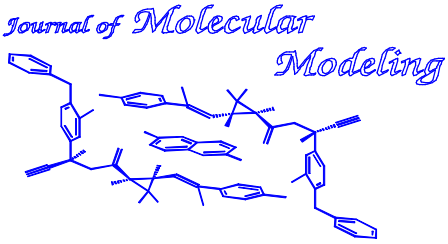
J. Mol. Model. **1996**, *2*, 46 - 50



Structural Evaluation of Distant Homology. A 3-D Model of the Ligand Binding Domain of the Nicotinic Acetylcholine Receptor Based on Acetylcholinesterase: Consistency with Experimental Data

Ajita A. Bhat and E. Will Taylor*

Computational Center for Molecular Structure and Design and Department of Medicinal Chemistry, The University of Georgia, Athens, GA 30602-2352, Tel: (706)-542-5391, Fax: (706)-542-5358 (WTAYLOR@RX.UGA.EDU)

Received:18 November 1995 / Accepted: 5 February 1996 / Published: 12 March 1996

Abstract

Acetylcholine is a ligand for both acetylcholinesterases and nicotinic acetylcholine receptors. Hence, at least some local sequence and structural similarities between the acetylcholinesterases and the receptors which recognize acetylcholine (ACh) might be expected. Peterson [2] produced an alignment of the ACh binding region between these two types of ACh-binding molecules, featuring a number of well conserved residues. The extent of this region of sequence similarity suggests the possible existence of a common ancenstral ACh binding module. To attempt to further validate Peterson's sequence alignment we have built a homology model of the ACh binding domain of the human neuromuscular nicotinic acetylcholine receptor based on the structure of acetylcholinesterase from *Torpedo californica*. Using this 3-D model we have examined the residues which were previously shown to interact with the endogenous ligand by various methods (mapping, site-directed mutagenesis). The consistency of such data with the model provides further support for a structural similarity and possibly a divergent evolutionary relationship between the ACh-binding domains of these two classes of proteins. Results suggest that this model may be able to contribute to an understanding of the structure and function of the ACh receptor. Using this case as an example, we propose that 3-dimensional computer modeling can be used as a tool to evaluate distant homologies when adequate experimental data (e.g., site-directed mutagenesis) is available.

Keywords: Acetylcholine Receptor, Acetylcholinesterase, Modeling, Receptor, Homology, 3-D Structure
Short Title: Structural Evaluation of Distant Homology

Introduction

Along with exon shuffling, gene duplication and subsequent divergence is a fundamental mechanism of protein evolution. Proteins or protein modules derived from a common ancestor can accumulate numerous point mutations leading to homologous proteins with varying degrees of sequence similarity, in some cases so low as to be insignificant despite conserved structural features. A number of techniques have been developed in order to evaluate such potential distant homologies, including the comparison of patterns of conserved sequence motifs in multiple sequence alignments, and the method of "inverse folding", which can be applied when at least one of the sequences has a known crystal structure [1]. When there is sufficient sequence similarity to at least permit a motif-based alignment of the protein sequences, there is another way to evaluate such distant relationships that has not been widely used. One can evaluate the potential homology by examining the consistency between the 3-D pro-

* *To whom correspondence should be addressed*

tein model based on the proposed alignment and the existing biochemical data, if that data is sufficiently specific in relation to the position of amino acid side chains and ligand or substrate binding sites. In this paper, we will use the method which we call structural evaluation of distant homology (SEDH), to assess a proposed relationship between the nicotinic acetylcholine receptor (nAChR) and acetylcholinesterase (AChase). We will demonstrate the approach by mutating a part of the substrate binding domain of acetylcholinesterase (a known 3-D structure) into the corresponding region of the nicotinic acetylcholine receptor ligand binding domain, based on an alignment previously proposed by Peterson [2]. We will show that the consistency of the model with known ligand binding data supports the proposed structural relationship.

Acetylcholine (ACh) is a peripheral and central nervous system neurotransmitter that acts at several types of ACh receptors, viz. muscarinic and nicotinic. After its action acetylcholine is acted upon by the extracellular enzyme acetylcholinesterase (AChase) whose biological role is termination of the impulse at the nerve endings by rapid hydrolysis of ACh. The nicotinic ACh receptor (nAChR) belongs to the ligand gated ion channel (LGIC) superfamily of proteins. nAChR is the most well studied receptor to date within this large family and is thus a prototype for understanding the binding of neurotransmitter to the receptor and the subsequent structural changes of the receptor. The musarinic ACh receptors, nAChRs and the AChases share the primary biological function of interacting with the acetylcholine molecule.[2] Therefore there is reason to expect that they may have some structural features in common, such as certain amino acid residues to which ACh binds.

The alignment presented by Peterson shows consensus residues among these two types of protein sequences.[2] A slightly truncated version of Peterson's alignment of the region of sequence similarity is shown in Figure 1. This region contains most of the residues involved in substrate binding in AChase and AChR. Based on this information we ask the question: can the known 3-D structure of the *Torpedo californica* AChase from the Brookhaven protein data bank (code 1ace.pdb) be mutated to give a resonable nAChR model that is consistent with the experimental data (i.e. amino acid residues involved in ACh binding)?

Structure of AChases

Within the acetylcholinesterase family, the most well studied members have been the proteins from *Torpedo californica* and *Torpedo marmorata*, where AChase exists in the form of a homodimer consisting of about 540 amino acids. The AChase structure has been resolved by X-ray crystallography at 2.8 Å.[3] AChase has a deep, narrow cavity 20 Å long in the center of the homodimer, also known as the "active site gorge".[4] An extensive lining made of aromatic residues (including Y70, W84, W114, Y121, Y130) forms the binding pocket for the quaternary amine which permits different possible orientations for ACh and various sites for agonist and antagonist interactions. The model proposed by J. Sussmann et al.[3] features an oxyanion hole formed by the main chain nitrogens of G118, G119 and A201 (not shown in the figure) which interact with the carbonyl oxygen of ACh. The conserved G117 and G118 probably also play a role in making the chain very flexible, facilitating the binding for its interaction.[3] (Figure 2).

```
Neuromuscular nAChrs:
   135       145       155       165       175       185       195       205       215      223
1  THFPFDQQNCIMKLGIWTYDGTKVSISPESDRPDLSTFMESGEWVMKDYRGWKHWVYYTCCPDTPYLDITYHFIMQRIPLYFVVNVIIPCL
2  THFPFDQQNCIMKLGIWTYDGTKVSISPESDRPDLSTFMESGEWVMKDYRGWKHWVYYTCCPDTPYLDITYHFIMQRIPLYFVVNVIIPCL
3  THFPFDEQNCSMKLGTWTYDGSVVAINPESDQPDLSNFMESGEWVIKEARGWKHWVFYSCCPTTPYLDITYHFVMQRLPLYFIVNVIIPCL
4  THFPFDEQNCSMKLGTWTYDGSVVAINPESDQPDLSNFMESGEWVIKESRGWKHSVTYSCCPDTPYLDITYHFVMQRLPLYFIVNVIIPCL

Acetycholinesterases:
   62        72        82        92       102       112       122       134      142      151
1  STYPNNCQQYVDEQFPGFSGSEMWNPNREMSEDCLYLNIWVPSPRPKSTTVMVWIYGGGFYSGSSTLDVYNGKYLAYTEEVVLVSLSYRVG
2  STYPNNCQQYVDEQFPGFPGSEMWNPNREMSEDCLYLNIWVPSPRPKSATVMVWIYGGGFYSGSSTLDVYNGKYLAYTEEVVLVSLSYRVG
```

Figure 1*: Comparison of AChase and nAChR sequences (after Peterson, 1989 [2]). Conserved residues are shown in red (in several cases they are offset by one position as indicated). In nAChR, residues involved in agonist binding are shown in green and those involved in antagonist binding are shown in blue. Amino acid Y190, involved in both agonist and antagonist binding, is shown in cyan. Amino acids involved in ACh binding in AChase are colored in magenta. Amino acid Y130 in AChase is conserved and takes part in binding.* ***1****: Torpedo californica* ***2****: Torpedo marmorata* ***3****: Mouse* ***4****: Human*

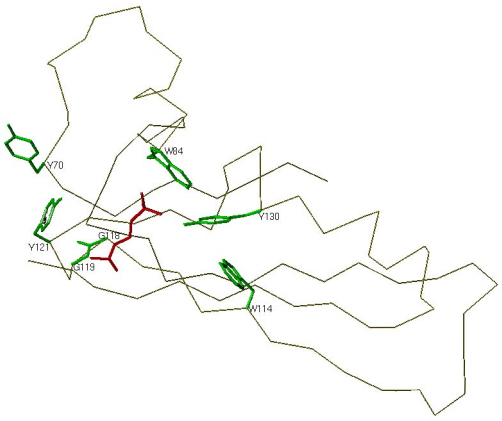


Figure 2. *Interaction between ACh and AChase in the experimental (X-ray) AChase crystal structure (1ace.pdb). The peptide backbone is shown as a Cα trace. ACh is shown in red without hydrogens. The side chains of some important residues that influence binding or interact withACh are shown (1-7). 1:Y70 2: W84 3: W114 4: G118 5: G119 6: Y121 7: Y130*

Structure of nAChR

The nAChR is a transmembrane protein consisting of four homologous subunits, $\alpha_2\beta\gamma\delta$. These subunits are arranged in the shape of a torus with a centrally located ion channel. The subunits are structurally related, implying that they may have evolved from a common ancestor.[5] Ligands of various structural classes are proposed to bind to the large region on the extracellular receptor surface towards the N-terminal end, i.e. the α subunit.

Residues Involved In Ligand Binding to nAChR: Experimental Data

Among these nAChR subunits mainly the α subunits have cholinergic binding sites. Some of these binding sites are different for agonists and competitive antagonists, as shown by site-directed mutagenesis and mapping.[6,8] The two main ligand binding segments are 134-153 and 181-200.[9]

Site-directed mutagenesis and photo-affinity labelling studies show that the antagonist binding site (e.g. for α-Bungarotoxin) is not a single residue or a continuous segment but rather discontinuous short segments (C142, H186, V188, Y189, Y190, P194, D195, Y198),[6-8] probably folded together within the AChR structure.[9] Residues involved in agonist binding are Y190, C192, C193, Y198.[6,8]

Thus residues involved in binding lie primarily within the α subunit including and flanking the region of the two cysteines 192 and 193 which form a disulfide bond, creating an unusual cis-peptide linkage.[10, 11] These two cysteines are conserved in all α subunits of the nicotinic acetylcholine receptors.[12] In the past, theoretical calculations have been done which have predicted that the L-Cys-L-Cys peptide bond can occur only in a cis-peptide form, [13, 14] which is supported by X-ray data on model peptides.[10] It has been suggested this Cys-Cys bond may act as a molecular switch to control receptor activation. Furthermore since proline residues have been observed to have a higher probability of having the cis isomer of the preceeding peptide bond compared to other residues,[15] the peptide bond between Cys-193 and Pro-194 has been suggested to be in the cis form.[8] Thus there is the potential for at least one or possibly two consecutive cis-peptide bonds in this region of AChR. These potential cis peptide bonds must be given due consideration when modeling this region of the AChR.

Materials and Methods

Sybyl 6.04 (Tripos Associates, St. Louis, MI 63144-2913) was used to model the receptor protein. The AChase structure from *Torpedo californica* containing the ACh molecule was obtained from the Brookhaven Protein data bank (PDB file 1ace), figure 2. The MUTATE command was used to mutate the residues between amino acids 61 and 151 according to the human neuromuscular nAChR. The rest of the molecule was truncated using the SPLIT command. Residues were renumbered from 133 to 223, to correspond with the nAChR sequence. Possibly due to an unusual bend in the protein backbone around position 198 involving proline 197 (not present in AChase), the phenyl ring of tyrosine 198 was not oriented to fit in the protein globular structure and faced the side opposite to that of ACh. The protein backbone structure between the residues Pro 197 and Asp 200 was thus remodeled using the BIO LOOP feature. From the protein model obtained, two versions were made, one containing a cis peptide bond between Cys 192 and Cys 193 (mono-cis) and the other containing two cis peptide bonds, between Cys192-Cys193 and between Cys 193-Pro 194 (di-cis). Both the protein strucutres were separately minimized using the following procedure: the amino acids Cys 192 to Pro 194 were defined as the 'hot region' using the function ANNEAL in Sybyl. A 'hot radius' of 4.5 Å and an 'interesting region' of 10 Å was defined. Minimization was carried out using steepest descent using KOLLMAN_ALL atom force field. The ACh molecule and the backbone was then defined as AGGREGATE and only the side chains were minimized using the steepest descent method with the KOLLMAN_ALL atom force field. Slowly the constraints were relieved and the entire molecule was minimized using the KOLLMAN_ALL atom force field, using the conjugate-gradient method with KOLLMAN charges, a non-bonded cut-off of 8 Å, 1,4-scaling equal to 0.5 and ΔE of 0.1kcal/mol. The residues which assist in agonist binding and those in-

volved in the interaction with the competitive antagonists are shown in figure 3.

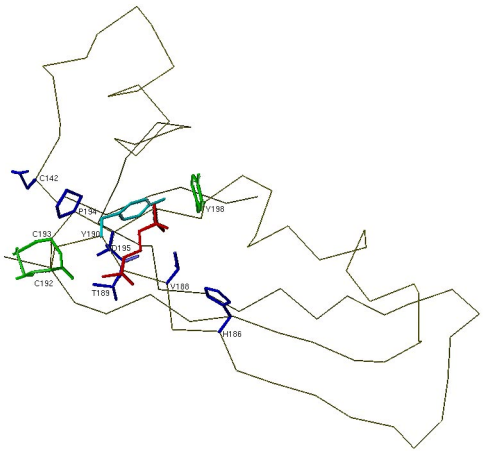


Figure 3. *Locations of known ACh-interacting residues in the nAChR model. The peptide backbone is shown as a Cα trace. ACh is shown in red without hydrogens. Residues which have been shown to interact with agonist ligands are numbered below as 5-7, 10 whereas residues numbered below as 1-5, 8, 9 have been shown to interact with antagonist ligands. The color scheme for the ligands is the same as that used in figure 1. In this model, all these residues clearly cluster around the ACh molecule, consistent with experimental results (see text). Residues C192, C193 and P194 have cis peptide bonds between them (see figure 4). 1: C142 2: H186 3: V188 4: T189 5: Y190 6: C192 7: C193 8: P194 9: D195 10:Y198*

Results and Discussion

We confirmed the alignment with the BESTFIT program [16] (gap weight 1.5/2.5, gap length 0.3/.25) using the sequences of Torpedo californica AChase versus those of Torpedo and human AChR. This gave pairwise alignments that were very close to Peterson's alignment in the region of interest (between amino acids T133 and L223). After minimization of the mono-cis and di-cis protein fragments we found that the protein with two cis peptide bonds (di-cis) was more stable by >5 kcal/mol, than the protein with only one cis peptide bond (mono-cis) between the two cysteines forming a disulfide bond. We also made another model (not shown) di-trans with trans peptide bonds between Cys192-Cys193-Pro194 which was also >5 kcal/mol higher in energy relative to the di-cis model. Considering the energy differences we concluded that the di-cis model is a more valid representation of the active domain, a result consistent with earlier observations.[9] A backbone fit of the AChR model to the AChase enzyme gave a weighted root mean squared difference of 1.6945 Å. Root mean square difference of the backbone of the two versions of the AChRs was 0.4188 Å which differed only at the region between Cys 192 and Pro 194 (Figure 4). Among all the interactive residues within the AChR, only one ligand-binding residue has been mutated from the *Torpedo californica* to the human spieces i.e., Y189T. However from the model even the side chain of threonine 189 is seen in close contact with the ACh molecule.

The idea was then to test this model by comparing with the experimental data available in the literature, such as site-directed mutagenesis. Many experiments have been performed in the past [7-13] where ligand binding sites of the AChR were identified. An attempt was made to rationalize the location and function of the binding sites to those given by the model, in order to validate any possible distant homology. It is immediately obvious from the model that, as shown in figure 3, all the residues previously shown to interact with cholinergic ligands are clustered in the immediate vicinity of the predicted ACh binding site. This suggests that the model is at least qualitatively correct.

Furthermore, the model can account for specific changes in activity and building associated with specific amino acid mutations. By mutagenesis Chaturvedi et. al.[8] and O'Leary et. al.[17] have indicated that the aromaticity of the Tyr 190 and Tyr 198 of nAChR were important since they form an electronegative subsite through the formation of a tyrosinate anion which could attract the positive nitrogen of acetylcholine. This is clearly seen in the model where the aromatic rings and their hydroxyl groups face the nitrogen and interact with it. In the AChase structure (1ace.pdb) the glycines 117, 118 and 119 are close to the ACh molecule, which correspond to T189 and Y190 of AChR in the alignment. The model also correctly predicts that Cys 192, Cys 193 and Pro 194 in AChR form a turn at the tip of the beta sheet as pointed out earlier by Chaturvedi et. al.[8] The rigid conformation induced by the three amino acids comprising the disulfide bridge and the cis form of proline (as shown in Figure 4) may uniquely contribute to the ability of this region of the nAChR to bind ligands and contribute in the process of channel gating.[17]

The sidechains of His 186, Val 188 and Asp 195 in nAChR point toward the ligand in the binding domain in this model. Since they are quite a distance (>6 Å) from ACh, they are not involved in agonist binding but could be readily involved in binding larger molecules like α-Bungarotoxin, which again, is precisely consistent with the experimental data [7, 9]. Mutations of H186A, V188T probably decrease the sidechain length and branching required for binding of the antagonist α-Bungarotoxin to its binding site.

The consistency of the model with the experimental results suggests that these two classes of ACh-binding proteins do have some common structural features, supporting the possibility that their ligand binding domains may have evolved from a common ancestor. The results suggest that the Torpedo AChase protein structure in the Brookhaven protein data bank (figure 2) is a reasonable starting point for

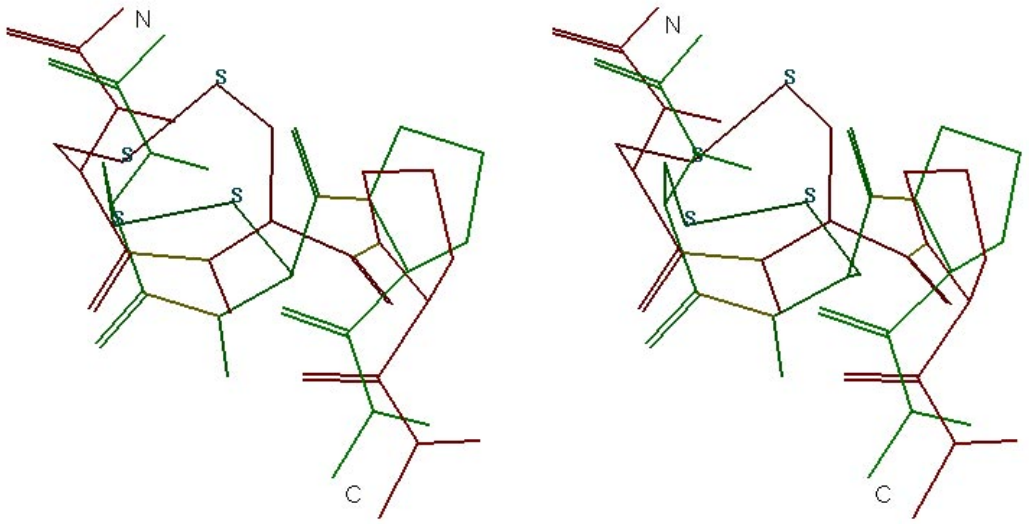


Figure 4. *Stereoview comparing the two versions of the protein model involving the Cys-Cys-Pro sequence, and the potential cis-peptide bonds (highlighted in yellow). The mono-cis model is shown in red and the di-cis model is shown in green. The amino and the carboxy ends of the fragments are denoted with the letters N and C respectively. The models differ only in the peptide bond between Cys 193 and Pro 194. Both the mono-cis model and the di-trans model (not shown) are >5kcal/mol higher in energy relative to the di-cis model.*

homology modeling of the ligand binding domain of the human ACh receptor (figure 3).

With this example, we have illustrated the approach of this method, which we have tentatively called structural evaluation of distant homology (SEDH). When adequate biochemical data (mutagenesis etc.) are available, this method may complement existing approaches like inverse folding, as a tool for validating speculative homology models.

References

1. Godzik, A.; Kolinski, A.; Skolnick, J. *J. Comput. Aided. Mol. Des.* **1993**, *7*, 397.
2. Peterson, G.L. *J. Neurochem. Res.* 1989, 22, 488.
3. Sussman, J.L.; Harel, M.; Silman, I. Chem. Biol. Inter. 1993, 87, 187.
4. Sussman, J.L.; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman, I. Science, 1991, 25, 872.
5. Raftery, M.A.; Hunkapiller, M.W.; Strader, C.D.; Hood, L.E. Science, 1980, 208, 1454.
6. Sine, S.M.; Quiram, P.; Papanikolaou, F.; Kreienkamp, H.; Taylor, P. J. Biol. Chem., 1994, 269, 8808.
7. Conti-Tronconi, B.M.; Diethelm, B.M.; Wu, X.; Tang, F.; Bertazzon, T. Biochem, 1991, 30, 2575.
8. Chaturvedi, V.; Donnelly-Roberts, D.L.; Lentz, T.L. Biochem, 1993, 32, 9570.
9. Conti-Tronconi, B.M.; Tang, F.; Diethelm, B.M.; Spencer, S.R.; Reinhardt-Maelicke, S.; Maelicke, A. Biochem, 1990, 29, 6221.
10. Capasso, S.; Mattia, C.; Mazzarella, L. Acta Cryst B, 1977, 33, 2080.
11. Kao, P.N.; Karlin, A. *J. Biol. Chem.*, **1986**, *261*, 8085.
12. Blount, P.; Merlie, J.P. *J. Cell. Biol.* **1990**, *111*, 2613.
13. Ramachandran, G.N.; Sasisekharan, V. *Adv. Protein. Chem.*, **1968**, *23*, 284.
14. Chandrasekaran, R.; Balasuramanian, R. *Biochim. Biophys. Acta*, **1969**, *188*, 1.
15. McArthur, M.W.; Thornton, J.M. *J. Mol. Biol.*, **1991**, *218*, 397.
16. Program Manual for the Wisconsin Package, Ver. 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, WI 53711.
17. O'Leary, M.E.; White, M.M. *J. Biol. Chem.*, **1992**, *267*, 8360.