# Improved Secondary Structure Predictions for a Nicotinic Receptor Subunit: Incorporation of Solvent Accessibility and Experimental Data into a Two-Dimensional Representation

Nicolas Le Novère, Pierre-Jean Corringer, and Jean-Pierre Changeux
Centre National de la Recherche Scientifique URA D1284 Neurobiologie Moléculaire, Institut Pasteur, 75015 Paris, France

Abstract A refined prediction of the nicotinic acetylcholine receptor (nAChR) subunits' secondary structure was computed with third-generation algorithms. The four selected programs, PHD, Predator, DSC, and NNSSP, based on different prediction approaches, were applied to each sequence of an alignment of nAChR and 5-$HT_3$ receptor subunits, as well as a larger alignment with related subunit sequences from glycine and GABA receptors. A consensus prediction was computed for the nAChR subunits through a "winner takes all" method. By integrating the probabilities obtained with PHD, DSC, and NNSSP, this prediction was filtered in order to eliminate the singletons and to more precisely establish the structure limits (only 4% of the residues were modified). The final consensus secondary structure includes nine $\alpha$-helices (24.2% of the residues, with an average length of 13.9 residues) and 17 $\beta$-strands (22.5% of the residues, with an average length of 6.6 residues). The large extracellular domain is predicted to be mainly composed of $\beta$-strands, with only two helices at the amino-terminal end. The transmembrane segments are predicted to be in a mixed $\alpha/\beta$ topology (with a predominance of $\alpha$-helices), with no known equivalent in the current protein database. The cytoplasmic domain is predicted to consist of two well-conserved amphipathic helices joined together by an unfolded stretch of variable length and sequence. In general, the segments predicted to occur in a periodic structure correspond to the more conserved regions, as defined by an analysis of sequence conservation per position performed on 152 superfamily members. The solvent accessibility of each residue was predicted from the multiple alignments with PHDacc. Each segment with more than three exposed residues was assumed to be external to the core protein. Overall, these data constitute an envelope of structural constraints. In a subsequent step, experimental data relative to the extracellular portion of the complete receptor were incorporated into the model. This led to a proposed two-dimensional representation of the secondary structure in which the peptide chain of the extracellular domain winds alternatively between the two interfaces of the subunit. Although this representation is not a tertiary structure and does not lead to predictions of specific $\beta$-$\beta$ interaction, it should provide a basic framework for further mutagenesis investigations and for fold recognition (threading) searches.

## INTRODUCTION

The nicotinic acetylcholine receptors (nAChRs) belong to the superfamily of ligand-gated ion channels (LGIC) (Cockcroft et al., 1992; Galzi and Changeux, 1994) that are allosteric transmembrane proteins responsible for fast ionic responses to neurotransmitters. These receptors are homo- or hetero-pentamers made from a set of 16 related subunits in vertebrates (9$\alpha$, 4$\beta$, $\gamma$, $\delta$, and $\epsilon$) (review in Le Novère and Changeux, 1995). Other receptors formed of polypeptides homologous to the nAChR subunits include 5-HT3, $GABA_A$, $GABA_C$, and glycine receptors of vertebrates, as well as their invertebrate counterparts. Despite their rather different pharmacological properties (Ortells and Lunt, 1995), these receptors likely possess a common quaternary structure (Eisele et al., 1993; Langosch et al., 1988). The vertebrate $GABA_A$ receptors are formed from a set of 16 related subunits (6$\alpha$, 4$\beta$, 4$\gamma$, 1$\delta$, and 1$\epsilon$) (MacDonald and Olsen, 1994). The $GABA_C$ receptors are homo-pentamers of $\rho$1–3 subunits (Bormann and Feigenspan, 1995). The vertebrate glycine receptors are made from a set of five related subunits (4$\alpha$ and 1$\beta$) (Bechade and Triller, 1994).

Every mature subunit of the nAChR family is assumed to follow the same transmembrane topology (Hucho et al., 1996). A large amino-terminal portion carrying the components of the acetylcholine (ACh) binding site faces the extracellular environment. The three subsequent segments cross the membrane, followed by a large intracellular domain and a fourth segment that again crosses the membrane. The relatively short carboxy-terminal domain is extracellular.

The sequence conservation varies along the subunits. The amino-terminal signal peptide and the middle of the cytoplasmic portion are highly variable, whereas the amino-terminal moiety, as well as the membrane flanking portions of the cytoplasmic part, are well conserved. The transmembrane segments are highly conserved. In humans, the size of the subunits vary from 457 aa ($\alpha$1) to 627 aa ($\alpha$4).

The overall low resolution structure of the nAChRs was initially determined by electron microscopy on single molecules (Cartaud et al., 1973), or bi-dimensional crystals (Kistler and Stroud, 1981) of *Torpedo* electric organ receptor. The nAChR molecules result from the assembly of five subunits arranged around an axis of symmetry perpendicular to the membrane. The length of the molecule is ~120 Å,

*Received for publication 27 May 1998 and in final form 3 February 1999.*
Address reprint requests to Jean-Pierre Changeux, CNRS URA D1284 Neurobiologie Moléculaire, Institut Pasteur, 25, rue du Dr Roux, 75015 Paris, France. Tel.: 33-(0)1-45-68-88-05; Fax: 33-(0)1-45-68-88-36; E-mail: lenov@pasteur.fr.

0006-3495/99/05/2329/17 $2.00

with an extracellular, funnel-shaped portion of 60 Å and a transmembrane portion of 30 Å. The diameter of the extracellular entry of the pore is ~25 Å wide, while the intracellular one is slightly smaller (Toyoshima and Unwin, 1988; Unwin, 1993a). A similar shape was proposed for $GABA_A$ receptors (Nayeem et al., 1994).

Affinity labeling and site-directed mutagenesis have shown that the ligand binding sites are located at the interface of two subunits, formed by residues belonging to two components (Galzi and Changeux, 1994; Table 1). The principal component [on the subunit surface that would be reached first if following the clockwise path when the structure is viewed from the extracellular surface (Machold et al., 1995)] is carried by the $\alpha$ subunits and comprises at least three segments or loops (Dennis et al., 1988; Galzi et al., 1990). Facing it, the complementary component includes three (or possibly four) different segments or loops (Corringer et al., 1995; Czajkowski et al., 1993; Prince and Sine, 1996; Sine et al., 1995; reviewed in Hucho et al., 1996; Tsigelny et al., 1997). Such composite ligand binding sites appear to be conserved throughout the superfamily of LGIC. Indeed, it has been shown that the binding sites for ACh, GABA, glycine, and benzodiazepines are homologous (Schmieden et al., 1992; Vandenberg et al., 1992; reviewed in Galzi and Changeux, 1994). The ionotropic glutamate receptors constitute a separate superfamily in which agonist sites probably do not occur at subunit interfaces (Paas, 1998).

Cryoelectron microscopy of the *Torpedo* electric organ receptor has provided three-dimensional (3D) images of the nAChR at a resolution of 9 Å (Unwin, 1993a). Such a resolution is not sufficient to resolve the spatial position and the secondary structure assignment of any particular amino acid. Although the extracellular domain has been successfully produced in a soluble form (Wells et al., 1998), the quantities obtained are still too low to permit the production of crystals suitable for x-ray diffraction. The NMR approach has been limited to small fragments (Basus et al., 1993). Some attempts were made with other methods, such as atomic force microscopy (Lal and Yu, 1993), though with a resolution lower than that of electron microscopy on two-dimensional (2D) crystals.

It is therefore of interest to obtain information on the receptor protein organization from the data currently available, i.e., the sequences of the subunits. Accordingly, in parallel with the experimental approaches, efforts have been made to predict the structure of the individual subunit with computational techniques. Two approaches have been used. The comparative modeling techniques sought to give a structural description of a protein provided that a plausible structural model can be identified. The problem resides in the identification of a suitable template from sequence information only. However, the lack of sufficient sequence similarity between an nAChR subunit and a protein of known structure requires fold recognition methods (Gready et al., 1997; Tsigelny et al., 1997) which, as known from test cases, are only partially successful in recognizing similar folds in the absence of sequence similitude (Rost and Sander, 1996). This approach also suffers from the fact that a plausible 3D model may not exist in the currently available protein structure database (Marchler-Bauer et al., 1997). The two models proposed so far are indeed different (Gready et al., 1997; Tsigelny et al., 1997). In parallel, ab initio secondary structure predictions were performed with first-generation algorithms (single amino acid-based, 50–60% accuracy) by Finer-Moore and Stroud (1984) and Ortells (1997).

Here we present a secondary structure prediction of the nAChR subunit based on third-generation algorithms (based on multiple alignments and that are capable of achieving >70% accuracy), in order to take into account the information derived from the wealth of cloned homologous subunit sequences. The combination of several independent first- and second-generation algorithms has been shown to in-

**TABLE 1 Summary of experimentally identified residues in the large amino-terminal hydrophilic domain**

| Amino Acid | Function in the oligomer | | |
|---|---|---|---|
| | Complementary | | Principal |
| T34 (83) | agonist site | | |
| W54 (104) | agonist site (D) | | |
| T61 (110)-V74 (124) | | main immunogenic region (MIR) | |
| W86 (135), Y93 (142) | | | agonist site (A) |
| L108 (159), Q116 (174), L118 (176) | agonist site (E) | | |
| t$\alpha$1N141 (198) | | glycosylation | |
| W 149 (206), Y151 (208) | | | agonist site (B) |
| D152 (209), T154 (211) | | glycosylation (experimentally induced) | |
| D164 (235), E173 (244) | agonist site (F) | | |
| c$\alpha$7D163 (235)-E172 (244) | | $Ca^{2+}$ binding site | |
| t$\alpha$1E187 (256) | | glycosylation (mongoose) | |
| t$\alpha$1F189 (258) | | glycosylation (cobra) | |
| Y188 (259) C190 (263), C191 (264), Y195 (269) | | | agonist site (C) |

Each amino acid is given according to the mature $\alpha$7 subunit of chick. The numbering of the alignment shown in Fig. 4 (AL1) is given in brackets. Refer to the alignment to find the corresponding residues in other subunits. For a more exhaustive review of amino acids identified in the muscle subunits, see Tsigelny et al., 1997 and Hucho et al., 1996.

crease the accuracy of secondary structure predictions (Biou et al., 1988; Nishikawa and Ooi, 1986; Zhang et al., 1992). We describe a program that integrates results from several prediction algorithms and multiple homologous proteins. We applied this program to the different members of the nAChR family and LGIC superfamily to increase the signal/noise ratio. In addition, the program furnished the consensus of predicted solvent accessibility and topology. By using these data in combination with information obtained from experimental sources, we integrated the results into a 2D representation of a typical nAChR subunit.

## MATERIALS AND METHODS

### Alignment

All sequences used in this study can be found in the ligand-gated ion channel (LGIC) subunit database at the URL (http://www.pasteur.fr/units/neubiomol/LGIC.html). For the secondary structure predictions, two multialignments were achieved with ClustalX software (Thompson et al., 1997; available at ftp-igbmc.u-strasbg.fr) (pairwise gap opening, 10; pairwise gap extension, 0.1; multiple gap opening, 5; multiple gap extension, 0.05; Blosum matrix series). One alignment was carried out with 18 subunit sequences of cationic channels (AL1). AL1 contains 5-HT3 from *Mus,* nicotinic $\alpha$1 of *Torpedo,* $\alpha$2–6, $\alpha$9, and $\beta$2–4 of *Rattus*, $\alpha$7–8 of *Gallus*, $\beta$1, $\gamma$, $\delta$, $\epsilon$, of *Mus* (one example of each paralog gene), and DEG3 of *Cænorhabditis* (which has still no uncovered vertebrates ortholog). Another alignment was constructed with 38 LGIC (cationic and anionic LGIC) sequences (AL2). AL2 contains AL1 subunit sequences plus GABA $\alpha$1–6, $\beta$1–3, $\gamma$1–3, $\rho$1–3, $\delta$, glycine $\alpha$1–3, and $\beta$ from *Rattus*. The aim was to determine whether the incorporation of information from more distantly related sequences would improve the predictions. We did not use more than one sequence per group orthology because of the high similarity between orthologs (and hence the lack of additional information brought from the use of multiple orthologs). The ASSP software (Russel and Barton, 1993; available at ftp://geoff.biop.ox.ac.uk/programs/assp/) allowed us to expect a $Q_3$ accuracy (i.e., a percentage of three-state comparison identity) of perfect prediction in the interval [83.45–100%] for AL1 and [82.74–100%] for AL2. To study the conservation of sequence at each position along the sequence, a third multiple alignment was constructed from 152 different LGIC subunit sequences. All these sequences correspond to subunits shown to be integrated in functional receptors (thus eliminating the putative members originating from large-scale genome projects).

### Secondary structure prediction by consensus average

A computer program was written in C to integrate secondary structure predictions based on different algorithms. SSPCA (for secondary structure prediction by consensus average) was designed to combine three-state predictions and probabilities from several prediction programs and several sequences (Fig. 1). The SSPCA program is also designed to treat other types of prediction such as solvent accessibility and topological arrangements for membrane proteins. The individual predictions were not weighted by sequence similitudes.

As input, SSPCA takes an alignment of amino acid sequences (in a Clustal format) and a file containing the predictions. The prediction file contains for each sequence and for each method (if available) the probability for helix, $\beta$, and coil [0–9], the resulting secondary structure prediction [H(elix) or E(xtended) or C(oil)], the probability of accessibility to solvent [0–9], the resulting accessibility to solvent (e(xposed) or b(uried)), and the topological state (o(utside), i(nside), T(ransmembrane)). The output of SSPCA is composed of (points 1–5 concern only the secondary structure prediction):

1. The *MxS* predictions $P_{(m_i,s_x)}$ where $M$ is the number of method, $S$ the number of sequences, $m_i$ the $i$th method, and $s_x$ the $x$th sequence, projected on the alignment (insertion of gap in the predictions when present in the

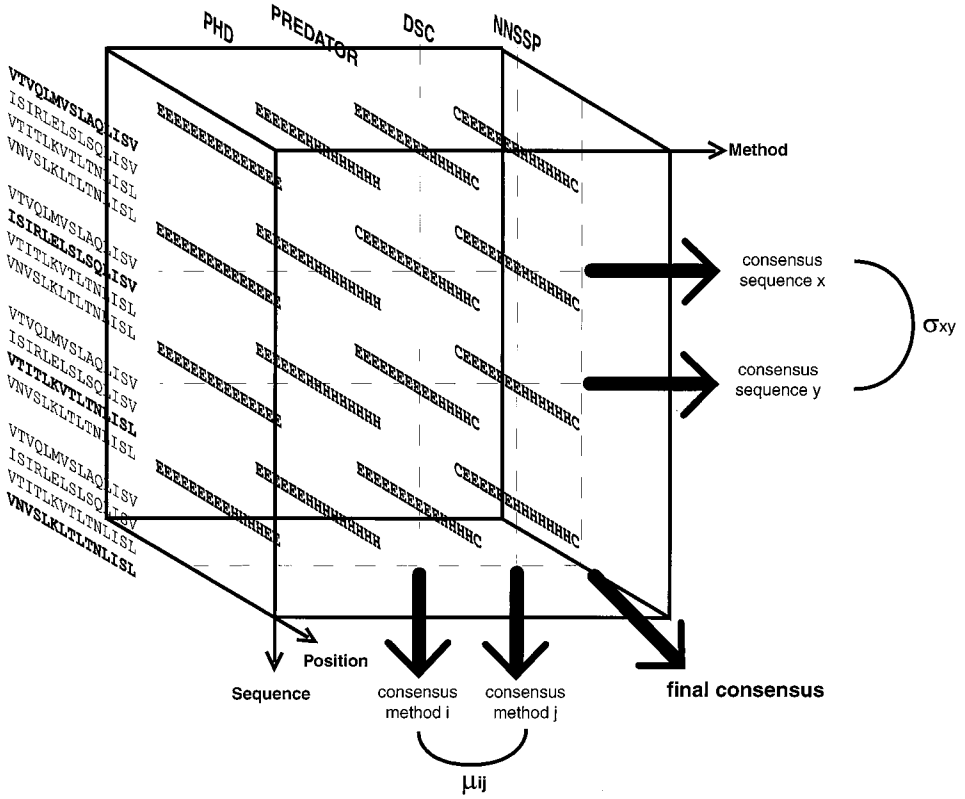


FIGURE 1 Scheme describing our approach. Only four sequences are displayed (the portion of sequence shown corresponds to $E_1$).

alignment). Each $P_{(m_i,s_x)}$ is then a character string with the length of alignment, each character belonging to {H, E, C,'-'}).

2. The $MxS(MxS-1)/2$ pairwise comparisons $C[P_{(m_i,s_x)}, P_{(m_j,s_y)}]$ of the predictions $P_{(m_i,s_x)}$ and $P_{(m_j,s_y)}$. If $\bar{G}$ is the set of positions of the alignment where neither $P_{(m_i,s_x)}$ nor $P_{(m_j,s_y)}$ contain a gap, that is, where both predictions are defined,

$$C[P_{(m_i,s_x)}P_{(m_j,s_y)}] = \frac{100}{\mathrm{card}(\bar{G})} \sum_{\alpha\in\bar{G}} \delta[P_{(m_i,s_x)}, P_{(m_j,s_y)}]\alpha \quad (1)$$

Where

$$\begin{cases}\delta(a,b)=1 & \text{if} \quad a=b\\ \delta(a,b)=0 & \text{if} \quad a\neq b\end{cases} \text{(Kronecker's } \delta\text{)}$$

and $\mathrm{card}(\bar{G})$ is the cardinal (the number of elements) of $\bar{G}$.

3. The congruence between methods $\mu_{i,j}$: this parameter represents the percent identity between the consensus of predictions for all the sequences of two methods.

$$\mu_{i,j} = \frac{1}{S}\sum_{s=1}^{S} C[P_{(m_i,s_x)}P_{(m_j,s_x)}] \quad (2)$$

for every pair $(i, j)$, $i \neq j$.

4. The congruence between sequences $\sigma_{x,y}$. This parameter represents the percent identity between the consensus of predictions for two sequences by all the methods.

$$\sigma_{x,y} = \frac{1}{M}\sum_{i=1}^{M} C[P_{(m_i,s_x)}P_{(m_i,s_y)}] \quad (3)$$

for every pair $(x, y)$, $x \neq y$. This parameter permits the comparison of the predictions for different homologous proteins.

5. The consensus predictions and the sum of probabilities: by sequences, by methods, and in toto (and the percent helix and strand for each consensus prediction). For each position, the consensus is computed as the major state. In case of identical cardinals, the arbitrary priority order is E > H > C > '-'. The percent helix and strand is given for the total nongapped consensus length.

6. The global consensus solvent accessibility. In case of identical cardinals, the arbitrary priority order is $b > e >$ '-'.

7. The global consensus topology. In case of identical cardinals, the arbitrary priority order is $T > i > o >$ '-'.

## Secondary structure prediction programs

PHDsec (Rost and Sander, 1993a, b; 1994a) is composed of several cascading neural networks (previously trained on proteins of known structures). A first network takes as input a set of vectors representing the amino acid composition at positions of the multiple alignment in a window sliding along it. Its output is composed of a vector representing the probabilities for each of the three states of the central residue of the window. Since the secondary structure of a residue is not independent of the structure of neighboring residues, a second step takes into account these local interactions. A neural network takes as input the vectors present in a window sliding along the previous output. Its own output is a refined three-state probabilities vector. Another step consists of averaging (for each state) the outputs from independently trained networks. Finally, a "winner takes all" decision assigns the secondary structure state. No explicit rules are included in the algorithm. PHD may generate its own alignment with the submitted sequence [with the MaxHom algorithm (Sander and Schneider, 1991)]. Therefore, for every sequence of AL1 and AL2, a different alignment was generated and used for the prediction. PHDsec is accessible at the URL (http://www.embl-heidelberg.de/predictprotein/predictprotein.html).

PREDATOR (Frishman and Argos, 1996, 1997) is based on the calculated propensities of every 400 amino acid pairs to interact inside an $\alpha$-helix or one upon three types of $\beta$-bridges. It then incorporates nonlocal interaction statistics. PREDATOR also uses propensities for $\alpha$-helix, $\beta$-strand, and coil derived from a nearest-neighbor approach (see below). To use information obtained from homologous proteins, PREDATOR relies on local pairwise alignments. PREDATOR is able to use Clustal alignment as input. The program was employed with the option '-a,' which furnishes a prediction for every sequence of the input set. The source code is kindly distributed by the authors. PREDATOR is also accessible at the URL (http://www.embl-heidelberg.de/cgi/predator serv.pl).

DSC (King and Sternberg, 1996) combines several explicit parameters in order to produce a "meaningful" prediction. It runs the GORIII algorithm [Gibrat et al. (1987), based on information theory applied to local interactions] on every sequence to provide mean potentials for the three states. In addition, DSC uses the presence of insertions/deletions, the distance from the end of the chain, the moment of conservation, and the moment of hydrophobicity (the two last parameters given an $\alpha$-helical structure and a $\beta$-strand structure). A linear combination of these different attributes gives an output that is subsequently filtered. The program was used with the following options: '-a' (to turn off removal of poorly aligned sections), '-i' (to stop removal of isolated predictions), '-f1' (to apply the filtering rules once), and '-w' (Clustalw alignment). The source code is kindly distributed by the authors. DSC is also accessible at the URL (http://bonsai.lif.icnet.uk/bmm/dsc/dsc_read_align.html).

NNSSP (Salamov and Solovyev, 1995) is based on the nearest-neighbor algorithm [sometimes improperly called the "homologue" method (Levin et al., 1986; Nishikawa and Ooi, 1986)]. The basic idea of the nearest-neighbor approach is the prediction of the secondary structure state of the central residue of a test segment, based on the secondary structure of similar segments from proteins with known 3D structure. The information provided by the different templates is scored according to their similarity (according to the sequence or other properties) with the test segment. NNSSP is an enhancement of the algorithm designed by Yi and Lander (1993), which selects the neighbors by the mean an environmental score (Bowie et al., 1991) and combine by the mean of a neural network predictions made with different parameters (environmental scores, length of nearest-neighbors . . . ). In addition to the latter program, it incorporates information from multiple aligned sequences (by averaging their scores for the weighting of each nearest-neighbor). An executable program was kindly provided by the authors. NNSSP is also accessible at the URL (http://dot.imgen.bcm.tmc.edu:9331/psspredict ion/pssp.html).

A program was written in C to convert a Clustal alignment into NNSSP alignments, clu2nnssp. This program is available at the URL (http://www.pasteur.fr/units/neubiomol/softwares.html) or upon request.

## Accessibility to solvent and topology program

PHDacc (Rost and Sander, 1994b) is able to compute the probable accessibility to solvent. It was used to refine the secondary structure predictions.

PHDhtm (Rost et al., 1995, 1996) was used to provide a more accurate prediction of the transmembrane segments position, rather than the original one, established with only a few subunits and only from hydropathy plots (Popot and Changeux, 1984).

## Conservation index

A computer program, ConsIndex, was written in C to compute the sequence conservation between homologous sequences at each position of a multiple alignment. The program takes as input an alignment of a Clustal-like format and a similarity matrix. It computes first the $N(N-1)/2$ global similarities $S_{ij}$ (identities if the identity matrix is input) of the $N$ sequences. Then for each position of the alignment, a conservation index CI is

computed as follows:

$$CI = \left( \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{s_{ij}}{S_{ij}} \middle/ \sum_{i=1}^{N} \sum_{j=i+1}^{N} S_{ij} \right) \quad (4)$$

where $s_{ij}$ is the relevant similarity matrix element for the sequences $i$ and $j$ at the considered position. In the present work, the default similarity matrix of the Wisconsin package program GAP (Devereux et al., 1984) was used. It was rescaled from [−1.2,1.5] to [0–100]. The gap was added as an independent amino acid, with every matrix element involving it considered as null. ConsIndex is available at the URL (http://www.pasteur.fr/units/neubiomol/softwares.html) or upon request.

## RESULTS

### Strategy

Previous works have shown that the accuracy of secondary structure predictions increases from the combination of several independent first- and second-generation algorithms (Biou et al., 1988; Nishikawa and Ooi, 1986; Zhang et al., 1992). Here we combined the prediction of several third-generation algorithms, using the information given by a set of aligned homologous sequences to compute the secondary structure of nAChR subunits.

The algorithms used in this study were chosen according to three criteria: 1) they analyze multiple alignments instead of single protein sequences; 2) they yield a better than 70% accuracy for three-state (H, E, C) prediction when tested on a set of proteins of known structure with sequence identities lower than 25% (Rost and Sander, 1994a) or during blind predictive situations (King, 1996; Rost, 1997); and 3) each of these algorithms is based on a different predictive approach. Each program was applied successively on every sequence of the alignments to increase the signal/noise ratio.

Two different sets of sequences were used to make the secondary structure predictions. The first (AL1) represented the entire group of cationic LGIC subunits in the acetylcholine receptor superfamily (5-HT3 and nicotinic receptors). A second set of sequences (AL2) contained the first set and, in addition, sequences covering the whole group of anionic LGIC subunits (GABA and glycine receptors).

### Consistency of the predictions between methods and sequences

The congruencies between methods $\mu_{i,j}$ for every pair of methods are listed in Table 2. The four methods gave all $\mu_{i,j}$ values >67%. The use of the larger set of sequences caused a decrease of $\mu$, which nevertheless remained >57%. The congruence between sequence consensus predictions $\sigma_{x,y}$ was also examined for every pair of sequences. The predictions for the cationic LGIC subunits were found consistent, the congruencies varying from $\sigma_{deg3,\alpha1} = 80.8\%$ to $\sigma_{\alpha3,\alpha6} = 95.3\%$. In the larger set, the lowest $\sigma$ occurred just above 64%, a value much larger than random (which is 33% for a nonbiased three-state comparison and 38% if the present bias of PDB is taken into consideration). The good congruency of the different predictions for the various members of the nAChR family is illustrated in Fig. 2 (*top*), where the peaks are sharp and 17 of 25 final structural elements are predicted in >90% of the cases. The sequence consensus predictions were very similar. The positions of the secondary structure were almost identical, with little variation of the assignments. The method consensus predictions were more variable, though similar. The assignment of the structures varied somewhat, as well as (but only very rarely) their occurrence.

The resemblance of protein 3D structures (rmsd) is proportional to their sequence identity (Chothia and Lesk, 1986; Flores et al., 1991). The incorporation of distant sequence information is expected to increase the reliability of predicted structures, although decreasing the consistency of the overall prediction (Russel and Barton, 1993; Sternberg, 1996). The $\sigma$ values of AL2 were plotted against the global amino acid similarities determined by the conservation index program. A correlation unambiguously occurred between the sequence similarities and the structure prediction similarities (Fig. 3) ($n = 703$, $r = 0.882$, $p < 0.001$). Two main components emerged from the comparisons: a lower similarity population representing the comparisons of anionic/cationic (e.g., $GABA_A$ vs. nAChR), with a higher similarity population representing the comparisons of anionic/anionic or cationic/cationic LGIC subunits. Together, these data show that the variations between the secondary structure predictions therefore were not random, as expected from algorithm imperfections. On the contrary, they relied on the variation of sequence. This reflects the fact that if the core structures are conserved between the different superfamily members, as supported by a large body of experimental evidence (Galzi and Changeux, 1994), then the structural assignation at the level of individual residue may vary (for instance at the extremities of the structures). This variation was indeed found to increase when the sequence

**TABLE 2 Congruence of the predictions given by the different methods**

| | PHD | PREDATOR | DSC | NNSSP |
|---|---|---|---|---|
| PHD | 100.00 (±0.00) | | | |
| PREDATOR | 73.58 (±5.85) | 100.00 (±0.00) | | |
| | 69.32 (±6.35) | | | |
| DSC | 73.20 (±3.69) | 67.14 (±2.39) | 100.00 (±0.00) | |
| | 66.67 (±4.7) | 57.12 (±2.98) | | |
| NNSSP | 76.64 (±4.84) | 76.44 (±3.83) | 78.07 (±2.11) | 100.00 (±0.00) |
| | 71.7 (±4.42) | 60.53 (±5.45) | 69.13 (±5.05) | |

$\mu_{ij}$ expressed as mean (±SD). Upper values are from AL1 analysis, lower values are from AL2 analysis.

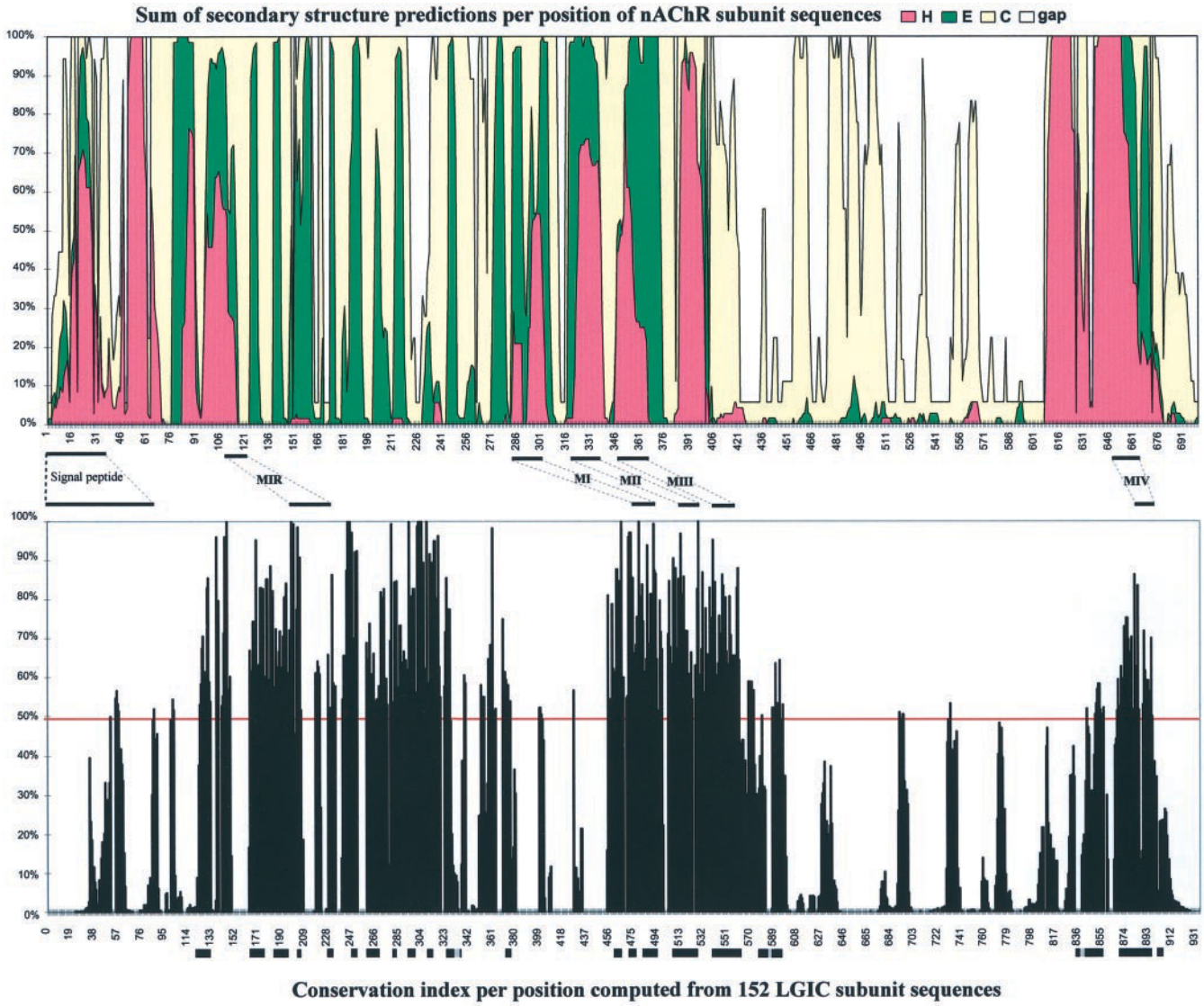


FIGURE 2 *Top:* relative incidence of the three-state populations per residue position. For each residue position of AL1,

$$\sum_{i=1}^{M}\sum_{x=1}^{S} P_{(m_i,s_x)} \tag{5}$$

the sum of all the state predictions for all methods and all sequences, is represented. The diagram is cumulative, i.e., the difference between the height of the peaks at a given position is informative, not the height of the peaks itself. *Bottom:* conservation index per residue position. For each residue position of an alignment of 152 LGIC subunit sequences, a conservation index was computed by ConsIndex. The black squares under the graph represent the final predicted structures ($\alpha$-helix or $\beta$-strand) in the mature subunit. Note that if an $\alpha$-helix is adjacent to a $\beta$-strand, there is only one black square.

relationship decreases. Another conclusion can be derived from Fig. 3. The more distant to our target protein are the homologs used to infer secondary structure, the less reliable is the information obtained. A trade-off is reached between the information obtained from multiple alignments (reliability of secondary structure position and assignment) and the mispredictions at the level of individual residues due to sequence divergence (Russel and Barton, 1993). There is no known method published up to now to establish the best compromise.

The final results obtained with the two alignments AL1 and AL2 were very similar, with only a few residues predicted to be in a different state. Every structure except one was equally predicted with both sets, and in all these cases the secondary structure assignment remained the same. Therefore, except when otherwise stated, we present the results obtained with AL1 (see Fig. 4).

## Raw secondary structure prediction

The proportions of the three-state populations in the entire set of predictions for each alignment position are presented in Fig. 2 (*top*). Fig. 4 shows the raw consensus prediction,

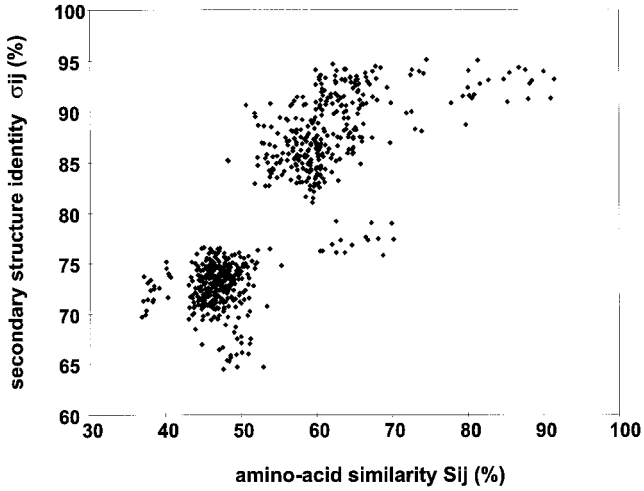


FIGURE 3 Correlation between sequence similarities and secondary structure prediction identities. For $N$ sequences, there are $(N-1)(N-2)/2$ amino acid similarities. Here are represented the 703 dots corresponding to the 38 sequences of AL1. The amino acid similarities $S_{ij}$ have been computed by the program ConsIndex. They are plotted against the secondary structure identities $\sigma_{ij}$, as described in the Methods section. Each point is therefore the comparison of one amino acid similarity versus 16 predictions (four prediction methods for each sequence). The bivariate regression analysis shows that the correlation is meaningful ($n = 703$, $r = 0.882$, $p < 0.001$). Note the two components: the bottom left concentration represents the comparisons of anionic/cationic (e.g., $GABA_A$ vs. nAChR), whereas the less dense upper right distribution represents the comparisons of anionic/anionic or cationic/cationic.

in plain text just below the alignments, with the designation of the structure above. In Fig. 2 (*bottom*), the conservation index determined on the full superfamily of LGIC (152 subunits) is plotted, along with the predicted secondary structures (*black squares below the graph*). In all instances but three ($E_9$, $H_F$, and $H_G$) the predicted structures were located in regions of high (>50%) conservation. The region of $E_9$ is in fact highly conserved except for the nematode subunit unc38. The region of $H_F$ and $H_G$ are highly conserved in cationic channel subunits but less in anionic channels. In summary, within the cationic channel subunit family, all predicted structures were located in regions of high conservation. This fact is important since the high sequence variation between family members is likely to occur in less well-structured regions. A structure predicted in a conserved region is therefore more likely to be accurate.

## Refined secondary structure predictions

PHD, DSC, and NNSSP provide prediction probabilities for the three states in addition to the predicted state. Combining these probabilities permits the correction of the threshold decisions at the level of single predictions, which may lead to false assignment, and offers the possibility of resolving some single-residue problems, such as singletons (isolated structured residue) or amino acid located at the borderline of the secondary structure motifs. The changes made in this way affect only 29 residues. The resulting refined prediction contains (without the signal peptide) nine $\alpha$-helices (mean length 13.9 amino acids) designated $H_A$ to $H_{H,}$ and 17 $\beta$-strands, designated $E_1$ to $E_{17}$ (mean length 6.6 amino acids). Their positions and lengths are summarized in Table 3. Except for two large helices surrounding a large $\beta$-strand at the amino-terminal extremity, the extracellular portion of the subunits was predicted to occur as an all-$\beta$ structure, formed by successive short strands.

The structure of the carboxy-terminal portion of $H_A$ is consistent with solvent accessibility patterns (described in the following by strings of 'e' for exposed and 'b' for buried) i.e., "bbeebbee," its amino-terminal portion being completely exposed.

The structure at the center of $E_1$ is also consistent with solvent accessibility patterns "bebebe"; its two extremities being predicted as completely buried. Its carboxy-terminal portion is less consistent, since it is predicted in an $\alpha$-helical state in every AL1 sequence consensus (see Fig. 2, *top*). An $\alpha$-helical structure for the last four residues could be envisioned. Indeed, for AL1 these $\alpha$-helical residues were predicted in every sequence consensus and in three of four methods' consensus (only PHD predicted all the residues under $\beta$-strand state). However, for AL2, only the PREDATOR consensus, nicotinic $\alpha$8 consensus, and nicotinic $\gamma$ consensus presented some residues predicted as $\alpha$-helical. This unique $\alpha$-helical turn could be a specific feature of the cationic channel subunits, since it does not appear in the AL2 sequence consensus, where the extended structure is consistently predicted.

The main immunogenic region (MIR) is located from the end of $H_B$ to the beginning of $E_3$ (Tzartos et al., 1990). This segment was already known to be exposed to the solvent since it is directly involved in numerous forms of the autoimmune disease myasthenia gravis (Tzartos et al., 1990). Accordingly, its central portion is predicted as totally accessible to the solvent.

The assignment of $H_B$ appeared consistent with all predictions except those of DSC for AL1 and AL2 as well as PHD on AL2 (only some residue predicted under $\beta$-strand state). The solvent accessibility pattern is more consistent with a $\beta$-strand in the carboxy-terminal portion. However, the glycosylation at $AL1_{102}$ (as well as at $AL1_{198}$ and $AL1_{242}$) implies that these residues are exposed to the exterior of the receptor. Since the residue $AL1_{104}$ is labeled by tubocurarine and presumably faces the binding site, a $\beta$-strand might cause steric hindrance between the ligand and the sugar, whereas an $\alpha$-helix would place the side chain of the two residues in opposite directions.

The structure of $E_2$ (three residues long) is not predicted by the analysis of AL2. It is the only structural element that differs between the two analyses. However, the assignment of $E_3$ is contradicted by a cross-linking experiment (Watty et al., 1998) showing that its two first residues should expose their side chain in the same direction.

$E_4$ is predicted to be completely buried. $E_5$ and $E_8$ are consistent with solvent accessibility "ebebebeb."

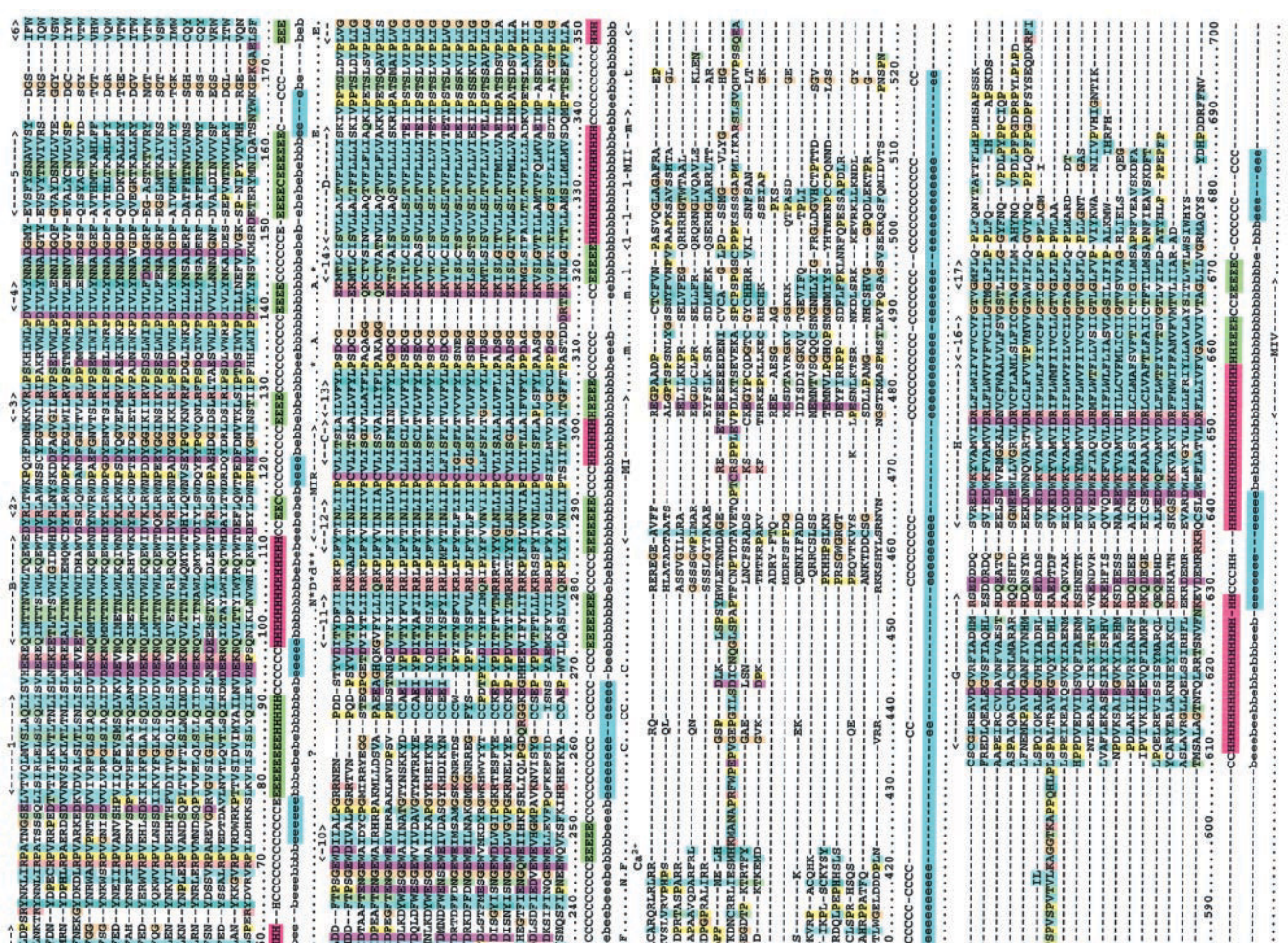

FIGURE 4 Alignment of the cationic channel subunits (AL1). The first line gives the limit of the refined secondary structures and their respective designation (letters for $\alpha$-helix and numbers for $\beta$-strand). The alignment, generated by ClustalX, follows. The colors are set according to the amino acid or the consensus at each position. For instance, a proline is always yellow, whereas a cysteine is purple if it is the consensus residue or can be blue if the consensus is a hydrophobic residue. Below the alignment and the ruler, the raw consensus secondary structure prediction is presented. The magenta and green boxes represent the refined predictions. Below the secondary structure prediction is the predicted accessibility to the solvent (b: buried, e: exposed). On the last line is reported the topology inferred from SSPCA output and experimental considerations. On this line are also reported the affinity labeling results (A–F represent the respective binding site segments) and glycosylation (N) natural or induced.

**TABLE 3 Summary of predicted structural element position and length**

| Helices | Position in AL1 (mature Gallus $\alpha$7) | Strands | Position |
|---|---|---|---|
| A | 50–61 (Phe-3–Asn-14) | 1 | 78–90 (Leu-28–Met-40) |
| B | 97–110 (Gln-47–Thr-60) | 2 | 113–115 (Tyr-63–Gln-65) |
| C | 295–300 (Leu-220–Ala-225) | 3 | 125–128 (Lys-75–Arg-78) |
| D | 324–337 (Val-245–Glu-258) | 4 | 139–142 (Ile-89–Tyr-92) |
| E | 348–355 (Leu-269–Ser-276) | 5 | 151–161 (Asp-100–Asn-110) |
| F | 385–400 (Pro-305–Leu-330) | 6 | 173–175 (Cys-115–Tyr-117) |
| G | 609–627 (Pro-408–Arg-425) | 7 | 186–191 (Tyr-128–Trp-133) |
| H | 637–656 (Ala-432–Val-451) | 8 | 200–203 (Asn-142–Phe-145) |
| | | 9 | 213–218 (Ser-154–Met-159) |
| | | 10 | 245–249 (Trp-173–Gly-177) |
| | | 11 | 272–278 (Ile-197–Met-203) |
| | | 12 | 284–290 (Tyr-209–Leu-215) |
| | | 13 | 301–305 (Leu-226–Leu-230) |
| | | 14 | 318–323 (Thr-244–Ile-243) |
| | | 15 | 356–373 (Thr-277–Tyr-294) |
| | | 16 | 657–662 (Phe-452–Ile-457) |
| | | 17 | 677–670 (Gly-462–Met-465) |

The position of the structures in the sequence of mature chick $\alpha$7 subunit is given for comparison with previous studies. The limits are included (i.e., 113–115 means 113–114–115).

The predictions of the $E_{12–15}$ and $H_{C-E}$ are probably less accurate than the extramembranous portions. Indeed, the secondary-structure prediction programs were not designed or tested with membrane proteins (see Discussion). The length of the predicted secondary structures varied considerably according to the set of sequences used. With AL2, $H_D$ is shorter (in MII), $H_E$ is longer, and $E_{15}$ shorter (in MIII). Finally, $H_F$ and $H_G$ are fully consistent with the solvent accessibility predictions "bbebbbebbbebbbebb" and "eebeebbebbebbbeeb," implying one face exposed, the other buried.

## 2D representation of the amino-terminal domain

Data may be added to the 1D structural assignments given by SSPCA. This defines an envelope of structural constraints (Fig. 5), which permits proposal of a 2D folding of the peptide chain. No data concerning the tertiary folding are included, since no $\beta$-$\beta$ interactions are known.

First, on the basis of electron microscopy images, we may locate the MIR at the distal end of the receptor, respectively to the membrane (Beroukhim and Unwin, 1995). As a consequence, $E_2$ and $E_3$ are also placed at the top of the fold. $E_{11}$ is likely to be close to the membrane since it is adjacent to MI (see below for the position of the transmembrane domains). Then, we may assume that each stretch of at least four residues predicted to be exposed to the solvent forms a loop at the surface of the subunit. This constraint implies bending of the 1D structure between $E_5$ and $E_6$, $E_8$ and $E_9$, $E_9$ and $E_{10}$, and $E_{10}$ and $E_{11}$. The beginning of $E_7$ and $E_8$ are linked by a disulfide bond, and are thus in close proximity. This disulfide bridge forces a new bend between $E_7$ and $E_8$. This so-called "Cys-loop" is the most conserved part of the LGIC subunit amino-terminal domains. Although half of it is not predicted to be folded into a periodic structure, we may reasonably hypothesize that the entire region adopts a strongly constrained conformation. Finally, a bend is introduced between $E_3$ and $E_5$ to respect the observed size of the subunit, which protrudes 60 Å from the membrane, with a diameter of ~40 Å.

Each subunit can then be artificially subdivided into two domains. One is formed by $H_A$, $E_1$, and $H_B$, the other by $E_{2–11}$. On the basis of the cryoelectron microscopy images of Unwin (1993b), $H_A$ and $H_B$ have been disposed perpendicularly to the membrane.

This representation is fully compatible with the body of experimental data concerning the nicotinic binding site. Indeed, affinity labeling and site-directed mutagenesis led to the identification of amino acids (see Table 1) that are distributed at the interface of the subunits on six different elements, referred to as A ($\alpha$7W85 and $\alpha$7Y92), B ($\alpha$7W148 and $\alpha$7Y150), and C ($\alpha$7Y187, $\alpha$7C189, $\alpha$7C190, and $\alpha$7Y194) for the "principal" component; and D ($\alpha$7W54), E ($\alpha$7L108, $\alpha$7N110, $\alpha$7Q116, and $\alpha$7L118), and F ($\alpha$7D163 and $\alpha$7E172) for the "complementary component" (note that all residues are quoted according to the mature chick $\alpha$7 subunit; please see the alignment for conversion. It does not mean that these residues have been identified only in or also in this subunit). Another residue has recently been identified on the complementary component ($\alpha$7T34). Since it is located in $E_1$, its position does not add further constraints on the 2D representation, though it possibly constrains the tertiary folding.

Affinity labeling experiments with toxin derivatives assigned the principal and complementary binding components to be carried by the clockwise and counterclockwise sides of the subunits, respectively, when the receptor is seen from the extracellular compartment (Machold et al., 1995).

The $H_AE_1H_B$ part must be folded onto the $E_{2–11}$ sheet in order to form a compact structure, contained in a 40 × 60

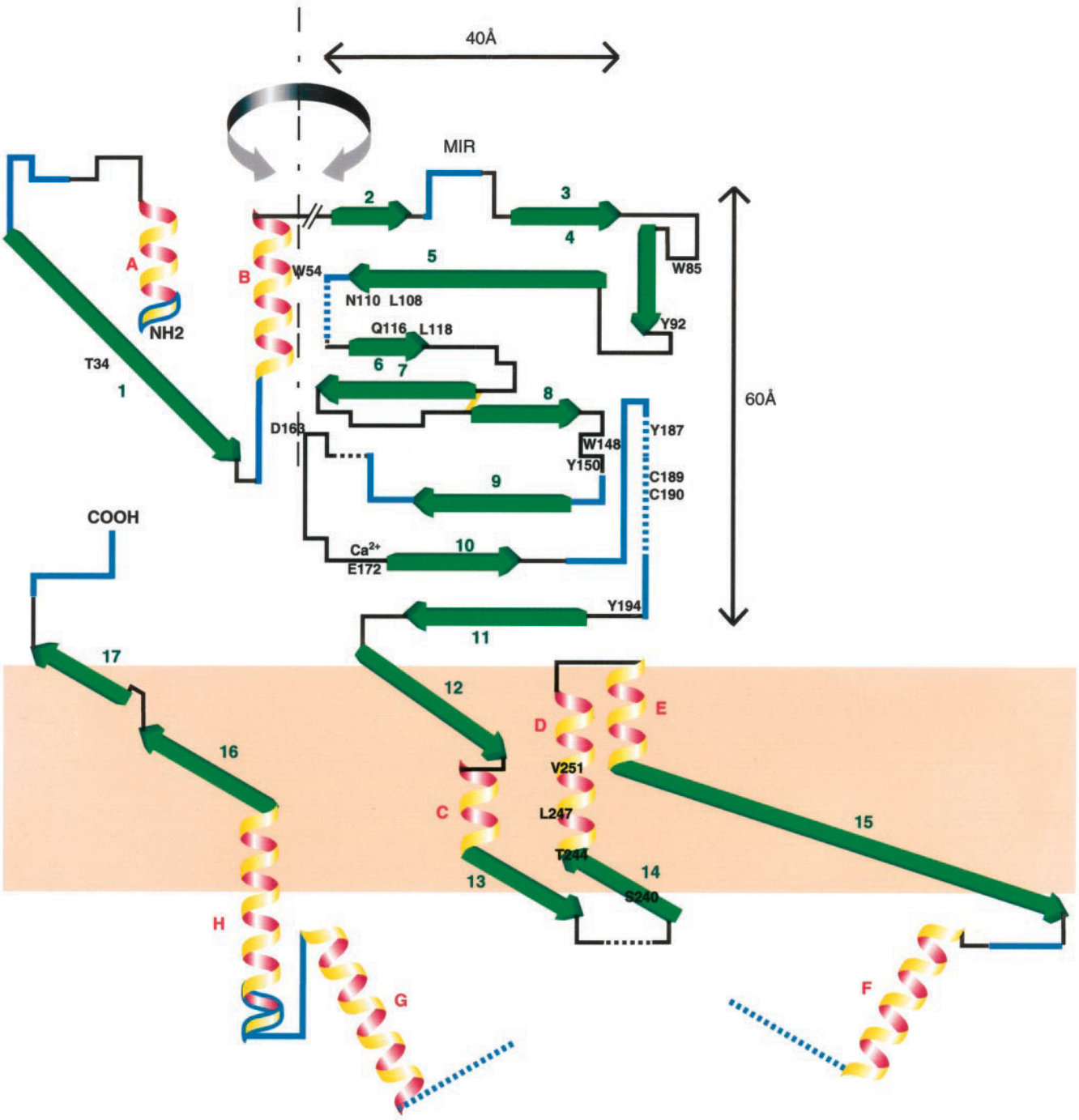


FIGURE 5 2D folding of a typical subunit. The perspective is from inside the pore, perpendicular to the membrane. The $\alpha$-helices and the length of $\beta$-strand are on the same scale. The length of the nonstructured part is approximately scaled. The gray double-arrow means that the two parts of the extracellular moiety have to be bent together. The blue segments represent the positions of the stretches (of length $>4$) of residues exposed to the solvent. The yellow link represents the disulfide bond. MIR: main immunogenic region. The amino acids identified by affinity labeling are noted (the numbering is that of mature chick $\alpha$7). Note that this fold is a 2D representation, and has nothing in common with a 3D model. Indeed, the $\beta$-strands are placed totally parallel and in a progressive order only for reasons of convenience. Note that the figure does not in any way imply any specific $\beta$-$\beta$ interaction. The positions of the strands are chosen arbitrarily, in the order of the primary sequence.

Å$^2$ surface, and to account for the possible contribution of residues homologous to mouse $\gamma$K34 to the active site. Yet, only a few data constrain the folding of the $H_AE_1H_B$ domain.

## The transmembrane and cytoplasmic domains

We used PHDhtm (Rost et al., 1995, 1996) to investigate the organization of membrane spanning segments. PHDhtm is

the only program that did not predict the signal peptide as transmembrane domain, probably because of its lack of conservation. In addition, it predicted the four transmembrane domains for each LGIC members. SSPCA provided the consensus of output from PHDhtm applied to all the sequences of AL1. The results, compiled in Table 5, yielded four transmembrane segments. The length of the consensus segments are 18 for MI, 17 for MII, 19 for MIII, and 17 for MIV. For comparison, four other programs were also used on $\alpha$1 and $\alpha$7. All of them predicted the four transmembrane segments of AL1 sequences, although in some cases other parts of the subunit were incorrectly predicted as crossing the membrane.

The length of the consensus transmembrane segments are smaller than depicted in the usual proposals (Popot and Changeux, 1984). However, this could be an artifact due to conservative prediction of PHDhtm. For comparison, four other programs were also used on $\alpha$1 and $\alpha$7. The results vary according to the method but also according to the sequences used. This fact supports the importance of using consensus of multiple analyses.

SSPCA predicts each of the transmembrane segments to fold in a mixed $\alpha$ helix/$\beta$ strand fashion, with almost no coiled structures. Fig. 5 shows an attempt to represent the transmembrane portion in 2D. Yet, since the present study gives no information about the precise orientation of the structures in the membrane, the represented angles are arbitrary, except in the case of the helix present in the MII segment shown to be orientated rather perpendicular to the membrane. In addition, the length of the predicted structures in the membrane is poorly accurate.

Except for $H_F$ and $H_G$, the cytoplasmic domain is predicted as totally accessible to solvent and in a nonperiodic structure. The solvent accessibility pattern of the two helices suggests that they possess one face buried and another exposed to the solvent (see Discussion).

## DISCUSSION

Previous works have shown that the accuracy of secondary structure predictions increases with the combination of several independent algorithms (Biou et al., 1988; Nishikawa and Ooi, 1986; Zhang et al., 1992). As reported here, in order to determine the best available prediction of the nAChR subunits' secondary structure, we integrated the results of several third-generation programs, using the information from a set of aligned homologous sequences. These programs were selected on the basis of their recognized efficiency on test sets of proteins with known secondary structure (Rost and Sander, 1994a) or during blind predictive situations (King, 1996; Rost, 1997). Moreover, each program was applied on every sequence of the alignments in order to increase the signal/noise ratio.

Two main ab initio predictions have been reported for nAChRs in the past two decades. Finer-Moore and Stroud (1984) used the algorithm of Garnier et al. (1978) for the extramembranous regions and an analysis (by Fourier transformations) of hydrophobicity periodicity for the putative transmembrane regions. Recently, Ortells (1997) presented a secondary structure prediction based on a Chou and Fasman-like algorithm (Chou and Fasman, 1978). The main difference between the initial method and the one used by Ortells resides in the definition of the secondary structure initiators. Instead of being predicted solely by the sequence (via statistical tables) as in the Chou and Fasman algorithm, these initiators were determined as follows: an initiator was defined as a residue that is constantly predicted in the same state, across different sets of LGIC subunit sequences, analyzed by first- and second-generation algorithms. Another difference resides in the fact that the propagation from the initiators was unidirectional (in Ortells, 1997), from the amino-terminal to the carboxy-terminal, while it is bi-directional in Chou and Fasman (1978). The expected prediction accuracy has already been discussed elsewhere (see Kabsch and Sander, 1983; and Nishikawa, 1983 for initial assessment, and Rost and Sander, 1994a; 1996 for recent reviews), but the difference of expected accuracy between these pioneering works and our own may reach 20%. On an identical test set, Chou and Fasman reached 49% in $Q_3$, whereas PHD2 reached 72.5% (Rost and Sander, 1994a).

**TABLE 4 Experimental and predicted $\alpha$-helix and $\beta$-strand content in an entire subunit**

| Source | Helix Content | Strand Content | Ratio |
|---|---|---|---|
| Yager et al., 1984 | 39% | 33% | 1.18 |
| Butler and McNamee, 1993 | 18.7% | 42% | 0.45 |
| Méthot et al., 1994 | 39% | 35% | 1.11 |
| West et al., 1997 | 48% | 26% | 1.85 |
| Mean of experiments | 36.2% | 34% | 1.15 |
| Finer-Moore and Stroud, 1984 | 44% | 27% | 1.63 |
| Ortells, 1997 | 29.7% | 24.9% | 1.19 |
| SSPCA consensus | 25.8% | 22.3% | 1.16 |
| Corrected consensus | 24.2% | 22.5% | 1.07 |

Note that Butler and McNamee (1993) is clearly an outlier, decreasing the mean $\alpha$-helix content, and increasing the mean $\beta$-strand content.

### $\alpha$-Helix and $\beta$-strand contents

The helix and strand content of the entire nAChR was measured by several groups using different spectroscopic measurement methods (Butler and McNamee, 1993; Méthot et al., 1994; Moore et al., 1974; Yager et al., 1984). The results showed a high variability, which cannot be due solely to the differences of receptor environment. Indeed, inferred helix content varied from 18.7% (Butler and McNamee, 1993) to 48% (West et al., 1997), inferred strand content (without $\beta$-turn) varied from 26% (West et al., 1997) to 42% (Butler and McNamee, 1993), and the calculated helix/strand ratio varied from 0.45 (Butler and McNamee, 1993) to 1.85 (West et al., 1997), with Méthot et al. (1994) and Yager et al. (1984) finding intermediate values of 1.11 and 1.18) (see Table 4). The corrected SSPCA

**TABLE 5 Determination of the transmembrane segment positions**

| | | MI | MII | MIII | MIV |
|---|---|---|---|---|---|
| Original pattern (presented in Ortells and Lunt, 1996 for $\alpha$7 and Popot and Changeux, 1984 for $\alpha$1) | $\alpha$7gg | 283–307 (208–232) | 317–337 (238–258) | 353–374 (274–295) | 651–675 (446–470) |
| | $\alpha$1tc | 281–308 (210–236) | 310–339 (239–265) | 348–375 (273–300) | 650–677 (407–433) |
| DAS (http://www.biokemi.su.se/~server/DAS/): Cserzo et al., 1997 | $\alpha$7gg | 287–309 (212–234) | 320–340 (241–261) | 352–373 (273–294) | 51–672 (446–467) |
| | $\alpha$1tc | 281–307 (210–235) | 319–339 (245–265) | 349–372 (274–297) | 647–671 (404–427) |
| Tmpred (http://ulrec3.unil.ch/software/TMPRED_form.html) (Hofman and Stoffel, 1993) | $\alpha$7gg | 290–309 (215–234) | 321–340 (242–261) | 354–372 (275–293) | 649–673 (448–468) |
| | $\alpha$1tc | 281–307 (210–235) | 319–339 (245–264) | 353–371 (278–296) | 652–672 (409–428) |
| Toppred2 (http://www.biokemi.su.se/~server/toppred2/) (von Heijne, 1992) | $\alpha$7gg | 289–309 (214–234) | 319–339 (240–260) | 353–373 (274–294) | 649–673 (448–468) |
| | $\alpha$1tc | 288–309 (217–237) | 317–337 (243–263) | 352–372 (277–297) | 651–672 (408–428) |
| SOSUI (http://www.tuat.ac.jp/~mitaku/adv_sosui/) (Hirokawa et al., 1998) | $\alpha$7gg | 286–308 (211–233) | 319–341 (240–262) | 351–372 (272–293) | 650–672 (445–467) |
| | $\alpha$1tc | 284–307 (213–235) | 319–341 (245–267) | 349–371 (274–296) | 649–672 (406–428) |
| PHDhtm (Rost et al., 1996) | $\alpha$7gg | 285–302 (210–227) | 322–339 (243–260) | 351–368 (272–289) | 647–668 (446–463) |
| | $\alpha$1tc | 284–303 (213–231) | 320–338 (246–264) | 350–368 (275–293) | 650–668 (407–424) |
| SSPCA consensus (on each AL1 member) | $\alpha$7gg | 285–303 (210–228) | 322–339 (243–260) | 349–368 (270–289) | 651–668 (446–463) |
| | $\alpha$1tc | – (214–231) | – (248–265) | – (274–393) | – (408–424) |

Note: the default parameters were used in all the following programs. A more careful usage would probably enhance the expected accuracy a bit. The numbering is the one of AL1 (between brackets is the correspondence with the mature peptides).

consensus yielded slightly lower values of helix and strand contents, although the ratio is consistent with the mean of experimentally determined ratios. In the amino-terminal portion (as defined by West et al., 1997 and not by our transmembrane segment determination), our consensus gives an equivalent predicted helix content (13.7% vs. 12%) and less strand (31.7% vs. 51%) compared to the experimentally observed one in the unique study of West et al. (1997).

## Comparison with other predictions of the amino-terminal domain

At the level of the extracellular amino-terminal domain, all approaches predicted a structure mainly folded in $\beta$-strand (Fig. 6 *A*). However, the position of the structures, as well as their number, differ considerably among the different studies. The high $\beta$ content is also consistent with the cryoelectron microscopy images (although three helices were proposed in these latter investigations; Unwin, 1993b, 1996). The structures predicted by Ortells (1997) are longer than those presented in the present work and longer than the value observed in the PDB. Notably, the two large $\alpha$-helices predicted in the amino-terminal half of the extracellular portion are 20 aa long, whereas we predict 12 and 14 aa and the PDB average is 9. Also, the mean length of $\beta$-strand is 7.2 in Ortells (1997), 5.8 in the present work, and 5.1 in the PDB. These discrepancies are likely due to the method Ortells used to propagate the structural elements. When initiated, each element is extended forward until a different initiator or a proline or a glycine is reached.

Fig. 5 *A* also provides a comparison with the secondary structures derived from threading methods (Gready et al., 1997; Tsigelny et al., 1997). In this case, not only the lengths of the motifs, but also their positions in the sequence are very different.

## Comparison with other predictions of the membrane-spanning segments

The location of the four putative transmembrane segments was originally performed by hydropathy plot analysis. This method, though of great interest and easy to use, does not apply satisfactorily in the case of membrane channels. Indeed, the residues lining the pore in the open state are not anticipated to be hydrophobic. Moreover, in a protein with multiple membrane crossings, such as the nAChR, the internal transmembrane segments may be isolated from the lipid environment. In addition, some hydrophobic stretches can be external to the membrane (in close proximity to, or embedded in, the core protein). As a consequence, some transmembrane segments were not correctly predicted. For instance, for the rat glycine $\alpha$1 subunit, the program SOSUI (Hirokawa et al., 1998), based on amino acid physical properties, did not predict the MII and MIV segments as transmembrane units, nor did the program TMpred (Hofman and Stoffel, 1993), based on the comparison with a database of known transmembrane segments.

The original predictions vary from one author to another (Fig. 7). The membrane-spanning position is set by PHDhtm with 95% accuracy. Such a precision is superior to the original variations.

Some structural prediction have already been made for the transmembrane domain based on analogy arguments. Unwin (1993b) suggested, on the basis of his images from electron microscopy, that the transmembrane region of the nAChR could have a folding similar to that of some pentameric enterotoxin domains. Ortells and Lunt (1996) further exploited this idea to model part of the LGIC transmembrane region based on the crystallographic structure of the *Escherichia coli* heat-labile enterotoxin domain B (Sixma et al., 1993). The resulting model presents a mixed $\alpha/\beta$ secondary structure, where MII is all-$\alpha$, MI is all-$\beta$, and


FIGURE 6 (*A*) Comparison of our secondary structure prediction with those of Gready et al. (1997), Tsigelny et al. (1997), and Ortells (1997). All the predictions are projected on the amino-terminal part of *Torpedo* α1. Numbering of the residues is that of mature *Torpedo* α1 and that of AL1. (*B*) Comparison of our secondary structure prediction with that of Ortells and Lunt (1996) based on analogy with enterotoxin. The predictions are projected on the transmembrane part of *Gallus* α7. Numbering of the residues is that of mature *Gallus* α7 and that of AL1.

MIII is α/β (Fig. 6 *B*), the MIII α-helical region being a posteriori added to the template. Several remarks can be make about this study apart from the fact that the template was never found, up to now, by any threading algorithm. First of all, the enterotoxin is not an integral membrane protein, and thus may not be an adequate template for the nAChR transmembrane domain. Ortells and Lunt (1996) removed the first strand that interacts with the fifth. The resulting template might then be less stable, one sheet being composed of only two antiparallel strands. As stated by the authors, the further addition of MIII and MIV, modeled as helices (partially for MIII) may result in a segregation of the enterotoxin-modeled moiety from the lipids. The secondary structure predictions presented here do not agree with those proposed by Ortells and Lunt (1996; Fig. 6 *B*). A three-state comparison between this study and the present prediction gives only 33% of identical residues.

## The nicotinic ligand-binding site

The 2D representation accounts for the basic information concerning the ligand binding site for ACh and competitive antagonists. Secondary structure predictions suggest that whereas binding elements B, F, and C are carried by segments without regular structural patterns, binding elements A, E, and D are at least partially carried by structured segments.

At the level of the complementary component, the affinity-labeled δW57 ($AL1_{104}$) is located near the center of helix B. Mutations at position $AL1_{106}$ have been shown to modulate agonist and antagonist pharmacology (Chiara and Cohen, 1997; Corringer et al., 1995; Harvey and Luetje, 1996). The side chains of residues $AL1_{102}$ and $AL1_{104}$, however, point outward from opposite faces of the helix, implying that $AL1_{104}$ mediates its effects indirectly, possibly through local alteration of the structure. At the level of element E, two successive β-strands are predicted, $E_5$ carrying the identified mouse γS111 ($AL1_{161}$) (Sine et al., 1995), and $E_6$ carrying mouse γY117 ($AL1_{175}$). One possibility could be that these β-strands interact in an antiparallel β-sheet, which would direct the side chains of these binding residues in the same direction and in close proximity. Finally, the segment carrying element F has been shown to contain the calcium binding site involved in agonist potentiation. The predicted arrangement of this region without a regular structure is consistent with the notion that the seg-

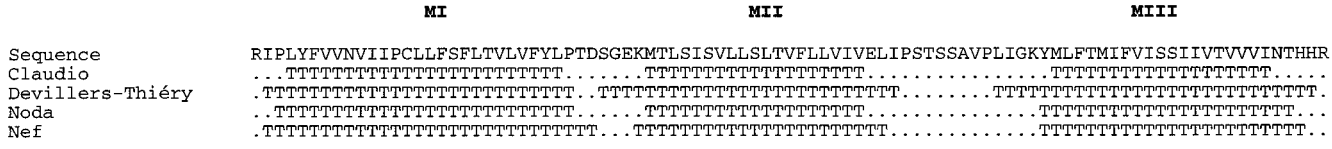


FIGURE 7 Comparison of the transmembrane segments as originally suggested by Claudio et al. (1983) (actual analysis is that of *Torpedo* γ), Devillers-Thiéry et al. (1983), Noda et al. (1983), and Nef et al. (1988) (actual analysis is that of chick γ and δ).

ment $AL1_{234}$–$AL1_{245}$ folds into a specific pocket that constitutes a calcium binding site, as observed for the corresponding synthetic peptide (Galzi et al., 1996).

At the level of the principal component, mutagenesis experiments have shown that several mutations, located at the vicinity of labeled residues from elements B and C, profoundly altered the pharmacological properties (regions $AL1_{210}$–$AL1_{214}$ and $AL1_{256}$–$AL1_{259}$) (Corringer et al., 1998). Since the entire corresponding regions are predicted to lack a regular secondary structure, they may fold into loops, such that the mutations could possibly alter agonist binding indirectly, through structural reorganization of these putatively flexible segments.

## Transmembrane segments as an $\alpha/\beta$ structure

Each transmembrane segment of the receptor is predicted to fold in a mixed $\alpha/\beta$ structure. This prediction should be taken with extreme caution, since, as noted above, the programs used were not designed to work on membrane proteins. Prediction methods based on analyses of globular proteins could incorrectly predict strands in helical transmembrane regions.

Direct transitions are seen at the end of MI, MIII, and MIV. Such transitions are impossible following a helix of more than four residues. Due to the low reliability of the predictions in these regions, a small hinge could in fact link the $\alpha$-helices and the following $\beta$-strands.

Also, affinity labeling experiments with a radioactive hydrophobic probe support an organization of the MIII and MIV transmembrane segments in $\alpha$-helix (Blanton and Cohen, 1994). MIII was predicted to be $\alpha$-helical until $AL1_{362}$ ($\alpha$7F283), while the $H_E$ was predicted to reach only $AL1_{355}$ ($\alpha$7S276), and MIV was predicted to be $\alpha$-helical until $AL1_{668}$ ($\alpha$7I463), whereas $H_H$ is predicted to reach only $AL1_{657}$ ($\alpha$7F452).

At the level of the MII segment, known to face the lumen of the ion channel, our predictions could lead to a reconsideration of the currently accepted architecture of the ion pathway. MII is predicted here to start at the level of amino acid $AL1_{323}$, four residues after the standard model. In addition, the MII helix is predicted to be slightly shorter. Much of the data coming from affinity labeling and site-directed mutagenesis experiments are readily represented by a helical structure (Akabas et al., 1994; Revah et al., 1990). However, recent results (Wilson and Karlin, 1998) support an elongated strand for the short segment-spanning residues $AL1_{310}$ ($\alpha$7S335) to $AL1_{319}$ ($\alpha$7S240). Moreover, it is thought that MI and MII are in close proximity (Akabas and Karlin, 1995). Consequently, the cytoplasmic portion linking MI and MII is predicted to be longer, and could fold into a $\beta$-hairpin ($E_{13}$–$E_{14}$), the length of loop linking the strands being variable according to the subunit. Recent mutagenesis experiments from this laboratory point to a major contribution of the center of this cytoplasmic portion to the selectivity filter of the ion channel. Furthermore, it was found that its conformation, rather than its precise amino acid sequence, had a critical effect on the selectivity properties of the ion channel (Corringer et al., 1999). This large cytoplasmic region could thus fold in such a way that some carbonyl of the peptide backbone would be exposed in the correct geometry for dehydration of specific ions, as observed in the case of a bacterial potassium channel (Doyle et al., 1998).

## The cytoplasmic portion and the oligomerization

$H_F$ and $H_G$ are predicted to be amphipathic, with one face exposed to the solvent and the other buried. The maximum hydrophobic moment (as determined with the program MOMENT of the Wisconsin Package (Devereux et al., 1984) with a window of eight residues) is 0.19 for $H_F$ (low) and 0.57 for $H_G$ (high). In addition, both helices present a leucine-zipper signature (on 79 sequences: at $AL1_{393}$, 61L; $AL1_{400}$, 62L, 14M; $AL1_{611}$, 30L, 6M; $AL1_{618}$, only 2L, but 21I and a conserved hydrophobic position in $AL1_{615}$; $AL1_{625}$, 17L, 30M, and a conserved hydrophobic position in $AL1_{622}$). These two cytoplasmic helices could interact in a coiled-coil arrangement, within the subunit or even between subunits. This motif could be critical for the oligomerization process. Indeed, Yu and Hall (1994) have demonstrated that

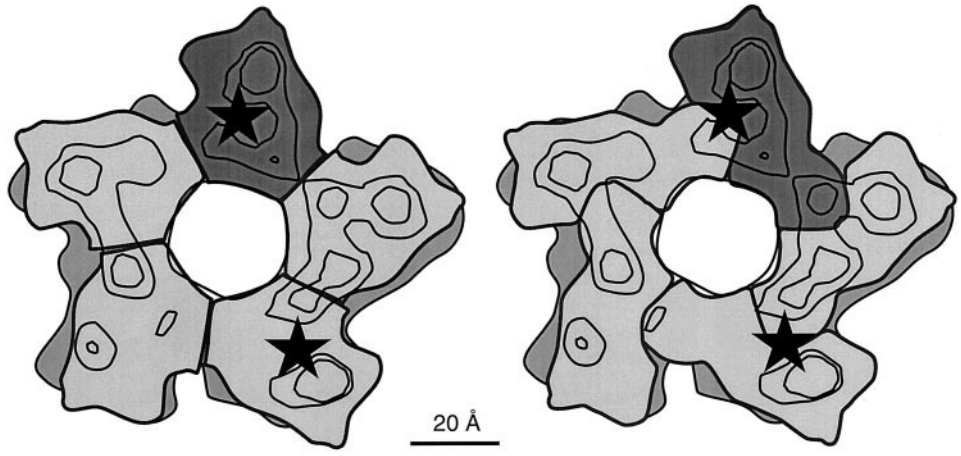


FIGURE 8 Hypothetical model of the pentameric arrangement of the subunits around the symmetry axis. The left image is based on the classical view of rod-shaped subunits. The right image is based on the present idea of flat-shaped subunits. The black curves are adapted from the isodensity lines of Unwin (1993a; Fig. 8). The stars represent the bungarotoxin binding site.

two deletions of amino acids belonging to $H_F$ and $H_G$ imply their intervention in the formation of the pentamer.

### Structure of the pentameric protein

On the basis of the predicted 2D representation, we may propose a hypothetical model for the assembly of the five subunits into the receptor oligomer. Each subunit is usually viewed as a vertical rod, with five of them forming the receptor (Unwin, 1993b). However, in volume reconstruction from electron diffraction images, the receptor molecule exhibits a clockwise torsion, each density group turning around the symmetry axis (Toyoshima and Nigel, 1990). Based on the 2D representation presented above, we may speculate that the extracellular portion of each subunit is not rod-shaped, but more flattened (Fig. 8). Its width would then be of the order of 40 Å and not of 25 Å. The agonist binding site would remain located between the three densities hypothesized to be intrasubunit $\alpha$-helices by Unwin (1993a). But in this reconstruction, which is supported by several lines of experimental evidence, the site would be placed at the interface between two subunits. At variance with the hypothesis of Unwin (1993a), in our view only the two main densities would correspond to $\alpha$-helices, but would belong to different subunits. This representation emphasizes the fundamental asymmetry characteristic in each individual "protomeric" subunit of the superfamily within the symmetrical oligomer (Changeux and Edelstein, 1998; Monod et al., 1965).

## CONCLUSIONS

We have presented novel secondary structure predictions of a typical nAChR subunit on the basis of an analysis of primary sequence data using a combination of third-generation algorithms. These predictions could serve as the basis for fold recognition methods. We incorporated additional predicted structural information, such as solvent accessibility, as well as available experimental data in order to formulate a 2D representation with the minimum number of unverified hypotheses, into the secondary structure assignments. This 2D representation may also serve as a framework to propose new experimental approaches for mutagenesis and construction of chimeric proteins within the superfamily to further relate the 3D organization of the receptor molecule and its physiological and pharmacological properties.

We thank Marc Delarue, Stuart Edelstein, and Clément Léna for critical reading of the manuscript, and Burkhard Rost for his comments.

This work was supported by the Collège de France, the Association Française Contre les Myopathies, Reynolds Pharmaceuticals, and the European Union (Biotech contract).

## REFERENCES

Akabas, M. H., and A. Karlin. 1995. Identification of acetylcholine receptor channel-lining residues in the M1 segment of the $\alpha$-subunit. *Biochemistry*. 34:12496–12500.

Akabas, M. H., C. Kaufmann, P. Ardeacon, and A. Karlin. 1994. Identification of acetylcholine receptor channel-lining residues in the entire M2 segment of the $\alpha$ subunit. *Neuron*. 13:919–927.

Basus, V. J., G. Song, and E. Hawrot. 1993. NMR solution of an $\alpha$-bungarotoxin/nicotinic receptor peptide complex. *Biochemistry*. 32: 12290–12298.

Bechade, C., and A. Triller. 1994. The inhibitory neuronal glycine receptor. *Bioessays*. 16:735–744.

Beroukhim, R., and N. Unwin. 1995. Three-dimensional location of the main immunogenic region of the acetylcholine receptor. *Neuron*. 15: 323–331.

Biou, V., J. Gibrat, J. Levin, B. Robson, and J. Garnier. 1988. Secondary structure prediction: combination of three different methods. *Protein Eng*. 2:185–191.

Blanton, M. P., and J. B. Cohen. 1994. Identifying the lipid-protein interface of the *Torpedo* nicotinic acetylcholine receptor: secondary structure implications. *Biochemistry*. 33:2859–2872.

Bormann, J., and A. Feigenspan. 1995. $GABA_C$ receptors. *Trends Neurosci*. 18:515–519.

Bowie, J. U., R. Lüthy, and D. Eisenberg. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. 253:164–170.

Butler, D. H., and M. G. McNamee. 1993. FITR analysis of nicotinic acetylcholine receptor secondary structure in reconstituted membranes. *Biochim. Biophys. Acta*. 1150:17–24.

Cartaud, J., L. Beneditti, J. B. Cohen, J.-C. Meunier, and J.-P. Changeux. 1973. Presence of a lattice structure in membrane fragments rich in nicotinic receptor protein from the electric organ of *Torpedo marmorata. FEBS Lett*. 33:109–113.

Changeux, J.-P., and S. Edelstein. 1998. Allosteric receptors after 30 years. *Neuron*. 21:959–980.

Chiara, D. C., and J. B. Cohen. 1997. Identification of amino acids contributing to high and low affinity d-tubocurarine sites in the *Torpedo* nicotinic acetylcholine receptor. *J. Biol. Chem*. 272:32940–32950.

Chothia, C., and A. M. Lesk. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J*. 5:823–826.

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

Claudio, T., M. Ballivet, J. Patrick, and S. Heinemann. 1983. Nucleotide and deduced amino acid sequences of *Torpedo californica* acetylcholine receptor $\gamma$ subunit. *Proc. Natl. Acad. Sci. USA*. 80:1111–1115.

Cockcroft, V. B., D. J. Osguthorpe, E. A. Barnard, A. E. Friday, and G. G. Lunt. 1992. Ligand-gated ion channels, homology and diversity. *Mol. Neurobiol.* 4:129–169.

Corringer, P.-J., S. Bertrand, S. Bohler, S. J. Edelstein, J.-P. Changeux, and D. Bertrand. 1998. Critical elements determining diversity in agonist binding and desensitization of neuronal nicotinic acetylcholine receptors. *J. Neurosci*. 18:648–657.

Corringer, P. J., S. Bertrand, J. L. Galzi, A. Devillers-Thiéry, J. P. Changeux, and D. Bertrand. 1999. Mutational analysis of the charge selectivity filter of the $\alpha$7 nicotinic acetylcholine receptor. *Neuron*. (in press).

Corringer, P., J. Galzi, J. Eisele, S. Bertrand, J. Changeux, and D. Bertrand. 1995. Identification of a new component of the agonist binding site of the nicotinic $\alpha$7 homooligomeric receptor. *J. Biol. Chem*. 279: 11749–11752.

Cserzo, M., E. Wallin, I. Simon, G. von Heijne, and A. Elofsson. 1997. Prediction of transmembrane alpha-helices in procaryotic membrane proteins: the dense alignment surface method. *Protein Eng*. 10:673–676.

Czajkowski, C., C. Kaufman, and A. Karlin. 1993. Negatively charged amino acid residues in the nicotinic receptor $\delta$ subunit that contribute to the binding of acetylcholine. *Proc. Natl. Acad. Sci. USA*. 90:6285–6289.

Dennis, M., J. Giraudat, F. Kotzyba-Hibert, M. Goeldner, C. Hirth, J. Chang, C. Lazure, M. Chrétien, and J. Changeux. 1988. Amino acids of the *Torpedo marmorata* acetylcholine receptor $\alpha$ subunit labeled by a photoaffinity ligand for the acetylcholine binding site. *Biochemistry*. 27:2346–2357.

Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res*. 12: 387–395.

Devillers-Thiéry, A., J. Giraudat, M. Bentaboulet, and J.-P. Changeux. 1983. Complete mRNA sequence of the acetylcholine binding $\alpha$-subunit of *Torpedo receptor* acetylcholine receptor: a model for the transmembrane organization of the polypeptide chain. *Proc. Natl. Acad. Sci. USA*. 80:2067–2071.

Doyle, D., J. Cabral, R. Pfuetzner, A. Kuo, J. Gulbis, S. Cohen, B. Chait, and R. MacKinnon. 1998. The structure of the potassium channel: molecular basis of $K^+$ conduction and selectivity. *Science*. 280:69–77.

Eisele, J. L., S. Bertrand, J. L. Galzi, A. Devillers-Thiéry, J. P. Changeux, and D. Bertrand. 1993. Chimeric nicotinic serotonergic receptor combines distinct ligand binding and channel specificities. *Nature*. 366: 479–483.

Finer-Moore, J., and R. M. Stroud. 1984. Amphipathic analysis and possible formation of the ion channel in an acetylcholine receptor. *Proc. Natl. Acad. Sci. USA*. 81:155–159.

Flores, T., C. Orengo, D. Moss, and T. JM. 1991. Comparison of conformational characteristics in structurally similar protein pairs. *Protein Sci*. 2:1811–1826.

Frishman, D., and P. Argos. 1996. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng*. 9:133–142.

Frishman, D., and P. Argos. 1997. Seventy-five percent accuracy in protein secondary structure prediction. *Proteins*. 27:329–335.

Galzi, J.-L., S. Bertrand, P.-J. Corringer, J.-P. Changeux, and D. Bertrand. 1996. Identification of calcium binding sites that regulate potentiation of a neuronal nicotinic acetylcholine receptor. *EMBO J*. 15:5824–5832.

Galzi, J., and J. Changeux. 1994. Neurotransmitter-gated ion channels as unconventional allosteric proteins. *Curr. Opin. Struct. Biol*. 4:554–565.

Galzi, J., F. Revah, D. Black, M. Goeldner, and C. C. Hirth, JP. 1990. Identification of a novel amino acid $\alpha$-tyrosine 93 within the cholinergic ligand-binding sites of the acetylcholine receptor by photoaffinity labeling. Additional evidence for a three-loop model of the cholinergic ligand-binding sites. *J. Biol. Chem*. 265:10430–10437.

Garnier, J., D. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol*. 120:97–120.

Gibrat, J., J. Garnier, and B. Robson. 1987. Further developments of protein secondary structure prediction using information theory. *J. Mol. Biol*. 198:425–443.

Gready, J. E., S. Ranganathan, P. R. Schofield, Y. Matsuo, and K. Nishikawa. 1997. Predicted structure of the extracellular region of ligand-gated ion-channel receptors shows SH2-like and SH3-like domains forming the ligand-binding site. *Protein Sci.* 6:983–998.

Harvey, S. C., and C. W. Luetje. 1996. Determinants of competitive antagonist sensitivity on neuronal nicotinic receptor $\beta$3 subunits. *J. Neurosci*. 16:3798–3806.

Hirokawa, T., S. Boon-Chieng, and S. Mitaku. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*. 14:378–379.

Hofman, K., and W. Stoffel. 1993. TMBASE: a database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler*. 374:166.

Hucho, F., V. I. Tsetlin, and J. Machold. 1996. The emerging three-dimensional structure of a receptor, the nicotinic acetylcholine receptor. *Eur. J. Biochem*. 239:539–557.

Kabsch, W., and C. Sander. 1983. How good are predictions of protein secondary structure? *FEBS Lett*. 155:179–182.

King, R. D. 1996. Prediction of secondary structure. *In* Protein Structure Prediction, Vol. 170. M. J. Sternberg, editor. IRL Press, Oxford. 79–99.

King, R. D., and M. J. E. Sternberg. 1996. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci*. 5:2298–2310.

Kistler, J., and R. M. Stroud. 1981. Crystalline arrays of membrane-bound acetylcholine receptor. *Proc. Natl. Acad. Sci. USA*. 78:3678–3682.

Lal, R., and L. Yu. 1993. Atomic force microscopy of cloned nicotinic acetylcholine receptor expressed in *Xenopus* oocytes. *Proc. Natl. Acad. Sci. USA*. 90:7280–7284.

Langosch, D., L. Thomas, and H. Betz. 1988. Conserved quaternary structure of ligand-gated ion channels: the postsynaptic glycine receptor is a pentamer. *Proc. Natl. Acad. Sci. USA*. 85:7394–7398.

Le Novère, N., and J.-P. Changeux. 1995. Molecular evolution of the nicotinic acetylcholine receptor subunit family: an example of multigene family in excitable cells. *J. Mol. Evol*. 40:155–172.

Levin, J., B. Robson, and J. Garnier. 1986. An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS Lett*. 205:303–308.

MacDonald, R. L., and R. W. Olsen. 1994. $GABA_A$ receptor channels. *Annu. Rev. Neurosci*. 17:569–602.

Machold, J., C. Weise, Y. Utkin, V. Tsetlin, and F. Hucho. 1995. The handedness of the subunit arrangement of the nicotinic acetylcholine receptor from *Torpedo californica. Eur. J. Biochem*. 234:427–430.

Marchler-Bauer, A., M. Levitt, and S. H. Bryant. 1997. A retrospective analysis of CASP2 threading predictions. *Proteins*. (Suppl.)1:83–91.

Méthot, N., M. P. McCarthy, and B. J. E. 1994. Secondary structure of the nicotinic acetylcholine receptor: implications for structural models of a ligand-gated ion channel. *Biochemistry*. 33:7709–7717.

Monod, J., J. Wyman, and J.-P. Changeux. 1965. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol*. 12:88–118.

Moore, W., L. Holliday, D. Puett, and R. Brady. 1974. On the conformation of the acetylcholine receptor protein from *Torpedo nobiliana*. *FEBS Lett*. 45:145–149.

Nayeem, N., T. Green, I. Martin, and E. Barnard. 1994. Quaternary structure of the native $GABA_A$ receptor determined by electron microscopic image analysis. *J. Neurochem*. 62:815–818.

Nef, P., C. Oneyser, C. Alliod, S. Couturier, and M. Ballivet. 1988. Genes expressed in the brain define three distinct neuronal nicotinic acetylcholine receptors. *EMBO J.* 7:595–601.

Nishikawa, K. 1983. Assessment of secondary-structure prediction of proteins: comparison of computerized Chou-Fasman method with others. *Biochim. Biophys. Acta*. 748:285–299.

Nishikawa, K., and T. Ooi. 1986. Amino acid sequence homology applied to the prediction of protein secondary structures, and joint prediction with existing methods. *Biochim. Biophys. Acta*. 871:45–54.

Noda, M., H. Takahashi, T. Tanabe, S. Kikyotani, Y. Furutani, T. Hirose, H. Takashima, S. Inayama, T. Miyata, and C. Numa. 1983. Structural homology of *Torpedo californica* acetylcholine receptor subunits. *Nature*. 302:528–532.

Ortells, M. O. 1997. Prediction of the secondary structure of the nicotinic acetylcholine receptor nontransmembrane regions. *Proteins*. 29: 391–398.

Ortells, M. O., and G. G. Lunt. 1995. Evolutionary history of the ligand-gated ion-channel superfamily of receptors. *Trends Neurosci*. 18: 121–126.

Ortells, M. O., and G. G. Lunt. 1996. A mixed helix-beta-sheet model of the transmembrane region of the nicotinic acetylcholine receptor. *Protein Eng*. 9:51–59.

Paas, Y. 1998. The macro and microarchitectures of the ligand-binding domain of glutamate receptors. *Trends Neurosci*. 21:117–125.

Popot, J.-L., and J.-P. Changeux. 1984. Nicotinic receptor of acetylcholine: structure of an oligomeric integral membrane protein. *Physiol. Rev.* 64:1162–1239.

Prince, R. J., and S. M. Sine. 1996. Molecular dissection of subunit interfaces in the acetylcholine receptor. *J. Biol. Chem*. 271: 25770–25777.

Revah, F., J.-L. Galzi, J. Giraudat, P.-Y. Haumont, F. Lederer, and J.-P. Changeux. 1990. The noncompetitive blocker [$^3$H]chlorpromazine labels three amino acids of the acetylcholine receptor $\gamma$ subunit: implications for the $\alpha$-helical organization of regions MII and for the structure of the ion channel. *Proc. Natl. Acad. Sci. USA*. 87:46765–4679.

Rost, B. 1997. Better 1D predictions by experts with machines. *Proteins*. (Suppl.)1:192–197.

Rost, B., R. Casadio, and P. Fariselli. 1996. Refining neural network predictions for helical transmembrane proteins by dynamic programming. *The Fourth International Conference on Intelligent Systems for Molecular Biology.* A. Press, editor. St. Louis, Missouri. 192–200.

Rost, B., R. Casadio, P. Fariselli, and C. Sander. 1995. Prediction of helical transmembrane segments at 95% accuracy. *Protein Sci*. 4:521–533.

Rost, B., and C. Sander. 1993a. Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl. Acad. Sci. USA*. 90:7558–7562.

Rost, B., and C. Sander. 1993b. Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol*. 232:584–599.

Rost, B., and C. Sander. 1994a. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*. 19: 55–72.

Rost, B., and C. Sander. 1994b. Conservation and prediction of solvent accessibility in protein families. *Proteins*. 20:216–226.

Rost, B., and C. Sander. 1996. Bridging the protein sequence-structure gap by structure predictions. *Annu. Rev. Biophys. Biomol. Struct*. 25: 113–136.

Russel, R., and G. Barton. 1993. The limits of protein structure prediction accuracy from multiple sequence alignment. *J. Mol. Biol*. 234:951–957.

Salamov, A. A., and V. V. Solovyev. 1995. Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments. *J. Mol. Biol*. 247:11–15.

Sander, C., and R. Schneider. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*. 9:56–68.

Schmieden, V., J. Kuhse, and H. Betz. 1992. Agonist pharmacology of neonatal adult glycine receptor $\alpha$ subunits: identification of amino acid residues involved in taurine activation. *EMBO J*. 11:2025–2032.

Sine, S. M., H.-J. Kreienkamp, N. Bren, R. Maeda, and P. Taylor. 1995. Molecular dissection of subunit interfaces in the acetylcholine receptor: identification of determinants of $\alpha$-conotoxin M1 selectivity. *Neuron*. 15:205–211.

Sixma, T. K., K. H. Kalk, B. A. van Zanten, Z. Dauter, J. Kingma, B. Witholt, and W. G. Hol. 1993. Refined structure of *Escherichia coli* heat-labile enterotoxin, a close relative of cholera toxin. *J. Mol. Biol*. 230:890–918.

Sternberg, M. J. 1996. Protein structure prediction: principles and approaches. *In* Protein Structure Prediction, Vol. 170. M. J. Sternberg, editor. IRL Press, Oxford. 1–30.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The Clustal-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 25:4876–4882.

Toyoshima, C., and U. Nigel. 1990. Three-dimensional structure of the acetylcholine receptor by cryoelectron microscopy and helical image reconstruction. *J. Cell. Biol*. 111:2623–2635.

Toyoshima, C., and N. Unwin. 1988. Ion channel of acetylcholine receptor reconstructed from images of postsynaptic membranes. *Nature*. 336: 247–250.

Tsigelny, I., N. Sugiyama, S. M. Sine, and P. Taylor. 1997. A model of the nicotinic receptor extracellular domain based on sequence identity and residue location. *Biophys. J*. 73:52–66.

Tzartos, S., H. V. Loutrari, F. Tang, A. Kokla, S. L. Walgrave, R. P. Milius, and B. M. Conti-Tronconi. 1990. Main immunogenic region of Torpedo electroplax and human muscle acetylcholine receptor: localization and microheterogeneity revealed by the use of synthetic peptides. *J. Neurochem*. 54:51–61.

Unwin, N. 1993a. Neurotransmitter action: opening of ligand-gated ion channels. *Cell*. 72(Suppl.):31–41.

Unwin, N. 1993b. Nicotinic acetylcholine receptor at 9-angstrom resolution. *J. Mol. Biol.* 229:1101–1124.

Unwin, N. 1996. Projection structure of the nicotinic acetylcholine receptor: distinct conformations of the $\alpha$ subunits. *J. Mol. Biol*. 257: 586–596.

Vandenberg, R. J., C. R. French, P. H. Barry, J. Shine, and P. R. Schofield. 1992. Antagonism of ligand-gated ion channel receptors: two domains of the glycine receptor $\alpha$ subunit form the strychnine-binding site. *Proc. Natl. Acad. Sci. USA*. 89:1765–1769.

von Heijne, G. 1992. Membrane protein structure prediction, hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol*. 225:487–494.

Watty, A., C. Weise, M. Dreger, P. Franke, and F. Hucho. 1998. The accessible surface of the nicotinic acetylcholine receptor. Identification by chemical modification and cross-linking with $^{14}$C-dimethyl suberimidate. *Eur. J. Biochem*. 252:222–228.

Wells, G. B., R. Anand, F. Wang, and J. Lindstrom. 1998. Water-soluble nicotinic acetylcholine receptor formed by $\alpha$7 subunit extracellular domains. *J. Biol. Chem*. 273:964–973.

West, A. P. J., P. J. Bjorkman, D. A. Dougherty, and H. A. Lester. 1997. Expression and circular dichroism studies of the extracellular domain of the $\alpha$ subunit of the nicotinic acetylcholine receptor. *J. Biol. Chem*. 272:25468–25473.

Wilson, G. G., and A. Karlin. 1998. The location of the gate in the acetylcholine receptor channel. *Neuron*. 20:1269–1281.

Yager, P., E. Chang, R. Williams, and A. Dalziel. 1984. The secondary structure of acetylcholine receptor reconstituted in a single lipid component as determined by Raman spectroscopy. *Biophys. J*. 45:26–28.

Yi, T.-M., and E. S. Lander. 1993. Protein secondary structure prediction using nearest-neighbor methods. *J. Mol. Biol*. 232:1117–1129.

Yu, X., and Z. Hall. 1994. A sequence in the main cytoplasmic loop of the subunit is required for assembly of mouse muscle nicotinic acetylcholine receptor. *Neuron*. 13:247–255.

Zhang, X., J. P. Mesirov, and D. L. Waltz. 1992. Hybrid system for protein secondary structure prediction. *J. Mol. Biol*. 225:1049–1063.