Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 2067–2071, April 1983
Neurobiology

Complete mRNA coding sequence of the acetylcholine binding α-subunit of *Torpedo marmorata* acetylcholine receptor: A model for the transmembrane organization of the polypeptide chain

(integral membrane protein/recombinant DNA/protein evolution)

ANNE DEVILLERS-THIERY, JÉRÔME GIRAUDAT, MARTINE BENTABOULET, AND JEAN-PIERRE CHANGEUX

Neurobiologie Moléculaire, L.A. N° 270 Centre National de la Recherche Scientifique, Interactions Moléculaires et Cellulaires, Institut Pasteur, 25, rue du Docteur Roux, 75724 Paris Cédex 15, France

Communicated by François Jacob, December 17, 1982

ABSTRACT **A 1,350-base-pair-long cDNA clone, named pα-2, was isolated by hybridization to the previously characterized clone pα-1 and found to be specific for the α-subunit of the *Torpedo marmorata* acetylcholine receptor. The nucleotide sequences of both cDNA inserts were analyzed and the sequence of the complete coding region and part of the 5′ and 3′ untranslated regions of the α-chain mRNA was determined. The complete amino acid sequence of the α-chain precursor is presented and used to develop a model for the transmembrane organization of the polypeptide.**

At the vertebrate neuromuscular junction and in the electric organ of the fish *Torpedo* and *Electrophorus*, the nicotinic acetylcholine receptor protein (AcChoR) plays a central role in synaptic transmission because it mediates the opening of the post-synaptic ion channel by acetylcholine (AcCho) (1). Reconstitution experiments (2–5) have shown that the minimal macromolecular unit required for this important regulatory interaction is a transmembrane oligomeric protein made up of four different polypeptide chains (apparent molecular masses: α, 40; β, 50; γ, 60; δ, 66 kilodaltons) present in the AcChoR light form with the stoichiometry $\alpha_2\beta\gamma\delta$ (6–9). Of these subunits, the α-chain has the most well-defined function. It carries at least part of the binding site for the neurotransmitter (7) and thus is involved in the opening of the ion channel by AcCho. In addition, it is labeled by noncompetitive agents that block the permeability response (10) and thus might contribute, together with other subunits (in particular the δ chain), to a single high-affinity binding site for these ligands. This site might plausibly be located in a central hydrophilic pit common to all five subunits (11). An attractive and probable hypothesis is that this central hydrophilic channel, also observed by electron microscopy of negatively stained preparations of the AcChoR molecule (9, 12), is the AcCho-regulated "ionophore" (6).

The α-subunit has thus been the subject of extensive structural investigations. Its NH_2-terminal amino acid sequence has been determined (8, 13), its transmembrane character has been demonstrated by selective proteolysis (14), and its cDNA has been cloned (15–17) concurrently with that of the γ-subunit (18).

In this paper, we describe the selection of a clone specific for the α-chain that cross-hybridizes with the previously isolated pα-1 clone (15). From the nucleotide sequences of these two clones, the entire mRNA coding sequence of the European *Torpedo marmorata* has been determined. The corresponding primary structure of the α-subunit allows prediction of some important features of the transmembrane organization of the polypeptide chain.

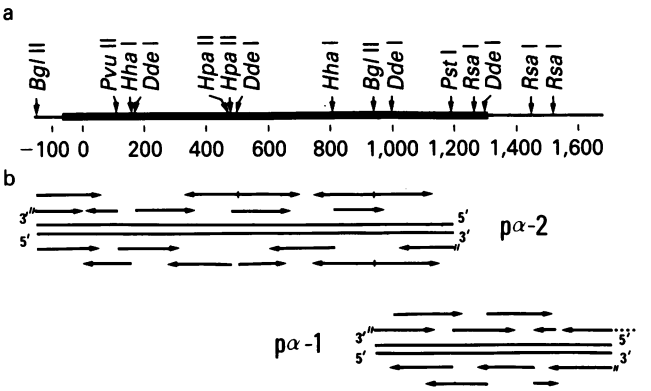

FIG. 1. Strategy used for sequence analysis of the cDNA inserts in clones pα-1 and pα-2. (*a*) Relative positions of restriction endonuclease sites used to generate fragments for analysis. The coding sequence is indicated by the thicker line. Nucleotide numbers are as in Fig. 2. (*b*) Horizontal arrows indicate extent and direction of sequence determinations. For each of the two clones, the upper arrows correspond to analysis of the noncoding strand and the lower arrows correspond to analysis of the coding strand. Slashes indicate fragments labeled at the *Pst* I site flanking the cDNA insert. The dotted line corresponds to a fragment analyzed from the *Hpa* II site located on pBR322 DNA.

MATERIALS AND METHODS

Construction and Identification of Plasmid pα-2. A cDNA library from adult *T. marmorata* electric organ was constructed as described in ref. 15 except that avian myeloblastosis virus reverse transcriptase was used for synthesis of both cDNA strands. Tetracycline-resistant and ampicillin-sensitive clones were screened by *in situ* hybridization (19) with the pα-1 DNA insert, ^{32}P-labeled by nick-translation (20) (2×10^7 cpm/μg).

Nucleotide Sequence Analysis. The chemical method of Maxam and Gilbert (21) was used for DNA sequence analysis. We labeled 3′ ends of *Pst* I and *Hha* I restriction fragments with terminal deoxynucleotidyltransferase and [α-^{32}P]ATP; 5′ ends of *Bgl* II, *Dde* I, *Hpa* II, *Pvu* II, and *Rsa* I fragments with T4 polynucleotide kinase by the phosphate exchange reaction; and 3′ ends of *Bgl* II, *Dde* I, and *Hpa* II fragments with the Klenow fragment of *Escherichia coli* DNA polymerase (22).

Materials. *T. marmorata* were supplied by the Marine Biological Station of Arcachon, France. T4 polynucleotide kinase

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: AcCho, acetylcholine; AcChoR, nicotinic acetylcholine receptor; bp, base pair(s).

was purchased from P-L Biochemicals, and terminal deoxynucleotidyltransferase and the Klenow fragment of *E. coli* DNA polymerase were obtained from Boehringer Mannheim. Restriction endonucleases were products of New England Biolabs, Boehringer Mannheim, or Bethesda Research Laboratories. ^{32}P-Labeled nucleotides were from the Radiochemical Centre (Amersham, England).

RESULTS

Construction and Identification of the pα-2 cDNA Clone. The cDNA insert of the previously isolated pα-1 clone (15) contained only 750 base pairs (bp) and thus did not include the complete coding sequence of the α-chain mRNA. To obtain longer clones, another cDNA library was constructed. Recombinant plasmids were then tested by *in situ* hybridization for their ability to hybridize with the nick-translated pα-1 DNA insert. From those giving a positive response, a clone with an insert of 1,350 bp was selected and named pα-2.

cDNA Nucleotide Sequence Determination. Restriction maps of the pα-1 and pα-2 cDNAs are shown in Fig. 1. These two cDNAs overlap by 270 bp, thus end-to-end representing a total unique DNA sequence of 1,828 bp. This length is close to that of the mRNA (about 2,000 nucleotides) coding for the α-chain (15).

The nucleotide sequences of both strands of pα-1 and pα-2 cDNAs were determined by the method of Maxam and Gilbert (21) following the strategy illustrated in Fig. 1. The nucleotide sequence of the α-subunit mRNA is shown in Fig. 2. It contains only one open reading frame. The coding sequence is 1,383 nucleotides long and extends from the AUG codon located in

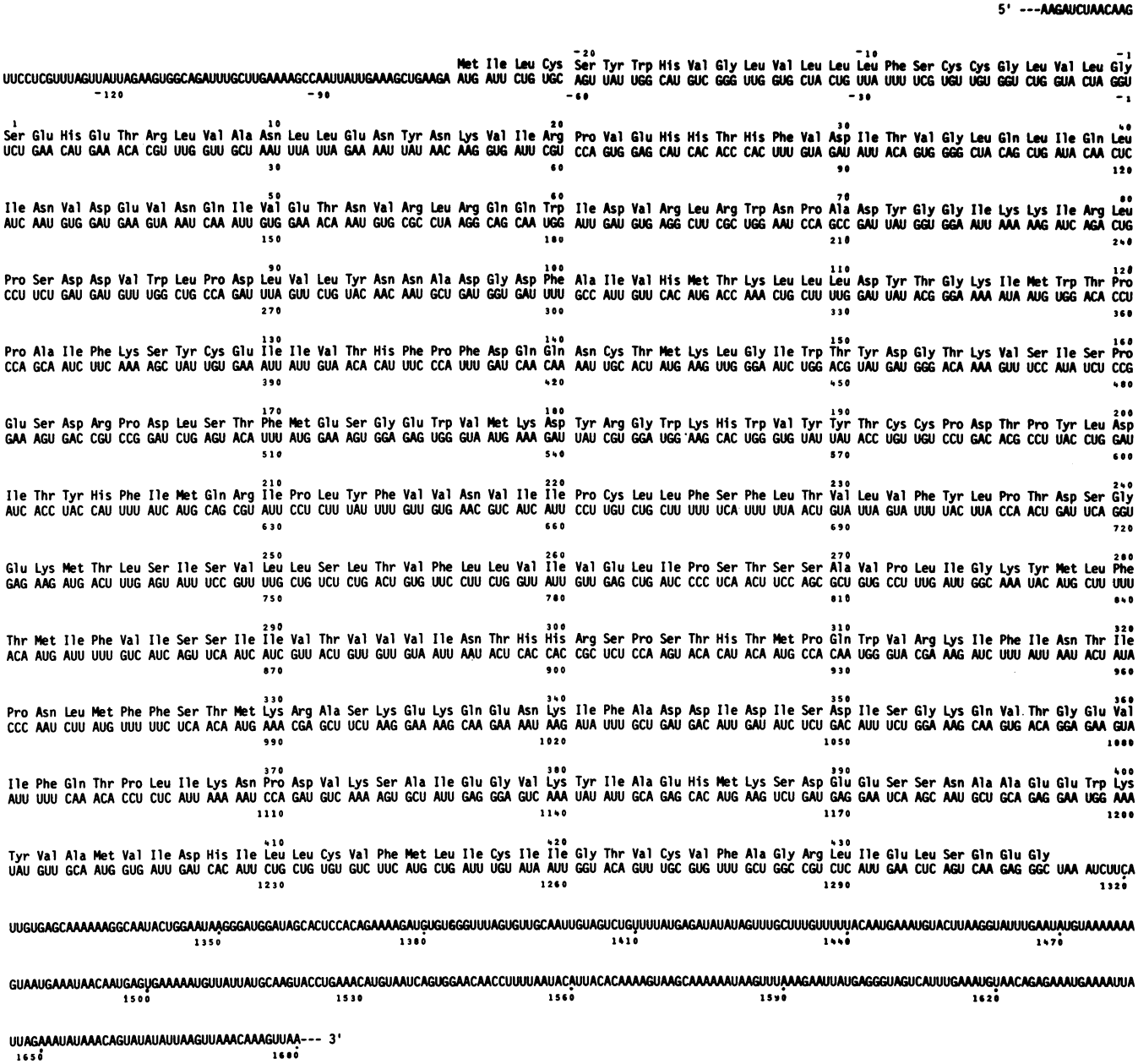

FIG. 2. Primary structure of the *T. marmorata* mRNA coding for the α-subunit of the acetylcholine receptor. The mRNA nucleotide sequence was deduced from those of the pα-1 and pα-2 cDNA inserts. Nucleotides are numbered in the 5′ to 3′ direction starting with the first residue of the codon corresponding to the NH_2-terminal residue of the mature α-subunit. Nucleotides of the 5′ untranslated region are indicated by negative numbers. The predicted amino acid sequence of the α-chain precursor is shown above the nucleotide sequence. Amino acid residues are numbered starting with the NH_2-terminal serine of the mature α-subunit. Residues of the transient signal peptide are indicated by negative numbers.

pα-2 cDNA to the UAA termination codon present in pα-1 cDNA. This sequence is flanked by 76 nucleotides of the 5′ untranslated region and 369 nucleotides of the 3′ untranslated region and is probably incomplete since we did not detect any poly(A) sequence or polyadenylylation signal.

Complete Amino Acid Sequence of the α-Subunit Precursor. The complete protein sequence encoded is 461 amino acids long (Fig. 2). In front of the already known NH_2-terminal sequence of the mature *T. marmorata* AcChoR α-subunit (13), 24 additional amino acids were found. This segment is highly hydrophobic, most of the residues have long side chains, and the last one is a small neutral amino acid (glycine). It is characteristic of a transient cleavable signal sequence (23) and is indeed cleaved off (in *Torpedo californica*) in the presence of microsomes (24). The mature subunit thus contains 437 amino acids and has a molecular mass of 50,200 daltons. It includes eight cysteine residues whereas three were estimated by amino acid composition analysis of the purified α-subunit (from *T. californica*) (25).

During the completion of this work, a partial sequence of the α-subunit mRNA from *T. marmorata* AcChoR was published (16). Except for one amino acid at position 323 and a silent nucleotide mutation at amino acid position 244, identity between the common fragments is noted. In another article, the mRNA sequence of *T. californica* AcChoR α-chain was reported (17). Interestingly, several nucleotide differences exist between the mRNAs from these two closely related species. Of 11 mutations in the coding region (Table 1), 5 are silent and 6 introduce changes in the amino acid sequence. Three of the latter occur at positions 230, 291, and 424 in a hydrophobic domain (see *Model*) and do not modify its hydrophobic character. The independent reports that amino acid 42 is asparagine in *T. marmorata* (this paper and ref. 16) and serine in *T. californica* (8, 17), strengthen the conclusion that at least some of the observed mRNA sequence differences are not due to individual variation but to species-specific differences. In the 3′ untranslated region, we found 11 nucleotide differences among 369 nucleotides (3%) compared with 0.8% in the coding region.

A MODEL FOR THE TRANSMEMBRANE ORGANIZATION OF THE α-SUBUNIT

As deduced from the sequence of the mRNA, the mature protein contains 437 amino acids and has a molecular mass of 50,200 daltons, significantly larger than the *apparent* molecular mass estimated from migration on $NaDodSO_4$/polyacrylamide gels. In the absence of determination of the COOH-terminal residue on the mature protein, one cannot exclude an eventual COOH-terminal post-translational cleavage. However, sequence data obtained on peptides from the purified α-chain have shown (unpublished data) that this cleavage could not occur before the glutamic acid at position 384. In any case, such a COOH-terminal maturation cannot explain the discrepancy noted between the apparent molecular mass of the *in vitro*-synthesized and unprocessed precursor of the α-subunit (38,000 daltons) on $NaDodSO_4$/polyacrylamide gels and its actual amino acid content (461 residues). A particular conformation of the polypeptide chain that might resist the usual denaturation conditions could explain such discrepancies for both the precursor and the mature α-chain.

Table 1. Comparison of nucleotide sequences of mRNAs coding for α-chain precursors of *T. marmorata* and *T. californica* AcChoR

Mutation type	Position	Amino acid		Codon	
		marmorata	*californica*	*marmorata*	*californica*
Silent	−16	Val	Val	GUC	GUA
	159	Ser	Ser	UCU	UCC
	196	Thr	Thr	ACG	ACU
	217	Asn	Asn	AAC	AAU
	300	His	His	CAC	CAU
Nonsilent	42	Asn	Ser	AAU	AGU
	230	Val	Gly	GUA	GGA
	291	Val	Ile	GUU	AUU
	318	Asn	Asp	AAU	GAU
	323	Leu	Val	CUU	GUU
	424	Cys	Ser	UGC	AGC

T. californica data are from ref. 17.

Straightforward inspection of the amino acid sequence discloses four major sequence domains within the mature α-subunit. Two of them (residues 1–209 and 299–408) are rich in hydrophilic amino acids, whereas the other two (residues 210–298 and 409–437) have a highly hydrophobic character. Structural data (26, 27) on several known membrane proteins have shown that regions that span the lipid bilayer are usually α-helices composed of about 20 consecutive uncharged and mostly hydrophobic amino acids flanked by charged residues. Four such uncharged regions can be identified within the hydrophobic domains of the AcChoR α-subunit: I (residues 210–237), II (residues 243–261), III (residues 277–298), and IV (residues 409–428). They are bounded by charged amino acids and often by proline residues, known as potential helix breakers. These regions are most likely organized in α-helices. A planar projection of these transmembrane helices was constructed (Fig. 3) using the conventional values for helical pitch and number of residues per turn. Charged residues are positioned close enough to a surface of the membrane so that their side chains can reach the aqueous environment. Helix IV is shown in dotted lines because post-translational maturation by cleavage of the COOH-terminal domain after residue 384 (see above) cannot be excluded. The α-helices I, II, and III are linked by very short polypeptides and thus should form a compact core.

Two orientations of the polypeptide chain are possible with respect to the synaptic and cytoplasmic surfaces of the membrane. Data on *in situ* proteolysis of the α-chain allowed us to select one of these arrangements. Purified AcChoR-rich membrane fragments from *T. marmorata* make sealed vesicles with a uniform right side-out orientation. As shown by Wennogle and Changeux (14), the AcChoR resists proteolysis unless the vesicles are opened. Then, trypsin progressively converts the α-chain into fragments of apparent molecular masses (as determined by electrophoresis on $NaDodSO_4$/polyacrylamide gels) of 38,000 daltons, 35,000 daltons, and finally 32,000 daltons. The 35,000-dalton fragment has the same NH_2-terminal sequence as the mature α-subunit (14). Proteolysis thus occurs at the cytoplasmic face near the COOH-terminus of the protein. Such cleavage sites can be found only in the hydrophilic fragment (residues 301–408), which contains 14 lysine or arginine residues. It must therefore face the cytoplasm. The most plausible orientation of the α-subunit in the subsynaptic membrane is illustrated in Fig. 3. In agreement with this interpretation, this domain contains several serine and threonine residues, which are possible sites for the phosphorylation known to occur at the cytoplasmic face of the membrane (28).

The amino acid sequence of the first large hydrophilic domain (residues 1–209) does not contain clear clues as to its three-dimensional organization. Being rather long, this segment could cross the membrane in a nonhelical manner. This seems unlikely. Indeed to do so, it would have to cross it twice since, to fit with the signal hypothesis (29), the NH_2-terminal

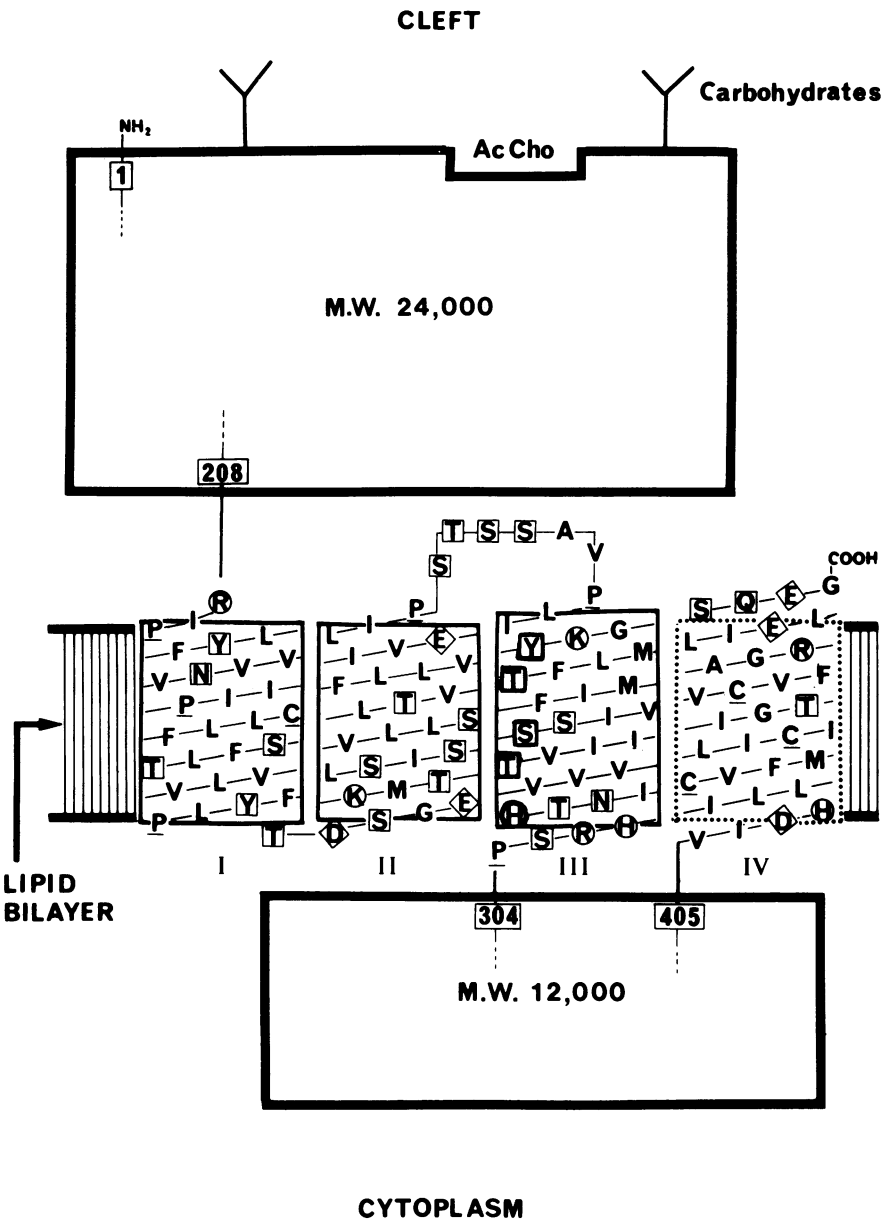

FIG. 3. Schematic model of the arrangment of the α-subunit in the postsynaptic membrane. The α-chain is subdivided into three major domains, represented by rectangles having surfaces proportional to their respective molecular weights. The upper rectangle facing the synaptic cleft is the NH_2-terminal, mostly hydrophilic, domain (residues 1–208). It carries carbohydrates and at least part of the AcCho binding site. The intermediate domain is hydrophobic and composed of four α-helices (I–IV) embedded in the lipid bilayer. The conventional parameters used to construct the helices are helix diameter = 9.6 Å, 3.6 residues per turn, and helical pitch = 5.4 Å. A planar projection is shown here, and the amino acids are indicated by the one-letter code with uncharged polar, positively charged, and negatively charged residues in squares, circles, and diamonds, respectively, and prolines and cysteines underlined. On helix III, the polar residues organized in the remarkable vertical line (see *Model*) are framed with darker symbols. Since a COOH-terminal post-translational cleavage cannot be excluded (see *Model*), helix IV is indicated by a dotted line. The third domain of the α-chain is drawn as a rectangle facing the cytoplasm; this long hydrophilic loop of 101 residues links helix III to IV. A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

amino acid has to be exposed to the synaptic side of the membrane. This implies that a single α-subunit carries 5 or 6 transmembrane spanning segments, which is difficult to reconcile with the physical data (9, 12) obtained by electron microscopy and x-ray diffraction on the AcChoR light form. We thus favor a model (Fig. 3) in which the entire NH_2-terminal domain from amino acid 1 to amino acid 209 is on the synaptic side of membrane. The size of this domain is consistent with physical data on the transmembrane organization and shape of the AcChoR light form (9, 12, 30), which indicate a total transverse height of 11 nm with a bulky moiety 5.5 nm high facing the synaptic cleft and a short 1.5-nm domain extending from the cytoplasmic face. Being the only domain exposed at the synaptic face of the membrane, this long hydrophilic sequence should carry at least the AcCho binding site and the glycosylated residues. Affinity labeling of the AcChoR site by several covalent reagents (6, 7) (but not all of them) requires the reduction of a disulfide bridge located in its vicinity. Of the eight cysteine residues present on the α-subunit, four are located on this hydrophilic domain (amino acids 128, 142, 192, and 193) and possibly participate in this disulfide bridge. The only possible site of N-glycosylation (asparagine-141) is also present on the same domain but several other residues could be O-glycosylated.

Helix III has a unique character. Inspection of its two-dimensional projection (Fig. 3) reveals a vertical line composed

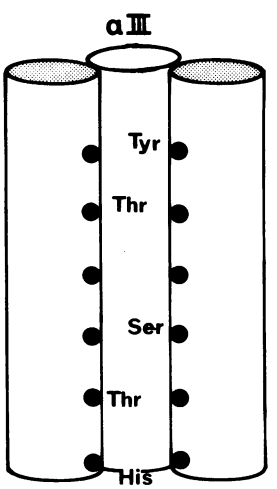

FIG. 4. Schematic model (transverse view) of the open hydrophilic channel of the AcChoR light form. The hydrophilic channel is proposed to be delimited by transmembrane α-helices of the different AcChoR subunits. Helix III of one α-subunit is represented with its characteristic line of hydrophilic residues that face the inside of the channel. Only two other α-helices are shown, on which putative similar polar amino acids are indicated by black dots.

exclusively of polar residues: tyrosine-277, threonine-281, serine-288, threonine-292, histidine-299. In three dimensions, this line corresponds to a hydrophilic surface that spans the lipid bilayer (Fig. 4). This surface in conjunction with similar ones that might be present on the other subunits could delimit a transmembrane hydrophilic channel. This channel could of course serve as the ionophore. If only five α-helices surround the channel, it would then have an internal diameter of 6.6 Å. This value is in agreement with the estimation of the smallest cross section of the open ionophore (31). However, the known cationic selectivity of the ionophore can hardly be explained solely on the basis of the uncharged amino acids present on the hydrophilic surface of the α-subunit helix III. In addition to the membrane-spanning portion of the physiological channel, it is possible that parts of the extrinsic synaptic and cytoplasmic domains are also involved in the selectivity process. The ionophore carries part of the high-affinity site for noncompetitive blocking agents to which all five subunits of the AcChoR light form are known to contribute (10, 11). It will be of particular interest to see which segment of the α-chain, the synaptic domain or the membrane spanning segment, is labeled by chlorpromazine (11). Interestingly, the line of hydrophilic residues begins with a tyrosine, which could photoreact with chlorpromazine (11).

Finally, the model discussed above is consistent with the concept of the AcChoR being an "allosteric" protein (6). Indeed, if the AcCho binding site and the ion channel are carried by distinct segments of the chain (the hydrophilic domain composed of residues 1–209 and helix III of the hydrophobic core, respectively), an "indirect" interaction between these two classes of sites might take place. Comparison of the complete sequences of the four chains, when available, should further indicate whether, as expected for an allosteric protein (32), some cryptic symmetry (6) of the molecule remains despite striking differences in subunit composition.

Note Added in Proof. After this paper was submitted, Noda *et al.* (33) reported the complete sequences of the *T. californica* β- and δ-subunits. Their data are fitted by the model of transmembrane organization proposed here for the α subunit; in particular, a vertical line of hydrophilic residues is also present on helix III from the β- and δ-subunits.

We thank Dr. J.-L. Popot for his collaboration in elaborating the transmembrane model; Drs. C. Henderson and D. M. Engelman for fruitful discussions; Dr. W. Beard for providing us with avian myeloblastosis virus reverse transcriptase; and B. Geoffroy, C. Jauneault, and H. Nuret for skillful technical assistance. This research received support from the Muscular Dystrophy Association of America, the Fondation de France, the Fondation pour la Recherche Médicale, the Collège de France, the Ministère de la Recherche et de l'Industrie, the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, and the Commissariat à l'Energie Atomique.

1. Katz, B. (1966) *Nerve Muscle and Synapse* (MacGraw-Hill, New York).
2. Anholt, R., Lindstrøm, J. & Montal, M. (1980) *Eur. J. Biochem.* **109,** 481–487.
3. Heidmann, T., Sobel, A., Popot, J. L. & Changeux, J. P. (1980) *Eur. J. Biochem.* **110,** 35–55.
4. Popot, J. L., Cartaud, J. & Changeux, J. P. (1981) *Eur. J. Biochem.* **118,** 203–214.
5. Wu, W. C. S. & Raftery, M. A. (1981) *Biochemistry* **20,** 694–701.
6. Changeux, J. P. (1981) *The Harvey Lectures* **75,** 85–254.
7. Karlin, A. (1980) in *Cell Surface Reviews,* eds. Poste, G., Nicolson, G. L. & Cotman, C. W. (Elsevier, Amsterdam), pp. 191–260.
8. Raftery, M. A., Hunkapiller, M. W., Strader, C. B. & Hood, L. E. (1980) *Science* **208,** 1454–1457.
9. Kistler, J., Stroud, R. M., Klymkowsky, W., Lalancette, R. A. & Fairclough, R. H. (1982) *Biophys. J.* **37,** 371–383.
10. Oswald, R. & Changeux, J. P. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3925–3929.
11. Heidmann, T., Oswald, R. & Changeux, J. P. (1982) *C. R. Acad. Sci. Paris* **295,** 345–349.
12. Cartaud, J., Benedetti, L., Sobel, A. & Changeux, J. P. (1978) *J. Cell Sci.* **29,** 313–337.
13. Devillers-Thiery, A., Changeux, J. P., Paroutaud, P. & Strosberg, D. (1979) *FEBS Lett.* **104,** 99–105.
14. Wennogle, L. & Changeux, J. P. (1980) *Eur. J. Biochem.* **106,** 381–393.
15. Giraudat, J., Devillers-Thiery, A., Auffray, C., Rougeon, F. & Changeux, J. P. (1982) *EMBO J.* **1,** 713–717.
16. Sumikawa, K., Houghton, M., Smith, J. C., Bell, L., Richards, B. M. & Barnard, E. A. (1982) *Nucleic Acids Res.* **10,** 5809–5822.
17. Noda, M., Takahashi, H., Tanabe, T., Toyosato, M. F., Furutami, Y., Hirose, T., Asai, M., Inayama, S., Miyata, T. & Numa, S. (1982) *Nature (London)* **299,** 793–797.
18. Ballivet, M., Patrick, J., Lee, J. & Heinemann, S. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4466–4470.
19. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3961–3965.
20. Rigby, P. M. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
21. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
22. Challberg, M. D. & Englund, P. T. (1980) *Methods Enzymol.* **65,** 39–43.
23. Von Hejne, G. (1981) *Eur. J. Biochem.* **116,** 419–422.
24. Anderson, D. J. & Blobel, G. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5592–5602.
25. Vandlen, R. L., Wu, W. C. S., Eisenach, J. C. & Raftery, M. A. (1979) *Biochemistry* **18,** 1845–1854.
26. Von Heijne, G. (1981) *Eur. J. Biochem.* **120,** 275–278.
27. Engelman, D. M., Henderson, R., McLachlan, A. D. & Wallace, B. A. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 2023–2027.
28. Davis, C. G., Gordon, A. S. & Diamond, I. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3666–3670.
29. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67,** 835–851.
30. Ross, M. J., Klymkowsky, M. W., Agard, D. A. & Stroud, R. M. (1977) *J. Mol. Biol.* **116,** 635–659.
31. Dwyer, T. M., Adams, D. & Hille, B. (1980) *J. Gen. Physiol.* **75,** 469–492.
32. Monod, J., Wyman, J. & Changeux, J. P. (1965) *J. Mol. Biol.* **12,** 88–118.
33. Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikoytani, S., Hiroshe, T., Asai, M., Takashima, H., Inayana, S., Miyata, T. & Numa, S. (1983) *Nature (London)* **301,** 251–255.