBSTRATE

sphaeroides (9). Although these proteins are not ion transporters, their structures yield general information about the relative positioning of helices within the membrane. The low resolution crystal structure of bacteriorhodopsin, a protein which translocates protons in response to light activation, also shows the relative orientation and number of transmembrane helices (10). The examination of crystals of the nicotinic acetylcholine receptor using electron diffraction has yielded information on the overall shape of this protein (11).

Aside from the cases just mentioned, all information on membrane protein topology has been derived from chemical experiments, which have located disulfide bonds, proteolytic sites, and the sites of labelling by various probes. Spectroscopic techniques have also provided information about the overall secondary structure and the environments of various regions of these proteins. In the absence of crystallographic data, however, these characterizations of the protein are rarely sufficient to define its three-dimensional structure. The only recourse in such cases is to attempt to model the protein using those experimental results in conjunction with predictive techniques based on the protein's sequence. At present, most of these predictive algorithms have been applied only to determine the location of transmembrane helices with respect to the primary sequence. The hydrophobic moment analysis of Eisenberg (12) is one of the few to attempt to predict the location of these helices with respect to hydrophilic and hydrophobic environments. In this work, another type of analysis to predict the arrangement of helices within the bilayer is proposed. Such predictive algorithms may allow the prediction of

some three-dimensional structure from sequence analysis.

The goal of the present work was to use various techniques to identify the transmembrane regions of band 3 and to arrange them in the bilayer in a manner consistent with all known structural information. This arrangement should determine possible mechanisms for ion transport and channel gating. Toward this end, the known sequences of band 3 proteins were examined with respect to hydropathy, hydrophobic moment, secondary structure (as predicted by the Chou-Fasman algorithm (13)), genetic structure, and a unique variability analysis. To test the usefulness of these methods, the nicotinic acetylcholine receptor and the bacteriorhodopsin superfamilies were also analyzed. These protein families were chosen because they are both involved in ion transport, have been studied extensively, and some information about their three-dimensional structures is available. Although it is not an anion transporter, the reaction center proteins were also included in these analyses because their membrane dispositions and structures are well-defined. A brief description of the known characteristics of these proteins will be given below.

The Reaction Center

The reaction center is a protein complex which is found in the membranes of photosynthetic bacteria. With the energy captured by light-collecting complexes, these proteins effect the transport of an electron across the membrane. The reaction center, therefore, converts the energy of a photon to electrical energy in the form of a transmembrane potential. This protein complex consists of three sub-

units designated as L, M, and H, and also includes a number of cofactors (8). The structure of this protein has been determined to high resolution for two different bacteria and one of these structures is shown in figure 1.2. In general, the L and M subunits form a core region, with the five helices from each subunit spiralling around the cofactors. The subunits are approximately related to one another through rotation around an axis of two-fold symmetry which is perpendicular to the plane of the bilayer. The H subunit has a transmembrane helix which probably serves only as a membrane anchor and so will not be discussed in detail in this work. The location of the membrane with respect to the protein has been located through energy calculations because lipid and detergent molecules could not be located in the crystal structures due to their mobility (14).

The subunits L and M are homologous, a fact that is reflected in their symmetrical arrangement. It has also been observed that the D1 and D2 subunits of photosystem II in green plants are homologous to L and M. Although they are distantly related, it has been proposed that these proteins form a family (15).

The Nicotinic Acetylcholine Receptor Superfamily

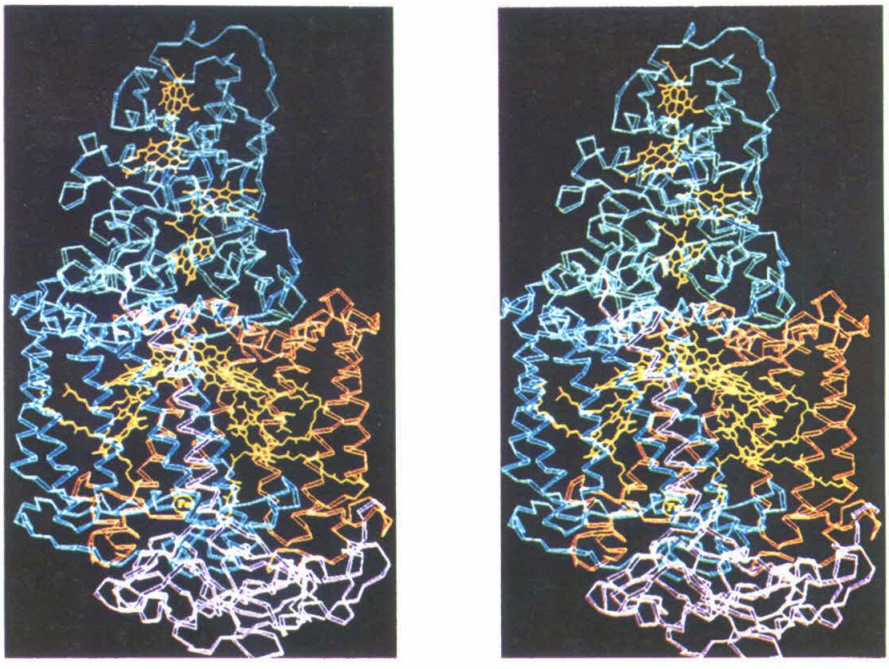
The nicotinic acetylcholine receptor is the most characterized of the neurotransmitter receptors. It is located on the post-synaptic cell membrane and is found both in neural and muscular tissue. When a signal is passed from cell to cell, acetylcholine is the substance released into the synaptic cleft. When it binds to the acetylcholine receptor, it induces the opening of a cation-selective channel. This

Figure 1.2

Structure of the Reaction Center

This stereodrawing of the reaction center of *Rhodopseudomonas viridis* shows the protein complex from a direction parallel to the plane of the bilayer. Color code: L subunit (orange), M subunit (blue), H subunit (purple), cytochrome (green), and prosthetic groups (yellow).

This figure was reproduced from reference 8.



opening results in an influx of Na^+ and K^+ ions and causes a change in the membrane potential. The net result is that the receptor transduces the chemical signal of the neurotransmitter to an electrical signal in the form of a change in the post-synaptic cell's membrane potential.

The acetylcholine receptor has a molecular weight of about 125 kDa and is composed of five homologous subunits. Neural receptors are composed of two types of subunits, α and β (16), and neuromuscular acetylcholine receptors are composed of four types of subunits with the stoichiometry $\alpha_2\beta\gamma\delta$ (5). The acetylcholine binding sites are located on the α subunits of neuromuscular receptors and on the β subunits of the neural receptor (17). This seeming discrepancy may be due to differences in nomenclature or perhaps to a misidentification of the neural α subunit. The α subunit, in particular, shows diversity in its primary structure depending on its tissue of origin. The existence of a gene family including the various subtypes has been proposed (18). Except for the presence of the acetylcholine-binding subunit, other subunit types can be substituted for those naturally occurring in the receptor. Although the properties of the modified receptor are generally quite different, the channel gating is functional and responds to acetylcholine (19).

Each of the subunits is believed to cross the membrane four or five times, with one helix from each subunit forming the channel. Although the number of helices is currently disputed in the literature, the extracellular location of the C-terminus favors the four-helix model (20). The transmembrane topology of the subunits will be

discussed in more detail in chapter 2, section C.

The overall three-dimensional shape of the acetylcholine receptor has been determined by electron microscopy (11) and is shown in figure 1.3. The five-fold symmetry of the receptor is apparent and suggests the equal participation of each of the subunits in channel formation. The central location and size of the channel are also indicated. Details of this channel structure will be discussed later in this work.

Other members of this superfamily include the γ-aminobutyric acid (GABA) receptor (21,22), and the strichnine-binding subunit of the glycine receptor (23). The GABA receptor is made up of only four subunits, in contrast to the five subunits of the acetylcholine receptor. The stoichiometry of the GABA subunits is $\alpha_2\beta_2$ (24) and both the α and β subunits are homologous to the α subunit of the acetylcholine receptor. One major difference between the two receptors is that agonist binding to the GABA receptor triggers the opening of an anion channel, rather than a cation channel. The glycine receptor is also believed to be composed of two types of homologous subunits (25) and like the GABA receptor prevents neural firing by hyperpolarizing the neural membrane. For this reason, an anion transport channel is probable, but this protein has not been well-characterized yet.

The "Bacteriorhodopsin" Superfamily

This superfamily is composed of the proton pump bacteriorhodopsin (26), the chloride pump halorhodopsin (29), the muscarinic acetylcholine receptor, the adrenergic receptor, and all known eucaryotic

Figure 1.3

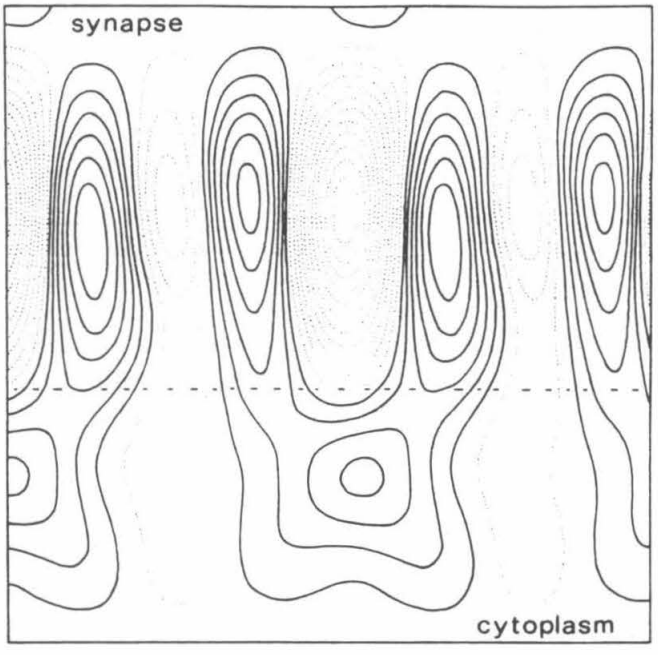
Structure of the Acetylcholine Receptor

A. Contour display of the acetylcholine receptor for a section which is perpendicular to the plane of the bilayer and passes through the receptor's center. The central plane of the bilayer is indicated by a broken line.

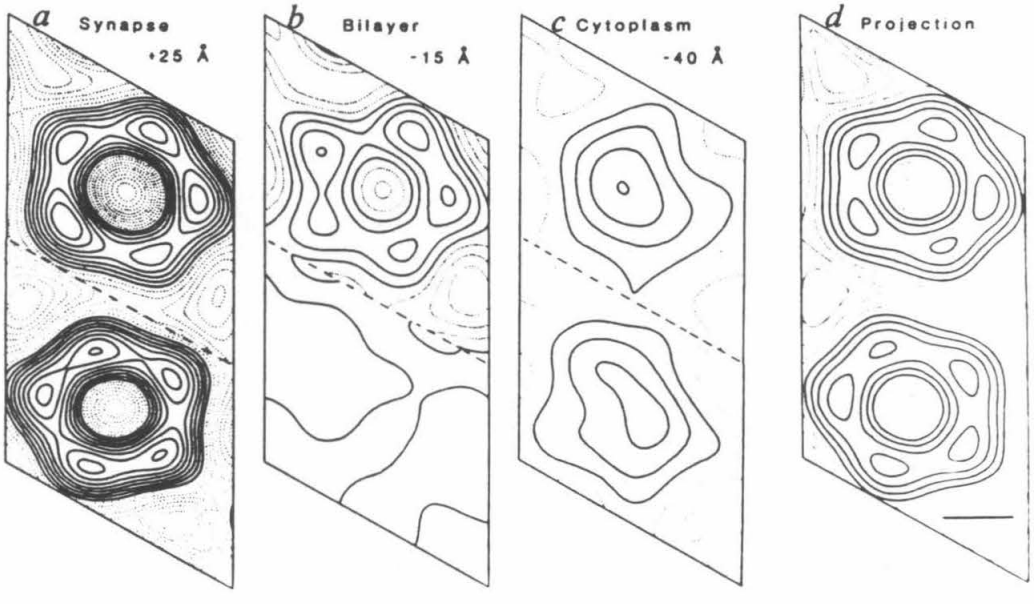
B. Cross sections of the receptor, which are parallel to the plane of the bilayer. The upper half of diagrams *a-c* represents frozen, unstained membranes and the lower half depicts the contours of receptors stained with phosphotungstate. *a*) 25 **A** above the synaptic surface of the membrane. *b*) 15 **A** into the membrane from its synaptic surface. *c*) the cytoplasmic surface of the membrane. *d*) projection of two cross-linked receptor molecules. The scale bar (bottom right) corresponds to 25 **A**.

This figure was reproduced from reference 11.

A.


synapse
cytoplasm

B.


a Synapse
+25 Å
b Bilayer
-15 Å
c Cytoplasm
-40 Å
d Projection

sensory rhodopsins (27,28). There is, however, apparently no sequence homology between the two ion pumps and rhodopsin (30), although the overall structures of bacteriorhodopsin and rhodopsin have been shown to be virtually identical (31). Because there is some question about the inclusion of bacteriorhodopsin and halorhodopsin in this superfamily on the basis of sequence homology and because more information is available on the other more closely related receptors, these two proteins will not be examined in detail in this work. Rhodopsin, the β-adrenergic receptor, and the muscarinic receptors will be examined, however, and used as test proteins because they are well-characterized membrane proteins.

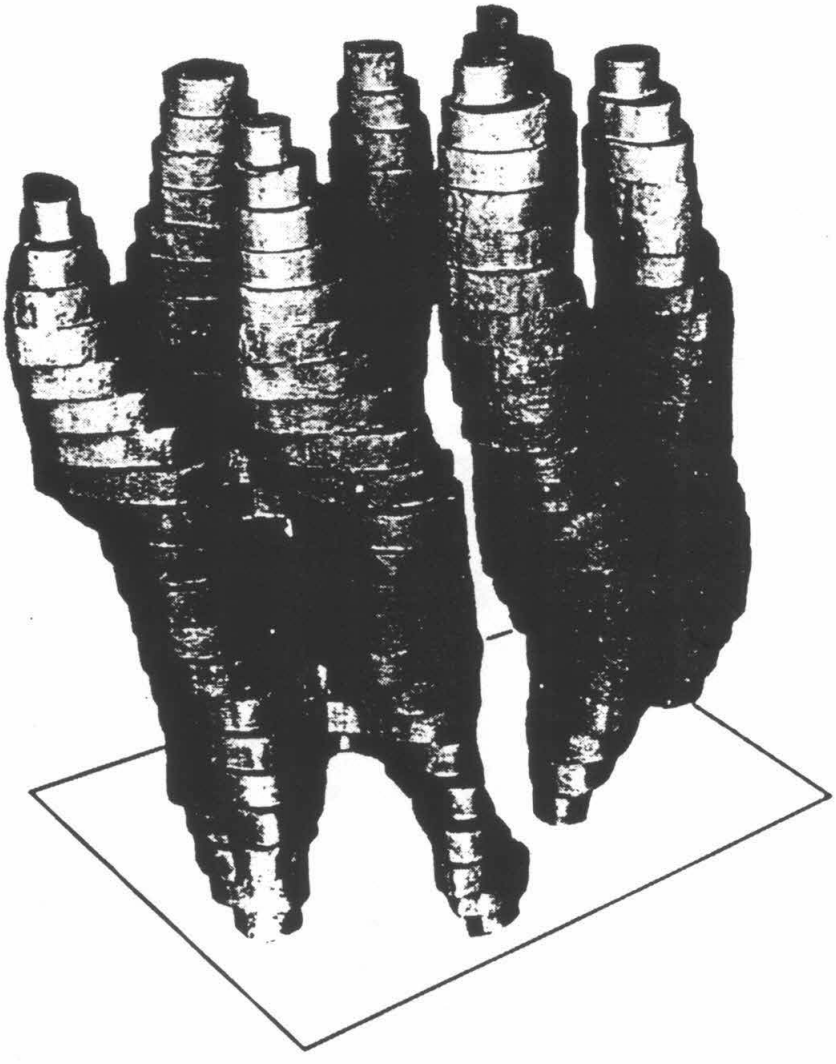
All of the above-mentioned proteins are believed to span the membrane seven times based upon sequence analysis, proteolytic information, and, most definitively, the low resolution crystal structures reported for bacteriorhodopsin (10) and rhodopsin (31). This structure is shown in figure 1.4. Rhodopsins and the two pumps both have a molecule of retinal covalently attached to a lysine residue located on one of the transmembrane helices (26). This cofactor allows these proteins to respond to activation by light. In contrast, the muscarinic acetylcholine receptor and the adrenergic receptors respond to the binding of agonist, acetylcholine in the case of the former, and epinephrine or norepinephrine in the case of the latter. Recent evidence indicates that the binding of these agonists occurs within the transmembrane region (32,33), implying an important role for the transmembrane region in signal induction for each of these proteins.

Figure 1.4

Structure of Bacteriorhodopsin

This balsa wood model of bacteriorhodopsin was derived from electron density maps obtained through x-ray diffraction studies. The protein is viewed roughly parallel to the plane of the membrane.

This figure was reproduced from reference 10.



The response invoked by all the above proteins, with the exception of bacteriorhodopsin and halorhodopsin, is mediated by one of the G proteins. G proteins are a class of GTP-binding proteins that may act to regulate secondary messengers like cAMP and cGMP, activate the phosphatidyl inositol cycle, or interact directly with certain ion channels. In general, when members of this receptor superfamily are activated, conformational changes allow the binding and subsequent activation of the corresponding G-proteins. One other feature that these proteins all have in common is that each can be phosphorylated in the region of its C-terminus, although the functional importance of this modification is not known at present (27).

Band 3

Band 3 is the final protein to be examined in this work and it is the most complex in many ways. Band 3 is the 95 kDa anion transporter found in the membrane of red blood cells. As described above, it effects the electroneutral exchange of anions across the membrane. Although it is selective for monovalent anions, divalent anions can also be translocated at a greatly reduced rate. It has been postulated that a proton is co-transported with these dianions (41). Physiologically, band 3 exchanges chloride and bicarbonate. This, in effect, increases the carbon dioxide carrying capacity of blood by preventing the build up of bicarbonate within the red cell. Band 3 also has several other functions within the cell. A part of the protein binds to cytoskeletal proteins to provide an important linkage between the cell membrane and cytoskeleton (36). Several cytoplasmic

proteins also bind to band 3 and include aldolase (37), glyceraldehyde-3-dehydrogenase (38), phosphofructokinase (39), and hemoglobin (40). In the membrane, band 3 exists as a dimer or tetramer (43), although the monomer is the unit of transport (44).

Recently, several excellent reviews of this protein's structure have been published (34,35). In general, band 3 can be viewed as having two distinct domains. The N-terminal 42 kDa is located in the cytoplasm and is the section of the protein which binds the cytoskeleton and the cytoplasmic proteins mentioned above. The C-terminal half of the protein contains the anion transport mechanism. This region of the protein has been proposed to span the membrane from 10-12 times. An even number of transmembrane spans is required because both the C- and N-termini are located within the cell (42). Band 3 is also glycosylated in this transmembrane region. All information about the topology of this protein has come from proteolytic or chemical labelling experiments. No x-ray or electron diffraction structures are available at this time.

References

1. Brasseur, R., Killian, J.A., De Kruijff, B., and Ruysschaert, J.M.(1987) *Biochim. Biophys. Acta* **903**, 11-17.
2. Kleffel, B., Garavito, R.M., Baumeister, W., and Rosenbusch, J.P.(1985) *EMBO J.* **4**, 1589-1592.
3. Schulten, Z., and Schulten, K.(1986) *Methods Enzymol.* **127**, 419-438.
4. Eisenman, G., and Dani, J.A.(1987) *Ann. Rev. Biophys. Biophys. Chem.* **16**, 205-226.
5. Stroud, R.M., and Finer-Moore, J.(1985) *Ann. Rev. Cell Biol.* **1**, 317-351.
6. Falke, J.J., Pace, R.J., and Chan, S.I.(1984) *J. Biol. Chem.* **259**, 6481-6491.
7. Gunn, R.B., and Frölich, O.(1979) *J. Gen. Physiol.* **74**, 351-374.
8. Diesenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H.(1985) *Nature* **318**, 618-624.
9. Allen, J.P., Feher, G., Yeates, T.O., Komiya, H., and Rees, D.C.(1987) *Proc. Natl. Acad. Sci. USA* **84**, 6162-6166.
10. Henderson, R., and Unwin, P.N.T.(1975) *Nature* **257**, 28-32.
11. Brisson, A., and Unwin, P.N.T.(1985) *Nature* **315**, 474-477.
12. Eisenberg, D.(1984) *Ann. Rev. Biochem.* **5**, 595-623.
13. Chou, P.Y., and Fasman, G.D.(1974) *Biochemistry* **13**, 222-245.
14. Yeates, T.O., Komiya, H., Rees, D.C., Allen, J.P., and Feher, G.(1987) *Proc. Natl. Acad. Sci. USA* **84**, 6438-6442.
15. Williams, J.C., Steiner, L.A., and Feher, G.(1986) *Proteins* **1**, 312-325.
16. Whiting, P.J., and Lindstrom, J.M.(1986) *Biochemistry* **25**, 2082-2093.
17. Whiting, P., and Lindstrom, J.(1987) *FEBS Lett.* **213**, 55-60.

18. Goldman, D., Deneris, E., Luyten, W., Kochhar, A., Patrick, J., and Heinemann, S.(1987) *Cell* **48**, 965-973.

19. Kurosaki, T., Fukuda, K., Konno, T., Mori, Y., Tanaka, K., Mishina, M., and Numa, S.(1987) *FEBS Lett.* **214**, 253-258.

20. McCrea, P.D., Popot, J.-L., and Engelman, D.M.(1987) *EMBO J.* **6**, 3819-3526.

21. Schofield, P.R., Darlison, M.G., Fujita, N., Burt, D.R., Stephenson, F.A., Rodriquez, H., Rhee, L.M., Ramachandran, J., Reale, V., Glencorse, T.A., Seeburg, P.H., and Barnard, E.A.(1987) *Nature* **328**, 221-227.

22. Barnard, E.A., Darlison, M.G., and Seeburg, P.(1987) *Trends. Neurosci.* 10, 502-509.

23. Grenningloh, G., Rienitz, A., Schmitt, B., Methfessel, C., Zensen, M., Beyreuther, K., Gundelfinger, E.D., and Betz, H.(1987) *Nature* **328**, 215-220.

24. Mamalaki, C., Stephenson, F.A, and Barnard, E.A.(1987) *EMBO J.* **6**, 561-565.

25. Pfeiffer, F., Simler, R., Grenningloh, G., and Betz, H.(1984) *Proc. Natl. Acad. Sci. USA* **81**, 7224-7227.

26. Ovchinikov, Y.A.(1982) *FEBS Lett.* **148**, 179-191.

27. Hall, Z.W.(1987) *Trends Neurosci.* **10**, 99-101.

28. Dohlman, H.G., Caron, M.G., and Lefkowitz, R.J.(1987) *Biochemistry* 26, 2657-2664.

29. Blanck, A., and Oesterhelt, D.(1987) *EMBO J.* **6**, 265-273.

30. Dunn, R., McCoy, G., Simsek, M., Mafumdar, A., Chang, S.H., Rajbhandary, U.L., and Khorana, H.G.(1981) *Proc. Natl. Acad. Sci. USA* **78**, 6744-6748.

31. Corless, J.M., McCaslin, D.R., and Scott, B.L.(1982) *Proc. Natl. Acad. Sci. USA* **79**, 116-120.

32. Yarden, Y., Rodriguez, H., Wong, S.K-F., Brandt, D.R., May, D.C., Burnier, J., Harkins, R.N., Chen, E.Y., Ramachandran, J., Ullrich, A., and Ross, E.M.(1986) *Proc. Natl. Acad. Sci. USA* **83**, 6795-6799.

33. Ratnam, M., Nguyen, D.L., Rivier, J., Sargent, P.B., and Lindstrom, J. (1986) *Biochemistry* **25**, 2633-2643.

34. Passow, H.(1986) *Rev. Physiol. Biochem. Pharmacol.* **103**, 61-223.

35. Jay, D., and Cantley, L.(1986) *Ann. Rev. Biochem.* **55**, 511-538.

36. Bennett, V., and Stenbuck, P.J.(1980) *J. Biol. Chem.* **255**, 6424-6432.

37. Murthy, S.N.P., Liu, T., Kaul, R.K., Köhler, H., and Steck, T.L.(1981) *J. Biol. Chem.* **365**, 9-17.

38. Tsai, I.H., Murthy, S.P., and Steck, T.L.(1982) *J. Biol. Chem.* **257**, 1438-1442.

39. Higashi, T., Richards, C.S., and Uyeda, K.(1979) *J. Biol. Chem.* **254**, 9542-9550.

40. Salhany, J.M., Cordes, K.A., and Gaines, E.D.(1980) *Biochemistry* **19**, 1447-1454.

41. Jennings, M.L.(1976) *J. Membr. Biol.* **28**, 187-205.

42. Lieberman, D.M., Natriss, M., and Reithmeier, R.A.F.(1987) *Biochim. Biophys. Acta* **903**, 37-47.

43. Macara, I.G. and Canley, L.C. (1983) *Cell Membr.* **1**, 41-87.

44. Knauf, P.A.(1979) *Curr. Top. Membr. Transp.* **12**, 249-363.

CHAPTER 2

TRANSMEMBRANE REGIONS OF INTEGRAL MEMBRANE PROTEINS

2A. Introduction: General Views of Transmembrane Segments

Now that the sequences of many transmembrane proteins have been determined, better methods for the identification of transmembrane helices from a knowledge of the primary structure are being sought. Toward that end, the characteristics of peptides which can be determined from their amino acid sequences are being examined with regard to the accurate prediction of transmembrane regions.

Some proposed characteristics of transmembrane helices have their basis in the thermodynamic stability of the helix once it is located in the membrane. Because membranes have a hydrophobic central region, transmembrane helices are believed to consist mainly of hydrophobic residues. The ends of these helices are probably in contact with the polar lipid head groups so their termini are thought to be less hydrophobic. In addition, the dipole moment of the head group region is oriented such that positive charges at the membrane surface should be especially stable (1). Positive charges have in fact been reported to terminate putative transmembrane segments for several proteins (2,3). The presence of a proline residue in the center of a helix has been proposed to improve interhelix packing due to an induced bend in the helix. These prolines have also been postulated to play a role in conformational transitions in transport proteins (4). The location of prolines at helix termini may also be indicative of a change from the α-helical structure of the transmembrane region to the less

regular structure of the extramembranous loops.

Some determinants of transmembrane topography have been proposed based on theories of membrane insertion. For instance, Engelman *et al.* (5) have proposed a prediction method based on the peptide's energy of insertion. The regions of a peptide which lie within the membrane are those that, when folded into an α-helix, have the lowest transfer energies in going from aqueous solution into the membrane environment. A similar suggestion was made by von Heijne, who characterized proteins of the bacterial inner membrane with regard to the charge distribution across the membrane (6). It was observed that for a set of well-characterized proteins, the cytoplasmic loops between transmembrane helices contained four times more basic residues than were found in the periplasmic loops. The distribution of negatively charged residues did not show this asymmetry. It was postulated that the activation energy to transfer the positive charges of the cytoplasmic loops past the dipolar headgroups would be high enough to prevent their translocation.

One final mechanism of helix insertion will be mentioned. Singer *et al.* (7) have proposed that transmembrane proteins can be inserted through the same mechanism as has been proposed for the secretion of hydrophilic proteins. This insertion would be mediated by special integral membrane proteins termed translocator proteins, which would provide an aqueous channel for the low-energy movement of the hydrophilic loops across the membrane as well as a pathway for the hydrophobic, transmembrane domains to exit from the translocator's channel into the bilayer. To date, such translocator proteins have

only been identified in the rough endoplasmic reticulum of eucaryotic cells (8), but the existence of similar proteins has been suggested for bacteria (9). Because this mechanism is highly speculative and the translocator protein virtually uncharacterized, the properties of a polypeptide chain, which would cause it to be transferred to the lipid environment, cannot be predicted. The determinants of peptide insertion into the membrane according to this model are protein-protein interactions and not protein-lipid interactions.

In the cases of several membrane proteins where more than one sequence is known, it has been noted that putative membrane regions have been better conserved than other parts of the protein (10,11). Whether this is a general characteristic of transmembrane helices suitable for aiding the identification of transmembrane regions of proteins has yet to be established. Another observation, which suggests that transmembrane regions are somewhat protected from evolutionary change, is that the genetic codings for transmembrane regions are rarely interrupted by introns (12).

In the following sections, some of these hypotheses about the general characteristics of transmembrane proteins will be examined, with a special emphasis on those which may be useful for the location of transmembrane helices within the peptide sequence. Several methods proposed for the assignment of transmembrane helices will also be evaluated.

Assignment of Transmembrane Helices

To date, many methods have been proposed for the identification of

transmembrane helices from a knowledge of the protein's amino acid sequence. By far, the most popular technique is the hydropathic analysis first proposed by Kyte and Doolittle (13). In this method, the hydrophobicity of the sequence is calculated and plotted as a function of sequence position. Many different hydrophobicity scales have been proposed for the calculation of sequence hydrophobicity and they are reviewed in reference 14. The ultimate goal of the hydropathic analysis is to identify hydrophobic regions which are long enough to traverse the hydrophobic center of the bilayer. In most analyses, these regions are determined by inspection, although a more quantitative method has been developed by Eisenberg *et al.* (15). In this method, the helices selected as transmembrane are 21 residues long and have an average hydrophobicity above a certain threshold. A criterion is then used to select between any overlapping regions. The putative transmembrane regions are further characterized in this procedure on the basis of their amphipathic nature, an examination which also helps to locate the segment with respect to the membrane. Regions that are both hydrophobic and amphipathic are suspect as transmembrane regions because they may actually lie along the surface of the membrane instead of spanning it. The Eisenberg method includes a procedure called the hydrophobic moment analysis, which allows protein segments to be characterized on the basis of their hydrophobicity and amphiphilicity. This procedure will be discussed in more detail in the next section. The reliability of hydrophobicity for the identification of transmembrane helices is questionable in the case of channel proteins. Channel-lining helices are probably amphiphilic,

implying that up to 50% of the residues could be non-hydrophobic. Such a region might not be detected in an analysis based solely on hydrophobicity.

Prediction of the secondary structure of transmembrane proteins has generally been unsuccessful, except perhaps in predicting the conformation of extramembranous loops (16).

In this work, several of these methods will be applied to the reaction center proteins to assess their value in predicting the locations of transmembrane helices within the primary sequence of a protein. From the results of these analyses, criteria for the assignment of transmembrane helices are developed and applied to the acetylcholine receptor, rhodopsin, and band 3. Athough the use of only one membrane protein as a guide for general protein behavior is not ideal, some new conclusions about the nature of transmembrane helices can be made.

The properties relating to the location of transmembrane regions which will be examined in this work include hydrophobicity, hydrophobic moment, secondary structure, overall variability, and charge and proline distributions. In addition, the proposal that the genetic coding of membrane-spanning regions is not interrupted by introns will be examined. The reliability of these methods will be improved by using the average characteristics of aligned, homologous sequences. Such averaging masks non-conserved properties of individual proteins.

2B. The Reaction Center

The reaction center is the only intrinsic membrane protein for which the crystal structure is known to high resolution. This detailed structural information has permitted the transmembrane regions of the primary sequence to be identified to within one or two residues (17). At present, therefore, the reaction center is the best protein on which to test predictive algorithms for their accuracy in identifying the exact locations of transmembrane helices.

Methods

Sequences were aligned with the aid of published alignments, where indicated, or by inspection. Ambiguous regions were aligned with the aid of the Needleman and Wunch algorithm (18). This method of sequence alignment maximizes the similarity between two sequences by inserting gaps where necessary. A penalty is assigned for inserting the first gap, and a different, usually lesser, penalty is assigned for each successive gap. The penalties used in these alignments were generally 1.0 and 0.3, respectively.

The hydrophobic analysis was done using the method and scale of Kyte and Doolittle (19). A seven-residue averaging window was generally used for this analysis. This hydrophobicity scale was chosen because the hydrophobic regions were well-defined and easily distinguishable from the hydrophilic regions.

The hydrophobic moment analysis of Eisenberg was applied to the determination of transmembrane helices as originally described (15).

The Eisenberg hydrophobicity scale (15) was used to assess the average hydrophobicity at each residue position of the aligned sequences. Possible transmembrane regions were identified, which were 21 residues long and had an average hydrophobicity greater than or equal to .42. The best non-overlapping regions were selected on the basis of their hydrophobicities, and these putative membrane helices were then characterized according to their amphiphilicity. The amphiphilicity is quantitated using a parameter called the average hydrophobic moment, which is defined below:

$$\mu_H = \frac{\left\{ \left[\sum_{n=1}^{N} H_n \sin(\delta n) \right]^2 + \left[\sum_{n=1}^{N} H_n \cos(\delta n) \right]^2 \right\}^{1/2}}{N}$$

where H_n is the hydrophobicity of the nth residue, N is the helix length, and δ is the angle between successive side chains. The angle that corresponds to a regular α-helix is 100° and this value of δ was used in these analyses. The higher the value calculated for the hydrophobic moment, the more amphipathic the peptide. The characterization of the transmembrane helices, once identified, was done by plotting their hydrophobic moments versus their average hydrophobicities. Different regions of this type of plot were tentatively assigned by Eisenberg to different locales in relation to the bilayer, *i.e.*, globular (extramembranous), surface, and transmembrane. These regions can be seen in figure 2.4. The two transmembrane regions differentiate between those helices which are multimeric, or likely to associate with other helices in the membrane, and those

which are monomeric. The monomeric helices are generally more hydrophobic and have lower hydrophobic moments. Although in his analyses, Eisenberg selected on the most amphipathic segment of each helix to characterize using these plots, the average for the complete 21-residue helix has been used in this work.

The prediction of secondary structure was done according to Chou and Fasman (20). The α-helix, β-sheet, and β-turn potentials were calculated for each position in the primary structure.

The variability analysis used to examine the proteins was that proposed by Wu and Kabat (21). This method takes into account both the variety of amino acids found at a particular position and the distribution of residues among these amino acid types. The expression for the variability is given below:

$$\text{Variability} = \frac{\text{number of different amino acids at a given position}}{\text{frequency of the most common amino acid at that position}}$$

A seven-residue smoothing window was used in this analysis.

This work was done using the *DNA Master* program in the laboratory of Professor Leroy Hood at the California Institute of Technology.

Results

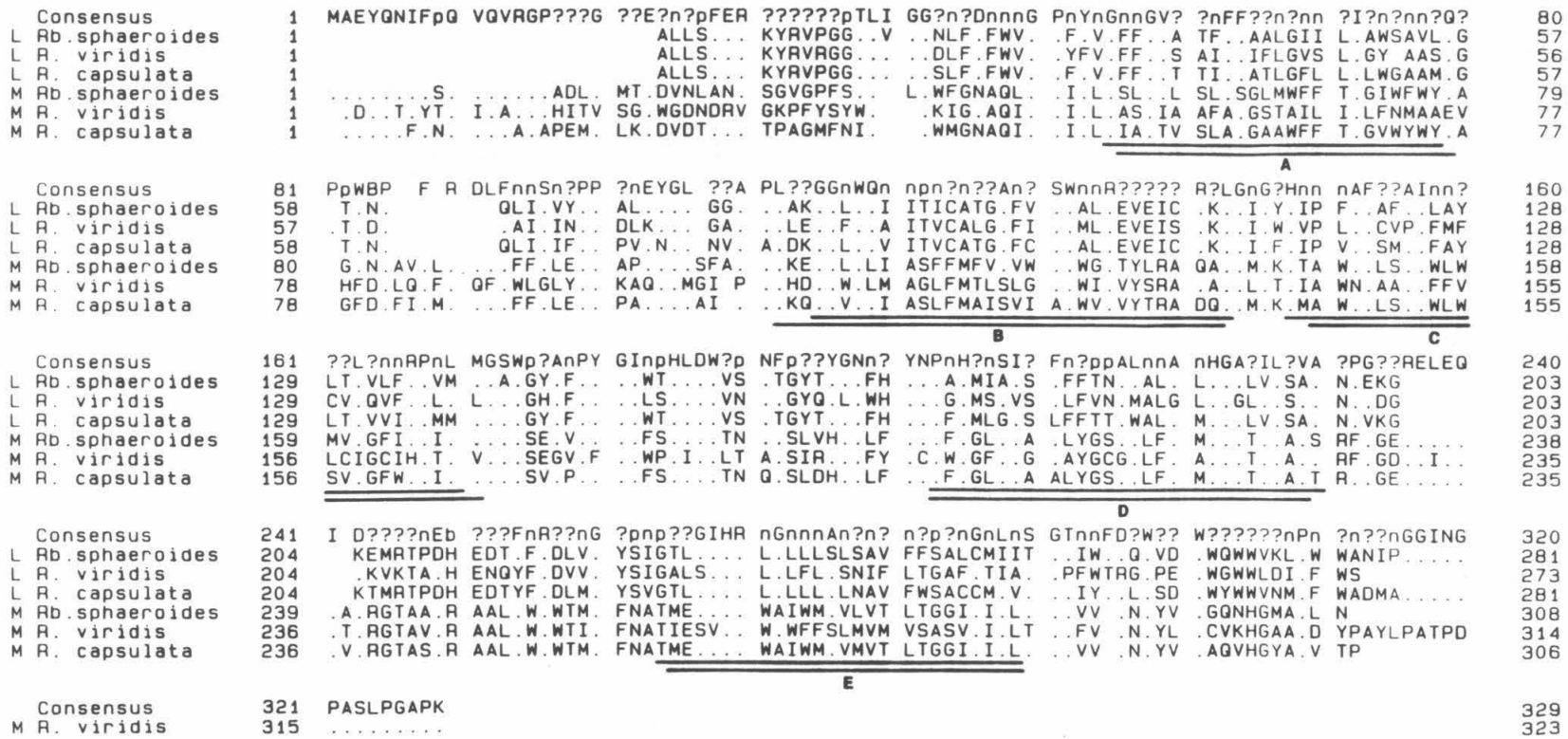
The alignment of the proteins used for the analysis of the reaction center is shown in figure 2.1. Sequences of both the L and M subunits were included because they are homologous. The actual positions of the helices as defined by Yeates *et al.* (17) are indicated in the figure. In general, the transmembrane regions are

Figure 2.1

Alignment of the Reaction Center Sequences

The positions of transmembrane helices as determined by Yeates *et al.* (17) have been indicated by heavy bars under the corresponding residues in the alignment. The upper bars indicate the transmembrane helices of the L subunit and the lower bars the transmembrane regions of the M subunit. The reaction center sequences in the alignment include the L and M subunits of the following microorganisms: *Rhodobacter sphaeroides* (22,23), *Rhodopseudomonas viridis* (24), and *Rhodopseudomonas capsulata* (25).

Reaction Center: L and M Subunits



aligned for the L and M subunits, confirming that these proteins can be treated as a homogeneous group. The helix ends for the two subunit types generally differ by only one or two residues. The largest discrepency between them is at the N-terminus of helix B, where the M subunit has three more residues than the L subunit. Because the MB helix has an irregular turn at the membrane surface, a case could be made for exclusion of the first few residues, which probably lie along the very surface of the membrane and interact with the lipid headgroups.

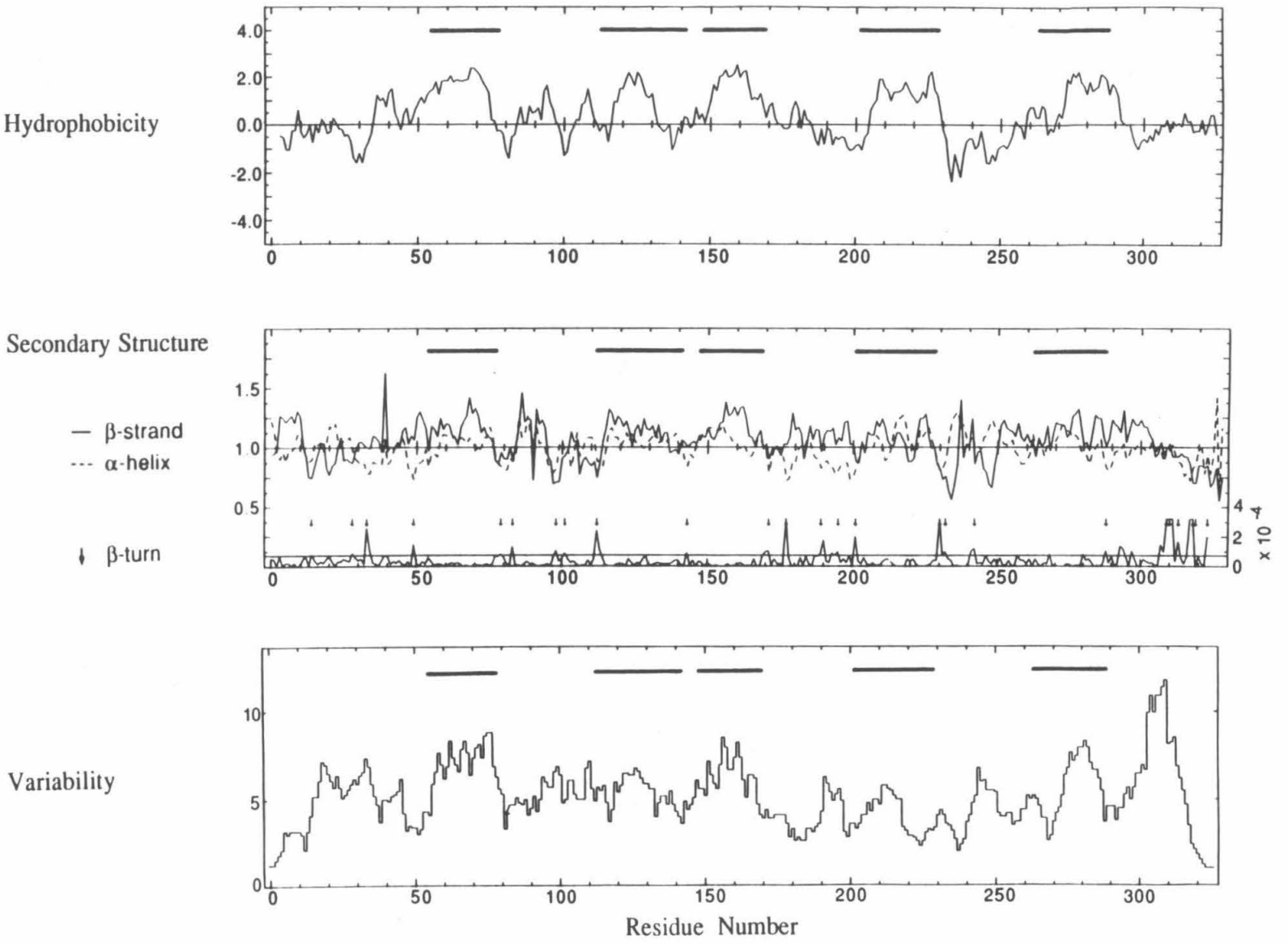
Hydrophobicity. The hydropathic plot is shown in figure 2.2. As expected, five hydrophobic domains are evident, each in a region designated as membrane-spanning. The shortest hydrophobic region corresponding to a transmembrane span is that of helix B, which is 16 residues long. This observation is of interest for determining the minimum hydrophobic length required for a transmembrane location. If a rise per residue of 1.5 Å is assumed, the length of the hydrophobic region is only 24 Å long. This value is about 5 Å shorter than the length of the hydrophobic region of the membrane, which was estimated to be about 29-30 Å (17). The longest hydrophobic stretch is that of helix D, which is about 24 residues long.

It is interesting to note that the helix ends do not correspond well to the edges of the hydrophobic regions. Only about 60% of the helix ends fall within 2 residues of the edge of a hydrophobic domain. If the protein traversing the headgroup region is more hydrophilic and the length of this headgroup region is about 5.5 Å (26,27), three to four residues on either end of the helix could be expected to be non-

Figure 2.2

Hydropathy, Secondary Structure Potential, and Variability Plots for the Reaction Center Protein

The hydropathy, secondary structure potential, and sequence variability of the aligned sequences are shown as a function of sequence position. A smoothing window of seven residues was used for all plots. Heavy bars indicate the positions of transmembrane helices with the upper bar corresponding to the L subunit and the lower to the M subunit.


Hydrophobicity
4.0
2.0
0.0
-2.0
-4.0
0
50
100
150
200
250
300
Secondary Structure
— β-strand
--- α-helix
β-turn
1.5
1.0
0.5
4
2
0
x 10 -4
Variability
10
5
0
Residue Number

hydrophobic. Even allowing for this region, the C-terminal region of helix B is more hydrophilic than anticipated, perhaps relating to the fact that this region of helix B interacts with a cytoplasmic loop, which partially inserts into the membranous region (28). A comparison of the average hydophobicity of the 155 buried residues to that of the 106 lipid-exposed residues showed that overall, both were hydrophobic, but the buried residues were relatively more hydrophilic. The hydrophobicity values according to the scale of Kyte and Doolittle (12) were 1.1 and 1.9 for the buried and exposed residues, respectively. Another unexpected observation is that the C-terminus of helix E is located five residues before the end of the hydrophobic domain.

The length of the hydophobic domain also does not correlate well to the length of the helix. Theoretically, as the length of the helix increases to allow for a greater inclination with respect to the bilayer normal or to allow a smaller radius of curvature, the length of the hydrophobic domain should also increase. In figure 2.3, the length of the hydrophobic stretch found within the actual helix is plotted against the number of helix residues. Discrepancies may result from helix-helix interactions, as mentioned above, although other explanations are possible.

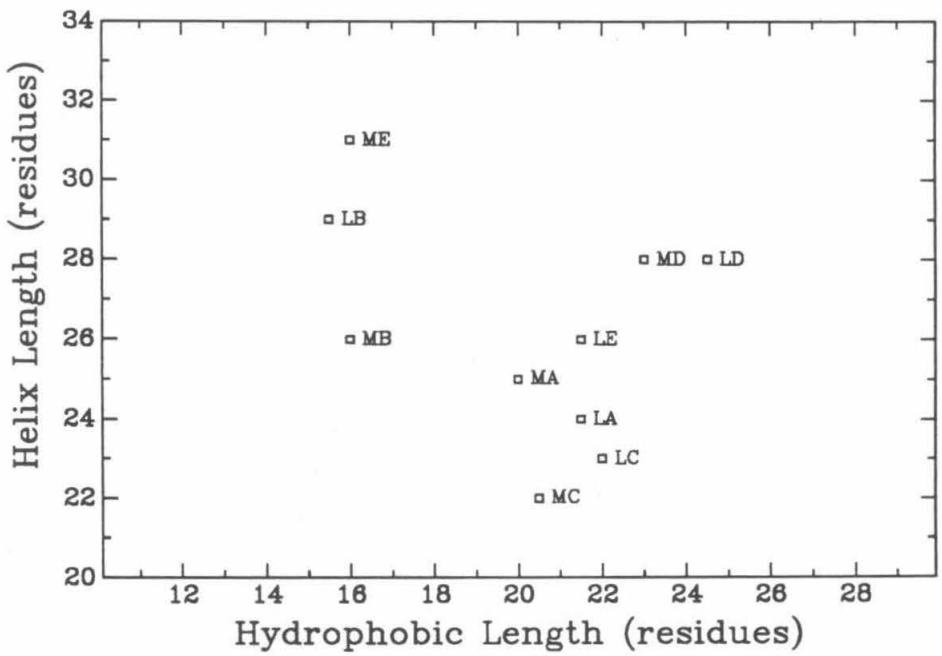
Hydrophobic Moment. The helices predicted to be transmembrane by the Eisenberg criteria corresponded to the most hydrophobic, 21-residue segment of each hydrophobic domain. All of the actual transmembrane regions of the reaction center were identified using the Eisenberg algorithm. This method did, however, predict an erroneous

Figure 2.3

Helix Length *vs.* Hydrophobic Length

The helix lengths as defined by Yeates, *et al.* (17) have been plotted *versus* the length of its hydrophobic domain as determined from the hydropathy plots. No line could be fitted to these points.

Helix Length vs. Hydrophobic Length



helix corresponding to position 169 of the alignment shown in figure 2.1. This position corresponds to residues 137 and 167 of the *Rb. sphaeroides* L and M chains, respectively. According to Allen *et al.* (28), half of this region corresponds to an amphiphathic surface helix, which might account for its selection by the algorithm. When the Eisenberg diagram, shown in figure 2.4, is examined, the average hydrophobicity of the erroneous helix maps to the globular section. Although this type of diagram was originally used by Eisenberg to characterize only the most amphiphilic 11-residue segment of each helix, his derived limits may be useful in helix identification and discrimination. All other selected helices map to the transmembrane regions and none map to the surface region. Helices A, C, and perhaps F map to the region of the plot that designates solitary helices or signal sequences, although the significance of this in relation to multiple-span proteins like the reaction center is unclear.

Secondary Structure. As previously noted, the prediction of secondary structure for membrane proteins has generally not been successful. There are, however, several results of interest upon the application of the Chou-Fasman algorithms to the aligned sequences. The most obvious feature in the secondary structure analysis shown in figure 2.2 is that transmembrane regions are often strongly predicted to be β-sheet. This outcome results from the fact that the residues with the highest β-sheet potential according to the Chou-Fasman assignments are also the residues with the greatest hydrophobicity according to Kyte and Doolittle. An examination of the predicted β-turns shows that predicted turns are located at or within five re-

Figure 2.4

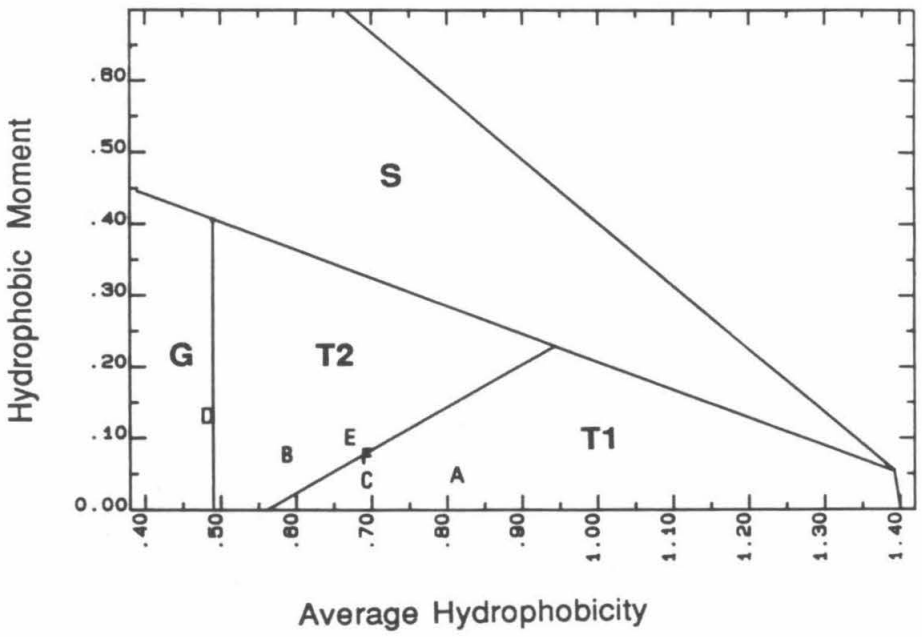
Hydrophobic Moment Plot of the Reaction Center

The average hydrophobic moment of each 21-residue helix selected by the Eisenberg algorithm (15) was plotted against its average hydrophobicity. The numerical helix positions refer to the numbering in the alignment of figure 2.1. Actual helices of the reaction center correspond to helices A, B, C, E, and F.

Symbols: G=gobular, S=surface, T1=monomeric transmembrane, T2=multimeric transmembrane.

Hydrophobic Moment Plot

Reaction Center



A = 54
B = 114
C = 146
D = 169
E = 209
F = 271

sidues of the helix ends for 9 out of the 10 positions. One of these turns occurs at the C-terminus of helix E and is located in a hydrophobic region. Perhaps the presence of a turn is a better indication of the helix end than the end of a hydrophobic domain. No turns were predicted within the helical regions when the aligned sequences were analyzed, although some were predicted if sequences were analyzed individually or in small groups.

Variability. The sequence variability of the transmembrane segments was found to be no less than that of the protein as a whole. This suggests that sequence conservation probably cannot be used as an indicator of transmembrane regions. Inspection of the variability plot shown in figure 2.2 shows that the most conserved helical regions are the C-terminus of helix D and the N-terminus of helix E. These regions are the most buried in the three-dimensional structure of the protein, suggesting that the variability of a helix may be related to its location. This possibility will be discussed in more detail later in this work.

Prolines and Basic Residues. The distribution of positively charged residues and prolines was also examined. Proline residues were found within five residues of 8 out of 10 helix termini, although they were not all conserved. This observation is probably related to the observed (and predicted) presence of turns at these helix ends. There seems to be no pattern in the distribution of positive charges except that, when in a transmembrane segment, they tend to be near the ends. The C-terminal region of helix B is again an exception, perhaps due to its interaction with the interhelix loop.

Genetic Structure. The DNA, which codes for the L and M subunits of *Rb. sphaeroides*, contains no introns in the transcribed region, as is expected for a prokaryote. The distantly related proteins of photosystem II do contain introns, however. These sequences were aligned to those of the reaction center and the positions of the introns were compared to the positions of the transmembrane helices. This alignment and the intronic positions are shown in figure 2.5. Of the eight intronic positions, two of them mapped within the membrane, one within helix A, and one within helix B. In the aligned sequences, the intron in helix A marked the beginning of a hydrophilic C-terminal region for the D1 protein, but the corresponding position in the reaction center is a continuation of the hydrophobic domain and six residues from the C-terminus. The intron in helix B is eight residues before the C-terminus and marks the beginning of a more hydrophilic region for both the proteins. This is also the region that interacts with the extramembranous loop. One intron also maps to the N-terminus of helix B.

Discussion

The number of transmembrane helices in this protein was indicated clearly by inspection of the hydropathic plot and by analysis using the Eisenberg algorithm. As a predictive procedure, the hydropathic analysis of Eisenberg is preferable for application to transport proteins because it is able to detect amphipathic helices.

The difficulty in accurately identifying helices is the determination of their endpoints. For instance, based on the hydrophobicity or

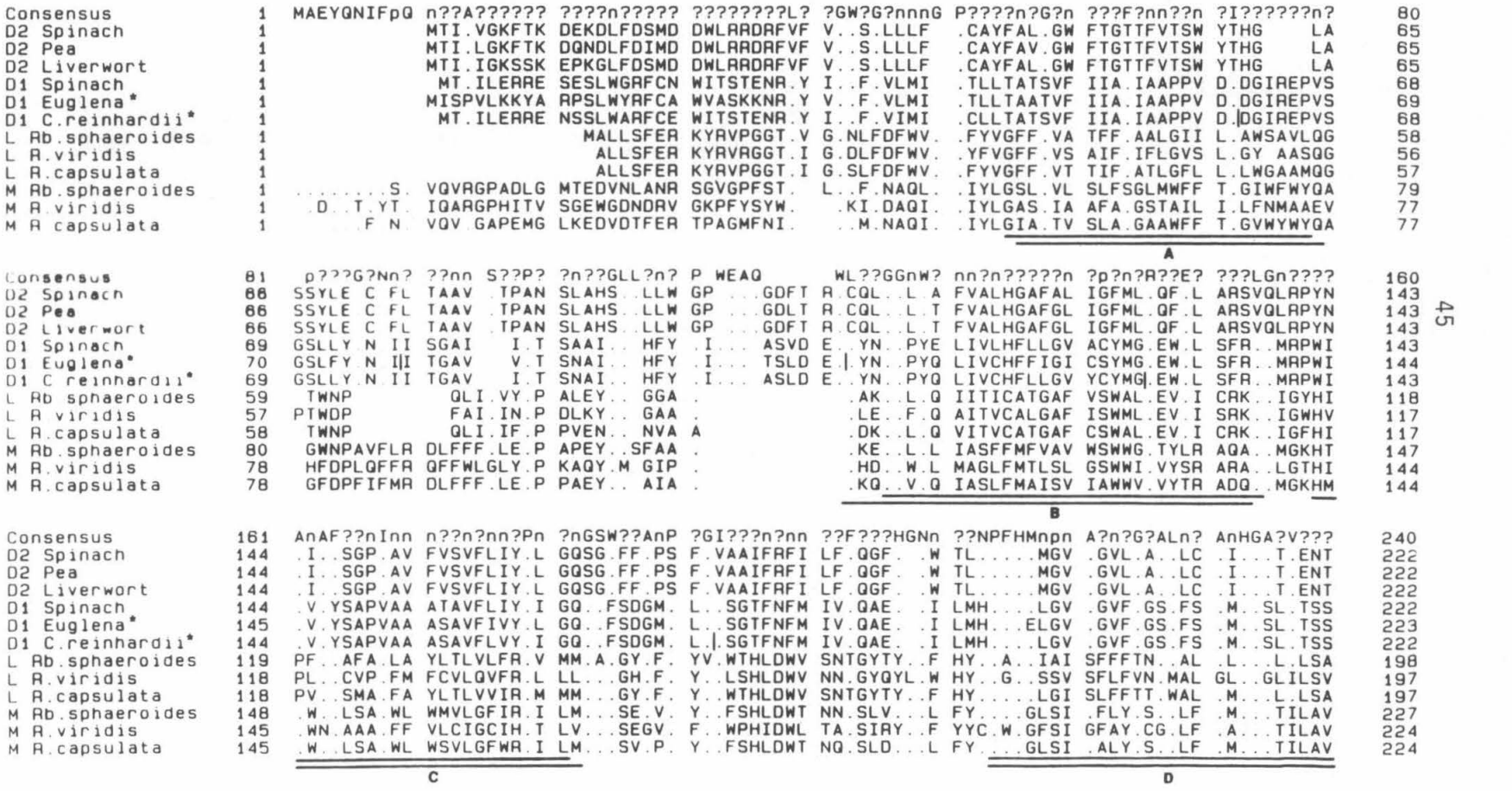
Figure 2.5

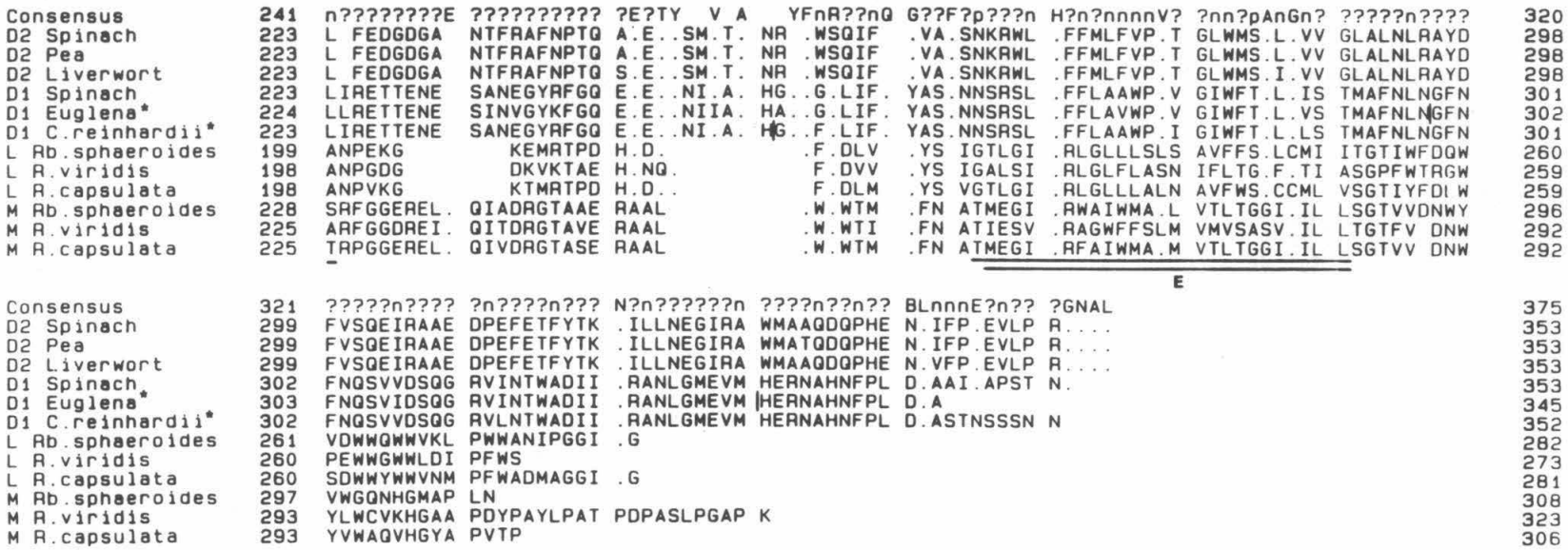
Sequence Alignment of the Reaction Center Family

The location of the reaction center's transmembrane helices are indicated in this alignment by heavy bars, the L subunit above the M. Sequences obtained from genomic DNA are marked with asterisks and the intronic positions within them are denoted by vertical lines between residues.

Sequences aligned: **D2** Spinach, *Spinachia oleracea* (29); Pea, *Pisum sativum* (30); Liverwort, *Marchantia polymorpha* (31); **D1** Spinach, *Spinachia oleracea* (32); Euglena, *Euglena gracili* (33); *Chlamydomonas reinhardii* (34); **LM** *Rhodobacter sphaeroides* (22,23); *Rhodopseudomonas viridis* (24); *Rhodopseudomonas capsulata* (25).

Reaction Center Family




Consensus 241 n????????E ?????????? ?E?TY V A YFnR??nQ G??F?p???n H?n?nnnnV? ?nn?pAnGn? ?????n???? 320
D2 Spinach 223 L FEDGDGA NTFRAFNPTQ A.E..SM.T. NR .WSQIF .VA.SNKRWL .FFMLFVP.T GLWMS.L.VV GLALNLRAYD 298
D2 Pea 223 L FEDGDGA NTFRAFNPTQ A.E..SM.T. NR .WSQIF .VA.SNKRWL .FFMLFVP.T GLWMS.L.VV GLALNLRAYD 298
D2 Liverwort 223 L FEDGDGA NTFRAFNPTQ S.E..SM.T. NR .WSQIF .VA.SNKRWL .FFMLFVP.T GLWMS.I.VV GLALNLRAYD 298
D1 Spinach 223 LIRETTENE SANEGYRFGQ E.E..NI.A. HG..G.LIF. YAS.NNSRSL .FFLAAWP.V GIWFT.L.IS TMAFNLNGFN 301
D1 Euglena* 224 LLRETTENE SINVGYKFGQ E.E..NIIA. HA..G.LIF. YAS.NNSRSL .FFLAVWP.V GIWFT.L.VS TMAFNLNGFN 302
D1 C.reinhardii* 223 LIRETTENE SANEGYRFGQ E.E..NI.A. HG..F.LIF. YAS.NNSRSL .FFLAAWP.I GIWFT.L.LS TMAFNLNGFN 301
L Rb.sphaeroides 199 ANPEKG KEMRTPD H.D. .F.DLV .YS IGTLGI .RLGLLLSLS AVFFS.LCMI ITGTIWFDQW 260
L R.viridis 198 ANPGDG DKVKTAE H.NQ. F.DVV .YS IGALSI .RLGLFLASN IFLTG.F.TI ASGPFWTRGW 259
L R.capsulata 198 ANPVKG KTMRTPD H.D.. F.DLM .YS VGTLGI .RLGLLLALN AVFWS.CCML VSGTIYFDIW 259
M Rb.sphaeroides 228 SRFGGEREL. QIADRGTAAE RAAL .W.WTM .FN ATMEGI .RWAIWMA.L VTLTGGI.IL LSGTVVDNWY 296
M R.viridis 225 ARFGGDREI. QITDRGTAVE RAAL .W.WTI .FN ATIESV .RAGWFFSLM VMVSASV.IL LTGTFV DNW 292
M R.capsulata 225 TRPGGEREL. QIVDRGTASE RAAL .W.WTM .FN ATMEGI .RFAIWMA.M VTLTGGI.IL LSGTVV DNW 292
E
Consensus 321 ?????n???? ?n????n??? N?n??????n ????n??n?? BLnnnE?n?? ?GNAL 375
D2 Spinach 299 FVSQEIRAAE DPEFETFYTK .ILLNEGIRA WMAAQDQPHE N.IFP.EVLP R.... 353
D2 Pea 299 FVSQEIRAAE DPEFETFYTK .ILLNEGIRA WMATQDQPHE N.IFP.EVLP R.... 353
D2 Liverwort 299 FVSQEIRAAE DPEFETFYTK .ILLNEGIRA WMAAQDQPHE N.VFP.EVLP R.... 353
D1 Spinach 302 FNQSVVDSQG RVINTWADII .RANLGMEVM HERNAHNFPL D.AAI.APST N. 353
D1 Euglena* 303 FNQSVIDSQG RVINTWADII .RANLGMEVM HERNAHNFPL D.A 345
D1 C.reinhardii* 302 FNQSVVDSQG RVLNTWADII .RANLGMEVM HERNAHNFPL D.ASTNSSSN N 352
L Rb.sphaeroides 261 VDWWQWWVKL PWWANIPGGI .G 282
L R.viridis 260 PEWWGWWLDI PFWS 273
L R.capsulata 260 SDWWYWWVNM PFWADMAGGI .G 281
M Rb.sphaeroides 297 VWGQNHGMAP LN 308
M R.viridis 293 YLWCVKHGAA PDYPAYLPAT PDPASLPGAP K 323
M R.capsulata 293 YVWAQVHGYA PVTP 306

the hydrophobic moment, the C-termini of both helix B and helix E would have been incorrectly predicted. The best parameter to use in the identification of likely helix termini, according to this study, is the location of predicted β-turns. The next best criterion is the presence of proline residues near the expected helix terminus. According to these data, the presence of basic residues at helix termini is not generally observed, although this possibility will be discussed again in the next section.

The location of introns within the primary sequence is probably not indicative of the actual helix termini, but may indicate the boundary of the region which actually determines the transmembrane location for that helix. This hypothesis predicts that the presence of an intron in the central ~13 residues of a transmembrane helix is highly unlikely. Any correlation between the locations of introns and the transmembrane sections of a peptide is tenuous at present and cannot be used in predictive schemes.

The variability observed for the transmembrane regions also suggests that these regions are not more conserved than the rest of the protein, and that a predictive scheme should not incorporate conservation as a predictive characteristic.

Based on the above observations, the method proposed for the prediction of transmembrane segments would consist of the following steps:

1. Identify possible transmembrane helices using the Eisenberg selection method.

2. Use β-turns as predicted by the Chou-Fasman method to identify probable helix ends. If no turns are predicted near the putative helix, proline residues, then basic residues (see next section) should be used to estimate the helix end's position.
3. After at least one helix end per transmembrane region has been located, the other end should be sought an average of 26 residues away, based on the average helix length of the reaction center subunits. Transmembrane helices are assumed to be between 20 and 33 residues in length, based upon the reaction center, and allowing for the possibility of thicker membranes or more tilted helices.
4. When none of the above criteria can locate a helix end, the position resulting in a 26-residue helix could be used.
5. The final step is to compare the putative helices to any known experimental evidence indicating protein topology.

These guidelines will be applied to the nicotinic acetylcholine receptor, rhodopsin, and band 3 in the following sections.

2C. The Nicotinic Acetylcholine Receptor

The criterion developed in the previous chapter was applied to the nicotinic acetylcholine receptor (AChR) in order to predict which regions of this protein lie within the membrane. Although some of the subunits have special roles in the binding of acetylcholine (35), for the purposes of this analysis all were assumed to have identical topologies with respect to their transmembrane regions. This treatment of the subunits as a homologous group is somewhat justified by their sequence homology and the experimental observation that some subunits can be exchanged without destroying channel activity (36).

In the current literature, models of the acetylcholine receptor with both four and five transmembrane helices per subunit have been proposed.

Methods

The methods used to analyze the sequences of the AChR subunits were as described in the last section. The criterion outlined in the discussion section was applied to the prediction of the AChR's transmembrane helices. First, the regions selected by the Eisenberg method were identified and the surrounding residue positions examined for predicted β-turns. Then, helix termini were assigned on the basis of these turns or defined by the presence of proline or basic residues.

Results

The sequence alignment used for the calculations is shown in figure 2.6. All residue numbers used in this discussion will refer to the numbering of the consensus sequence in this figure.

The first step toward assigning the transmembrane regions was to identify the helices predicted by the Eisenberg method. These helices included residues 231-251, 265-285, 298-318, and 493-513. The hydrophobic moment plot of these selected helices is shown in figure 2.7. None of the predicted helices fall within the surface or globular regions so all are predicted to be transmembrane regions of the protein.

The hydrophobicity, secondary structure potential, and variability profiles for the sequence alignment can be seen in figure 2.8. The four hydrophobic areas, which correspond to the predicted transmembrane regions, are evident in the hydropathy plot. Predicted β-turns were located near one end of each of the four helices, and again, none were found within the putative transmembrane regions. The locations of these predicted β-turns were used to assign one helix terminus for each of the four regions.

The other end of each helix was located through examination of the sequence composition in the regions which were about 25 amino acids from the known helix end. For two out of the four cases, a proline could be found to indicate the helix terminus. In the remaining two helices, the presence of conserved basic residues was observed in the regions near the likely helix termini and these basic residues were used to define the helix ends. Overall, there were conserved positive

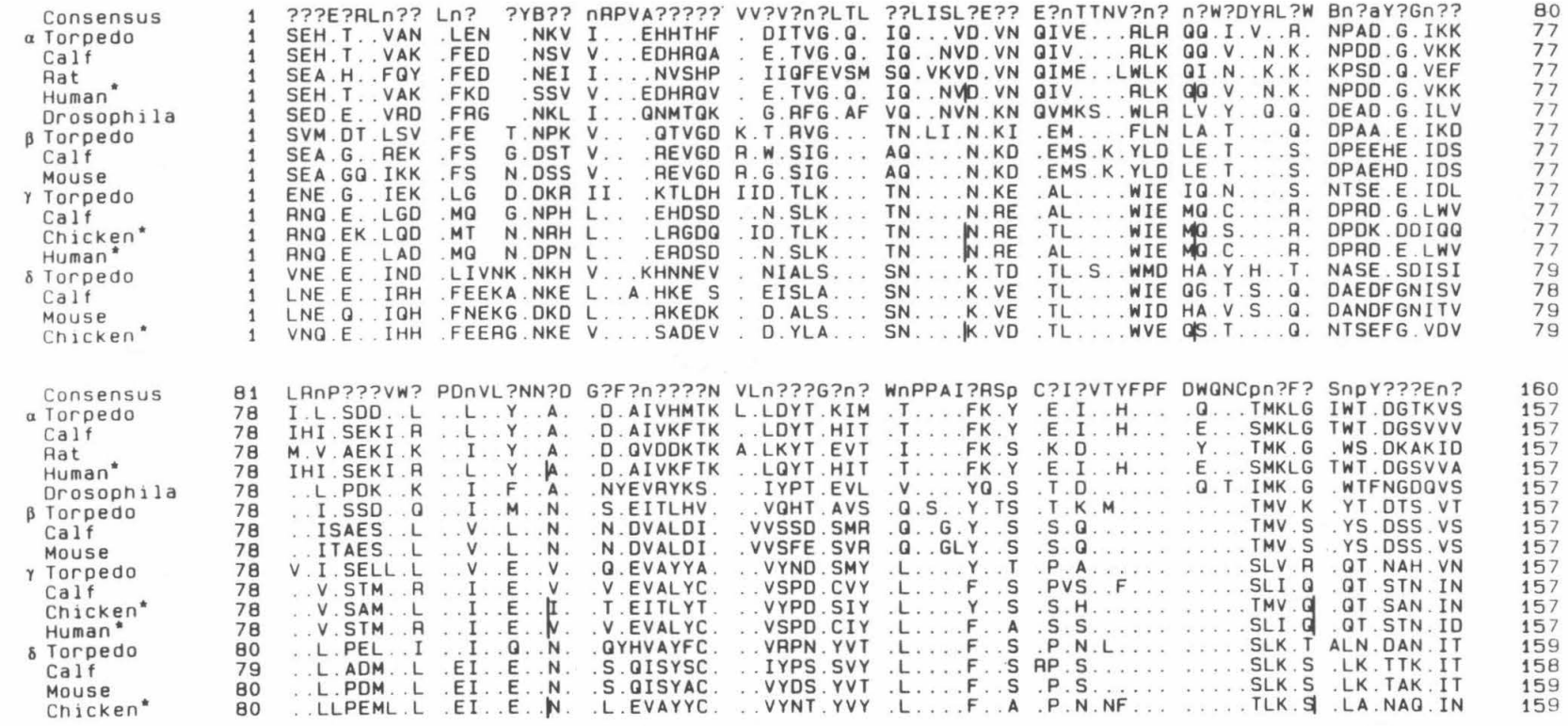
Figure 2.6

Sequence Alignment of the Acetylcholine Receptor Subunits

The sequences in this alignment with known genetic structures have been denoted using asterisks and the intronic positions within these sequences were indicated by vertical lines between residues. The predicted transmembrane regions have been underlined with heavy bars.

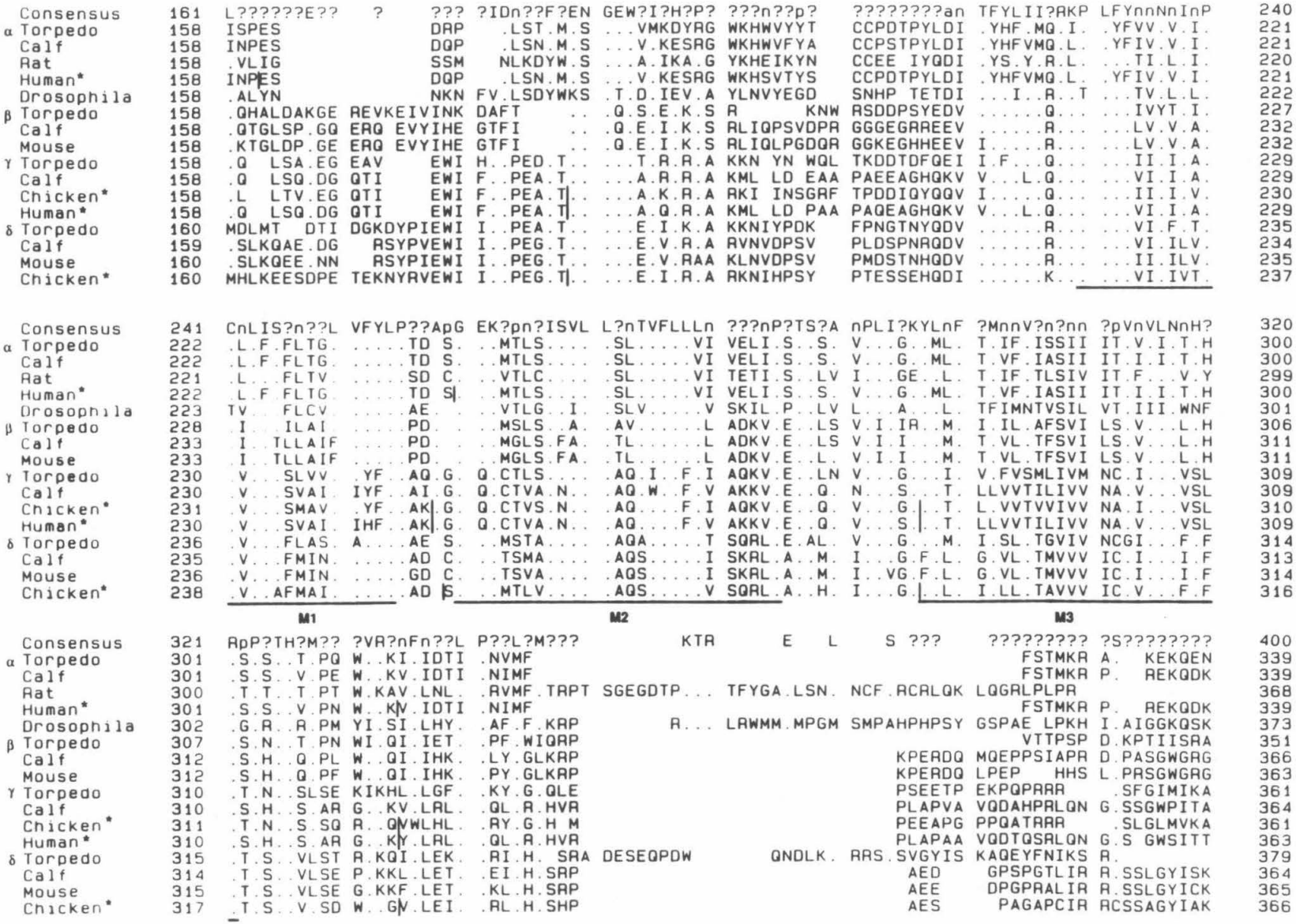
Sequences included: α *Torpedo californica* (37), Calf (38), Rat (39), Human (38), *Drosophila* (40); β *Torpedo californica* (37), Calf (41), Mouse (42); γ *Torpedo californica* (37), Calf (43), Chicken (44), Human (45); δ *Torpedo californica* (37), Calf (46), Mouse (47), Chicken (44).

Acetylcholine Receptor



```
                1 ???E?RLn?? Ln?  ?YB?? nRPVA????? VV?V?n?LTL ??LISL?E?? E?nTTNV?n? n?W?DYRL?W Bn?aY?Gn??   80
Consensus       1 SEH.T..VAN .LEN   .NKV I...EHHTHF . DITVG.Q. IQ...VD.VN QIVE...RLR QQ.I.V..R. NPAD.G.IKK   77
α Torpedo       1 SEH.T..VAK .FED   .NSV V...EDHRQA . E.TVG.Q. IQ..NVD.VN QIV....RLK QQ.V..N.K. NPDD.G.VKK   77
Calf            1 SEH.T..VAK .FED   .NSV V...EDHRQV . E.TVG.Q. IQ..NVD.VN QIV....RLK QQ.V..N.K. NPDD.G.VKK   77
Rat             1 SEA.H..FQY .FED   .NEI I....NVSHP . IIQFEVSM SQ.VKVD.VN QIME..LWLK QI.N..K.K. KPSD.Q.VEF   77
Human*          1 SEH.T..VAK .FKD   .SSV V...EDHRQV . E.TVG.Q. IQ..NVD.VN QIV....RLK QQ.V..N.K. NPDD.G.VKK   77
Drosophila      1 SED.E..VRD .FRG   .NKL I...QNMTQK . G.RFG.AF VQ..NVN.KN QVMKS..WLR LV.Y..Q.Q. DEAD.G.ILV   77
β Torpedo       1 SVM.DT.LSV .FE  T.NPK V..  .QTVGD K.T.RVG... TN.LI.N.KI .EM....FLN LA.T....Q. DPAA.E.IKD   77
Calf            1 SEA.G..REK .FS  G.DST V..  .REVGD R.W.SIG... AQ....N.KD .EMS.K.YLD LE.T....S. DPEEHE.IDS   77
Mouse           1 SEA.GQ.IKK .FS  N.DSS V..  .REVGD R.G.SIG... AQ....N.KD .EMS.K.YLD LE.T....S. DPAEHD.IDS   77
γ Torpedo       1 ENE.G..IEK .LG  D.DKR II.  .KTLDH IID.TLK... TN....N.KE .AL....WIE IQ.N....S. NTSE.E.IDL   77
Calf            1 RNQ.E..LGD .MQ  G.NPH L..  .EHDSD ..N.SLK... TN....N.RE .AL....WIE MQ.C....R. DPRD.G.LWV   77
Chicken*        1 RNQ.EK.LQD .MT  N.NRH L..  .LRGDQ .ID.TLK... TN....N.RE .TL....WIE MQ.S....R. DPDK.DDIQQ   77
Human*          1 RNQ.E..LAD .MQ  N.DPN L..  .ERDSD ..N.SLK... TN....N.RE .AL....WIE MQ.C....R. DPRD.E.LWV   77
δ Torpedo       1 VNE.E..IND .LIVNK.NKH V...KHNNEV . NIALS... SN....K.TD .TL.S..WMD HA.Y.H..T. NASE.SDISI   79
Calf            1 LNE.E..IRH .FEEKA.NKE L..A.HKE S . EISLA... SN....K.VE .TL....WIE QG.T.S..Q. DAEDFGNISV   78
Mouse           1 LNE.Q..IQH .FNEKG.DKD L....RKEDK . D.ALS... SN....K.VE .TL....WID HA.V.S..Q. DANDFGNITV   79
Chicken*        1 VNQ.E..IHH .FEERG.NKE V....SADEV . D.YLA... SN....K.VD .TL....WVE QS.T....Q. NTSEFG.VDV   79

                81 LRnP???VW? PDnVL?NN?D G?F?n????N VLn???G?n? WnPPAI?RSp C?I?VTYFPF DWQNCpn?F? SnpY???En?  160
Consensus       78 I.L.SDD..L ..L..Y..A. .D.AIVHMTK L.LDYT.KIM .T....FK.Y .E.I..H... .Q...TMKLG IWT.DGTKVS  157
α Torpedo       78 IHI.SEKI.R ..L..Y..A. .D.AIVKFTK ..LDYT.HIT .T....FK.Y .E.I..H... .E...SMKLG TWT.DGSVVV  157
Calf            78 IHI.SEKI.R ..L..Y..A. .D.QVDDKTK A.LKYT.EVT .I....FK.S .K.D..... .Y...TMK.G .WS.DKAKID  157
Rat             78 M.V.AEKI.K ..I..Y..A. .D.AIVKFTK ..LQYT.HIT .T....FK.Y .E.I..H... .E...SMKLG TWT.DGSVVA  157
Human*          78 IHI.SEKI.R ..L..Y..A. .NYEVRYKS. ..IYPT.EVL .V....YQ.S .T.D..... .Q.T.IMK.G .WTFNGDQVS  157
Drosophila      78 ..L.PDK..K ..I..F..A. .S.EITLHV. ..VQHT.AVS .Q.S..Y.TS .T.K.M..... .....TMV.K .YT.DTS.VT  157
β Torpedo       78 ..I.SSD..Q ..I..M..N. .N.DVALDI. .VVSSD.SMR .Q..G.Y..S .S.Q..... .....TMV.S .YS.DSS.VS  157
Calf            78 ..ISAES..L ..V..L..N. .N.DVALDI. .VVSFE.SVR .Q..GLY..S .S.Q..... .....TMV.S .YS.DSS.VS  157
Mouse           78 ..ITAES..L ..V..L..N. .Q.EVAYYA. ..VYND.SMY .L....Y..T .P.A..... .....SLV.R .QT.NAH.VN  157
γ Torpedo       78 V.I.SELL.L ..V..E..V. .V.EVALYC. ..VSPD.CVY .L....F..S .PVS..F... .....SLI.Q .QT.STN.IN  157
Calf            78 ..V.STM..R ..I..E..V. .T.EITLYT. ..VYPD.SIY .L....Y..S .S.H..... .....TMV.Q .QT.SAN.IN  157
Chicken*        78 ..V.SAM..L ..I..E..I. .V.EVALYC. ..VSPD.CIY .L....F..A .S.S..... .....SLI.Q .QT.STN.ID  157
Human*          78 ..V.STM..R ..I..E..V. .QYHVAYFC. ..VRPN.YVT .L....F..S .P.N.L.... .....SLK.T .LN.DAN.IT  159
δ Torpedo       80 ..L.PEL..I ..I..Q..N. .S.QISYSC. ..IYPS.SVY .L....F..S RP.S..... .....SLK.S .LK.TTK.IT  158
Calf            79 ..L.ADM..L .EI..E..N. .S.QISYAC. ..VYDS.YVT .L....F..S .P.S..... .....SLK.S .LK.TAK.IT  159
Mouse           80 ..L.PDM..L .EI..E..N. .L.EVAYYC. ..VYNT.YVY .L....F..A .P.N.NF... .....TLK.S .LA.NAQ.IN  159
Chicken*        80 ..LLPEML.L .EI..E..N. .L.EVAYYC. ..VYNT.YVY .L....F..A .P.N.NF... .....TLK.S .LA.NAQ.IN  159
```


Consensus 161 L??????E?? ? ??? ?IDn??F?EN GEW?I?H?P? ???n??p? ????????an TFYLII?AKP LFYnnNnInP 240
α Torpedo 158 ISPES DAP .LST.M.S ...VMKDYAG WKHWVYYT CCPDTPYLDI .YHF.MQ.I. .YFVV.V.I. 221
Calf 158 INPES DQP .LSN.M.S ...V.KESAG WKHWVFYA CCPSTPYLDI .YHFVMQ.L. .YFIV.V.I. 221
Rat 158 .VLIG SSM NLKDYW.S ...A.IKA.G YKHEIKYN CCEE IYQDI .YS.Y.R.L. ...TI.L.I. 220
Human* 158 INPES DQP .LSN.M.S ...V.KESAG WKHSVTYS CCPDTPYLDI .YHFVMQ.L. .YFIV.V.I. 221
Drosophila 158 .ALYN NKN FV.LSDYWKS .T.D.IEV.A YLNVYEGD SNHP TETDI ...I..R..T ...TV.L.L. 222
β Torpedo 158 .QHALDAKGE AEVKEIVINK DAFT .. .Q.S.E.K.S R KNW RSDDPSYEDV ......Q... ...IVYT.I. 227
Calf 158 .QTGLSP.GQ ERQ EVYIHE GTFI .. .Q.E.I.K.S RLIQPSVDPR GGGEGAREEV ......R... ...LV.V.A. 232
Mouse 158 .KTGLDP.GE ERQ EVYIHE GTFI .. .Q.E.I.K.S RLIQLPGDQR GGKEGHHEEV I.....R... ...LV.V.A. 232
γ Torpedo 158 .Q LSA.EG EAV EWI H..PED.T.. ...T.R.R.A KKN YN WQL TKDDTDFQEI I.F...Q... ...II.I.A. 229
Calf 158 .Q LSQ.DG QTI EWI F..PEA.T.. ...A.R.R.A KML LD EAA PAEEAGHQKV V...L.Q... ...VI.I.A. 229
Chicken* 158 .L LTV.EG QTI EWI F..PEA.T.. ...A.K.R.A RKI INSGRF TPDDIQYQQV I.....Q... ...II.I.V. 230
Human* 158 .Q LSQ.DG QTI EWI F..PEA.T.. ...A.Q.R.A KML LD PAA PAQEAGHQKV V...L.Q... ...VI.I.A. 229
δ Torpedo 160 MDLMT DTI DGKDYPIEWI I..PEA.T.. ...E.I.K.A KKNIYPDK FPNGTNYQDV ......R... ...VI.F.T. 235
Calf 159 .SLKQAE.DG RSYPVEWI I..PEG.T.. ...E.V.R.A RVNVDPSV PLDSPNRQDV ......R... ...VI.ILV. 234
Mouse 160 .SLKQEE.NN RSYPIEWI I..PEG.T.. ...E.V.RAA KLNVDPSV PMDSTNHQDV ......R... ...II.ILV. 235
Chicken* 160 MHLKEESDPE TEKNYRVEWI I..PEG.T.. ...E.I.R.A RKNIHPSY PTESSEHQDI ......K... ...VI.IVT. 237
Consensus 241 CnLIS?n??L VFYLP??ApG EK?pn?ISVL L?nTVFLLLn ???nP?TS?A nPLI?KYLnF ?MnnV?n?nn ?pVnVLNnH? 320
α Torpedo 222 .L.F.FLTG. .....TD S. ..MTLS.... .SL.....VI VELI.S..S. V...G..ML. T.IF.ISSII IT.V.I.T.H 300
Calf 222 .L.F.FLTG. .....TD S. ..MTLS.... .SL.....VI VELI.S..S. V...G..ML. T.VF.IASII IT.I.I.T.H 300
Rat 221 .L...FLTV. .....SD C. ..VTLC.... .SL.....VI TETI.S..LV I...GE..L. T.IF.TLSIV IT.F...V.Y 299
Human* 222 .L.F.FLTG. .....TD S. ..MTLS.... .SL.....VI VELI.S..S. V...G..ML. T.VF.IASII IT.I.I.T.H 300
Drosophila 223 TV FLCV .....AE. ..VTLG..I. .SLV.....V SKIL.P..LV L...A...L. TFIMNTVSIL VT.III.WNF 301
β Torpedo 228 I...ILAI .....PD. ..MSLS..A. .AV.....L ADKV.E..LS V.I.IR..M. I.IL.AFSVI LS.V...L.H 306
Calf 233 I..TLLAIF .....PD. ..MGLS.FA. .TL.....L ADKV.E..LS V.I.I...M. T.VL.TFSVI LS.V...L.H 311
Mouse 233 I..TLLAIF .....PD. ..MGLS.FA. .TL.....L ADKV.E..L. V.I.I...M. T.VL.TFSVI LS.V...L.H 311
γ Torpedo 230 V...SLVV. .YF..AQ.G. Q.CTLS.... .AQ.I..F.I AQKV.E..LN V...G...I. V.FVSMLIVM NC.I...VSL 309
Calf 230 V...SVAI. IYF..AI.G. Q.CTVA.N.. .AQ.W..F.V AKKV.E..Q. N...S...T. LLVVTILIVV NA.V...VSL 309
Chicken* 231 V...SMAV. .YF..AK|.G. Q.CTVS.N.. .AQ....F.I AQKV.E..Q. V...G.|..T. L.VVTVVIVV NA.I...VSL 310
Human* 230 V...SVAI. IHF..AK|.G. Q.CTVA.N.. .AQ....F.V AKKV.E..Q. V...S.|..T. LLVVTILIVV NA.V...VSL 309
δ Torpedo 236 V...FLAS. A....AE S. ..MSTA.... .AQA.....T SQRL.E.AL. V...G...M. I.SL.TGVIV NCGI...F.F 314
Calf 235 V...FMIN. .....AD C. ..TSMA.... .AQS.....I SKRL.A..M. I...G.F.L. G.VL.TMVVV IC.I...I.F 313
Mouse 236 V...FMIN. .....GD C. ..TSVA.... .AQS.....I SKRL.A..M. I..VG.F.L. G.VL.TMVVV IC.I...I.F 314
Chicken* 238 .V..AFMAI. .....AD S. ..MTLV.... .AQS.....V SQRL.A..H. I...G.|..L. I.LL.TAVVV IC.V...F.F 316
M1
M2
M3
Consensus 321 RpP?TH?M?? ?VR?nFn??L P??L?M??? KTR E L S ??? ????????? ?S????????? 400
α Torpedo 301 .S.S..T.PQ W..KI.IDTI .NVMF FSTMKR A. KEKQEN 339
Calf 301 .S.S..V.PE W..KV.IDTI .NIMF FSTMKR P. REKQDK 339
Rat 300 .T.T..T.PT W.KAV.LNL. .RVMF.TRPT SGEGDTP... TFYGA.LSN. NCF.RCRLQK LQGRLPLPR 368
Human* 301 .S.S..V.PN W..KV.IDTI .NIMF FSTMKR P. REKQDK 339
Drosophila 302 .G.R..R.PM YI.SI.LHY. .AF.F.KRP R... LRWMM.MPGM SMPAHPHPSY GSPAE LPKH I.AIGGKQSK 373
β Torpedo 307 .S.N..T.PN WI.QI.IET. .PF.WIQRP VTTPSP D.KPTIISRA 351
Calf 312 .S.H..Q.PL W..QI.IHK. .LY.GLKRP KPERDQ MQEPPSIAPR D.PASGWGRG 366
Mouse 312 .S.H..Q.PF W..QI.IHK. .PY.GLKRP KPERDQ LPEP HHS L.PRSGWGRG 363
γ Torpedo 310 .T.N..SLSE KIKHL.LGF. .KY.G.QLE PSEETP EKPQPRRR .SFGIMIKA 361
Calf 310 .S.H..S.AR G..KV.LRL. .QL.R.HVR PLAPVA VQDAHPRLQN G.SSGWPITA 364
Chicken* 311 .T.N..S.SQ R..QVWLHL. .RY.G.H M PEEAPG PPQATRRR .SLGLMVKA 361
Human* 310 .S.H..S.AR G..KV.LRL. .QL.R.HVR PLAPAA VQDTQSRLQN G.S GWSITT 363
δ Torpedo 315 .T.S..VLST R.KQI.LEK. .RI.H. SRA DESEQPDW QNDLK. RRS.SVGYIS KAQEYFNIKS R. 379
Calf 314 .T.S..VLSE P.KKL.LET. .EI.H.SRP AED GPSPGTLIR R.SSLGYISK 364
Mouse 315 .T.S..VLSE G.KKF.LET. .KL.H.SRP AEE DPGPRALIR R.SSLGYICK 365
Chicken* 317 .T.S..V.SD W..GV.LEI. .RL.H.SHP AES PAGAPCIR RCSSAGYIAK 366

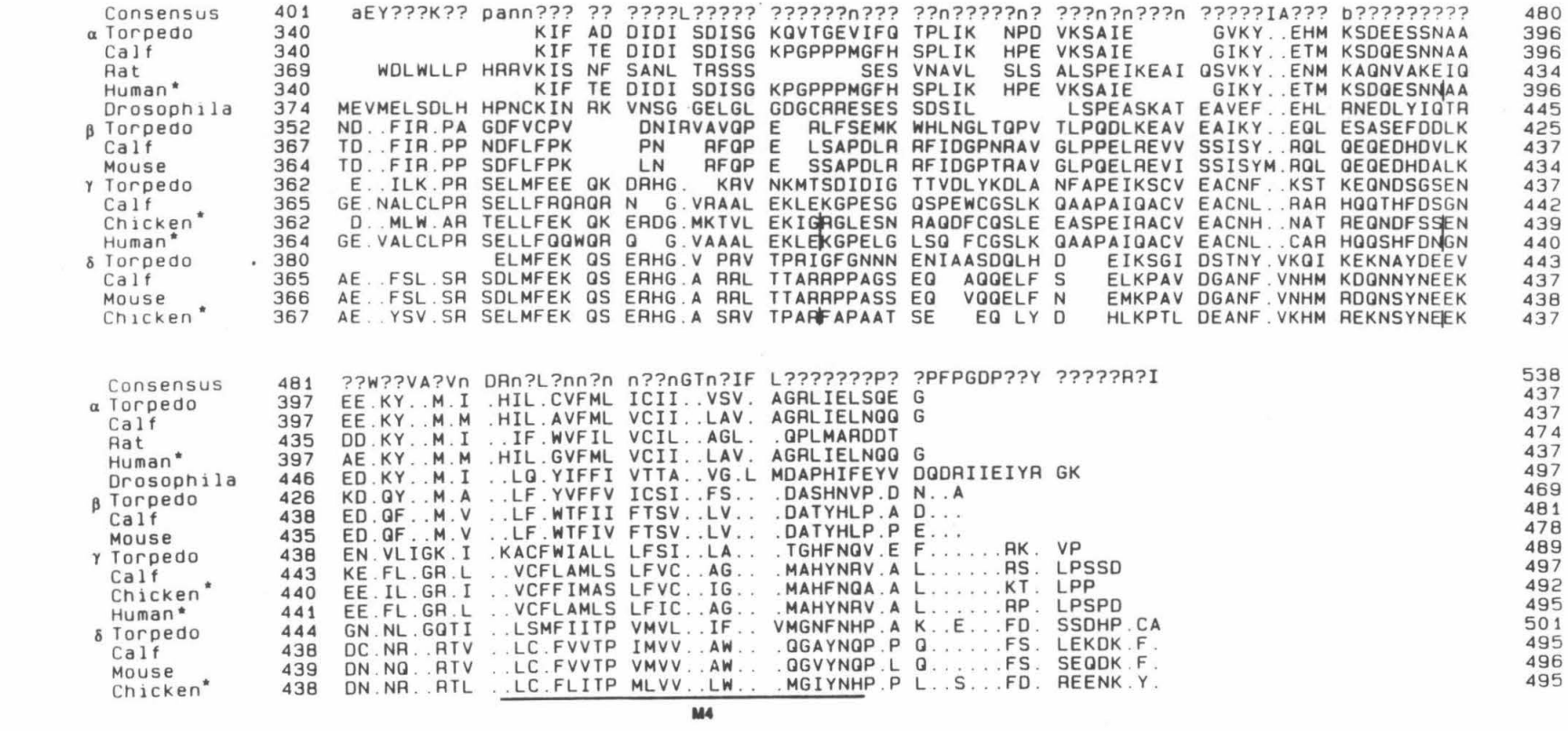

Consensus 401 aEY???K?? pann??? ?? ????L????? ??????n??? ??n?????n? ???n?n???n ?????IA??? b?????????? 480
α Torpedo 340 KIF AD DIDI SDISG KQVTGEVIFQ TPLIK NPD VKSAIE GVKY..EHM KSDEESSNAA 396
Calf 340 KIF TE DIDI SDISG KPGPPPMGFH SPLIK HPE VKSAIE GIKY..ETM KSDQESNNAA 396
Rat 369 WDLWLLP HRRVKIS NF SANL TRSSS SES VNAVL SLS ALSPEIKEAI QSVKY..ENM KAQNVAKEIQ 434
Human* 340 KIF TE DIDI SDISG KPGPPPMGFH SPLIK HPE VKSAIE GIKY..ETM KSDQESNNAA 396
Drosophila 374 MEVMELSDLH HPNCKIN RK VNSG GELGL GDGCRRESES SDSIL LSPEASKAT EAVEF..EHL RNEDLYIQTR 445
β Torpedo 352 ND..FIR.PA GDFVCPV DNIRVAVQP E RLFSEMK WHLNGLTQPV TLPQDLKEAV EAIKY..EQL ESASEFDDLK 425
Calf 367 TD..FIR.PP NDFLFPK PN RFQP E LSAPDLR RFIDGPNRAV GLPPELREVV SSISY..RQL QEQEDHDVLK 437
Mouse 364 TD..FIR.PP SDFLFPK LN RFQP E SSAPDLR RFIDGPTRAV GLPQELREVI SSISYM.RQL QEQEDHDALK 434
γ Torpedo 362 E..ILK.PR SELMFEE QK DRHG. KRV NKMTSDIDIG TTVDLYKDLA NFAPEIKSCV EACNF..KST KEQNDSGSEN 437
Calf 365 GE.NALCLPR SELLFRQRQR N G.VRAAL EKLEKGPESG QSPEWCGSLK QAAPAIQACV EACNL..RAR HQQTHFDSGN 442
Chicken* 362 D..MLW.AR TELLFEK QK ERDG.MKTVL EKIGRGLESN RAQDFCQSLE EASPEIRACV EACNH..NAT REQNDFSSEN 439
Human* 364 GE.VALCLPR SELLFQQWQR Q G.VAAAL EKLEKGPELG LSQ FCGSLK QAAPAIQACV EACNL..CAR HQQSHFDNGN 440
δ Torpedo 380 ELMFEK QS ERHG.V PRV TPRIGFGNNN ENIAASDQLH D EIKSGI DSTNY.VKQI KEKNAYDEEV 443
Calf 365 AE..FSL.SR SDLMFEK QS ERHG.A RRL TTARRPPAGS EQ AQQELF S ELKPAV DGANF.VNHM KDQNNYNEEK 437
Mouse 366 AE..FSL.SR SDLMFEK QS ERHG.A RRL TTARRPPASS EQ VQQELF N EMKPAV DGANF.VNHM RDQNSYNEEK 438
Chicken* 367 AE..YSV.SR SELMFEK QS ERHG.A SRV TPARFAPAAT SE EQ LY D HLKPTL DEANF.VKHM REKNSYNEEK 437
Consensus 481 ??W??VA?Vn DRn?L?nn?n n??nGTn?IF L????????P? ?PFPGDP??Y ?????R?I 538
α Torpedo 397 EE.KY..M.I .HIL.CVFML ICII..VSV. AGRLIELSQE G 437
Calf 397 EE.KY..M.M .HIL.AVFML VCII..LAV. AGRLIELNQQ G 437
Rat 435 DD.KY..M.I ..IF.WVFIL VCIL..AGL. .QPLMARDDT 474
Human* 397 AE.KY..M.M .HIL.GVFML VCII..LAV. AGRLIELNQQ G 437
Drosophila 446 ED.KY..M.I ..LQ.YIFFI VTTA..VG.L MDAPHIFEYV DQDRIIEIYR GK 497
β Torpedo 426 KD.QY..M.A ..LF.YVFFV ICSI..FS.. .DASHNVP.D N..A 469
Calf 438 ED.QF..M.V ..LF.WTFII FTSV..LV.. .DATYHLP.A D... 481
Mouse 435 ED.QF..M.V ..LF.WTFIV FTSV..LV.. .DATYHLP.P E... 478
γ Torpedo 438 EN.VLIGK.I .KACFWIALL LFSI..LA.. .TGHFNQV.E F.....RK. VP 489
Calf 443 KE.FL.GR.L ..VCFLAMLS LFVC..AG.. .MAHYNRV.A L.....RS. LPSSD 497
Chicken* 440 EE.IL.GR.I ..VCFFIMAS LFVC..IG.. .MAHFNQA.A L.....KT. LPP 492
Human* 441 EE.FL.GR.L ..VCFLAMLS LFIC..AG.. .MAHYNRV.A L.....RP. LPSPD 495
δ Torpedo 444 GN.NL.GQTI ..LSMFIITP VMVL..IF.. VMGNFNHP.A K..E...FD. SSDHP.CA 501
Calf 438 DC.NR..RTV ..LC.FVVTP IMVV..AW.. .QGAYNQP.P Q.....FS. LEKDK.F. 495
Mouse 439 DN.NQ..RTV ..LC.FVVTP VMVV..AW.. .QGVYNQP.L Q.....FS. SEQDK.F. 496
Chicken* 438 DN.NR..RTL ..LC.FLITP MLVV..LW.. .MGIYNHP.P L..S...FD. REENK.Y. 495
M4

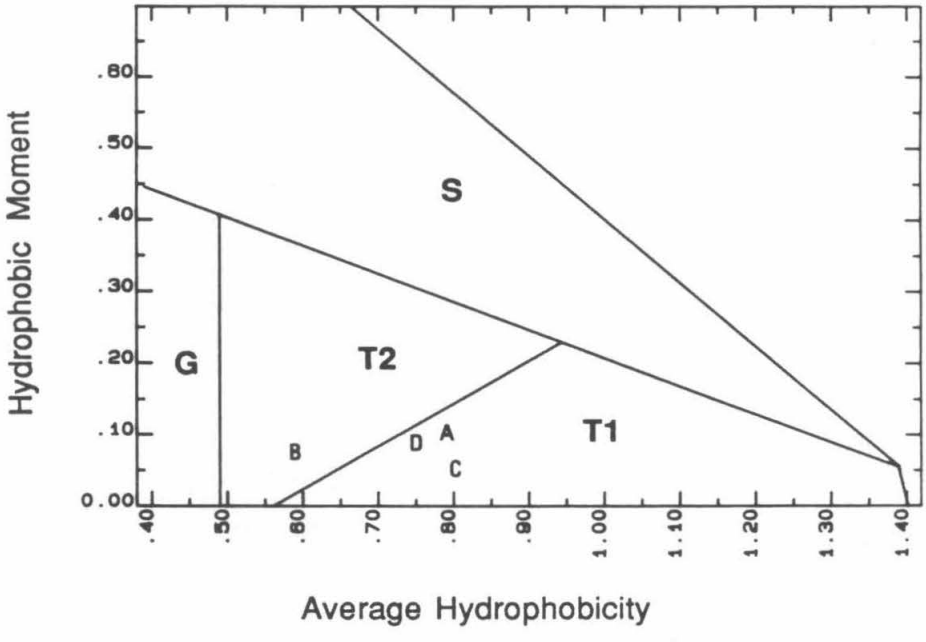
Figure 2.7

Hydrophobic Moment Plot of the Acetylcholine Receptor

The four helices selected by the Eisenberg algorithm have been plotted according to their average hydrophobicity and average hydrophobic moment. All four regions are believed to span the membrane.

Hydrophobic Moment Plot

Acetylcholine Receptor

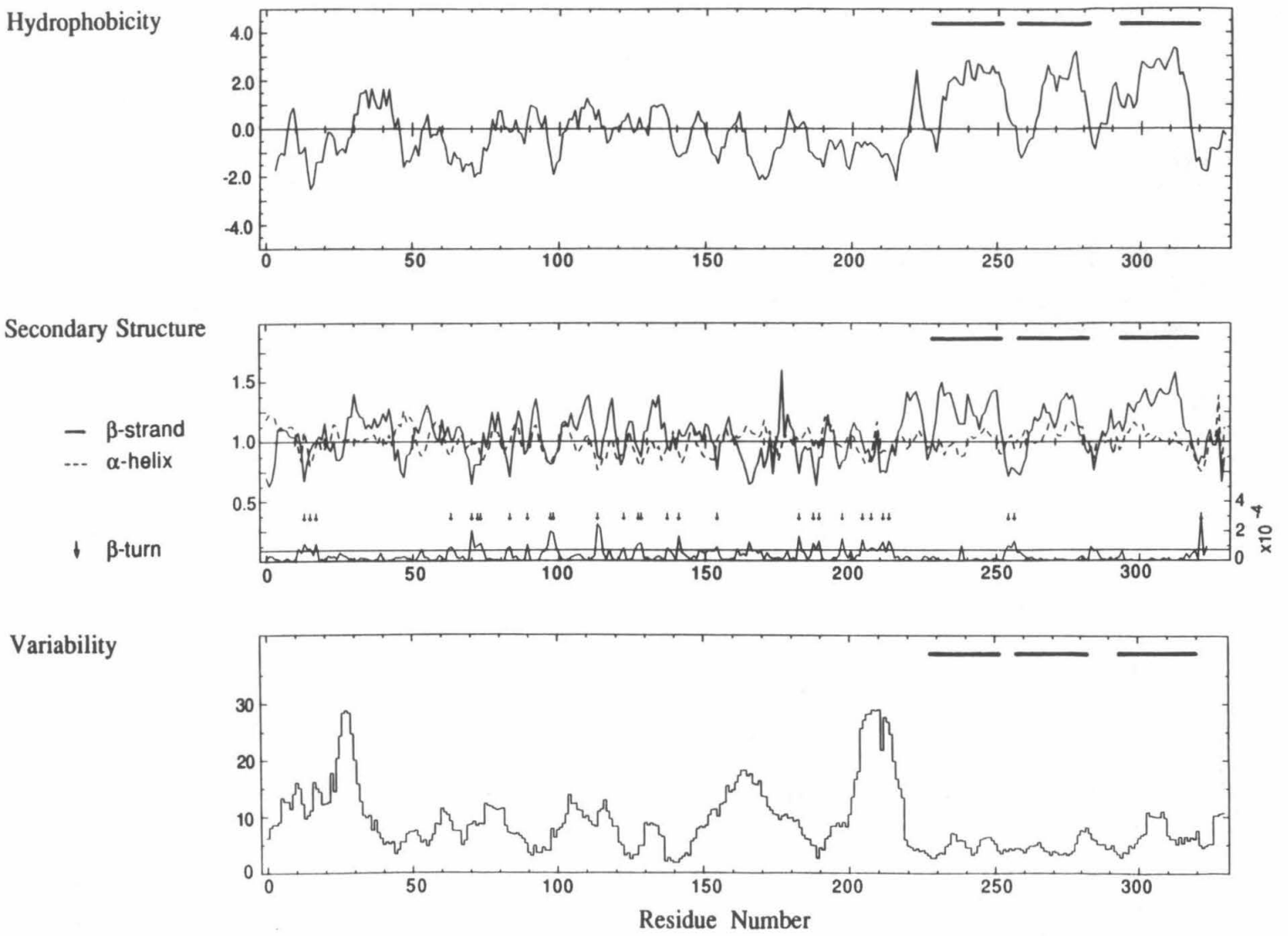


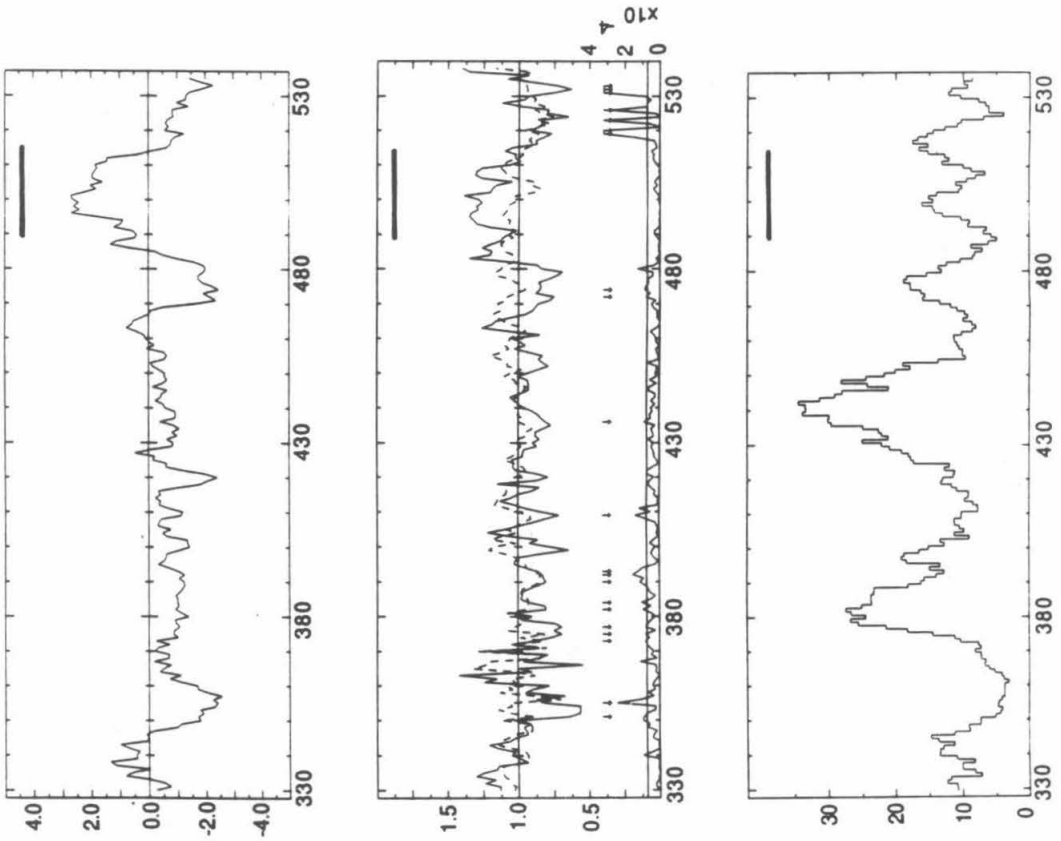
A = 231
B = 265
C = 298
D = 493

Figure 2.8

Hydropathy, Secondary Structure Potential, and Variability Plots for the Acetylcholine Receptor

The aligned sequences of the acetylcholine receptor were analyzed with regard to their hydrophobicity, secondary structure potential, and variability. The heavy bars indicate the position of putative transmembrane regions.


Hydrophobicity
4.0
2.0
0.0
-2.0
-4.0
0
50
100
150
200
250
300
Secondary Structure
β-strand
α-helix
β-turn
1.5
1.0
0.5
4
2
0
x10 -4
Variability
30
20
10
0
Residue Number


4.0
2.0
0.0
-2.0
-4.0
330
380
430
480
530
1.5
1.0
0.5
x10 -4
4
2
0
330
380
430
480
530
30
20
10
0
330
380
430
480
530

charges at five of the eight helical ends. Although it may be of no significance, four of these regions were at the N-terminus of each helix.

The regions finally selected as transmembranous correspond to residues 230-254, 260-284, 297-321, and 492-517 of the aligned sequences.

The variability plot for this group of sequences indicates that the transmembrane regions are generally more conserved than the protein as a whole.

Discussion

The use of positive charges to indicate the termini of transmembrane helices may not be wholly justified on the basis of the sequences analyzed above, but previous studies using other proteins have also indicated that basic residues often occur at the membranous interface (2,3). For this reason their presence has been included as a means of identifying the helix ends, even though the reaction center and photosystem II proteins do not show a tendency to have positive charges located at the termini of their transmembrane helices. Perhaps the presence of positive charge at a helix terminus relates to the membrane composition or the method of membrane insertion.

The helices defined for the acetylcholine receptor using this method of assignment generally agree with those chosen by other authors on the basis of hydrophobicity alone (37). The presence of the amphipathic helix proposed by some to span the bilayer and to line the cation channel (48-51) was not predicted as a transmembrane

region. The uncertainty in the number of transmembrane regions persists in the literature for several reasons. In their review of the AChR, Popot and Changeux (52) point out that both the four- and five-helix models place the N-terminus and potential glycosylation sites on the external side of the membrane. In addition, the known phosphorylation sites of *Torpedo californica*, γSer354 and δSer361, are on the cytoplasmic side. Knowledge of the location of the C-terminus would, however, distinguish between these two models. Biochemical experiments have recently localized the C-terminus of the δ subunit to the synaptic side of the membrane (53), but immunolocalization experiments indicate a cytoplasmic location for it (54). The use of immunolocalization in the determination of transmembrane topology has recently been questioned, especially in light of conficting information about the receptor which resulted from similar immunolocalization experiments (55). The biochemical determination of the location of the C-terminus may be more reliable and, therefore, the four helices selected in this work are assumed to represent the transmembrane regions of the acetylcholine receptor.

2D. Rhodopsin

Rhodopsin is homologous to both the β-adrenergic and muscarinic acetylcholine receptors. Within the rhodopsin family are both types of photoreceptors, those found in the membranes of cone cells, sometimes referred to as the iodopsins, and those found in the photoreceptor disks of rod cells, the rhodopsins. Because rhodopsin is located in the membrane of the photoreceptor disc, its intradiscal loops correspond to the extracellular portions of the iodopsin, the β-adrenergic receptor, and the muscarinic acetylcholine receptor.

All proteins in the rhodopsin superfamily are believed to contain seven transmembrane helices from hydrophobic analyses. X-ray diffraction studies of frog disk rhodopsin have shown that this protein does indeed have seven membrane-spanning helices and that the three-dimensional shape of this rhodopsin is virtually identical to that of bacteriorhodopsin (57). The regions in the primary sequence which correspond to these helices have not been explicitly identified, although proteolysis (reviewed in 58) and photolabelling (59,60) experiments have helped to identify the locations of certain residues.

Methods

The methods used for assigning the transmembrane helices of rhodopsin were as described in the previous sections of this chapter.

Results

The alignment used for these analyses is shown in figure 2.9. The

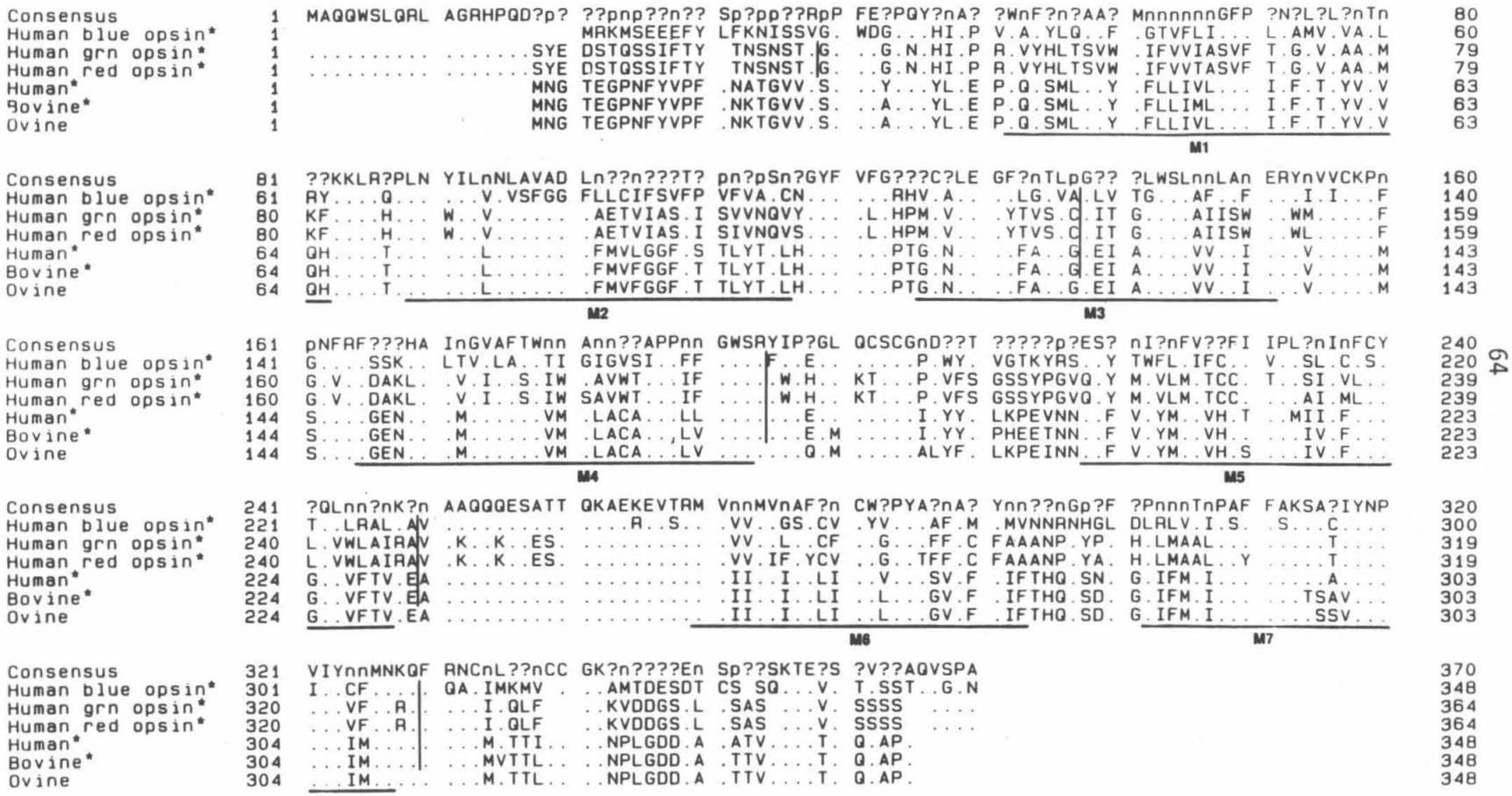
Figure 2.9

Sequence Alignment of Rhodopsin

Proteins in this alignment for which the gene structure is known have been indicated with asterisks. Within those sequences, the intron locations have been denoted by vertical lines between residues. The heavy bars underlining portions of the alignment indicate putative transmembrane regions.

Sequences included are human blue, green, and red opsin (61), human rhodopsin (62), bovine rhodopsin (63), and ovine rhodopsin (64,65).

Rhodopsin


Consensus 1 MAQQWSLQRL AGRHPQD?p? ??pnp??n?? Sp?pp??RpP FE?PQY?nA? ?WnF?n?AA? MnnnnnnGFP ?N?L?L?nTn 80
Human blue opsin* 1 MRKMSEEEFY LFKNISSVG. WDG...HI.P V.A.YLQ..F .GTVFLI... L.AMV.VA.L 60
Human grn opsin* 1 .....SYE DSTQSSIFTY TNSNST.G. ..G.N.HI.P R.VYHLTSVW .IFVVIASVF T.G.V.AA.M 79
Human red opsin* 1 .....SYE DSTQSSIFTY TNSNST.G. ..G.N.HI.P R.VYHLTSVW .IFVVTASVF T.G.V.AA.M 79
Human* 1 MNG TEGPNFYVPF .NATGVV.S. ..Y...YL.E P.Q.SML..Y .FLLIVL... I.F.T.YV.V 63
Bovine* 1 MNG TEGPNFYVPF .NKTGVV.S. ..A...YL.E P.Q.SML..Y .FLLIML... I.F.T.YV.V 63
Ovine 1 MNG TEGPNFYVPF .NKTGVV.S. ..A...YL.E P.Q.SML..Y .FLLIVL... I.F.T.YV.V 63
M1
Consensus 81 ??KKLR?PLN YILnNLAVAD Ln??n???T? pn?pSn?GYF VFG???C?LE GF?nTLpG?? ?LWSLnnLAn ERYnVVCKPn 160
Human blue opsin* 61 RY....Q... ...V.VSFGG FLLCIFSVFP VFVA.CN... ...RHV.A.. ..LG.VA.LV TG...AF..F ...I.I...F 140
Human grn opsin* 80 KF....H... W..V..... .AETVIAS.I SVVNQVY... .L.HPM.V.. .YTVS.C.IT G....AIISW ..WM.....F 159
Human red opsin* 80 KF....H... W..V..... .AETVIAS.I SIVNQVS... .L.HPM.V.. .YTVS.C.IT G....AIISW ..WL.....F 159
Human* 64 QH....T... ...L..... .FMVLGGF.S TLYT.LH... ...PTG.N.. ..FA..G.EI A....VV..I ...V.....M 143
Bovine* 64 QH....T... ...L..... .FMVFGGF.T TLYT.LH... ...PTG.N.. ..FA..G.EI A....VV..I ...V.....M 143
Ovine 64 QH....T... ...L..... .FMVFGGF.T TLYT.LH... ...PTG.N.. ..FA..G.EI A....VV..I ...V.....M 143
M2
M3
Consensus 161 pNFAF???HA InGVAFTWnn Ann??APPnn GWSAYIP?GL QCSCGnD??T ?????p?ES? nI?nFV??FI IPL?nInFCY 240
Human blue opsin* 141 G....SSK.. LTV.LA..TI GIGVSI..FF ....F..E.. .....P.WY. VGTKYAS..Y TWFL.IFC.. V..SL.C.S. 220
Human grn opsin* 160 G.V..DAKL. .V.I..S.IW .AVWT...IF ....W.H.. KT...P.VFS GSSYPGVQ.Y M.VLM.TCC. T..SI.VL.. 239
Human red opsin* 160 G.V..DAKL. .V.I..S.IW SAVWT...IF ....W.H.. KT...P.VFS GSSYPGVQ.Y M.VLM.TCC. ...AI.ML.. 239
Human* 144 S....GEN.. .M.....VM .LACA...LL ....E.. .....I.YY. LKPEVNN..F V.YM..VH.T ..MII.F... 223
Bovine* 144 S....GEN.. .M.....VM .LACA..,LV ....E.M .....I.YY. PHEETNN..F V.YM..VH.. ...IV.F... 223
Ovine 144 S....GEN.. .M.....VM .LACA...LV .....Q.M .....ALYF. LKPEINN..F V.YM..VH.S ...IV.F... 223
M4
M5
Consensus 241 ?QLnn?nK?n AAQQQESATT QKAEKEVTRM VnnMVnAF?n CW?PYA?nA? Ynn??nGp?F ?PnnnTnPAF FAKSA?IYNP 320
Human blue opsin* 221 T..LRAL.AV .....R..S.. .VV..GS.CV .YV...AF.M .MVNNRNHGL DLRLV.I.S. .S...C.... 300
Human grn opsin* 240 L.VWLAIRAV .K..K..ES. .....VV..L..CF ..G...FF.C FAAANP.YP. H.LMAAL.....T.... 319
Human red opsin* 240 L.VWLAIRAV .K..K..ES. .....VV.IF.YCV ..G...TFF.C FAAANP.YA. H.LMAAL..Y .....T.... 319
Human* 224 G..VFTV.EA ....II..I..LI ..V...SV.F .IFTHQ.SN. G.IFM.I..... .....A.... 303
Bovine* 224 G..VFTV.EA ....II..I..LI ..L...GV.F .IFTHQ.SD. G.IFM.I..... ...TSAV... 303
Ovine 224 G..VFTV.EA ....II..I..LI ..L...GV.F .IFTHQ.SD. G.IFM.I..... .....SSV... 303
M6
M7
Consensus 321 VIYnnMNKQF RNCnL??nCC GK?n????En Sp??SKTE?S ?V??AQVSPA 370
Human blue opsin* 301 I..CF....| QA.IMKMV . ..AMTDESDT CS SQ...V. T.SST..G.N 348
Human grn opsin* 320 ...VF..R.| ...I.QLF ..KVDDGS.L .SAS ...V. SSSS 364
Human red opsin* 320 ...VF..R.| ...I.QLF ..KVDDGS.L .SAS ...V. SSSS 364
Human* 304 ...IM....| ...M.TTI.. ..NPLGDD.A .ATV....T. Q.AP. 348
Bovine* 304 ...IM....| ...MVTTL.. ..NPLGDD.A .TTV....T. Q.AP. 348
Ovine 304 ...IM..... ...M.TTL.. ..NPLGDD.A .TTV....T. Q.AP. 348

positions of known introns in the protein-encoding region of the DNA have been indicated in this figure. Residue numbers used in the following discussion will refer to the numbering of the consensus sequence.

The Eisenberg analysis identified seven possible transmembrane regions and the hydrophobic moment plot for these regions is shown in figure 2.10. Of these regions, six could be defined on one end by a β-turn as shown by the secondary structure prediction in figure 2.11. The other ends could all be positioned such that a proline or basic residue bordered the helix. The regions predicted to span the membrane, M1 to M7, correspond to residues 52-82, 89-116, 126-151, 165-193, 218-247, 270-293, and 302-327. The β-turn analysis predicted turns in the centers of helices 1 and 6 as so defined. The current assignments were selected because alternate assignments would have incorporated many more charged residues into the membrane. These predicted β-turns may be artifacts, the result of a limited data set.

Discussion

The predicted transmembrane helices of rhodopsin are generally longer than those predicted for the nicotinic acetylcholine receptor or those determined for the reaction center. The photoreceptor disc membrane is believed to be thicker than other membranes due to the presence of long-chain lipids and fatty acids (66), a fact that may explain this difference. Because the β-adrenergic receptor and the muscarinic acetylcholine receptor are located in the cell membrane, which is thinner, it is unlikely that the rhodopsin helix termini will

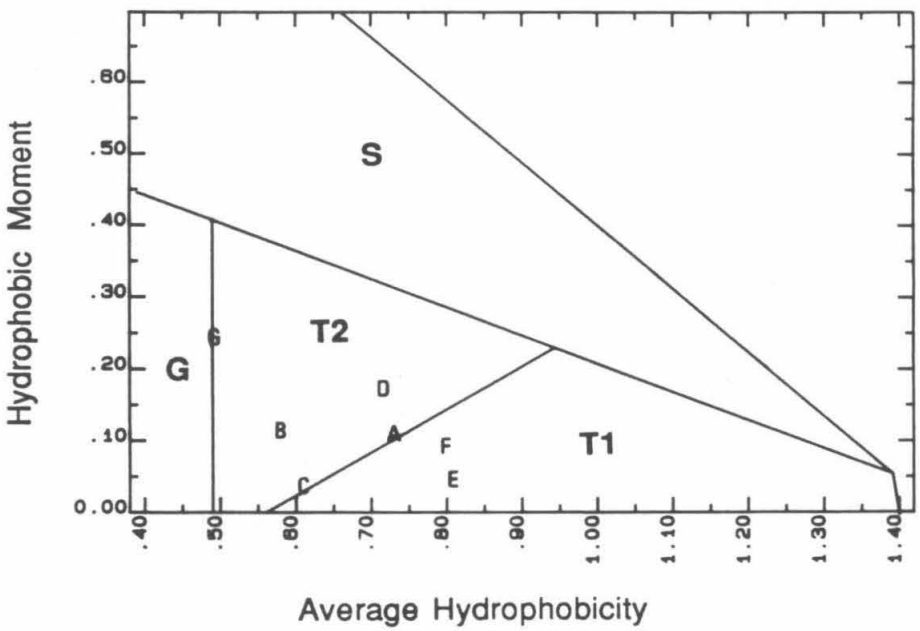
Figure 2.10

Hydrophobic Moment Plot for Rhodopsin

The seven regions selected by the Eisenberg criteria have been plotted according to their average hydrophobicity and average hydrophobic moment. Each region is believed to be part of a membrane-spanning region.

Hydrophobic Moment Plot

Rhodopsin

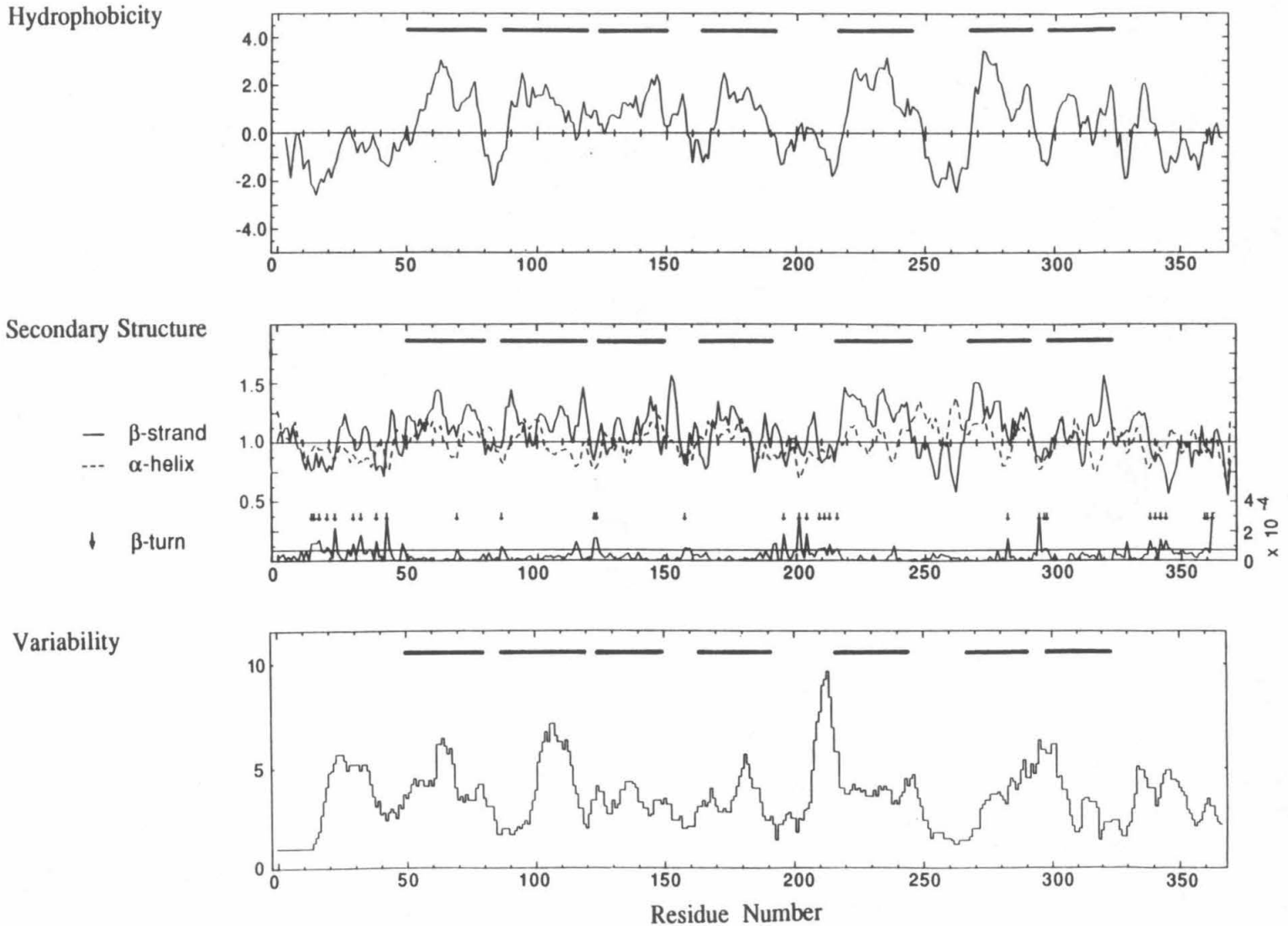


A = 58
B = 92
C = 130
D = 170
E = 220
F = 270
G = 304

Figure 2.11

Hydropathy, Secondary Structure Potential, and Variability Plots for Rhodopsin

The hydropathy, secondary structure potential, and sequence variability has been plotted as a function of sequence position. Heavy bars indicate the positions of putative membrane sequences.


Hydrophobicity
4.0
2.0
0.0
-2.0
-4.0
0
50
100
150
200
250
300
350
Secondary Structure
β-strand
α-helix
β-turn
1.5
1.0
0.5
4
2
0
x 10 -4
Variability
10
5
0
Residue Number

correspond well with those of the other receptors when the sequences are aligned.

The predicted helices do not contain any of the sites that have been shown to be extramembranous. These sites are conveniently summarized by Findlay (58), who modeled ovine rhodopsin. The CNBr cleavage sites included in that summary were not considered in evaluating the current transmembrane topology because cyanogen bromide may be membrane permeant, allowing cleavage to occur within the membrane. In addition, the conditions used for such cleavages are harsh and likely to destroy membrane integrity. All sites shown to be in the hydrophobic domain, as defined by photolabelling with the hydrophobic probe azido-iodobenzene (59,60), map to the transmembrane region with the exceptions of two lysine residues at positions 83 and 248. These lysines are immediately adjacent to assigned transmembrane helices, helices M1 and M6, and there is evidence in both cases that the following amino acid is extramembranous (58). Together, this information helps to confirm the helix assignments for M1 and M6, the two helices which each have a predicted β-turn within them. The cysteine residue at position 333 was also labelled by azido-iodobenzene, but has been shown by other evidence to be hydrophilic and exposed to the cytoplasm (67).

Like the reaction center, the sequences of the predicted helices were as variable as the protein as a whole, although a large fraction of this protein is located in the membrane. Those loops of the protein exposed to the cytoplasm were fairly well conserved, however.

Three of the five intron locations correspond to helix termini.

Of the remaining two, one is located in the N-terminal, extra-membranous portion of the protein, and the other is located in the center of M3. This latter location is contrary to the idea that introns do not occur in the center of the DNA coding for membrane-spanning regions.

2E. Band 3

The anion-exchange protein, band 3, is found in the erythrocyte membrane. It is composed of two sections, a cytoplasmic N-terminal region which is not essential to ion transport (68) and a membrane-associated, C-terminal domain which contains the transport machinery. Both the N- and C-termini of this protein are believed to be located on the cytoplasmic side of the membrane (69), indicating that the protein must span the membrane an even number of times.

The assignment of transmembrane helices for band 3 may be somewhat different than for the other proteins examined thus far. Band 3 is an anion transporter and may, therefore, contain basic residues within the transmembrane region. The use of positively charged residues to indicate helix ends may lead to erroneous assignments when such residues are located near the channel mouth but within the membrane.

Methods

The methods were applied as described earlier. Because only three sequences for this protein are known in their entirety, all currently known fragments of human erythrocyte band 3 were also included in the analyses. The inclusion of partial sequences in these procedures does not introduce artifacts because the quantities examined are averages. The one exception is the variability analysis, for which the partial sequence was not included.

Although the analytical techniques were the same as those described earlier, a modification of the procedure outlined in section

2B was required and will be described below.

Results

The sequences known for band 3 are shown aligned in figure 2.12. The numbering of the consensus sequence will be used to denote residue positions in the following discussion.

The Eisenberg algorithm predicted 12 possible transmembrane regions in band 3. These 21-residue regions begin at residues 330, 505, 548, 592, 622, 694, 730, 785, 832, 887, 909, and 970 of the aligned sequences. In figure 2.13 the hydrophobic moment plot of these regions shows that the putative helix at 330 maps to the globular region of the hydrophobic moment diagram, indicating that this region is probably not membrane-associated. This supposition has been experimentally verified since this region is within the cytoplasmic N-terminal region of the protein (74). The hydrophobic moment analysis, therefore, suggests that band 3 has eleven transmembrane helices. The requirement that there must be an even number of membrane-spanning regions for this protein indicates that either one of these predicted helices does not traverse the membrane, or a region traverses the membrane more than once. An examination of the hydrophobicity plot shown in figure 2.14 shows two hydrophobic domains which could, conceivably, span the membrane more than once. In order to define the transmembrane topology of band 3, it was necessary to incorporate known experimental evidence concerning the transmembrane disposition of the protein after possible transmembrane regions were identified. Reviews of such topological information have

Figure 2.12

Sequence Alignment of Band 3

The sequences of chick erythroid (70), mouse erythroid (71), and human non-erythroid (72) band 3 have been aligned. The human non-erythroid band 3 has been denoted as K562, after the cell line from which it was derived. Partial sequences of human erythroid band 3 (73-79) have also been included. It should be noted that the numbering of the protein fragments of human erythroid band 3 in this figure is meaningless after the first 201 residues.

The locations of introns in the DNA coding for the murine sequence have been determined by Kopito *et al.* (80) and are denoted within the protein sequence by vertical lines between residues. Heavy bars underlining portions of the aligned sequences indicate putative membrane helices.

This alignment is patterned after that of Demuth *et al.* (72), especially for the 500 N-terminal residues.

```
Consensus    1 M?an?D?YEa VLEIPDRD?E E?LENIIGQ? ?Y?Dn?IP?p ?M?aP?A??T EnTATDYVPS ?TST?HPSSG ?HKVYVELQE  80
Human        1 .EELQ.D..D         .M. .N..     .E E.E.PD..ES Q.EE.A.HD. .A.....     H. .S..   T.........  63
Mouse*       1 .GDMR.H .E ........S. .E.......I A.R.LT..VT E.QD.E.LP. .Q........ S...P.....   Q ........  77

Consensus   81 LnMD?EbNQE L?WMEAARWn ?LEENL?E?G AWGRPHLS?L TFWSLLEL?R ?F?KG?VLLD L??TSLAGVA ??LLD?nI?E 160
Human       64 .V..E.K... .R.......V Q.....G.N. ........H. .........R. V.T..T.... .QE..... NQ...RF.F. 143
Chick        1          ...G..L H...SMEPG. ... S..PL. .YH.....H. A.A..V.... VAA....A.. HV...QL.Y. 66
Mouse*      78 .M..Q R... .Q.V...H.I G.....R.D. V.......Y. ........QK V.S..TF..G .AE..... NH...CF.Y. 156
K562         1                                          ..R. TLAH.A.... .DQQT.P... QVVEQMV.S  33

Consensus  161 DQIbPQDR?? nLRALLLKHS Hn?aEKDFSF PRNISA SLG ?CWGITMVRG LAVTPTSPSL SWEVFLATRL EGVKERDV  240
Human      144 ...R....EE L.........  .AGE              .E A                    .  G...          175
Chick       67 G.LK..H.DD V......R.K .PSE           .E.                                  VW  95
Mouse*     157 ...R....EE L.......R. .AED             .. N                    .  ....          188
K562        34 ...KAL..NA V.........  .PSD.....  .....G... S.........  .........  .........  .VER....PP 113

Consensus  241    PPA?LTRS ???????LP? ??SLETQL?       G          LKIQL HQQLPED?EA TLVLVGCA?F LE?P?LAFVR 320
Human      176      ..V....  GDPSQPL..Q HS.....F                                                    201
Chick       98 TL.A.Q.QC. DGEQKDAEQR ALLR.QRAVE MRELH.AGES PSRAQ.GP.. .......T.. .....A..A. ..Q.L..L.. 175
Mouse*     189      ..V....  GGASEPL..H QP......YC GQCEG.SEGP STSGT...        .P.S.T ......R.N. ..K.V.G... 259
K562       114 PA...GI...  KSKHELK.       ..                          ..       ..NA.. NV.....VE. .SR.TM.... 161

Consensus  321 L?EAV L DA VL?VPVPVRF LL?LLGP??P ?nDY???GRA AATLM?DRVF R??AYLA??R EELL??L??F LDCSnVLPP  400
Chick      176 .AGL.  R..  ..A..L.... V.TV...RQ. APQLPRDR.. ...V.A.... .RD...CGG. A...GG.QG. .EA.I....Q 253
Mouse*     260 .K...P.E.L  ..PE..G  .  ..V....EA. HV..TQL... .....TE... .IT.SM.HN. ....RS.ES. ....L....T 337
K562       162 .R...E. ..  ..E.....  .FL....SSA NM..HEI..S IS...S.KQ. HEA....DE. .D..TAINA. ....V....  239

Consensus  401    PSE  L   L PVQ?EELL RRRYHFQRQM LKKREEQGRL LP??A???P? ??????L?G? KG?nGQ?DDP L?RTGRnFGG 480
Human      202                                                    LD.B.G                        207
Chick      254 EV...QH.HA .I.L.R HAV  ....                       QHPDTNRT.G GPTAPKDT.D ..QAP.D... .L..R.P... 316
Mouse*     338 DA...KA.LN .V...K  ...  ....                     ..SP.KPD.N LYNTLD.N.G ..GP.DE... .R....I... 400
K562       240    ..          ..G....  .SVA.....  .........  ..TG.GLE.K SAQDKA.LQM VERQ..LKMI PSAD.AA...  308

Consensus  481 LIRDIRRRYP ?YLSDn?DAL ?PQ?LAAVIF IYFAALSPAI TFGGLLGEKT R?LMGVSELL nSTAVQGnLF ?LLGAQPLLV 560
Human      208                T..F S..V....FY SFI.....  ...F.....       .....  I.....I..  A.........  265
Chick      317 .V.........  K....VR...  N..C.....  .........  ..........  .GM.....  L..S..CL..  S..S.....  396
Mouse*     401 ..........  Y....IT...  S..V.....  .........V  ......|....  .N.....  I.....I..  A.........  480
K562       309 ....V.....  H....FR...  D..C.....  .........  ..........  QD.I.....I  M...L..VV.  C.........  388
                                      M1                                  M2
```

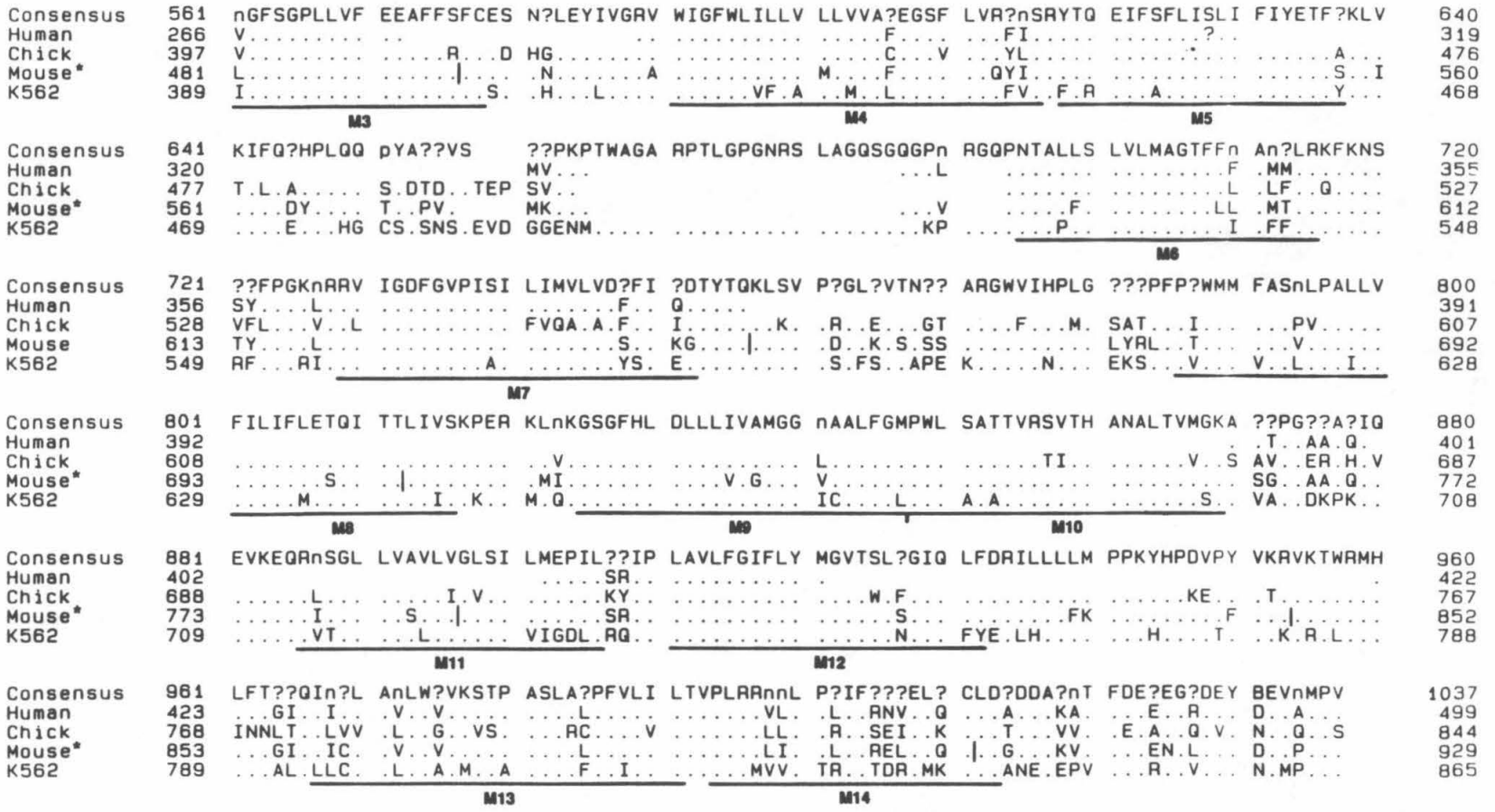

Consensus 561 nGFSGPLLVF EEAFFSFCES N?LEYIVGRV WIGFWLILLV LLVVA?EGSF LVR?nSRYTQ EIFSFLISLI FIYETF?KLV 640
Human 266 V 319
Chick 397 V 476
Mouse* 481 L 560
K562 389 I 468
M3
M4
M5
Consensus 641 KIFQ?HPLQQ pYA??VS ??PKPTWAGA RPTLGPGNRS LAGQSGQGPn RGQPNTALLS LVLMAGTFFn An?LRKFKNS 720
Human 320 MV 355
Chick 477 T.L.A S.DTD..TEP SV 527
Mouse* 561 DY T..PV. MK 612
K562 469 E HG CS.SNS.EVD GGENM KP P I .FF 548
M6
Consensus 721 ??FPGKnRRV IGDFGVPISI LIMVLVD?FI ?DTYTQKLSV P?GL?VTN?? ARGWVIHPLG ???PFP?WMM FASnLPALLV 800
Human 356 SY L Q 391
Chick 528 VFL V L FVQA.A.F I K. .R..E..GT F M. SAT I PV 607
Mouse 613 TY L S.. KG D..K.S.SS LYRL T V 692
K562 549 RF RI A YS. E .S.FS..APE K N EKS V V..L I.. 628
M7
Consensus 801 FILIFLETQI TTLIVSKPER KLnKGSGFHL DLLLIVAMGG nAALFGMPWL SATTVRSVTH ANALTVMGKA ??PG??A?IQ 880
Human 392 .T..AA.Q. 401
Chick 608 V L TI V S AV..ER.H.V 687
Mouse* 693 S MI V.G V SG..AA.Q.. 772
K562 629 M I..K.. M.Q IC L A.A S VA..DKPK.. 708
M8
M9
M10
Consensus 881 EVKEQRnSGL LVAVLVGLSI LMEPIL??IP LAVLFGIFLY MGVTSL?GIQ LFDRILLLLM PPKYHPDVPY VKRVKTWRMH 960
Human 402 SR 422
Chick 688 L I.V KY W.F KE T 767
Mouse* 773 I S SR FK F 852
K562 709 VT L VIGDL.RQ N FYE.LH H T K.R.L 788
M11
M12
Consensus 961 LFT??QIn?L AnLW?VKSTP ASLA?PFVLI LTVPLRRnnL P?IF???EL? CLD?DDA?nT FDE?EG?DEY BEVnMPV 1037
Human 423 GI I V V L VL. .L..RNV..Q A KA E R D..A 499
Chick 768 INNLT LVV L G VS RC V LL. .R..SEI..K T VV. E.A..Q.V. N..Q..S 844
Mouse* 853 GI IC V V L LI. .L..REL..Q G KV EN.L D..P 929
K562 789 AL LLC L A.M A F I MVV. TR..TDR.MK ANE.EPV R V N.MP 865
M13
M14

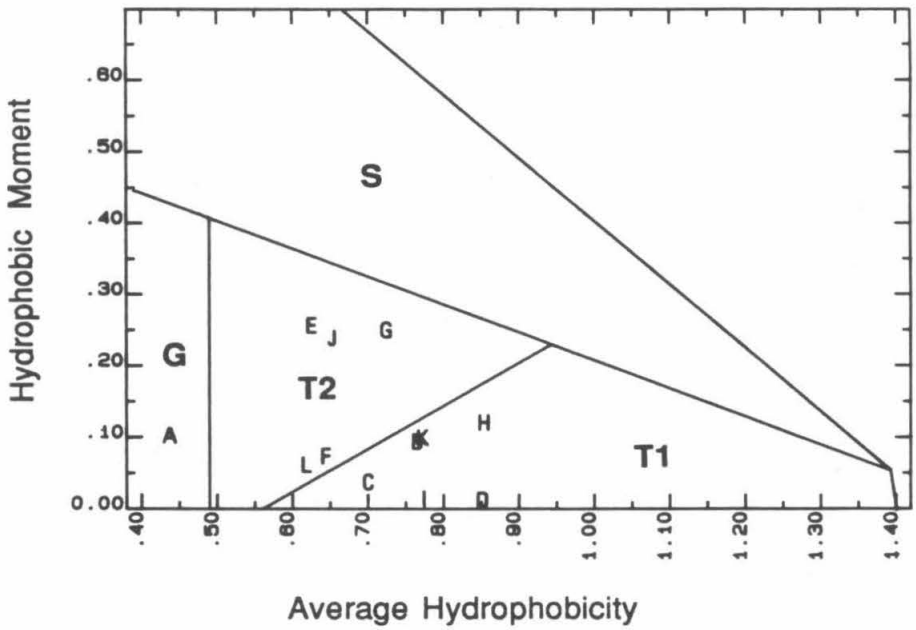
Figure 2.13

Hydrophobic Moment Plot of Band 3

In this diagram, the helices selected by the Eisenberg algorithm have been plotted according to their average hydrophobicity and hydrophobic moment. Region A is not located within the membrane (74), as suggested by this plot.

Hydrophobic Moment Plot

Band 3

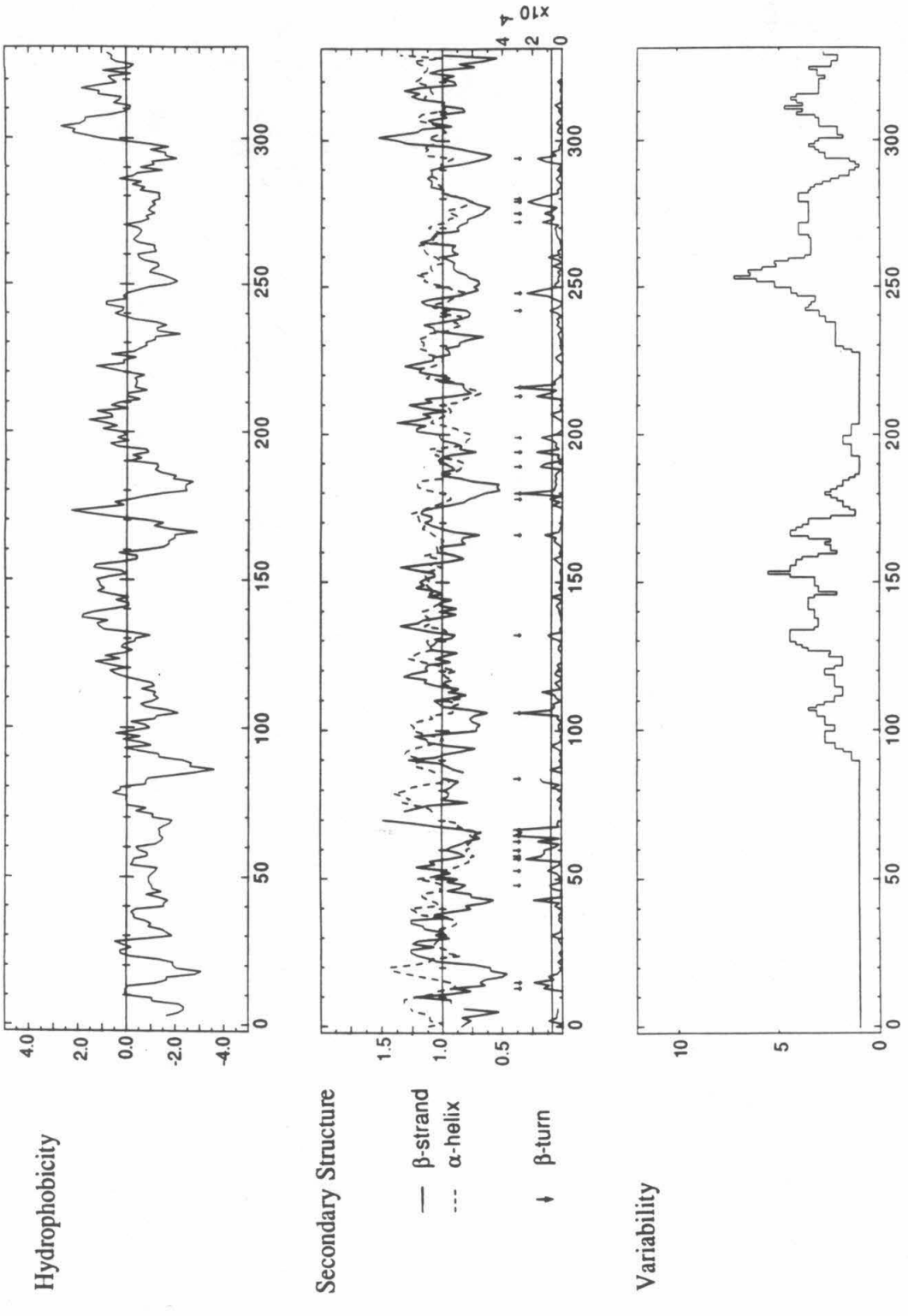


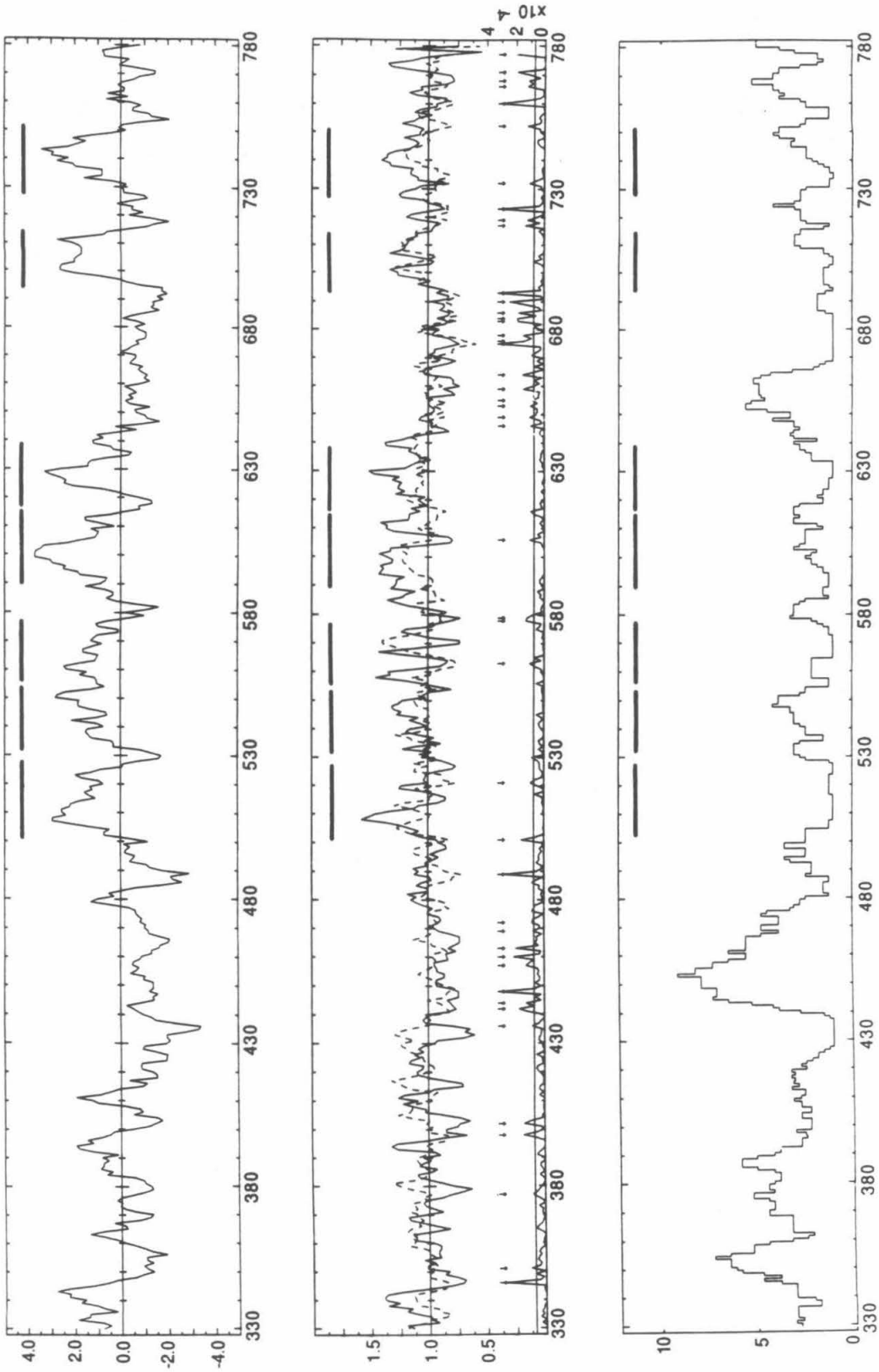
A	= 330	G	= 730
B	= 505	H	= 785
C	= 548	I	= 832
D	= 592	J	= 887
E	= 622	K	= 909
F	= 694	L	= 970

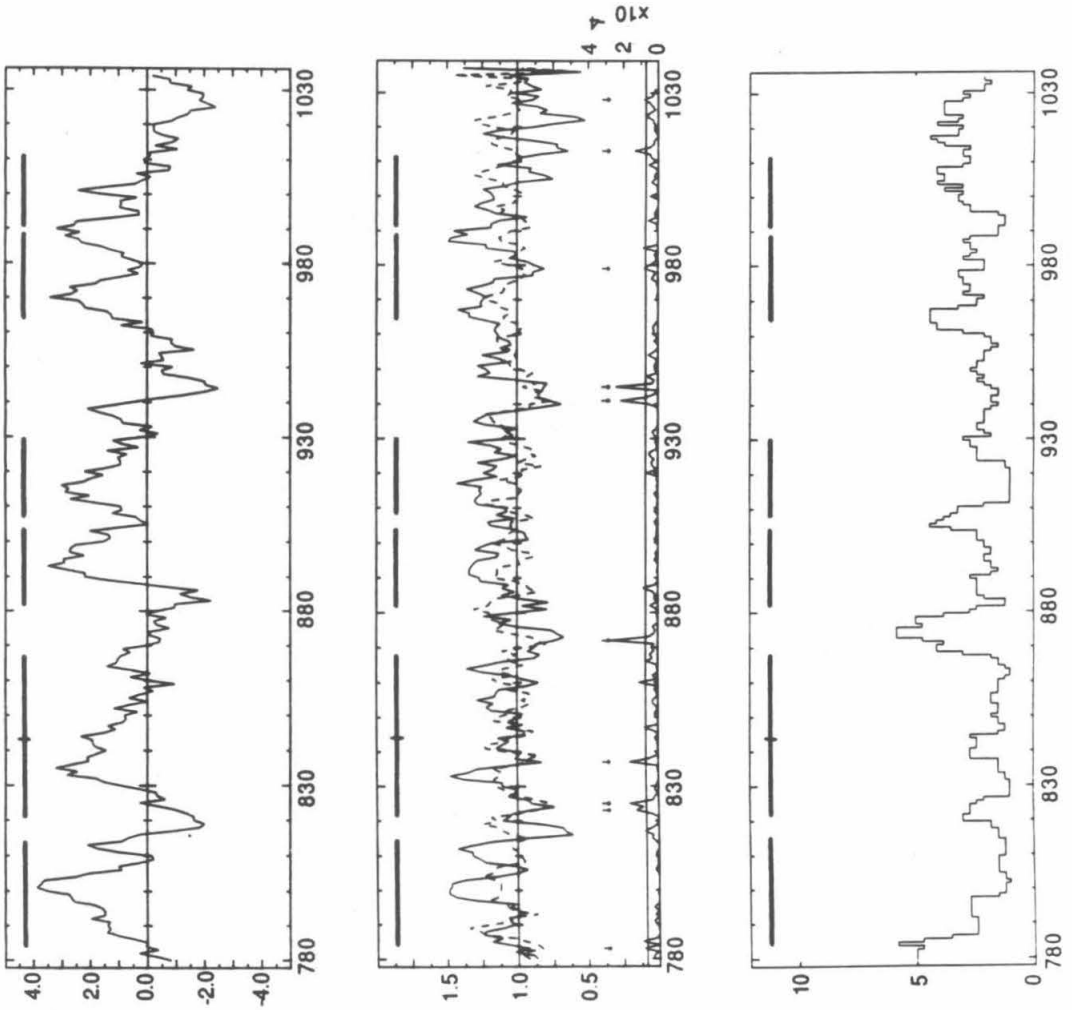
Figure 2.14

Hydropathy, Secondary Structure Potential, and Variability Plots for Band 3

The hydrophobicity, secondary structure potential, and sequence variability for the band 3 sequence alignment has been displayed as a function of residue position. Heavy bars indicate the positions of putative transmembrane helices.


Hydrophobicity
4.0
2.0
0.0
-2.0
-4.0
0
50
100
150
200
250
300
Secondary Structure
β-strand
α-helix
β-turn
1.5
1.0
0.5
x10 4
4
2
0
Variability
10
5
0
Residue Number


4.0
2.0
0.0
-2.0
-4.0
330
380
430
480
530
580
630
680
730
780
1.5
1.0
0.5
x10 4
2
0
10
5
0


4.0
2.0
0.0
-2.0
-4.0
780
830
880
930
980
1030
1.5
1.0
0.5
x10 4
4
2
0
10
5
0

been published (81,82). In figure 2.15, known proteolytic information and other pertinent topological data have been depicted in a schematic diagram.

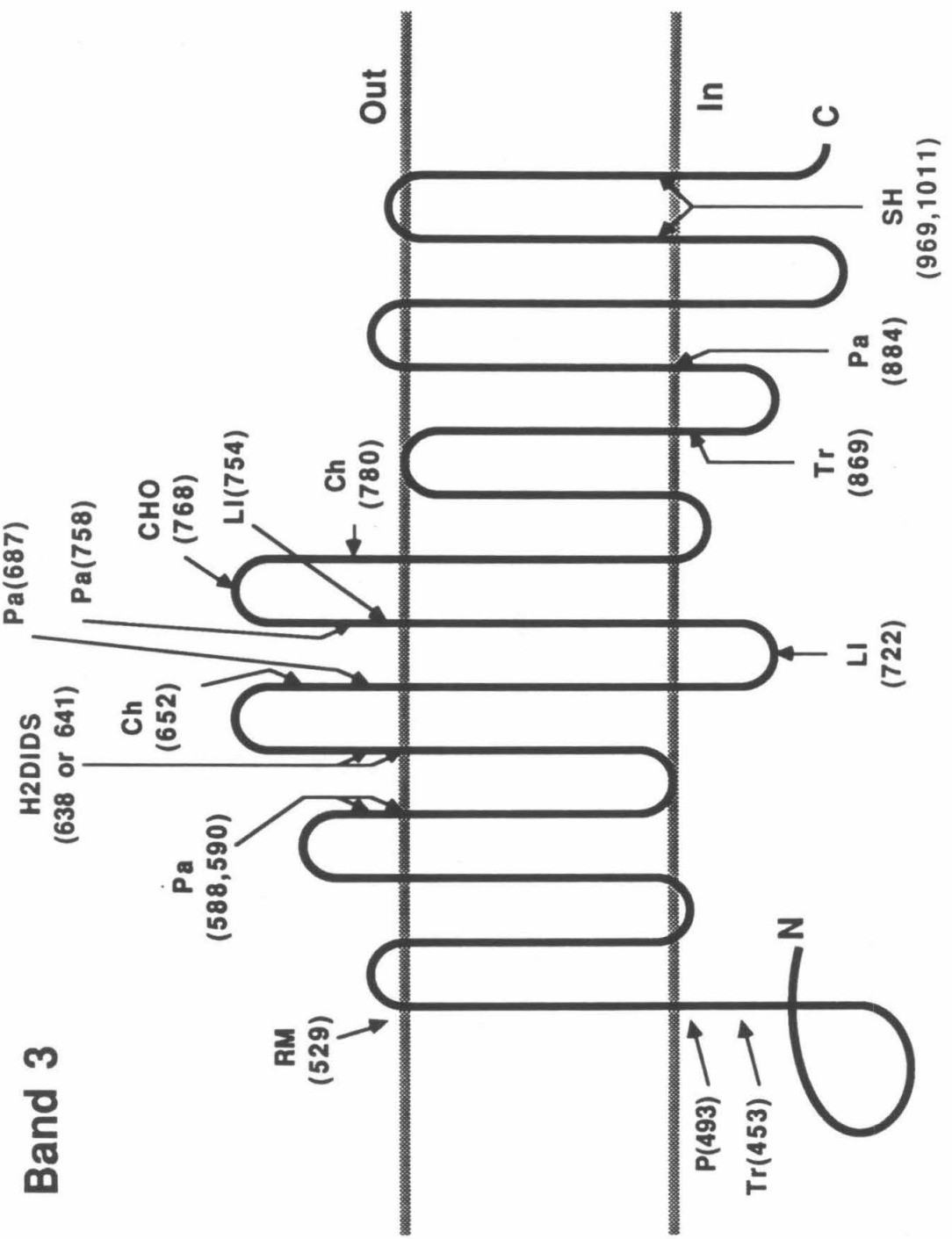
The N-terminus of the protein is known to be intracellular. Intracellular tryptic cleavage at residue 453 removes most of this intracellular domain and leaves the transporting part of band 3 intact within the membrane (68,74). The pepsin cleavage site, which occurs at residue 493 (75), is also assumed to be intracellular because there is no intervening hydrophobic region. A lysine residue corresponding to residue 529 was shown to be extracellular by reductive methylation (83), so the first transmembrane helix is located between residues 493 and 529. This helix, M1, was assigned to 503-529.

An extracellular chymotryptic cleavage site occurs at residue 652 (77), so between residues 529 and 652, the peptide must traverse the membrane an even number of times. Only three helices are predicted for this region by the Eisenberg anaysis, so either one region spans the membrane twice or one region does not span the membrane. The hydrophobic region between 530 and 653 could span the bilayer twice so helices M2 and M3 will be assigned as residues 534-555, and 558-578, respectively. Helices M4 and M5 as defined by the protocol are 591-616 and 618-640. In confirmation of these assignments, papain cleavage sites have recently been identified at residues 588 and 590 (76). One of the lysine residues at positions 638 or 641 can be labelled from outside the cell by 4,4'-diisothiocyano-dihydrostilbene-2,2'-disulfonate (84,85), which is also in accord with these assignments.

Figure 2.15

Schematic Diagram of Band 3 Topology

The topology and sites of proteolytic cleavage have been indicated in this schematic diagram. Some sites of chemical labelling have also been included. No attempt was made to represent the topology of the cytoplasmic domain. Numbers in parentheses refer to residue positions according to the alignment in figure 2.12. Abbreviations used are: **C**, C-terminus; **Ch**, chymotryptic cleavage site; **CHO**, glycosylation site; **H_2DIDS**, covalent labelling site of 4,4'-diisothiocyano-dihydrostilbene-2,2'-disulfonate; **LI**, site of lactoperoxidase-catalyzed radioiodination; **N**, N-terminus; **P**, pepsin cleavage site; **Pa**, papain cleavage site; **RM**, site of reductive methylation; **SH**, cysteine; **Tr**, trypsin cleavage site.


Band 3
Out
In
C
N
SH
(969,1011)
Pa
(884)
Tr
(869)
LI
(722)
LI(754)
CHO
(768)
Ch
(780)
Pa(758)
Pa(687)
H2DIDS
(638 or 641)
Ch
(652)
Pa
(588,590)
RM
(529)
P(493)
Tr(453)

An extracellular papain cleavage site at residue 687 has been identified (86). Because the region between the extracellular chymotryptic cleavage at 652 and this site is short and hydrophillic, it is unlikely to span the membrane. An extracellular papain cleavage site after residue 758 (87,88), indicates that the region between residues 687 and 758 must also contain an even number of transmembrane helices. These regions, corresponding to 695-715 and 729-752, compose helices M6 and M7. These assignments correlate well with the evidence from lactoperoxidase-catalyzed radioiodination, which indicates that Tyr 754 is extracellular and Tyr 722 is intracellular (77,81,82).

A chymotryptic cleavage site occurs at position 780 (75), and because, again, there is no hydrophobic domain between this site and the extracellular papain site at 758, residue 780 is presumed to be extracellular. The intracellular trypsin cleavage site at 869 (78) indicates that the peptide must traverse the membrane an odd number of times, although the Eisenberg analysis and inspection of the hydropathic plot indicate that two spans are more likely. The region from residues 820 to 860 is generally more hydrophobic than the region from 780 to 810, so this region will be chosen to traverse the membrane twice. Helices M8, M9, and M10 are therefore assigned to the residues 786-816, 825-847, and 848-869.

A papain cleavage site is located at position 884 (76) and is assumed to be intracellular because no hydrophobic stretch separates it from the intracellular tryptic cleavage site at 869. The intracellular C-terminus requires that the remaining section of the peptide span the membrane an even number of times. Again, a region of the

peptide, which is not predicted as transmembranous, must cross the bilayer or a section predicted to reside within the membrane must be extramembranous. It seems unlikely that three helices could be contained in the region between the papain cleavage site at 884 and the end of this hydrophobic region at 935, a region for which the Eisenberg algorithm predicts two helices. A more likely location is between residues 955 and 1000. The helices for this final region of the protein are predicted to be between residues 886-906, 911-932, 967-991, and 994-1013.

The positions of introns within the protein coding is indicated in figure 2.12. Of the 12 introns, five are located within predicted transmembrane regions and four of these are within five residues of the helix terminus. Only the intron which interrupts the DNA coding for M11 is within the hydrophobic center of the helix.

The variability plot in figure 2.13 indicates that, in general, the putative transmembrane regions of band 3 are more conserved than those regions which are proposed to be extramembranous.

Discussion

In total, 14 helices were assigned to the transmembrane region of band 3, three more than would have been predicted in the absence of experimental data. These helices correspond to: M1(503-529), M2(534-555), M3(558-578), M4(591-616), M5(618-637), M6(695-715), M7(729-752), M8(786-816), M9(825-847), M10(848-869), M11(886-906), M12(911-932), M13(967-991), and M14(994-1013). Of the helices predicted, eight are bordered on at least one end by β-turns. Five contain predicted

β-turns, which may result from the limited number of sequences included. All the other edges are bordered by either prolines or basic residues with the exception of the N-terminus of helix 13, which was assigned based on the distance from the assigned C-terminus. The assignment of helices 2-3 , 9-10, and 13-14 were somewhat difficult on the basis of predicted β-turns, prolines, and basic residues because more than one assignment could be made. Helices 2-3 and 9-10 could be assigned equally well as a short loop which would not reach the opposite face of the membrane. At this time, the longer loops will be chosen because the existence of a β-turn, or any other type of turn, within the hydrocarbon region of the lipid bilayer is thermodydamically unfavorable. Turns within the hydrophobic region of the bilayer would leave polar, backbone carbonyl moieties exposed to the hydrocarbon region without stabilization from hydrogen-bonding. Such turns might be protected from the lipid bilayer by other trans-membrane helices, but this possibility will not be considered in this work. The assignment of helices 13-14 is consistent with the accessibility of cytosolic reagents to the cysteine residues which are located at positions 969 and 1011. The positions of these residues within the helices also is consistent with the observation that under some conditions, they form a disulfide linkage (89). The recent identification of the pyridoxal phosphate binding-site at residue 977 (79) also helps to confirm the membranous location of this residue. This site has been shown to be an integral part of band 3's transport mechanism.

The helices, as assigned, accommodate all known proteolytic in-

formation and correctly place residue 768, the probable place of glycosylation (90,91), in an extracellular location.

The assignment scheme of section 2B failed to identify three of the transmembrane helices. This may be due to the fact that the Eisenberg algorithm searches for 21-residue helices. If this length were shortened a little to 18 residues, the two helices assigned as M2 and M3 may both have been identified. The other two helices assigned to be transmembrane, M10 and M14, are significantly hydrophilic and would not have been predicted by the Eisenberg algorithm, no matter what helix-length was sought. It is possible in a protein with so many transmembrane helices that some of them will be completely buried. In such a case, it is likely that the buried helices will have properties quite different from those that are lipid-exposed, and will not be easily identified with the current, or perhaps any, scheme for their prediction.

2F. Conclusions

Variability

The variability of the transmembrane helices with respect to the overall variability of the membrane protein seems to depend upon the arrangement of the helices within the bilayer. Proteins such as rhodopsin and the reaction center have transmembrane helices which are all in contact with the lipid. These regions tend to be as variable as the protein as a whole. On the other hand, proteins like the acetylcholine receptor and band 3 may have a substantial number of buried helices, *i.e.*, helices that are not in contact with the lipid. Both of these proteins possess a large number of transmembrane helices, 20 in the case of the assembled acetylcholine receptor, and 28 in the case of a band 3 dimer. It is probable that these helices will pack in such a way that lipid molecules will be excluded as much as possible, leaving some helices buried. The calculated volume of atoms within the reaction center (17) and the helix spacing observed for bacteriorhodopsin (92) suggests that, indeed, lipids are excluded from between adjacent helices. The transmembrane helices of band 3 and the acetylcholine receptor generally were found to be less variable than protein as a whole, a condition which may relate to more stringent requirements of helix-helix packing compared to helix-lipid interactions.

Genetic Structure

In total, thirty transmembrane helices have been examined in this

work, five from the reaction center and 25 putative membrane helices from the nicotinic acetylcholine receptor, rhodopsin, and band 3. Of these helices, only two are interrupted in their central regions by introns. If the intron positions of the D1 and D2 proteins of photosystem II are assumed to correlate to the L and M proteins, this number increases to three. A study of several diverse membrane proteins (12) also indicated that transmembrane regions are not generally interrupted by introns in their hydrophobic core regions, although there was at least one exception. One explanation of these observations lies in the theory that exons correspond to the functional and/or structural domains of a protein. This theory (93-95) was developed shortly after the discovery of introns and RNA splicing and included the idea that exon shuffling in the DNA or variations in splicing would aid the evolution of new proteins through the addition and deletion of functional domains. In the case of membrane proteins, these domains might include an extramembranous loop, a transmembrane helix, or groups of transmembrane helices, but probably not half of a helix. The observation that only about ten percent of the membrane helices examined are split by introns supports these ideas.

Another explanation based on probability and the DNA sequence is possible, however. The consensus sequences, which indicate the beginning and end of an intron, have been determined for eucaryotic organisms. The sequence denoting the start of an intron is $\binom{\mathbf{C}}{\mathbf{A}}$**AGGT**$\binom{\mathbf{A}}{\mathbf{G}}$**AGT** and that denoting the end of an intron is $\binom{\mathbf{T}}{\mathbf{C}}_{\geq 11}$**N**$\binom{\mathbf{C}}{\mathbf{T}}$**AGG**, where the boldface type indicates exonic bases (96). An analysis of the exonic sequence $\binom{\mathbf{C}}{\mathbf{A}}$ **AGG** to determine which amino acids are likely to be

associated with an intron, reveals that approximately 54% of splice junctions should contain positively charged residues, which often occur at the termini of transmembrane helices, and approximately 92% of the junctions should contain at least one hydrophilic residue. If no factors other than the consensus sequence operate to determine the location of splice junctions, one would expect them to be located in the more hydrophilic regions of the protein. Because hydrophobic interactions are believed to stabilize the structures of globular proteins and to determine the topologies of membrane proteins, splice junctions may have developed in this manner so that these regions were avoided. In support of this view is the observation that splice junctions map to the hydrophilic surfaces of many globular proteins where turn regions occur (97). The consensus sequence also would allow for about 50% of the splice junctions to contain one of the four residues with the highest β-turn potential according to Chou and Fasman (20).

This statistical analysis of the consensus sequence indicates that the occurence of introns within the coding regions of transmembrane helices is unfavored, but not forbidden. Approximately 8% of the splice junctions could exclusively contain hydrophobic residues and thus be permitted within a hydrophobic core region. In general, splice junctions are expected to map to hydrophilic regions, especially those involving turns or perhaps positive residues. Such regions are similar to those that can be found at the ends of some transmembrane helices.

The above discussion was based solely upon the analysis of a con-

sensus sequence. A more rigorous investigation would take common variants of this sequence into account, as well as the frequency of hydrophilic residues within transmembrane regions. Unfortunately, there is not much experimental evidence available for an accurate assessment of the latter in membrane proteins with multiple membrane-spanning regions. Only predicted assignments such as those made herein are available in most cases.

Assignment of Transmembrane Helices

In general, the criterion developed in chapter 2 worked well for the prediction of transmembrane helices for the acetylcholine receptor and rhodopsin. For band 3, however, experimental evidence had to be incorporated during the assignment procedure. The reason band 3 is anomalous is probably that this protein contains enough transmembrane helices that some of them may be completely protected from the lipid environment. These helices might have properties different from lipid-exposed helices. If helices were close-packed, any protein with six or more predicted transmembrane regions might also have an unidentified, buried helix. For such proteins, experimental evidence concerning the protein's topology should be included in the prediction procedure whenever possible.

References

1. Flewelling, R.F., and Hubbel, W.L.(1986) *Biophys. J.* **49**, 541-542.
2. von Heijne, G.(1981) *Eur. J. Biochem.* **120**, 275-278.
3. von Heijne, G., and Blomberg, C.(1979) *Eur. J. Biochem.* **97**, 175-181.
4. Brandl, C.J., and Deber, C.M.(1986) *Proc. Natl. Acad. Sci. USA* **83**, 917-921.
5. Engelman, D.M., Steitz, T.A., and Goldman, A.(1986) *Ann. Rev. Biophys. Biophys. Chem.* **15**, 32-53.
6. von Heijne, G.(1986) *EMBO J.* **5**, 3021-3027.
7. Singer, S.J., Maher, P.A., and Yaffe, M.P.(1987) *Proc. Natl. Acad. Sci. USA* **84**, 1960-1964.
8. Meyer, D.I., Krause, E., and Dobberstein, B.(1982) *Nature* **297**, 647-650.
9. Wickner, W.T., and Lodisch, H.F.(1985) *Science* **230**, 400-407.
10. Fahnestock, S.R., Alexander, P., Nagle, J., and Filpula, D.(1986) *J. Bacteriol.* **167**, 870-880.
11. Goldman, D., Dennis, E., Luyten, W., Kochhar, A., Patrick, J., and Heinemann, S.(1987) *Cell* **48**, 965-973.
12. Argos, P. and Rao, J.K.M.(1985) *Biochim. Biophys. Acta* **827**, 283-297.
13. Kyte, J. and Doolittle, R.F.(1982) *J. Mol. Biol.* **157**, 105-132.
14. Cornette, J.L., Cease, K.B., Margalit, H., Spouge, J.L., Berzofsky, J.M., and DeLisi, C.(1987) *J. Mol. Biol.* **195**, 659-685.
15. Eisenberg, D., Swartz, E., Komaromy, M., and Wall, R.(1984) *J. Mol. Biol.* **179**, 125-142.
16. Wallace, B.A., Cascio, M., and Mielke, D.L.(1986) *Proc. Natl. Acad. Sci. USA* **83**, 9423-9427.

17. Yeates, T.O., Komiya, H., Rees, D.C., Allen, J.P., and Feher, G.(1987) *Proc. Natl. Acad. Sci. USA* **84**, 6438-6442.

18. Needleman, S.B., and Wunsch, C.D.(1970) *J. Mol. Biol.* **48**, 443-453.

19. Kyte, J., and Doolittle, R.F.(1982) *J. Mol. Biol.* **157**, 105-132.

20. Chou, P.Y., and Fasman, G.D.(1974) *Biochemistry* **13**, 222-245.

21. Wu, T.T., and Kabat, E.A.(1970) *J. Exp. Med.* **132**, 211-250.

22. Williams, J.C., Steiner, L.A., Feher, G., and Simon, M.I.(1984) *Proc. Natl. Acad. Sci. USA* **81**, 7303-7307.

23. Williams, J.C., Steiner, L.A., Ogden, R.C., Simon, M.I., and Feher, G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6505-6509.

24. Youvan, D.C., Bylina, E.J., Alberti, M., Begusch, H., Hearst, J.E.(1984) *Cell* **37**, 949-957.

25. Michel, H., Weyer, K.A., Gruenberg, H., Dunger, I., Oesterhelt, D., and Lottspeich, F.(1986) *EMBO J.* **5**, 1149-1158.

26. Simon, S.A., and McIntosh, T.J.(1986) *Methods Enzymol.* **127**, 511-521.

27. Lewis, B.A., and Engelman, D.M.(1983) *J. Mol. Biol.* **166**, 203-210.

28. Allen, J.P., Feher, G., Yeates, T.O., Komiya, H., and Rees, D.C.(1987) *Proc. Natl. Acad. Sci. USA* **84**, 6162-6166.

29. Holschuh, K., Bottomly, W., and Whitfeld, P.R.(1984) *Nucl. Acids Res.* **12**, 8819-8834.

30. Rasmussen, O.F., Bookjans, G., Stummann, B.M., and Henningsin, K.W.(1984) *Plant Mol. Biol.* **3**, 191-199.

31. Ohyama, K.(1986) Sequence submitted to Genbank.

32. Zurawski, G., Bonner, H.J., Whitfeld, P.R., and Bottomly, W.(1982) *Proc. Natl. Acad. Sci. USA* **79**, 7699-7703.

33. Karabin, G.D., Farley, M., and Hallick, R.B.(1984) *Nucl. Acids Res.* **12**, 5801-5812.

34. Erickson, J.M., Rahire, M., and Rochaix, J.-D.(1984) *EMBO J.* **3**, 2753-2762.

35. Karlin, A.(1980) *The Cell Surface and Neuronal Function*, Poste, G., Nicolson, G.L., and Cotman, C.W., eds., Elsevier: New York, 191-260.

36. Kurosaki, T., Fukuda, K., Konns, T., Mori, Y., Tanaka, K., Mishina, M., and Numa, S.(1987) *FEBS Lett.* **214**, 253-258.

37. Noda, M., Takahashi, H., Tsutomu, T., Toyosato, M., Kikyotani, S., Furitani, Y., Hirose, T., Takashima, H., Imayama, S., Miyata, T., and Numa, S.(1983) *Nature* **302**, 528-532.

38. Noda, M., Furitani, Y., Takahashi, H., Toyosato, M., Tanabe, T., Shimizu, S., Kikyotani, S., Kayano, T., Hirose, T., Inayama, S., and Numa, S. (1983) *Nature* **305**, 818-823.

39. Boulter, J., Evans, K., Goldman, D., Martin, G., and Treco, D.(1986) *Nature* **319**, 368-374.

40. Hermans-Borgmeyer, I., Zopf, D., Ryseck, R.-P., Hovemann, B., Betz, H., and Gundelfinger, E.D.(1986) *EMBO J.* **5**, 1503-1508.

41. Tanabe, T., Noda, M., Furutani, Y., Takai, T., Takahashi, H., Tanaka, K., Hirose, T., Inayama, S., and Numa, S.(1984) *Eur. J. Biochem.* **144**, 11-17.

42. Buonanno, A., Mudd, J., Shah, V., and Merlie, J.P.(1986) *J. Biol. Chem.* **261**, 16451-16458.

43. Takai, T., Noda, M., Furutani, Y., Takahashi, H., Notake, M., Shimizu, S., Kayano, T., Tanabe, T., Tanaka, K., Hirose, T., Inayama, S., and Numa, S. (1984) *Eur. J. Biochem.* **143**, 109-115.

44. Nef, P., Mauron, A., Stalder, R., Alliod, C., and Ballivet, M.(1984) *Proc. Natl. Acad. Sci. USA* **81**, 7975-7979.

45. Shibara, S., Kubo, T., Perski, H.J., Takahashi, H., Noda, M., and Numa, S. (1985) *Eur. J. Biochem.* **146**, 15-22.

46. Kubo, S., Noda, M., Takai, T., Tanabe, T., Kayano, T., Shimizu, S., Tanaka, K., Takahashi, H., Hirose, T., Inayama, S., Kikuno, R., Miyata, T., and Numa, S.(1985) *Eur. J. Biochem.* **149**, 5-13.

47. LaPolla, R.J., Mayne, K.M., and Davidson, N.(1984) *Proc. Natl. Acad. Sci. USA* **81**, 7970-7974.

48. Stroud, R.M., and Finer-Moore, J.(1985) *Ann. Rev. Cell Biol.* **1**, 317-351.

49. Kosower, E.M.(1987) *Eur. J. Biochem.* **168**, 431-449.

50. Guy, H.R.(1984) *Biophys. J.* **45**, 249-261.

51. Finer-Moore, J., and Stroud, R.M.(1984) *Proc. Natl. Acad. Sci. USA* **81**, 155-159.

52. Popot, J.-L., and Changeux, J.-P.(1984) *Physiol. Rev.* **64**, 1162-1239.

53. Huganir, R.L., Delcour, A.H., Greegard, P., and Hess, G.P.(1986) *Nature* **321**, 774-777.

54. McCrea, P.D., Popot, J.-L., and Engelman, D.M.(1987) *EMBO J.* **6**, 3619-3626.

55. Ratnam, M., Sargent, P.B., Sarin, V., Fox, J.L., and Lindstrøm, J.(1986) *Biochemistry* **25**, 2621-2632.

56. McCrea, P.D., Engelman, D.M., and Popot, J.-L.(1988) *Trends Biochem. Sci.* **13**(8), 289-290.

57. Corless, J.M., McCastin, D.R., and Scott, B.L.(1982) *Proc. Natl. Acad. Sci. USA* **79**, 116-120.

58. Findlay, J.B.C.(1986) *Photobiochem. Photobiophys.* **13**, 213-228.

59. Davidson, M.D., and Findlay, J.B.C.(1986) *Biochem. J.* **234**, 413-420.

60. Davidson, M.D., and Finlay, J.B.C.(1986) *Biochem. J.* **236**, 389-395.

61. Nathans, J., Thomas, D., and Hogness, D.(1986) *Science* **232**, 193-202.

62. Nathans, J., and Hogness, D.S.(1984) *Proc. Natl. Acad. Sci. USA* **81**, 4851-4855.

63. Nathans, J., and Hogness, D.S.(1983) *Cell* **34**, 807-814.

64. Brett, M., and Findlay, J.B.C.(1983) *Biochem. J.* **211**, 661-670.

65. Pappin, D.J.C., Eliopoulos, E., Brett, M., and Finlay, J.B.C.(1984) *Int. J. Biol. Macromol.* **6**, 73-76.

66. Aveldaño, M.I.(1988) *Biochemistry* **27**, 1229-1239.

67. Barclay, P.L., and Findlay, J.B.C.(1984) *Biochem. J.* **220**, 75-84.

68. Grinstein, S., Ship, S., and Rothstein, A.(1978) *Biochim. Biophys. Acta* **507**, 294-304.

69. Lieberman, D.M., Nattriss, M., and Reithmeier, R.A.F.(1987) *Biochim. Biophys. Acta* **903**, 37-47.

70. Cox, J.V., and Lazarides, E.(1988) *Mol. Cell. Biol.* **8**(3), 1327-1335.

71. Kopito, R.R., and Lodish, H.F.(1985) *J. Cell Biochem.* **29**, 1-17.

72. Demuth, D.R., Showe, L.C., Ballantine, M., Palumbo, A., Fraser, P.J., Cioe, L., Rovera, G., and Curtis, P.J.(1986) *EMBO J.* **5**, 1205-1214.

73. Kaul, R.K., Murthy, S.N.A., Reddy, A.G., Steck, T.L., and Kohler, H.(1983) *J. Biol. Chem.* **258**, 7981-7990.

74. Mawby, W.J., and Findlay, J.B.C.(1983) *Biochem J.* **205**, 465-475.

75. Brock, C.J., and Tanner, M.J.A.(1986) *Biochem. J.* **235**, 899-901.

76. Kanes, K.J., Vogelaar, N.J., and Chan, S.I. (unpublished results).

77. Brock, C.J., Tanner, M.J.A., and Kempf, C.(1983) *Biochem. J.* **213**, 577-586.

78. Jennings, M.L., Anderson, M.P., and Monaghan, R.(1986) *J. Biol. Chem.* 261, 9002-9010.

79. Kawano, Y., Okubo, K., Tokunaga, F., Miyata, T., Iwanaga, S., and Hamasaki, N.(1988) *J. Biol. Chem.* 263, 8232-8238.

80. Kopito, R.R., Andersson, M., and Lodish, H.F.(1982) *J. Biol. Chem.* **262**, 8035-8040.

81. Jay, D., and Cantley, L.(1986) *Ann. Rev. Biochem.* **55**, 511-538.

82. Passow, H.(1986) *Rev. Physiol. Biochem. Pharmacol.* **103**, 61-223.

83. Jennings, M.L., and Nicknish, J.S.(1984) *Biochemistry* **23**, 6432-6436.

84. Jennings, M.L., and Passow, H.(1979) *Biochim. Biophys. Acta* **554**, 498-519.

85. Rudloff, V., Lepke, S., and Passow, H.(1983) *FEBS Lett.* **163**, 14-21.

86. Jennings, M.L., Adams-Lackey, M., and Denney, G.H.(1984) *J. Biol. Chem.* **259**, 4652-4660.

87. Drickamer, L.K.(1976) *J. Biol. Chem.* **251**, 5115-5123.

88. Jennings, M.L., and Adams, M.F.(1981) *Biochemistry* **20**, 7118-7123.

89. Rao, A.(1979) *J. Biol. Chem.* **254**, 3503-3511.

90. Jay, D.G.(1986) *Biochemistry* **25**, 554-556.

91. Kopito, R.R., and Kodish, H.F.(1985) *Nature* **316**, 234-238.

92. Henderson, R., and Unwin, P.N.T.(1975) *Nature* **257**, 28-32.

93. Gilbert, W.(1978) *Nature* **271**, 501.

94. Blake, C.C.F.(1978) *Nature* **273**, 267.

95. Blake, C.C.F.(1979) *Nature* **277**, 598.

96. Mount, S.M.(1982) *Nucl. Acids Res.* **10**, 459-472.

97. Craik, C.S., Rutter, W.J., and Flettrick, R.(1983) *Science* **220**, 1125-1129.

CHAPTER 3

VARIABILITY ANALYSIS OF TRANSMEMBRANE REGIONS

As mentioned in the previous chapter, various algorithms designed to predict the secondary structure of proteins have been developed, although none have been very successful when applied to membrane proteins (1). From the x-ray crystal structures of bacteriorhodopsin (2), rhodopsin (3), and the reaction center (4-5), and from various studies employing circular dichroism, however, it appears that for most membrane proteins, the secondary structure of the peptide chain within the membrane is α-helical. The three-dimensional structure of the protein within the membrane would be obtained if the spacial relationships between the helices could be determined. Significant progress toward this end could be made even if the disposition of each helix with respect to other helices and the bilayer lipids could be determined. A method for assigning such helix orientations on the basis of sequence variation has been developed and is described below.

In the last chapter it was suggested that those helices which are exposed to lipids tend to be more variable than those which are buried within the protein. Yeates *et al.* (6) have made a similar observation on the level of the primary structure. In their work, it was found that residues of the reaction center which face the lipid bilayer were, in general, less conserved than those which interacted with other helices. If the lipid's hydrocarbon region is considered to serve as a solvent, this variability is analogous to that observed

for globular proteins (7), which are variable on surfaces that are solvent-exposed. Presumably, the requirements for protein-protein interactions are more stringent than those for protein-lipid interactions.

The method described below results in the identification of variable residues, which are probably lipid-exposed in the assembled protein. The positions of these variable residues within their α-helices will only be discussed briefly in this section. The use of this derived information will be dealt with in the next chapter where it plays a significant role in the three-dimensional modelling of the proteins.

Methods

The variability analysis of Wu and Kabat (8) was used to quantitate the diversity of the sequences at each residue position. The definition of variability is:

$$\text{Variability} = \frac{\text{number of different amino acids at a given position}}{\text{frequency of the most common amino acid at that position}}$$

The alignments for the reaction center, the nicotinic acetylcholine receptor, rhodopsin, and band 3 used for these calculations are those found in figures 2.1, 2.6, 2.9, and 2.12, respectively. The regions examined were the experimentally determined transmembrane helices in the case of the reaction center, and the portions of the other proteins predicted in the preceding chapter to be transmembrane.

Residues were designated as either variable or conserved based on the variability values calculated above. The ranges of these values

used to distinguish variable and conserved residues necessarily varied with the protein under study, due to the different numbers of sequences in each analysis and their evolutionary diversity. The ranges used are given in the table below.

Protein	Variable	Conserved
Reaction Center	≥10.0	≤ 2.5
Acetylcholine Receptor	≥ 8.0	≤ 5.0
Rhodopsin	≥ 6.0	≤ 2.5
Band 3	≥ 3.0	≤ 2.9

These values were selected somewhat qualitatively based upon the fraction of amino acids classified. No universal criterion could be derived because of the varying degrees of evolutionary divergence within each sequence group, and because of the complex dependency of the variance parameter upon the number of sequences analyzed.

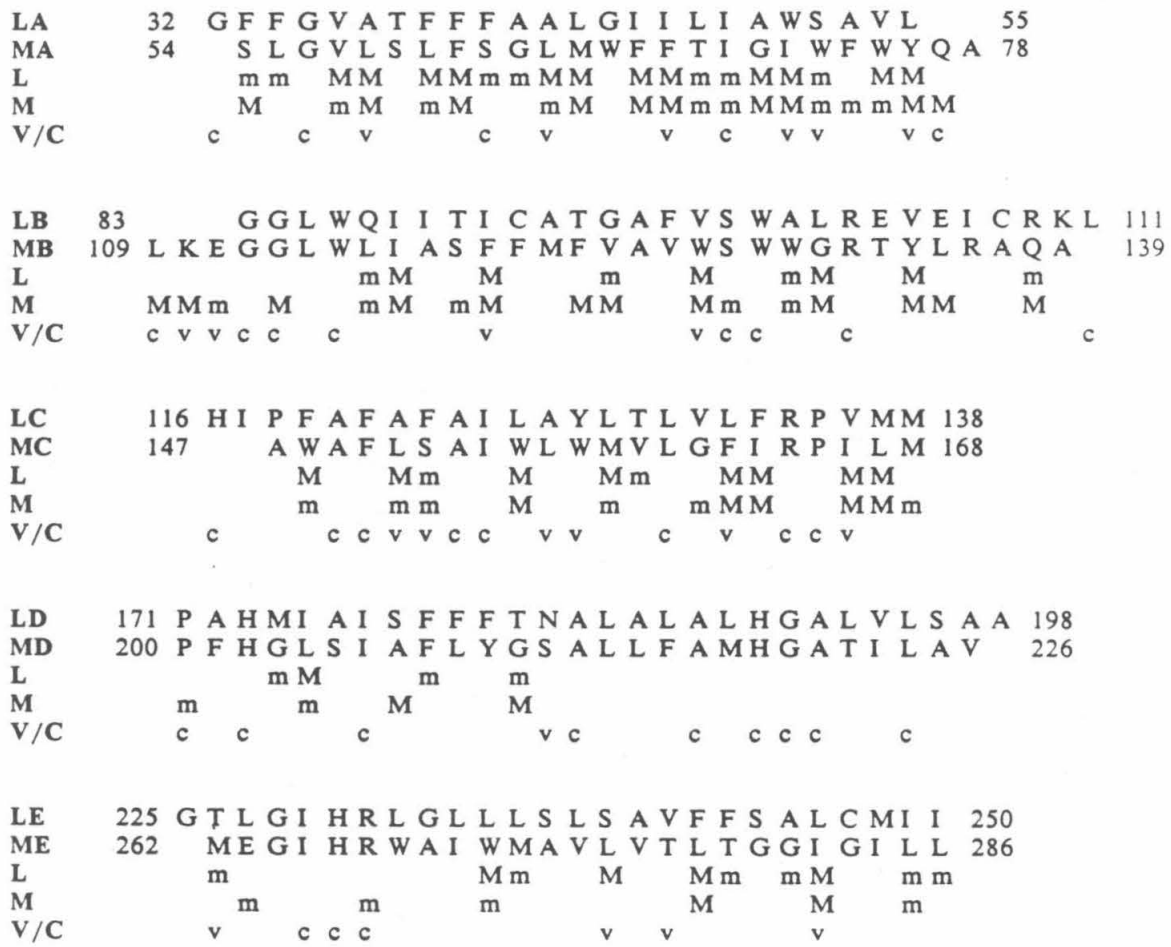
Results

Reaction Center. The residues of the reaction center's transmembrane helices were designated as conserved or variable. The correspondence between these assignments and the exposure of the residues to lipid molecules, as defined by Yeates, *et al.* (6), is shown in figure 3.1. Of the residues designated as conserved, 87-91% correspond to residue positions, which are more than half buried within the protein. The residues designated as conserved, therefore, correspond well to residue positions with less than a 50% lipid-exposure,

Figure 3.1

Variability Analysis of the Reaction Center's Transmembrane Helices

The residues in the transmembrane regions of the reaction center were examined with regard to their evolutionary conservation and assigned as conserved or variable. Shown in this figure are the transmembrane helices and the membrane exposures of each residue as determined by Yeates, *et al.* (6). The sequences of the reaction center of *Rb. sphaeroides* (9) are shown above the lipid accessibility of the residues. Lipid exposure: **M**, >50%, **m**, 20-50%, **blank**, 0-20%. The residues designated conserved, **c**, or variable, **v**, are indicated beneath the lipid accessibility.


LA 32 G F F G V A T F F F A A L G I I L I A W S A V L 55
MA 54 S L G V L S L F S G L M W F F T I G I W F W Y Q A 78
L m m M M M M m m M M M M m m M M m M M
M M m M m M m M M M m m M M m m m M M
V/C c c v c v v c c v v c
LB 83 G G L W Q I I T I C A T G A F V S W A L R E V E I C R K L 111
MB 109 L K E G G L W L I A S F F M F V A V W S W W G R T Y L R A Q A 139
L m M M m m M m m
M M M m M m M m M M M m M M M M M
V/C c v v c c c v v c c c c
LC 116 H I P F A F A F A I L A Y L T L V L F R P V M M 138
MC 147 A W A F L S A I W L W M V L G F I R P I L M 168
L M M m M m M M M M
M m m m M m m M M M M m
V/C c c c v v c c v v c c v
LD 171 P A H M I A I S F F F T N A L A L A L H G A L V L S A A 198
MD 200 P F H G L S I A F L Y G S A L L F A M H G A T I L A V 226
L m M m m
M m m M M
V/C c c c v c c c c c c
LE 225 G T L G I H R L G L L L S L S A V F F S A L C M I I 250
ME 262 M E G I H R W A I W M A V L V T L T G G I G I L L 286
L m M m M m M m m m
M m m m M M m
V/C v c c c v v v

as expected. The variable residues do not correspond as well to the lipid-exposed residues. Of the 21 variable residues, 17, or 81%, coincide with residue positions that are exposed to some extent to the lipid bilayer. The remaining four residue positions are located near the center of helices C, D, and E. Two cofactors of the reaction center, bacteriochlorophyll *b* and bacteriopheophytin *b*, lie in pockets formed, in part, by these three helices. It is impossible to determine from published literature whether or not these four residues face the cofactors, but it is likely that the phytyl and isoprenoid chains of the cofactors are similar to the hydrocarbon chains of the lipids and do not impose many constraints on their interfacial amino acids. Residues which face these cofactors may, therefore, be more variable than residues involved in helix-helix interactions, although they would still be considered "buried" in that they are not exposed to membrane lipids. In support of this idea is the observation that only 52-71% of the variable residues are exposed to the lipid membrane over half of their surfaces, although the location of these residues with respect to the cofactors could not be determined from the literature.

The conserved and variable residues for the reaction center's transmembrane helices were also determined using the sequence alignment of figure 2.5, which includes the D1 and D2 proteins of photosystem II. The assignments did not correspond as well to the structural domains as the designations reported above. This is probably due to the fact that the proteins of this sequence alignment are more diverse. Proteins which are less related can be expected to contain more suppressor mutations, or pairs of compensating mutations, which

may occur in functionally significant regions of the protein. Such mutations would tend to make this type of analysis based on sequence variance less accurate.

Acetylcholine Receptor, Rhodopsin, and Band 3. Residues of these three proteins were assigned as conserved or variable and these assignments are shown in figures 3.2, 3.3, and 3.4. It is interesting to note that in the case of the acetylcholine receptor, residues that are well-conserved within a certain subunit are often found to be variable when all the subunits are taken into consideration. (The sequence alignment for the acetylcholine receptor can be found in figure 2.6.) The degree of evolutionary divergence in the sequences used as a data set may therefore be of great importance in this type of analysis. In addition, it is good to have a large number of sequences to avoid erroneous assignments. If only the α subunit of the acetylcholine receptor was used in this analysis, the residue corresponding to position 232 of the alignment in figure 2.6 might have been assigned as variable but more sequences indicate that this position is actually conserved.

No conclusions about the correlation of variable residues to lipid-exposed regions or the correspondence of conserved residues to buried residues can be made in the case of these proteins until more is known about their three-dimensional structures. These residues do, however, tend to map to one face of an α-helix, as will be shown in the next chapter, in which the three-dimensional modelling of each of these proteins will be discussed. One other observation that comes from three-dimensional computer modelling will be mentioned here.

Figure 3.2

Variability Analysis of the Acetylcholine Receptor's

Transmembrane Helices

The residues designated as conserved, c, or variable, v, have been indicated below the sequences of the transmembrane helices for this receptor. Sequences of the mouse acetylcholine receptor (10-13) are shown.

α M1	211	P	L	Y	F	I	V	N	V	I	I	P	C	L	L	F	S	F	L	T	S	L	V	F	Y	L	235
β M1	222	P	L	F	Y	L	V	N	V	I	A	P	C	I	L	I	T	L	L	A	I	F	V	F	Y	L	246
γ M1	219	P	L	F	Y	V	I	N	I	I	A	P	C	V	L	I	S	S	V	A	I	L	I	Y	F	L	243
δ M1	225	P	L	F	Y	I	I	N	I	L	V	P	C	V	L	I	S	F	M	I	N	L	V	F	Y	L	249
V/C		c	c	c	c	v	c	c	v	c	v	c	c		c	c	c		c	v	v	c	c	c	c	c				

α M2	240	G	E	K	M	T	L	S	I	S	V	L	L	S	L	T	V	F	L	L	V	I	V	E	L	I	264
β M2	251	G	E	K	M	G	L	S	I	F	A	L	L	T	L	T	V	F	L	L	L	L	A	D	K	V	275
γ M2	249	G	Q	K	C	T	V	A	T	N	V	L	L	A	Q	T	V	F	L	F	L	V	A	D	D	V	273
δ M2	254	G	E	K	T	S	V	A	I	S	V	L	L	A	Q	S	V	F	L	L	L	I	S	K	R	L	278
V/C		c	c	c	v	c		v	c	c	c	c	c				c	c	c	c	c	v	v	v	v					

α M3	277	Y	M	L	F	T	M	V	F	V	I	A	S	I	I	I	T	V	I	V	I	N	T	H	H	T	201
β M3	288	Y	L	M	F	T	M	V	L	V	T	F	S	V	I	L	S	V	V	V	L	N	L	H	H	R	312
γ M3	286	Y	L	T	F	L	M	V	V	T	I	L	I	V	V	N	S	V	V	V	L	N	V	S	L	R	310
δ M3	291	F	L	L	F	G	M	V	L	V	T	M	V	V	V	I	C	V	I	V	L	N	I	H	F	R	315
V/C		c	c	v	c	v	c	v	v		v	v		c	v	v	v	c		c	c	c	v	c	v	c				

α M4	408	H	I	L	L	G	V	F	M	L	V	C	L	I	G	T	L	A	V	F	A	G	R	L	I	E	L	433
β M4	446	R	L	F	L	W	T	F	I	V	F	T	S	V	G	T	L	V	I	F	L	D	A	T	Y	H	L	471
γ M4	454	R	V	C	F	L	A	M	L	S	L	F	I	C	G	T	A	G	I	F	L	M	A	H	Y	N	Q	479
δ M4		R	L	C	L	F	V	V	T	P	V	M	V	V	G	T	A	W	I	F	L	Q	G	V	Y	N	Q	475
V/C		c	v	v	c	v	v	v	v	v	v	v	v	v	v	c	c	v	v	c	c		v	v	v	v	v	v		

Figure 3.3

Variability Analysis of Rhodopsin's Transmembrane Helices

The residues designated as conserved, c, or variable, v, have been indicated below the sequences of the transmembrane helices. The sequence of ovine rhodopsin (14,15) is shown.


M1 35 W Q F S M L A A Y M F L L I V L G F P I N F L T L Y V T V Q H 65
V/C c v v v v c v v v v v c v c c v c v v v
M2 72 L N Y I L L N L A V A D L F M V F G G F T T T L Y T S L 99
V/C c c c c c c c c c c v v v v v v v v c v v v v v
M3 109 G C N L E G F F A T L G G E I A L W S L V V L A I E 134
V/C v c v c c c v v c c v v v v c c c c v v c
M4 148 F G E N H A I M G V A F T W V M A L A C A A P P L V G W S 176
V/C c v v c c v c c c c v v v v v c c c c v v c c c
M5 201 E S F V I Y M F V V H F S I P L I V I F F C Y G Q L V F T V 230
V/C c v c v v c v c c v v c v c c v c c v v v v
M6 253 M V I I M V I A F L I C W L P Y A G V A F Y I F 276
V/C c c c c v c c v c c c c v c v v v
M7 285 P I F M T I P A F F A K S S V Y N P V I Y I M M N 310
V/C c v v v c c c c c c c v c c c c c c c v v v c c

Figure 3.4

Variability Analysis of Band 3's Transmembrane Helices

The residues in the transmembrane regions which have been designated as variable, v, are shown beneath the sequence of murine band 3 (16). Because there were few sequences in this analysis, residues which were not strictly conserved were assigned to be variable. For the sake of clarity, these conserved residues have not been indicated on the diagram.

```
M1  423  Q V L A A V I  F I  Y F A A L S P A V T F G G L L G E K 449
V/C        v                                 v

M2  454  M G V S E L L I S T A V Q G I L F A L L G A 475
V/C        v           v v     v v   v v v   v     v

M3  478  L L V L G F S G P L L V F E E A F F S F C 498
V/C           v                             v

M4  511  W I  G F W L I L L V M L V V A F E G S F L V Q Y I S 536
V/C                 v v   v v   v     v       v     v v v

M5  538  Y T Q E I  F S F L I S L I  F F I  Y E T F S 557
V/C        v   v       v                         v

M6  587  N T A L F S L V L M A G T F L L A M T L R 607
V/C           v v                   v v   v v

M7  621  R V I  G D F G V P I S I  L I  M V L V D S F I  K G 644
V/C        v                 v   v v v v   v   v v   v v

M8  678  P T W M M F A S V L P A L L V F I  L I  F L E S Q I  T T L I  V S 708
V/C           v       v     v       v                 v   v             v

M9  717  G S G F H L D L L L V V G M G G V A A L F G M 739
V/C                        v   v       v v         v

M10 740  P W L S A T T V R S V T H A N A L T W M G K 761
V/C           v   v       v v                 v v

M11 778  R I S G L L V S V L V G L S I  L M E P I  L 798
V/C        v v         v v   v   v     v v v v v

M12 803  L A V L F G I  F L Y M G V T S L S G I  Q L F 824
V/C                              v   v       v v

M13 859  I I C L A V L W V V K S T P A S L A L P F V L I  L 883
V/C      v v v v   v     v   v v v v v       v v     v   v

M14 886  P L R R L I L P L I  F R E L E L Q C L D 905
V/C           v v v   v v     v v v   v v
```

Variable residues are often found near the assigned helix termini, but do not neccessarily correlate with the sidedness of the center of the helix. This sequence variance may reflect a less-constrained environment near the membrane surface.

Discussion

Conserved residues in transmembrane helices correspond well to those regions of the protein which are not exposed to the lipid environment in the case of the reaction center. There is also a strong possibility that those residues which are variable could be used to identify residues exposed to the lipid milieu, although the lipid-like co-factors involved in the reaction center complex make the interpretation of this analysis more difficult. It is anticipated that residues in rhodopsin which are not lipid-exposed, but which are in contact with retinal's hydrocarbon region may be more variable than those residues involved in helix-helix interactions.

This type of analysis seems quite sensitive to the sequences used in the variance determination. Too much diversity within the alignment seems to decrease the correspondence between the variability assignment and the structural role of the residue. Sequences, however, need a certain level of variability in order to increase the probability of making the correct variability assignments. The statistical significance of these assignments might also be improved by increasing the number of sequences included in the analysis. In the application of this type of analysis it is recommended that as many sequences as possible be included, but only sequences which share at

least 25% identity with the other sequences in the group and a group which, on average, contains sequences that are about 40% identical.

References

1. Wallace, B.A., Cascio, M., and Mielke, D.K.(1986) *Proc. Natl. Acad. Sci. USA* **83**, 9423-9427.
2. Henderson, R., and Unwin, P.N.T.(1975) *Nature* **257**, 28-32.
3. Corless, J.M., McCaslin, D.R., and Scott, B.L.(1982) *Proc. Natl. Acad. Sci. USA* **79**, 116-120.
4. Diesenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H.(1985) *Nature* **318**, 618-624.
5. Allen, J.P., Feher, G., Yeates, T.O., Komiya, H., and Rees, D.C.(1987) *Proc. Natl. Acad. Sci. USA* **84**, 6162-6166.
6. Yeates, T.O., Komiya, H., Rees, D.C., Allen, J.P., and Feher, G.(1987) *Proc. Natl. Acad. Sci. USA* **84**, 6438-6442.
7. Smith, E.L.(1970) *The Enzymes, 3rd ed.* **1**, Boyer, P., ed., Academic Press: New York, 267-339.
8. Wu, T.T., and Kabat, E.A.(1970) *J. Exp. Med.* **132**, 211-250.
9. Williams, J.C., Steiner, L.A., and Feher, G.(1986) *Proteins* **1**, 312-325.
10. Boulter, J., Luyten, W., Evans, K., Mason, P., Ballivet, M., Goldman, D., Stengelin, S., Martin, G., Heinemann, S., and Patrick, J.(1985) *J. Neurosci.* **5**, 2545-2552.
11. Buonanno, A., Mudd, J., Shah, V., and Merlie, J.P.(1986) *J. Biol. Chem.* **261**, 16451-16458.
12. Boulter, J., Evans, K., Martin, G., Mason, P., Stengelin, S., Goldman, D., Heinemann, S., and Patrick, J.(1986) *J. Neurosci. Res.* **16**, 37-49.
13. LaPolla, R.J., Mixter-Mayne, K., and Davidson, N.(1984) *Proc. Natl. Acad. Sci. USA* **81**, 7970-7974.

14. Pappin, D.J.C., Eliopoulos, E., Brett, M., and Findlay, J.B.C.(1984) *Int. J. Biol. Macromol.* **6**, 73-76.

15. Brett, M., and Findlay, J.B.C.(1983) *Biochem. J.* **211**, 661-670.

16. Kopito, R.R., and Lodish, H.F.(1985) *Nature* **280**, 468-473.

CHAPTER 4

THREE-DIMENSIONAL MODELLING OF MEMBRANE PROTEINS

4A. Introduction

In the three-dimensional modelling of membrane proteins, several characteristics of protein conformation must be considered. The most basic of these is the packing of helices within the bilayer. Thirty years ago, Crick proposed that α-helices pack together as coiled coils, a structure in which adjacent helices are about 18^{o} from parallel (1). Such packing allows the side chains of one helix to interdigitate with those of its neighbor in what is often called a "knobs-into-holes" packing. This arrangement maximizes the van der Waals attractive forces between the helices. Because the backbone twist of an α-helix is right-handed, the best packing results in left-handed coiled coils. These left-handed coils are a general motif in globular proteins (2-4) and have been identified in the transmembrane regions of bacteriorhodopsin (5), rhodopsin (6), and the reaction center (7,8). The low-resolution structure of the gap junction, which was obtained using electron image analysis, also shows a general left-handed twist in the transmembrane region (9), although the acetylcholine receptor does not (10). In their work, Cohen and Parry (2) discuss the properties of interlocking helices in terms of a heptad (*a*, *b*, *c*, *d*, *e*, *f*, and *g*), which repeats along the length of interhelix contact. The residues as positions *a* and *d* are proposed to be apolar and to form the surfaces where the helices mesh. Because many of the residues in a membrane protein are apolar, this character-

istic is probably not useful in the identification of regions of helix contact. The left-handed coiling of transmembrane helices may, however, be an important structural motif in membrane proteins.

A second characteristic of α-helices, which may affect their packing within the membrane, is their dipole moment. It has been proposed by several authors (11-13) that the nearly parallel alignment of the backbone amide linkages results in a large dipole moment. The electrostatic interactions of adjacent, antiparallel helices are, therefore, believed to result in a significant net stabilization of the protein and adjacent, parallel helices in a net destabilization of the protein. It has been calculated that for a 4-helix bundle, the antiparallel grouping results in a net stabilization of 5-7 kcal. per mole whereas the parallel arrangement results in a destabilization of about 20 kcal/mole (14). These electostatic interactions have been suggested to play a major role in the structural stabilization of the reaction center (15).

Interactions between specific side chains may dictate the arrangement of helices within the bilayer in some instances. Such interactions could include the formation of hydrogen bonds and salt bridges. Salt bridges have been proposed to play a major role in the stabilization of the bacteriorhodopsin molecule (16), although they do not seem important in the stabilization of the reaction center (15).

One obvious limitation in the positioning transmembrane helices with respect to one another is the length of the extramembranous loops that separate them. Engelman *et al.* (16) used this constraint to reduce the number of helical arrangements possible in the modelling of

bacteriorhodopsin.

Finally, mechanisms relating to the method of protein insertion or assembly could operate to influence the protein's final structure. Any such effects are not predictable from the primary structure at this point in time, but it can be said that transmembrane regions which are sequential in the sequence are not necessarily adjacent. The reaction center is an example of a protein that has a non-sequential packing of the helices (7,8,15).

The presence of lipid molecules between transmembrane helices is not probable for either the reaction center (15) or bacteriorhodopsin (5). Their exclusion may result if helix-helix interactions are more favorable than helix-lipid interactions.

In the following sections, the modelling of band 3, rhodopsin, and the acetylcholine receptor will be discussed. These proteins have been characterized well enough that their predicted three-dimensional structures, when combined with the high resolution structure of the reaction center, may yield some insight into the validity of the theories discussed above.

4B. Methods and Methodology of Membrane Protein Modelling

Before the actual modelling of the proteins is described, the procedure followed and its implicit assumptions and limitations will be discussed. First of all, only the transmembrane region of the protein will be modelled. The reason for this is that the secondary structure of the membranous region, unlike that of the interhelical loops, is probably a regular secondary structure, *i.e.*, α-helical. Another reason is that the functionally important conformational changes which occur in these proteins are transmitted through motions in the membranous region of the protein even though they may be triggered at an extramembranous location. The examination of only the transmembrane region also simplifies to some extent the range of motions that must be considered. Because helices within the membrane probably do not move more than a few angstroms out of the bilayer, the motions of these helices are essentially within its two-dimensional plane.

The extramembranous loops of the protein will not be disregarded completely, however, because their lengths must be considered when arranging the helices, as mentioned above. In their work, Engelman *et al.* (16) assumed the maximum possible extension for each residue, about 3.6 Å. If loops of the reaction center are typical of all extramembranous loops, a better estimate might be about 2.2 Å per residue, between the 3.6 Å of an extended-chain and the 1.5 Å of an α-helix. A minimum of four residues is probably necessary for the loop unless a significant distortion of the helix termini is permitted.

Other factors considered in arranging the bilayer helices should include any structural information from experimental work, such as the locations of disulfide bonds, binding sites, and areas of chemical cross-linking. Once this structural information is taken into account, several groupings of nearby helices can be made and the number of possible structures reduced.

Once the helices have been grouped in this manner, such characteristics as their lipid-exposure can be used to arrange the helices. Because such data is often lacking, the variability analysis discussed in the last chapter was developed. The conserved residues identified by that procedure are assumed to be involved in interhelical interactions, and the variable residues are assumed to be involved in interactions with lipids. The distribution of these particular residues should, therefore, provide information on both the orientation of the helices with respect to other helices and the degree to which each helix is exposed to the membrane. It is also assumed by this procedure that helix-helix interactions are more favorable than helix-lipid interactions and that lipids do not reside between adjacent helices.

In the arrangement of these helices, the interactions of helix dipoles can be used to decide between otherwise equivalent arrangements.

Methods. The helices of each protein were grouped together on the basis of known structural information and the length of their interconnecting loops. The distribution of variable and conserved residues was then considered for each helix and the helix was placed in an

appropriate position with regard to the implied lipid-exposure. The interaction of adjacent helix dipoles was the next factor to be considered in determining the final arrangement of helices.

The three-dimensional modelling of the proteins was done using the program *Biograf*, a product of BioDesign of Pasadena, CA. Helices were built using the residue assignments made in chapter 2 and were constructed initially as regular α-helices. Energy minimization was done on each helix to allow any unfavorable interactions to be relieved. When arranging the helices, care was taken to avoid significant overlap of their van der Waals surfaces and generally resulted in interhelical distances of about 11 Å.

4C. The Acetylcholine Receptor

There are three acetylcholine receptors for which the sequences of all four subunits are known. These are the calf (17-21), mouse (22-25), and *Torpedo californica* (26, and references therein) receptors. The mouse receptor was chosen for study, although most of the conclusions reached in this work are applicable to the other receptors.

The numbering used in this section corresponds to the consensus sequence of figure 4.4.

Helix groupings. The first step in the modelling procedure was to use structural information and experimental data to group the helices. From the positions of the helices within the primary sequence, it is evident that the loops connecting helices M1 to M2 and M2 to M3 are quite short. These three helices are, therefore, probably adjacent to one another in the protein. Helix M4 was shown to be labelled by a hydophobic probe, indicating that this helix is exposed to the lipid mileu (27).

Helix M2 has been shown to form at least a part of the ion channel. Photolabelling experiments using compounds that enter the open channel have labelled residues on the α, β, and δ chains corresponding to position 267 of the alignment (28,29). The residue at position 270 of the mouse receptor was also labelled (29). These experimental results indicate that the ion channel is formed of the homologous M2 helices of the different subunits and that the residues analogous to positions 267 and 270 of the α subunit face the channel.

The above information, combined with the five-fold symmetry of the

receptor (10), implies that the channel is formed of the five M2 helices arranged pentagonally and that the M1 and M3 helices are adjacent. The M4 helices are probably in an outer position.

Helix positions and orientations. The M2 helices were arranged as a pentagon with the residues mentioned above oriented toward the channel. The distribution of variable and conserved residues along these helices showed no clear sidedness so this characteristic could not be used to determine their orientations with respect to the channel. The variability assignments made in the last chapter were then used to arrange the other helices with respect to one another by placing their variable faces toward the lipid and their conserved faces toward other helices. The best helical arrangement is shown in figure 4.1 and the variable and conserved regions are shown in figure 4.2a-c. The variable and conserved residues of M2 have been shown separately for clarity. The residues of these channel-lining helices are well-conserved in spite of the fact that many of them are exposed to the solvent, water.

The resulting receptor shape is very similar to the structure observed by electron scattering (shown in figure 1.3), and the arrangement agrees well with known structural information in that the M2 helices line the channel and the M4 helices are exposed to the lipid mileu. The residue corresponding to mouse α416 was labelled by a lipophilic reagent (27). This position is correctly predicted to face the lipid according to this model. M1, M2, and M3 are also adjacent as predicted on the basis of their connecting loops.

The successive helices of each subunit are assumed to be arranged

Figure 4.1

Helical Arrangement of the Acetylcholine Receptor

This figure depicts the most likely arrangement of the acetylcholine receptor's transmembrane helices on the basis of the variability analysis. It cannot be determined if this is the arrangement as viewed from the cytoplasm or as seen from the synaptic cleft. Evidence from the packing of α-helices in globular proteins suggests that this is the helical arrangement as viewed from the synaptic cleft (see text).

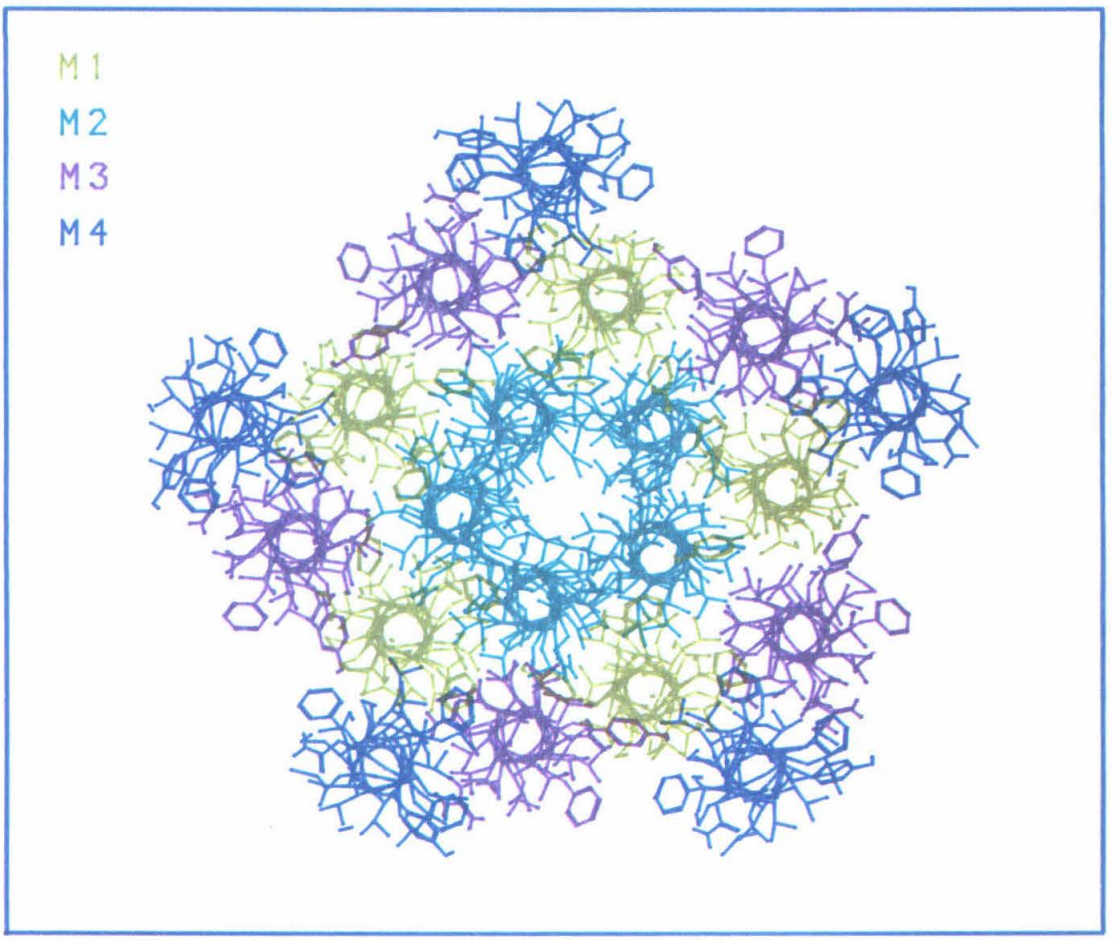

M1
M2
M3
M4

Figure 4.2

Distribution of Variable and Conserved Residues

in the Acetylcholine Receptor

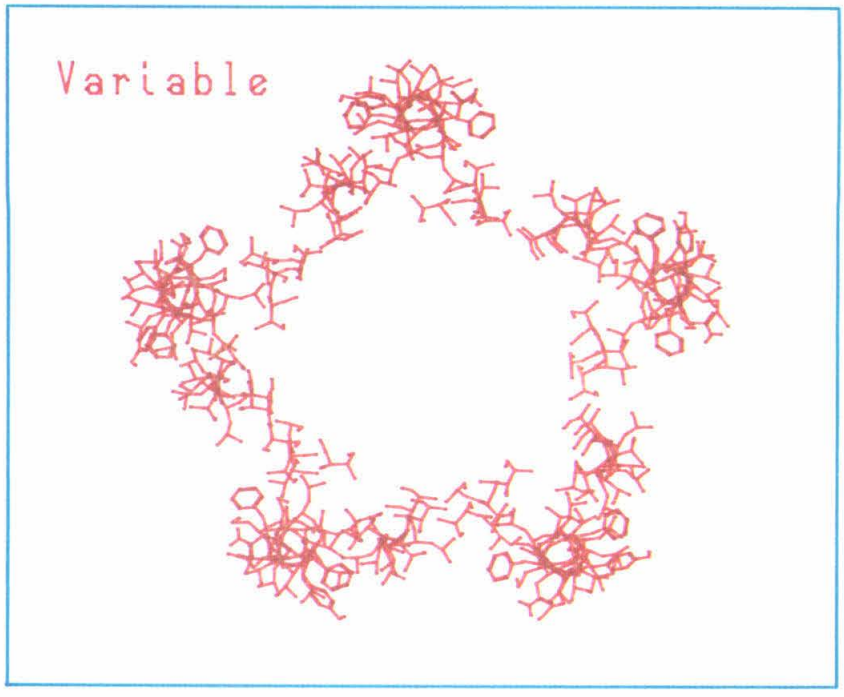
The variable and conserved residues for this receptor are shown. A few variable residues, which occur within five residues of helix termini, have been omitted so that the overall pattern of variability can be seen clearly for each helix.

4a.) The variable residues of helices M1, M3, and M4 are shown. M2 has been omitted from this representation for clarity. The variable residues line the outer boundary of the receptor.

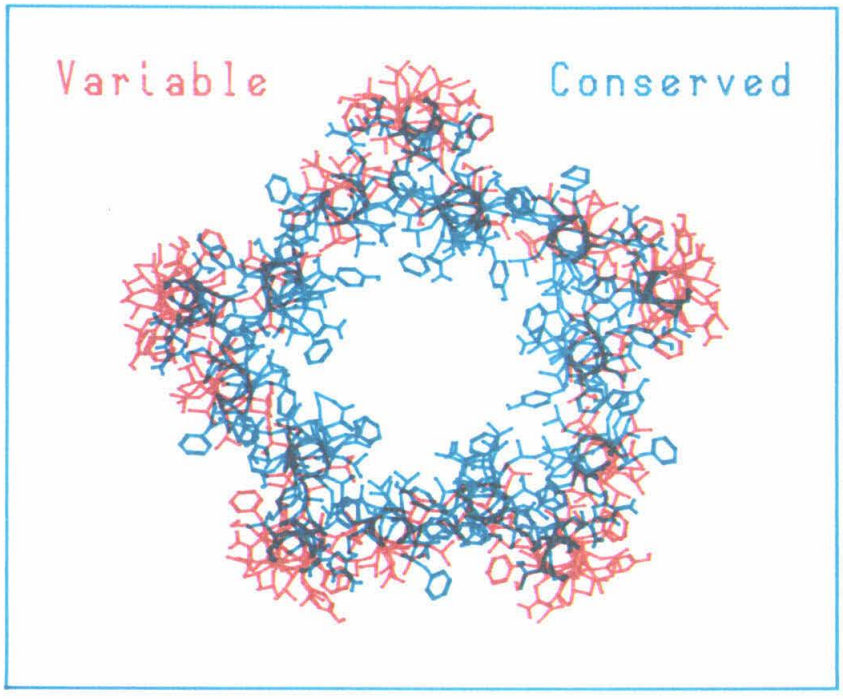
4b.) Both the variable (red) and conserved (blue) residues for the M1, M3, and M4 helices can be seen. The preference of conserved residues for the protein interior is evident.

4c.) The variability profile of M2 is shown with the channel-exposed face of the helix facing to the right and the synaptic helix terminus toward the top.

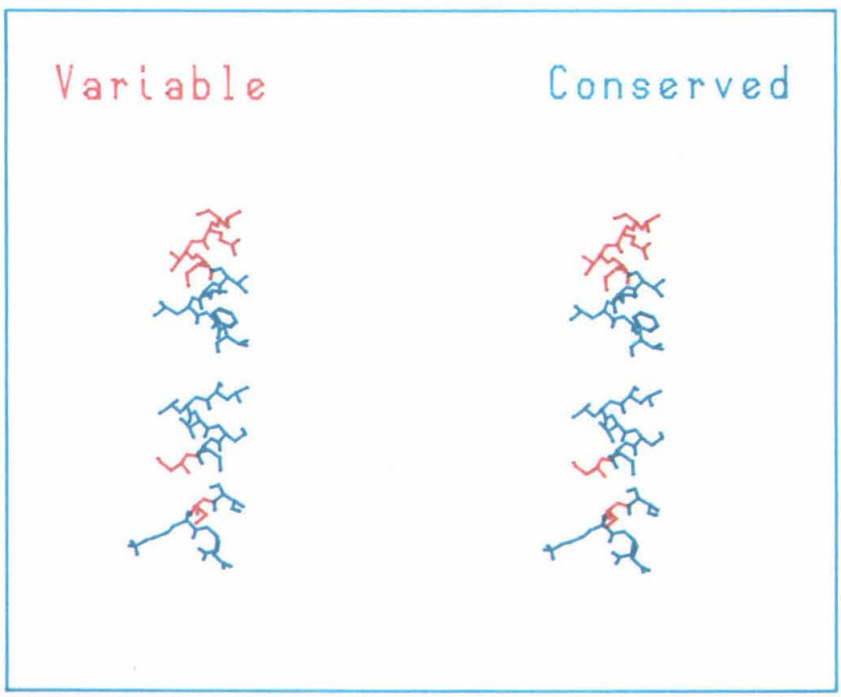
A.



B.



C.



in a "diamond" shape which forms a "point" of the star-shaped receptor, although it is possible that they trace a path somewhat like that of a check mark (✓).

It should be noted that although the helices can be oriented with respect to one another, the structure shown in figure 4.1 is only one of an enantiomeric pair. In other words, it cannot be determined whether the helical arrangement of figure 4.1 is being viewed from the cytoplasm or from the cell's exterior. The resolution of the known three-dimensional structure of the receptor (10) is not sufficient to show which form is correct. A similar asymmetry might be expected for the gap junction protein, which is structurally analogous to this receptor (30,31). Unfortunately, the resolution of the gap junction (9) is also too poor to suggest which of the two enantiomers is likely.

The examination of globular proteins with four antiparallel α-helices does enable the prediction of which of these enantiomers is more likely, however. These proteins were all found to contain 4-α-helical bundles with the same enantiomeric arrangement of successive helices (3). In terms of the acetylcholine receptor's subunits, this arrangement would correspond to a clockwise arrangement of the helices as viewed from the synaptic side of the membrane.

Subunit arrangements. Recent work in image analysis by Kubalek *et al.* (32) has shown that the subunits are arranged in the order $\alpha\beta\alpha\gamma\delta$ clockwise around the receptor as viewed from the synaptic side of the membrane. The receptor has been modelled according to this arrangement.

Ion channel. The channel is proposed to be lined by the residues corresponding to positions 259-260, 263-264, 266-267, 270-271, 274-275, 278-279, and 282. Recent results from mutagenesis studies (33), as well as the channel-labelling experiments mentioned above show that positions 263, 267, 270, and 271 are within the channel, in confirmation of the proposed helix orientation.

The nature of these channel-lining residues can be seen in figure 4.3, where the hydrophilic, hydrophobic, and charged residues are indicated. At the synaptic lip of the channel is a ring of charged residues. Although this region appears to have a large net positive charge, there are a number of negatively charged residues in the following peptide that are suitably positioned for ion-pairing should the α-helical structure continue. In fact, differences in this ion-pairing may be responsible for the shorter open-channel times of the *Torpedo* receptor when compared to the murine and bovine receptors (34,35). In figure 4.4, the aligned sequences of these three receptors can be seen. The negative charges at position 285 are proposed to interact with the positive channel residues, which are at residue number 282. The *Torpedo* receptor contains a negative charge at this position in each of the chains which contain a positive charge within the channel. This results in there being no unpaired charge at the channel mouth. The murine and bovine receptors, on the other hand, do not have a negative charge in the δ subunit to pair with the positive charge of the channel, and a net charge at the mouth of the channel results. This positive charge would repel the lysine side chains within the channel, destabilizing the closed channel relative

Figure 4.3

Nature of the Acetylcholine Receptor's

Channel-Lining Residues

Only the proposed channel-lining residues (defined in text) are shown in this figure. The side-chains have been color coded according to their nature. Hydrophobic residues: I, L, M, F, W, Y, and V. Hydrophilic residues: C, S, T, N, and Q. Positive residues: R and K. Negative residues: E and D.

Channel-Lining Residues

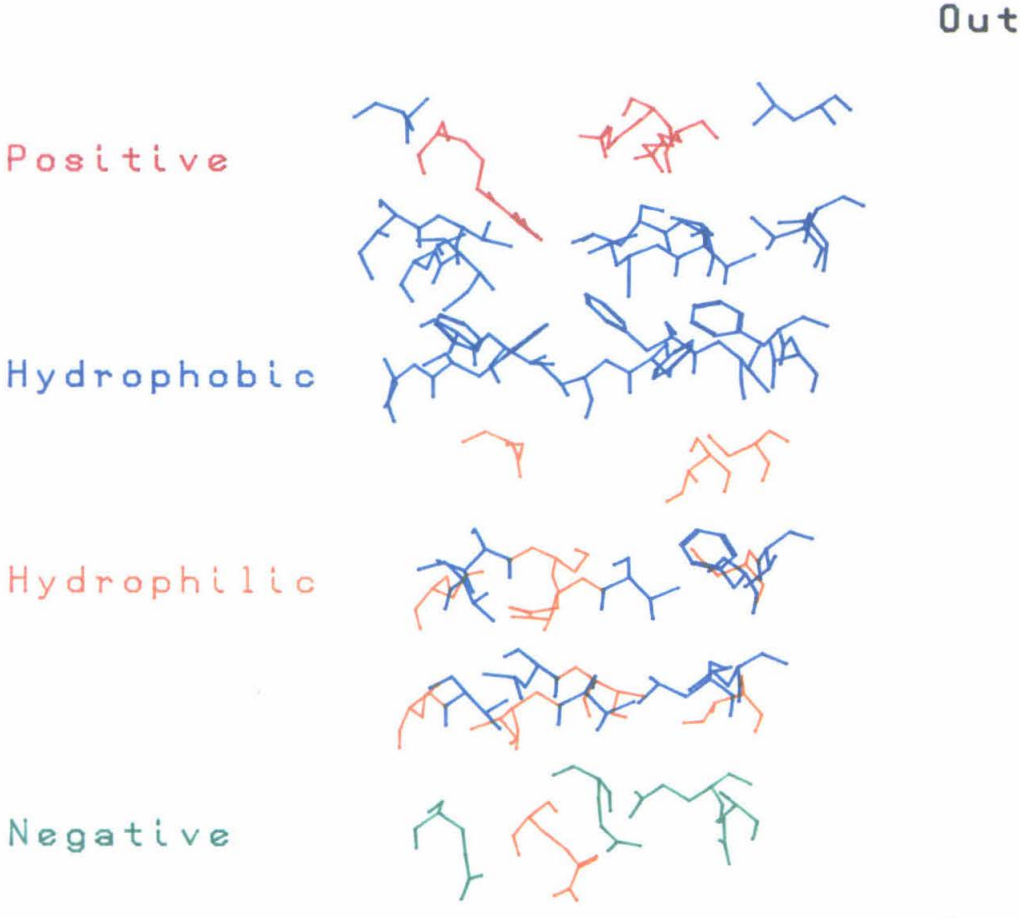


Figure 4.4

Aligned Sequences of Complete Acetylcholine Receptors

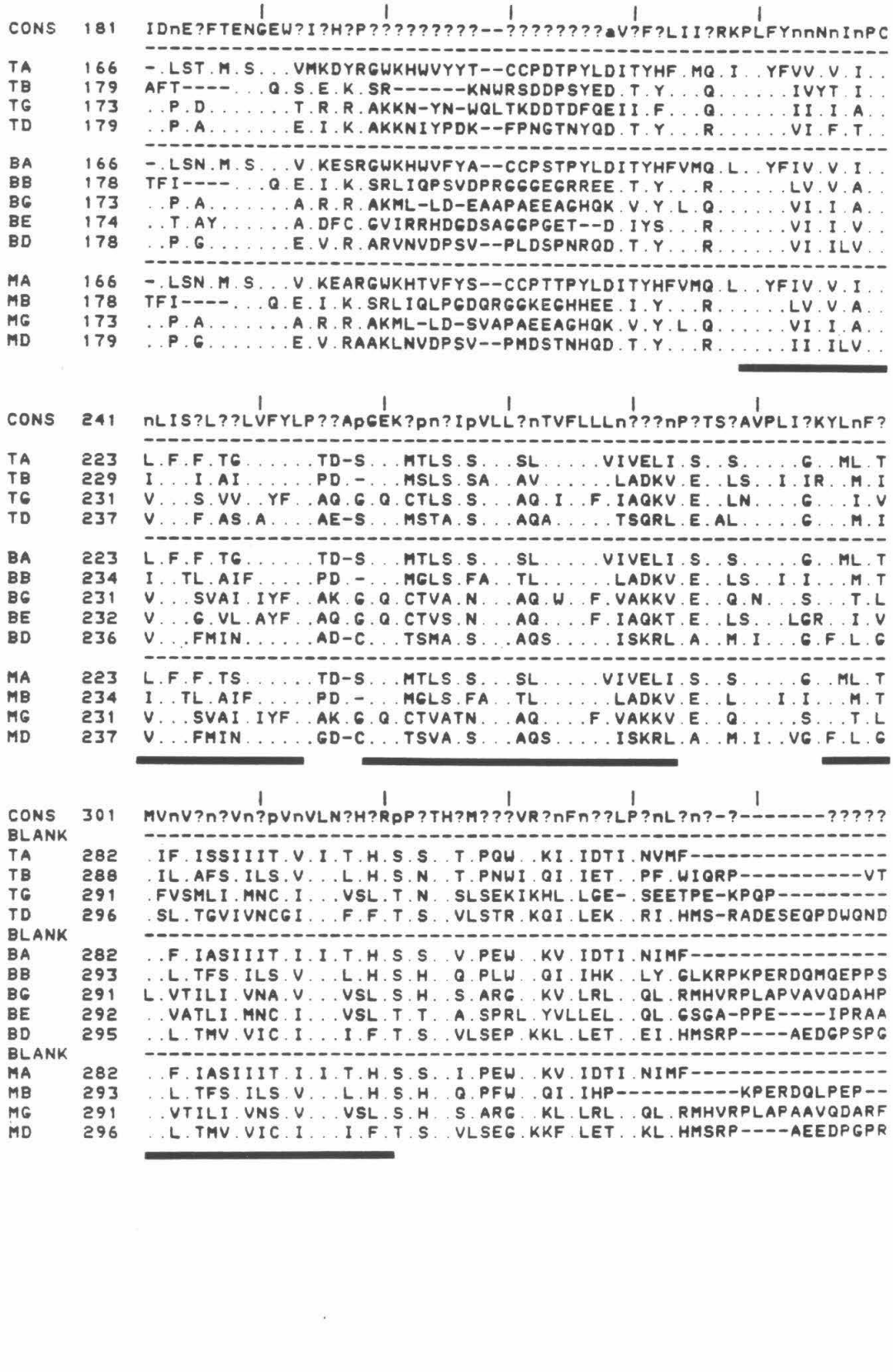
The sequences for the *Torpedo* (26, and references therein), calf (17-21) and mouse (22-25) receptors have been aligned. The sequences have been denoted as TA-D, BA-E, and MA-D, respectively. All four subunits of each recepter have been included, along with the ε subunit of the adult bovine receptor.

Complete AChR Sequences

```
                    |         |         |         |         |
CONS      1  ???E?RLn??Ln?-K?YB??nRPn??????V?V?n?LTL??LISL?E??E?nTTNV?n?n   60
             ------------------------------------------------------------
TA        1  SEH.T..VAN.LEN--.NKVI..VEHHTHF.DITVG.Q.IQ...VD.VNQIVE...RLRQ   58
TB        1  SVM.DT.LSV.FE--T.NPKV..AQTVGDK.T.RVG...TN.LI.N.KI.EM....FLNL   58
TG        1  ENE.G..IEK.LG--D.DKRII.AKTLDHIID.TLK...TN....N.KE.AL....WIEI   58
TD        1  VNE.E..IND.LIVNK.NKHV..VKHNNEV.NIALS...SN....K.TD.TL.S..WMDH   60
             ------------------------------------------------------------
BA        1  SEH.T..VAK.FED--.NSVV..VEDHRQA.E.TVG.Q.IQ..NVD.VNQIV....RLKQ   58
BB        1  SEA.G..REK.FS--G.DSTV..AREVGDR.W.SIG...AQ....N.KD.EMS.K.YLDL   58
BG        1  RNQ.E..LGD.MQ--G.NPHL..AEHDSDV.N.SLK...TN....N.RE.AL....WIEM   58
BE        1  KNE.L..YHY.FD--T.DPGR..VQEPEDT.TISLK...TN....N.KE.TL..S.WIGI   58
BD        1  LNE.E..IRH.FEE.A.NKEL..AAHKE-S.EISLA...SN....K.VE.TL....WIEQ   59
             ------------------------------------------------------------
MA        1  SEH.T..VAK.FED--.SSVV..VEDHREI.Q.TVG.Q.IQ..NVD.VNQIV....RLKQ   58
MB        1  SEA.GQ.IKK.FS--N.DSSV..AREVGDR.G.SIG...AQ....N.KD.EMS.K.YLDL   58
MG        1  RNQ.E..LAD.MR--N.DPHL..AERDSDV.N.SLK...TN....N.RE.AL....WIEM   58
MD        1  LNE.Q..IQH.FNE.G.DKDL..VARKEDK.D.ALS...SN....K.VE.TL....WIDH   60

                    |         |         |         |         |
CONS     61  ?W?DYRL?WBn?a??Gn??LRnP???VW?PDnVL?NN?DG?F?n????NVLV???G?n?W  120
             ------------------------------------------------------------
TA       59  Q.I.V..R.NPADYG.IKKI.L.SDD..L..L..Y..A..D.AIVHMTKL.LDYT.KIM.  118
TB       59  A.T....Q.DPAAYE.IKD..I.SSD..Q..I..M..N..S.EITLHV....QHT.AVS.  118
TG       59  Q.N....S.NTSEYE.IDLV.I.SELL.L..V..E..V..Q.EVAYYA....YND.SMY.  118
TD       61  A.Y.H..T.NASEYSDISI..L.PEL..I..I..Q..N..QYHVAYFC....RPN.YVT.  120
             ------------------------------------------------------------
BA       59  Q.V..N.K.NPDDYG.VKKIHI.SEKI.R..L..Y..A..D.AIVKFTK..LDYT.HIT.  118
BB       59  E.T....S.DPEEHE.IDS..ISAES..L..V..L..N..N.DVALDI..V.SSD.SMR.  118
BG       59  Q.C....R.DPRDYG.LWV..V.STM..R..I..E..V..V.EVALYC....SPD.CVY.  118
BE       59  D.Q....NYSKGDFG.VET..V.SEL..L.EI..E..I..Q.GVAYEA....SEG.YLS.  118
BD       60  G.T.S..Q.DAEDFGNISV..L.ADM..L.EI..E..N..S.QISYSC...IYPS.SVY.  119
             ------------------------------------------------------------
MA       59  Q.V..N.K.NPDDYG.VKKIHI.SEKI.R..V..Y..A..D.AIVKFTK..LDYT.HIT.  118
MB       59  E.T....S.DPAEHD.IDS..ITAES..L..V..L..N..N.DVALDI..V.SFE.SVR.  118
MG       59  Q.C....R.DPKDYE.LWI..V.STM..R..I..E..V..V.EVALYC....SPD.CIY.  118
MD       61  A.V.S..Q.DANDFGNITV..L.PDM..L.EI..E..N..S.QISYAC....YDS.YVT.  120

                    |         |         |         |         |
CONS    121  ?PPAI?RSpC?I?VTYFPFDWQNCSn?F?S?pY???En?n??????E?????----????  180
             ------------------------------------------------------------
TA      119  T....FK.Y.E.I..H....Q...TMKLGIWT.DGTKVSISPES------------DRP-  165
TB      119  Q.S..Y.TS.T.K.M.......TMV.K.YT.DTS.VTLQHALDAKGEREVKEIVINKD  178
TG      119  L....Y..T.P.A.........LV.R.QT.NAH.VNLQ--LSA.EGEAV----EWIH  172
TD      121  L....F..S.P.N.L.......LK.TALN.DAN.ITMDLMT--DTIDGKDYPIEWII  178
             ------------------------------------------------------------
BA      119  T....FK.Y.E.I..H....E...MKLGTWT.DGSVVVINPES------------DQP-  165
BB      119  Q..G.Y..S.S.Q.........TMV.S.YS.DSS.VSLQTGLSP.GQERQ-EVYIHEG  177
BG      119  L....F..S.PVS..F......LI.Q.QT.STN.INLQ--LSQ.DGQTI----EWIF  172
BE      119  L....Y..T.AVE.........LV.R.QT.NAE.VEFV-FAVDDEGKTIS----KID  173
BD      120  L....F..SRP.S.........LK.S.LK.TTK.ITLSLKQAE.DG--RSYPVEWII  177
             ------------------------------------------------------------
MA      119  T....FK.Y.E.I..H....E...MKLGTWT.DGSVVAINPES------------DQP-  165
MB      119  Q..GLY..S.S.Q.........TMV.S.YS.DSS.VSLKTGLDP.GEERQ-EVYIHEG  177
MG      119  L....F..S.S.S.........LI.Q.QT.STS.INLQ--LSQ.DGQAI----EWIF  172
MD      121  L....F..S.P.S.........LK.S.LK.TAK.ITLSLKQEE.NN--RSYPIEWII  178
```


CONS 181 IDnE?FTENGEW?I?H?P?????????--????????aV?F?LII?RKPLFYnnNnInPC 240
TA 166 -.LST.M.S...VMKDYRGWKHWVYYT--CCPDTPYLDITYHF.MQ.I..YFVV.V.I.. 222
TB 179 AFT----...Q.S.E.K.SR------KNWRSDDPSYED.T.Y...Q.....IVYT.I.. 228
TG 173 ..P.D.....T.R.R.AKKN-YN-WQLTKDDTDFQEII.F...Q.....II.I.A.. 230
TD 179 ..P.A.....E.I.K.AKKNIYPDK--FPNGTNYQD.T.Y...R.....VI.F.T.. 236
BA 166 -.LSN.M.S...V.KESRGWKHWVFYA--CCPSTPYLDITYHFVMQ.L..YFIV.V.I.. 222
BB 178 TFI----...Q.E.I.K.SRLIQPSVDPRGGGEGRREE.T.Y...R.....LV.V.A.. 233
BG 173 ..P.A.....A.R.R.AKML-LD-EAAPAEEAGHQK.V.Y.L.Q.....VI.I.A.. 230
BE 174 ..T.AY.....A.DFC.GVIRRHDGDSAGGPGET--D.IYS...R.....VI.I.V.. 231
BD 178 ..P.G.....E.V.R.ARVNVDPSV--PLDSPNRQD.T.Y...R.....VI.ILV.. 235
MA 166 -.LSN.M.S...V.KEARGWKHTVFYS--CCPTTPYLDITYHFVMQ.L..YFIV.V.I.. 222
MB 178 TFI----...Q.E.I.K.SRLIQLPGDQRGGKEGHHEE.I.Y...R.....LV.V.A.. 233
MG 173 ..P.A.....A.R.R.AKML-LD-SVAPAEEAGHQK.V.Y.L.Q.....VI.I.A.. 230
MD 179 ..P.G.....E.V.RAAKLNVDPSV--PMDSTNHQD.T.Y...R.....II.ILV.. 236
CONS 241 nLIS?L??LVFYLP??ApGEK?pn?IpVLL?nTVFLLLn???nP?TS?AVPLI?KYLnF? 300
TA 223 L.F.F.TG.....TD-S...MTLS.S...SL.....VIVELI.S..S.....G..ML.T 281
TB 229 I...I.AI.....PD.-...MSLS.SA..AV.....LADKV.E..LS..I.IR..M.I 287
TG 231 V...S.VV..YF..AQ.G.Q.CTLS.S...AQ.I..F.IAQKV.E..LN....G...I.V 290
TD 237 V...F.AS.A....AE-S...MSTA.S...AQA.....TSQRL.E.AL.....G...M.I 295
BA 223 L.F.F.TG.....TD-S...MTLS.S...SL.....VIVELI.S..S.....G..ML.T 281
BB 234 I..TL.AIF.....PD.-...MGLS.FA..TL.....LADKV.E..LS..I.I...M.T 292
BG 231 V...SVAI.IYF..AK.G.Q.CTVA.N...AQ.W..F.VAKKV.E..Q.N...S...T.L 290
BE 232 V...G.VL.AYF..AQ.G.Q.CTVS.N...AQ....F.IAQKT.E..LS...LGR..I.V 291
BD 236 V...FMIN.....AD-C...TSMA.S...AQS.....ISKRL.A..M.I...G.F.L.G 294
MA 223 L.F.F.TS.....TD-S...MTLS.S...SL.....VIVELI.S..S.....G..ML.T 281
MB 234 I..TL.AIF.....PD.-...MGLS.FA..TL.....LADKV.E..L...I.I...M.T 292
MG 231 V...SVAI.IYF..AK.G.Q.CTVATN...AQ....F.VAKKV.E..Q.....S...T.L 290
MD 237 V...FMIN.....GD-C...TSVA.S...AQS.....ISKRL.A..M.I..VG.F.L.G 295
CONS 301 MVnV?n?Vn?pVnVLN?H?RpP?TH?M???VR?nFn??LP?nL?n?-?-------????? 360
BLANK
TA 282 .IF.ISSIIIT.V.I.T.H.S.S..T.PQW..KI.IDTI.NVMF---------- 325
TB 288 .IL.AFS.ILS.V...L.H.S.N..T.PNWI.QI.IET..PF.WIQRP----------VT 337
TG 291 .FVSMLI.MNC.I...VSL.T.N..SLSEKIKHL.LGE-.SEETPE-KPQP--------- 339
TD 296 .SL.TGVIVNCGI...F.F.T.S..VLSTR.KQI.LEK..RI.HMS-RADESEQPDWQND 354
BLANK
BA 282 ..F.IASIIIT.I.I.T.H.S.S..V.PEW..KV.IDTI.NIMF---------- 325
BB 293 ..L.TFS.ILS.V...L.H.S.H..Q.PLW..QI.IHK..LY.GLKRPKPERDQMQEPPS 352
BG 291 L.VTILI.VNA.V...VSL.S.H..S.ARG..KV.LRL..QL.RMHVRPLAPVAVQDAHP 350
BE 292 ..VATLI.MNC.I...VSL.T.T..A.SPRL.YVLLEL..QL.GSGA-PPE----IPRAA 346
BD 295 ..L.TMV.VIC.I...I.F.T.S..VLSEP.KKL.LET..EI.HMSRP----AEDGPSPG 350
BLANK
MA 282 ..F.IASIIIT.I.I.T.H.S.S..I.PEW..KV.IDTI.NIMF---------- 325
MB 293 ..L.TFS.ILS.V...L.H.S.H..Q.PFW..QI.IHP----------KPERDQLPEP-- 340
MG 291 ..VTILI.VNS.V...VSL.S.H..S.ARG..KL.LRL..QL.RMHVRPLAPAAVQDARF 350
MD 296 ..L.TMV.VIC.I...I.F.T.S..VLSEG.KKF.LET..KL.HMSRP----AEEDPGPR 351

```
                 |         |         |         |         |
CONS  361  ?????S??????????E?n??KPRSaLnFPE-----E?n??????????n?????-?????  420
           ------------------------------------------------------------
TA    326  ----F.TMKRASKEKQ.N-------------------KIFADDIDISDISGKQV-TGEVI  361
TB    338  TPSPD.KPTIISRAND.YFIR..AGDFVC.---------VDNIRVAVQPERLFVEMKWHL  388
TG    340  --RRR.SFGIMIKA-E.YILK....E.M.E.QKDRHGFEEQKDRHGLKRVNKMTSDIDIG  396
TD    355  LKLRR.SSVGYISKAQ.YFNI.S..E.M.-------.KQSERHGLVPRVTPRIG-FRVNN  406
           ------------------------------------------------------------
BA    326  -----.TMKRPSREKQDK-------------------KIFTEDIDISDISGK--------  353
BB    353  IAPRD.PASGWGRGTD.YFIR..PNDFL..-------------KPNRFQPELSADLRRFI  399
BG    351  RLQNG.SSGWPITAGE.NALCL...E.L.--------RQRQRNGLVRAALEKLEKGPESG  402
BE    347  SPPRRASSLGLLLRAE.LILK....E.V.-------.QQRHRHGTWTATL-CQN-LGAAA  397
BD    351  TLIRR.SSLGYISKAE.YFSL.S..D.M.-------.KQSERHGLARRLT--TA-RRPAA  400
           ------------------------------------------------------------
MA    326  ----F.TMKRPSRDKQ.K-------------------RIFTEDIDISDISGK--------  354
MB    341  -HHSL.PRSGWGRGTD.YFIR..P.DFL.--IRKPPSDFLFPKLNRFQPESSAHSPLIKH  397
MG    351  RLQNG.SSGWPIMARE.GDLCL...E.L.--------RQRQRNGLVQAVLEKLENG-EVR  401
MD    352  ALIRR.SSLGYICKAE.YFSL.S..D.M.-------.KQSERHGLARRLT--TA-RRP-A  400


                 |         |         |         |         |
CONS  421  ????n????KQA???nK??na????IA??????????????W??V??VnDRn?n?nn??  480
           ------------------------------------------------------------
TA    362  FQTPLIKN-----PDV.SAIEGVKY..EHMKSDEESSNDAEE.KY.AM.I.HILLCVFML  416
TB    389  NGLTPLIKN---PQDL.EAVEAIKY..EQLESASEFDDAKKD.QY.AM.A..LFLYVFFV  445
TG    397  TTVDLYKDLANFAPEI.SCVEACNF..KSTKEQNDSGSENEN.VLIGK.I.KACFWIALL  456
TD    407  ENIAASDQL---HDEI.SGIDSTNY.VKQIKEKNAYDEEVGN.NL.GQTI..LSMFIITP  463
           ------------------------------------------------------------
BA    354  PPPPMGFH-----SEV.SAIEGIKY..ETMKSDQESNNAAEE.KY.AM.M.HILLAVFML  408
BB    400  DGPNFHSPL---IKHPREVVSSISY..RQLQEQEDHDVLKED.QF.AM.V..LFLWTFII  456
BG    403  QSPEWCELL...APAIQACVEACNL..RARHQQTHFDSGNKE.FL.GR.L..VCFLAMLS  462
BE    398  P-EIRCCVD---------AVDAVNFV.SSTRDQEATGEEVKE.FL.GR.L..VCFLAMLS  447
BD    401  GSEQAQQVD---ASEL.PAVDGANF.VNHMKDQNNYNEEKDC.NR.ARTV..LCLFVVTP  457
           ------------------------------------------------------------
MA    355  PPPPMGFH-----S-V.SAIEGVKY..ETMKSDQESNNAAEE.KY.AM.M.HILLGVFML  408
MB    398  REPTRAVGL---PQELREVISSISYM.RQLQEQEDHDALKED.QF.AM.V..LFLWTFIV  454
MG    402  QSQEFCGDG...SPAIQACVDACNLM.RARRQQSHFDSGNEEQLL.GR.L..VCFLAMLS  461
MD    401  SSEQVQQCG---SLEM.PAVDGANF.VNHMRDQNSYNEEKDNQNQ.ARTV..LCLFVVTP  457
                                                          ▬▬▬▬▬▬▬▬▬▬▬▬

                 |         |         |         |         |
CONS  481  n???GTn?IFL???????P??PFPGDP??Y??????FI                        518
           ------------------------------------
TA    417  ICII..VCV.AGRLIELSQEG                                        437
TB    446  ICSI..FS...DASHNVP.DN..AP                                    470
TG    457  LFSI..LA...TGHFNQV.EF.....RK.VPD                            489
TD    464  VMVL..IF..VMGNFNHP.AK..E...FD.SSDHPRCA                      501
           ------------------------------------
BA    409  VCII..LAV.AGRLIELNQQGG                                       431
BB    457  FTSV..LV...DATYHLP.AD...S                                    481
BG    463  LFVC..AG...MAHYNRV.AL.....RS.LPSSDE                         498
BE    448  LFVC..SL...GAYFNRV.QL.Y                                      471
BD    458  IMVV..AW...QGAYNQP.PQ.....FS.LEKDKR..                        495
           ------------------------------------
MA    409  VCLI..LAV.AGRLIELHQQG                                        429
MB    455  FTSV..LV...DATYHLP.PE                                        479
MG    462  LFIC..AG...MAHYNQV.DL.....RP.LPLPDE                          497
MD    458  VMVV..AW...QGVYNQP.LQ.....FS.SEQDKR..                        495
           ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬
```

to the open channel. The bovine and murine receptors would, therefore, stay open longer than the *Torpedo* receptor. The bovine receptor has been observed to have a longer open-channel duration than the *Torpedo* receptor (35).

The second region encountered by a cation on its travel through the channel is a hydrophobic region, which is formed by residues on two successive turns of the helix and is about 7.5 A long. These hydrophobic residues are relatively well-conserved in all the known subunits, indicating that they may be important to receptor activity. This hydrophobic ring seems a rather unusual feature for an ion channel and will be discussed further below.

The third region is made up of the next three successive helical turns and contains a mixture of hydrophilic and hydrophobic residues. It is in this region that the open-channel blockers can label the receptor.

The helices have been positioned in the model such that at the channel's narrowest point, it is about 6.5 A wide. This width corresponds to the channel diameter determined by permeability experiments (37). The channel has been built with an angle of about 18^{o} between helices. Such positioning allows a better packing of the side chains. The helices have also been angled with respect to the bilayer normal to accomodate large channel-blockers like chlorpromazine and triphenylmethylphosphonium, which enter the channel from the synaptic side of the membrane. The low resolution structure also indicates that the cytoplasmic end of the channel is narrower than the synaptic end, as modelled here.

Mechanism of channel gating. When acetylcholine binds to the receptor, it interacts with residues of the α subunit, which are near the N-terminus of M1 (38-40). Although great progress has been made in the characterization of this binding site, it is still uncertain which parts of the transmembrane region are pressured to change conformation in response to acetylcholine binding. The close-packed structure proposed for this protein suggests that all of the helices might move in response to a localized pressure. One study of this receptor does suggest that the conformational changes which occur upon ligand binding are of a global nature (41). The position of a localized force created by acetylcholine binding is likely be at αM1 because it is the membrane helix closest to the ligand binding site, although an indirect effect on M2 through the M2-M3 loop is possible.

The thought of the αM1 helices triggering a global conformational change is interesting in light of the fact that they are in direct contact with the M2 helix of every subunit except one. In the recently reported structure of the desensitized acetylcholine receptor (31), one subunit, the γ subunit, was shown to deviate from the observed symmetry of the normal receptor and to project further into the bilayer (and away from the channel). It could be speculated that some sort of "decoupling" of this subunit occurs upon receptor desensitization, preventing the usual channel gating in response to acetylcholine. If it is not directly in contact with the αM1 helices, the motions of the γ subunit might indeed be more independent than those of the other subunits. According to the present arrangement, the subunit that is not directly in contact with the αM1 helices is

the δ subunit, however. If the helices of each subunit were arranged sequentially counterclockwise as viewed from the synaptic side of the membrane, the γ subunit would not be adjacent to the αM1 helices. This arrangement seems rather improbable in light of the many proteins that demonstrate the opposite configuration, including the reaction center (3,7,8,15).

With the assumptions that a force is acting to move the transmembrane helices and that the movements of the helices are concerted, there are only a limited number of motions which could cause the channel to open. The most obvious of these is a radial rotation of the helices away from the channel, physically widening the channel. This type of motion, if limited, would keep most of the side chain packing between helices intact and would not expose new regions of the protein to the membrane. Two types of motion can be envisioned, one in which the synaptic ends of the channel helices are pulled slightly toward the channel, forcing the cytoplasmic ends outward, and one in which the synaptic ends are forced outward, resulting, perhaps, in the narrowing of the cytoplasmic portion of the channel, but the widening of the synaptic side. A schematic diagram of these complementary types of motion is shown in figure 4.5.

The type of motion which is more probable, however, is a motion that is tangential to the ring of the channel and perpendicular to that just described. This type of transition might be viewed as a clockwise or counterclockwise motion of the synaptic helix termini with fixed cytoplasmic termini. This motion would keep the side chain interactions between helices intact. Within the channel, the side

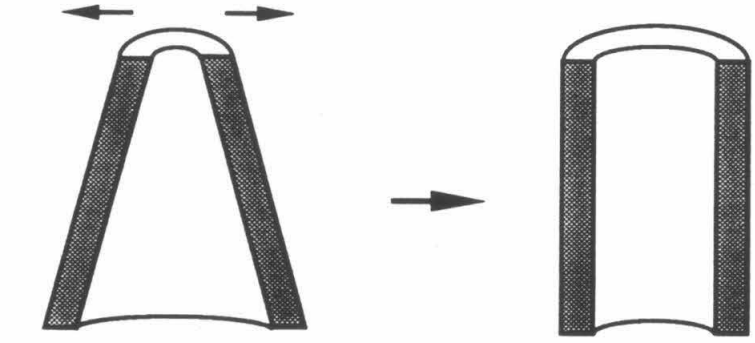
Figure 4.5

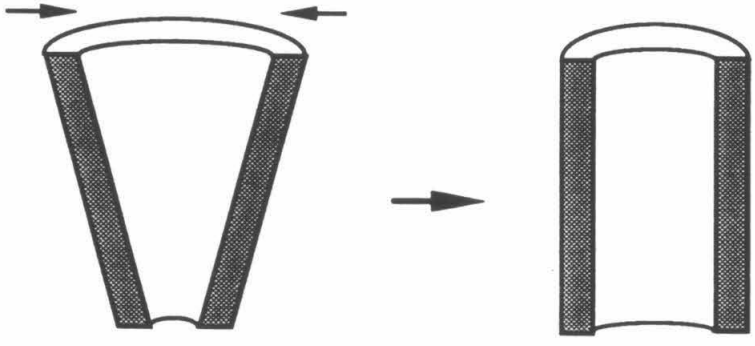
Radial Motions of Channel Helices

These schematic diagrams depict possible helix movements which could result in channel-opening. The smaller arrows indicate the direction of the applied force.

Radial Helix Motions

Closed Channel **Open Channel**





chains of the residues would effectively rotate with respect to the channel center. Side chains, which had been oriented directly toward the channel could be rotated toward the side and *vice versa.* In addition, a small component of the radial motion described in the last paragraph could be included, resulting in an additional widening of the channel.

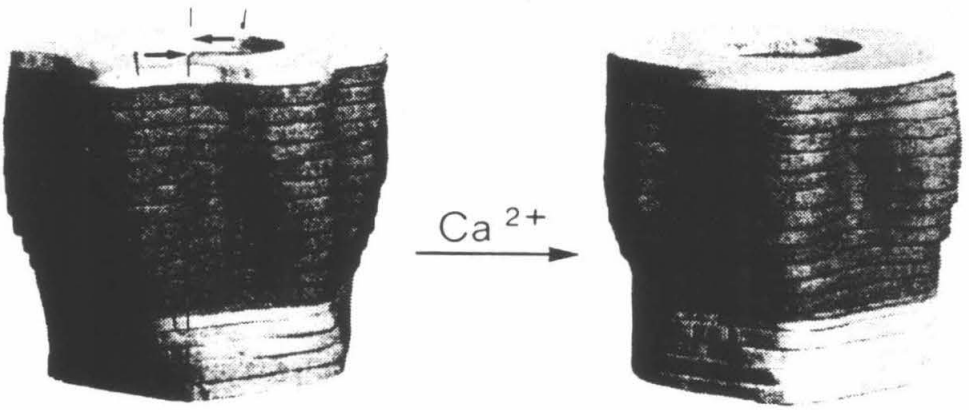
A tangential rotation of subunits has already been observed in the opening and closing of the gap junction protein (42). This protein, like the acetylcholine receptor, is made up of homologous subunits but the gap junction contains six subunits rather than five. Subunits of both proteins are arranged symmetrically around the channel and in both cases they have been proposed to span the membrane four times (30,43). Unlike the acetylcholine receptor, however, ligand binding induces the channel of the gap junction to close, rather than open. The structures of the protein in both the open and closed states were determined by electron scattering and found to differ from one another by a 7^{o} tilt in a direction toward the bilayer normal (42). These two structures can be seen in figure 4.6. The subunits of the closed channel are more perpendicular to the plane of the bilayer than those of the open channel. It should be noted that some controversy has developed about whether Ca^{+2}, the ligand used to induce the closed-channel state of this protein, can induce channel-closing unaided, or whether calmodulin must be present (42). Although the "closed-channel" structure displayed in figure 4.6 was prepared in the absence of calmodulin (unless it was present as a contaminant), the observed conformational change demonstrated that such rotations are not

Figure 4.6

The Open and Closed States of the Gap Junction

These balsa wood models of the open and closed gap junction illustrate the tangential motion of subunits upon channel closure. Although slightly modified, this figure is from reference 42.

Gap Junction



open channel **closed channel**

forbidden by high energy requirements.

The known structure of the acetylcholine receptor suggests that its subunits are oriented similarly to those of the closed gap junction, where they appear nearly parallel to the bilayer normal (31). The structure of the desensitized receptor, in which the channel is probably closed, also has an orientation similar to the closed gap junction (31). The conductive (open channel) acetylcholine receptor cannot be imaged until a method is found to lock the ion channel open.

There are two regions of the acetylcholine channel, which could conceivably be involved in the gating of the channel. One of these regions is located on the cytoplasmic side of the residues labelled by the open-channel blockers and would include the residues near position 267. Because the blockers only penetrate this far into the channel it is probable that the channel narrows to its smallest diameter in this region. The gating in this case would effect a widening of the channel at this position. It is surprising to find that these residues are not very well conserved.

The other possible region for gating to occur is more intriguing. It was noted above that a surprising hydrophobic region occurred near the synaptic mouth of the channel. Even more unusual is the observation that a phenylalanine residue occurs within this region at position 275 of the alignment in figure 4.4. This phenylalanine is conserved in every sequence of every subunit currently known and faces directly into the channel when the helices are oriented nearly perpendicular to the plane of the bilayer, the orientation suggested by the receptor's low-resolution structure. A tangential rotation of the

helices by about 8^{o} with respect to the bilayer normal causes these phenylalanine residues to be directed toward the sides of the channel, allowing the channel to pass ions at least as large as 7.5 Å. The M2 helices are not required to be parallel to the bilayer normal for the channel to be in its closed state. About six degrees of helix tilt is allowable.

The presence of such a ring of phenylalanine residues within the channel may explain the substates which are observed over the course of a single channel-opening event (49). The sporadic rotation of phenylalanine side chains away from the sides and into the center of the open channel may temporarily impede the flow of ions and result in the observed *pseudo*closed-channel substate. These substates occur on the timescale of about 50 μs. The duration of an open channel is between 2 and 10 milliseconds (35,36). The nature of nearby channel-lining helices may be critical for holding the phenyl rings toward the channel walls and may account for the conserved, hydrophobic residues that are located in this region of the channel.

In figure 4.7, the channel is shown modelled with the narrowest point at the position discussed above. The proposed conformations of the open and closed channel are both shown. The effects of helix tilting on the conserved ring of phenylalanine residues are shown in figure 4.8. The channel has been modelled as a left-handed supercoil, and it is postulated that the other helices of each subunit will be aligned roughly parallel to their respective M2 helices. This type of three-dimensional arrangement would allow the favorable packing of the M1 and M3 helices of neighboring subunits, but would not permit super-

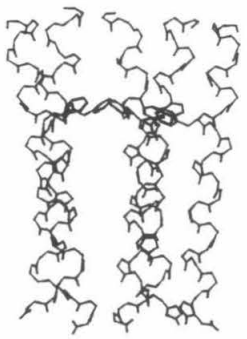
Figure 4.7

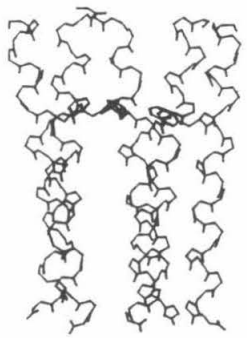
Open and Closed Conformations of the Cation Channel

The relative positioning of helices with respect to one another is depicted for both the open and closed channel conformations. Only the backbone traces of the helices have been displayed. The synaptic side of the membrane corresponds to the top of each stereo drawing and the bilayer normal is parallel to the receptor's axis of rotational symmetry.

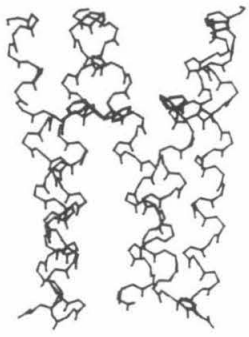
The position of the ring of phenylalanine residues has also been indicated in this diagram. Their side chains are evident near the synaptic mouth of the channel.

Closed





Open



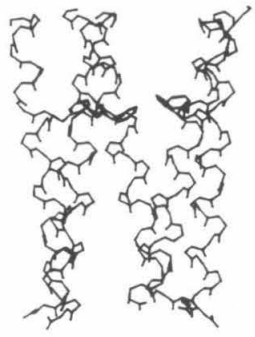
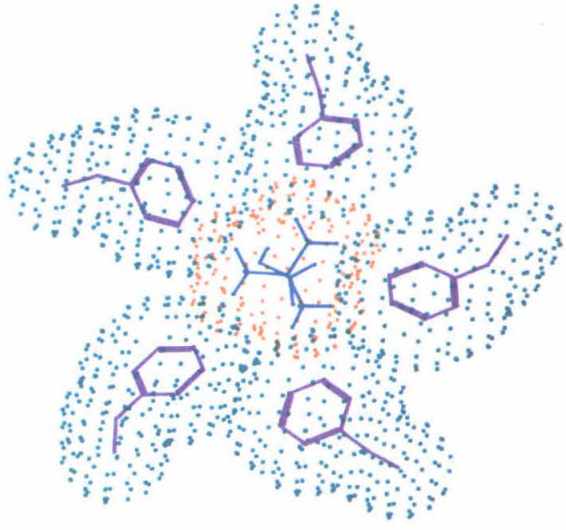


Figure 4.8

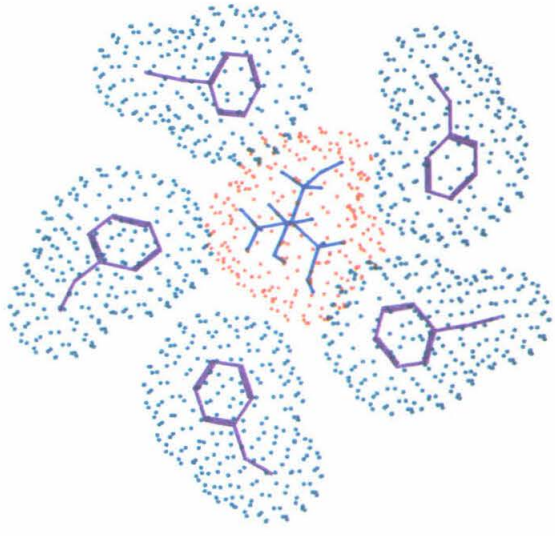
Phenylalanine Orientations in the Open and Closed Channel

The ring of phenylalanine residues corresponding to position 275 of the alignment is shown as a slice taken perpendicular to the bilayer normal. Orientations proposed for both the open and closed channels are shown. Phenylalanine side chains are shown in purple and their van der Waals surfaces have been indicated by light blue dots. A molecule of protonated tris(hydroxymethyl)aminomethane (dark blue) has been positioned within the channel and its van der Waals surface shown with orange dots. This cation is one of the larger ions transported by the channel and is shown to approximate the size of a hydrated sodium ion.

Closed



Open



coiling of helices within individual subunits. If supercoiling within the helices of each subunit is dominant, the channel helices would have a right-handed, coiled structure. It is impossible to tell at the present time which, if either, of these two types of helix packing is likely for the receptor. The M4 helices would have the same general orientation in either case, matching the tilt of the open gap junction, and the rotational mechanisms of channel gating discussed above would be equally valid.

Implications for other membrane receptors and channels. An examination of the proposed helical arrangement indicates that the helices of the acetylcholine receptor are positioned such that their interacting dipole moments serve to stabilize the structure. The requirements for this stabilization may explain to some extent the large number of transmembrane helices proposed for channel proteins such as the gap junction (30), the sodium and calcium channels (45,46), and, of course, the acetylcholine receptor. Sheridan *et al.* (14) have estimated that two adjacent, antiparallel α-helices stabilize each other by about 2.5 kcals/mole and that two adjacent, parallel helices destabilize one another by the same amount. The energy of interaction for non-adjacent helices in a 4-helix bundle were estimated to be 1.5 kcal/mole. Because the distance between non-adjacent helices of the channel is comparable to this separation, this value was also used to estimate the dipolar stabilization of the channel's transmembrane helices. Calculations used to approximate the stabilization afforded to the receptor for various combinations of helices are shown in figure 4.9. For the purposes of illustration, the net stabilization

(positive energy) or destabilization (negative energy) has been calculated for each helix type, each subunit, and for the receptor as a whole. Because two helices are involved in each interaction, values of 1.25 and .75 kcal/mole were used in estimating the energy contributions to each helix for the short and long-range interactions described above.

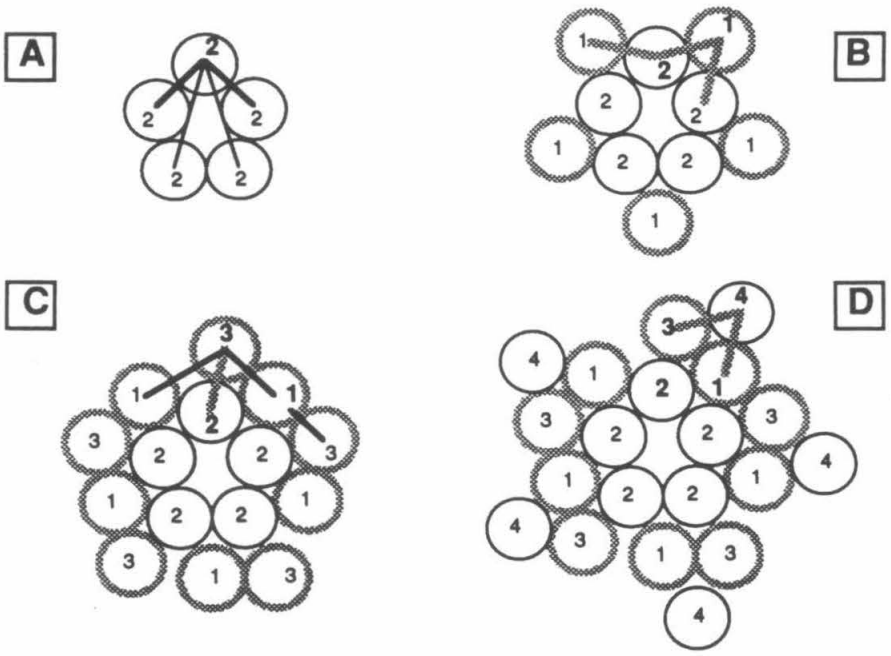
Viewed as an isolated group, the M2 helices which make up the channel would be destabilized by about 20 kcal/mole compared to the energy of five non-interacting helices. With the addition of the M1 helices, the net stabilization of the resulting receptor would be only 5 kcal/mole. The M2 helices are still somewhat destabilized with only two helices per subunit. If M3 is added to the structure, each M2 helix has a net destabilization of only .25 kcal/mole, in spite of the fact that they are parallel to one another. The structure as a whole, however, is destabilized by about 7.5 kcal/mole, due to the unfavorable interactions between helices M1 and M3. When the M4 helix is added, the net stabilization of the receptor increases to about 17.5 kcal/mole and the M2 helices are stabilized to the same extent as in the 3-helix case. On the basis of these estimations, it is postulated that the three non-channel helices of each subunit are necessary for the formation of a stable receptor. It should be noted that the need for this additional stabilization is a result of the parallel arrangement of the channel-lining helices. Dipolar stabilization from additional helices in proteins such as the sodium channel, which has been proposed to have 28 helices surrounding the four parallel ones that line the channel (36), may be necessary to offset other destabilizing

Figure 4.9

Helix Dipole Interactions in the Acetylcholine Receptor

The stabilization and destabilization energies for the helix combinations shown in A through D have been calculated and are summarized in the table. Each successive case builds upon the previous one. Stabilizing interactions have been denoted by gray bars and destabilizing ones by black bars. The longer-ranged interactions have been denoted by thinner lines. The borders of the circles, which denote helices, have also been shaded to indicate their relative orientations.

Helix Dipole Interactions



	HELIX	INTERACTION ENERGY (kcal/mol)			/receptor
A	M2	2(-1.25) + 2(-.75)	=	-4.0	
	NET			-4.0	-20.0
B	M1	2(1.25)	=	2.5	
	M2	-4.0 + 2(1.25)	=	-1.5	
	NET			1.0	5.0
C	M1	2.5 + 2(-1.25)	=	0.00	
	M2	-1.5 + 1.25	=	-.25	
	M3	1.25 + 2(-1.25)	=	-1.25	
	NET			-1.5	-7.5
D	M1	0 + 1.25	=	1.25	
	M2	-.25 + 0	=	-.25	
	M3	-1.25 + 1.25	=	0.00	
	M4	2(1.25)	=	2.50	
	NET			3.5	17.5

interactions such as charge-charge repulsions within the channel, but this is quite speculative and other stabilizing mechanisms may be present in such proteins.

The use of evolutionary variance to arrange the helices of the acetylcholine receptor was successful, and the resulting structure agrees well with known characteristics of the receptor. The application of the results obtained for this receptor to the homologous GABA receptor is less certain, because the conserved residues of the acetylcholine receptor are often not conserved in the GABA sequences. This difference is emphasized in the sequence alignment shown in figure 4.10, where sequences of the acetylcholine and GABA receptors have been aligned. It is evident from the figure that particular residues, which have been conserved in the acetylcholine receptor, have not been conserved in the GABA receptor, although the homology between the proteins is obvious. The insertion of gaps within the putative transmembrane regions was also necessary for the optimal alignment of the sequences. The helices of the GABA receptor have been arranged using a similar variability analysis to the one used in this modelling and the resulting structure was quite different than the structure proposed herein for the acetylcholine receptor (47). In that work, the GABA receptor was proposed to form two channels. There is some experimental evidence which also suggests that the structures of these two receptors are fundamentally different, although their sequences are homologous. The GABA receptor, for instance has a different subunit stoichiometry than the acetylcholine receptor, with four subunits instead of five (48).

Figure 4.10

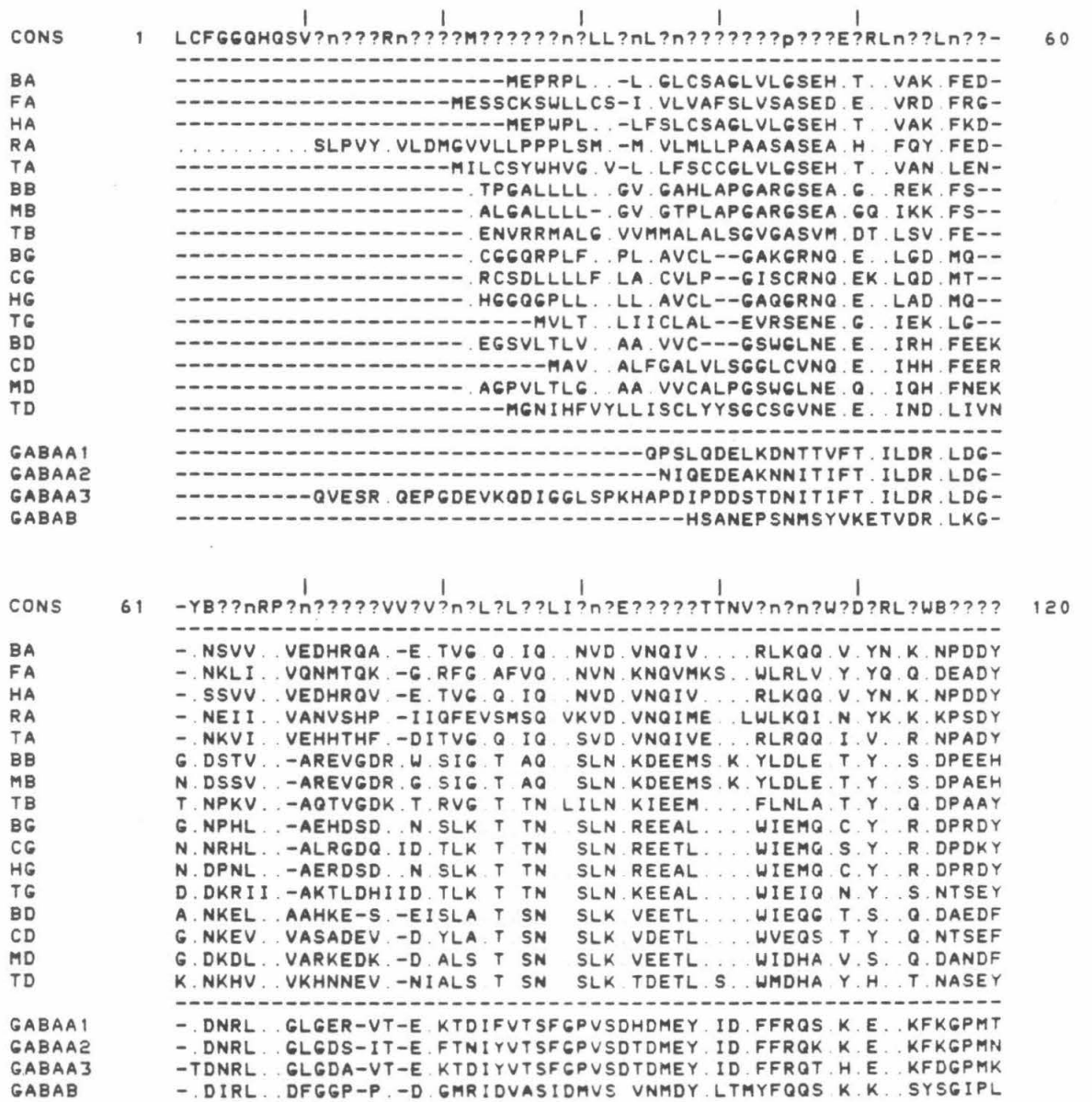
Alignment of GABA and Acetylcholine Receptor Sequences

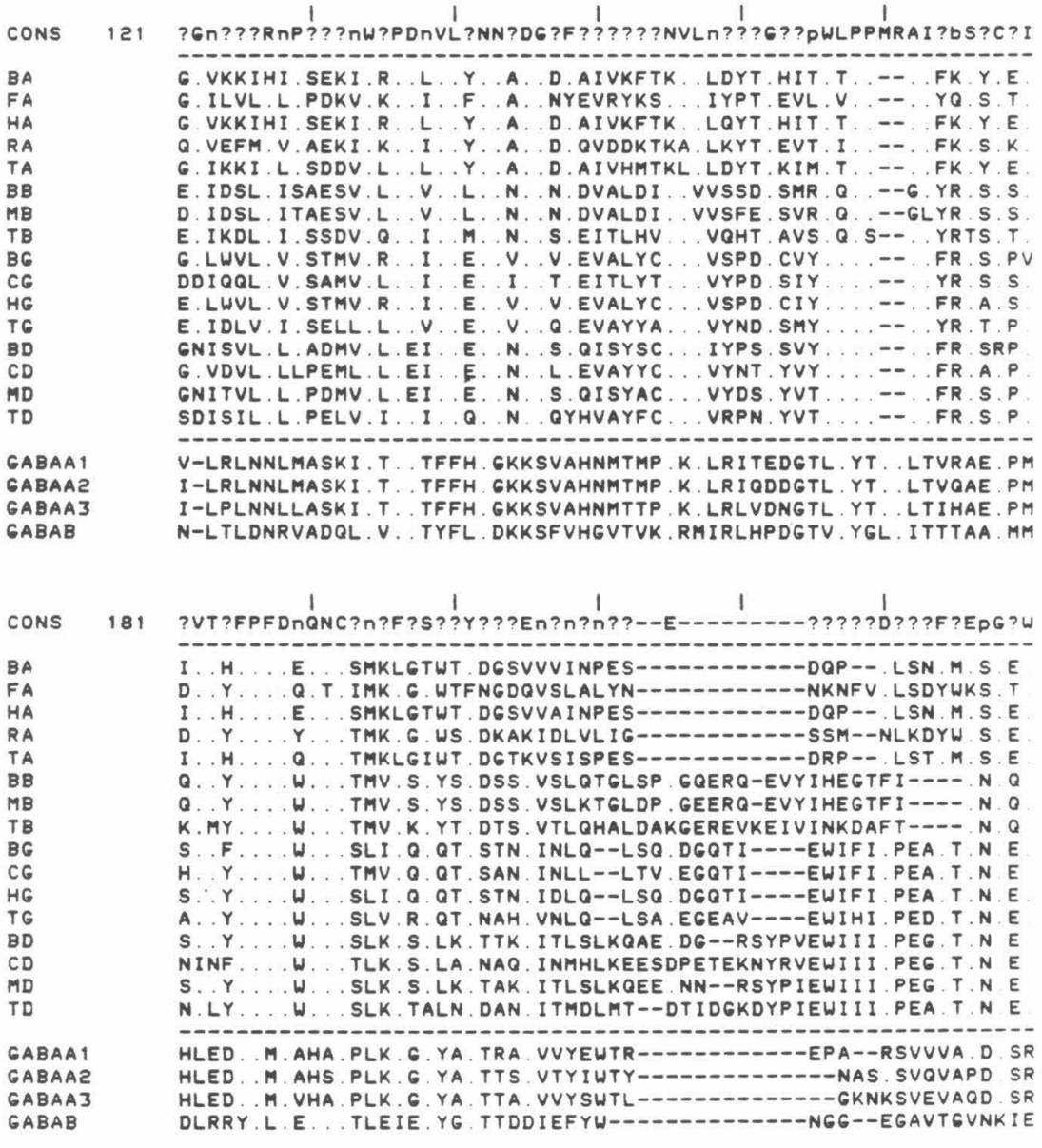
Sequences of subunits from both the GABA and acetylcholine receptors have been aligned. The putative transmembrane regions of the acetylcholine receptor have been underlined.

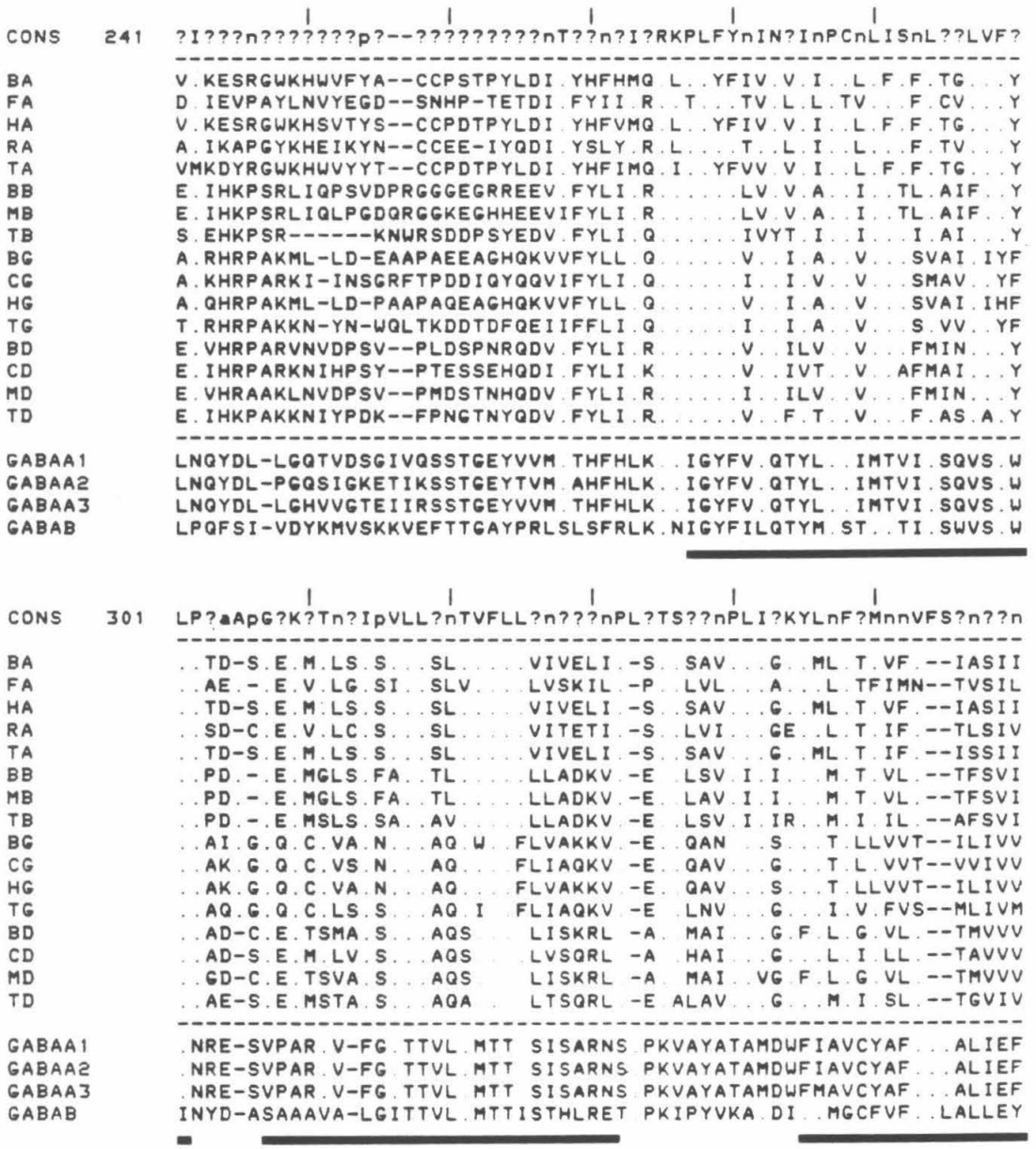
AChR sequences included: **BA-BD** bovine α-δ (50-53); **CG, CD** chick γ,δ (54); **FA** fly α (55); **HA, HG** human α,γ (50,56); **MB, MD** mouse β,δ (57,58); **RA** rat α (59); and **TA-TD** *Torpedo californica* α-δ (60, and references therein).

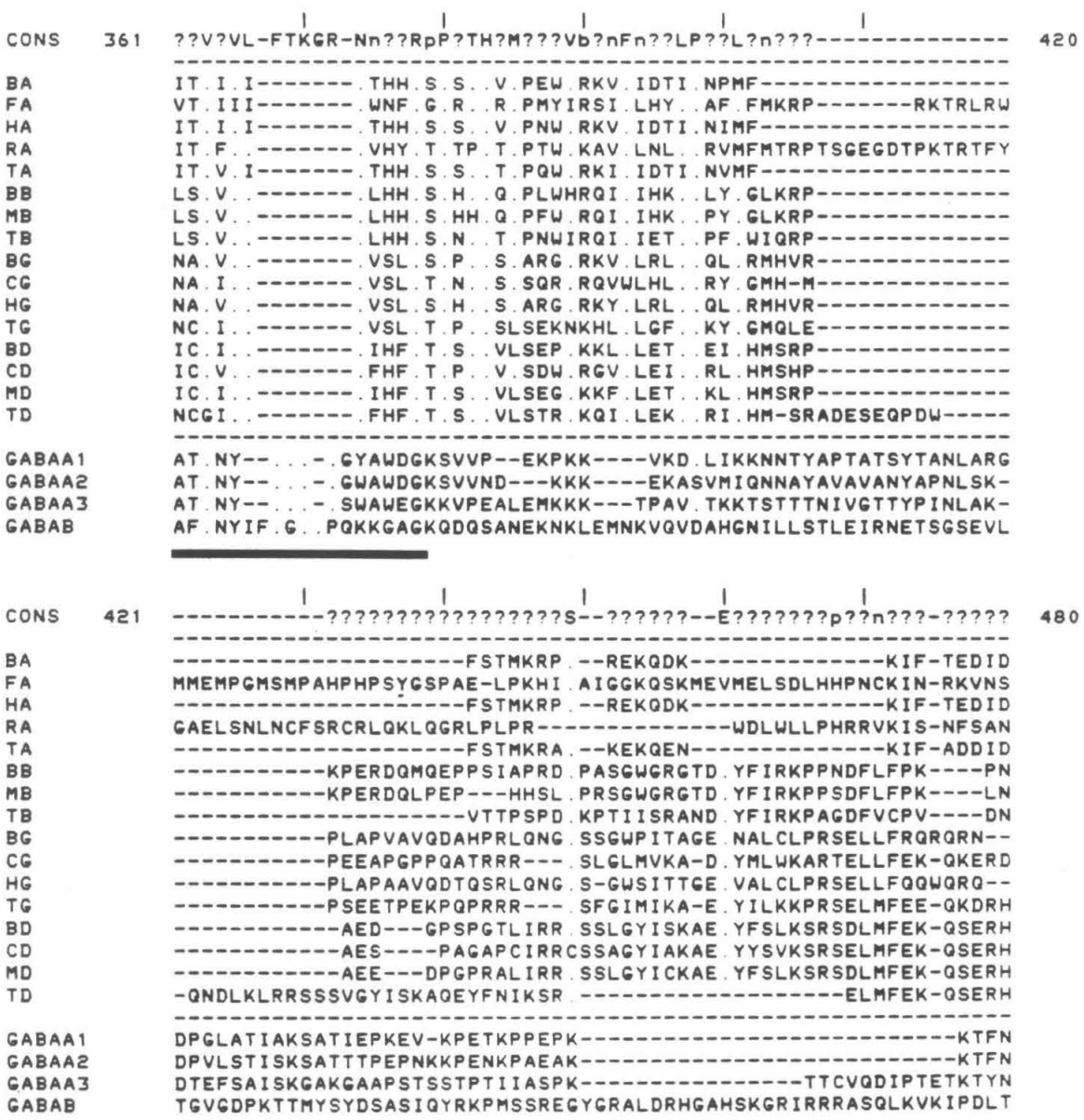
GABA sequences included: **GABAA1, GABAA2, GABAA3** bovine α (61,62); and **GABAB** bovine β (61).

GABA / AChR Sequence Alignment




CONS 121 ?Gn???RnP???nW?PDnVL?NN?DG?F??????NVLn???G??pWLPPMRAI?bS?C?I 180
BA G.VKKIHI.SEKI.R..L..Y..A..D.AIVKFTK..LDYT.HIT.T..--..FK.Y.E.
FA G.ILVL.L.PDKV.K..I..F..A..NYEVRYKS...IYPT.EVL.V..--..YQ.S.T.
HA G.VKKIHI.SEKI.R..L..Y..A..D.AIVKFTK..LQYT.HIT.T..--..FK.Y.E.
RA Q.VEFM.V.AEKI.K..I..Y..A..D.QVDDKTKA.LKYT.EVT.I..--..FK.S.K.
TA G.IKKI.L.SDDV.L..L..Y..A..D.AIVHMTKL.LDYT.KIM.T..--..FK.Y.E.
BB E.IDSL.ISAESV.L..V..L..N..N.DVALDI..VVSSD.SMR.Q..--G.YR.S.S.
MB D.IDSL.ITAESV.L..V..L..N..N.DVALDI..VVSFE.SVR.Q..--GLYR.S.S.
TB E.IKDL.I.SSDV.Q..I..M..N..S.EITLHV...VQHT.AVS.Q.S--..YRTS.T.
BG G.LWVL.V.STMV.R..I..E..V..V.EVALYC...VSPD.CVY....--..FR.S.PV
CG DDIQQL.V.SAMV.L..I..E..I..T.EITLYT...VYPD.SIY....--..YR.S.S.
HG E.LWVL.V.STMV.R..I..E..V..V.EVALYC...VSPD.CIY....--..FR.A.S.
TG E.IDLV.I.SELL.L..V..E..V..Q.EVAYYA...VYND.SMY....--..YR.T.P.
BD GNISVL.L.ADMV.L.EI..E..N..S.QISYSC...IYPS.SVY....--..FR.SRP.
CD G.VDVL.LLPEML.L.EI..E..N..L.EVAYYC...VYNT.YVY....--..FR.A.P.
MD GNITVL.L.PDMV.L.EI..E..N..S.QISYAC...VYDS.YVT....--..FR.S.P.
TD SDISIL.L.PELV.I..I..Q..N..QYHVAYFC...VRPN.YVT....--..FR.S.P.
GABAA1 V-LRLNNLMASKI.T..TFFH.GKKSVAHNMTMP.K.LRITEDGTL.YT..LTVRAE.PM
GABAA2 I-LRLNNLMASKI.T..TFFH.GKKSVAHNMTMP.K.LRIQDDGTL.YT..LTVQAE.PM
GABAA3 I-LPLNNLLASKI.T..TFFH.GKKSVAHNMTTP.K.LRLVDNGTL.YT..LTIHAE.PM
GABAB N-LTLDNRVADQL.V..TYFL.DKKSFVHGVTVK.RMIRLHPDGTV.YGL.ITTTAA.MM
CONS 181 ?VT?FPFDnQNC?n?F?S??Y???En?n?n??--E---------?????D???F?EpG?W 240
BA I..H....E...SMKLGTWT.DGSVVVINPES----------DQP--.LSN.M.S.E
FA D..Y....Q.T.IMK.G.WTFNGDQVSLALYN----------NKNFV.LSDYWKS.T
HA I..H....E...SMKLGTWT.DGSVVAINPES----------DQP--.LSN.M.S.E
RA D..Y....Y...TMK.G.WS.DKAKIDLVLIG----------SSM--NLKDYW.S.E.
TA I..H....Q...TMKLGIWT.DGTKVSISPES----------DRP--.LST.M.S.E
BB Q..Y....W...TMV.S.YS.DSS.VSLQTGLSP.GQERQ-EVYIHEGTFI----.N.Q
MB Q..Y....W...TMV.S.YS.DSS.VSLKTGLDP.GEERQ-EVYIHEGTFI----.N.Q
TB K.MY....W...TMV.K.YT.DTS.VTLQHALDAKGEREVKEIVINKDAFT----.N.Q
BG S..F....W...SLI.Q.QT.STN.INLQ--LSQ.DGQTI----EWIFI.PEA.T.N.E
CG H..Y....W...TMV.Q.QT.SAN.INLL--LTV.EGQTI----EWIFI.PEA.T.N.E
HG S..Y....W...SLI.Q.QT.STN.IDLQ--LSQ.DGQTI----EWIFI.PEA.T.N.E
TG A..Y....W...SLV.R.QT.NAH.VNLQ--LSA.EGEAV----EWIHI.PED.T.N.E
BD S..Y....W...SLK.S.LK.TTK.ITLSLKQAE.DG--RSYPVEWIII.PEG.T.N.E
CD NINF....W...TLK.S.LA.NAQ.INMHLKEESDPETEKNYRVEWIII.PEG.T.N.E
MD S..Y....W...SLK.S.LK.TAK.ITLSLKQEE.NN--RSYPIEWIII.PEG.T.N.E
TD N.LY....W...SLK.TALN.DAN.ITMDLMT--DTIDGKDYPIEWIII.PEA.T.N.E
GABAA1 HLED..M.AHA.PLK.G.YA.TRA.VVYEWTR----------EPA--RSVVVA.D.SR
GABAA2 HLED..M.AHS.PLK.G.YA.TTS.VTYIWTY----------NAS.SVQVAPD.SR
GABAA3 HLED..M.VHA.PLK.G.YA.TTA.VVYSWTL----------GKNKSVEVAQD.SR
GABAB DLRRY.L.E...TLEIE.YG.TTDDIEFYW----------NGG--EGAVTGVNKIE


CONS 241 ?I???n???????p?--?????????nT??n?I?RKPLFYnIN?InPCnLISnL??LVF? 300
BA V.KESRGWKHWVFYA--CCPSTPYLDI.YHFHMQ L..YFIV.V.I..L.F.F.TG...Y
FA D.IEVPAYLNVYEGD--SNHP-TETDI.FYII.R..T...TV.L.L.TV...F.CV...Y
HA V.KESRGWKHSVTYS--CCPDTPYLDI.YHFVMQ.L..YFIV.V.I..L.F.F.TG...Y
RA A.IKAPGYKHEIKYN--CCEE-IYQDI.YSLY.R.L....T..L.I..L...F.TV...Y
TA VMKDYRGWKHWVYYT--CCPDTPYLDI.YHFIMQ.I..YFVV.V.I..L.F.F.TG...Y
BB E.IHKPSRLIQPSVDPRGGGEGRREEV.FYLI.R.....LV.V.A..I..TL.AIF..Y
MB E.IHKPSRLIQLPGDQRGGKEGHHEEVIFYLI.R.....LV.V.A..I..TL.AIF..Y
TB S.EHKPSR------KNWRSDDPSYEDV.FYLI.Q.....IVYT.I..I...I.AI...Y
BG A.RHRPAKML-LD-EAAPAEEAGHQKVVFYLL.Q.....V..I.A..V...SVAI.IYF
CG A.KHRPARKI-INSGRFTPDDIQYQQVIFYLI.Q.....I..I.V..V...SMAV..YF
HG A.QHRPAKML-LD-PAAPAQEAGHQKVVFYLL.Q.....V..I.A..V...SVAI.IHF
TG T.RHRPAKKN-YN-WQLTKDDTDFQEIIFFLI.Q.....I..I.A..V...S.VV..YF
BD E.VHRPARVNVDPSV--PLDSPNRQDV.FYLI.R.....V..ILV..V...FMIN...Y
CD E.IHRPARKNIHPSY--PTESSEHQDI.FYLI.K.....V..IVT..V..AFMAI...Y
MD E.VHRAAKLNVDPSV--PMDSTNHQDV.FYLI.R.....I..ILV..V...FMIN...Y
TD E.IHKPAKKNIYPDK--FPNGTNYQDV.FYLI.R.....V..F.T..V...F.AS.A.Y
GABAA1 LNQYDL-LGQTVDSGIVQSSTGEYVVM.THFHLK..IGYFV.QTYL..IMTVI.SQVS.W
GABAA2 LNQYDL-PGQSIGKETIKSSTGEYTVM.AHFHLK..IGYFV.QTYL..IMTVI.SQVS.W
GABAA3 LNQYDL-LGHVVGTEIIRSSTGEYVVM.THFHLK..IGYFV.QTYL..IMTVI.SQVS.W
GABAB LPQFSI-VDYKMVSKKVEFTTGAYPRLSLSFRLK.NIGYFILQTYM.ST..TI.SWVS.W
CONS 301 LP?aApG?K?Tn?IpVLL?nTVFLL?n???nPL?TS??nPLI?KYLnF?MnnVFS?n??n 360
BA ..TD-S.E.M.LS.S...SL.....VIVELI.-S..SAV...G..ML.T.VF.--IASII
FA ..AE.-.E.V.LG.SI..SLV....LVSKIL.-P..LVL...A...L.TFIMN--TVSIL
HA ..TD-S.E.M.LS.S...SL.....VIVELI.-S..SAV...G..ML.T.VF.--IASII
RA ..SD-C.E.V.LC.S...SL.....VITETI.-S..LVI...GE..L.T.IF.--TLSIV
TA ..TD-S.E.M.LS.S...SL.....VIVELI.-S..SAV...G..ML.T.IF.--ISSII
BB ..PD.-.E.MGLS.FA..TL.....LLADKV.-E..LSV.I.I....M.T.VL.--TFSVI
MB ..PD.-.E.MGLS.FA..TL.....LLADKV.-E..LAV.I.I....M.T.VL.--TFSVI
TB ..PD.-.E.MSLS.SA..AV.....LLADKV.-E..LSV.I.IR..M.I.IL.--AFSVI
BG ..AI.G.Q.C.VA.N...AQ.W..FLVAKKV.-E..QAN...S...T.LLVVT--ILIVV
CG ..AK.G.Q.C.VS.N...AQ....FLIAQKV.-E..QAV...G...T.L.VVT--VVIVV
HG ..AK.G.Q.C.VA.N...AQ....FLVAKKV.-E..QAV...S...T.LLVVT--ILIVV
TG ..AQ.G.Q.C.LS.S...AQ.I..FLIAQKV.-E..LNV...G...I.V.FVS--MLIVM
BD ..AD-C.E.TSMA.S...AQS.....LISKRL.-A..MAI...G.F.L.G.VL.--TMVVV
CD ..AD-S.E.M.LV.S...AQS.....LVSQRL.-A..HAI...G...L.I.LL.--TAVVV
MD ..GD-C.E.TSVA.S...AQS.....LISKRL.-A..MAI..VG.F.L.G.VL.--TMVVV
TD ..AE-S.E.MSTA.S...AQA.....LTSQRL.-E.ALAV...G...M.I.SL.--TGVIV
GABAA1 .NRE-SVPAR.V-FG.TTVL.MTT.SISARNS.PKVAYATAMDWFIAVCYAF...ALIEF
GABAA2 .NRE-SVPAR.V-FG.TTVL.MTT.SISARNS.PKVAYATAMDWFIAVCYAF...ALIEF
GABAA3 .NRE-SVPAR.V-FG.TTVL.MTT.SISARNS.PKVAYATAMDWFMAVCYAF...ALIEF
GABAB INYD-ASAAAVA-LGITTVL.MTTISTHLRET.PKIPYVKA.DI..MGCFVF..LALLEY


CONS 361 ??V?VL-FTKGR-Nn??RpP?TH?M???Vb?nFn??LP??L?n???---------- 420
BA IT.I.I-------.THH.S.S..V.PEW.RKV.IDTI.NPMF----------
FA VT.III-------.WNF.G.R..R.PMYIRSI.LHY..AF.FMKRP-------RKTRLRW
HA IT.I.I-------.THH.S.S..V.PNW.RKV.IDTI.NIMF----------
RA IT.F..-------.VHY.T.TP.T.PTW.KAV.LNL..RVMFMTRPTSGEGDTPKTRTFY
TA IT.V.I-------.THH.S.S..T.PQW.RKI.IDTI.NVMF----------
BB LS.V..-------.LHH.S.H..Q.PLWHRQI.IHK..LY.GLKRP----------
MB LS.V..-------.LHH.S.HH.Q.PFW.RQI.IHK..PY.GLKRP----------
TB LS.V..-------.LHH.S.N..T.PNWIRQI.IET..PF.WIQRP----------
BG NA.V..-------.VSL.S.P..S.ARG.RKV.LRL..QL.RMHVR----------
CG NA.I..-------.VSL.T.N..S.SQR.RQVWLHL..RY.GMH-M----------
HG NA.V..-------.VSL.S.H..S.ARG.RKY.LRL..QL.RMHVR----------
TG NC.I..-------.VSL.T.P..SLSEKNKHL.LGF..KY.GMQLE----------
BD IC.I..-------.IHF.T.S..VLSEP.KKL.LET..EI.HMSRP----------
CD IC.V..-------.FHF.T.P..V.SDW.RGV.LEI..RL.HMSHP----------
MD IC.I..-------.IHF.T.S..VLSEG.KKF.LET..KL.HMSRP----------
TD NCGI..-------.FHF.T.S..VLSTR.KQI.LEK..RI.HM-SRADESEQPDW-----
GABAA1 AT.NY--...-.GYAWDGKSVVP--EKPKK----VKD.LIKKNNTYAPTATSYTANLARG
GABAA2 AT.NY--...-.GWAWDGKSVVND---KKK----EKASVMIQNNAYAVAVANYAPNLSK-
GABAA3 AT.NY--...-.SWAWEGKKVPEALEMKKK---TPAV.TKKTSTTTNIVGTTYPINLAK-
GABAB AF.NYIF.G..PQKKGAGKQDQSANEKNKLEMNKVQVDAHGNILLSTLEIRNETSGSEVL
CONS 421 ----------????????????????S--??????--E???????p??n???-????? 480
BA ----------FSTMKRP.--REKQDK----------KIF-TEDID
FA MMEMPGMSMPAHPHPSYGSPAE-LPKHI.AIGGKQSKMEVMELSDLHHPNCKIN-RKVNS
HA ----------FSTMKRP.--REKQDK----------KIF-TEDID
RA GAELSNLNCFSRCRLQKLQGRLPLPR----------WDLWLLPHRRVKIS-NFSAN
TA ----------FSTMKRA.--KEKQEN----------KIF-ADDID
BB ----------KPERDQMQEPPSIAPRD.PASGWGRGTD.YFIRKPPNDFLFPK----PN
MB ----------KPERDQLPEP---HHSL.PRSGWGRGTD.YFIRKPPSDFLFPK----LN
TB ----------VTTPSPD.KPTIISRAND.YFIRKPAGDFVCPV----DN
BG ----------PLAPVAVQDAHPRLQNG.SSGWPITAGE.NALCLPRSELLFRQRQRN--
CG ----------PEEAPGPPQATRRR---.SLGLMVKA-D.YMLWKARTELLFEK-QKERD
HG ----------PLAPAAVQDTQSRLQNG.S-GWSITTGE.VALCLPRSELLFQQWQRQ--
TG ----------PSEETPEKPQPRRR---.SFGIMIKA-E.YILKKPRSELMFEE-QKDRH
BD ----------AED---GPSPGTLIRR.SSLGYISKAE.YFSLKSRSDLMFEK-QSERH
CD ----------AES----PAGAPCIRRCSSAGYIAKAE.YYSVKSRSELMFEK-QSERH
MD ----------AEE---DPGPRALIRR.SSLGYICKAE.YFSLKSRSDLMFEK-QSERH
TD -QNDLKLRRSSSVGYISKAQEYFNIKSR.----------ELMFEK-QSERH
GABAA1 DPGLATIAKSATIEPKEV-KPETKPPEPK----------KTFN
GABAA2 DPVLSTISKSATTTPEPNKKPENKPAEAK----------KTFN
GABAA3 DTEFSAISKGAKGAAPSTSSTPTIIASPK----------TTCVQDIPTETKTYN
GABAB TGVGDPKTTMYSYDSASIQYRKPMSSREGYGRALDRHGAHSKGRIRRRASQLKVKIPDLT

```
                                    |         |         |         |         |
CONS      481   ?n?-????--???n?????--?--?L??---?n---V-????IA???b???????????W   540
                ------------------------------------------------------------
BA              I-SDISGKPGPPPMGFHSPLIK--HPEVKSAIE-----GIKY..ETMKSDQESNNAAEE.
FA              G-GELGLGDGCRRESESSDSIL------LSPEASKATEAVEF..EHLRNEDLYIQTRED.
HA              I-SDISGKPGPPPMGFHSPLIK--HPEVKSAIE-----GIKY..ETMKSDQESNNAAAE.
RA              L-TRSSS-------SESVNAVL--S.SALSPEIKEAIQSVKY..ENMKAQNVAKEIQDD.
TA              I-SDISGKQVTGEVIFQTPLIK--NPDVKSAIE-----GVKY..EHMKSDEESSNAAEE.
BB              ---RFQPE--LSAPDLRRFIDGPNRAVGLPPELREV.SSISY..RQLQEQEDHDVLKED.
MB              ---RFQPE--SSAPDLRRFIDGPTRAVGLPQELREVISSISYM.RQLQEQEDHDALKED.
TB              IRVAVQPE--RLFSEMKWHLNGLTQPVTLPQDLKEA.EAIKY..EQLESASEFDDLKKD.
BG              GLVRAALEKLEKGPESGQSPEWCGS.KQAAPAIQAC.EACNL..RARHQQTHFDSGNKE.
CG              GLMKTVLEKIGRGLESNRAQDFCQS.EEASPEIRAC.EACNH..NATREQNDFSSENEE.
HG              GLVAAALEKLEKGPELGLSQ-FCGS.KQAAPAIQAC.EACNL..CARHQQSHFDNGNEE.
TG              GL--KRVNKMTSDIDIGTTVDLYKD.ANFAPEIKSC.EACNF..KSTKEQNDSGSENEN.
BD              GLA-RRLTTARRPPAGSEQ--AQQE.FS---ELKPA.DGANF.VNHMKDQNNYNEEKDC.
CD              GLA-SRVTPARFAPAATSE--;EQ-.YD---HLKPTLDEANF.VKHMREKNSYNEEKDN.
MD              GLA-RRLTTARRPPASSEQ--VQQE.FN---EMKPA.DGANF.VNHMRDQNSYNEEKDN.
TD              GLV-PRVTPRIGFGNNNENIAASDQ.HD---EIKSGIDSTNY.VKQIKEKNAYDEEVGN.
                ------------------------------------------------------------
GABAA1          SVS---------------------------------------------------------
GABAA2          SVS---------------------------------------------------------
GABAA3          SVS---------------------------------------------------------
GABAB           DVN---------------------------------------------------------


                                    |         |         |         |         |
CONS      541   ??VA??nDRn??nnFnnn??nGTn??Fn???????n??PF?GDP??Y?????R?I        595
                ------------------------------------------------------------
BA              KY..MVM.HILLAV.MLVCII..LAV.AGRLIELNQQG
FA              KY..MVI..LQLYI.FIVTTA..VGILMDAPHIFEYVDQDRIIEIYRGK
HA              KY..MVM.HILLGV.MLVCII..LAV.AGRLIELNQQG
RA              KY..MVI..IFLWV.ILVCIL..AGL.LQPLMARDDT
TA              KY..MVI.HILLCV.MLICII..VSV.AGRLIELSQEG
BB              QF..MVV..LFLWT.IIFTSV..LVI.LDATYHLPPAD..P
MB              QF..MVV..LFLWT.IVFTSV..LVI.LDATYHLPPPE..P
TB              QY..MVA..LFLYV.FVICSI..FSI.LDASHNVPPDN..A
BG              FL.GRVL..VCFLAMLSLFVC..AGI.LMAHYNRVPAL..P...RS.LPSSD
CG              IL.GRVI..VCFFIMASLFVC..IGI.LMAHFNQAPAL..P...KT.LPP
HG              FL.GRVL..VCFLAMLSLFIC..AGI.LMAHYNRVPAL..P...RP.LPSPD
TG              VLIGKVI.KACFWIALLLFSI..LAI.LTGHFNQVPEF..P...RK.VP
BD              NR..RTV..LCLFVVTPIMVV..AWI.LQGAYNQPPPQ..P...FS.LEKDK.F.
CD              NR..RTL..LCLFLITPMLVV..LWI.LMGIYNHPPPL..S...FD.REENK.Y.
MD              NQ..RTV..LCLFVVTPVMVV..AWI.LQGVYNQPPLQ..P...FS.SEQDK.F.
TD              NL.GQTI..LSMFIITPVMVL..IFI.VMGNFNHPPAK..E...FD.SSDHP.CA
                ------------------------------------------------------------
GABAA1          -----KI..LSRIA.PLLFGIFNLVYWATYLNREPQLKAPTPHQ
GABAA2          -----KI..MSRIV.PVLFGTFNLVYWATYLNREREPVLGV
GABAA3          -----KV.KISRII.PVLFAIFNLVYWATYVNRESAIKGMIRKQ
GABAB           -----SI.KWSRMF.PITFSLFNVVYWLYYVH
```

Summary

This modelling of the acetylcholine receptor has shown that the use of evolutionary variability can be used to aid in the arrangement of transmembrane helices within the bilayer and to orient them with respect to each other. The resulting model agrees with all current structural information about the receptor and could explain some observed differences between certain receptors. Some possible mechanisms for channel gating were suggested by this structure and may serve to motivate future experiments.

4D. Rhodopsin

This protein was initially chosen for modelling because it is a well-characterized membrane protein. It was hoped that the current knowledge about this protein's structure would allow further testing of the ideas on variability, which have been discussed in the last few sections of this work. The three-dimensional arrangement of rhodopsin's helices within the membrane has been determined by electron diffraction (63) and each helix has tentatively been correlated to a transmembrane region identified in the sequence (64). This correlation was based on the work of Engelman *et al.* (65), who assigned the helices of bacteriorhodopsin through the examination of their interconnecting loop lengths, and on the results of neutron scattering work, which indicated the probable locations of M1 and M2, the first two helices in the sequence (66). Findlay (64) used this information along with that obtained from labelling studies using a lipophilic probe (67,68) to orient the faces of the helices with respect to the membrane. He also did some computer modelling of the protein in that work.

Rhodopsin was modelled using the procedure described in section 4B. First, helices were grouped on the basis of experimentally derived structural information and through the examination of their primary sequences. Then the results of the variability analysis, which was described in Chapter 3, were applied to the arrangement and orientation of the helices. Some structural information derived from studies of the β-adrenergic receptor and the muscarinic acetylcholine

receptor has also been considered in this modelling. Ovine rhodopsin has been modelled in particular, although the other opsins will also be discussed.

Residue positions given in this section correspond to the numbering of the consensus sequence in figure 2.9.

Helix groupings. Many of the interhelical loops of rhodopsin are short, with less than 15 residues. Helices connected by these short residues are probably adjacent. For this reason, helices M1 and M2, M2 and M3, M3 and M4, and M6 and M7 are assumed to be adjacent to one another.

A disulfide bond between cysteines at positions 127 and 204 in the sequence alignment of figure 2.9 (69,70,89) constrains the middle of the extracellular M4-M5 loop to the N-terminal end of M3. When the number of residues involved in these linkages is taken into account, it appears likely that M4 and M5 are adjacent to M3, although the 15-residue length between the disulfide bond at the terminus of M3 and the N-terminus of M5 makes it possible, but unlikely, that these two helices are non-adjacent.

Helix arrangement. The arrangement of rhodopsin's helices based upon their variability profiles proved to be a difficult task. The biggest problem arose from the fact that all the helices have a similar orientation with respect to the bilayer and with respect to the other helices. The low-resolution structure (63) shows that each helix has one face exposed to the lipid bilayer, and one face buried within the helix cluster. The extent of lipid contact is also similar for each helix. Recent information about the rhodopsin family's

mechanism of signal transduction has implicated the formation of receptor oligomers (71,72), although such oligomers have not been characterized. If the protein does form oligomers, the regions of protein contact should be conserved. A further complication to the variability analysis would then arise if these regions lie within the membrane. The outer faces of interlocking helices would be conserved and any helical arrangement based upon variability would be invalid. For these reasons, the variability analysis could not be used to correlate the transmembrane regions identified in the sequence with the known three-dimensional structure.

Helices were arranged as shown in figure 4.11a and their facial orientations determined on the basis of their variability. The distributions of variable and conserved residues can be seen in figure 4.11a and b. (The residues assigned as conserved or variable for ovine rhodopsin are shown in figure 3.3.) These orientations were then compared to the sites labelled with the lipophilic probe, azidoiodobenzene (67,68) and the results can be seen in figure 4.11c. Included in this figure are the positions of several residues that were not labelled by the hydrophobic probe and, therefore, proposed to be protected from the lipid mileu. Although there is some ambiguity in the orientation of helices 2 and 4 on the basis of variability, one can see that, in general, the sites of labelling do not correspond to the regions of rhodopsin predicted to be lipid-exposed.

There are several possible reasons for this discrepancy aside from a conclusion that variability cannot be used to facially orient transmembrane helices. It is possible that too few sequences have

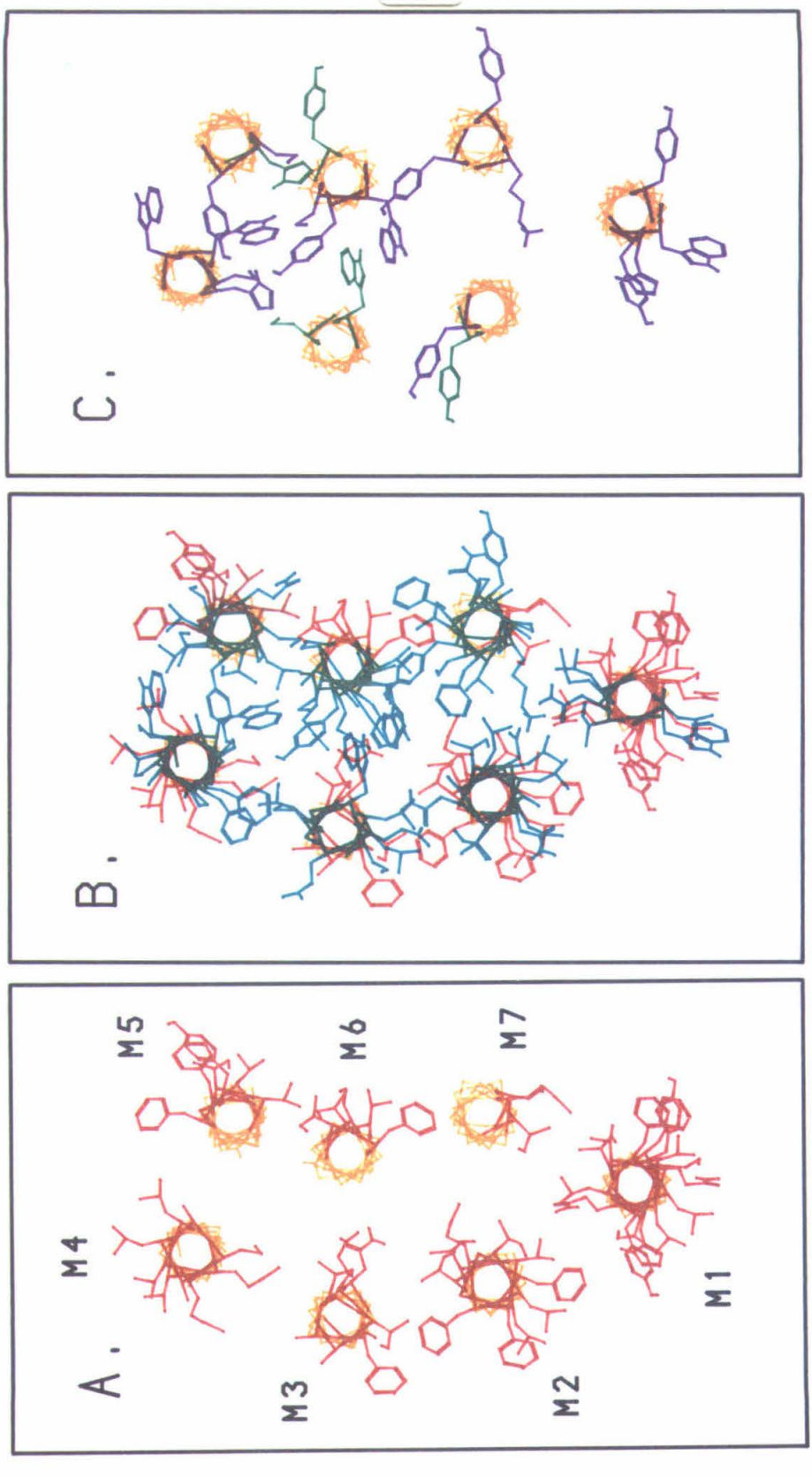
Figure 4.11

Facial Orientation of Rhodopsin's Helices:

Lipophilic Probe *vs.* Variability Analysis

The helical arrangement chosen for the modelling of rhodopsin is shown as viewed from the photoreceptor disc.

A. The variable residues are indicated in red around the yellow helix backbones. **B.** The conserved residues are indicated here in blue and the variable residues in red. **C.** The positions of the residues labelled by the hydrophobic reagent, azido-iodobenzene, are indicated in purple, and residues which were protected from the label are shown in green.


A.
M4
M5
M6
M7
M3
M2
M1
B.
C.

been incorporated into this analysis to yield the correct assignment of variable and conserved residues. To see if this was the case, the variability analysis was also done on a set of sequences, which included muscarinic and β-adrenergic receptors (results not shown). Although the orientation of M2 was clarified, the results were not significantly different from those shown in figure 4.11. As discussed in Chapter 3, evolutionary divergence may decrease the accuracy of this type of variability analysis. Indeed, the sequence alignment of the rhodopsin family required the insertion of gaps within putative transmembrane regions, unlike the alignment of the opsins alone. These gaps may introduce an effective rotation of 100^{o} around the helical axis for succeeding residue positions, making the direct comparison of gapped and ungapped sequences ambiguous. One interesting feature of all the receptors, however, is that the N-terminal portion of M2 is well-conserved while the C-terminal half is variable, without a conserved face.

Although the lipid-exposed faces of these helices are expected to be variable and the buried faces conserved, the nature of rhodopsin also makes the use of this criterion to orient the helices suspect. Retinal, which is covalently linked into the center of the helical cluster, may have the same effect on the buried residues that the hydrocarbon chains of the lipid bilayer have on the exposed side chains, *i.e.*, these buried regions may be variable due to a non-specific interaction with retinal rather than conserved through the constraints of helix-helix interactions. It is also worth noting that the lipophilic probe used for the labelling studies is similar to

retinal in that it is planar, unsaturated, hydrophobic, and about the size of retinal's β-ionone ring.

Overall, rhodopsin does not seem to be a suitable protein for the application of the variability analysis. Some insights into the possible mechanism of signal transmission by rhodopsin can be gained from the current model, however.

Helical arrangement and implications to protein function. This helical arrangement was chosen for this modelling because it incorporated all the features discussed above in the section on helix groupings. In addition, the interactions of helix dipoles in this arrangement is estimated to stabilize the receptor by about 10 kcal/mole. This estimate was made using the process and values described in the last section.

The use of variability to determine the facial orientation of the helices allows lysine 313 to point toward the center of the protein, although a rotation of 20^{o} is required for it to be directed into the helical cluster. This lysine is the retinal attachment point (73). Both the variability analysis and the labelling work indicate that Asp 100 faces into the helix cluster. This residue is located in the center of M2 and is conserved in all sequences with the exception of the blue-sensitive iodopsin. A second acidic residue, Glu 139, is located in the middle of the membrane region and is predicted to face into the helical cluster by both helix variabiliy and hydrophobic labelling. This charge is conserved only in the rod receptors. It is proposed that these two charges are the two negative point charges predicted by Kakitani, *et al.* (74) to account for the "opsin shift" of

the rhodopsins. As predicted, these residues both fall on the same side of the plane defined by the chromophore. One charge is located near the protonated Schiff base and the other is near C12 of the chromophore. In recently reported work, Kosower (75) also proposed that these two residues are responsible for the opsin shift and postulated the formation of an ion pair between Glu 139 and His 228. The present work also indicates that such ion-pairing is possible. The current helical arrangement can accomodate the specific arrangement of charges proposed by Kakitani (74). The charged residues of ovine rhodopsin that are located in the center of the bilayer are shown in figure 4.12.

The positions of charged residues located within the red and green iodopsins are shown in figure 4.13. Although the negative charge is conserved at position 100, the one at residue 139 has been replaced by Glu 103 on M2, which is also likely to effect the required bathochromic shift in the chromophore's absorption spectrum. This observation has also been made by Kosower (75). There are no charge differences within the transmembrane region of the red and green iodopsins, although there are some differences in the distribution of polar residues. The red iodopsin has more polar residues surrounding the chromophore than the green iodopsin does, as can also be seen in figure 4.13. These polar residues probably stabilize the excited state of the receptor by interacting with the positive charge, which becomes distributed further along the chromophore upon its isomerization. The residues most likely to effect the bathochromic shift of red-sensitive iodopsin occur at positions 181, 278, and 310. It is

Figure 4.12

Charge Distribution Within Rhodopsin's Transmembrane Region

The locations of charges located near the center of the rhodopsin's transmembrane region have been indicated in this stereo drawing. Green denotes negatively charged residues and red denotes positively charged side chains. Histidine residues have been indicated in blue and a possible orientation of cis-retinal has been drawn in purple. The protein is oriented as in figure 4.11.

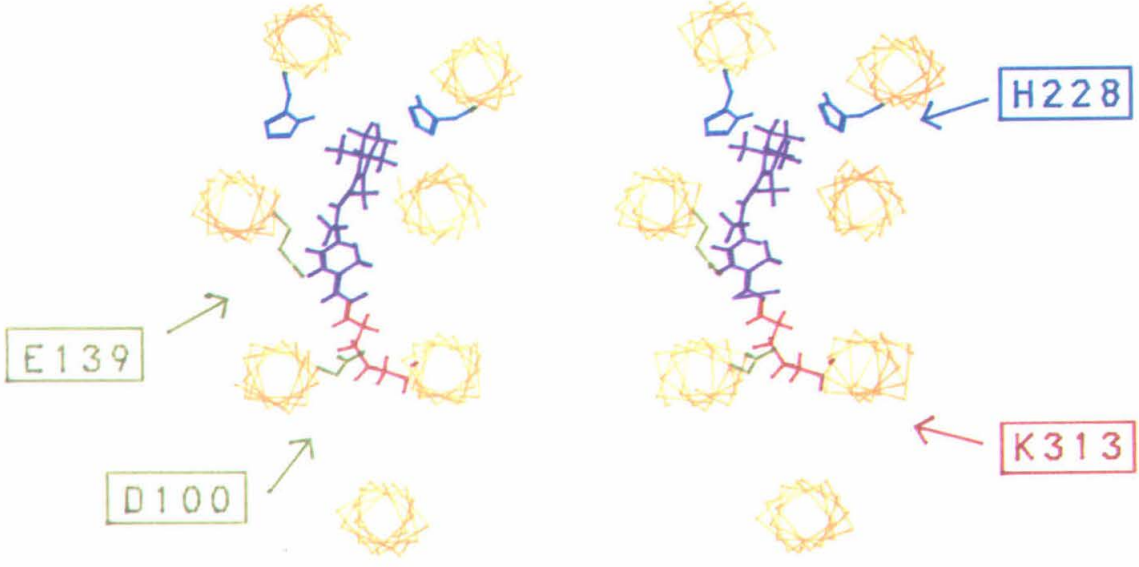
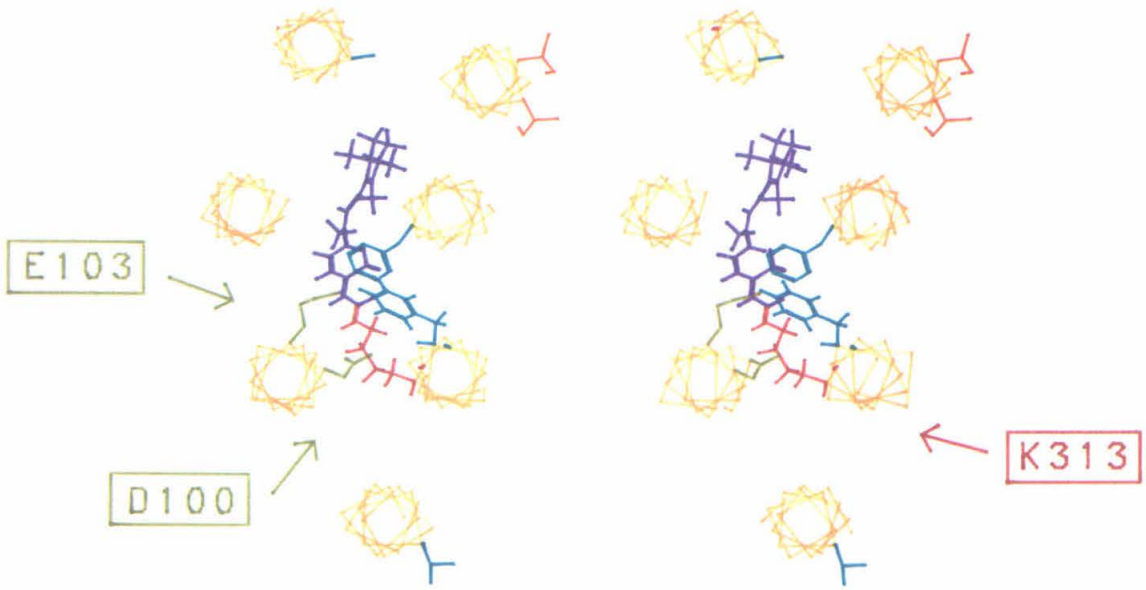

E139
D100
H228
K313

Figure 4.13

Charge Distributions Within the Transmembrane Regions

of the Iodopsins

The charge distribution of red and green iodopsin is shown and the positions of polar residues which occur in only one receptor type have been indicated. Hydrophilic residues that occur in the red iodopsin but not in the green iodopsin have been indicated in blue, and hydrophilic residues that occur only in the green iodopsin are shown in orange. Only the polar residues which fall near the retinal moiety have been drawn. It should be noted that only the residue positions have been indicated, the actual residues shown are those of the ovine rhodopsin. The orientation of the protein is the same as that of figure 4.11.


E103
D100
K313

possible that the two polar residues of green-sensitive iodopsin actually face into the helical cluster. Even in this orientation, they would not interact with the conjugated region of the chromophore and, therefore, would probably not influence the wavelength of chromophore absorption. The identification of the residues responsible for the different sensitivities of these proteins might best be addressed through site-specific mutagenesis.

The blue-sensitive iodopsin has no charged residue in the center of the bilayer other than the retinal-binding lysine residue. It has been suggested that the protonated Schiff bases associated with blue pigments may not require the influence of other charges to yield the required wavelength sensitivity (74).

Possible mechanisms of signal transduction. The rhodopsin family's receptors effect signal transduction through the activation of homologous G proteins, which bind to the cytoplasmic region of the receptor. Specific regions of the receptors have been implicated in this binding and include the loop between M3 and M4, certain residues of the M5-M6 loop, and the C-terminal portion of the protein (76-78). This C-terminal region plays a role in receptor regulation because its phosphorylation by various kinases, including receptor-specific kinases, results in receptor desensitization (79-83).

It has been postulated from studies of G-protein/receptor interactions that at least two portions of the G-protein interact with the receptor (84). If the corresponding regions of the receptor move relative to one another, the induced conformational changes in the G-protein could result in signal transduction.

The most obvious helix motion, which would result in the required displacements of G-protein binding sites, is a hinge-like movement of the helices relative to one another. This type of motion has been suggested by Wessling-Resnick *et al.* (85), although the details of their mechanism differ from those described below. A hinge-like motion is easily visualized for the opsins, where the isomerization of rhodopsin to the 11-trans form might cause the helices near the β-ionone ring to move away from the other helices of the cluster, especially in the region between M5 and M6. Helices M3, M4, and M5 would probably move as a unit because of the disulfide bond that holds them together. This type of motion is depicted schematically in figure 4.14. On the cytoplasmic surface of the protein, such a rotation would cause the cytoplasmic M3-M4 and M5-M6 loops to change their positions relative to the other regions of the receptor involved in G-protein binding. The possible movements of these regions has also been schematically illustrated in figure 4.14.

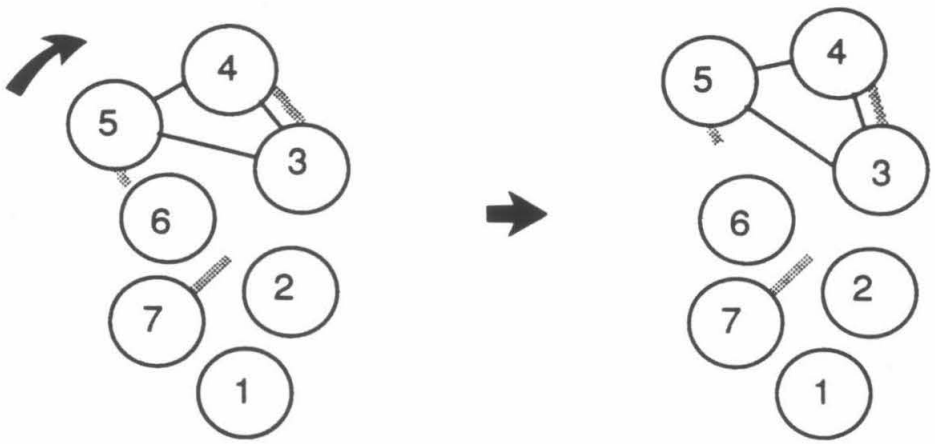
The specific interactions which could effect a similar motion in the muscarinic and β-adrenergic receptors are more difficult to identify. These receptors are activated by the binding of certain positively-charged agonists within their transmembrane regions (70,86-89). Site-directed mutagenesis and chemical labelling experiments have identified regions of these receptors which are involved in ligand binding. One residue, implicated in agonist binding for both receptor types, corresponds to position 100 of the rhodopsin alignment (87,89). In the β-adrenergic receptor, agonist binding was reduced when this residue was modified, but antagonist binding was not (89).

Figure 4.14

Possible Helix Movement During Receptor Activation

The mechanism which is proposed to result in G-protein activation has been illustrated in this schematic diagram. The helical arrangement is depicted as viewed from the cytoplasm. Regions of the cytoplasmic loops, which have been implicated in G-protein binding, have been indicated by shaded bars and the three helices linked to one another through the disulfide bond have been connected by solid black lines. The M3-M4-M5 group is proposed to move as a unit, with the least amount of movement occuring at M3. It is possible that M6 will reorient to maintain the M5-M6 interhelical contact.

Proposed Helix Motions for Rhodopsin



A second negatively charged residue, which corresponds to alignment position 134, has been shown to be essential for both agonist and antagonist binding in the β-adrenergic receptor (86). Because this residue is conserved in the muscarinic receptors, it is likely to play a role in these proteins as well. Residue 100 is in the center of M2 and corresponds to the counterion of the protonated Schiff base of rhodopsin. Residue 134 lies on M3, about one third of the way through the membrane from the extracellular surface. It is possible that agonist binding causes a change in the interaction between M2 and M3 and results in the opening of the "hinge" as described above. Because M3 is closely associated with M4 and M5, the positions of these helices should also be affected. Their arrangement may, in fact, amplify a small movement in the M3 region into a larger one near M5.

Residue 134, which plays a role in agonist binding, and 139, which has been proposed in this work to regulate the absorption maximum of rhodopsin's chromophore, are not on the same face of M3. Glu 130, which is conserved in all the opsins studied, does lie on the same helical face as 134, however. This observation suggests that either the positioning of M3 differs in the various receptors, or that Glu 130 may regulate the opsin shift. The latter possibility implies that the facial orientation of M3 within the protein sequence is incorrect in the present model. Studies similar to those of Kakitani *et al.* (74) may provide insight into which of these two orientations, if either, is likely.

According to this model, interactions of antagonists with the receptor should prevent the opening of the hinge. One type of interac-

tion is suggested by the roles played by the two acidic residues described above. The observation that antagonists do not interact with the negative charge on helix 2, but do interact with helix 3 indicates that the binding of these compounds may not affect the interactions between M2 and M3 and would not, therefore, trigger receptor response. Because these antagonists often have a large, usually planar and unsaturated hydrophobic region, it is possible that residues on M3, M4, and M5 are involved in their binding. These helices are the same as those proposed to interact with the β-ionone ring of retinal. An alternate mechanism of antagonist inhibition might be through tethering the hinge in a closed position. A strong interaction of antagonist with M6 and any of the three helices M3, M4, or M5 would prevent their movement.

The differences in agonist affinity between the β-adrenergic and muscarinic receptors is presumably related to differences in the distributions of polar and charged residues within the helix cluster. Although these distributions have not been analyzed in detail here, many of the helices show different hydrophobic/hydrophilic profiles when the aligned sequences are examined (results not shown). The muscarinic receptor contains an additional negative residue in helix 3, but its position is not on the same helical face as residue 134 and it is, therefore, probably not involved in the agonist-receptor interaction unless helix M3 is somewhat distorted.

Summary

The presented model of rhodopsin can account for several known features of this receptor. First, it predicts that all the charged

residues which occur near the centers of transmembrane helices are oriented into the helical cluster, rather than toward the membrane lipids. The positioning of these charges can account for the differences in wavelength selectivity for the various photoreceptors. This model also suggests a hinge-type mechanism, which would change the relative positioning of M3, M4, and M5 with respect to the other receptor helices. The resultant movement of the cytoplasmic M3-M4 loop relative to the receptor's C-terminal region is proposed to cause activation of the receptor-associated G-protein. This mechanism was also postulated to account for the activity of the muscarinic acetylcholine receptor and the β-adrenergic receptor in response to agonist binding.

4E. Band 3

Because so little is known about the structure of band 3 and because very few of its sequences are known, its three-dimensional modelling was more difficult than for either of the preceding proteins. The modelled protein is also less likely to correspond to the actual structure.

Residue numbers given in this section refer to the alignment numbering of figure 2.12. Murine erythrocyte band 3 was chosen for this modelling.

Helix groupings. Many of the transmembrane regions are connected by short lengths of the peptide chain, so most successive helices are predicted to be adjacent. The few exceptions are helices 5 and 6, 7 and 8, 12 and 13, and perhaps 10 and 11. A disulfide bond, probably between residues 969 and 1011, forms during the preparation of ghost membranes (90). This further reinforces the prediction that helices 13 and 14 are adjacent because each contains one of these two cysteines.

Because band 3 exists as a dimer within the membrane, and sometimes as a tetramer (91-93), the helical groupings in the region of the interface must also be considered. Bis-sulfosuccinimidyl sulfate cross-links the two subunits by reacting with the 5-6 extramembraneous loop (94). These two helices are, therefore, probably located at the interface between the dimers.

Helical arrangement. As before, the variable residues were used to arrange the helices within the bilayer. The best arrangement can

be seen in figure 4.15 and the distribution of variable residues is displayed in figure 4.16. Although most of the helices on the perimeter of the helical cluster have a variable face, those in the center were quite variable, contrary to the hypothesis on variability and helical packing. Experimental evidence or the ionic character of these regions required their location within the helical cluster, as explained below.

The anion transport region is proposed to be made up of helices M4, M5, M7, M10, M13, and M14. Of these, M5 is the most conserved, perhaps due to a close interaction between subunits at this location. A lysine residue near the C-terminus of this helix is covalently modified by H_2DIDS, 4,4'-diisothiocyanodihydrostilbene-2,2'-disulfonate (98-100). H_2DIDS, like many stilbene disulfonate derivatives, inhibits band 3 by interacting at the outward-facing, high-affinity anion binding site (95-99). This lysine is not believed to be an integral part of the high-affinity site, however, because the isothiocyanate moiety is removed from the two sulfonate groups that are responsible for the high binding affinity of the compound. M4 is remarkable only in that the face oriented toward the outside of the helix cluster contains two tryptophans, three phenylalanines, and one tyrosine. M7, although having no conserved face, is probably a part of the ion transport machinery because it contains four charged residues which are cofacial and extend into the hydrophobic region of the bilayer. A carboxylate group in this region can be modified by an extracellular carbodiimide, inhibiting anion transport (101). M10 has been included as part of the anion-transporting region because it con-

Figure 4.15

Helical Arrangement of Band 3

The proposed helical arrangement for a band 3 dimer is shown. The protein is viewed from the cytoplasm.

Helix Arrangement for Band 3

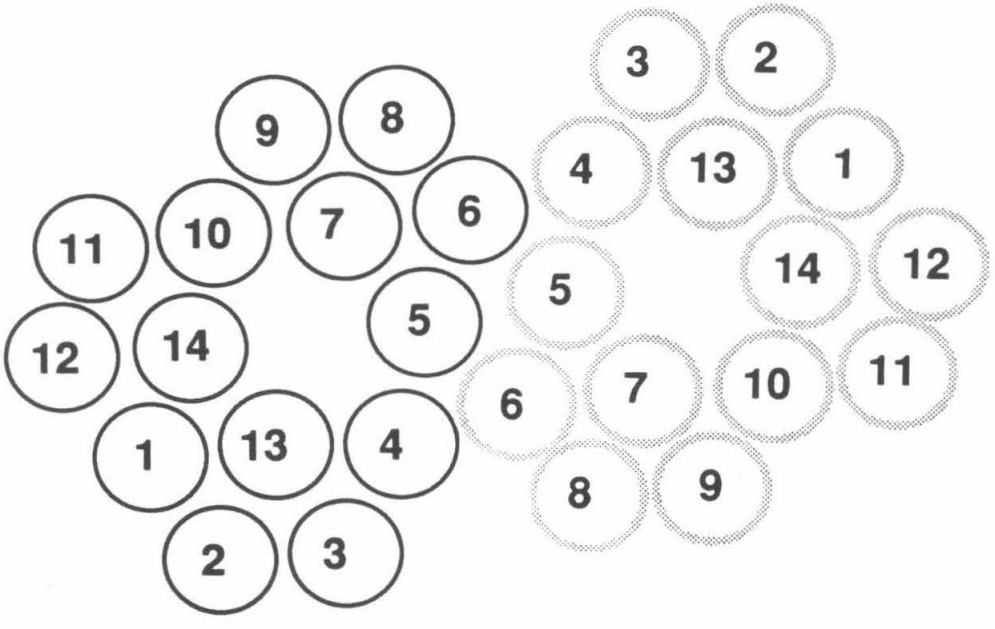


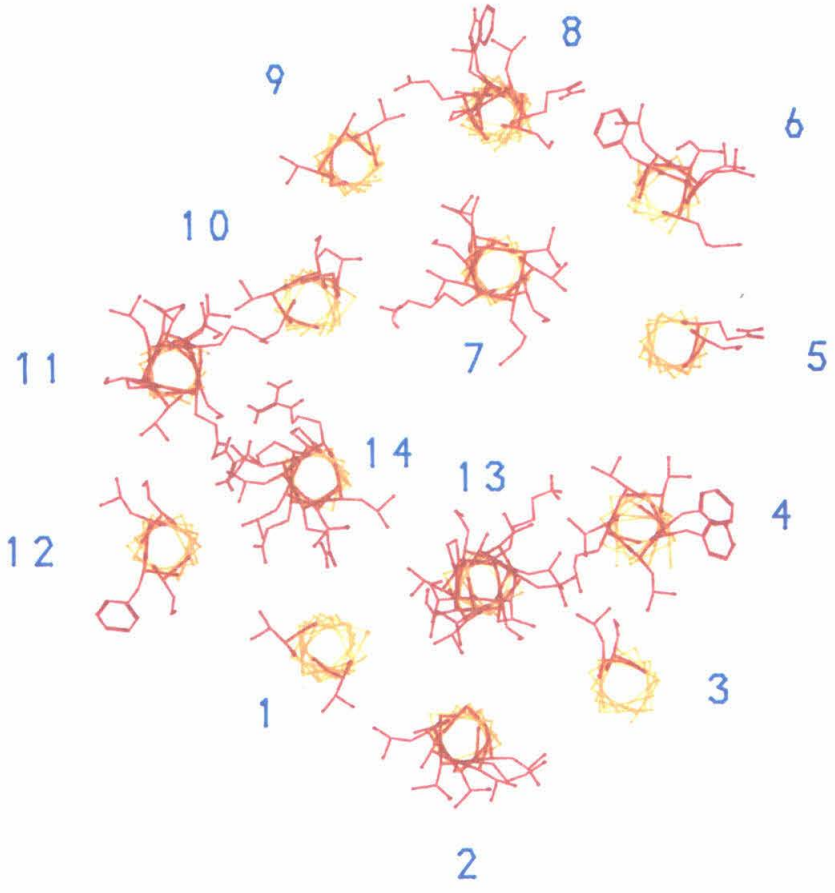
Figure 4.16

Variable Residues of Band 3

The side chains of band 3's variable residues are shown in red. The protein is viewed from the cytoplasm.

Variable Residues

Band 3



tains two cofacial, positive charges in the center if the bilayer. Finally, M13 has recently been shown to contain a binding site for pyridoxal phosphate, a transportable ion (102).

Other charged residues occur within the central region of the bilayer, although with the exception of the histidine residue on M9, all of them are negatively charged. Virtually every one of these acidic residues is glutamic acid, which has a higher pK_a than aspartic acid. These residues may, therefore, exist in their neutral, protonated form within the bilayer.

The charge distribution within the bilayer can be seen in figure 4.17. The current helix arrangement results in a "hydrophobic pocket" made up of helices M4 and M5. This uncharged region may account for the ability of band 3 to accomodate large, hydrophobic anions.

An examination of the interactions of helix dipoles, similar to that done in the last two sections, indicates that the proposed structure may be stabilized by about 30 kcal/mole. Every helix is stabilized to some extent by its surrounding helices.

Possible transport site residues. From various experiments (reviewed in 113), band 3 is believed to contain at least two types of high-affinity anion binding sites, referred to as transport sites. One of these would be accessible to the extracellular mileu and another to the cytoplasm. As schematically illustrated in figure 1.1, an anion must occupy one of these sites in order to effect the conformational change that allows it to be translocated. The translocation results in the anion's occupancy of the second transport site and subsequent release into the bulk medium. Before possible trans-

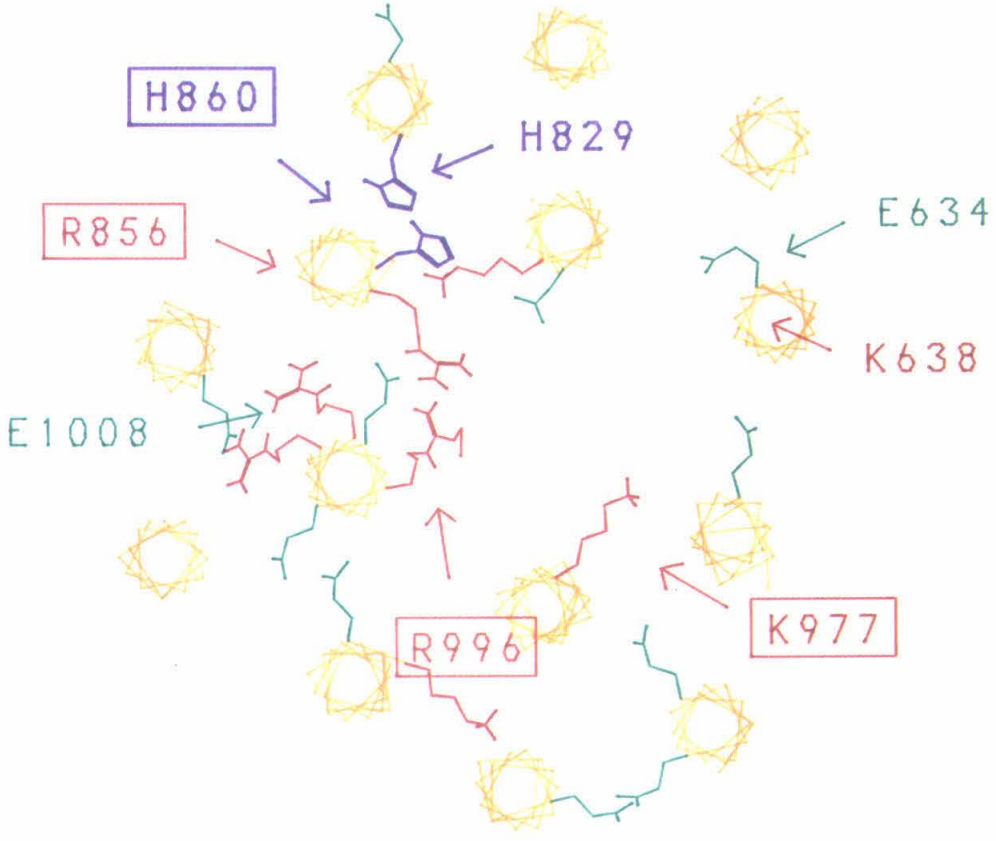
Figure 4.17

Charge Distribution in Band 3

The charged residues located within the membranous regions of band 3 have been indicated. Positive residues are red and negative ones are green in this figure. The locations of some specific residues, which are discussed in the text, have been indicated. Boxed residues are those that may play a significant role in the transport of anions. Although it was not drawn in this figure, the position of Lys 638 was indicated by an arrow. This lysine was not assigned to be part of the membranous region so the side chain itself was not included.

Charged Residues

Band 3



port mechanisms are discussed, likely transport sites in the model will be identified on the basis of current biochemical information. The locations of some of the residues mentioned below have been indicated on figures 4.17 and 4.18. Figure 4.18 shows the positions of charged residues on helices M7, M10, and M14. These helices contain most of the charged residues that are located within the proposed transport region.

More is currently known about the outward-facing transport site than the inner transport site. The outer site is believed to contain an arginine residue from the pH dependence of anion binding and transport, and there is some evidence that a second arginine may be nearby (103). Anion transport is regulated by a residue with a pK_a between 11 and 12 (103,104). This could only correspond to an arginine residue. External modification of the protein by phenylglyoxal, an arginine-selective reagent, causes complete inhibition of both anion transport and chloride binding to the high-affinity transport site (104-106). The modified arginine residue(s) responsible for transport inhibition can be protected from phenylglyoxal by the presence of chloride or DNDS. DNDS (4,4'-dinitrostilbene-2,2'-disulfonate) is an inhibitor of anion transport, which is also believed to interact with band 3 at this extracellular transport site (104,106).

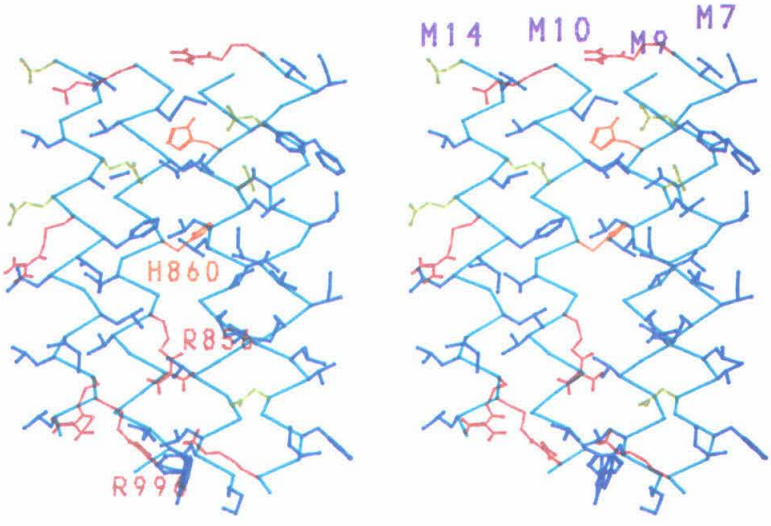
The arginine residues which are most likely to be involved in this high-affinity binding site are residues 856, which is on helix 10, and 996 on helix 14. Both of these two residues are conserved in all sequences, and both fall in the extracellular half of the bilayer, although residue 856 is more buried. Both residues may contribute to

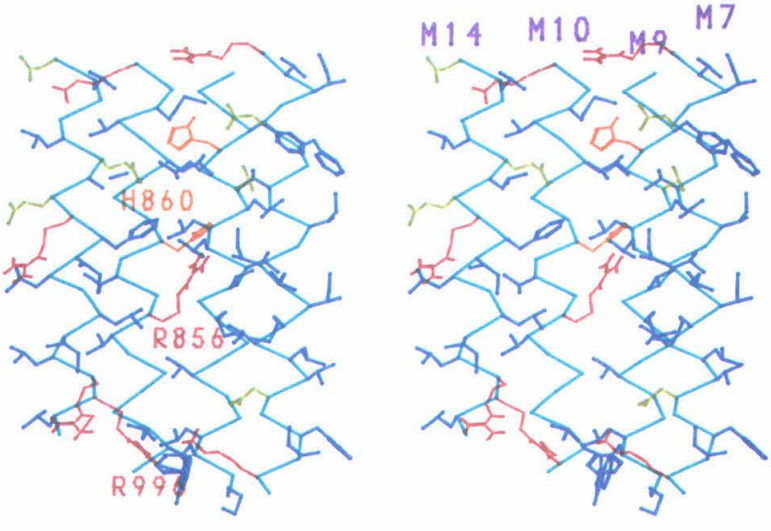
Figure 4.18

Charge Distribution in the Transport Region of Band 3

The charge distribution on helices M7, M9, M10, and M14 have been shown in these stereo drawings. The view is from the center of the helical cluster and is along the plane of the bilayer. The top of the figure corresponds to the cytoplasmic side of the membrane.

Positive residues have been drawn in red , negative ones in green, and histidine residues in orange. Hydrophobic residues (I, L, M, F, W, Y, and V) are shown in dark blue. The upper figure corresponds to the configuration proposed for the outward-facing transport site, and the lower to the inward-facing transport site (see text).


M14
M10
M9
M7
H860
R856
R996


M14
M10
M9
M7
H860
R856
R996

the affinity of the site for anions. The proposed orientation of DNDS when bound to band 3 is indicated in figure 4.19. If these two arginine residues define the binding site for stilbene disulfonates, the lysine residues modified and cross-linked by H_2DIDS might be residues 977 and 638 or 641. Experimental evidence has already implicated one of these latter two residues in covalent modification by H_2DIDS (98-100). The proposed orientation of H_2DIDS when covalently bound to band 3 is also indicated in figure 4.19. The residues covalently crosslinked by H_2DIDS are known to fall on opposite sides of the extracellular chymotrypsin cleavage site that occurs at position 652 (98,107). The lysine at position 977 is the same one identified by Kawano, *et al.* (102) to be the pyridoxal phosphate binding-site and is on the intracellular side of the membrane according to the current helix assignments. Because this residue is not conserved in all the band 3 proteins, it may not be essential to anion transport.

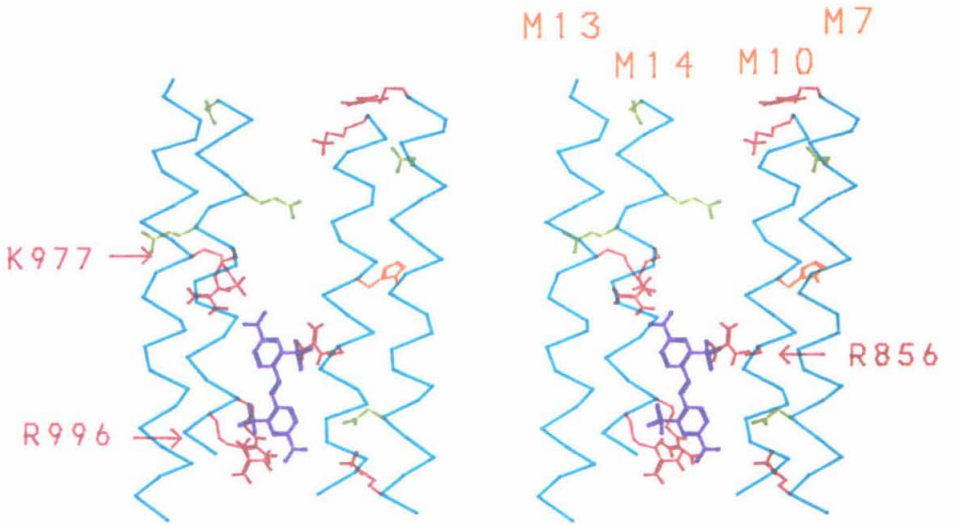
One residue which is likely to regulate ion transport is the histidine at position 860. It is located near the center of the bilayer, toward the cytoplasmic side, and is conserved in all sequences. Chloride self-exchange across the bilayer is inhibited by the protonation of a group accessible from the inside of the cell which has a pK_a of about 6.1 (108). The pK_a of histidine corresponds well to this value. In the transport of dianions, a "titratable carrier" model has been proposed by Gunn (109), which when combined with experimental data (110) also implicates the involvement of a histidine residue. This model requires the protonation of a residue on band 3 for dianion transport. The model has recently gained

Figure 4.19

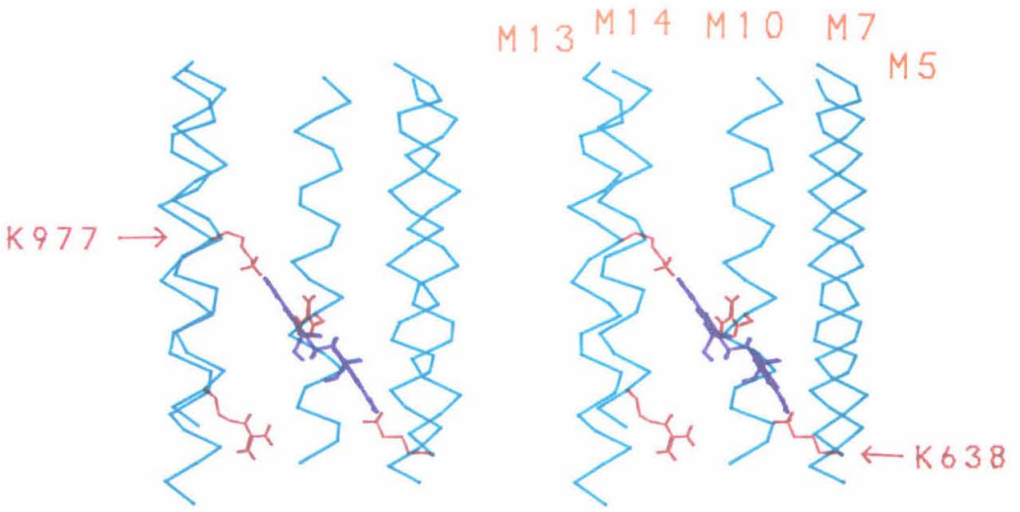
Stilbene Disulfonate Interactions with Band 3

The non-covalent interaction of stilbene disulfonates is represented by the positioning of DNDS in the transport region. The most likely orientation of DNDS with band 3 has been indicated in the upper drawing. The lower stereo drawing indicates a possible orientation for H_2DIDS, the stilbene disulfonate that covalently cross-links two lysine residues of band 3. The two most likely residues according to the current model have been indicated.

DNDS Orientation



H2DIDS Orientation



support from the study of "titratable anions," which are phosphate analogs which differ in the number and acidity of their titratable protons (111). The transport of these anions was found to depend upon the protonation of a residue on the protein that has a pK_a of about 5.5. This value probably also corresponds to a histidine residue, perhaps the same residue responsible for the deactivation of monovalent transport. (The differences in the pK_a values determined by the two groups may result from differences in the temperature or method of measurement.) It should be noted that the pH profiles of divalent anions also depend upon characteristics of the anion. As the pH is decreased, the protonation of the titratable residue on the protein will increase, favoring the transport of dianions, while the dianion may become protonated and have a low affinity for the now protonated form of the protein. The histidine at postition 809 may be the histidine responsible for one or both of these characteristics of band 3 transport, but because it has a peripheral location in this model, residue 860 seems more probable. In addition, His 860 is nearest to Lys 977, the point of attachment for pyridoxal phosphate (102). Pyridoxal phosphate is also believed to interact with a histidine (115) and, when positioned on the proposed transport region, can only interact with His 860. This positioning is indicated in figure 4.20.

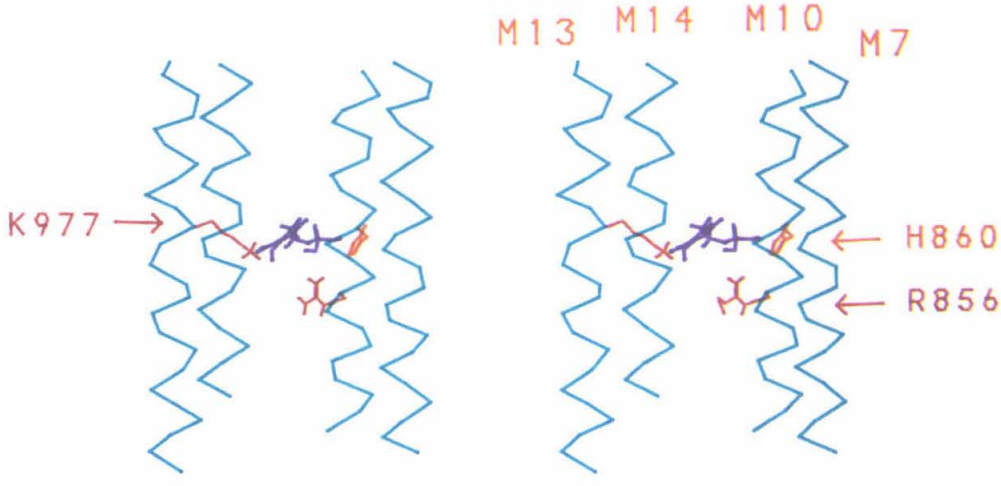
The only other residue type that has been implicated to partake in anion transport is glutamic acid. Recent evidence has indicated that two glutamate (and no aspartate) residues occur near the stilbenedisulfonate binding site (112). The two most likely residues correspond

Figure 4.20

Orientation of Pyridoxal Phosphate within Band 3

Pyridoxal phosphate has been oriented within the proposed transport region so that its phosphate group can interact with both His 860 and Arg 856. This placement also allows its aldehyde moiety to covalently modify Lys 977.

Pyridoxal Phosphate Binding



to positions 634 and 1008, which are on helices 5 and 14, respectively. Although these residues are on different sides of the extracellular chymotriptic cleavage site at position 652 in agreement with experimental evidence, residue 1008 is quite near the cytoplasm and is unlikely to be protected by DNDS. The current helical arrangement also does not account for the observation that Asp 737 is not modified, even though it is quite near the predicted stilbene binding site.

Possible transport mechanism. Several mechanisms for band 3 mediated anion transport have been proposed (104,107,113). The critical part of these transport models is the method of ion gating. One of these models, termed the "swinging arm" model (107) is depicted schematically in figure 4.21. This mechanism requires n positive residues and $n-1$ negative residues within the anion pathway. Each negative side chain must be able to interact with its two nearest positively charged neighbors. The anion is transported from the side of the membrane with the free positive side chain. As the ion travels to the other side of the membrane it is carried by each successive positive chain. When the anion reaches its destination and dissociates, the final positive residue remains unpaired and waits for the binding of another ion. One problem with this model is that there seems to be no large energy barrier to the reorientation of side chains from one conformation to another in the absence of anions. Barriers to the movement of charged, *i.e.*, not ion-paired, species at any point may prevent this translocation of the empty carrier. No simple modification of this model can explain the transport of divalent anions,

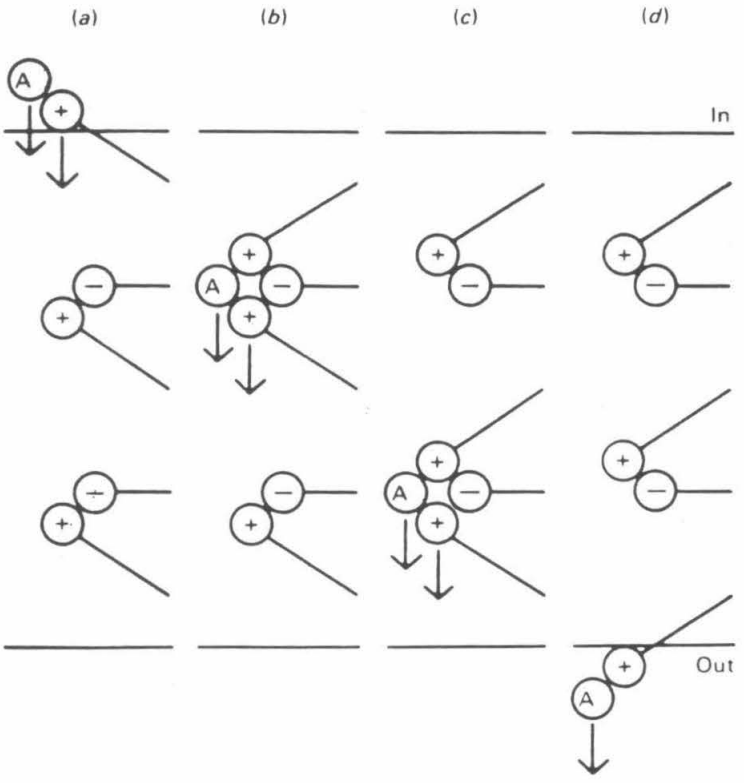
Figure 4.21

Schematic Diagram of the Swinging Arm Transport Mechanism

The steps required to transport an anion from one side of the membrane to the other are depicted. The anion (A) from one side of the membrane is passed from side chain to side chain until its release on the opposite side of the membrane. The side chains are then positioned so that an anion must be transported in the opposite direction.

This figure was reproduced from reference 107.

Swinging Arm Model



however.

An examination of the charge distribution for the proposed transport region, shown in figure 4.18, shows that the swinging arm model is not probable. The charges cannot interact in the manner described above, especially on the cytoplasmic side of the membrane.

Another mechanism proposed for band 3 mediated transport originated in this laboratory and has been termed the "hydrophobic barrier" model (104). In this model, channel helices surround a peptide chain which contains a bulky hydrophobic region near the charged transport site. When an anion binds to this positively charged transport site from the accessible compartment, the charge is neutralized and the hydrophobic barrier may slide past. The anion may then be released on the opposite side of the membrane. This mechanism is illustrated schematically in figure 4.22. This model could account for the transport of dianions if a histidine was also located near the positive transport residue. The presence of a centrally located peptide chain may not permit the transport of larger anions, however.

Although the current helical arrangement does not have a centrally located peptide chain, certain aspects of the hydrophobic barrier suggest a mechanism of anion transport which could be operable. A careful examination of figure 4.18 shows a hydrophobic turn on helix 14 at the level of Arg 856. The hydrophobic nature of this region is conserved for all currently known band 3 proteins. The residues on M7 are also hydrophobic in this region. Although it is possible that M14 travels in a direction perpendicular to the plane of the membrane in order to position its hydrophobic barrier on alternating sides of the

Figure 4.22

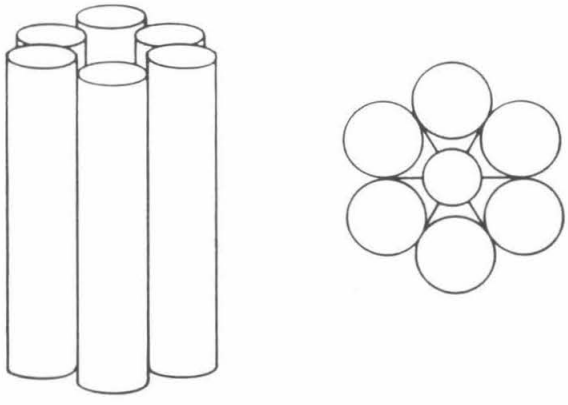
Schematic Diagram of the Hydrophobic Barrier Model

The helical arrangement and transport mechanism proposed in this model are indicated. The hydrophobic barrier is located on the central helix and is denoted by a small cylinder on the peptide chain in the lower diagrams. Anions are indicated by filled circles.

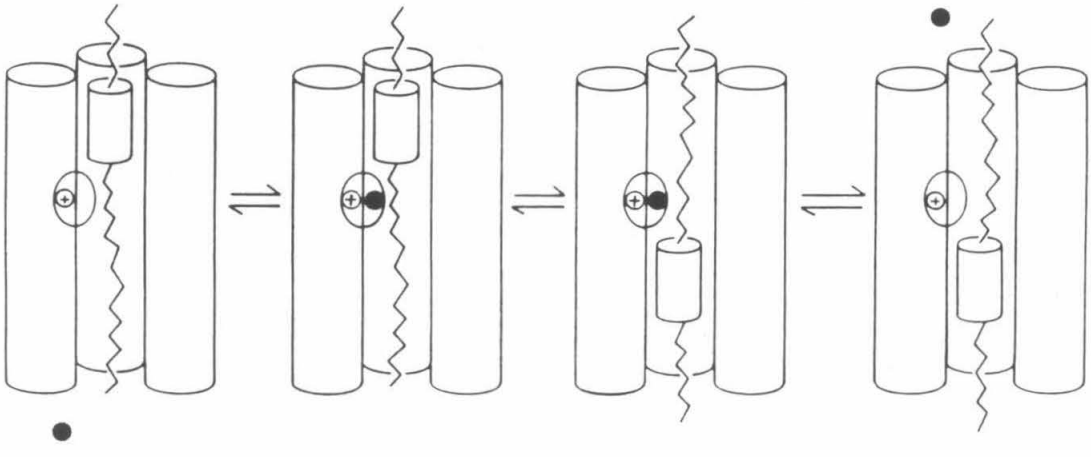
This diagram was reproduced from reference 104.

Hydrophobic Barrier Model

Helical Arrangement



Transport Mechanism



arginine residue, it seems more probable that the arginine side chain, neutralized by the anion, merely rotates across the barrier. These two positions of the arginine side chain were indicated in figure 4.18. When the anion is released, the arginine side chain is stuck on that side of the barrier until another anion is bound and translocated. The transport of dianions would be permitted with the protonation of His 860. The anion would then bind to both side chains in order for transport to occur. Eosin 5-isothiocyanate, a competitive inhibitor of dianion transport, shows interactions with both a histidine residue and an arginine when covalently bound to band 3 (114), also suggesting that the interaction between these two residue types may occur during dianion transport. Because the proposed mechanism incorporates certain aspects of the two previously proposed mechanisms, it might be termed the "stationary barrier/swinging arm" model. The mechanism is shown schematically in figure 4.23.

Summary

Although there are some uncertainties in the proposed model, it could account for many of the general features of anion transport by band 3. In particular, the pH dependance of monovalent and dianion transport could be explained. The arginine and histidine residues on helix 10 are probably the residues essential for anion transport and the arginine residue on helix 14 probably makes up a part of the stilbene disulfonate binding site.

Figure 4.23

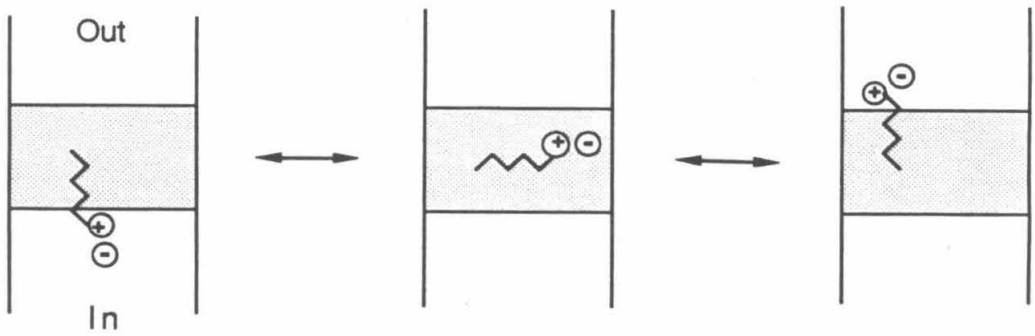
Schematic Diagram of the

Stationary Barrier/Swinging Arm Mechanism

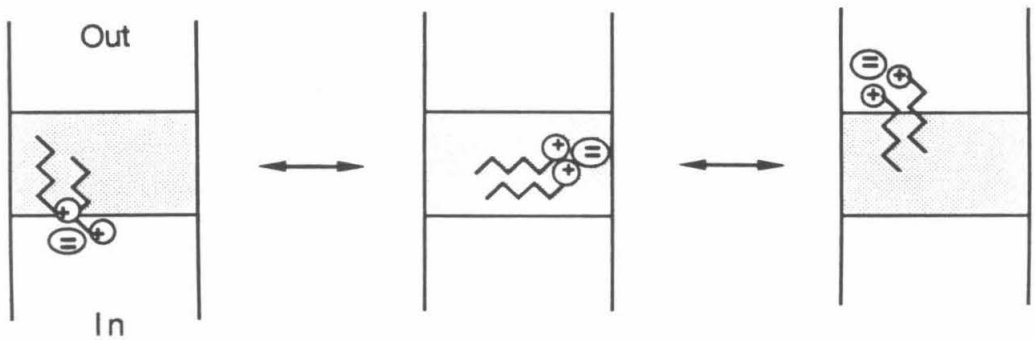
The mechanism proposed in this work is illustrated schematically in this figure. The stationary hydrophobic barrier is represented by a shaded horizontal band. Mechanisms for the transport of both mono- and divalent anions have been drawn.

Stationary Barrier / Swinging Arm Mechanism

Monovalent Anion Transport



Divalent Anion Transport



References

1. Crick, F.H.C.(1953) *Acta Crystallog.* **6**, 689-697.
2. Cohen, G., and Parry, D.A.D.(1986) *Trends Biochem. Sci.* **11**, 245-248.
3. Weber, P.C., and Salemme, F.R.(1980) *Nature* **287**, 82-84.
4. Richardson, J.S.(1981) *Adv. Prot. Chem.* **34**, 167-339.
5. Henderson, R., and Unwin, P.N.T.(1975) *Nature* **257**, 28-32.
6. Corless, J.M., McCastin, D.R., and Scott, B.L.(1982) *Proc. Natl. Acad. Sci. USA* **79**, 116-120.
7. Diesenhofer, J., Epp, O., Miki, K., Huber, K., and Michel, H.(1985) *Nature* **318**, 618-624.
8. Allen, J.P., Feher, G., Yeates, T.O., Komiya, H., and Rees, D.C.(1987) *Proc. Natl. Acad. Sci. USA* **84**, 6162-6166.
9. Unwin, P.N.T., and Zampighi, G.(1980) *Nature* **283**, 545-549.
10. Brisson, A., and Unwin, P.N.T.(1985) *Nature* **315**, 474-477.
11. Wada, A.(1976) *Adv. Biophys.* **9**, 1-63.
12. Hol, W.G.J., van Duijnen, P.Th., and Derendsen, H.J.C.(1978) *Nature* **273**, 443-446.
13. Sheridan, R.P., and Allen, L.C.(1980) *Biophys. Chem.* **11**, 133-136.
14. Sheridan, R.P., Levy, R.M., and Salemme, F.R.(1982) *Proc. Natl. Acad. Sci. USA* **79**, 4545-4549.
15. Yeates, T.O., Komiya, H., Rees, D.C., Allen, J.P., and Feher, G.(1987) *Proc. Natl. Acad. Sci. USA* **84**, 6438-6442.
16. Engelman, D.M., Henderson, R., McLachlan, A.D., and Wallace, B.A.(1980) *Proc. Natl. Acad. Sci.* **77**, 2023-2027.

17. Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Tanabe, T., Shimizu, S., Kikotani, S., Kayano, T., Hirose, T., Inayama, S., and Numa, S. (1983) *Nature* **305**, 818-823.

18. Tanabe, T., Noda, M., Furutani, Y., Takai, T., Takahashi, H., Tanaka, K., Hirose, T., Inayama, S., and Numa, S.(1984) *Eur. J. Biochem.* **144**, 11-17.

19. Takai, T., Noda, M., Furutani, Y., Takahashi, H., Notake, M., Shimizu, S., Kayano, T., Tanabe, T., Tanaka, K., Hirose, T., Inayama, S., and Numa, S.(1984) *Eur. J. Biochem.* **143**, 109-115.

20. Kubo, S., Noda, M., Takai, T., Tanabe, T., Kayano, T., Shimizu, S., Tanake, K., Takahashi, H., Hirose, T., Inayama, S., Kikuno, R., Miyata, T., and Numa, S.(1985) *Eur. J. Biochem.* **149**, 5-13.

21. Takai, T., Noda, M., Mishina, M., Shimizu, S., Furutani, Y., Kayano, T., Ikeda, T., Kubo, T., Takahashi, H., Takahashi, T., Kuno, M., and Numa, S.(1985) *Nature* **315**, 761-764.

22. Boulter, J., Luyten, W., Evans, K., Mason, P., Ballivet, M., Goldman, D., Stengelin, S., Martin, G., Heinemann, S., and Patrick, J.(1985) *J. Neurosci.* **5**, 2545-2552.

23. Buonanno, A., Mudd, J., Shah, V., and Merlie, J.P.(1986) *J. Biol. Chem.* **261**, 16451-16458.

24. Boulter, J., Evans, K., Martin, G., Mason, P., Stengelin, S., Goldman, D., Heinemann, S., and Patrick, J.(1986) *J. Neurosci. Res.* **16**, 37-47.

25. LaPolla, R.J., Mayne, K.M., and Davidson, N.(1984) *Proc. Natl. Acad. Sci. USA* **81**, 7970-7974.

26. Noda, M., Takahashi, H., Tsutomu, T., Toyosato, M., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H., Imayama, S., Miyata, T., and Numa, S.(1983) *Nature* **302**, 528-532.

27. Giraudat, J., Montecucco, C., Bisson, R., and Changeux, J.-P.(1985) *Biochemistry* **24**, 3121-3127.

28. Hucho, F., Oberthür, W., and Lottspeich, F.(1986) *FEBS Lett.* **205**, 137-142.

29. Giraudet, J., Dennis, M., Heidmann, T., Haumont, P.-Y., Lederer, F., and Changeux, J.-P.(1987) *Biochemistry* **26**, 2410-2418.

30. Unwin, N.(1986) *Nature* **323**, 12-13.

31. Unwin, N.(1987) *Chem. Scr.* **27B**, 47-51.

32. Kubalek, E., Ralston, S., Lindstrom, J., and Unwin, N.(1987) *J. Cell Biol.* **105**, 9-18.

33. Leonard, R.J., Labarca, C.G., Charnet, P., Davidson, N., and Lester, H.A. (1988) *Science*, in press.

34. White, M.M., Mayne, K.M., Lester, H.A., and Davidson, N.(1985) *Proc. Natl. Acad. Sci. USA* **82**, 4852-4856.

35. Sakmann, B., Methfessel, C., Mishina, M., Takahashi, T., Takai, T., Kurasaki, M., Fukuda, K., and Numa, S.(1985) *Nature* **318**, 538-543.

36. Greenblatt, R.E., Blatt, Y., and Montal, M.(1985) *FEBS Lett.* **193**, 125-134.

37. Dwyer, T.M., Adams, D.J., and Hille, B.(1980) *J. Gen. Physiol.* **75**, 469-492.

38. Mishina, M., Tobimatsu, T., Imoto, K., Tanaka, K., Fujita, Y., Fukuda, K., Kurasaki, M., Takahashi, H., Morimoto, Y., Hirose, T., Inayama, S., Takahashi, T., Kuno, M., and Numa, S.(1985) *Nature* **313**, 364-369.

39. Mulac-Jericevic, B., and Atassi, M.Z.(1987) *Biochem. J.* **248**, 847-852.

40. Dennis, M., Giraudat, J., Kotzyba-Hilbert, F., Goeldner, M., Hirth, C., Chang, J.-Y., Lazure, C., Chrétien, M., and Changeux, J.-P.(1988) *Biochemistry* **27**, 2346-2357.

41. Witzemann, V., and Raftery, M.(1978) *Biochemistry* **17**, 3598-3604.

42. Unwin, P.N.T., and Ennis, P.D.(1984) *Nature* **307**, 609-613.

43. Kumar, N.M., and Gilula, N.B.(1986) *J. Cell Biol.* **103**, 767-776.

44. Girsch, S.J., and Peracchia, C.(1985) *J. Membr. Biol.* **83**, 217-225.

45. Greenblatt, R.E., Blatt, Y., and Montal, M.(1985) *FEBS Lett.* **193**, 125-134.

46. Tanabe, T., Takeshima, H., Mikami, A., Flockerzi, V., Takahashi, H., Kangawa, K., Kojima, M., Matsuo, H., Hirose, T., and Numa, S.(1987) *Nature* **328**, 313-318.

47. Guy, R.H., and Raghunathan, G.(1988) *Biophys. J.* **53**, 639a.

48. Mamalaki, C., Stephenson, F.A., and Barnard, E.A.(1987) *EMBO J.* **6**, 561-565.

49. Colquhoun, D., and Sakmann, B. (1981) *Nature* **294**, 464-466.

50. Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Tanabe, T., Shimizu, S., Kikyotani, S., Kayano, T., Hirose, T., Inayama, S., and Numa, S. (1983) *Nature* **305**, 818-823.

51. Tanabe, T., Noda, M., Furutani, Y., Takai, T., Takahashi, H., Tanaka, K., Hirose, T., Inayama, S., and Numa, S.(1984) *Eur. J. Biochem.* **144**, 11-17.

52. Takai, T., Noda, M., Furutani, Y., Takahashi, H., Notake, M., Shimizu, S., Kayano, T., Tanabe, T., Tanaka, K., Hirose, T., Inayama, S., and Numa, S.(1984) *Eur. J. Biochem.* **143**, 109-115.

53. Kubo, S., Noda, M., Takai, T., Tanabe, T., Kayano, T., Shimizu, S., Tanaka, K., Takahashi, H., Hirose, T., Inayama, S., Kikuno, R., Miyata, T., and Numa, S.(1985) *Eur. J. Biochem.* **146**, 15-22.

54. Nef, P., Mauron, A., Stalder, R., Alliod, C., and Ballivet, M.(1984) *Proc. Natl. Acad. Sci. USA* **81**, 7975-7979.

55. Hermans-Borgmeyer, I., Zopf, D., Ryseck, R.-P., Hovemann, B., Betz, H., and Gundelfinger, E.D.(1986) *EMBO J.* **5**, 1503-1508.

56. Shibara, S., Kubo, T., Perski, H.J., Takahashi, H., Noda, M., and Numa, S. (1985) *Eur. J. Biochem.* **146**, 15-22.

57. Buonanno, A., Mudd, J., Shah, V., and Merlie, J.P.(1986) *J. Biol. Chem.* **261**, 16451-16458.

58. LaPolla, R.J., Mayne, K.M., and Davidson, N.(1984) *Proc. Natl. Acad. Sci. USA* **81**, 7970-7974.

59. Boulter, J., Evans, K., Goldman, D., Martin, G., and Treco, D.(1986) *Nature* **319**, 368-374.

60. Noda, M., Takahashi, T., Tsutomu, T., Toyosato, M., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H., Imayama, S., Miyata, T., and Numa, S.(1983) *Nature* **302**, 528-532.

61. Schofield, P.R., Darlison, M.G., Fujita, N., Burt, D.R., Stephenson, F.A., Rodriguez, H., Rhee, L.M., Ramachandran, J., Reale, V., Glencorse, T.A., Seeburg, P.H., and Barnard, E.A.(1987) *Nature* **328**, 221-227.

62. Levitan, E.S., Schofield, P.R., Burt, D.R., Rhee, L.M., Wisden, W., Köhler, M., Fujita, N., Rodriguez, H.F., Stephenson, A., Darlison, M.G., Barnard, E.A., and Seeburg, P.H.(1988) *Nature* **335**, 76-79.

63. Corless, J.M., McCastin, D.R., and Scott, B.L.(1982) *Proc. Natl. Acad. Sci. USA* **79**, 116-120.

64. Findlay, J.B.C.(1986) *Photochem. Photobiophys.* **13**, 213-228.

65. Engelman, D.M., Henderson, R., McLachlan, A.D., and Wallace, B.A.(1980) *Proc. Natl. Acad. Sci. USA* **77**, 2023-2027.

66. Trewhella, J., Popot, J.-L., Zacai, G., and Engelman, D.M.(1986) *EMBO J.* **5**, 3045-3049.

67. Davison, M.D., and Findlay, J.B.C.(1986) *Biochem. J.* **234**, 413-420.

68. Davison, M.D., and Findlay, J.B.C.(1986) *Biochem. J.* **236**, 389-395.

69. Al-Saleh, S., Gore, M., and Akhtar, M.(1987) *Biochem. J.* **246**, 131-137.

70. Dixon, R.A.F., Sigal, I.S., Candelore, M.R., Register, R.B., Scattergood, W., Rands, E., and Strader, C.D.(1987) *EMBO J.* **6**, 3269-3275.

71. Wessling-Resnick, M., and Johnson, G.L.(1987) *Biochemistry* **26**, 4316-4326.

72. Fraser, C.M., Kerlavage, A.R., Mariani, A.P., and Venter, J.C.(1987) *Proteins* **2**, 34-41.

73. Wang, J.K.,McDowell, J.H., and Hargrave, P.A.(1980) *Biochemistry* **19**, 5111-5117.

74. Kakitani, H., Kakitani, T., Rodman, H., and Honig, B.(1985) *Photochem. Photobiol.* **41**, 471-479.

75. Kosower, E.M.(1988) *Proc. Natl. Acad. Sci. USA* **85**, 1076-1081.

76. Weiss, E., Kelleher, D.J., and Johnson, G.L.(1988) *J. Biol. Chem.* **263**, 6150-6154.

77. Strader, C.D., Dixon, R.A.F., Cheung, A.H., Candelore, M.R., Blake, A.D., and Sigal, I.S.(1987) *J. Biol. Chem.* **262**, 16439-16443.

78. Franke, R.R., Sakmar, T.P., Oprian, D.D., and Khorana, H.G.(1988) *J. Biol. Chem.* **263**, 2119-2122.

79. Findlay, J.B.C., Brett, M., and Pappin, D.J.C.(1981) *Nature* **293**, 314-316.

80. Sitaramayya, A.(1986, publ. 1987) *Membrane Proteins: Proceedings of the Membrane Protein Symposium*, Goheen, S.C., ed., Biorad Laboratories: San Diego, CA., 601-610.

81. Dohlman, H.G., Bouvier, M., Benovic, J.L., Caron, M.G., and Lefkowitz, R.J.(1987) *J. Biol. Chem.* **262**, 14282-14288.

82. Benovic, J.L., Staniszewski, C., Mayor, F., Caron, M.G., and Lefkowitz, R.J.(1988) *J. Biol. Chem.* **263**, 3893-3897.

83. Wilden, U., Hall, S.W., and Kühn, H.(1986) *Proc. Natl. Acad. Sci. USA* **83**, 1174-1178.

84. Sullivan, K.A., Miller, R.T., Masters, S., Beiderman, B., Heideman, W., and Bourne, H.R.(1987) *Nature* **330**, 758-760.

85. Wessling-Resnick, M., Kelleher, D.J., Weiss, E.R., and Johnson, G.L.(1987) *Trends. Biochem. Sci.* **12**, 473-477.

86. Strader, C.D., Sigal, I.S., Register, R.B., Candelore, M.R., Rands, E., and Dixon, R.A.F.(1987) *Proc. Natl. Acad. Sci. USA* **84**, 4384-4388.

87. Chung, F.-Z., Wang, C.-D., Potter, P.C., Venter, J.C., and Fraser, C.M. (1988) *J. Biol. Chem.* **263**, 4052-4055.

88. Dohlman, H.G., Caron, M.G., Strader, C.D., Amlaiky, N., and Lefkowitz, R.J. (1988) *Biochemistry* **27**, 1813-1817.

89. Wheatley, M., Hulme, E.C., Birdsall, N.J.M., Curtis, C.A.M., Eveleigh, P., Pedder, E.K., and Poyner, D.(1988) *Trends Pharm. Sci. Suppl.: Subtypes of Muscarinic Receptors III*, 19-24.

90. Rao, A.(1979) *J. Biol. Chem.* **254**, 3503-3511.

91. Reithmeier, R.A.F., and Rao, A.(1979) *J. Biol. Chem.* **254**, 6151-6155.

92. Macara, I.G., and Cantley, L.C.(1981) *Biochemistry* **20**, 5095-5105.

93. Benz, R., Tosteson, M.T., and Schubert, D.(1984) *Biochim. Biophys. Acta* **775**, 347-355.

94. Jennings, M.L., and Nicknish, J.S.(1985) *J. Biol. Chem.* **260**, 5472-5479.

95. Falke, J.F., Pace, R.J., and Chan, S.I. (1984) *J. Biol. Chem.* **259**, 6472-6480.

96. Cabantchik, Z.I., and Rothstein, A.(1974) *J. Membr. Biol.* **15**, 207-226.

97. Lepke, S., Fasold, H., Pring, M., and Passow, H.(1976) *J. Membr. Biol.* **29**, 147-177.

98. Jennings, M.L., and Passow, H.(1979) *Biochim. Biophys. Acta* **554**, 498-519.

99. Grinstein, S., Ship, S., and Rothstein, A.(1978) *Biochim. Biophys. Acta* **507**, 294-304.

100. Jay, D., and Cantley, L.(1986) *Ann. Rev. Biochem.* **55**, 511-538.

101. Bjerrum, P.J.(1983) *Structure and Function of Membrane Proteins*, Quagliariello, E., and Palmieri, F., eds., Elsevier: Amsterdam, 107-115.

102. Kawano, Y., Okubo, K., Tokunaga, F., Miyata, T., Iwanaga, S., and Hamasaki, N.(1988) *J. Biol. Chem.* **263**, 8232-8238.

103. Wieth, J.O., and Bjerrum, P.J.(1982) *J. Gen. Physiol.* **79**, 253-282.

104. Falke, J.J.(1985) *Doctoral Thesis* (California Institute of Technology).

105. Zaki, L.(1984) *FEBS Lett.* **169**, 234-240.

106. Zaki, L., and Julien, T.(1985) *Biochim. Biophys. Acta* **818**, 325-332.

107. Brock, C.J., Tanner, M.J.A., and Kempf, C.(1983) *Biochem. J.* **261**, 577-586.

108. Wieth, J.O., Brahm, J., and Funder, J.(1980) *Ann. N.Y. Acad. Sci.* **341**, 394-418.

109. Gunn, R.B.(1973) in *Erythrocytes, Throbocytes, Leukocytes*, Gerlach, E., Moser, K., Deutsch, E., and Wilmanns, W., eds. (Thieme, Stuttgart), 77-79.

110. Milanick, M.A., and Gunn, R.B.(1984) *Am. J. Physiol.* **247**, C247-C259.

111. Labotka, R.J., and Omachi, A.(1988) *J. Biol. Chem.* **263**, 1166-1173.

112. Jennings, M.L., and Anderson, M.P.(1987) *J. Biol. Chem.* **262**, 1691-1697.

113. Passow, H.(1986) *Rev. Physiol. Biochem. Pharmacol.* **103**, 61-223.

114. Chiba, T., Sato, Y., Suzuki, Y.(1986) *Biochim. Biophys. Acta* **858**, 107-117.

115. Chiba, T., Sato, Y., Suzuki, Y.(1987) *Biochim. Biophys. Acta* **897**, 14-24.

116. Hamasaki, N., Kawano, Y., and Inoue, H.(1987) *Biomed. Biochim. Acta* **46**, S51-S54.

CHAPTER 5

CONCLUSIONS AND A LOOK INTO THE FUTURE

Conclusions

In this work, a criterion was developed for identifying the transmembrane regions of integral membrane proteins. Assignments based on this criterion fit all experimental data, except in the case of band 3. Topological information from experimental evidence had to be used during the assignment procedure for that protein, probably the result of some helices being completely buried. For this reason, it was suggested that such topological information be incorporated into the helix assignment procedure whenever six or more helices are predicted using hydropathic analyses.

The use of residue variability to arrange the transmembrane helices was unambiguous only in the case of the acetylcholine receptor. As discussed in the last chapter, the complications arising in the case of rhodopsin may be due to the nature of its prosthetic group, retinal, and the nature of its oligomeric structure. The identification of variable residues in band 3 is uncertain due to the limited number of sequences available for analysis. The helices positioned in the center of the helical cluster were also quite variable, in contrast to the theory that buried helices will be more conserved. Structural deviations from the α-helical structure used in the modelling or a loose association with the lipid-exposed helices may be responsible. As more sequences and high resolution structures of membrane proteins become available, the use of variability in the

arrangement of transmembrane helices may be better assessed.

The models developed for the proteins were all successful to some degree. At the very least, specific residues in the primary sequence were identified which may play a role in the protein function. In the case of the acetylcholine receptor, specific residues that may make up the cation channel have been identified and a novel method for channel gating has been proposed. The three-dimensional modelling of rhodopsin has suggested which residues may be responsible for the regulation of wavelength sensitivity in the different family members. In addition, a specific conformational change of the receptor has been proposed which could effect the activation of rhodopsin's G-protein, transducin. The same conformational change might also occur in the muscarinic and β-adrenergic receptors to activate their G-proteins. The model of band 3 could account for some of the observed characteristics of the protein, but not others. It did, however, suggest specific residues which may be involved in the transport of anions. The best mechanism that could be proposed involved changes in side chain orientations, instead of helical motions.

Future Work on Membrane Proteins

Membrane proteins represent a large class of proteins with properties which are likely to be very much different than soluble proteins. At the moment, very little is known about their structures, especially if they have more than a few transmembrane helices. High-resolution structures for more proteins will be needed in order to gain insight into helix packing motifs, the characteristics of helix-helix contact

regions, the nature of subunit-subunit interactions, and the importance of various types of interactions in protein stabilization. The possible occurence within the membrane of peptide structures other than a-helices has not been considered in this work, although their existence within a circle of lipid-adjacent α-helices is not inconceivable. Helices, or helical pairs that do not traverse the complete length of the bilayer have also not been permitted in this modelling, although they may actually occur in membrane proteins. Knowledge about the mechanisms of protein insertion into the membrane may also provide insight into which regions of the protein sequence are membrane imbedded in the final structure.

It is hoped that advances in the techniques used to study these important proteins will soon permit detailed analyses of their structures and provide insights into the molecular aspects of protein function.