Introduction: Chapter 1

Chemical Scale Neuroscience with Unnatural Amino Acids and Computational Modeling

Ligand Gated Ion Channels

The human brain is perhaps Nature's most complex and powerful creation. It is made up of roughly 10^{12} nerve cells, or neurons.[1] (Fig. 1, Top Left) These neurons can be thought of as biological wires, which function by transmitting electrical signals, just like wires in man-made devices. The junctions between neurons are known as synapses, and are generally chemical in nature. (Fig. 1, Bottom Left) When the electrical signal in one neuron reaches a synapse, it causes the release of small molecule neurotransmitters such as acetylcholine (ACh) from vesicles at the terminus of this presynaptic neuron. (Fig. 1, Top Right) These small molecules diffuse across the synaptic cleft and bind to proteins called ligand gated ion channels (LGICs) on the postsynaptic side. These proteins are large, integral membrane proteins that respond to neurotransmitter (ligand) binding by undergoing a conformational change that opens a channel in their center that allows ions to pass through. (Fig. 1, Bottom Right) Thus, they convert a chemical signal, ligand release, into an electrical one, the flow of ions across an otherwise ion-impermeable membrane. Amazingly, this complex process takes place on the millisecond timescale.[2]

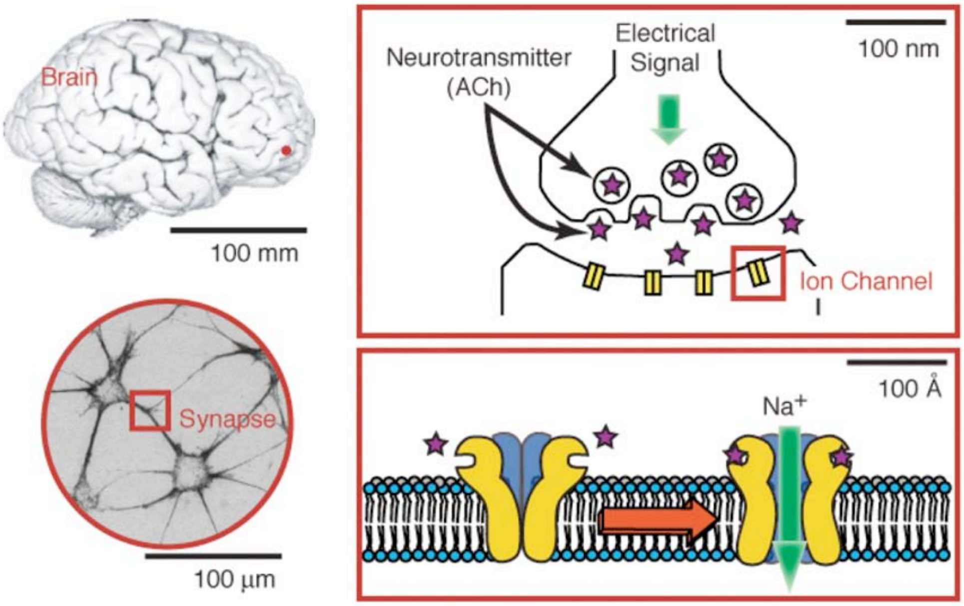


Figure 1. Synaptic transmission. Top left: A human brain. Bottom left: A synaptic connection between two mammalian neuronal cells in culture. Top right: Schematic representation of a single synapse. Bottom right: Ion channel gating.

Ion flux through the LGIC causes a change in the voltage across the membrane which opens other, voltage gated ion channels (VGICs). More VGICs further down the neuron respond to the opening of these channels, initiating a cascade of VGIC opening that establishes an electrical signal in the postsynaptic neuron. This, in broad strokes, is the mechanism by which neurons function and communicate with each other.[2] LGICs lie at the center of this process. We wish to understand the structure and function of these complex molecules. The complexity of LGICs make their study an intriguing problem for the chemist; their essential role in thought and learning make their study a problem of tremendous import to the life sciences, from basic biology to applied medicine.

Early Ion Channel Characterization

Although the "ionic" nature of nerve transmission had been known since Helmholtz's time,[3] the study of the ion channel as a molecular entity began in the 1950s with Hodgkin and Huxley's groundbreaking studies of the squid giant axon.[4, 5] Some of the techniques that they employed were essentially the same as the two electrode voltage clamp (TEVC, described below) electrophysiology methods that will be employed throughout the work in this dissertation. Hodgkin and Huxley were able to establish the fact that separate mechanisms were responsible for sodium and potassium transport by separately blocking the currents from each type of ion.[4] Tetrodotoxin, isolated from the Fugu fish, was used to selectively block Na^+ currents and tetraethylammonium ions were used to block K^+ currents. These studies gave rise to the idea of molecular channels that were selective for different stimuli and for specific ions: ion channels.

The first structural characterization of ion channels came from 1971 work by Hille.[6] Electrical measurements made on frog nerve cells showed Hille that Li^+, Na^+, NH_3OH^+, and $NH_3NH_2^+$ cations could pass through the sodium channel, but K^+ could not. This lead Hille to develop a model of the pore that gave it a 5Å diameter with a hydrogen bond acceptor that must be satisfied. Li^+ and Na^+ were small enough to pass through the channel with a hydrating water, but K^+ was not (although K^+ itself is smaller than 5Å).

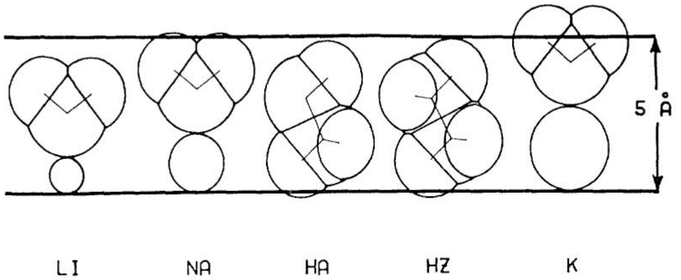


Figure 2. Hille's 1971 Model of the Sodium Channel Pore. Hydrated Li^+ (LI) and Na^+ (NA) ions can pass through the proposed 5Å channel, but hydrated K^+ (K) cannot. Hydroxylamine (HA) and hydrazine can also pass because they can fulfill hydrogen bond donor requirement. Reproduced from Hille, 1971.[6]

The organic cations were roughly the same size as a Li^+ or Na^+ ion with one bound water molecule and were capable of displaying a H-bond donor. This is a landmark study in the history of ion channel study, but it also illustrates a historical difficulty in the study of ion channel structure and function. Hille was able to precisely manipulate the chemical nature of the ion, but he had to use this to make assumptions about the sodium channel's chemical nature, which he could not alter.

Contemporary Ion Channel Studies

Since Hille's time, the essential ion channels have been cloned, and manipulation of their DNA through molecular biology permits us to alter the identity of one or several amino acids. Today, the most common method for studying ion channels is the structure/function study, which involves mutagenesis of the LGIC DNA to change the channel's structure, followed by electrophysiological characterization of the mutant's function. In TEVC electrophysiology, one electrode is used to measure the voltage difference across a cell membrane (relative to an extracellular ground), and the other injects sufficient current into the cell to "clamp" the voltage at a certain value *via* a feedback circuit. When ion channels in the cell open, the ion flux across the membrane demands that the current electrode must inject more current to maintain the voltage clamp. Thus, the record of flux through the current electrode is a direct measure of the ion channel gating.[7]

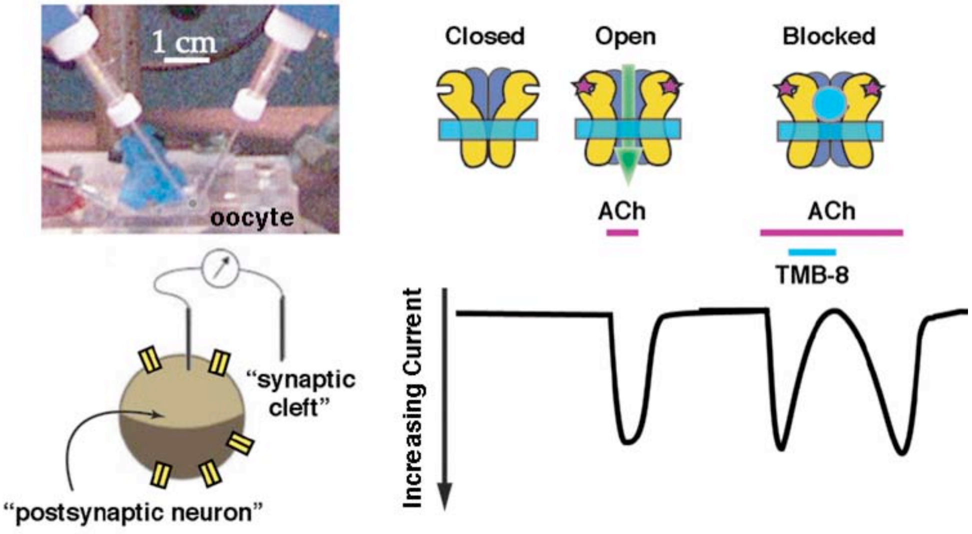


Figure 3. Oocyte electrophysiology. Top left: A *Xenopus* oocyte set up for TEVC electrophysiology. Bottom left: Schematic representation of TEVC set up. Left: Sample electrophysiological trace with images of the receptor in the states corresponding to the given electrophysiological data.

TEVC electrophysiology is often used in conjunction with the heterologous expression system of the *Xenopus laevis* oocyte, in which mammalian ion channels can be expressed (Fig. 3).[8] The oocyte is a particularly large single cell (1 mM in diameter) that can easily be injected with DNA or mRNA encoding an ion channel. The membrane-bound channel is expressed on the surface of the oocyte after an appropriate incubation period. The oocyte is then perfused with media that mirrors the fluid of the synaptic cleft in composition. Thus, the oocyte serves as a mimic of the postsynaptic neuron (Fig. 3). The binding of ligands like ACh to the LGICs expressed on the oocyte's surface causes channel opening and ion flux, which can be measured with the electrodes. Electrophysiology can be used to characterize the responses of agonists (ligands that open channels, such as ACh, shown above), antagonists (ligands that prevent channel opening, not shown), and blockers (molecules that prevent ion flow through open channels, such as TMB-8 shown above).

The drawings in Figures 1 and 3 represent ion channels as blobs. Of course they are not blobs, they are organic macromolecules; and their precise atomic structure is crucial to their proper function. As chemists, we are interested in atomic scale details, and we hope to understand LGICs with this level of precision. In recent years, ion channel structural information has begun to emerge in the form of the structures of several potassium channels,[9] two mechanosensitive channels,[10] and analogs of LGIC ligand binding domains; but to date, there is no high resolution structure of a LGIC.[11] However, even the high resolution images that are available do not make for a complete understanding of channel function, because LGICs are very large, inherently dynamic molecules. For example, the nicotinic acetylcholine receptor (nAChR) binds the 146 Da. ACh ligand, initiating the gating process that causes structural changes 50 Å away in the 50, 000 Da. protein.[12] Structure/function studies using mutations can give us hints as to which parts of the protein are important, but the limited scope of change available from the 20 natural amino acids cannot give us an understanding of LGICs at the chemical level, the level of the functional group or hydrogen bond. The recently developed technique of unnatural amino acid mutagenesis can provide us with this high precision insight.

Protein Studies with Unnatural Amino Acid Mutagenesis

For over a decade, it has been possible to site-specifically incorporate unnatural amino acids into a protein through nonsense suppression.[13] To a chemical biologist this represents a tool of limitless potential, promising systematic structure/function studies with a precision far beyond that which is possible with conventional site-directed mutagenesis. For example, in order to explore the determinants of ligand binding to the nAChR (Fig. 4 A - D), we will perform the following alterations: 1) To investigate the existence of a cation-π interaction between ACh and a Trp in the binding site (Fig. 4E), we can incorporate a fluorinated Trp analog that is a weaker cation binder. (Fig. 4F) 2) To probe the local pK_a of the binding site, we can incorporate an unnatural amino acid with a tethered amine. Since this should only activate the channel when it is protonated to form a cation, its protonation in the receptor binding site can be measured by channel activation. (Fig. 4G) 3) To probe the role of a carbonyl in binding the ligand, we can convert the backbone amide to an ester, which has a smaller carbonyl dipole. (Fig. 4H) Within the constraints of conventional mutagenesis, we would be left with the following options: 1) To mutate Trp to Phe or Tyr to attempt to verify the existence of the cation-π interaction. 2) To mutate binding site residues to Lys to try to produce the same effect as our designed tethered agonist. 3) No conventional mutant can probe backbone carbonyl interactions. Clearly, the unnatural amino acids listed above can provide insights that these mutants cannot.

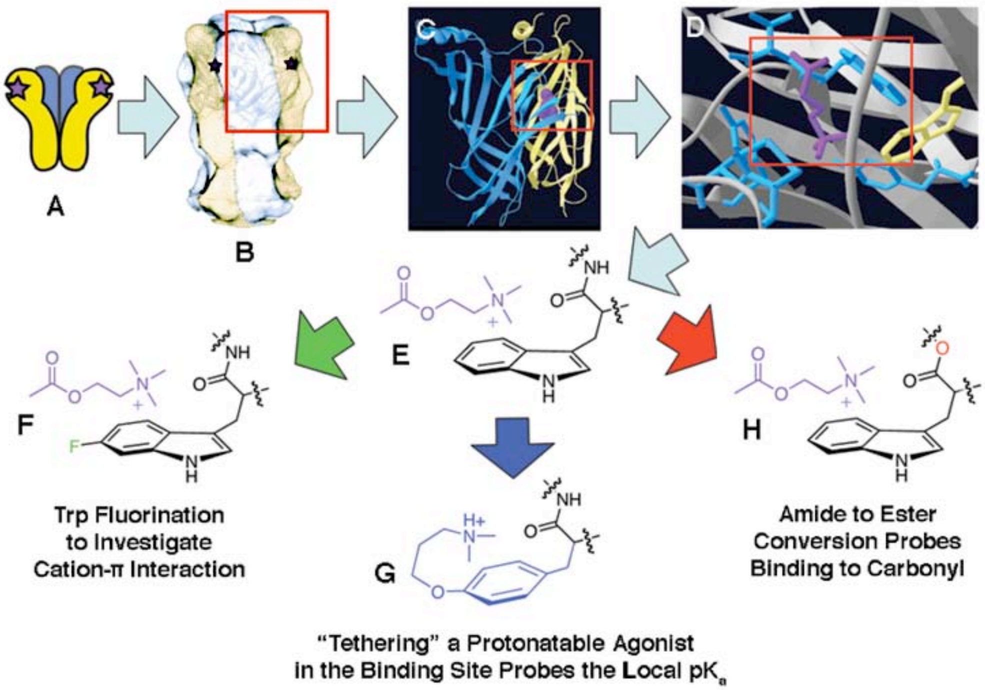


Figure 4. Unnatural Amino Acid Mutagenesis. A. An astructural model of a receptor. B-D. Increasingly high resolution structural images of the nAChR and its binding domain homolog AChBP. E. Schematic representation of ACh binding to Trp 149. F-H. Examples of unnatural mutations to probe the ACh binding environment.

The fundamental approach to unnatural amino acid incorporation through nonsense suppression is outlined in Figure 5.[14, 15] Briefly, a UAG stop codon is incorporated at the mRNA position coding for the residue of interest through mutagenesis. Separately, a $\mathrm{tRNA_{CUA}}$ that recognises the UAG codon – a suppressor tRNA – is prepared and chemically acylated with the desired unnatural amino acid. These two species are then added to an appropriate translation system. Without the $\mathrm{tRNA_{CUA}}$, translation would halt at the stop codon and the “nonsensical” protein would be released from the ribosome. In our system, the $\mathrm{tRNA_{CUA}}$ suppresses the nonsense. The unnatural amino acid is incorporated into the protein just like any other amino acid, and translation continues to give a full-length protein with an unnatural amino acid at a single, specific site.

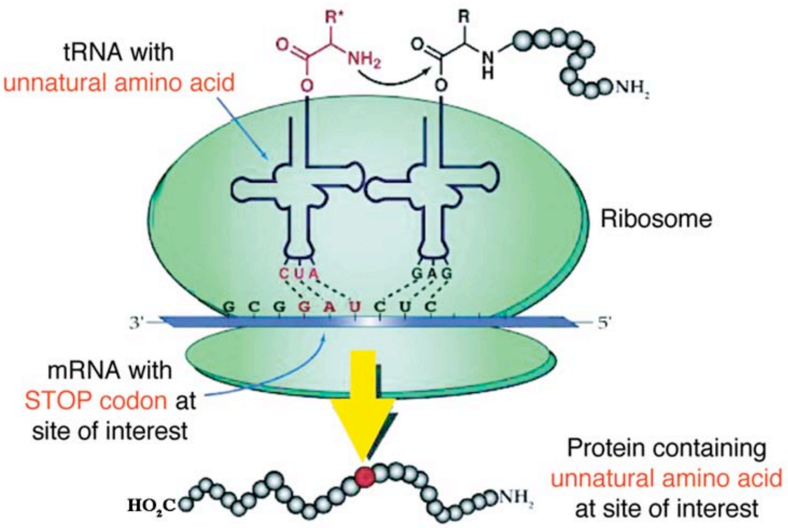


Figure 5. Unnatural Amino Acid Mutagenesis by Nonsense Suppression. The unnatural amino acid is incorporated at the site of interest by the ribosome, using the suppressor tRNA.

Naturally occurring nonsense suppression had been known for some time.[16] Hecht and coworkers developed key methodology for chemically synthesizing misacylated tRNAs in 1978.[17] Schultz put these two concepts together in 1989, reporting the first successful experiments in Noren *et al.*[18] Both Schultz[14, 18-20] and Hecht[21] expanded on this work with a large number of publications in the 1990s, which delineated the scope of amino acids that could be incorporated (See below) and showed the generality of the method through incorporation in many different proteins. Since then, the Sisido[22] and Yokoyama[23] groups have also made significant contributions. Sisido, in particular, has made important contributions to mapping the steric limitations on the incorporation of amino acids.[24]

Unnatural Amino Acid Mutagenesis Methods

The difficulty in implementing the method lies in the preparation of the aminoacyl tRNA (aa-tRNA). The preparation of the mRNA simply involves conventional site-directed mutagenesis. The semisynthesis of the aa-tRNA begins with the chemical synthesis of the amino acid as a α-N-protected cyanomethyl ester. This is coupled to a dCA dinucleotide in DMF as shown in Figure 6. The amine must be protected during the coupling reaction, and it remains protected throughout the aa-tRNA semisynthesis, as free amines destabilize the bond to the terminal adenosine. The aminoacyl-dCA (aa-dCA) is then ligated to a truncated 74 base suppressor tRNA ($tRNA_{CUA}^{-CA}$) to give a full-length, 76 base aa-$tRNA_{CUA}$. This is performed by *in vitro* treatment with T4 RNA ligase. The aa-$tRNA_{CUA}$ can be stored like this for years at –80° C. The final step in the semisynthesis of the aa-$tRNA_{CUA}$ for nonsense suppression is the deprotection of the α-amine. We typically employ 6-nitroveratryloxycarbonyl (NVOC) or 4-pentenoyl (4-PO) groups, which can be removed with methods that are chemically orthogonal to the functional groups present on the tRNA.[15] The NVOC group can be cleaved by irradiation with 350 nm light, and the 4-PO group can be cleaved by iodolactonization. The 4-PO is commonly used to protect amino acids with photochemically active sidechains so that they are not damaged before they are incorporated into the protein.

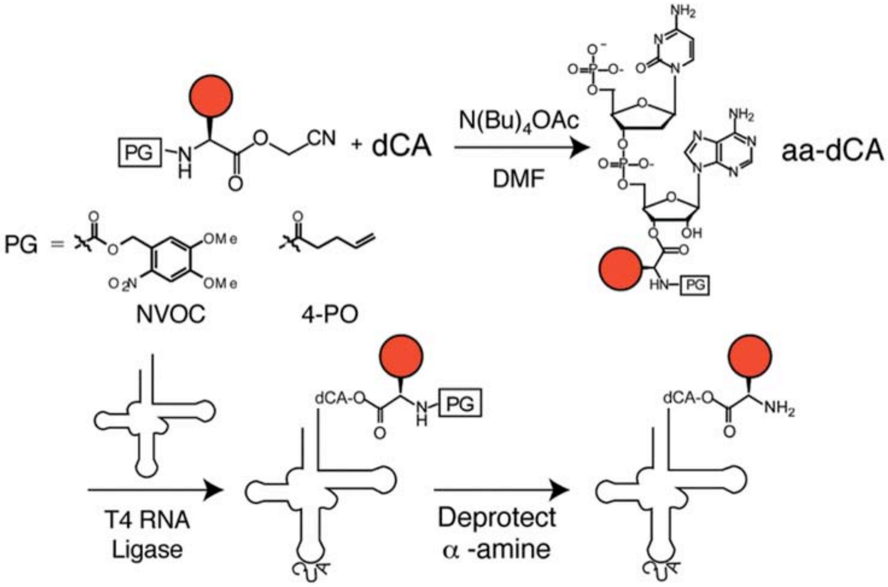


Figure 6. Semisynthesis of the aminoacylated suppressor tRNA. N-protected (with the NVOC or 4-PO group) cyanomethyl esters are chemically coupled to dCA, and this product is enzymatically ligated to a truncated tRNA with T4 RNA ligase. The full-length aminoacyl tRNA is deprotected prior to use.

To make use of this method in studying ion channels, one must simply inject the deprotected aa-tRNA and the mRNA bearing a UAG codon at the site of interest into a *Xenopus* oocyte.[25] As in conventional electrophysiology studies, we incubate these oocytes to allow expression of the channels on the cell surface. These are then assayed with TEVC techniques to characterize the results of the structural perturbation of the unnatural amino acid on the function of the ion channel.

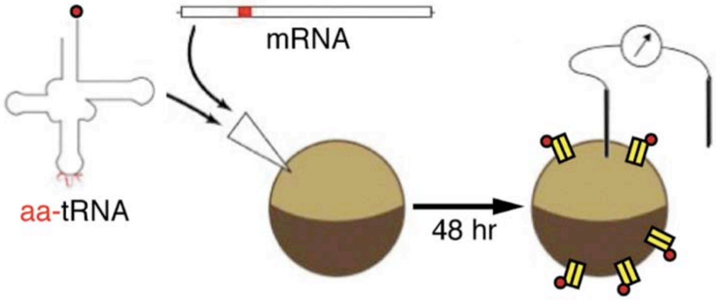


Figure 7. *in vivo* Nonsense Suppression in oocytes. mRNA with the stop codon at the site of interest and aminoacyl suppressor tRNA (aa-tRNA) are injected into *Xenopus* oocytes. After incubation, expressed receptors are assayed electrophysiologically.

This method has proven to be extraordinarily general. Practitioners of suppression methods have incorporated over 100 different amino acids into proteins in a variety of translation systems, both *in vitro* and *in vivo*.[13] In fact, nonsense suppression is not limited to incorporating amino acids: hydroxy acids, *N*-hydroxylamines, and hydrazides are also viable.[19, 26] D-amino acids and β-amino acids do not seem to be viable in translation.[27] In our own labs, we have incorporated a great number of amino acids into proteins, Figure 8 shows a nearly comprehensive list. Among these are amino acids that subtly differ from natural amino acids; like *allo*-isoleucine, which differs from isoleucine only in its stereochemistry; or like *O*-methylthreonine, which is isosteric with isoleucine, but differs in its polarity. We can completely remove specific non-covalent interactions with amino acids that delete a hydrogen bond donor such as *O*-methyltyrosine. We can also rationally modulate the non-covalent interaction as we do when we incorporate fluorinated Trp derivatives to probe cation-π interactions. We can turn on these interactions in the course of an experiment with caged amino acids whose photolyzable protecting group can be removed to reveal a wild type amino acid. We can use amino acids bearing the post-translational modifications that are often made enzymatically, like phosphorylation. We can also incorporate amino acids that are wildly different than the natural amino acids; a benzophenone-based amino acid can be used to photo-crosslink the protein in which it lies to other regions of the same protein or other proteins with which it interacts. As one can see, there are few limitations to the amino acids that can be incorporated by nonsense suppression.

Most methodological concerns involve issues with the tRNA_{CUA}. The first issue is “orthogonality.” The suppressor tRNA must be orthogonal to (not be recognized by) the translation system’s aminoacyl tRNA synthetases, the enzymes that charge tRNAs with their cognate amino acids. If this is not achieved, once tRNA_{CUA} has lost its unnatural amino acid, it will be charged with a natural amino acid and return to the protein synthesis cycle. This will lead to a mixture of proteins that would be of little value in biophysical studies. To avoid this, Schultz originally employed a yeast phenylalanine tRNA with a modified anticodon ($\text{tRNA}_{\text{CUA}}{}^{\text{Phe}}$) for use with an *Escherichia coli in vitro* translation system.[18] This solved the orthogonality problem because it was known that yeast tRNAs were not recognized by *E. coli* synthetases. For the *in vivo* nonsense suppression methodology, a variant of a *Tetrahymena thermophila* tRNA is used because $\text{tRNA}_{\text{CUA}}{}^{\text{Phe}}$ is not sufficiently orthogonal to the *Xenopus* synthetases.[28] This tRNA is also viable in the *E. coli in vitro* system. The *T. thermophila* tRNA is often called THG73, but will simply be referred to as tRNA_{CUA} as it is the only tRNA employed in the work described here.

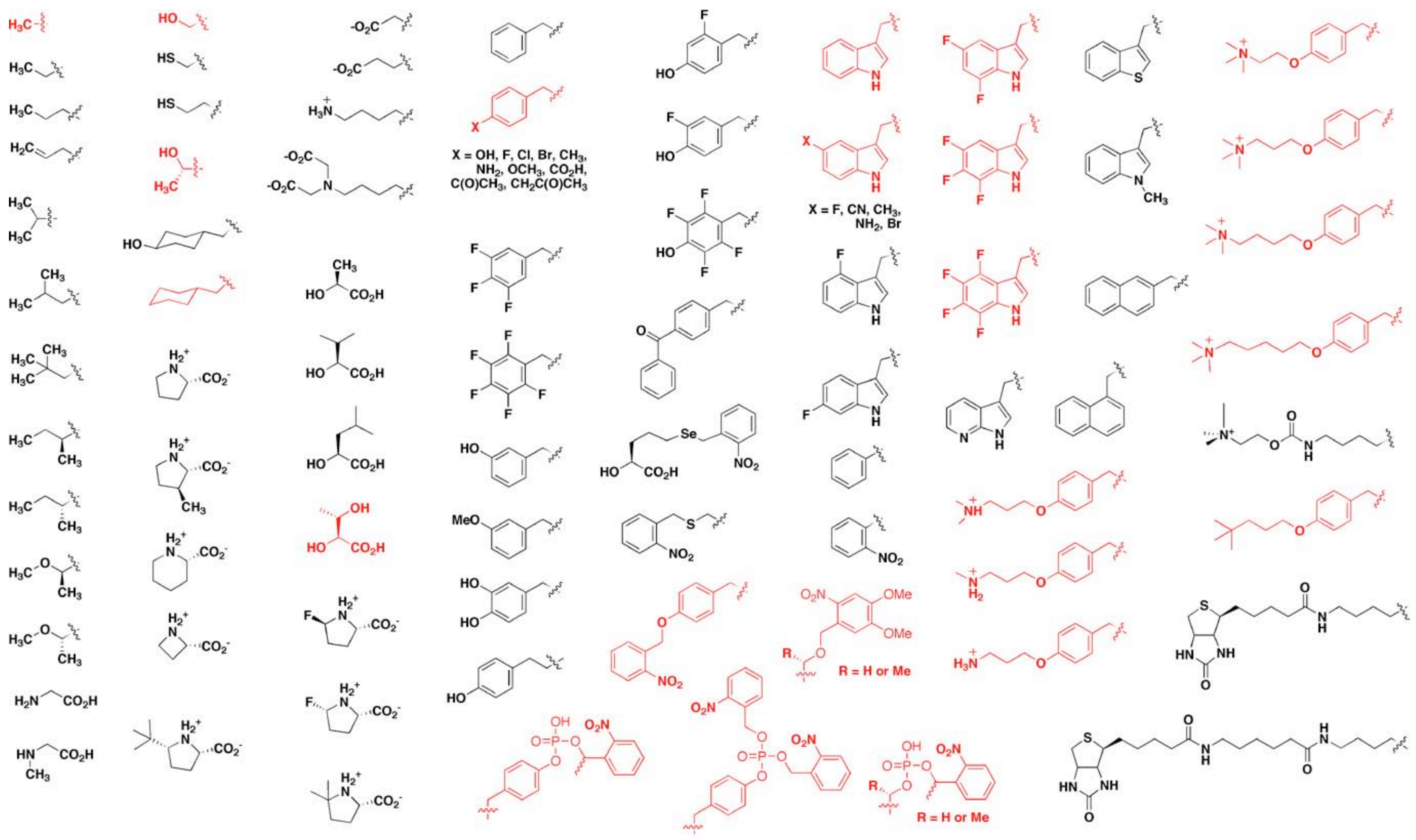


Figure 8. Amino and hydroxy acids incorporated through nonsense suppression by Dougherty and coworkers. Those amino acids that are used in the work described here are shown in red.

Another issue concerning the tRNA, and indeed the most severe limitation of the nonsense suppression methodology, is the fact that the aa-tRNA$_{CUA}$ is consumed as a "stoichiometric reagent." That is, because the aa-tRNA$_{CUA}$ is not reacylated with the amino acid inside the cell, one can obtain at most one protein per aa-tRNA$_{CUA}$ introduced. This represents a maximum; in practice, the yield is always lower than this. tRNA can lose its amino acid to both enzymatic and non-enzymatic hydrolysis before that amino acid is used. Thus, although the synthesis of this aa-tRNA reagent is well worked out, it represents a sufficiently large effort to generate proteins with unnatural amino acids that very sensitive assays must be sought. Fortunately, using TEVC electrophysiology as an assay, as little as 10 attomol of protein can be detected.[13]

The Role of Computational Modeling in Ion Channel Studies

As one can see, since Hille's 1971 experiments, we have gained the ability to manipulate the protein with the same precision that he manipulated the ionic charge carriers. Unnatural amino acids enable us to perform extremely precise structure/function studies, but the data generated in these studies must be interpreted in terms of structural models. Our ability to model chemical structures has also developed since Hille's study, and this constitutes an important part of our research. Several types of modeling will be discussed, varying in accuracy, and chosen to suit the scale of the problem at hand. The ones used most commonly are *ab initio* quantum mechanical (QM) calculations, molecular dynamics (MD) simulations using molecular mechanics (MM) force fields, and mixed method QMMM calculations. We employ these calculations in a relatively standard fashion without modifications to the code (with the exception of the QMMM study, which is described in detail in Chapter 5), so we will only describe the manner in which we use them.

QM calculations allow us to optimize the geometry of complexes of molecules and calculate interaction energies with very high precision.[29] By using numerical solutions to solve approximations of the full Schrödinger equation for each system, QM calculations take into account the polarization of electrons[30] that are often crucial to accurately representing the non-covalent interactions that control ligand binding and gating in LGICs. Unfortunately, the computational time required to perform these calculations scales with the size of the molecule, so QM calculations are limited to relatively small molecules. They cannot be applied to full ion channels. We use them to calculate the interaction energies of model systems that mimic certain key interactions of ligand or ion and protein. Because we use only model systems, the absolute energies calculated have little meaning, but energy differences and trends can be used to compare with data from our experiments. Unfortunately, these measurements do not take into account explicit solvent interactions or that binding to the protein can result from a dynamic ensemble of interaction geometries.

MM calculations are lower precision, but much faster, permitting simulations of whole proteins in boxes of tens of thousands of solvent molecules.[31] We can also use MM force fields to run MD simulations. The specificity of ligand binding can come from the sum of a number interactions of small individual energetic consequence, interactions both within the protein, shaping the binding pocket, and direct interactions with the ligand. MD simulations allow us to let the protein move between different conformations and sample the relevant ligand binding geometries. The drawback of MD simulations is that they are based on MM forcefields, which fail to accurately represent many forms of non-covalent interactions.[32] These are often rooted in subtle rearrangements of electrons; MM force fields do not explicitly treat electrons.[33]

QMMM methods, as their name implies, are methods which employ both types of calculations.[34] Direct interactions with the ligand can be modeled with high level QM theory, and the rest of the protein and solvent with MM theory. These represent the best way of modeling protein-ligand interactions currently available to us. We can even improve on this by calculating ligand binding energies using the ensemble of protein conformations resulting from MD simulations of the protein. It is prohibitively expensive (in terms of computational time) to perform an MD simulation with each unnatural amino acid substitution. Instead, we perform MD simulations with wild type structures and then perform *in silico* mutagenesis to replace the amino acid in question with an unnatural amino acid and calculate a QMMM binding energy. This is the highest level calculation that we can perform, one that is in some ways conceptually similar to our unnatural amino acid experiments. However, it requires some sort of knowledge of the protein's structure, so it cannot be performed in many cases. Therefore, we choose from among the computational methods to apply the most appropriate method given the structural information available to us. We use these methods in combination with unnatural amino acid mutagenesis to make the highest precision experimental and theoretical measurements of ligand binding to ion channels that are currently conceivable.

Dissertation Summary

The work described here can be divided into three sections. The first section discusses biophysical studies of ligand binding to the nAChR. The second section will describe two experiments undertaken to develop the tools for studying important events in learning and memory, processes that primarily involve a different type of LGIC, the glutamate-gated receptor (GluR). The final section collects a miscellany of experiments: some methodological developments for characterizing the aa-tRNA, some experiments done in support of other researchers' work, and some incomplete experiments.

Section 1 comprises Chapters 2, 3, 4, and 5. Chapter 2 outlines the problem tackled in the following chapters, that of understanding the mechanisms by which the nAChR discriminates between three of its principle ligands: ACh and nicotine (Nic), from which it gets its name, and the frog toxin epibatidine (Epi), which is a high affinity compound similar to Nic. (Fig. 9, Center) Chapter 3 describes unnatural amino acid experiments which explored the local pK_a and dielectric of the binding site to understand the protonation state of agonists there. Tethered agonists, unnatural amino acids that have a sidechain that mimics the agonist, are used. Some of these can only activate the channel when protonated, and therefore provide a probe of the binding site pK_a. In Chapter 4, we used QM computational models of interactions of ACh, Nic, and Epi with a binding site indole and carbonyl in concert with unnatural amino acids that probe these interactions to better understand discrimination between the three agonists. The experimental work in Chapter 4 was performed by Amanda Cashin. Chapter 5 describes our state-of-the-art computational model in which a structural model of the extracellular ligand binding domain of the nAChR has been generated, subjected to MD simulations, and then used in QMMM calculations of the binding of nicotinic and cholinergic agonists in the presence of unnatural amino acids in the binding site. These four chapters delineate a very exciting time period in the study of nAChRs. During the course of these studies, undertaken between 1999 and 2005, the model of the nAChR binding site has evolved from the simple, non-atomistic representation at the left of Figure 9 to the precise computational model of Nic binding shown on the right.

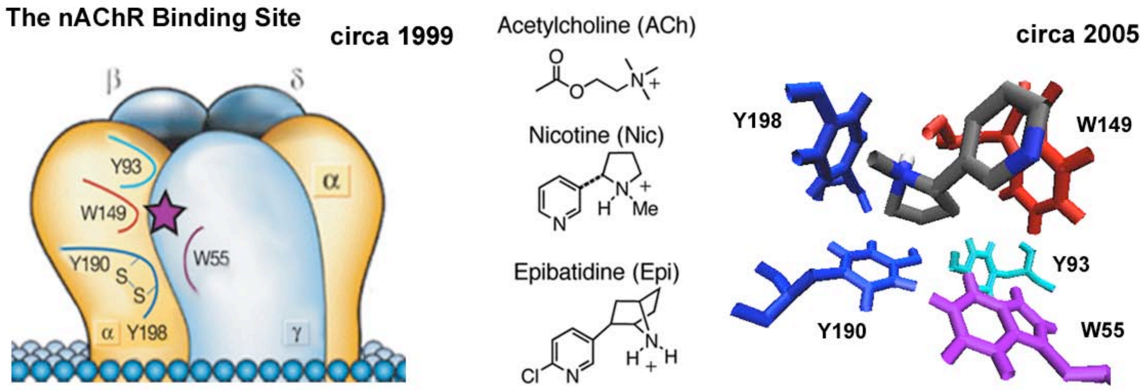


Figure 9. Models of the nAChR Binding Site. Left: Simple, non-atomistic model. Center: The agonists studied. Right: Homology model with bound Nic for QMMM calculations.

The nAChR is essential to neurotransmission at the junction between nerve and muscle cells, and it plays an important role in many central nervous system processes. However, its role in learning and memory is limited, at least in our current molecular models of these events. In a sense, the formation of a memory consists of the strengthening of some synaptic connections and the weakening of others. These processes, termed long term potentiation (LTP) and depression (LTD) respectively, are primarily governed by modifications to GluRs. Two major types of changes are believed to underlie LTP and LTD: alterations to the functional properties of a single glutamate channel and changes in the number of GluRs present at a synapse (trafficking). One common mechanism for initiating both of these changes is the phosphorylation of Ser, Thr, and Tyr hydroxyl groups. Many such residues are present in a GluR and are targets for phosphorylation, making it difficult to understand the effects of phosphorylation at any one residue. While phosphorylation can act as a functional group signal to alter protein function and trafficking, it appears that the initial trigger for both LTP and LTD involves removal of a Mg^{2+} ion from the channel of the NMDA-type GluR. This allows Ca^{2+} flow through the receptor, and the rise Ca^{2+} concentrations leads to changes in the phosphorylation states of GluRs.

Section 2 consists of Chapters 6, 7, and 8. Chapter 6 describes LTP and LTD in more detail and places in context the studies of phosphorylation in Chapter 7 and of the NMDA receptor in Chapter 8. In Chapter 7, we describe the first incorporation of "caged" phosphoamino acids that should permit precise temporal control of the onset of phosphorylation. This cage consists of a photocleavable protecting group applied either to the wild type amino acid or to the chemically synthesized, phosphorylated amino acid (Ser is given as an example in Fig. 10). These studies were performed in conjunction with Deborah Rothman and Barabara Imperiali at Massachusetts Institute of Technology. Chapter 8 details our studies of the molecular determinants of Mg^{2+} binding to the NMDA receptor, which is central to LTP and LTD onset. Katie McMenimen collaborated on this experiment and will continue to characterize this receptor. Our studies of the NMDA receptor are interesting in themselves, and provide us with entrée into the study of GluRs, new to these labs. We hope to incorporate the caged amino acids into GluRs to study the effects of these site-specific modifications in molecular models of learning and memory.

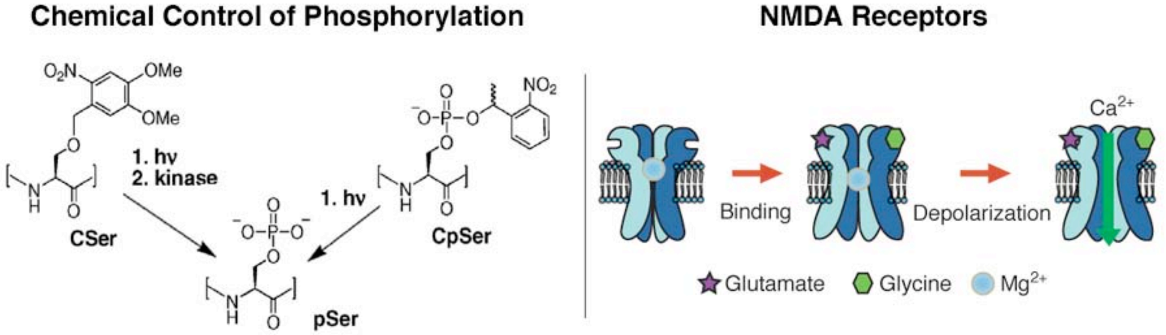


Figure 10. Biochemical Tools for Studying Learning and Memory. Left: Caged amino acids for controlling protein phosphorylation. Right: NMDA receptor gating.

Section 3 is made up of Chapters 9, 10, and 11. This section collects several experiments that do not fit cleanly into the storylines of Sections 1 and 2. Chapter 9 describes the novel application of matrix-assisted laser desorption ionization (MALDI) mass spectrometry to analyzing the purity of the aa-tRNA reagents which are so crucial to the success of unnatural amino acid experiments. In Chapter 10 we report the syntheses of several photoactivatable compounds made in support of David Dahan and Mohammed Dibas' work in using spectroscopic probes to monitor electrophysiologically silent changes in nAChR structure, changes which occur before the pore opens to allow ion flow. Chapter 11 details attempts to apply the tethered agonists from Chapter 3 to a receptor related to the nAChR, the serotonin-gated 5-HT$_3$ receptor.

Conclusion

The study of ion channels has advanced considerably since Hille's 1971 experiments, but despite the impressive efforts of several crystallography labs, ion channel study operates in a structural vacuum unlike that experienced by any other field of chemistry. The dynamic nature of these proteins ensures that what structural insights there are do not tell the whole story of the protein's function. It is in this environment that unnatural amino acid structure/function studies and high level computational modeling can make possible true chemical scale studies of ion channels.

LGICs represent an astoundingly complex target for physical chemistry and are interesting in this context alone. However, the implications of such studies are far-reaching. The understanding of the brain is an exercise with such inherently grand philosophical implications that it should need no additional fanfare as a goal for basic science. It is the role of the chemical characterization of ion channels in the field of neurological medicine that bears further consideration, as it may be even more important.

Neurological disorders cost Americans more quality days lost from their lives than most fatal diseases. Using the measure of disability adjusted life years (DALYs), a study published by the Harvard School of Public Health found that mental illness ranked second only to cardiac diseases in its disease burden on society.[35] The future of the pharmaceutical industry surely lies in this field, but it will require a change in the manner in which drugs are targeted to proteins. Since these diseases are disorders of crucial, endogenous proteins, drugs that completely disable these proteins will not provide effective treatments and will probably be lethal. Neurological disorders require precision drugs that can restore normal protein function. A detailed understanding of the protein's functional subtleties is necessary in developing such drugs. Research like the experiments described here can provide this understanding.

References

(1) Green, T.; Heinemann, S. F.; Gusella, J. F., *Neuron* **1998,** 20, 427-444.

(2) Kandel, E. R.; Schwartz, J. H.; Jessel, T. M., Ed.; "Principles of Neural Science." 4; McGraw-Hill: New York, NY, 2000

(3) http://opal.msu.montana.edu/cftr/IonChannelPrimers/ion_channel_history.htm

(4) Hodgkin, A. L.; Huxley, A. F., *J. Physiol.* **1952,** 116, 449-472.

(5) Hodgkin, A. L.; Huxley, A. F., *J. Physiol.* **1952,** 117, 500-544 Hodgkin, A. L.; Huxley, A. F., *J. Physiol.* **1952,** 116, 497-506 Hodgkin, A. L.; Huxley, A. F., *J. Physiol.* **1952,** 116, 473-496.

(6) Hille, B., *Proc. Natl. Acad. Sci. USA* **1971,** 68, 280-282.

(7) Sherman-Gold, R., Ed.; "The Axon Guide." Axon Instruments, Inc.: Foster City, CA, 1993 Hille, B., *Ion Channels of Excitable Membranes*; Sinauer Associates, Inc.: Sunderland, MA, 2001.

(8) Quick, M. W.; Lester, H. A., Methods for expression of excitability proteins in *Xenopus* oocytes. In *Ion Channels of Excitable Cells*; Narahashi, T., Academic Press: San Diego, 1994; 261-279 Lester, H. A., *Science* **1988,** 241, 1057-1063.

(9) Jiang, Y.; Lee, A.; Chen, J.; Ruta, V.; Cadene, M.; Chait, B. T.; MacKinnon, R., *Nature* **2003,** 423, 33-41 Jiang, Y.; Lee, A.; Chen, J.; Cadene, M.; Chait, B. T.; MacKinnon, R., *Nature* **2002,** 417, 515-522 Dutzler, R.; Campbell, E. B.; Cadene, M.; Chait, B. T.; MacKinnon, R., *Nature* **2002,** 415, 287-294 Zhou, Y.; Morais-Cabral, J. H.; Kaufman, A.; MacKinnon, R., *Nature* **2001,** 414, 43-48 Doyle, D. A.; Morais Cabral, J.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R., *Science* **1998,** 280, 69-77.

(10) Chang, G.; Spencer, R. H.; Lee, A. T.; Barclay, M. T.; Rees, D. C., *Science* **1998,** 282, 2220-2226 Bass, R. B.; Strop, P.; Barclay, M.; Rees, D. C., *Science* **2002,** 298, 1582-1587.

(11) Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van Der Oost, J.; Smit, A. B.; Sixma, T. K., *Nature* **2001,** 411, 269-276 Celie, P. H. N.; van Rossum-Fikkert, S. E.; van Dijk, W. J.; Brejc, K.; Smit, A. B.; Sixma, T. K., *Neuron* **2004,** 41, 907-914 Armstrong, N.; Sun, Y.; Chen, G. Q.; Gouaux, E., *Nature* **1998,** 395, 913-917 Mayer, M. L.; Olson, R.; Gouaux, E., *J. Mol. Biol.* **2001,** 311, 815-836.

(12) Grutter, T.; Changeux, J. P., *Trends Biochem. Sci.* **2001,** 26, 459-463 Karlin, A., *Nat. Rev. Neurosci.* **2002,** 3, 102-114.

(13) Dougherty, D. A., *Curr. Opin. Chem. Biol.* **2000,** 4, 645-652.

(14) Cornish, V. W.; Mendel, D.; Schultz, P. G., *Ang. Chem. Int. Ed.* **1995,** 34, 621-633.

(15) Petersson, E. J.; Brandt, G. S.; Zacharias, N. M.; Dougherty, D. A.; Lester, H. A., *Methods Enzymol.* **2003,** 360, 258-273.

(16) Saks, M. E., *Proc. Natl. Acad. Sci. USA* **2001,** 98, 2125-2127.

(17) Hecht, S. M.; Alford, B. L.; Kuroda, Y.; Kitano, S., *J. Biol. Chem.* **1978,** 253, 4517-4520.

(18) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G., *Science* **1989,** 244, 182-188.

(19) Koh, J. T.; Cornish, V. W.; Schultz, P. G., *Biochem.* **1997,** 36, 11314-11322.

(20) Ellman, J.; Mendel, D.; Anthonycahill, S.; Noren, C. J.; Schultz, P. G., *Methods Enzymol.* **1991,** 202, 301-336.

(21) Killian, J. A.; Van Cleve, M. D.; Shayo, Y. F.; Hecht, S. M., *J. Am. Chem. Soc.* **1998,** 120, 3032-3042 Mamaev, S. V.; Laikhter, A. L.; Arslan, T.; Hecht, S. M., *J. Am. Chem. Soc.* **1996,** 118, 7243-7244 Arslan, T.; Mamaev, S. V.; Mamaeva, N. V.; Hecht, S. M., *J. Am. Chem. Soc.* **1997,** 119, 10877-10887.

(22) Sisido, M., *Nucleic Acids Symp Ser* **2000**, 275-276 Ninomiya, K.; Minohata, T.; Nishimura, M.; Sisido, M., *J. Am. Chem. Soc.* **2004,** 126, 15984-15989.

(23) Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S., *Nucleic Acids Res.* **2002,** 30, 4692-4699 Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S., *Nat. Methods* **2005,** 2, 201-206.

(24) Hohsaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M., *J. Am. Chem. Soc.* **1999,** 121, 34-40.

(25) Nowak, M. W.; Gallivan, J. P.; Silverman, S. K.; Labarca, C. G.; Dougherty, D. A.; Lester, H. A., *Methods Enzymol.* **1998,** 293, 504-529 Beene, D. L.; Dougherty, D. A.; Lester, H. A., *Curr. Opin. Neurobiol.* **2003,** 13, 264-270.

(26) England, P. M.; Lester, H. A.; Dougherty, D. A., *Tetrahedron Lett.* **1999,** 40, 6189-6192.

(27) Amer Chemical Soc Tan, Z. P.; Forster, A. C.; Blacklow, S. C.; Cornish, V. W., *J. Am. Chem. Soc.* **2004,** 126, 12752-12753.

(28) Saks, M. E.; Sampson, J. R.; Nowak, M. W.; Kearney, P. C.; Du, F.; Abelson, J. N.; Lester, H. A.; Dougherty, D. A., *J. Biol. Chem.* **1996,** 271, 23169-23175.

(29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; J. A. Montgomery, J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; et al. *Gaussian 98 (Revision A.9)*, Gaussian, Inc.: Pittsburgh PA, 1998.

(30) Foresman, J. B.; Frisch, E., *Exploring Chemistry with Electronic Structure Methods*. ed.; Gaussian, Inc.: Pittsburgh, PA, 1996.

(31) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A., *AMBER 6*. ed.; University of California: San Fransisco, CA, 1999 Jorgensen, W. L.; Tirado-Rives, J. J., *J. Am. Chem. Soc.* **1988,** 110, 1657 Springer-Verlag Lindahl, E.; Hess, B.; van der Spoel, D., *J. Mol. Modeling* **2001,** 7, 306-317 Mayo, S. L.; Olafcon, B. D.; Goddard, W. A., III, *J. Phys. Chem.* **1990,** 94, 8897-8909.

(32) Chipot, C.; Maigret, B.; Pearlman, D. A.; Kollman, P. A., *J. Am. Chem. Soc.* **1996,** 118, 2998-3005 Felder, C.; Jiang, H.-L.; Zhu, W.-L.; Chen, K.-X.; Silman, I.; Botti, S. A.; Sussman, J. L., *J. Phys. Chem.* **2001,** 105, 1326-1333.

(33) Minoux, H.; Chipot, C., *J. Am. Chem. Soc.* **1999,** 121, 10366-10372 Caldwell, J. W.; Kollman, P. A., *J. Am. Chem. Soc.* **1995,** 117, 4177-4178 Cubero, E.; Luque, F. J.; Orozco, M., *Proc. Natl. Acad. Sci.* **1998,** 95, 5976-5980.

(34) Schrödinger *QSite*, 3.5; Schrödinger, LLC: New York, NY, 2005 Gherman, B. F.; Goldberg, S. D.; Cornish, V. W.; Friesner, R. A., *J. Am. Chem. Soc.* **2004,** 126, 7652-7664.

(35) Murray, C.; Lope, Z. A., Ed.; "The global burden of disease and injury series, volume 1: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020." Harvard University Press: Cambridge, MA, 1996; Vol. 1.

Section 1: Chapter 2

Understanding Ligand Discrimination at the Nicotinic Acetylcholine Receptor Binding Site

The Nicotinic Acetylcholine Receptor

The nicotinic acetylcholine receptor (nAChR) family of ligand-gated ion channels (LGICs) underlies transmission at the neuromuscular synapse and in the central nervous system.[1, 2] It serves as the target for nicotine (Nic) and a number of promising pharmaceuticals for pain-relief, memory enhancement, and the treatment of Parkinson's disease.[3] The nAChR has long served as a prototype for the understanding of ion channels in general. Its abundance in the electric organ of the *Torpedo* ray *electroplax* (nAChR currents give rise to the ray's electric shock) has permitted preparations of large amounts of receptor for various biochemistry applications. This allowed it to be cloned before any other LGIC (simultaneously by three groups in 1982[4]) and for electron microscopy studies of frozen nAChR preparations by Unwin[5], giving us images like that shown on the left of Figure 1 (with computational refinement, resolution has increased, see Chapter 5).

The nAChR has a cylindrical quaternary structure made up of five subunits, each of which is roughly 400 amino acids in size.[1, 6] There are several types of subunits: α, which are always required for channel function, and β, γ, δ, and ε which are also required, but in different combinations depending on the subtype of α subunit in the pentamer. To date, ten α subtypes (α1 - 10) have been identified in vertebrates, as well as four β subtypes (β1 - 4), and one each of the other types.[1] The pentameric assembly can be formed from one type of subunit, as in the α7 nAChR; two types, as in the α4β2 nAChR; or as many as four types, as in the muscle-type nAChR, composed of two α1, one β1, one δ, and one γ (Fig. 1). All of the experimental work described here will be performed on the mouse muscle nAChR; some of the computational work in Chapter 5 will involve the human α7 receptor.

In addition to the overall pentameric topology of the receptor, there are, of course, secondary and tertiary structural elements within each subunit. The N-terminal ligand-binding domain (LBD) is extracellular, four helical transmembrane (TM1 – 4) domains weave through the membrane, and the short C-terminal domain is also extracellular. (Fig. 1) Our understanding of the LBD structure comes primarily from the structure of the homologous acetylcholine binding protein[7], described below. (Fig. 6) The arrangement of the TM regions about the pore is shown on the left of Figure 1. While there is some uncertainty as to the relative positions of TMs 1 and 3, several biochemical studies have shown clearly that TM2 lines the channel pore and that TM4 faces into the membrane.[8]

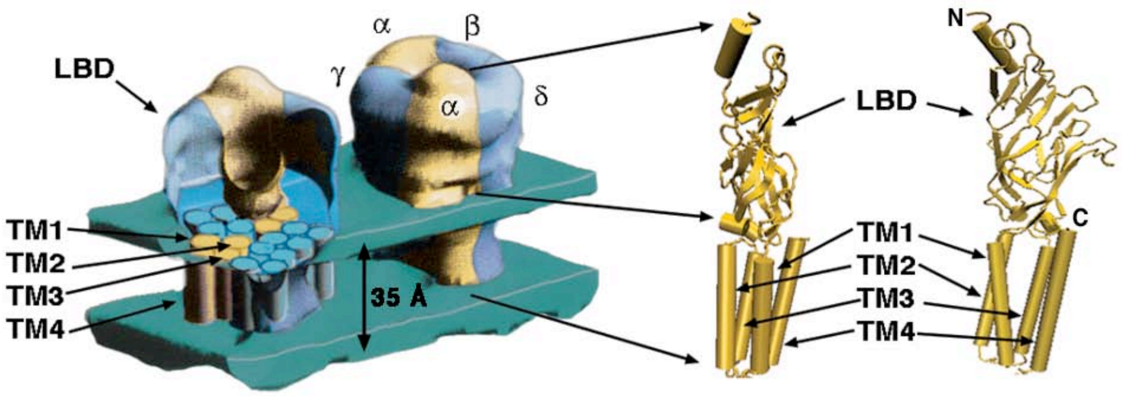


Figure 1. Pentameric assembly of the nAChR. Left: Electron microscopy image of nAChR shown cutaway to reveal schematic representation of the transmembrane helix assembly. Second image shown whole with mouse muscle nAChR subunits labeled. Adapted from Stroud, 1990.[9] Right: Structure of an individual α subunit from Unwin, 2005,[10] with N- and C- termini indicated. Ligand-binding domains (LBDs), and transmembrane helices 1 – 4 (TM1 – TM4) shown in both representations.

When the first unnatural amino acid investigations of the nAChR were undertaken, much of our knowledge of the receptor came from classical biochemical studies. Many of these studies focused on the agonist binding site, which is the area of interest for all of the studies described herein. The pentameric receptor has two binding sites, which are localized primarily on the α subunits. Pioneering work by Karlin and colleagues established that a conserved disulfide bond in the α subunit (Cys 192-193, mouse muscle numbering) was near the binding site.[11] Photoaffinity labeling studies by Changeux, and several radioligand binding studies identified a large number of aromatic residues near the binding site.[6, 12] Later work implicated residues in the γ and δ subunits that may contribute to acetylcholine (ACh) binding.[13] The residues identified in these studies, and the binding domain loops which contain them, are indicated in Figure 2. (Loop designations are given explicitly with the sequence alignments in Chapter 5 Supporting Information.) Though the α/γ binding site is depicted, one must remain aware that there are two binding areas, an α/δ and an α/γ, and that it is the cooperative binding of two ACh molecules that opens the channel.

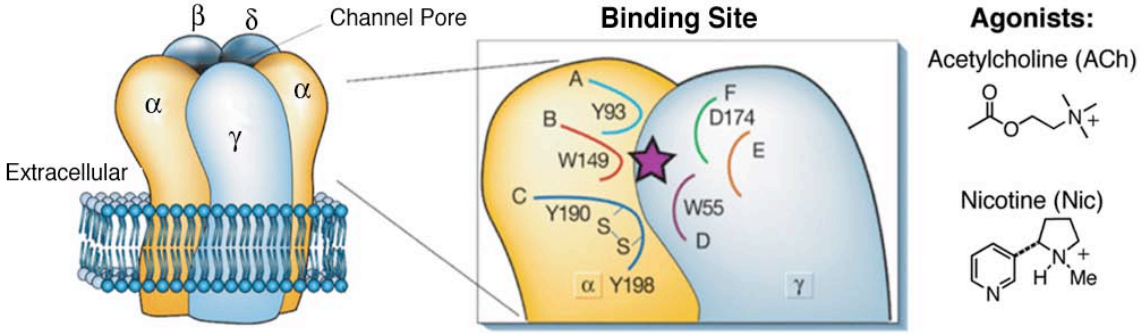


Figure 2. The Agonist Binding Site of the nAChR. Far left: Global layout. Center: Schematic of the agonist binding site, showing residues from the α and γ subunits thought to contribute to binding. The loops that contain the residues are indicated. Right: Common nAChR agonists addressed in our studies.

ACh Binds Through a Cation-π Interaction

The preponderance of aromatic sidechains among binding site residues led us to posit that a cation-π interaction with the quaternary ammonium of ACh may be important to ligand binding. A cation-π interaction involves the interaction of the positive charge on the cation with the electron-rich face of an aromatic moiety.[14] Although aromatic compounds like benzene have no net dipole, the highly polar C(sp^2)-H bonds form a quadrupole, the negatively charged center of which can bind cations. (Fig. 3, Left) This interaction is impossible to investigate in a rational way through conventional mutagenesis; the best that can be done is to demonstrate that one aromatic amino acid can be replaced by another (i.e. that a Trp to Phe mutation is not substantially perturbing). In contrast, unnatural amino acid methods allow us to perturb the electrostatic portion of the cation-π interaction without substantially altering the shape of the residue. We accomplish this by replacing the aromatic, such as Trp, with fluorinated derivatives. (Fig. 3, Right) The polarity of the C(sp^2)-F bond places the partial negative charge on the fluorine. Thus, successively substituting fluorine atoms for the ring hydrogens removes the partial negative charge from the center of the quadrapole and abrogates cation binding in a stepwise fashion. The electrostatic potential surface (red indicates a more favorable interaction with a positive point charge, blue a more unfavorable interaction) provides a colormetric measure of the effects of fluorination on the cation binding ability of the F-Trp series members.

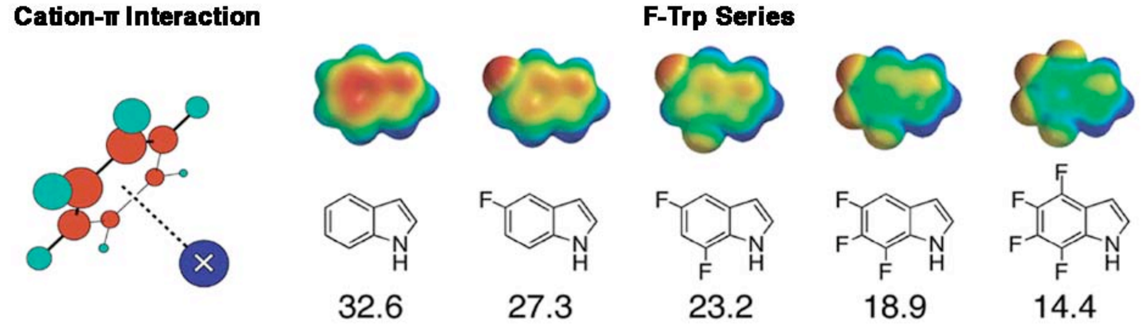


Figure 3. The Cation-π Interaction. Left: Schematic representation of the interaction of a cation with benzene. Right: The F-Trp series: Electrostatic potential surfaces shown on a colormetric scale from – 25 (red) to + 25 (blue) kcal/mol. Calculated, gas phase Na$^+$-binding energies given. Surfaces and energies from HF/6-31G** calculations. Surfaces determined from MO coefficients with a 0.002 e/Å^3 cutoff.

Both the Tyr and Trp residues were explored with unnatural amino acids, but we found compelling evidence for a cation-π interaction only at Trp 149 of the α subunit.[15] In a study described in Zhong *et al.*, the F-Trp series shown above was incorporated at Trp α149, and each successive fluorine resulted in a rightward shift of the dose-response relation.[16] For example, a comparison of the response of nAChRs containing either Trp or F$_2$-Trp to ACh shows that the potency of ACh at a given concentration is decreased for the

F_2-Trp receptor. (Fig. 4, Left) The effects of the full F-Trp series can be seen when we plot the response of each mutant receptor as a function of agonist concentration. (Fig. 4) While agonist potency is a composite of binding site affinity and ability to induce receptor gating, the stepwise effect of fluorination on ACh potency is consistent with our cation-π model, so we interpret this as a binding effect.

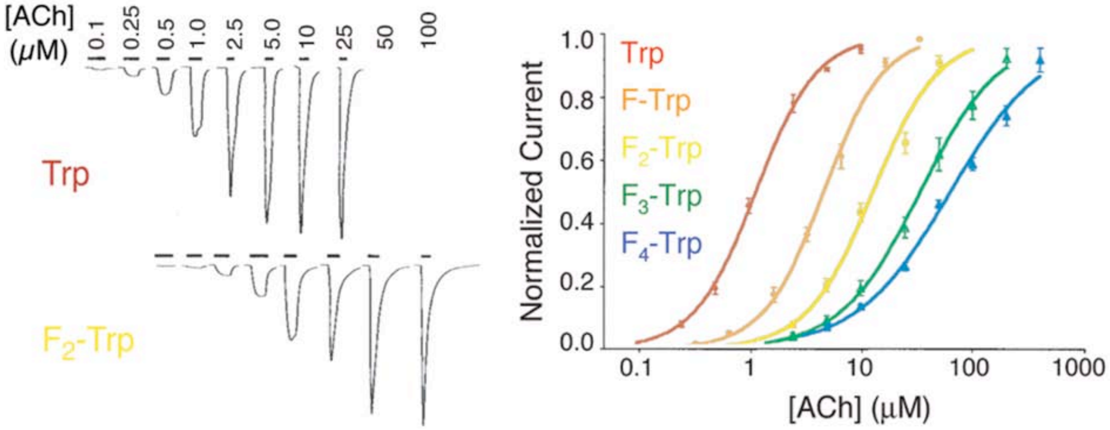


Figure 4. The F-Trp Series at α149. Left: Representative dose-response relations for Trp (wild type) and F2-Trp. Right: Averaged dose-response curves for each of the F-Trp series members at αTrp 149.

Dose-response relations are often characterized by an EC_{50}, the concentration of agonist at which half-maximal activation is evoked. The EC_{50} is determined by a fit to the Hill equation: $I_{Normalized} = 1/(1 + (EC_{50}/[ACh])^n$.[17] The Hill coefficient, n, is often interpreted as a measure of the cooperativity of binding, and we generally find Hill coefficients greater than 1 for the two-agonist process of nAChR gating. We plotted the ratio of the EC_{50} of the mutant receptor to the EC_{50} of the wild type receptor (in fact, the logarithm of this ratio, to put this on an "energy" scale) against the calculated cation-π binding ability of the F-Trp series member sidechain. Gratifyingly, these data could appear linear. (Fig. 5)

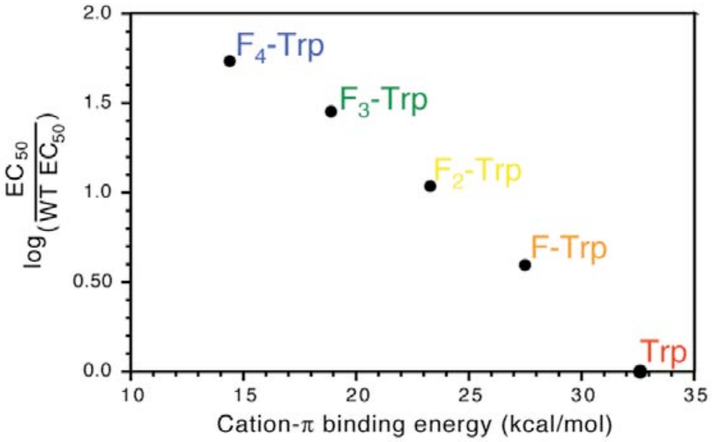


Figure 5. Fluorination Plot for ACh at αTrp 149. The logarithm of the ratio of the experimental EC_{50}s of mutant to wild type (Trp) receptors versus the calculated cation-π binding energies (from above).

The linear relationship observed between *ab initio*, quantum mechanical calculations of cation-π binding and *in vivo* measures of nAChR activation supports a model of binding in which the quaternary ammonium of ACh makes van der Waals contact with the face of Trp 149. The primacy of Trp 149 was further confirmed by tethered agonist studies in which unnatural amino acids bearing a mimic of the ammonium center were able to generate self-gating receptors.[16, 18] (A detailed description of this work is given in Chapter 3.) The most effective tethered agonist was one which could place the ammonium in the same position that it would adopt if it were placed above the six-membered ring of Trp 149.

The Structure of a nAChR Binding Domain Homolog

Subsequent to these studies, the Sixma group solved the crystal structure of a protein related to the extracellular LBD of the nAChR. The acetylcholine binding protein (AChBP) is a soluble protein secreted from snail glia that acts as a "sponge" to soak up excess ACh in the synapse and nAChR-targeted toxins released by other snails.[19] AChBP is a homopentameric protein in which each monomer is made up of an immunoglobulin domain-like fold of β-sheets. The overall size and shape of the protein is a good fit to the LBD of Unwin's cryo-electron microscopy structure.[20] The primary sequence is 20-26% identical to the LBD of members of the nAChR family, and 68% homologous to the LBD of α7, the homopentameric member to which it is most similar.[7]

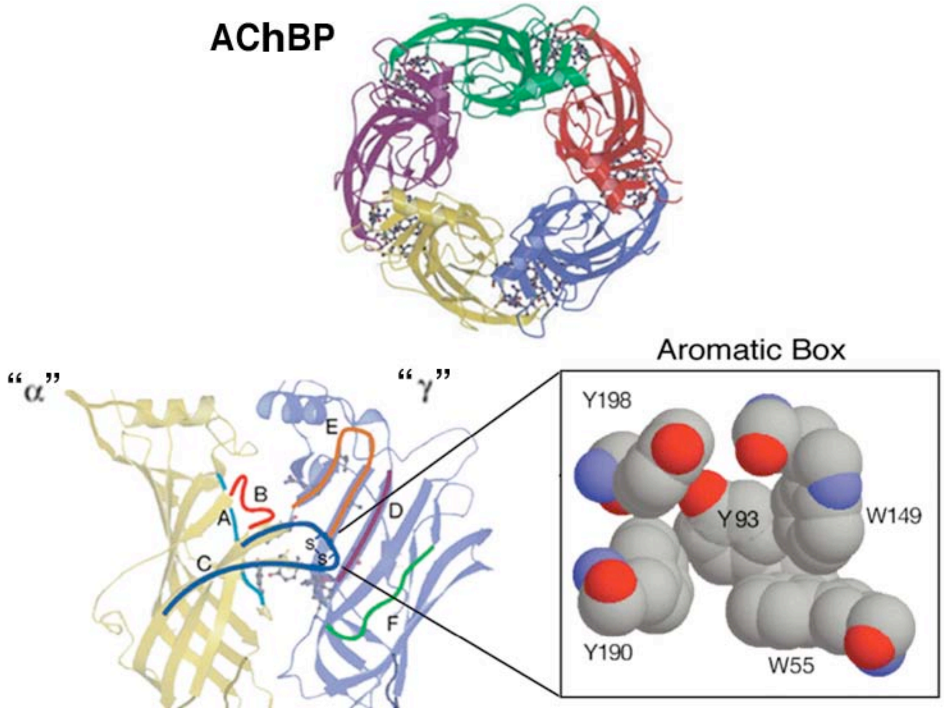


Figure 6. nAChR Structural Insights from AChBP. Top: The homopentameric structure of AChBP. Left: Two subunits, shown in a mimic of the α/γ binding site layout of Figure 2. The binding site loops are colored in the same fashion. Right: The Trp and Tyr residues of the binding site form an aromatic box in which ligands will bind.

Crystal structures must always be vetted by comparison to functional biochemical experiments, so it was important that the AChBP structure was consistent with the wealth of previous data. The loops of primary sequence that had been identified as contributing to the binding site (colored in Fig. 6 as they are in Fig. 2) were indeed found to be at the subunit interface where binding occurs. While AChBP is a homopentamer, residues from the γ and δ subunits map well to the side of the interface opposite the α subunit binding determinants. The aromatic residues labeled in the photoaffinity studies form an "aromatic box" (Fig. 6, Right), ready to bind the agonist.[21]

Although the original AChBP structure was determined without agonist bound, it contained some evidence that our assignment of the cation binding site to Trp 149 was correct. A molecule of HEPES, a buffer molecule that contains ammonium centers at physiological pH, was found in each of the five AChBP binding sites.[7] It was found with one of these ammoniums directly above Trp 143 of AChBP, which aligns with muscle α Trp 149. The degree to which our high-precision nAChR studies agreed with the crystallographic images of AChBP gave us confidence in using the F-Trp series to characterize the cation binding of other nAChR agonists (some of which we studied previous to the emergence of the AChBP structure).

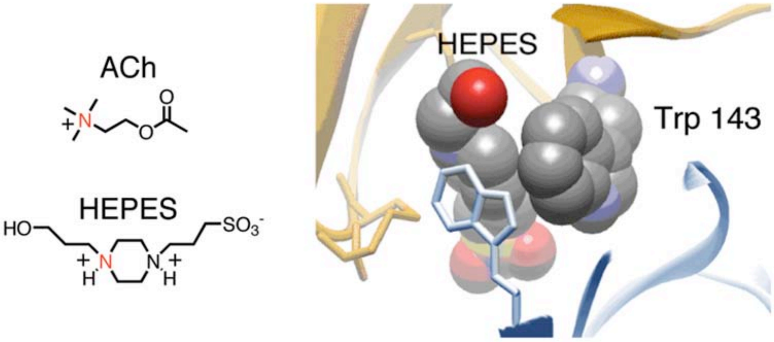


Figure 7. HEPES Ammonium Binds to AChBP Trp 143. Left: ACh and HEPES, ammonium of import highlighted. Right: HEPES interaction with AChBP Trp 143.

Nic Binding to the nAChR

The first molecule to be studied was nicotine, the other agonist from which the nAChR gets its name. Experiments by Cohen had shown that a Nic-based photoaffinity probe labeled the same residues as ACh analogs, so the two were believed to bind identically.[22] However, unlike ACh, Nic did not show a systematic change in binding with the F-Trp series.[23] (Nic is a poor agonist, so we made 9'S mutations in both the β and γ subunit TM2 regions to lower EC_{50}s.) While incorporation of F-Trp did cause a rightward shift in the Nic dose-response curve, subsequent fluorination had little further effect.

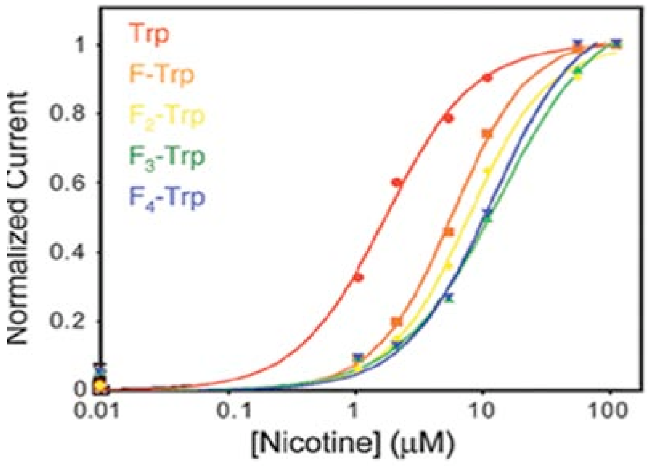


Figure 8. Dose-Response Relations for Nic at nAChRs (β/γ 9'S) with the F-Trp Series at α149.

When the data is presented in the form of a fluorination plot, we can see that the slope of the Nic fluorination plot changes between the F-Trp and F_2-Trp points. We interpret this as a change in mechanism, akin to the change in mechanism assigned to a change in slope in a Hammet plot.[24] In fact, our Trp fluorination plot is essentially a Hammet plot. The log EC_{50} reflects changes in ΔΔG for the binding and gating process, and Trp fluorination constitutes a perturbation series for this process (the calculated energy of an idealized Na^+ cation-π binding event is analogous to σ). For ACh, there is a linear relationship between these quantities, consistent with our assignment of a cation-π "mechanism" to the binding/gating process. For Nic, there are at least two different slopes, possibly indicative of a change in the interaction of Nic with the functional groups in the binding pocket, altering the "mechanism" of its binding.

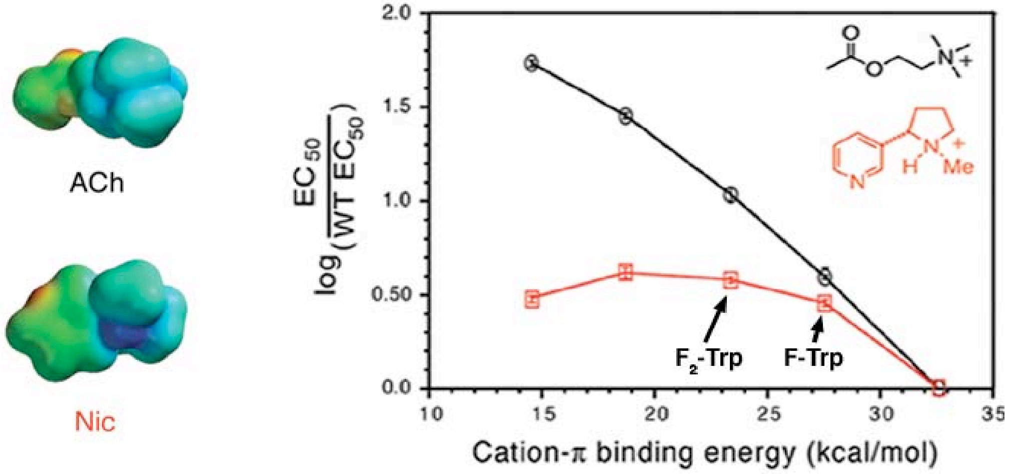


Figure 9. Comparison of ACh and Nic Binding to the nAChR. Left: Electrostatic potential surfaces shown on a colormetric scale from - 5 (Red) to + 160 (Blue) kcal/mol. Surfaces determined from MO coefficients of AM1 calculations with a 0.002 e/Å^3 cutoff. Right: Fluorination plot for Nic and ACh at α149 of the nAChR.

Trp fluorination is a nearly isosteric perturbation to the electrostatic component of the cation-π interaction, but it still may have the effect of altering the mechanism of ligand binding. Therefore, it is only in the stepwise alteration by several fluorinations that we can clearly demonstrate a cation-π interaction. The fact that Nic is a nearly 100-fold lower affinity ligand for the muscle nAChR than ACh may help to explain why the fluorination series affects the two ligands differently.[23] It may be that there is some weak attraction to the Trp ring for Nic, and that one fluorine is sufficient to eliminate this and cause Nic to bind to another moiety in the binding site. For ACh, although the interaction is weakened, the geometry of interaction with Trp 149 is more or less maintained because the rest of its interactions with the binding pocket are sufficiently strong to hold it in place. As a result, further fluorination continues to affect the cation-π interaction through the same mechanism. Regardless of the precise mechanistic differences, the clear implication of this study was that ACh and Nic bind differently to the nAChR. Although Cohen's work had shown that Nic and ACh bound to the same area of the protein in a global sense, the precise manner of their binding seemed to be different.[22] It is exactly this sort of subtle difference that we seek to explain with our studies. Furthermore, it is this sort of difference that can be exploited by the pharmaceutical industry to develop drugs which may compete effectively with Nic to treat addiction while still allowing ACh to bind.[25]

Not only does Nic not bind to Trp 149 through a cation-π interaction, but no correlation with fluorination was found with the other aromatic residues in the aromatic box.[26] (Not all were probed.) The fact that Nic does not simply bind to another residue with its cation lead us to examine the two agonists to explain the differences in binding mechanism. While Nic and ACh may appear quite different in their schematic representations, (Fig. 9, Inset) an examination of their electrostatic potential surfaces (Fig. 9, Left) shows that their shape and charge distribution are actually quite similar. This would make them well suited to bind to the same site on the nAChR, consistent with binding and structure function studies that had implicated the same residues in Nic and ACh binding. Thus, our demonstration that the precise mechanism of binding was not the same for the cholinergic and nicotinic agonists flew in the face of all previous models of their binding.

The Binding of Nic and ACh Analogs

We sought to further understand the differences in the binding of cholinergic and nicotinic agonists by employing analogs of the agonists. We suspected that the subtle differences in binding might come from the subtle differences in the cationic centers of the agonists. The ammonium centers of Nic and ACh differ in that Nic contains a tertiary amine while ACh contains a quaternary ammonium. Despite the aforementioned similarities in the overall electrostatic potential surfaces of the two agonists, one can see that nicotine's tertiary center has a focused positive charge that lies on the proton. The positive charge on the ACh ammonium is more diffuse. We employed a tertiary analog of ACh, norACh, and a quaternary analolg of Nic, MeNic, to test the import of charge distribution at the cation center. The affinities of these agonists were tested with F-Trp series at Trp 149 as with Nic and ACh.[23] The data from these studies are shown in Figure 10 along with their electrostatic potential surfaces.

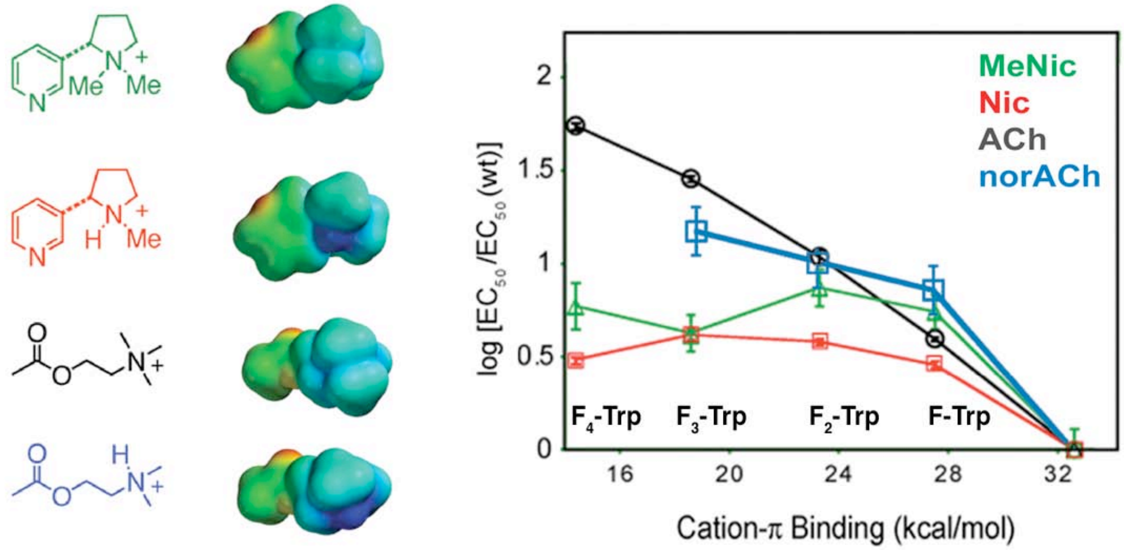


Figure 10. norACh and MeNic Binding to the nAChR. Left: Electrostatic potential surfaces of MeNic, Nic, ACh and norACh shown on a colormetric scale from - 5 to + 160 kcal/mol. Surfaces determined from MO coefficients of AM1 calculations with a 0.002 e/Å^3 cutoff. Right: Fluorination plot for MeNic, Nic, ACh, and norACh at α149 of the nAChR.

The results of this study are difficult to interpret. The binding affinities of norACh and MeNic are similarly compromised for mono- and difluorotryptophan, but their binding curves are divergent for trifluorotryptophan. The MeNic curve is relatively flat, and appears to be Nic-like. The norACh fluorination series is not so nearly linear as the ACh data, but straighter than for either Nic compound. Our Hammet plot model for understanding this data would dictate that a change in mechanism occurs at the F-Trp point for all ligands but ACh. Nic and MeNic alter binding modes to one in which fluorination of Trp149 has essentially no further effect. Although the slope of the norACh fluorination plot has an inflection point at the F-Trp datum, it does show a monotonic change in potency with Trp fluorination. Recall that the slope of the fluorination plot reflects the change in agonist potency relative to the calculated change in the binding of a Na^+ cation in an idealized geometry. It is possible that the slope of the norACh plot for F_1-Trp, F_2-Trp, and F_3-Trp is indicative of a norACh binding geometry in which norACh still interacts with the π-face of Trp 149 but in a geometry for which fluorination of the Trp ring has less of an effect. ACh is the only agonist that shows a clear cation-π interaction, and it is the only high affinity agonist among those in Figure 10. It may be that only in the case of ACh are the other ligand binding determinants sufficiently strong to hold the cationic center in the same orientation throughout the F-Trp series.

One conclusion from these studies seems clear: that agonist affinity is not related in a simple way to the degree of substitution at the cationic nitrogen center. norACh (EC_{50} = 140 μM with one 9'S) is a poor agonist; its EC_{50} is more than 100 times greater than the EC_{50} of ACh (EC_{50} = 1.2 μM with one 9'S).[23] It is therefore tempting to conclude that the degree of alkylation at the ammonium center is important, and that quaternary ammoniums bind better than tertiary amines. However, this does not prove true for the Nic compounds: MeNic (EC_{50} = 0.8 μM with two 9'S) has an EC_{50} comparable to Nic (EC_{50} = 1.4 μM with two 9'S).[23] It is somewhat surprising that methylation at the pyrolidine center has as little effect as it does, given that this site is prochiral. One might expect that only one protonated "enantiomer" of Nic would bind – that the binding site would have a preference for whether the *N*-methyl group of Nic was up or down. However, it seems that this does not matter, as MeNic would effectively place the methyl groups in both possible locations, but this has little effect on its EC_{50} relative to Nic.

In short, the MeNic and norACh studies were data-rich, but largely inconclusive. They did suggest the following other experiments:

Perhaps the reason that norACh is so much weaker an agonist than ACh is that the tertiary norACh is not protonated in the binding site of the receptor. Many proteins have pockets with local environments very different from the bulk solvent in which they are surrounded.[27] It is possible that although one would expect a tertiary amine like norACh or Nic to be protonated physiological pH (7.5; the agonists were also studied at 6.5 and no difference was observed in the size of currents, an issue that will be revisited in Chapter 3), they are not actually protonated in the binding site. Since the cationic moiety appears to be important for channel activation, these would be poor agonists. They would certainly not show a cation-π interaction in the F-Trp series because there would be no cation. In Chapter 3, we investigate the local pK_a of the binding site using an unnatural amino acid with an amine moiety attached to a Tyr sidechain by a carbon tether, termed a tethered agonist. Since we believe that placing a cation in the binding site is crucial for channel activation, this should activate the channel when the amine is protonated, making current a measure of tethered agonist protonation. Unlike normal agonists, the tethered agonist cannot diffuse out of the binding pocket when it is deprotonated, only its protonation state will change. This allows us to use the tethered agonist as a probe of the local pK_a of the binding site.

The only high affinity agonist we have examined thus far, ACh, has exhibited a clear cation-π interaction, while three weak agonists have not. We sought to understand this in terms of the alkylation state of the agonist. However, although converting ACh to a tertiary amine made it a weak agonist (norACh), while converting Nic to a quaternary ammonium (MeNic) did not make it a high affinity agonist or establish a cation-π interaction. In Chapter 4, we examine a high-affinity nicotine-like agonist, epibatidine, to determine whether it binds *via* a cation-π interaction. Epibatidine is a secondary amine, so testing it with the F-Trp series should tell us whether a non-quaternary amine is capable of forming a cation-π interaction in the nAChR binding site.

Nic does not bind to Trp 149 or any other nAChR binding site aromatic through a cation-π interaction.[23] However, it does bind to the nAChR, albeit with low affinity. Therefore, it must interact with some element in the binding site. This other binding determinant may take on a greater importance when a fluorinated Trp is present at Trp 149, part of the change in mechanism in our Hammet plot interpretation of the Nic data. A possible culprit is the carbonyl of Trp 149, which aims prominently into the binding site. This could act as a hydrogen bond acceptor for the protonated amine of Nic, so that it would be bound in a very similar place (and thus produce similar photoaffinity labeling results) but it would not make van der Waals contact with the face of Trp 149. In Chapter 4, we will probe this hydrogen bond by substituting a hydroxy acid at position 150, converting this backbone amide to an ester. Esters are weaker hydrogen bond acceptors than amides, so this should weaken Nic binding.[28] We will also test the effects of this alteration on epibatidine and ACh binding.

The Nic, norACh, and MeNic data do not show a cation-π interaction in the studies described here, but Trp fluorination does have some effect. The fact that the measured changes in EC_{50} do not match the calculated binding energies in some ways reflects the simplicity of the computational model. The calculated binding energies reported here reflect the highly idealized binding of a Na^+ cation to fluorinated indoles.[16] A perfect computational model of agonist binding should recapitulate the observed effects of our unnatural mutations on agonist binding. In Chapters 4 and 5 we attempt to generate more complex computational models to describe the binding site data obtained with our structure/function studies. Chapter 4 describes some *ab initio* quantum mechanical (QM) models that attempt to take into account both the aromatic face of Trp 149 and the carbonyl studied by ester substitution. In Chapter 5, we generate a model of the nAChR binding site that lies at the forefront of computational studies of ligand binding. We use the AChBP structure to generate a homology model of the LBD of the nAChR, which is then minimized and allowed to equilibrate with or without ligand in molecular dynamics simulations with explicit solvent. These structures are then used in quantum mechanics/ molecular mechanics (QMMM) calculations of ligand binding affinity. In this mixed method QMMM work, we calculate the change in agonist binding energy in response to *in silico* substitution with the unnatural amino acid in question. In this way we perform the highest precision calculation of binding affinity that can be reasonably performed, just as we perform the highest precision experimental measure of nAChR ligand binding that can reasonably be performed.

References

(1) Corringer, P. J.; Le Novere, N.; Changeux, J. P., *Annu. Rev. Pharmacol. Toxicol.* **2000,** 40, 431-458.

(2) Dajas-Bailador, F.; Wonnacott, S., *Trends Pharmacol. Sci.* **2004,** 25, 317-324.

(3) Newhouse, P.; Singh, A.; Potter, A., *Curr. Top. Med. Chem.* **2004,** 4, 267-282 Sher, E.; Chen, Y.; Sharples, T. J.; Broad, L. M.; Benedetti, G.; Zwart, R.; McPhie, G. I.; Pearson, K. H.; Baldwinson, T.; De Filippi, G., *Curr. Top. Med. Chem.* **2004,** 4, 283-297 Decker, M. W.; Rueter, L. E.; Bitner, R. S., *Curr. Top. Med. Chem.* **2004,** 4, 369-384.

(4) Ballivet, M.; Patrick, J.; Lee, J.; Heinemann, S., *Proc. Natl. Acad. Sci. USA* **1982,** 79, 4466-4470 Devillers-Thiery, A.; Giraudat, J.; Bentaboulet, M.; Changeux, J. P., *Proc. Natl. Acad. Sci. USA* **1983,** 80, 2067-2071 Noda, M.; Takahashi, H.; Tanabe, T.; Toyosato, M.; Furutani, Y.; Hirose, T.; Asai, M.; Inayama, S.; Miyata, T.; Numa, S., *Nature* **1982,** 299, 793-797.

(5) Toyoshima, C.; Unwin, N., *Nature* **1988,** 336, 247-250.

(6) Arias, H. R., *Brain Res. Rev.* **1997,** 25, 133-191.

(7) Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van Der Oost, J.; Smit, A. B.; Sixma, T. K., *Nature* **2001,** 411, 269-276.

(8) Leite, J. F.; Blanton, M. P.; Shahgholi, M.; Dougherty, D. A.; Lester, H. A., *Proc. Natl. Acad. Sci. USA* **2003,** 100, 13054-13059 Leite, J. F.; Rodrigues-Pinguet, N.; Lester, H. A., *J. Clin. Invest.* **2003,** 111, 436-437 Wilson, G. G.; Karlin, A., *Proc. Natl. Acad. Sci. USA* **2001,** 98, 1241-1248.

(9) Stroud, R. M.; McCarthy, M. P.; Shuster, M., *Biochem.* **1990,** 29, 11009-11023.

(10) Unwin, N., *J. Mol. Biol.* **2005,** 346, 967-989.

(11) Damle, V. N.; Karlin, A., *Biochem.* **1980,** 19, 3924-3932.

(12) Galzi, J. L.; Bertrand, D.; Devillers-Thiery, A.; Revah, F.; Bertrand, S.; Changeux, J. P., *FEBS Lett.* **1991,** 294, 198-202.

(13) Czajkowski, C.; Kaufmann, C.; Karlin, A., *Proc. Natl. Acad. Sci .USA* **1993,** 90, 6285-6289.

(14) Ma, J. C.; Dougherty, D. A., *Chem. Rev.* **1997,** 97, 1303-1324.

(15) Kearney, P. C.; Nowak, M. W.; Zhong, W.; Silverman, S. K.; Lester, H. A.; Dougherty, D. A., *Mol. Pharmacol.* **1996,** 50, 1401-1412 Kearney, P. C.; Nowak, M. W.; Zhong, W.; Silverman, S. K.; Lester, H. A.; Dougherty, D. A., *Mol. Pharmacol.* **1996,** 50, 1401-1412.

(16) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A., *Proc. Natl. Acad. Sci. USA* **1998,** 95, 12088-12093.

(17) Hille, B., *Ionic Channels of Excitable Membranes*; Sinauer Associates, Inc.: Sunderland, MA, 1992.

(18) Li, L. T.; Zhong, W. G.; Zacharias, N.; Gibbs, C.; Lester, H. A.; Dougherty, D. A., *Chem. Biol.* **2001,** 8, 47-58.

(19) Smit, A. B.; Syed, N. I.; Schaap, D.; van Minnen, J.; Klumperman, J.; Kits, K. S.; Lodder, H.; van der Schors, R. C.; van Elk, R.; Sorgedrager, B.; Brejc, K.; Sixma, T. K.; Geraerts, W. P., *Nature* **2001,** 411, 261-268.

(20) Unwin, N.; Miyazawa, A.; Li, J.; Fujiyoshi, Y., *J Mol Biol* **2002,** 319, 1165-1176 Unwin, N., *FEBS Lett.* **2003,** 555, 91-95.

(21) Dougherty, D. A.; Lester, H. A., *Nature* **2001,** 411, 252-253, 255.

(22) Sullivan, D. A.; Cohen, J. B., *J. Biol. Chem.* **2000,** 275, 12651-12660.

(23) Beene, D. L.; Brandt, G. S.; Zhong, W.; Zacharias, N. M.; Lester, H. A.; Dougherty, D. A., *Biochem.* **2002,** 41, 10262-10269.

(24) Carrol, F., *Perspectives on Structure and Mechanism in Organic Chemistry*; Brooks/Cole Publishing Co.: Pacific Grove, CA, 1998.

(25) Bunnelle, W. H.; Dart, M. J.; Schrimpf, M. R., *Curr. Top. Med. Chem.* **2004,** 4, 299-334.

(26) Zhong, W.; Unpublished Work.

(27) Highbarger, L. A.; Gerlt, J. A.; Kenyon, G. L., *Biochem.* **1996,** 35, 41-46.

(28) Koh, J. T.; Cornish, V. W.; Schultz, P. G., *Biochem.* **1997,** 36, 11314-11322.

Section 1: Chapter 3

Probing the Local pK_a of the Nicotinic Acetylcholine Receptor Binding Site with Tethered Agonist Unnatural Amino Acids

Nicotine and Acetylcholine Bind Differently

Several previous investigations of ligand binding to the nicotinic acetylcholine receptor (nAChR) have implied that the two principle agonists, nicotine (Nic) and acetylcholine (ACh), bind to the nAChR through different non-covalent interactions. (See Chapter 2) ACh exhibited a linear correlation between potency at the muscle nAChR and calculated measures of idealized cation-π binding ability at Trp 149 of the α subunit. Nic showed a more complex behavior when we perturbed its cation-π binding to Trp 149. Nic also did not seem to enjoy a cation-π interaction with any of the other aromatic binding site residues. We attempted to test a model of ligand binding that attributed these differences to the degree of substitution at the ammonium center. Nic and ACh analogs for which the degree of substitution was altered (MeNic and NorACh) were employed. However, the results of studies that tested the cation-π binding ability of these analogs failed to clarify the issue.

We posited that the lack of a cation-π interaction for Nic may be attributable to a perturbed pK_a at the nAChR binding site: a deprotonated Nic is not a cation, and cannot participate in a cation-π interaction. Many proteins have pockets with local environments very different from the bulk solvent by which they are surrounded.[1] It is possible that although one would expect the pyrolidine nitrogen of Nic to be protonated at physiological pH, it is not actually protonated in the binding site.[2] To test this theory, we needed to place the ammonium center in the binding site of the nAChR in such a way that it would always be there, regardless of its protonation state. We used tethered agonists, which had previously been used as tests of the binding site cation-π interaction, to place an ammonium moiety in the binding site. If a cation-π interaction is necessary for nAChR activation, channel activation should report on the protonation state of the tethered agonist in the binding site and allow us to measure the local pK_a.

Previous Tethered Agonist Studies

Zhong *et al.* gave compelling evidence for a cation-π interaction with αTrp149 in the binding of ACh as part of the gating process of the nAChR.[3] On binding, the cationic, quaternary ammonium group of ACh makes van der Waals contact with the face of the Trp sidechain. Specifically, calculations indicate that the geometry that produces the strongest binding positions the quaternary ammonium over the six-membered ring of the indole moiety. (Fig. 1) These are supported by experiments in which Trp is replaced by Tyr (whose ring lines up with Trp's 5-membered ring) and a dramatic increase in the EC_{50} is observed.[4] This suggests that placing a cationic group in the 6-membered ring position in the binding pocket of the nAChR will produce a constitutively active receptor, i.e. one that opens in the absence of agonist. (Fig. 1) Initial work by Wenge Zhong showed that a quaternary amine linked to a tyrosine ring by a three carbon chain (TyrO3Q) does indeed yield constitutive activity.[3]

This was not the first time that the tethered agonist strategy was employed with the nAChR. In pioneering experiments by Karlin and later work by Lester, tethered agonists were introduced by chemically modifying the reduced Cys 192-193 disulfide bond, producing a constitutively active receptor.[5] Cohen expanded on this methodology by introducing Cys residues using site-directed mutagenesis and then attaching tethered agonists with MTS reagents.[6] However, at α149 the Cys mutation made the receptor unresponsive to ACh, which is not surprising given that residue's crucial role.

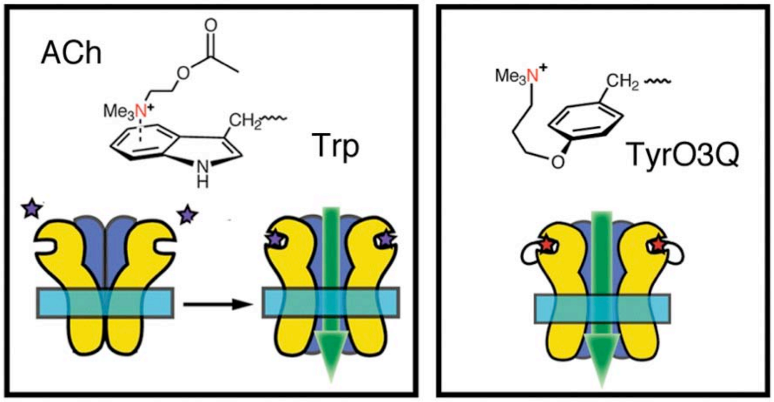


Figure 1. Tethered Agonism. Left: Quaternary ammonium of ACh positioned over the six-membered ring of αTrp149, and induced channel opening. Right: Structure of TyrO3Q, showing that the tethered quaternary ammonium can be positioned in roughly the same location as the ammonium of ACh.

It is important to appreciate that gating the nAChR is a complex process in which the agonist must bind to the receptor and initiate opening of an ion channel gate that is far from the binding site.[7] It is envisioned that agonist binding induces a conformational change that shifts a pre-existing equilibrium between closed and open states of the channel toward the open state. The structural perturbation of introducing a tethered agonist near the binding site could disrupt the protein in a manner that changes the equilibrium and promotes gating. While the tethered agonist's effects are observed in terms of an ion flux through a pore 50 Å away, subtle changes in its geometry and positioning can be crucial to its efficacy. The following work fostered an understanding of the binding site by establishing constraints on both the properties of the tethered agonist and its location.

Probing the Binding Site of the nAChR with a Series of Tethered Agonists

Zhong observed constitutive activity by incorporation of TyrO3Q (Fig. 2) at α149, but Li *et al.* added depth to the study by incorporating tethered agonists of varying lengths (TyrOnQ; n= 2, 3, 4 and 5) at Trp α86, Tyr α93, Trp α149, Trp α184, Tyr α190, Cys α192, Cys α193, Pro α194, Tyr α198, Trp γ55/δ57 and Asp γ174/δ180, positions implicated in agonist binding by previous work.[8] Constitutively active receptors were observed at only three of the sites evaluated - Trp α149, Tyr α93 and Trp γ55/δ57. (Fig. 3)

Successful suppression experiments in *Xenopus* oocytes yielded large standing currents (Fig. 2, identified as **a**) in the absence of ACh. The standing currents are reduced (**b**) in the presence of the open-channel blocker 8-(*N*, *N*-diethylamino)octyl-3,4,5-trimethoxybenzoate (TMB-8). This establishes that the observed current is specifically due to an open nAChR, rather than a non-specific basal current. In addition to TMB-8, Zhong *et al.* used channel blockers QX-314 (Lidocaine N-ethyl bromide) and NMDG (N-methyl-D-glucamine); the antagonist curare; desensitization of the standing current by longer application of ACh; and single channel measurements to establish that the standing current was indeed due to open nAChRs.[3] The single channel conductance for the tethered system is identical to that of the native receptor, demonstrating that the tethered agonist increases the open probability of the channel but probably does not affect the pore. It is also worth noting that standing currents are only observed when a L262S mutation is made in the channel (M2) region of the β subunit (designated β9'S). It is well established that this residue is far from the binding site and that the serine mutant facilitates channel opening.[9]

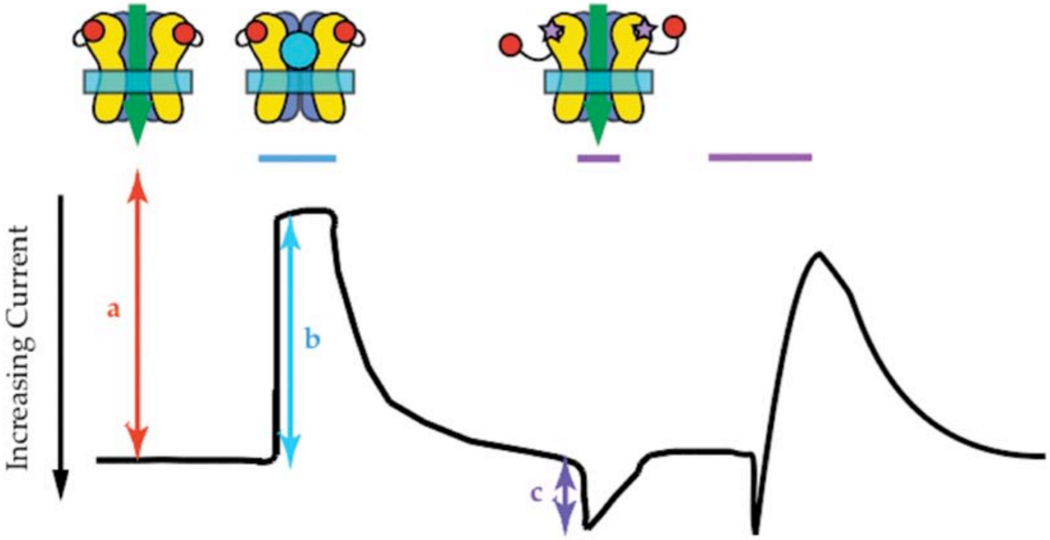


Figure 2. Efficacy. Representative recording traces of voltage-clamp currents for an individual oocyte expressing mutant nAChR with TyO3Q incorporated at α149. The horizontal bars indicate bath application of ACh or TMB-8. **a**: standing current due to constitutively active nAChR; **b**: standing current that is blocked by TMB-8; **c**: ACh-induced current.

After TMB-8 wash-off, addition of 5 μM ACh causes a small increase in current (**c**), indicating that the tethered agonist is a fairly weak agonist that never fully activates the receptor. The ratio of the constitutive current that is blocked by TMB-8 to the ACh-induced current, **b/c**, is seen as a measure of the efficacy of the tethered agonist. The ratio compares the extent to which the tethered agonist opens the receptor to the maximal response elicited by ACh. Using this ratio minimizes complications due to variations in the protein expression level. In short: the larger the **b/c** ratio, the more effective the tethered agonist. Of all the unnatural amino acids incorporated at the various sites, TyrO3Q at α149 was most effective. (Fig. 3) TyrO3*t*-Bu, an uncharged analog of TyrO3Q in which the quaternary nitrogen has been replaced by a carbon, also gave weakly constitutively active receptors at α93 and γ55/δ57 (but not at α149). This established the importance of delivering a positive charge at the 149 position for receptor activation.

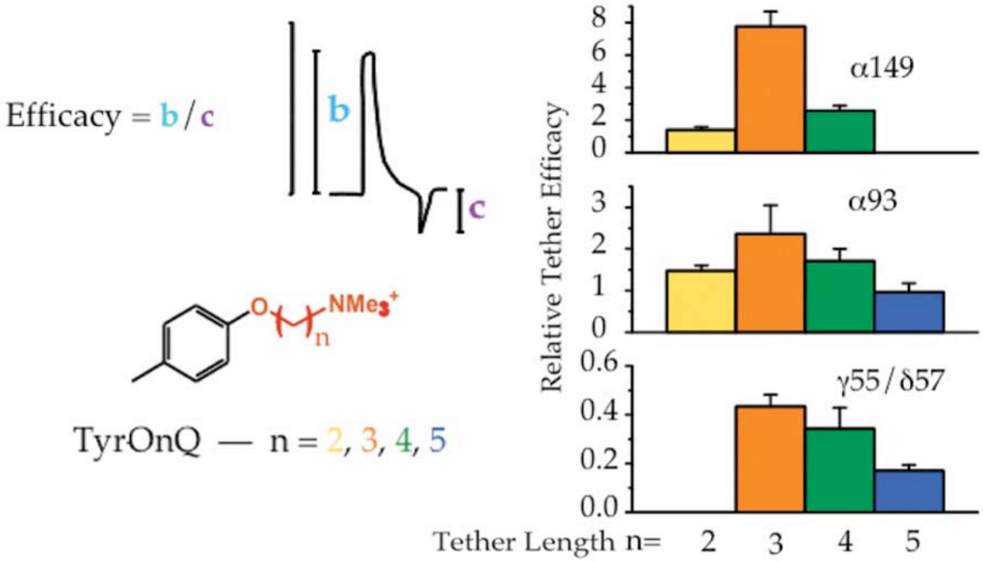


Figure 3. Probing the Binding Site with Tethered Agonists. TyrOnQ structure and tethered agonist relative efficiencies (**b/c** per Figure 2) for TyrOnQ as a function of n (the number of methylene groups in the side chain) at positions α149, α93, and γ55/δ57.

Both position and tether length significantly influence the effectiveness of the tethered agonist at α149. (Fig. 3) TyrO3Q represents the most effective chain length. Li *et al*. note that the constitutive currents are consistently larger when TyrOnQ is incorporated at α149 than at α93 and γ55/δ57. This is not simply because the receptor expresses more efficiently with the tether at α149.[8] As Figure 3 shows, the ratio of TMB-8-blocked current (as in Fig. 2, **b**) to ACh-induced current (Fig. 2, **c**) shows the same trends. Thus, the efficacy of the tethered agonist, not just the absolute size of the current, is larger at α149.

This library of experiments established that incorporating a propyl tethered agonist at α149 yielded the greatest degree of constitutive activity. It is posited that a three carbon chain is optimal because it can achieve the same positioning of the ammonium center as the cation-π interaction with the six-membered ring of Trp. It must be noted that it is not necessarily believed that the quaternary ammonium participates in an auto-cation-π interaction with its own aromatic ring. The efficacy of TyrO3Q at α149 led us to choose this tether length and position for secondary and tertiary tethered agonist explorations.

Secondary and Tertiary Tethered Agonists

Incorporation of a secondary or tertiary amine tethered agonist at α149 could yield a constitutively active receptor that can be modulated by pH. Despite our inability to verify a cation-π interaction in Nic binding, it is presumed that tertiary agonists such as nicotine activate the receptor in their protonated state *via* interactions analagous to those described for quaternary agonists. The fact that all known nAChr agonists contain a cationic (or potentially cationic) moiety seems to support this.[10] (The fact that the channel can be gated by TyrO3*t*-Bu can gate the channel does not necessarily indict this, delivering steric bulk with this unnatural amino acid is not the same as binding an agonist.) TyrO3T (Fig. 4) and its secondary and primary analogs (TyrO3S and TyrO3P) should gate the receptor only at pHs below their pK_a, so raising the pH would turn off the receptor. Thus, titration of the oocyte's media will allow us to probe the microenvironment of the receptor by establishing the pK_a of the tethered amine in the nAChR binding site. Suppression at α149 furthers our understanding of the binding of non-quaternary amine ligands and establishes a novel method for converting a ligand-gated ion channel into a pH sensor.

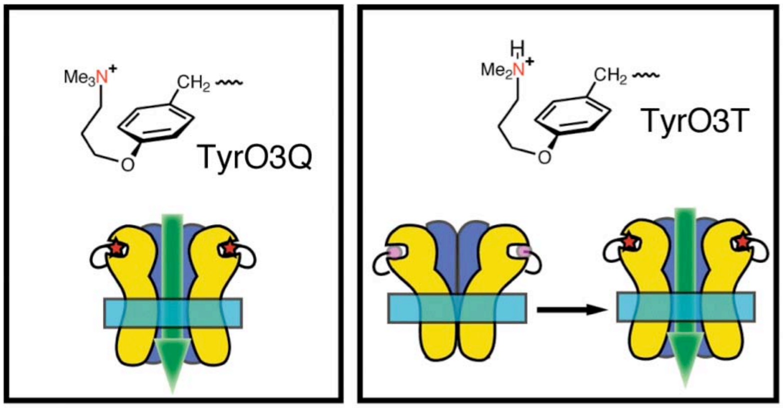


Figure 4. Tertiary Tethered Agonists. Left: Quaternary ammonium of ACh positioned over the six-membered ring of αTrp149, and induced channel opening. Right: Structure of TyrO3T, showing that the tethered tertiary amine could open the channel when protonated.

Tethered Agonist Syntheses

The protonatable tethered agonist study began with the syntheses of the novel amino acids TyrO3T, TyrO3S, and TyrO3P. The syntheses of TyrO3P and TyrO3S in appropriate forms for nonsense suppression were straightforward. Both the α-amine of the amino acid and the side chain amine were protected as nitroveratryoxycarbonyl (NVOC) groups. The tertiary amine of TyrO3T cannot be protected as an amide. Interestingly, the *N*-nitroveratryl side-chain protecting group, which has been used successfully in similar contexts[11], was not viable for TryO3T because of inefficient photo-deprotection of the quaternary group. Other studies with simpler model compounds confirm that nitrobenzyl-type photocleavage reactions are not efficient when converting a quaternary ammonium to a tertiary amine. (See Supporting Information) We have found, however, that the dimethoxycoumarin (DMCm) group is an effective photocleavable protecting group for tertiary amines, consistent with several previous observations.[12]

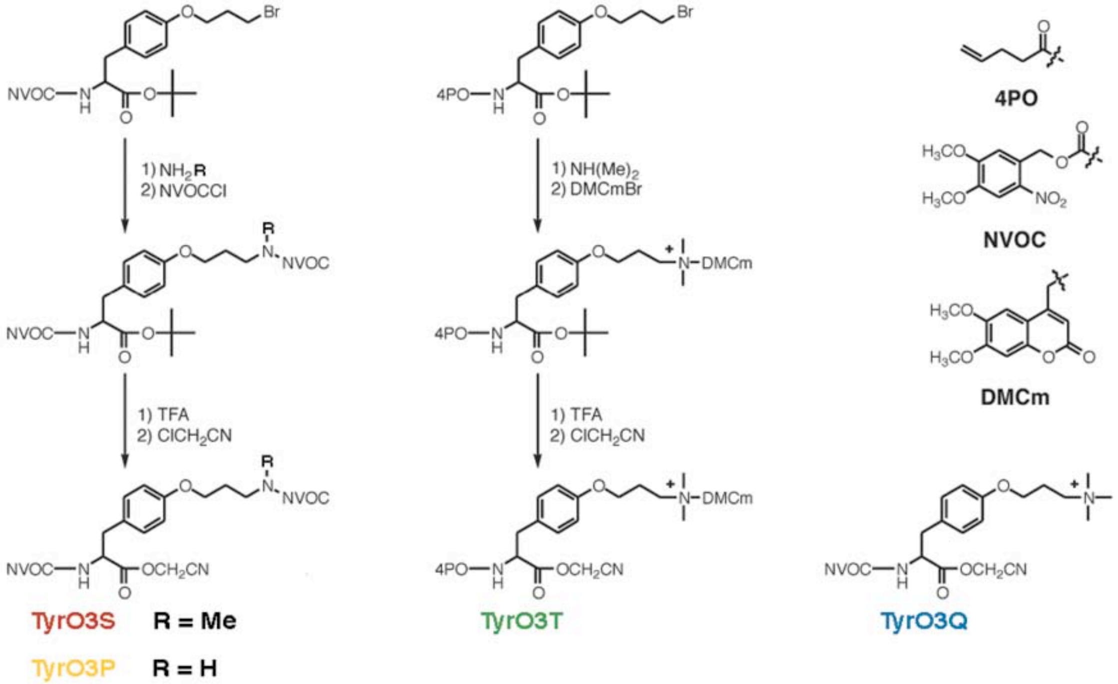


Figure 5. Syntheses of the tethered agonist amino acid monomers.

TyrO3T Constitutive Activity Can Be Reversibly Modulated by pH

The tethered amine unnatural amino acids were incorporated into nAChRs expressed in *Xenopus laevis* oocytes using now well-established protocols.[13, 14] Channel activity was monitored with standard two-electrode, voltage-clamp electrophysiology. Along with the constitutive (standing) current, responses to added ACh, known channel-blockers such TMB-8 and QX-314, or agonist-free solutions of various pH, were measured. It was found that the constitutive activity of the tethered agonists could in fact be modulated in a reversible fashion by changes in the pH of the media. An example of primary data for TyrO3T is shown in Figure 6. In agonist-free media of pH 7.5, a substantial standing current is observed. Lowering the pH to 6.5 increases receptor activity in the expected manner. Application of the channel blocker TMB-8 eliminates most of the current, showing that the standing current is specific to nAChR activation and not due to non-specific effects on the oocyte membrane. Importantly, both of these effects can be reversed by returning to agonist-free media of pH 7.5.

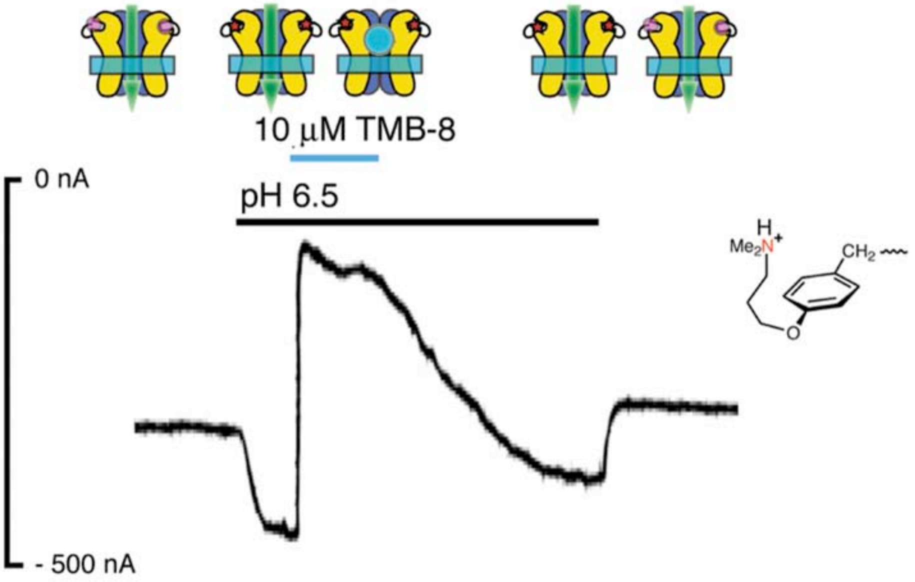


Figure 6. TyrO3T tethered agonism is reversibly modulated by pH. Where no bars are shown, pH 7.5 agonist-free media is applied. Changing the media pH to 6.5 potentiates receptor current, application of TMB-8 blocks all of the current from nAChRs. Both block and pH potentiation are reversible.

TyrO3T, TyrO3S, and TyrO3P Suppression and pH Modulation

Examples of electrophysiological recordings for TyrO3S, TyrO3T, and TyrO3Q suppressed at α149 are shown in Figures 7 and 8. For each tethered agonist, there is a substantial standing current even at pH 7.5, which can be blocked by TMB-8. As seen in previous tethered agonist studies, added ACh causes an increase in current (followed by desensitization), indicating that the tethers are "partial agonists." In fact, tethered agonist efficacy seems to map with steric bulk at the cationic amine center (See below). For TyrO3S and TyrO3T, application of low pH agonist-free solutions presumably potentiates constitutive activity by protonating the tether. For TyrO3Q, which is always charged, changes in constitutive activity with pH merely mirror the inherent changes in channel conductance as seen for wild type responses to 10 µM ACh at various pHs (Figure 8, bottom).

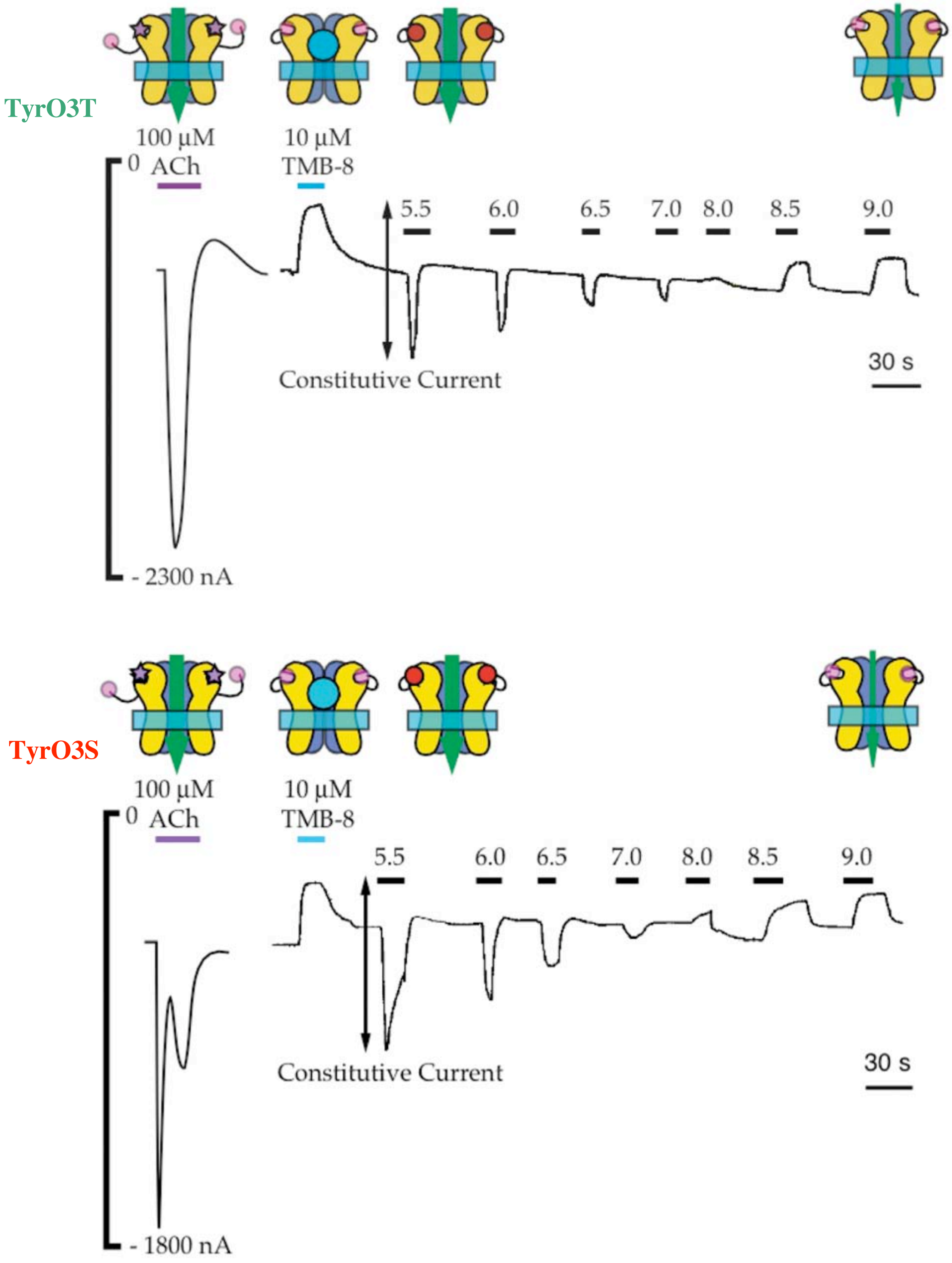


Figure 7. Examples of primary electrophysiological data: TyrO3T and TyrO3S at α149. Tethered agonist responses to ACh, TMB-8, and agonist-free solutions of varying pH. Upper bars indicate agonist or blocker application. Lower bars indicate application of solution of given pH. Arrows indicate blockable constitutive currents.

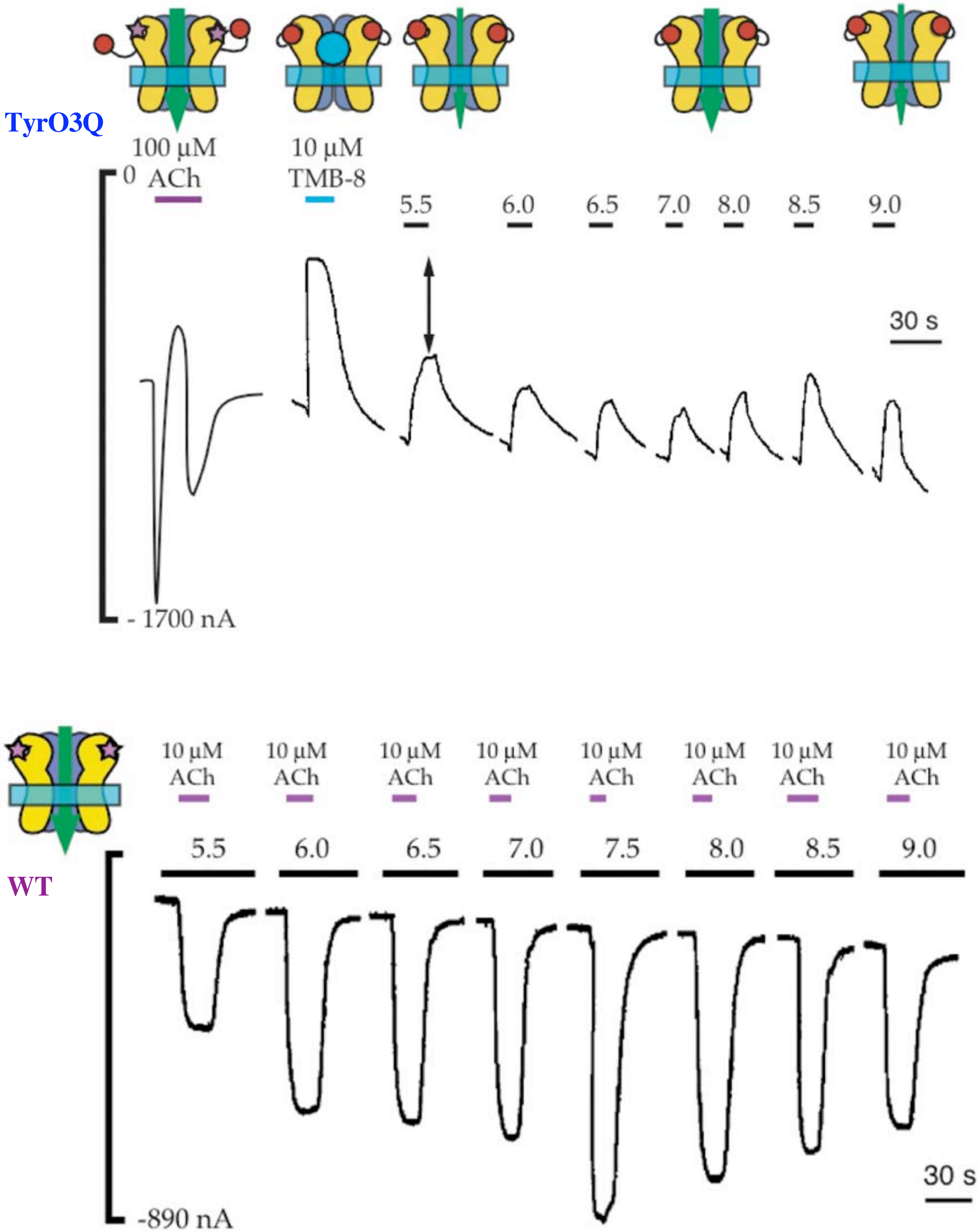


Figure 8. Examples of primary electrophysiological data: TyrO3Q at α149. Tethered agonist responses to ACh, TMB-8, and agonist-free solutions of varying pH. Wild Type nAChR responses to 10 µM ACh at differing pHs. Upper bars indicate agonist or blocker application. Lower bars indicate application of solution of given pH. Arrows indicate blockable constitutive currents.

The primary tether incorporates and gives functional nAChRs, as seen by the response to ACh, but shows no constitutive activity, even at pH 5.5. (Fig. 9) It would be surprising that the tether remains unprotonated, as the pK_a of aminopropanol is 10.2, higher than that of *N*,*N*-diethylaminopropanol (9.3).[15] TyrO3P may be protonated, but lack the steric bulk to activate the receptor.

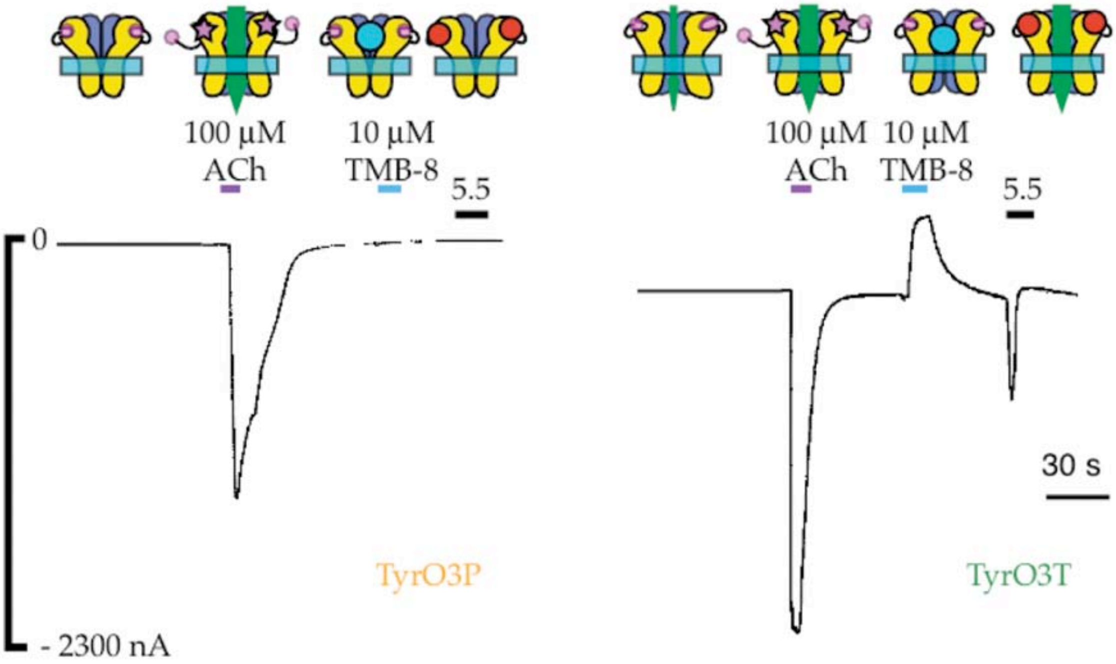


Figure 9. TyrO3P shows no constitutive activity. TyrO3P and TyrO3T at α149, their current responses to ACh, TMB-8, and pH 5.5.

TyrO3T and TyrO3S Efficacy – Background-corrected with TyrO3Q Data

Receptors that contain TyrO3T at position 149 of the α subunit (α149) showed a systematic increase in constitutive current as the pH was lowered. These currents can be blocked by TMB-8, showing that they arise from the opening of nAChRs. In contrast, the response of the wild type nAChR to saturating concentrations of ACh shows a small inherent dependence on pH, with maximal conductance at pH 7.5, falling at lower and higher pH. (Consistent with previous studies, See Supporting Information)[16, 17] In an important control, we find that the pH dependence of the constitutive current of the system with the tethered quaternary ammonium, TyrO3Q, mirrors that of wild type receptor. (Fig. 10) TyrO3Q differs from TyrO3T by only one methyl group, so it is stericly similar, but always cationic. The similarity of the TyrO3Q and WT pH effects establishes that the tether does not inherently introduce any pH modulation to receptor activity. These observations establish that the increase in blockable current observed for TyrO3T at lower pH is due to the specific protonation of the side-chain amine. TyrO3S constitutive currents also increase with decreasing pH. However, receptors containing TyrO3P showed no constitutive activity, even at pH 5.5. Note that attempts to incorporate Lys at α149 led to no surface nAChR expression; only unnatural amino acid incorporation enables these studies. (See Supporting Information) We use the WT and TyrO3Q data to correct the TyrO3S and TyrO3T results for effects on receptor activity that are not specific to the protonation of the tether and evaluate the background-corrected efficacy of the tethered agonists as a function of pH.

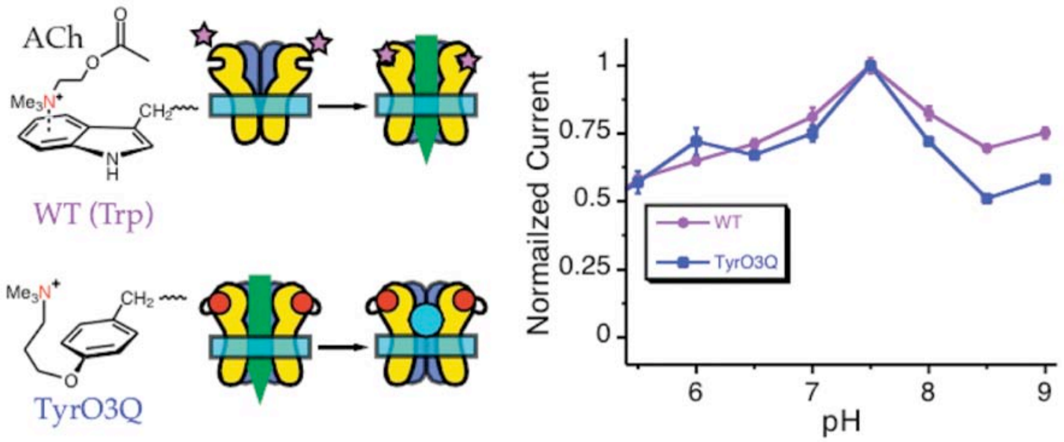


Figure 10. Comparison of pH effects on WT receptor activation by ACh and TyrO3Q constitutive activity (TMB-8-blockable current).

As in the TyrO3Q studies, we define the efficacy of a tethered agonist as the ratio of the constitutive current that can be blocked by TMB-8 to the maximum current induced by saturating concentrations of ACh (corrected for basal conductance changes with pH as determined from the data in Figure 10). By considering only constitutive current that can be blocked, we eliminate any background (non-nAChR specific) leak current. The efficacies of TyrO3Q, TyrO3T, and TyrO3S at α149 are shown in Figure 11.

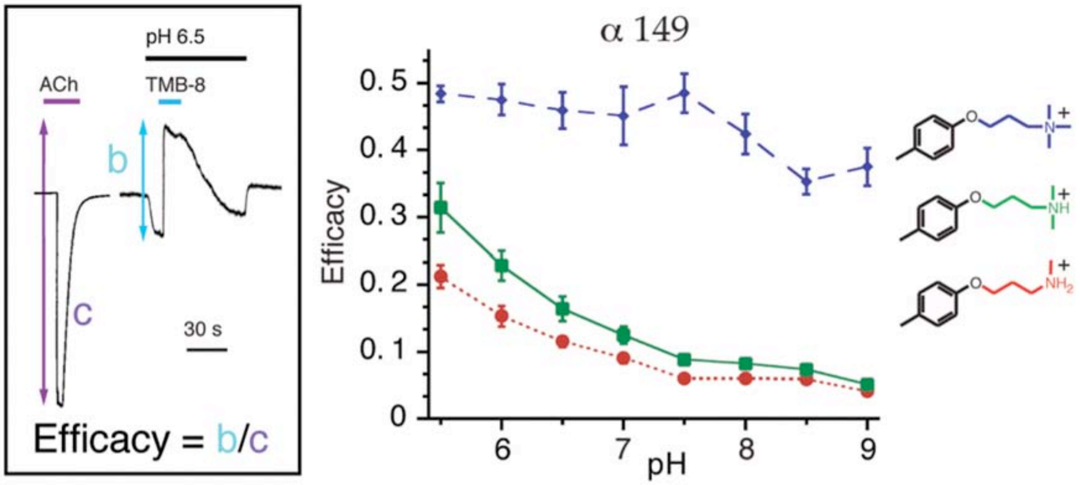


Figure 11. Tethered Agonists at α149: Tethered agonist efficacy as a function of solution pH. ACh efficacy = 1 at all pHs.

Plots of Tether Efficacy vs. pH Establish a Perturbed Binding Site pK_a

TyrO3Q's efficacy, as expected, is insensitive to pH changes, when corrected for basal pH dependence (by normalizing its efficacy to the changes in ACh efficacy with pH). TyrO3T and TyrO3S exhibit strong titration behavior when incorporated at α149. Previous studies with the uncharged tether TyrO3*t*-Bu showed that delivery of a cation to the α149 site was essential for channel activation.[6, 8] Thus, our adjusted measurments of receptor activation provide a straightforward way to assay the degree of protonation of a tethered amine at the binding site. The curve in Figure 11 can be interpreted to give a phenomenological pK_a for the side chain. Since pH ≤ 5 compromises the oocyte membrane integrity, we are unable to study the presumed plateau at low pH values. Nevertheless, it is clear that the side-chain pK_a of TyrO3T is ≤ 6 when incorporated at α149 of the nAChR, substantially shifted from its value in solution (~ 9.3).[15] This is discussed further below.

Tethered Agonists at α93 and γ55/δ57

An examination of the efficacies of TyrO3S, TyrO3T, and TyrO3Q at α93 and γ55/δ57 (Fig. 12) shows that all three tethers are less potent at these sites than at α149. TyrO3T and TyrO3Q curves are similar in shape to the α149 curves. Surprisingly, although TyrO3S yields constitutive activity at all three sites, it cannot be potentiated by pH at α93 and γ55/δ57.

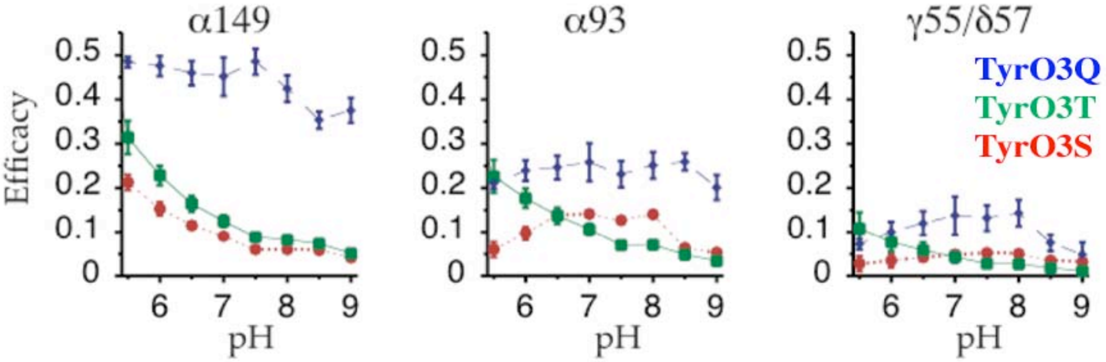


Figure 12. Efficacy as a function of solution pH for tethered agonists at α149, α93, and γ55/δ57.

TyrO3*t*-Bu also gave weakly constitutively active receptors at α93 and γ55/δ57 (but not at α149). TyrO3*t*-Bu showed no pH-dependent potentiation over background. This indicates that charge may be less of a factor at these positions than at α149, and may help to explain the puzzling TyrO3S behavior. In combination with the lack of constitutive activity for TyrO3P at α149 and the failure of the Lys mutagenesis experiments, this indicates that the geometric requirements of the binding site are quite precise. We will focus on the TyrO3T data at α149 in the analysis below, as it was the most efficacious tether.

Implications for the Character of the nAChR Binding Site

A comparison of the phenomenological pK_a for TyrO3T (≤ 6) and the free solution pK_a of *N*,*N*-dimethylaminopropanol (9.3) implies that a shift of at least 3 pK_a units is induced by the environment of the nAChR binding site. pK_a shifts of this magnitude are precedented. For example, Lys 115 in *Clostridium acetobutylicum* acetoacetate decarboxylase has an effective pK_a of 6.0, shifted by 4.5 units.[18] Such a large shift for the tethered agonist may seem surprising, given that the currents from norACh and Nic at pHs 6.5 and 7.5 were similar (Chapter 2 and below). The pK_as of these agonists are lower than for dimethylaminopropanol (Fig. 13), so one might expect that they would not be protonated in the binding site until the pH was even lower than 6.[2, 19] This is not inconsistent with the current data. The raw current data does not take into account the fact that the inherent conductance of the receptor decreases between pH 6.5 and 7.5 (Fig. 10). Thus, the efficacy of norACh and Nic actually does increase at lower pH (See below).

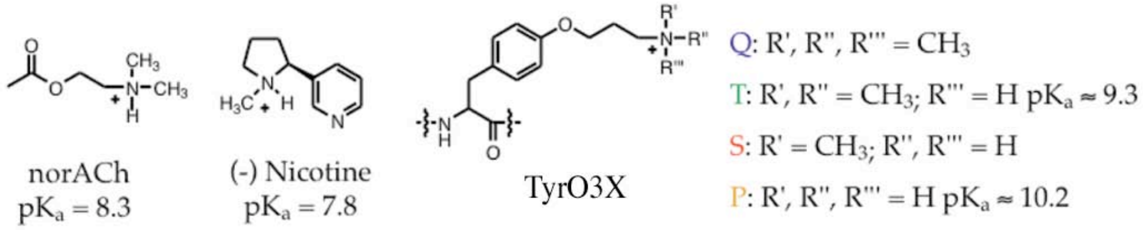
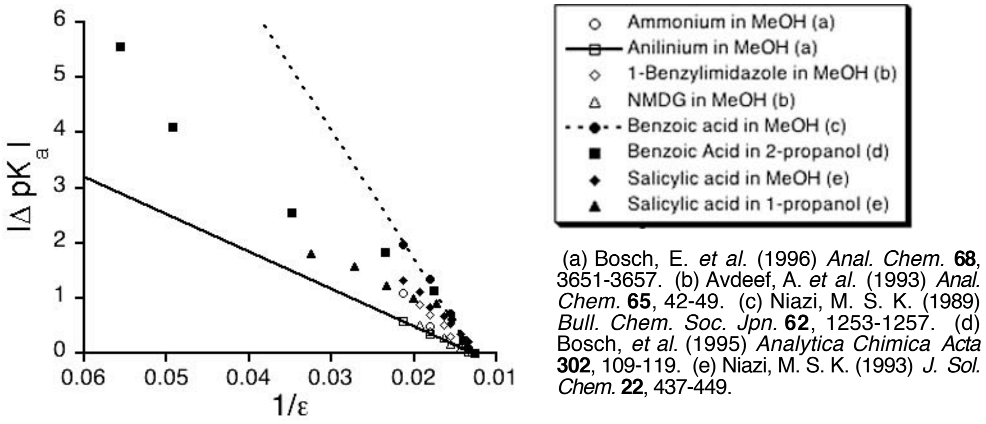


Figure 13. Protonatable nAChR agonists. Left: NorACh and Nic, with free solution pK_as listed. Right: Tethered agonists, with free solution pK_as of tether portion listed.

Our results suggest that the agonist binding site of the nAChR is relatively hydrophobic, consistent with the fact that the binding site is primarily formed by aromatic residues. A literature survey of the pK_a shifts of organic acids and bases in water/solvent mixtures of varying dielectric constants shows that a pK_a shift of 3 units is consistent with a dielectric constant (ε) of about 25. (Fig. 14)



(a) Bosch, E. *et al.* (1996) *Anal. Chem.* **68**, 3651-3657. (b) Avdeef, A. *et al.* (1993) *Anal. Chem.* **65**, 42-49. (c) Niazi, M. S. K. (1989) *Bull. Chem. Soc. Jpn.* **62**, 1253-1257. (d) Bosch, *et al.* (1995) *Analytica Chimica Acta* **302**, 109-119. (e) Niazi, M. S. K. (1993) *J. Sol. Chem.* **22**, 437-449.

Figure 14. pK_a shifts of acids and bases in water/solvent mixtures of varying dielectric constants.

The binding site environment was investigated computationally by calculating the solvent accessible surface areas of the residues in acetylcholine binding protein (AChBP, see Chapter 2) using NACCESS.[20] Trp 143 (Trp 149 in muscle nAChR numbering) was found to be 10.8% solvent accessible. (Fig. 15) The mid-range dielectric constant ($\varepsilon = 25$) implied by the shift in amine pK_a that we observe for TyrO3T seems reasonable for a moderately water accessible site such as Trp 149 ($\varepsilon(H_2O) = 78.5$, $\varepsilon(\text{hydrocarbon}) \sim 2$). It may be the lack of water accessibility that causes the shift in pK_a that we observe. We will use this information about the binding site environment in the computational modeling studies described in the following chapters.

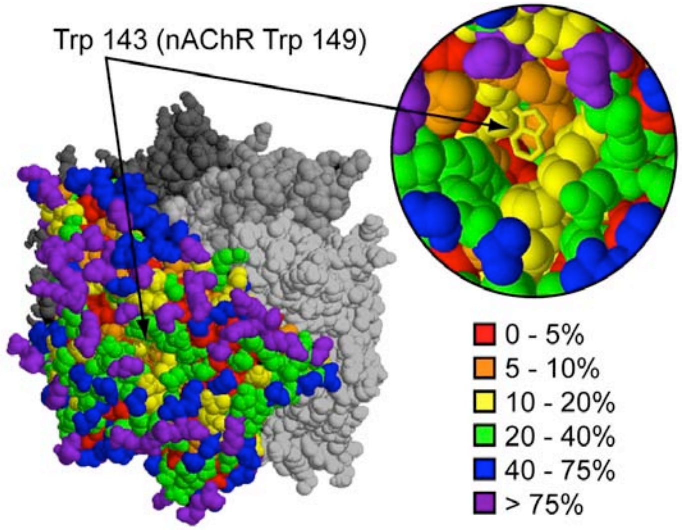


Figure 15. AChBP residue Solvent Accessible Surface Area (SASA). Two AChBP subunits shown colored according to SASA. Inset shows binding site with residues 73, 106, 112, 187, 188, 189, 190, and 192 removed for clarity. Trp 143 shown as stick representation. Residue solvent accessibilities were evaluated for a solvent of radius 1.40 Å. Accessibility is defined as the percentage of sidechain surface area accessible to the probe in the protein relative to the surface area accessible in an A-X-A tripeptide.

Implications for Agonist Binding

What, then, are the implications of our studies for the binding of tertiary agonists like Nic and norACh? We have measured the pH-dependence of Nic and norACh efficacy in the present context (Fig. 16). Efficacy is computed for non-tethered agonists by normalizing the current from applications of saturating agonist concentrations to the maximal ACh-induced current. Interestingly, there is no pK_a shift for Nic; the phenomenological pK_a is not measurably different from the solution pK_a ($pK_a = 7.8$) of the drug.[2] In contrast, norACh, ($pK_a = 8.3$) the closest possible protonatable analogue of ACh, shows a pK_a shift of ~ 1 unit, noticeable, although not as large as for the tethered amines.[19] Obviously the agonists are protonated while in solution; deprotonation occurs as the agonist binds, a result of the destabilizing binding site environment.

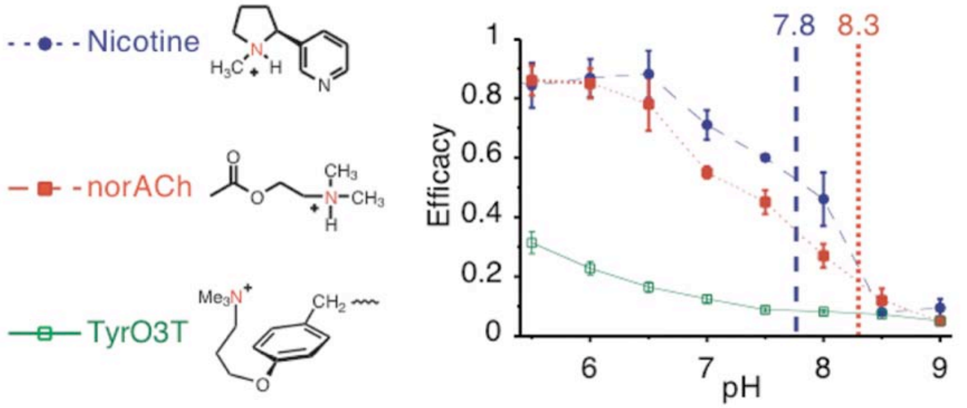


Figure 16. Tertiary Agonists: norACh and nicotine efficacy as a function of solution pH. ACh efficacy = 1 at all pHs.

These differences are interpreted as follows: We assume that only the cationic forms of agonists can activate the receptor. For the amine tethers and norACh, the protonatable amine can equilibrate with the medium when at the agonist binding site, and the local microenvironment of the protein produces a pK_a shift. This is lesser in extent for norACh than for the tethered agonists, a phenomenon for which we suggest an explanation below. For Nic, however, no pK_a shift is seen because the protonated amine does not equilibrate with the medium once it has bound to the receptor. The degree of receptor activation is dependent only on the amount of protonated Nic or norACh available to bind to the receptor. Therefore the pH dependence of activation mirrors nicotine's normal pK_a, and that of norACh is shifted.

Why would a receptor that binds cations have a binding site environment that is destabilizing toward cation formation from secondary and tertiary amines? The answer may have to do with selectivity. There are numerous biogenic amines present in the extracellular solution. Synaptic transmission would not work if the nAChR responded to every cationic, protonated amine. The implication of the ~ 3 pK_a unit shift observed for our tethered agonists is that their protonated states are destabilized by roughly 5.6 kcal/mol (R•300° K•ln(10^4)) relative to that state in bulk water. A binding site that is destabilizing to cation formation in general, but has functional groups that can stabilize the binding of "intended" agonists, would be very useful. These intended agonists, either endogenous ACh, or Nic, which the tobacco plant has evolved to bind to the nAChR, have structures that permit the interaction of their cationic moiety with the stabilizing functional groups. Thus, they do not exhibit as strongly perturbed a pK_a as the tethered agonists, which have not had the benefit of evolution in tuning their potency. Recall that these tethered agonists are relatively weak partial agonists: TyrO3T and TyrO3Q have maximal efficacies of 0.5 and 0.3 respectively. norACh and Nic have efficacies of 0.9 when fully protonated.

The same evolutionary argument may explain the differences between the two agonists. Nic does not exhibit a shift in pK_a in the binding site because it can interact with a specific functional group that stabilizes its protonated, tertiary amine. An excellent candidate for this functional group is the carbonyl of the Trp 149/ Thr 150 backbone amide, which we will study in the next chapter. NorACh is not as precisely aligned to attain this interaction, thus its effective pK_a in the binding site is shifted by roughly 0.7 pK_a units from its value in free solution. This is not as large as the pK_a shift for the tethered agonists, which cannot seem to find a functional group to stabilize their cation. This ability to stabilize the protonated amine in the binding site may be reflected in the 3.3-fold higher affinity for Nic than for norACh at the muscle nAChR.

In summary, we envision a model of the binding site that is moderately hydrophobic ($\varepsilon = \sim 25$) in its global character but has specific functional groups that can stabilize certain cations in the binding site. It appears that Nic is better able to attain the proper alignment for cation stabilization than norACh. This is the second recent line of evidence from our labs that indicates a difference between the binding modes of cholinergic and nicotinic agonists. ACh demonstrated a clear relationship between cation-π binding ability with αTrp 149 and agonist potency. In order to interact with Trp 149 *via* a cation-π interaction, the Nic ammonium must be stabilized by roughly 5 kcal/mol so that it remains protonated in the nAChR binding site. ACh, of course, must also find stabilization for its cationic moiety in the binding site, but the more diffuse quaternary ammonium may be more easily "solubilized" in the binding site. We attribute the change in slope in the Nic fluorination plot to a change in its binding mechanism when cation-π binding is weakened beyond a certain point. (See Chapter 2) If Nic must be protonated to open the channel, then it may need to adjust its binding mode to a more optimal interaction with other functional groups in the binding site in order stabilize its protonated state. ACh, with its more spherically symmetric cation moiety, does not seem to need to reposition itself (or this repositioning does not affect its cation-π binding ability). Our emerging model of ligand binding to the nAChR is therefore one in which subtle differences in ACh and Nic structure dictate differences in their binding modes to the receptor. Remember that the receptor is inherently dynamic because it must gate. Ligand binding is not a "lock and key" process, but one in which both ligand and receptor adjust in a cooperative manner to reach an energetic minimum which we call "binding." The beauty of the process is that although Nic and ACh fold the same residues about themselves differently, the folded structures are sufficiently similar that both agonists can gate the channel.

Supporting Information

NV-Protected TyrO3T

The initial attempt to synthesize a tertiary analog of TyrO3Q following previously established procedures met with several difficulties. Addition of dimethylamine to the cyanomethyl ester NVOC-TyrO3Br-OCH_2CN (See Materials and Methods) led to formation of the amide rather than S_N2 substitution for the terminal bromide. Cyanomethylation of the *N*-protected amino acid (produced by deprotection of NVOC-TyrO3T-O*t*-Bu) with chloroacetonitrile yielded alkylation of the sidechain amine as well as esterification of the carboxy terminus. Protection of the side chain with a photocleavable nitroveratryl group prior to formation of the cyanomethyl ester seemed to be a way around this problem. It was assumed that this NV group would be cleaved along with the NVOC group when the aminoacyl $tRNA_{CUA}$ was irradiated with 350 nm light.

Unnatural amino acid suppression with NVOC-TyrO3T(NV) $tRNA_{CUA}$ and α149TAG β9' mRNA gave mutant proteins which were expressed in oocytes. NVOC-TyrO3T(NV) gave ACh currents of up to 563 nA, but no discernable TMB-8 effects. Changes in pH also failed to change the standing current produced by NVOC-TyrO3T(NV). The EC_{50} obtained for NVOC-TyrO3T(NV)/α149β9' is 4.5 μM, not substantially greater than the wild type EC_{50} of 1.2 μM (αWTb9'). The Hill coefficient (n), which gives an indication of the cooperativity of agonist binding, is expected to be roughly 1.5 (n = 2 ideally), indicating that the binding of two agonist molecules is required for the opening of the channel. As one can see from the plot below, n = 0.7, perhaps indicating that the binding site has been altered.

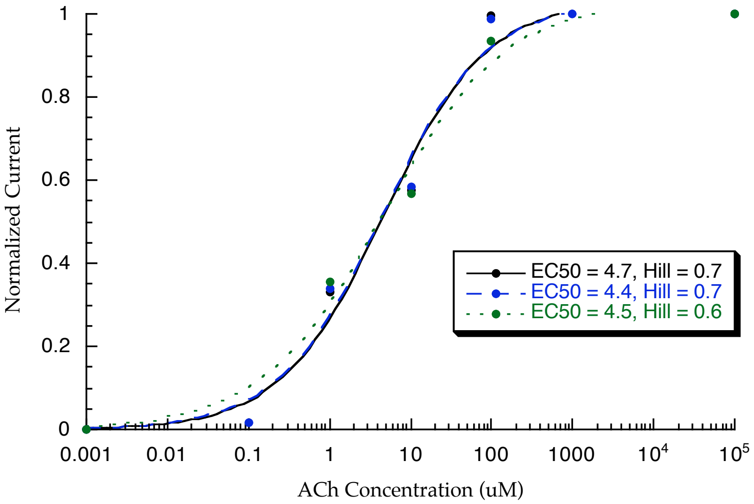


Figure 17. Dose-response relations and fits to the Hill equation for NVOC-TyrO3T(NV) suppression at α149β9'. Data from three ooctyes shown.

It is possible that the amino acid was being incorporated with its side-chain still protected, a group which is bulky enough to prevent the quaternary amine from positioning itself to properly activate the receptor, but still permits ACh access to the binding site. While it seemed that proper folding of the binding pocket was unlikely in the presence of such a bulky residue, the following experiments lead credence to this hypothesis.

The first question investigated was whether the side-chain NV group was being removed by irradiation. Generally, the aminoacyl $tRNA_{CUA}$ is irradiated for 5 minutes with a 400 W Hg arc lamp with a filter set permitting a relatively narrow band around 350 nm.[13] It has been shown previously that the quantum yield for photolysis of an NV protecting group on a quaternary ammonium is much lower than less substituted amines. While Peng *et al.* claim microsecond deprotection times elsewhere, their enzymatic assay for released nor-butyrylcholine clearly shows that full deprotection is only acheived after an hour of irradiation.[11]

The protection status of the amino acid was determined by an HPLC MS assay of irradiated samples of the dCA-coupled product in the same 1mM NaOAc solution in which the aminoacyl $tRNA_{CUA}$ is photolyzed. Monitoring the UV spectrum at 260 nm, the absorption maximum of the dCA nucleotides, and 350 nm, for NV groups, allowed us to observe the starting materials and products separated with the HPLC gradient. In addition, the tandem electrospray mass spectrometery system allowed for firm identification of all these compounds. NVOC-Ala-OdCA was photolyzed as a proof-of-principle experiment: after 5 minutes of irradiation, roughly 70% deprotection of the amino acid was observed, with products easily characterized by MS. Irradiation of NVOC-TyrO3T(NV)-OdCA cleaved the α-amino NVOC group fairly easily (producing TyrO3T(NV)-OdCA), but only very small amounts of fully deprotected TyrO3T-OdCA were produced, even after 60 minutes of irradiation. There were also many side-products which could not be readily identified from their MS fragmentation patterns as photolytic or hydrolytic products of NVOC-TyrO3T(NV)-OdCA. Maintaining the sample at 0° C while irradiating decreased degradation of the material, but did not improve the yield of fully deprotected product. Considering that the injected samples were only photolyzed for five to ten minutes, it seems likely that the residue incorporated was Tyr-O3T(NV), deprotected only at the α-amino end, and thus using this residue does not seem feasible.

We assumed that orthogonally protecting the amino terminus with a 4-pentenoyl (4-PO) group would make it possible to isolate a side-chain deprotected TyrO3T. After photolysis, the NV-less 4-PO-protected tertiary amine would be HPLC purified and ligated to $tRNA_{CUA}$ for injection. It was also thought that removal of the NV group could be easier without a 350 nm absorbtive α-amino NVOC moiety. 4-PO-TyrO3T(NV) synthesis and dCA coupling were facile as they were analagous to the previous synthesis. However, photolysis of the dCA-amino acid was extremely poor, almost no discernable product was present after 60 minutes. Apparently the difficulty observed in photocleaving at the quaternary ammonium is independent of the presence of other nitroveratryl groups.

When the initial attempt to deprotect 4-PO-TyrO3T(NV)-OdCA failed, a model compound, *N*-nitroveratryl-dimethylamino-1-propanol, was synthesized to test photolysis conditions. However, no deprotection was observed for this compound, despite varying temperature, concentration, and solution composition. Clearly photolysis of a NV group from a quaternary ammonium was not practical. This certainly calls into question the minimal literature precedent available.[11] In contrast, the photolysis of DMCm-protected TyrO3T was complete after 5 min. as assayed by HPLC-MS.

α149 Lys mutant

No ACh-induced or TMB-8-blockable currents were observed, and fluorescent labelling studies with tetramethylrhodamine-conjugated bungarotoxin (BuTx-TMR, an antagonist) showed no surface expression of α149 Lys mutant nAChRs in oocytes. This is consistent with work by Cohen and coworkers in which they mutated residues to Cys and created tethered agonists by modifying them with MTS-reagent based tethers.[6] They were not able to observe activity with Cys149 mutants, indicating that the nAChR may not fold properly without an aromatic residue at this position. Our tethered agonist, unlike our Lys or their Cys mutants, has an aromatic moiety. Six days following mRNA injection, oocytes were incubated (60 min, 4° C) in ND96 solution containing BuTx-TMR (100 nM) and bovine serum albumin (5 mg/ml). After three washes with ND96 the fluorescence intensity of the animal pole was determined using an inverted epifluorescent microscope (IX-70 FLA; Olympus Corp.) equipped with a photomultiplier tube (R928P; Hamamatsu Photonics) attached to the side port. This microscope (described previously) was fitted with an oil-immersion objective of 40X, NA 1.35.[8]

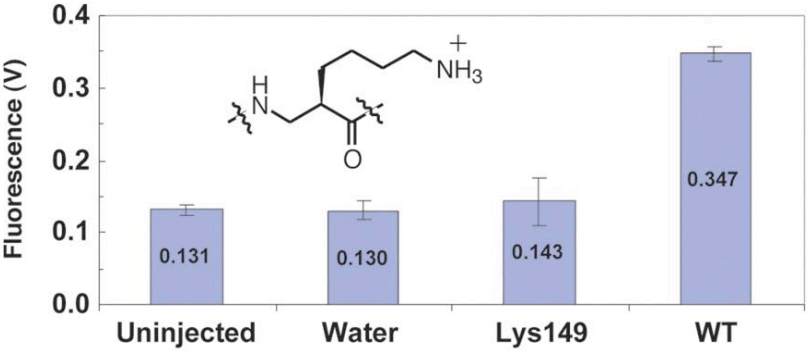


Figure 18. Fluorescent labelling of α149 Lys mutant nAChRs.

Previous Studies of pH Effects on nAChR Currents

Palma *et al*. examined 1 mM ACh currents over a range of pH's from 5.0 to 9.0 using the same buffers used in our experiments (MES: 5.5-6.5, HEPES: 6.5-8.5, CHES: 9.0).[17] They found that there is an order of magnitude change in the size of the ACh currents between 5.5 and 9.0 and that at pH 5.0 there was an almost complete abolition of current. In single-channel experiments, they found that pH affects both conductance and open times (Fig. 19 a and b). Pappone *et al*., working in BC3H-1 cells found similar variations in mouse muscle single channel conductances and open times.[16] While there has been much speculation about acidic residues in the pore region, no study to date has systematically identified the residues responsible for these pH-related changes.

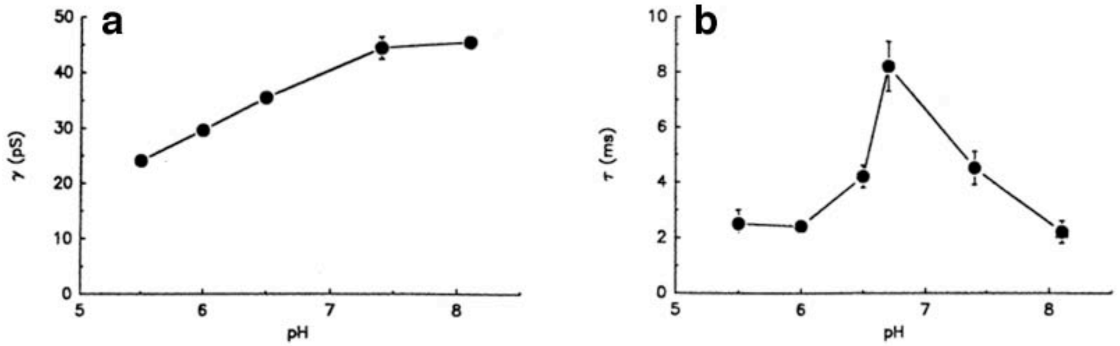


Figure 19. Previous Studies of pH Effects on nAChR Currents. **a**: Single-channel conductance as a function of pH. **b**: Channel open times as a function of pH. Reproduced from Palma *et al*., 1991.

Materials and Methods

General

Reagents were purchased from Aldrich, Sigma, or other commercial sources. TMB-8 was purchased from RBI (Natick, MA). ACh chloride and QX-314 were purchased from Sigma. Anhydrous THF and methylene chloride were obtained from J. T. Baker solvent kegs; anhydrous DMF (Puris) was obtained from Fluka. Flash chromatography was on 230-400 mesh silica gel with the solvent indicated. All NMR shifts are reported as δ ppm downfield from TMS. ^{1}H NMR and ^{13}C NMR spectra were recorded at 300 MHz in $CDCl_3$ or CD_3CN using a Varian QE-300 spectrometer. Electrospray (ESI) ionization and matrix-assisted laser-desorbed ionization (MALDI) quadrupole mass spectrometry were performed at the Caltech Protein/Peptide Micro Analytical Laboratory or at the Caltech Division of Chemistry and Chemical Engineering Mass Spectrometry Facility. Nitroveratryloxycarbonyl chloride (NVOCCl), NVOC-Tyrosine *t*-butyl ester (NVOC-Tyr-O*t*Bu, **6**), and NVOC-Tyrosine-*O*-propylbromide *t*-butyl ester (NVOC-TyrO3Br-O*t*Bu, **7**) were prepared as previously described.

N-NVOC-Bromopropylamine

0.15 g 3-bromopropylamine were dissolved in 15 mL *p*-dioxane and 15 mL water. 0.11 g Na_2CO_3 were dissolved in 15 mL water. 3-bromopropylamine solution was combined with the Na_2CO_3 solution. 0.19 g NVOC-Cl were completely dissolved in 25 mL dioxane, and were added to the other mixture and stirred for 1 h. The reaction mixture was extracted with 3 X 25 mL CH_2Cl_2, the organic layers were combined, dried with sodium sulfate, and run on a column in methylene chloride to give 0.25 g (37%) of product. ^{1}H NMR ($CDCl_3$): δ 7.66 (s, 1H), 6.97 (s, 1H), 5.46 (s, 2H), 5.16 (br s, 1H), 3.95 (s, 3H), 3.92 (s, 3H), 3.43 (t, J=6.3 Hz, 2H), 3.36 (td, J=6.3 Hz, 6.3 Hz, 2H), 2.07 (m, 2H); ^{13}C NMR ($CDCl_3$): 155.7, 153.2, 147.8, 127.7, 110.1, 107.9, 63.5, 56.3, 56.3, 39.4, 32.3, 30.6.

NVOC-TyrO3P(NVOC)-O*t*Bu

To a mixture of 0.12 g NVOC-Tyr-O*t*Bu and 0.166 g Cs_2CO_3 (two equivalents) dissolved in 10 mL DMF were added 0.096 g *N*-NVOC-bromopropylamine in 13 mL of anhydrous DMF, under Ar. After the reaction was stirred for 1.5 hours, the reaction mixture was extracted using 3 X 25 mL methylene chloride, the organic layers were combined, dried with sodium sulfate, and run on a flash column in 1:1 petroleum ether / ethyl acetate to give 75 mg (38%) of product. ^{1}H NMR ($CDCl_3$): δ 7.70 (s, 1H), 7.49 (s, 1H), 7.08 (d, J=4.2 Hz, 2H), 7.00 (s, 1H), 6.97 (s, 1H), 6.81 (d, J=4.2 Hz, 2H), 5.52 and 5.47 (AB, J=11.4 Hz, 11.4 Hz, 2H), 5.50 (s, 2H), 5.41 (d, 1H), 5.26 (br s, 1H), 4.52 (m, 1H), 4.02 (t, J=5.7 Hz, 2H), 4.98 (s, 3H), 4.98 (s, 3H), 4.98 (s, 3H), 4.98 (s, 3H), 3.44 (dt, J=6.3 Hz, 6.3 Hz, 2H), 3.05 (m, 2H), 2.02 (m, 2H), 1.44 (s, 9H); ^{13}C NMR ($CDCl_3$): 170.4, 157.4, 155.7, 155.0, 153.4, 153.2, 147.7, 147.5, 141.0, 139.2, 130.2, 128.1, 128.0, 114.2, 114.0, 109.9, 109.5, 107.9, 107.9, 82.3, 65.6, 63.6, 63.4, 56.4, 56.3, 56.1, 55.2, 38.7, 37.2, 29.3, 27.9.

NVOC-TyrO3P(NVOC)-OCH$_2$CN

0.075 g NVOC-Tyr-O3P(NVOC)-O*t*Bu were dissolved in 5 mL methylene chloride. 3 mL TFA were added to the mixture, using a glass pipette. The reaction was stirred for 1 hour. Volatiles were removed on the vacuum pump with a dry ice/acetone trap. 10 mL anhydrous DMF, 5 mL chloroacteonitrile, and 1 mL diisopropylethylamine (DIPEA) were added to the flask, under argon. After the reaction was stirred for 3 hours, the volatiles in the mixture were removed on the vacuum pump with a dry ice/acetone trap. The reaction mixture was run on a silica column in 1:1 petroleum ether / ethyl acetate to give 53 mg (72%) product. ^{1}H NMR (CDCl$_3$): 7.70 (s, 1H), 7.70 (s, 1H), 7.07 (d, J=4.2 Hz, 2H), 7.01 (s, 1H), 6.94 (s, 1H), 6.85 (d, J=4.2 Hz, 2H), 5.53 and 5.50 (AB, J=14.7 Hz, 14.8 Hz, 2H), 5.50 and 5.47 (AB, J=15.1 Hz, 14.7 Hz, 2H), 5.41 (s, 1H), 5.28 (d, J=6.3 Hz, 1H), 4.78 and 4.70 (AB, J=21.3 Hz, 21.3 Hz, 2H), 4.69 (m, 1H), 4.03 (t, J=5.7 Hz, 2H), 3.95 (s, 3H), 3.95 (s, 3H), 3.95 (s, 3H), 3.95 (s, 3H), 3.44 (dt, J=6.3 Hz, 6.3 Hz, 2H), 3.10 (m, 2H, J=6.3 Hz, 6.3 Hz), 2.02 (m, J=6.3 Hz, 2H); ^{13}C NMR (CDCl$_3$): 170.2, 157.8, 155.7, 155.0, 153.3, 153.2, 147.8, 147.6, 141.0, 139.4, 130.0 127.9, 127.4, 126.7, 114.6, 113.5, 110.0, 109.8, 107.9, 65.6, 64.0, 63.4, 56.4, 56.2, 56.3, 56.1, 54.8, 48.9, 38.6, 36.9, 29.7, 29.2.

NVOC-TyrO3S-O*t*Bu

268 mg NVOC-TyrO3Br-O*t*Bu were dissolved in 20 mL anhydrous THF under Ar in a 3-neck round-bottom flask with a CO$_2$(s) / acetone condenser. This was cooled to - 20 °C in a 30% KCl / ice bath and then NH$_2$Me gas was bubbled through the solution until the drip rate from the condenser tip was about 2 s^{-1}. The setup was allowed to warm to RT and then the reaction mixture was rotoevaporated to ensure removal of the dissolved NH$_2$Me. Chromatographic purification was achieved by elution of the starting material in CH$_2$Cl$_2$ followed by collection of the product in CH$_2$Cl$_2$ with 5% triethylamine. 164 mg of flaky yellow solid (67%) were obtained in this manner. ^{1}H NMR (CDCl$_3$) δ 7.67 (s, 1H), 7.09 (d, J = 8.5 Hz, 2H), 6.96 (s, 1H), 6.81 (d, J = 8.5 Hz, 2H), 5.81 (d, J = 8.1 Hz, 1H), 5.52 and 5.44 (AB, J = 15.2 Hz, 2H), 4.49 (m, 1H), 3.99 (t, J = 6.3 Hz, 2H), 3.95 (s, 3H), 3.93 (s, 3H), 3.03 (m, 2H), 2.75 (t, J = 6.9 Hz, 2H), 2.44 (s, 3H), 1.95 (m, 2H), 1.43 (s, 9H)

NVOC-TyrO3S(NVOC)-O*t*Bu

65 mg Na$_2$CO$_3$ (2 equiv.) were dissolved in 15 mL water and added to a solution of 164 mg NVOC-TyrO3S-O*t*Bu in 15 mL *p*-dioxane. 168 mg NVOC-Cl (2 equiv.) in 15 mL *p*-dioxane were added to this and the reaction was stirred overnight. The product was purified by flash chromatography in 1:1 EtOAC / petroleum ether, giving 104 mg (44%) of

a dark yellow solid. ^{1}H NMR ($CDCl_3$) δ 7.72 (s, 1H), 7.68 (s, 1H), 7.07 (d, J = 8.4 Hz, 2H), 7.03 (s, 1H), 7.01 (s, 1H), 6.79 (d, J = 8.4 Hz, 2H), 5.56 and 5.51 (AB, J = 15.0 Hz, 2H), 5.47 (br s, 2H), 5.35 (d, J = 7.5 Hz, 1H), 4.51 (m, 1H), 3.96 (br t, J = 6.3 Hz, 2H), 3.95 (s, 3H), 3.95 (s, 3H), 3.95 (s, 3H), 3.95 (s, 3H), 3.52 (t, J = 12.5 Hz, 2H), 3.05 (m, 2H), 3.00 (d, J = 17.5 Hz, 2H), 2.05 (m, 2H), 1.44 (s, 9H)

NVOC-TyrO3S(NVOC)-OCH$_2$CN

After 64 mg NVOC-TyrO3S(NVOC)-O*t*Bu were stirred with 1.00 mL trifluoroacetic acid for 1 hr. in 10 mL CH_2Cl_2, the reaction mixture was pumped on for 2 h. with a dry ice / acetone trap. The crude amino acid was coevaporated with toluene and redissolved in 10 mL DMF under Ar for esterification. 0.50 mL chloroacetonitrile was added with 0.10 mL DIPEA and the mixture was stirred at RT. The next morning, the volatiles were removed under vacuum and the mixture was run on a column in 3:1 petroleum ether / EtOAc. 24 mg of an orange solid were obtained in a 40% yield. ^{1}H NMR ($CDCl_3$) δ 7.70 (s, 1H), 7.67 (s, 1H), 7.07 (d, J = 8.4 Hz, 2H), 7.03 (s, 1H), 6.99 (s, 1H), 6.81 (d, J = 8.4 Hz, 2H), 5.55 and 5.49 (AB, J = 15.1 Hz, 2H), 5.47 (br d, J = 15.1 Hz, 2H), 5.33 (d, J = 7.5 Hz, 1H), 4.98 (br s, 2H), 4.38 (m, 1H), 3.97 (br t, J = 6.3 Hz, 2H), 3.94 (s, 3H), 3.94 (s, 3H), 3.94 (s, 3H), 3.94 (s, 3H), 3.56 (t, J = 10.5 Hz, 2H), 3.04 (m, 2H), 2.98 (d, J = 16.0 Hz, 2H), 2.08 (m, 2H), 1.46 (s, 9H)

4PO-Tyr-O*t*Bu

148 mg Na_2CO_3 (1.4 equiv.) were stirred with 237 mg Tyr-O*t*Bu in 25 mL H_2O and 20 mL *p*-dioxane. 256 μL 4-pentenoic anhydride (4P) anhydride, 2.1 equiv., Aldrich), dissolved in 5 mL *p*-dioxane, were added. The reaction was quenched after 1 hr. with 25 mL each CH_2Cl_2 and 1 M $NaHSO_4$. The aqueous layer was extracted with 3 X 25 mL CH_2Cl_2 and purified on a column in 3:1 petroleum ether / EtOAc. 249 mg of an oily yellow solid obtained in a 73% yield. ^{1}H NMR ($CDCl_3$) δ 6.90 (d, J = 8.4 Hz, 2H), 6.68 (d, J = 8.5 Hz, 2H), 6.21 (d, J = 6.6 Hz, 1H), 5.66 (m, 1H), 4.94 (d, J = 18.0 Hz, 1H), 4.89 (d, J = 11.2 Hz, 1H), 4.66 (m, 1H), 2.90 (m, 2H), 2.18 (m, 2H), 2.21 (m, 2H), 1.37 (s, 9H) ^{13}C NMR ($CDCl_3$): 171.6, 171.2, 155.6, 139.3, 130.5, 128.5, 115.5, 114.5, 82.6, 53.4, 37.5, 35.7, 32.6, 28.2.

4PO-TyrO3Br-O*t*Bu

750 mg Cs_2CO_3 (2 equiv.) were combined with 2.00 mL 1,3-dibromopropane in 20 mL dry DMF in a flame-dried flask under Ar. 249 mg 4PO-Tyr-O*t*Bu were dissolved in 10 mL DMF under Ar, added to the Cs_2CO_3 slurry through a septum, and stirred overnight. The reaction mixture was stirred 10 min. with 30 mL water and extracted with 3 X 25 mL CH_2Cl_2. Flash chromatography of the combined organics with 3:1 petroleum ether / EtOAc gave 157 mg of sticky yellow solid, a 47% yield. ^{1}H NMR ($CDCl_3$) δ 7.00 (d, J = 8.9 Hz,

2H), 6.75 (d, J = 8.7 Hz, 2H), 6.24 (d, J = 6.8 Hz, 1H), 5.68 (m, 1H), 4.96 (d, J = 17.9 Hz, 1H), 4.91 (d, J = 11.3 Hz, 1H), 4.67 (m, 1H), 3.99 (t, J = 5.8 Hz, 2H), 3.51 (t, J = 6.6 Hz, 2H), 2.94 (m, 2H), 2.27 (m, 2H), 2.23 (m, 2H), 2.20 (m, 2H), 1.36 (s, 9H) ^{13}C NMR ($CDCl_3$): 171.8, 171.1, 157.8, 137.1, 130.7, 128.7, 115.7, 114.5, 82.4, 65.5, 53.8, 37.5, 35.9, 32.7, 30.4, 29.7, 28.3.

4PO-TyrO3T-O*t*Bu

236 mg 4PO-TyrO3Br-O*t*Bu were dissolved in 25 mL dry THF in a 3-neck round-bottom flask with a dry ice / acetone condenser under Ar. The setup was cooled in ice and $NH(Me)_2$ gas was bubbled through the yellow solution until it was dripping vigorously from the condenser tip. The cooling apparatus was maintained for 2 hrs. and then the reaction mixture was allowed to warm to RT. After rotoevaporation to remove the volatile components, the mixture was chromatographed. Starting material was eluted with EtOAc and then the product was collected with 5% triethylamine in CH_2Cl_2, giving 210 mg (96%) of a sticky yellow solid. ^{1}H NMR ($CDCl_3$) δ 7.01 (d, J = 8.0 Hz, 2H), 6.79 (d, J = 8.0 Hz, 2H), 5.95 (d, J = 7.2 Hz, 1H), 5.76 (m, 1H), 5.02 (d, J = 17.7 Hz, 1H), 4.97 (d, J = 12.3 Hz, 1H), 4.71 (m, 1H), 3.96 (t, J = 6.3 Hz, 2H), 3.00 (m, 2H), 2.44 (t, J = 7.5 Hz, 2H), 2.35 (t, J = 9.0 Hz, 2H), 2.20 (s, 6H), 1.93 (m, 2H), 1.35 (s, 9H), 1.27 (m, 2H) ^{13}C NMR ($CDCl_3$): 171.3, 170.6, 157.7, 136.7, 130.2, 127.8, 115.3, 114,1, 82.0, 66.0, 56.3, 53.4, 45.5, 37.0, 35.6, 29.3, 27.9, 27.5.

4PO-TyrO3T(DMCm)-O*t*Bu

762 mg 4PO-TyrO3T-O*t*Bu and 675 mg dimethoxycoumarin methylbromide (DMCmBr) were stirred overnight in 200 mL CH_3CN at 60 °C under Ar. After rotoevaporation, the reaction mixture was purified by flash chromatography on silica gel. First, starting material was eluted using 1:1 petroleum ether / EtOAc with 5% MeOH, then product was eluted with 7:1:1:1 EtOAc / MeOH / AcOH / H_2O. 857 mg of an intensely yellow acetate salt were obtained in a 67% yield. ^{1}H NMR ($CDCl_3$) δ 7.58 (s, 1H), 7.05 (d, J = 9.4 Hz, 2H), 6.86 (s, 1H), 6.78 (d, J = 9.6 Hz, 2H), 6.63 (s, 1H), 6.11 (d, J = 7.4 Hz, 1H), 5.79 (m, 1H), 5.20 (br s, 2H), 5.04 (d, J = 18.0 Hz, 1H), 4.99 (d, J = 10.0 Hz, 1H), 4.72 (m, 2H), 4.05 (t, J = 8.2 Hz, 2H), 3.96 (s, 3H), 3.94 (s, 3H), 3.88 (br t, 2H), 3.30 (s, 6H), 3.01 (m, 2H), 2.35 (t, J = 7.8 Hz, 2H), 2.31 (m, 2H), 2.28 (m, 2H), 2.05 (s, 3H), 1.43 (s, 9H); ^{13}C NMR ($CDCl_3$): 177.1, 171.6, 170.6, 159.6, 156.8, 153.6, 149.9, 147.0, 142.0, 136.6, 130.4, 128.9, 119.8, 115.4, 114.0, 111.1, 105.9, 100.1, 82.3, 64.0, 63.1, 62.2, 56.9, 56.4, 53.5, 51.0, 37.1, 35.5, 29.3, 28.0, 23.2, 22.1.

4PO-TyrO3T(DMCm)-OCH_2CN

857 mg 4PO-TyrO3T(DMCm)-O*t*Bu were dissolved in 25 mL methylene chloride with. 5 mL TFA. When reaction was complete by TLC, volatiles were removed on the

vacuum pump with a dry ice / acetone trap. 15 mL anhydrous $ClCH_2CN$ and 1 mL triethylamine were added to the flask, under argon. After stirring overnight, the volatiles in the mixture were removed on the vacuum pump with a dry ice/acetone trap. The reaction mixture was run on a column in 10:1:1:1 EtOAc / MeOH / AcOH / H_2O. acetate. Collected fractions were rotovapped to dryness, redissolved in minimal CH_2Cl_2, and extracted against H_2O to remove triethylamine salt. 651 mg of an acetate salt, a 78% yield. ^1H NMR ($CDCl_3$) δ 7.73 (s, 1H), 6.98 (d, J = 8.5 Hz, 2H), 6.79 (s, 1H), 6.70 (d, J = 8.5 Hz, 2H), 6.65 (s, 1H), 6.11 (d, J = 7.3 Hz, 1H), 5.77 (m, 1H), 5.44 (br s, 2H), 5.03 (d, J = 15.0 Hz, 1H), 4.97 (d, J = 10.5 Hz, 1H), 4.81 (m, 2H), 4.04 (t, J = 6.8 Hz, 2H), 4.02 (s, 3H), 4.00 (s, 3H), 3.96 (t, J = 6.7 Hz, 2H), 3.42 (s, 6H), 3.07 (m, 2H), 2.42 (m, 2H), 2.36 (m, 2H), 2.31 (m, 2H), 2.02 (s, 3H), 1.43 (s, 9H); ^{13}C NMR ($CDCl_3$): 177.1, 172.2, 170.2, 159.5, 157.9, 163.6, 149.8, 147.1, 141.8, 136.6, 130.1, 128.0, 120.2, 115.4, 114.5, 114.0, 111.0, 106.2, 100.0, 64.1, 63.5, 62.1, 57.6, 56.4, 53.2, 50.5, 48.8, 36.6, 25.0, 29.2.

General procedure for coupling of cyanomethyl esters to dCA

The dinucleotide was prepared as previously reported with a few modifications: 1) the dinucleotide coupling, oxidation, and deprotection with *p*-toluenesulfonic acid were done in one pot; 2) the desalting of dCA was accomplished by redissolving it in Millipore water, freezing, and lyophilization to obtain a fluffy material; 3) the tetrabutylammonium salt of the dinucleotide was formed by mixing the proper amount 1 M $N(n\text{-}Bu)_4OH$ in MeOH with a solution of dCA in water. Freezing and lyophilization provided a white fluffy solid which was then stored at -80 °C.

NVOC-TyrO3P(NVOC)-OdCA

The synthesis of NVOC-TyrO3P(NVOC)-OdCA is described as a general procedure. 23 mg of dCA (tetrabutylammonium salt) were dissolved in 0.5 mL anhydrous DMF, and were stirred with 53 mg NVOC-TyrO3P(NVOC)-OCH_2CN (four equivalents) dissolved in 0.5 mL anhydrous DMF in a 10 mL pear-shaped flask, under argon. After 4 hours, a small amount of tetrabutylammonium acetate was added using a metal spatula. The reaction was monitored by analytical HPLC, using a Waters NOVA-Pak C_{18} (150 x 3.9 mm) reverse-phase column with a gradient from 25 mM NH_4OAc (pH 4.5) to CH_3CN. When the reaction was judged complete after 24 hours, the mixture was purified by semi-preparative HPLC with a Waters NOVA-Pak C_{18} (300 x 7.8 mm) using a similar gradient. The appropriate fractions were combined, frozen, and lyophilized overnight. To remove ammonium ions, which inhibit T4 RNA ligase in the ligation of the product to $tRNA_{CUA}$, the product was redissolved in 10 mM acetic acid, frozen, and lyophilized again. This yielded 5.6 mg (6%; yields can be as low as 1%) of the desired product as a pale yellow solid. Small amounts of material were quantified by their UV-Vis spectra in solution of 10 mM acetic acid, assuming ε_{350} = 6336 M^{-1} per nitroveratryl group. ESI^+-MS: calculated for 755.7; found $[M+Na]^+$: 778.4.

NVOC-Tyr-O3S(NVOC)-dCA

Coupling procedure differed in that N(*n*-Bu)$_4$OAc salt was added to a cloudy mixture of dCA in 500 μL DMF until dissolution was complete and then this was combined with a 500 μL solution of the cyanomethyl ester in DMF. Reaction was complete after 2 h. 2 mg obtained in a 5% yield. from 24 mg of cyanomethyl ester and 32 mg dCA. ESI^+-MS: calculated for $C_{52}H_{63}N_{12}O_{27}P_2^+$: 1349.3 ; found $[M+H]^+$: 1349.3

4PO-TyrO3T(DMCm)-dCA

Prepared by the general coupling procedure. After 36 h., reaction appeared to have reached steady-state at ~ 60% completion. 6 mg obtained in a 31% yield from 16 mg of cyanomethyl ester and 20 mg dCA. Quantification was performed assuming ε_{350} = 13,000 M^{-1} for the coumarin group. ESI^+-MS: calculated for $C_{50}H_{63}N_{10}O_{20}P_2^+$: 1185.4 ; found $[M+]^+$: 1185.4

Unnatural amino suppression in *Xenopus* oocytes

The site-directed mutagenesis of the nAChR TAG mutants, gene construction and synthesis of suppressor $tRNA_{CUA}$ and ligation of aminoacyl-dCA to $tRNA_{CUA}$ have been described previously.[13, 14] Plasmid DNAs were linearized with NotI, and mRNA was transcribed using the Ambion (Austin, TX) T7 mMESSAGE mMACHINE Kit.

Oocytes were removed from *Xenopus laevis* as described and maintained at 18 °C, in ND96 solution (96 mM NaCl/2 mM KCl/1.8 mM CaCl2/1 mM MgCl2/5 mM HEPES/2.5 mM sodium pyruvate/0.5 mM theophyline/10 g/ml Gentamycin, pH 7.5, with NaOH). Before microinjection, the NVOC-aminoacyl-$tRNA_{CUA}$ was deprotected by irradiating the sample for 5 or 10 min. with a 1000 W Hg/Xe arc lamp (Oriel) operating at 400 W equipped with WG-335 and UG-11 filters (Schott). 4PO-protected $tRNA_{CUA}$-aa was mixed 1:1 with a solution of saturated I_2 in water and allowed to sit for ten minutes at room temperature. Each oocyte was injected with a 1:1 mixture of deprotected aminoacyl-$tRNA_{CUA}$ (25-50 ng) and mRNA (12.5–18 ng of total at a concentration ratio of 20:1:1:1 for α:β:γ:δ subunits) in a volume of 50 nL.

Electrophysiological recordings

Voltage-clamped electrophysiological recordings were carried out 24-72 hours after injection. Whole-cell currents from oocytes were measured using a Geneclamp 500 amplifier and pCLAMP software (Axon Instruments, Foster City, CA) in the two-electrode voltage-clamp configuration. Microelectrodes were filled with 3 M KCl and had resistances ranging from 1.0 to 2.5 MΩ. Oocytes were continuously perfused with a nominally Ca^{2+}-

free bath solution consisting of 96 mM NaCl, 2 mM KCl, 1 mM $MgCl_2$, and 5 mM HEPES (pH 7.5). Microscopic ACh-induced and TMB-8 or QX-314-blocked currents were recorded in response to bath application of ACh and TMB-8 at a holding potential of -80 mV. Low (5.5 - 6.5) and high (8.5 and 9.0) pH solutions were of the same composition as Ca^{2+}-free bath with MES (low) or CHES (high) substituted for HEPES buffer. To ensure that changes in buffer were not responsible for the observed changes in channel conductance, recordings were taken at pH 7.0 and 8.0 in HEPES alongside recordings in MES and CHES.

References

(1) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R., *Biophys. J.* **2002,** 83, 1731-1748 Battistuzzi, G.; Borsari, M.; Cowan, J. A.; Ranieri, A.; Sola, M., *J. Am. Chem. Soc.* **2002,** 124, 5315-5324 Schutz, C. N.; Warshel, A., *Proteins* **2001,** 44, 400-417 Ferreira, G. C.; Zhang, J. S., *Cell. Mol. Biol.* **2002,** 48, 827-833.

(2) Seeman, J. I.; Whidby, J. F., *J. Org. Chem.* **1976,** 41, 3824-3826.

(3) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A., *Proc. Natl. Acad. Sci. USA* **1998,** 95, 12088-12093.

(4) Corringer, P. J.; Le Novere, N.; Changeux, J. P., *Annu. Rev. Pharmacol. Toxicol.* **2000,** 40, 431-458.

(5) Silman, I.; Karlin, A., *Science* **1969,** 164, 1420-1421.

(6) Sullivan, D. A.; Cohen, J. B., *J. Biol. Chem.* **2000,** 275, 12651-12660.

(7) Arias, H. R., *Brain Res. Rev.* **1997,** 25, 133-191.

(8) Li, L.; Zhong, W.; Zacharias, N.; Gibbs, C.; Lester, H. A.; Dougherty, D. A., *Chem. Biol.* **2001,** 8, 47-58.

(9) Labarca, C. G.; Nowak, M. W.; Zhang, Y.; Tang, L.; Deshpande, P.; Lester, H. A., *Nature* **1995,** 376, 514-516.

(10) Bunnelle, W. H.; Dart, M. J.; Schrimpf, M. R., *Curr. Top. Med. Chem.* **2004,** 4, 299-334 Decker, M. W.; Rueter, L. E.; Bitner, R. S., *Curr. Top. Med. Chem.* **2004,** 4, 369-384 Grutter, T.; Le Novere, N.; Changeux, J. P., *Curr. Top. Med. Chem.* **2004,** 4, 645-650.

(11) Peng, L.; Wirz, J.; Goeldner, M., *Tet. Let.* **1997,** 38, 2961-2964 Peng, L.; Wirz, J.; Goeldner, M., *Angew. Chem. Int. Ed.* **1997,** 36, 398-400.

(12) Schade, B.; Hagen, V.; Schmidt, R.; Herbrich, R.; Krause, E.; Eckardt, T.; Bendig, J., *J. Org. Chem.* **1999,** 64, 9109-9117 Sarker, A. M.; Kaneko, Y.; Neckers, D. C., *J. Photochem. Photobiol. A* **1998,** 117, 67-74.

(13) Petersson, E. J.; Brandt, G. S.; Zacharias, N. M.; Dougherty, D. A.; Lester, H. A., *Methods Enzymol.* **2003,** 360, 258-273.

(14) Nowak, M. W.; Gallivan, J. P.; Silverman, S. K.; Labarca, C. G.; Dougherty, D. A.; Lester, H. A., *Methods Enzymol.* ***1998,*** 293, 504-529.

(15) Perrin, D. D., *Dissociation Constants of Orgainc Bases in Aqueous Solution: Supplement 1972.*; Butterworth & Co.: London, U. K., 1972.

(16) Pappone, P. A.; Barchfeld, G. L., *J. Gen. Physiol.* **1990,** 96, 1-22.

(17) Palma, A.; Li, L.; Chen, X.; P., P.; McNamee, M., *J. Membrane Biol.* **1991,** 120, 67-73.

(18) Highbarger, L. A.; Gerlt, J. A.; Kenyon, G. L., *Biochem.* **1996,** 35, 41-46.

(19) Cho, A. K.; Jenden, D. J.; Lamb, S. I., *J. Med. Chem.* **1972,** 15, 391-394.

(20) Hubbard, S. J.; Thornton, J. M. *NACCESS*, Department of Biochemistry and Molecular Biology, University College: London, U.K., 1993.

Section 1: Chapter 4

Probing Nicotine Binding with Backbone Mutagenesis and Comparison to Epibatidine, a High Affinity Nicotine Analog

Seeking an Affirmative "Nicotinic" Binding Determinant

Studies of the muscle-type nicotinic acetylcholine receptor (nAChR) using unnatural amino acid mutagenesis showed that a cation-π interaction with a key tryptophan, Trp α149, is essential to determining acetylcholine (ACh) potency at the nAChR.[1] The other titular agonist, nicotine (Nic), does not exhibit this strong dependence on the cation-π interaction, although biochemical studies have shown that it binds in the same pocket.[2] These findings suggested that agonists of the nAChR could fall into two classes, which we term "cholinergic," binding like ACh, and "nicotinic," binding like Nic. The nicotinic class was largely defined by its lack of a cation-π interaction, a negative binding determinant. Investigations of the effect of the local protein environment on the binding of Nic and a tertiary ACh analog, norACh, further differentiated the binding of the two classes of compounds. While the pK_a of norACh was affected by the local environment of the binding site, the pK_a of Nic was not. We hypothesized that the interaction of the Nic amine proton with a specific functional group in the binding site was able to compensate for the hydrophobic character of the aromatic binding site. We sought to identify candidates for this functional group interaction by examining the crystal structure of acetylcholine binding protein (AChBP, a homolog of the nAChR binding domain).

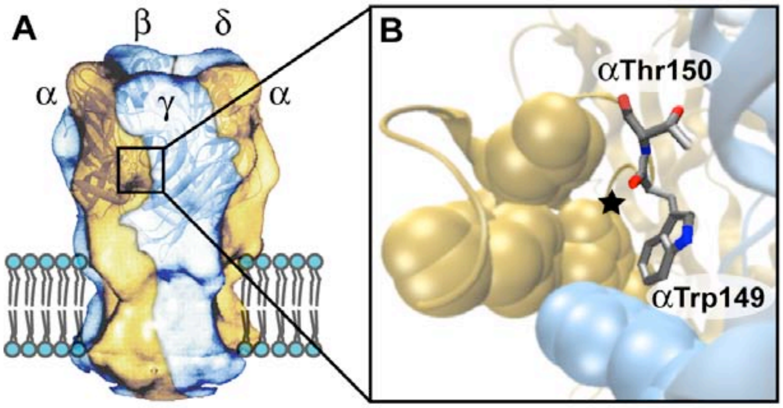


Figure. 1. Images of the nAChR. (A) The overall layout of the muscle receptor, indicating the arrangements of five subunits around a central pore. The receptor electron density from cryoelectron microscopy[3] is shown superimposed over a ribbon diagram of AChBP, which corresponds to the extracellular domain of the receptor. (B) The agonist binding site from AChBP with nicotinic numbering. Aromatic residues lining the binding pocket are shown as space-filling models. Residues and ribbons from the α subunit are gold; those from the γ subunit are blue. The star marks the backbone carbonyl that participates in a hydrogen bond with agonists.

Several modeling studies based on the original structure of AChBP suggested a hydrogen bonding interaction from the N$^+$–**H** of Nic to the backbone carbonyl of Trp α149.[4, 5] This carbonyl is denoted by a star in Fig. 1. ACh cannot make a hydrogen bond of this sort. Thus, this hydrogen bond could be a second discriminator between ACh and Nic, an affirmative determinant of Nic binding rather than the negative "lack of a cation-π interaction." While this work was nearing completion, Sixma reported the crystal structure of AChBP in the presence of bound Nic,[6] confirming the proposed hydrogen bond between Nic and the backbone carbonyl of Trp α149 at the agonist-binding site.

Epibatidine, A High Affinity Nic Analog

One challenge in studying the activity of Nic at the nAChR is that Nic has low agonist potency at the muscle receptor subtype.[7] Nic is a more potent agonist at some neuronal nAChR subtypes.[8] As a low potency agonist, the binding of Nic may be more flexible than the binding of ACh. It may be that the reason we do not observe a linear correlation between our perturbations of cation-π binding and Nic potency is that when the cation-π interaction is weakened, Nic alters its binding geometry so that it is no longer interacting as substantially with Trp 149. Nic clearly interacts with the channel in a sufficiently specific way to gate the channel, despite its low affinity. However, this low affinity may mean that we would never firmly identify a single interaction that is highly deterministic of Nic binding. For this reason, this study also examined epibatidine (Epi), a potent agonist at both muscle and neuronal nAChRs.[8, 9] Epi, while structurally similar to Nic, has a potency comparable to ACh.[10, 11] As such, Epi perhaps serves as a more meaningful probe of nicotinic interactions at the muscle nAChR than Nic itself. The structures and electrostatic potential surfaces of the agonists are presented in Fig. 2.

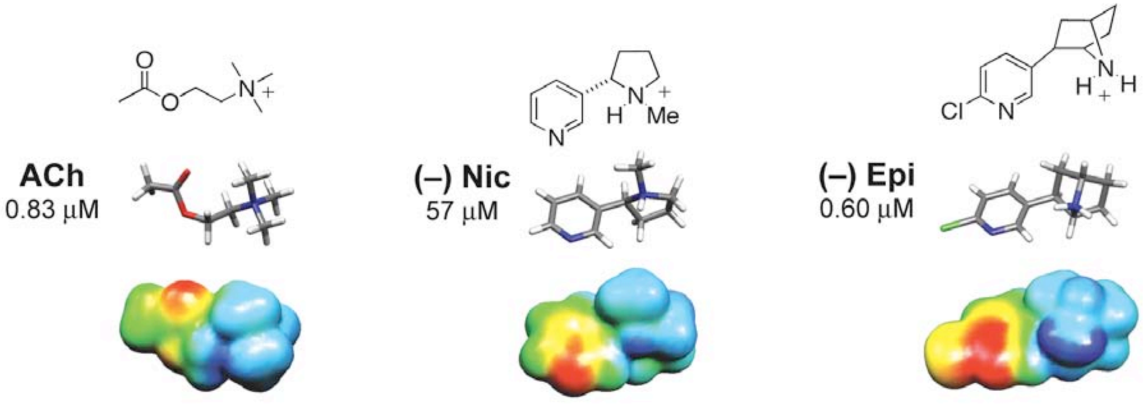


Figure 2. nAChR agonists examined in this study. Shown are calculated geometries and EC_{50} values for activation of the wild type nAChR . HF/6-31G electrostatic surfaces calculated using Molekel contrast the focused N$^+$–**H** positive charge on Nic and Epi with the diffuse ACh ammonium charge. Electrostatic surfaces correspond to an energy range of + *10 to* + *130* kcal/mol, where blue is highly positive and red is less positive.

Experimental Approach

The goals of this study were thus two-fold. First, we wished to evaluate the significance of the apparent hydrogen bond between Nic and the backbone carbonyl of Trp α149 to Nic binding. Second, we wished to evaluate the factors that render Epi almost 100-fold more potent than Nic, despite the clear structural similarity of the two. Studies employing fluorinated Trp derivatives at α149 identical to those performed with Nic and ACh were performed to probe the Epi cation-π interaction. In addition, we probe the functional significance of the interaction with the backbone carbonyl at Trp α149 with both Nic and Epi by weakening the hydrogen bonding ability of the backbone carbonyl through a backbone amide-to-ester mutation. This experimental work was conducted by Amanda Cashin. Modeling based on these data expands on previous computational models of cation-π binding in which Na^+ was used in an idealized interaction with fluorinated indoles. These more elaborate models suggest precise interactions that differentiate the agonists.

Unnatural amino acids were incorporated into the nAChR using *in vivo* nonsense suppression methods, and mutant receptors were evaluated electrophysiologically.[12] In studies of weak agonists and/or receptors with diminished binding capability, it is necessary to introduce another mutation that independently decreases EC_{50}. We accomplished this via a Leu-to-Ser mutation in the β subunit at a site known as 9' in the M2 transmembrane region of the receptor. This M2-β9’ site is almost 50 Å from the binding site, and previous work has shown that a Leu9'Ser mutation lowers the EC_{50} by a factor of roughly 10 without altering trends in EC_{50} values.[2, 13] Measurements of EC_{50} represent a functional assay; all mutant receptors reported here are fully functioning ligand-gated ion channels. It is important to note that the EC_{50} value is not a binding constant, but a composite of equilibria for both binding and gating.

Epibatidine Binds with a Potent Cation-π Interaction at Trp α149

The existence of a cation-π interaction between Epi and Trp α149 was evaluated using our previously developed strategy, the incorporation of a series of fluorinated Trp derivatives (5-F-Trp, 5,7-F_2-Trp, 5,6,7-F_3-Trp and 4,5,6,7-F_4-Trp). The EC_{50} values for the wild type and mutant receptors are shown in Table 1.

Table 1. Mutations Testing Cation-π Interactions at α149

	Trp	F-Trp	F_2-Trp	F_3-Trp
Epi[a]	0.83 ± 0.08	4.8 ± 0.1	9.3 ± 0.5	18 ± 2
Cation-π	32.6	27.5	23.3	18.9

[a] EC_{50} (μM) ± standard error of the mean. The receptor has a Leu9’Ser mutation in M2 of the β subunit.

Attempts to record dose-response relations from 4,5,6,7-F$_4$-Trp at α149 were unsuccessful, because this mutant required Epi concentrations above 100 μM. At these concentrations Epi becomes an effective open channel blocker,[9] confounding efforts to obtain an accurate dose-response curve. A clear trend can be seen in Table 1: each additional fluorine produces an increase in EC$_{50}$. As in previous work, our measure for the cation-π binding ability of the fluorinated Trp derivatives is the calculated binding energy of a generic probe cation (Na$^+$) to the corresponding substituted indole.[1, 2] This method provides a convenient way to express the trend in the dose-response data in a quantitative way. A "fluorination plot" of the logarithmic ratio of the mutant EC$_{50}$ to the wild type EC$_{50}$ versus the cation-π binding ability for Trp α149 reveals a compelling linear relationship (Fig. 3). These data demonstrate that the secondary ammonium group of Epi makes a cation-π interaction with Trp α149 in the muscle-type nAChR.

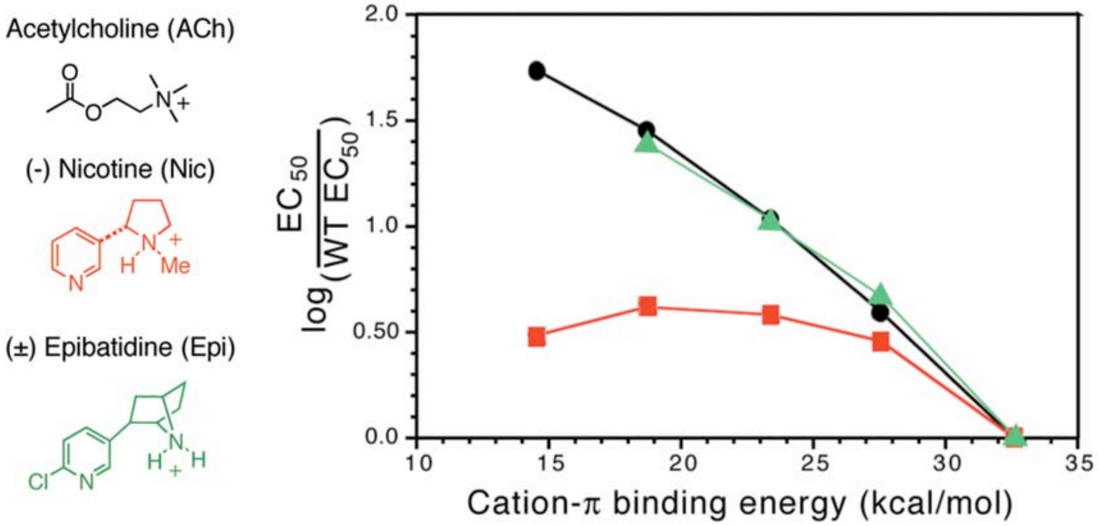
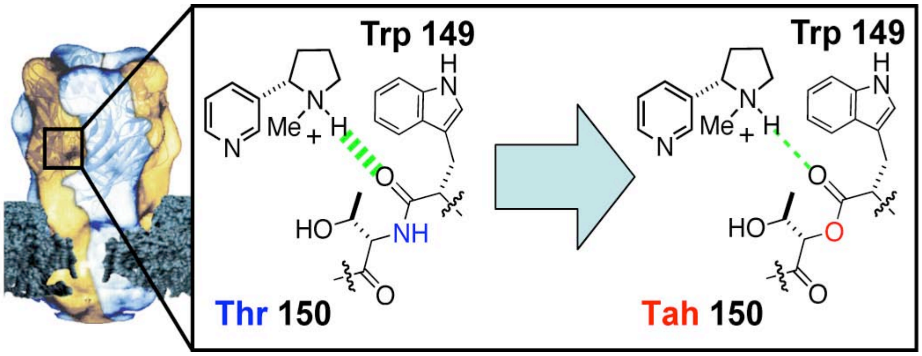


Figure 3. Fluorination plot for nAChR agonists; Epi data from Table 1; ACh data from reference,[1] Nic data from reference.[2, 14] The log [EC$_{50}$ /EC$_{50}$ (wt)] versus calculated cation-π ability is plotted for the series of fluorinated Trp derivatives at Trp α149.

Nicotine and Epibatidine Hydrogen Bond to the Carbonyl Oxygen of Trp α149

The recently reported crystal structure of AChBP with Nic bound indicated a hydrogen bond between the pyrrolidine N$^+$–**H** of Nic and the backbone carbonyl of Trp α149,[6] an interaction that had been anticipated by several modeling studies. To evaluate this possibility, the backbone amide at this position was converted to an ester by replacing Thr α150 with the analog α-hydroxy threonine (Tah) using the nonsense suppression methodology (Scheme 1). Converting an amide carbonyl to an ester carbonyl weakens the hydrogen bonding ability of the oxygen, an effect that has been valued at ~ 0.9 kcal/mol.[15]



Scheme 1. Conversion of Thr150 to Tah and its effect on H-bonding to nicotine.

The results of the incorporation of Tah at α150 are shown in Table 2. Upon ester substitution, the EC_{50} for Nic increases 1.6 fold. The change is larger for the more potent agonist Epi; conversion of the backbone carbonyl of Trp α149 to an ester leads to a 3.7-fold increase in EC_{50} (Fig. 4). In contrast, ACh, lacking a proton at the cationic center, shows a 3.3 fold *decrease* in EC_{50} (Fig. 4). These results further highlight the distinction between nicotinic and cholinergic agonists.

Table 2. Mutations Testing H-bond Interactions at α150[a]

Agonist	Thr[b]	Tah	Tah/Thr
ACh	0.83 ± 0.04	0.25 ± 0.01	0.31
Nic	57 ± 2	92 ± 4	1.6
Epi	0.60 ± 0.04	2.2 ± 0.2	3.7

[a] EC_{50} (μM) ± standard error of the mean. The receptor has a Leu9'Ser mutation in M2 of the β subunit.
[b] Rescue of wild type by nonsense suppression.

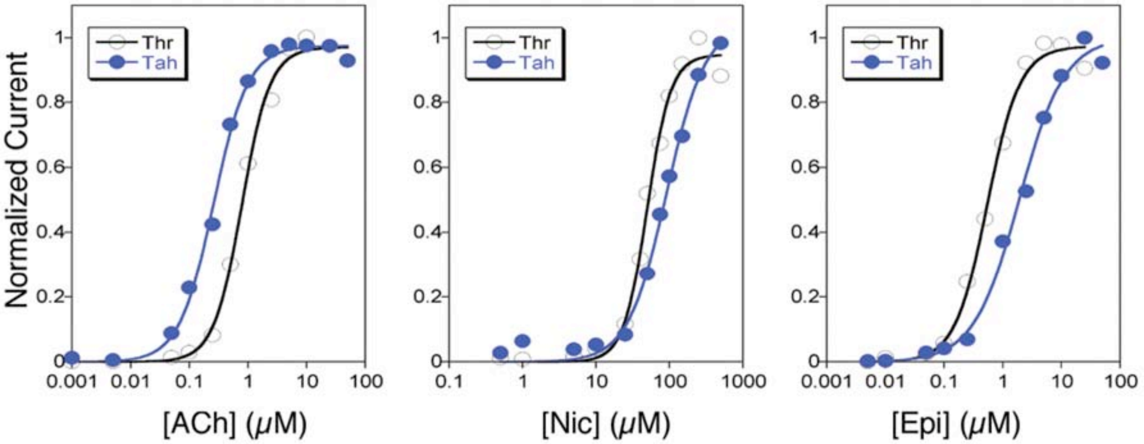


Figure 4. Representative acetycholine, (-) nicotine, and (±) epibatidine dose-response relations for nAChR suppressed with Thr (○) and Tah (●). Studies incorporate a βLeu9'Ser mutation.

Computational Modeling

Computational modeling was used to understand the variations in binding properties among the three agonists. Focusing on the interactions with Trp α149, we docked the ligands using *ab initio* (HF/6-31G) calculations taking into account both the cation-π interaction and the carbonyl hydrogen bond. Initial tryptophan and ligand coordinates were taken from the AChBP-based homology models of Changeux.[5] Geometry optimizations, counterpoise corrections, and zero point energy corrections were all performed in the gas phase. The optimized geometries for free ACh and Nic are in keeping with previous calculations at higher levels of theory and with solution NMR studies (bent "tg" structures are favored for ACh).[16] The calculated binding energies (Table 3) are consistent with those from previous computational studies of metal binding complexes with both cation-π and cation-carbonyl interactions[17] and studies of hydrogen bonds to protonated Nic.[18]

Table 3. Calculated Binding Energies* (kcal/mol)

Agonist	Amide Trp	Ester Trp	Ester – Amide
ACh	- 21.7	- 16.7	5.0
Nic	- 30.4	- 24.3	6.1
(+) Epi	- 34.8	- 26.6	8.2
(–) Epi	- 35.5	- 27.7	7.8

* HF/6-31G, Gas phase.

The calculated binding energies are summarized in Table 3. Experimentally, the EC_{50}s of (+) and (-) Epi are nearly identical for a given acetylcholine receptor subtype,[19] and we find that the calculated binding energies and the key geometrical parameters (Fig. 5) are indeed very similar for the two enantiomers. Epibatidine binds the amide more strongly than Nic by ~ 5 kcal/mol. As expected, conversion of the Trp α149 amide to an ester weakens the binding interactions to both Epi and Nic. The calculated energetic effect of ester conversion is larger for Epi than for Nic (8 kcal/mol vs. 6 kcal/mol). Using the PCM solvation model,[20] we studied these interactions in solvents of differing polarity (Table 4).

Table 4. Solvent Effects on Binding Energy Differences[a]

Agonist	Ester Binding Energy – Amide Binding Energy (kcal/mol)			
	Gas	THF	Ethanol	Water
ACh	5.0	0.6	- 1.7	- 2.0
Nic	6.1	3.1	1.2	- 0.8
Epi[b]	8.0	7.0	5.0	4.7

[a] ε(THF) = 7.6, ε(ethanol) = 24.3, ε(water) = 78.5.

[b] Average of energies for epibatidine enantiomers.

In each solvent, Epi favors amide binding over ester binding to a greater degree than Nic. The changes in hydrogen bonding energies observed in different solvent systems are consistent with similar calculations published by Houk and coworkers.[21]

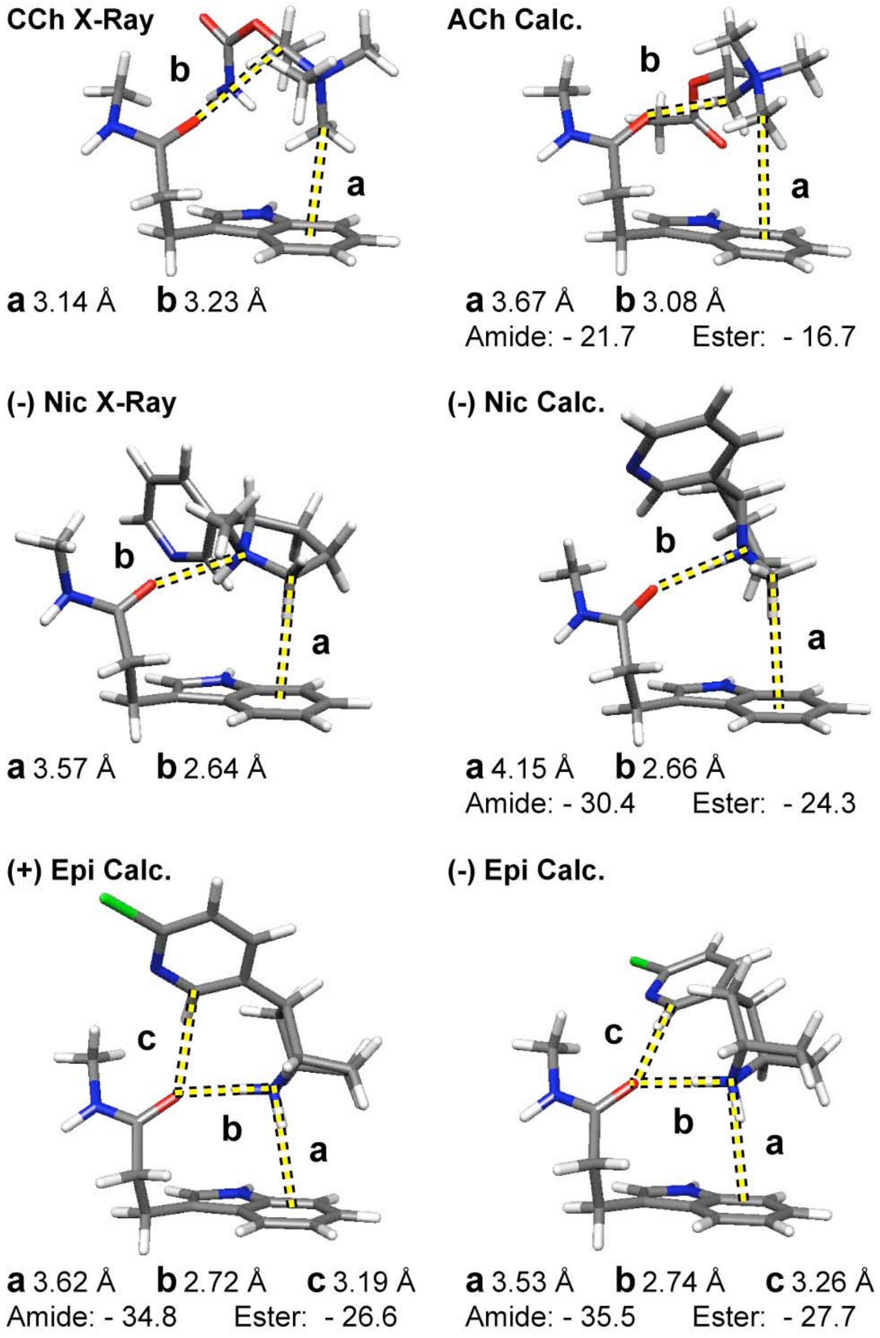


Figure 5. Crystal structure data (X-Ray) and computational modeling (Calc.) of agonist binding. Crystal structures for CCh and Nic were taken from Celie *et al.* (PDB ID 1UW6 (Nic) and 1UV6 (CCh)) [6]. Calculations were performed for ACh, (-) Nic, (+) Epi and (-) Epi. Distance **a** represents a cation-π interaction; **b** represents an N^+–**H** or N^+C–**H** hydrogen bond with the backbone carbonyl and **c** represents $C_{aromatic}$–**H•••**O=C hydrogen bond with the backbone carbonyl. Gas phase HF/6-31G optimized geometries (Å) are reported. Hydrogens were added to the x-ray structures using Gaussview.

The geometries of Figure 5 are consistent with the energetic trends observed. The cation-π interaction is expected to be much stronger for Epi than for Nic. The calculated N^+ to π-centroid distance is substantially shorter for Epi (**a** in Fig. 5). In addition, Epi points an N^+–**H** cationic center towards the Trp indole ring, vs. the N^+CH_2–**H** center of Nic (Fig. 5). The cationic center of Epi has a much more positive electrostatic potential than that of Nic (+ 139 kcal/mol for Epi, + 112 for Nic). These potentials, indicators of cation-π binding strength, are consistent with the experimental observation that epibatidine has a much stronger cation-π interaction than Nic.

Nic and Epi make significant hydrogen bonds to the Trp α149 carbonyl oxygen with an N^+–**H** group (**b** in Fig. 5). The geometrical parameters for interaction **b** with the two agonists are very similar, suggesting the two hydrogen bonds are comparably strong. In addition, the calculations suggest a second, previously unanticipated interaction between the $C_{aromatic}$–**H** of the carbon adjacent to the pyridine N and the same carbonyl (**c** in Fig. 5). Based on both the distance (**c** in Fig. 5) and angle (C-H-O = 168° in Epi vs. 145° in Nic), we expect the $C_{aromatic}$–**H•••**O=C interaction to be stronger for Epi than for Nic.

Comparison of Acetylcholine, Nicotine, and Epibatidine Binding

Previously, we observed an intriguing result: Nic and ACh use different noncovalent interactions to bind the muscle-type nAChR. ACh forms a strong cation-π interaction with Trp α149; Nic does not. Although known as the nicotinic receptor, the form we study here, that found in the peripheral nervous system, is relatively insensitive to Nic. At this muscle-type receptor ACh is over 70-fold more potent than Nic. The behavioral and addictive effects of Nic arise exclusively from interactions with one or more neuronal subtypes of nAChR found in the central nervous system, where Nic and ACh are generally comparably potent. We therefore wanted to probe a nicotinic-type agonist that is potent at the muscle receptor, and Epi was the logical choice. This alkaloid natural product possesses potent analgesic properties[19] and has served as a lead compound for a number of pharmaceutical programs targeted at the nAChR.[11] In the present work, we find two specific interactions that distinguish among the three agonists considered here, ACh, Nic, and Epi.

First, we now find that Epi makes a strong cation-π interaction with Trp α149 of the muscle-type nAChR. This result contrasts sharply to Nic, and this observation helps to explain the much higher affinity of Epi for this receptor relative to Nic. The apparent magnitudes of the cation-π interactions, indicated by the slopes of the fluorination plots in Fig. 3, are comparable for ACh and Epi. This similarity is surprising, because the cationic centers of the two agonists are chemically quite different (quaternary ammonium for ACh; protonated secondary ammonium for Epi). The computer modeling summarized in Fig. 5

nicely rationalizes the observed cation-π binding behavior. Epi, like ACh, makes much closer contact with the indole ring than Nic. Both the interaction distance (**a** in Fig. 5) and the electrostatic potential on the cationic hydrogen (Fig. 2: N^+–**H** in Epi; vs. N^+CH_2–**H** in Nic) favor the cation-π interaction in Epi over Nic.

The second discriminator we have probed is hydrogen bonding. A newer crystal structure of the AChBP includes Nic at the binding site.[6] The structure confirms the existence of a hydrogen bond between Nic and the backbone carbonyl of Trp α149, an interaction anticipated by modeling studies. In efforts to probe this non-covalent interaction, we studied the effects of decreasing the hydrogen bond acceptor ability of the backbone carbonyl of Trp α149. In such studies, the clear distinction between ACh and nicotinic agonists is strengthened. Nic and Epi, containing tertiary and secondary cationic center srespectively, both show increases in EC_{50} compared to the native receptor in response to the amide-to-ester modification (Table 2). The effect is larger with the more potent agonist, Epi. Thus, the experimental data support the suggestion that Nic and Epi interact with the nAChR through a hydrogen bond with the backbone carbonyl of Trp α149.

Understanding the Effects of Ester Conversion on Acetylcholine Binding

ACh, with a quaternary cationic center that cannot make a conventional hydrogen bond, shows a decrease in EC_{50} at the ester-containing receptor compared to the native receptor. We had anticipated that the binding of ACh would be unaffected by such a subtle change. The origin of this effect is unclear at present and is the object of further investigation. Here we consider two possibilities.

In the recently reported crystal structure of AChBP binding to carbamoylcholine (CCh), a cholinergic analogue of ACh, the backbone carbonyl oxygen of interest here makes contact with a CH_2 group adjacent to the $N^+(CH_3)_3$ group (CCh: $NH_2C(O)OCH_2\mathbf{CH_2}N^+(CH_3)_3$). This N^+CH_2 carries a significant positive charge, like the N^+CH_3 groups, and so a favorable electrostatic interaction is possible. This interaction with CCh would be much weaker than the N^+–**H** hydrogen bonds of Nic and Epi, but perhaps not negligible. Assuming that norACh binds to the receptor as CCh does in the crystal structure and ACh does in our calculations, it is hard to imagine how it could not also utilize a hydrogen bonding interaction with the carbonyl in question to stabilize its protonated state. It would require only a very minor geometry adjustment. This makes it difficult to explain why norACh has a perturbed pK_a in the binding site, while Nic does not.

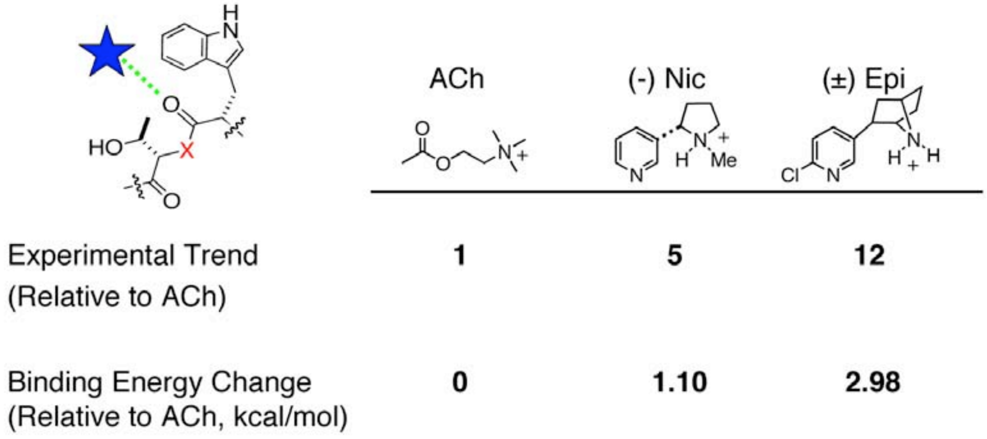
Interestingly, Sixma and coworkers noted that the binding of CCh to AChBP is less enthalpically favorable than that of Nic. They attribute this observation to the net unfavorable burial of the carbonyl oxygen by CCh – the weak interaction with the CH_2

group cannot compensate for the loss of hydrogen bonding, presumably to water molecules. With Nic, a strong hydrogen bond compensates this desolvation penalty more effectively, at least in the case of AChBP.

Remarkably, the relatively simple model calculations we conducted recapitulate this effect. In the gas phase, it is better to bind to the backbone amide than the ester for all three agonists. However, as solvation is introduced, the trend is reversed (Table 4). Interestingly, when a solvent of moderate polarity – ethanol – is used, ACh prefers the ester backbone, while Nic and Epi prefer the amide. The ethanol environment is defined in these calculations by a dielectric constant of 24.3. Two lines of evidence indicate that this is a reasonable estimate of the effective dielectric of the binding pocket of the AChBP or nAChR. (See Chapter 3) First, it is consistent with previous experimental measurements of a perturbed local pK_a in the nAChR binding site.[22] Second, calculations of the solvent accessible surface area of the binding site residues show that Trp 149 is 11% solvent-accessible. A moderate dielectric of 24.3 seems reasonable for the partially-exposed binding site. Thus, it may be that the EC_{50} for ACh decreases when the ester is introduced because the desolvation penalty of the ester carbonyl oxygen is less severe than the amide.

While the agreement between experiment and theory presented here and Sixma's thermodynamic measurements is gratifying, we also propose a second possible explanation. Highly conserved Asp α89 (Asp 85 AChBP numbering) makes a number of significant contacts with nearby residues, suggesting it plays a key structural role in shaping the agonist binding site.[6, 23] One such interaction is a hydrogen bond between the Asp α89 carboxylate side chain and the NH group of the backbone amide of Trp α149. The amide-to-ester mutation of the present study eliminates the NH and so removes this interaction. A possible outcome of this would be a structural change that would affect gating, biasing the conformational change in the direction of the open channel.

Regardless of the origin of the effect, it is reasonable to propose that the effect of ester substitution we see with ACh can be considered as the "background" for the Thr150Tah mutation. That is, if the magnitude of the cholinergic N^+CH_2•••O=C interaction is small, then both the desolvation and gating effects proposed are "generic" and should occur with all agonists. In this case, the changes in EC_{50} we measure for Nic or Epi actually represent the product of two terms: a generic 3.3-fold decrease evidenced by ACh, and a term specific to Nic or Epi. As such, the drop in hydrogen bonding strength is calculated to be 1.6*3.3 or ~5-fold for Nic, and 3.7*3.3 or ~12-fold for Epi. Energetically, these factors correspond to 1.0 and 1.5 kcal/mol, respectively, which although they do not correlate precisely with the calculated binding energies, correlate well with the trend. (Fig. 6)



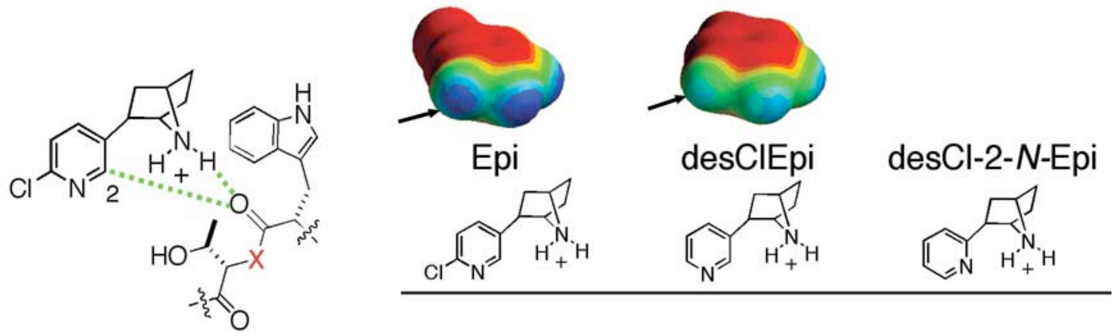
	ACh	(-) Nic	(±) Epi
Experimental Trend (Relative to ACh)	**1**	**5**	**12**
Binding Energy Change (Relative to ACh, kcal/mol)	**0**	**1.10**	**2.98**

Figure 6. Summary of Hydrogen Bond Study

Understanding Epibatidine’s High Affinity from Computational Models

The larger amide/ester effect seen for Epi vs. Nic suggests a stronger N^+–**H•••**O=C hydrogen bond in the former. However, these hydrogen bonds (**b** in Fig. 5) are geometrically very similar in the two complexes, suggesting that they are of comparable strengths. We propose here an alternative rationalization invoking the novel $C_{aromatic}$–**H** hydrogen bond revealed by our modeling studies. Aromatic hydrogens intrinsically carry a significant positive electrostatic potential (+ 18 kcal/mol in benzene). This effect is amplified when the carbon is *ortho* to a pyridine-type N (+ 24 kcal/mol in pyridine) and *meta* to an electron-withdrawing Cl (+ 31 kcal/mol in 2-chloropyridine). (See Fig. 7) Thus, we expect interaction **c** to be energetically significant. Geometrically, the $C_{aromatic}$–**H** hydrogen bond to the carbonyl (**c** in Fig. 5) is much tighter and better aligned for Epi than Nic. The computations thus suggest that it is this unconventional hydrogen bond (**c**), rather than the anticipated hydrogen bond (**b**), that rationalizes the slightly greater response of Epi vs. Nic to the backbone change. Thus, the small structural differences between Epi and Nic nicely explain their differing affinities. The secondary ammonium of Epi provides two N^+–**H**s that can undergo strong electrostatic interactions – a cation-π interaction and a hydrogen bond to a carbonyl. The tertiary ammonium of Nic can only make the hydrogen bond. Second, the slightly different positioning of the pyridine group in Epi allows for a more favorable $C_{aromatic}$–**H•••**O=C hydrogen bond than for Nic.

Studies by Spang *et al.* nicely support our results.[24] They studied the EC_{50}s of a series of Epi analogs with different substitution patterns at various neuronal nAChRs. They found that removal of the chloro group (desClEpi) had a 10-fold effect on binding in two cases, and that changing the position of the pyridine nitrogen had large effects in all cases (desCl-2-*N*-Epi). Their results are summarized in Figure 7. Positioning the pyridine nitrogen at the 2-poistion in the ring would eliminate the $C_{aromatic}$–**H** hydrogen bond and establish an unfavorable electronic clash between the lone pairs of the carbonyl oxygen on the protein and the pryidine nitrogen on desCl-2-*N*-Epi. This leads to a 10- to 100-fold decrease in affinity relative to the isodesmic desClEpi.



	Epi	desClEpi	desCl-2-*N*-Epi
α7 EC$_{50}$ (μM, (±) Avg.)	**2.3**	**4.9**	**40**
α3β4 EC$_{50}$ (μM, (±) Avg.)	**0.03**	**0.4**	**470**
α4β2 EC$_{50}$ (μM, (±) Avg.)	**0.02**	**1.7**	**50**

Figure 7. The Effects of the Position of the Pyridine Nitrogen on Epi Potency. From Spang *et al.*, 2000.[24] HF/6-31G* electrostatic surfaces calculated using Molekel show the increased positive electrostatic potential on the hydrogens of chloropyridine. Electrostatic surfaces correspond to an energy range of 0 to + 22 kcal/mol, where blue is highly positive and red is less positive.

Conclusions

Since we found that Nic did not seem to enjoy a strong cation-π interaction with any of the aromatic residues in the muscle nAChR binding site, we sought to find out what Nic did interact with. This lead us to develop the concept of nicotinic and cholinergic categories of agonists. The nicotinic agonists Nic and Epi both experience a favorable hydrogen bonding interaction with the carbonyl of Trp α149, which is qualitatively distinct from the interaction (if any) of ACh with this group. Epi is a much more potent agonist than Nic at the muscle-type nAChR because, along with hydrogen bonding, Epi experiences a cation-π interaction comparable to that seen with ACh, while Nic does not. In addition, Epi picks up a subtle $C_{aromatic}$–**H•••**O=C hydrogen bond that Nic does not.

It seems that our previous criterion of separating agonists on the basis of a cation-π interaction with Trp 149 is an overly simplistic interpretation. The linearity of the fluorination plot for an agonist is always reflective of the role of the cation-π interaction in conjunction with the agonist's other binding determinants. The agonist's binding interactions must be sufficiently strong and specific to permit our studies. That is, we cannot study the effects of Trp fluorination past the point at which we have eliminated all cation-π binding for that agonist. Part of our difficulty in understanding the results of our experimental observations may come from the simplicity of our computational models. Our calculations of cation-π binding were performed on an idealized interaction, Na^+ binding to an indole. A sufficiently realistic model of the binding site environment should yield a linear relationship with the observed changes in potency regardless of the perturbation we make. This may be required to understand the binding of low affinity agonists like nicotine. They may not have any interaction that is sufficiently strong that we can perturb it in a successive way as we did for the Trp cation-π interaction with ACh and Epi. Low affinity agonists may rely on the flexibility of the receptor binding site to provide several possible binding modes. In the following chapter, we will develop more complex computational models which take into account some of this flexibility to try to understand Nic binding.

Materials and Methods

This study was performed in collaboration with Amanda Cashin, who performed the electrophysiological experiments. Those methods have been described elsewhere, thus we only describe the computational work here.[25]

Acetylcholine, (-) nicotine, (+) epibatidine, (-) epibatidine, 3-(1H-Indol-3-yl)-N-methyl-propionamide, 3-(1H-Indol-3-yl)-O-methyl-propionate and the hydrogen-bonded complexes shown in Fig. 5 were optimized at the HF/6-31G level of theory. For the acetylcholine, (-) nicotine, and (-) epibatidine complexes, the starting coordinates of the ligand and Trp 147 (α7 numbering) were taken from the docked structures of Changeux and coworkers available at http://www.pasteur.fr/recherche/banques/LGIC/LGIC.html. The optimized geometries were fully characterized as minima by frequency analysis. Energies were calculated at the HF/6-31G level. Basis set superposition error (BSSE) corrections were determined in the gas phase at the HF/6-31G level, using the counterpoise correction method of Boys and Bernardi.[26] Zero point energy (ZPE) corrections were included by scaling the ZPE correction given in the HF/6-31G level frequency calculation by the factor of 0.9135 given by Foresman and Frisch.[27] All calculations were carried out with the Gaussian 98 program.[28] Binding energies were determined by comparing the BSSE- and ZPE-corrected energies of the separately optimized ligand and tryptophan analog to the energy of the complex. Solvent effects were added to the gas phase structures using the polarizable continuum model (PCM) self-consistent reaction field[20] with ε(THF) = 7.6, ε(EtOH) = 24.3, and ε(H2O) = 78.5. The optimized geometries are reported below.

Electrostatic potential surfaces were created with Molekel, available at www.cscs.ch/molekel/.[29] The electrostatic potential for each structure was mapped onto a total electron density surface contour at 0.002 e/Å^3. These surfaces were color-coded so that red signifies a value less than or equal to the minimum in positive potential and blue signifies a value greater than or equal to the maximum in positive potential. Benzene, pyridine, and 2-chloropyridine were also optimized at the HF/6-31G level of theory and their electrostatic potential surfaces were calculated.

HF/6-31G geometries in Gaussian "standard" orientation

Amide

1	6	-1.672610	-1.559835	-0.325739
2	6	-2.535131	-0.633875	0.500882
3	8	-2.395842	-0.522298	1.716081
4	6	-0.861909	-0.847868	-1.432500
5	6	0.156787	0.121171	-0.910069
6	6	0.043749	1.468676	-0.836109
7	6	1.459625	-0.210436	-0.366618
8	7	1.187643	2.012479	-0.282106
9	6	2.067982	0.995913	0.014298
10	6	2.152177	-1.410650	-0.181714
11	6	3.342474	1.037026	0.575254
12	6	3.414351	-1.377688	0.372266
13	6	4.003356	-0.160517	0.748996
14	7	-3.493141	0.058494	-0.164532
15	6	-4.400096	0.958787	0.527433
16	1	-1.005191	-2.057000	0.362004
17	1	-1.534637	-0.336256	-2.113443
18	1	-0.365956	-1.611541	-2.023592
19	1	-0.768358	2.094771	-1.132976
20	1	1.335283	2.974018	-0.105934
21	1	1.711128	-2.347795	-0.464722
22	1	3.795115	1.966472	0.863835
23	1	3.956696	-2.291270	0.521789
24	1	4.985975	-0.164045	1.180275
25	1	-5.064878	0.418005	1.188457
26	1	-4.988565	1.492343	-0.205053
27	1	-3.844002	1.669482	1.121073
28	1	-3.622814	-0.093243	-1.134482
29	1	-2.309104	-2.316582	-0.777692

Ester

1	6	-1.684484	-1.577555	-0.415491
2	6	-2.513618	-0.710739	0.483565
3	8	-2.574518	-0.807227	1.691013
4	6	-0.901484	-0.793091	-1.485381
5	6	0.117099	0.150216	-0.916329
6	6	0.002745	1.494814	-0.797146
7	6	1.416704	-0.198913	-0.374618
8	7	1.142023	2.019965	-0.214704
9	6	2.021597	0.994238	0.051892
10	6	2.109995	-1.404030	-0.227304
11	6	3.292124	1.016866	0.622919
12	6	3.368284	-1.389279	0.336437
13	6	3.953194	-0.185476	0.759574
14	8	-3.251352	0.194726	-0.194077
15	6	-4.133045	1.079798	0.532457
16	1	-1.019904	-2.148157	0.216624
17	1	-1.604169	-0.247297	-2.100393
18	1	-0.410745	-1.513493	-2.132575
19	1	-0.807215	2.127710	-1.084779
20	1	1.295025	2.977060	-0.019780

21	1	1.673123	-2.330721	-0.548593
22	1	3.742287	1.936207	0.945832
23	1	3.911239	-2.306844	0.456458
24	1	4.933123	-0.202998	1.196600
25	1	-4.867965	0.507497	1.075701
26	1	-4.601874	1.690490	-0.218195
27	1	-3.565465	1.683551	1.222545
28	1	-2.359340	-2.273847	-0.905465

ACh

1	6	-2.718473	0.518649	1.224882
2	7	-1.843253	0.092946	0.069036
3	6	-1.528263	1.320023	-0.762562
4	6	-2.595107	-0.912825	-0.763243
5	6	-0.577083	-0.518292	0.649577
6	6	0.379345	-1.123445	-0.385577
7	8	1.729726	-0.989025	0.084883
8	6	2.357284	0.216056	-0.044646
9	6	3.788800	0.163141	0.347961
10	8	1.738011	1.182690	-0.440701
11	1	-3.616582	0.970293	0.836081
12	1	-2.181137	1.233680	1.826480
13	1	-2.970183	-0.346391	1.816842
14	1	-0.953317	1.043954	-1.627204
15	1	-0.953596	2.009343	-0.168425
16	1	-2.460246	1.766826	-1.068509
17	1	-2.007982	-1.184936	-1.624273
18	1	-3.522545	-0.471393	-1.090821
19	1	-2.799401	-1.785877	-0.163632
20	1	-0.080435	0.272683	1.185252
21	1	-0.892425	-1.269461	1.358028
22	1	0.298830	-0.650257	-1.349922
23	1	0.234815	-2.183758	-0.493710
24	1	4.322987	-0.505403	-0.314948
25	1	4.217805	1.149867	0.292299
26	1	3.885268	-0.228958	1.351406

ACh + Amide

1	6	-2.001453	2.822840	0.622837
2	6	-0.512980	2.838647	0.863266
3	8	0.069291	1.899914	1.437360
4	6	-2.373196	2.420873	-0.829046
5	6	-1.940379	1.037796	-1.214315
6	6	-0.944368	0.706009	-2.073148
7	6	-2.508951	-0.212142	-0.748897
8	7	-0.835898	-0.666162	-2.173160
9	6	-1.790697	-1.247395	-1.372125
10	6	-3.557261	-0.537930	0.119870
11	6	-2.080370	-2.593233	-1.143625
12	6	-3.848330	-1.866269	0.351800
13	6	-3.113264	-2.886442	-0.276496
14	7	0.165232	3.902236	0.412248
15	6	1.606204	4.042198	0.571203

16	6	-0.448109	-1.191442	2.280381
17	7	1.060926	-1.316178	2.187070
18	6	1.449008	-2.730190	2.506296
19	6	1.668371	-0.364948	3.187891
20	6	1.462283	-0.932816	0.769080
21	6	2.923414	-1.229022	0.435725
22	8	3.336556	-0.417320	-0.683866
23	6	2.966246	-0.728725	-1.950039
24	6	3.659958	0.112104	-2.963919
25	8	2.152168	-1.602445	-2.172197
26	1	-2.435211	2.124016	1.322113
27	1	-1.949005	3.140357	-1.521487
28	1	-3.450425	2.510512	-0.922434
29	1	-0.304130	1.345927	-2.639875
30	1	-0.150935	-1.150789	-2.702028
31	1	-4.141237	0.230265	0.590525
32	1	-1.537949	-3.371833	-1.645245
33	1	-4.658177	-2.129854	1.003866
34	1	-3.375692	-3.910597	-0.093659
35	1	1.879918	4.046488	1.616841
36	1	1.913294	4.975347	0.125528
37	1	2.127829	3.231306	0.082366
38	1	-0.742655	-1.414353	3.293712
39	1	-0.900595	-1.892432	1.599473
40	1	-0.715029	-0.183123	2.017774
41	1	2.522352	-2.820616	2.539634
42	1	1.045266	-3.385746	1.750809
43	1	1.041805	-2.994421	3.469131
44	1	2.742111	-0.445973	3.164055
45	1	1.313895	-0.631431	4.171166
46	1	1.356000	0.634577	2.934646
47	1	0.812633	-1.474662	0.102571
48	1	1.258551	0.121154	0.674842
49	1	3.078007	-2.267116	0.186887
50	1	3.602125	-0.939244	1.218401
51	1	4.723755	-0.085541	-2.931420
52	1	3.281579	-0.112664	-3.947572
53	1	3.517901	1.160268	-2.738481
54	1	-0.333138	4.647774	-0.011942
55	1	-2.415540	3.802446	0.835890

ACh + Ester

1	6	-2.216953	2.611972	0.850796
2	6	-0.739758	2.700577	1.067618
3	8	-0.060985	1.874431	1.668000
4	6	-2.566870	2.322417	-0.630864
5	6	-2.018623	1.019868	-1.134706
6	6	-1.001952	0.853271	-2.017884
7	6	-2.466888	-0.311939	-0.775590
8	7	-0.768235	-0.490844	-2.230281
9	6	-1.661115	-1.221760	-1.482120
10	6	-3.477730	-0.803454	0.059210
11	6	-1.826214	-2.602967	-1.368504

12	6	-3.646638	-2.167601	0.177000
13	6	-2.824848	-3.060521	-0.532716
14	8	-0.199884	3.786118	0.514631
15	6	1.215084	4.077882	0.641056
16	6	-0.377080	-1.404677	2.204030
17	7	1.136556	-1.433058	2.110914
18	6	1.602170	-2.851573	2.274985
19	6	1.687733	-0.568640	3.217034
20	6	1.521183	-0.875664	0.747564
21	6	2.999029	-1.041778	0.398343
22	8	3.347689	-0.108699	-0.645193
23	6	3.021677	-0.354707	-1.939843
24	6	3.654552	0.614324	-2.875776
25	8	2.287345	-1.274282	-2.238558
26	1	-2.603365	1.834623	1.492756
27	1	-2.200701	3.139074	-1.239914
28	1	-3.648022	2.330710	-0.716608
29	1	-0.431123	1.595225	-2.532035
30	1	-0.056625	-0.867649	-2.809066
31	1	-4.128095	-0.133454	0.589027
32	1	-1.218669	-3.283989	-1.933441
33	1	-4.427997	-2.557403	0.799937
34	1	-2.994005	-4.115930	-0.439297
35	1	1.351379	5.013325	0.131466
36	1	1.796212	3.297334	0.176126
37	1	1.479589	4.166053	1.682223
38	1	-0.662183	-1.757635	3.182361
39	1	-0.788425	-2.046137	1.443279
40	1	-0.703811	-0.390620	2.057518
41	1	2.678689	-2.885066	2.308840
42	1	1.238496	-3.439325	1.446934
43	1	1.206483	-3.243341	3.198296
44	1	2.764233	-0.587409	3.195095
45	1	1.344847	-0.960762	4.161434
46	1	1.323527	0.435560	3.076819
47	1	0.909431	-1.379727	0.018177
48	1	1.257351	0.168592	0.762861
49	1	3.223773	-2.041175	0.061064
50	1	3.656826	-0.774407	1.206546
51	1	3.312425	0.429739	-3.880812
52	1	3.417548	1.627493	-2.580725
53	1	4.730953	0.507644	-2.834431
54	1	-2.660864	3.555243	1.142304

(-) Nic

1	6	-0.712854	-0.130382	0.550103
2	6	-1.437896	-1.477457	0.545398
3	7	-1.509071	0.667310	-0.507848
4	6	-2.920618	-1.103876	0.423697
5	6	-2.949690	0.151501	-0.465309
6	6	-1.390600	2.154553	-0.364601
7	6	0.762267	-0.115508	0.255660
8	6	1.291247	-0.757533	-0.860797

9	7	2.587048	-0.758045	-1.142551
10	6	3.436256	-0.138862	-0.322900
11	6	3.007609	0.516855	0.821292
12	6	1.652571	0.531007	1.109595
13	1	-0.908084	0.381042	1.483350
14	1	-1.116842	-2.084752	-0.291641
15	1	-1.220092	-2.030946	1.447164
16	1	-3.509022	-1.901006	-0.005890
17	1	-3.336920	-0.882167	1.397449
18	1	-3.254461	-0.053828	-1.479524
19	1	-3.573021	0.935663	-0.068189
20	1	-1.940867	2.631506	-1.160208
21	1	-0.351088	2.433188	-0.420935
22	1	-1.801479	2.444421	0.589163
23	1	0.673223	-1.296754	-1.552893
24	1	4.470183	-0.176324	-0.592700
25	1	3.712962	0.999955	1.465568
26	1	1.297390	1.029364	1.992070
27	1	-1.117474	0.424003	-1.402811

(-) Nic + Amide

1	6	1.170807	2.197630	-1.633995
2	6	0.330009	2.342097	-0.390984
3	8	-0.319620	1.382385	0.078505
4	6	2.692861	2.245862	-1.347496
5	6	3.175946	1.171739	-0.419353
6	6	3.626688	1.335928	0.848535
7	6	3.276774	-0.246111	-0.709325
8	7	3.990320	0.116976	1.386047
9	6	3.791696	-0.866309	0.441443
10	6	3.007438	-1.026385	-1.840345
11	6	4.029184	-2.238990	0.500863
12	6	3.240951	-2.385722	-1.788234
13	6	3.745687	-2.986475	-0.622731
14	7	0.303769	3.537009	0.197842
15	6	-0.467756	3.815695	1.404019
16	6	-2.432480	-1.335717	0.955781
17	6	-2.188759	-0.682137	2.321591
18	7	-1.059152	-1.222831	0.311466
19	6	-0.752721	-1.095655	2.712607
20	6	-0.082969	-1.572687	1.410145
21	6	-0.879552	-2.024067	-0.929402
22	6	-3.542340	-0.771727	0.108575
23	6	-3.579767	0.571463	-0.257378
24	7	-4.563424	1.089477	-0.984692
25	6	-5.569791	0.308469	-1.377500
26	6	-5.623692	-1.038451	-1.058503
27	6	-4.592787	-1.583091	-0.308699
28	1	0.909621	1.255651	-2.091573
29	1	2.956760	3.217800	-0.944683
30	1	3.203232	2.167962	-2.301743
31	1	3.731132	2.233939	1.416635
32	1	4.415092	-0.010483	2.270941

33	1	2.650046	-0.576162	-2.747482
34	1	4.435122	-2.697422	1.382088
35	1	3.059520	-2.992906	-2.653909
36	1	3.930731	-4.042998	-0.615086
37	1	-0.249774	4.821805	1.725407
38	1	-0.197938	3.128613	2.192113
39	1	-1.528339	3.731398	1.213539
40	1	-2.591669	-2.399121	1.088291
41	1	-2.265588	0.392403	2.233505
42	1	-2.928043	-1.007854	3.038924
43	1	-0.214982	-0.263326	3.141704
44	1	-0.757018	-1.892570	3.443227
45	1	0.853586	-1.087209	1.194335
46	1	0.060408	-2.642798	1.390203
47	1	0.136790	-1.914433	-1.272301
48	1	-1.562543	-1.668710	-1.683530
49	1	-1.079716	-3.062765	-0.714389
50	1	-2.801529	1.248761	0.026658
51	1	-6.339054	0.777114	-1.954347
52	1	-6.445656	-1.641546	-1.385864
53	1	-4.612490	-2.624552	-0.047443
54	1	0.800893	4.288048	-0.218946
55	1	-0.876971	-0.239651	0.087169
56	1	0.912169	2.985357	-2.333818

(-) Nic + Ester

1	6	1.235743	2.213994	-1.620128
2	6	0.418624	2.288837	-0.370020
3	8	-0.318205	1.393361	0.044434
4	6	2.753884	2.280611	-1.328127
5	6	3.232976	1.206376	-0.397012
6	6	3.707781	1.381556	0.861129
7	6	3.288133	-0.219604	-0.659574
8	7	4.045180	0.163466	1.416903
9	6	3.800712	-0.832040	0.497245
10	6	2.978444	-1.015189	-1.769767
11	6	3.993420	-2.210540	0.583039
12	6	3.167022	-2.380323	-1.691005
13	6	3.668194	-2.972576	-0.519588
14	8	0.555184	3.434248	0.277606
15	6	-0.172371	3.723639	1.503314
16	6	-2.471623	-1.448434	0.771400
17	6	-2.092868	-1.044455	2.201379
18	7	-1.147247	-1.268662	0.039634
19	6	-0.651044	-1.562384	2.396846
20	6	-0.100324	-1.821615	0.981917
21	6	-1.097263	-1.863664	-1.324739
22	6	-3.615847	-0.713965	0.124813
23	6	-3.628495	0.673788	0.011447
24	7	-4.638453	1.340018	-0.536011
25	6	-5.697466	0.668887	-0.988968
26	6	-5.779727	-0.712137	-0.915375
27	6	-4.721751	-1.409800	-0.353853

28	1	0.981769	1.296509	-2.129042
29	1	2.985121	3.253485	-0.914921
30	1	3.270781	2.213114	-2.279467
31	1	3.844969	2.288257	1.407987
32	1	4.476283	0.040631	2.299439
33	1	2.629538	-0.571938	-2.683675
34	1	4.398462	-2.663005	1.467735
35	1	2.957236	-2.999184	-2.542042
36	1	3.820197	-4.034071	-0.491691
37	1	0.159667	4.702438	1.793514
38	1	0.074113	2.991938	2.255018
39	1	-1.231695	3.721117	1.305674
40	1	-2.665018	-2.513594	0.736619
41	1	-2.131368	0.031376	2.303638
42	1	-2.788090	-1.464769	2.913475
43	1	-0.046492	-0.837785	2.922118
44	1	-0.636340	-2.475962	2.974606
45	1	0.835054	-1.329855	0.775072
46	1	0.003398	-2.874664	0.768404
47	1	-0.102096	-1.739939	-1.720746
48	1	-1.812221	-1.364122	-1.957588
49	1	-1.333558	-2.914789	-1.262065
50	1	-2.809658	1.270613	0.355841
51	1	-6.485770	1.253829	-1.413687
52	1	-6.643978	-1.224894	-1.284902
53	1	-4.763104	-2.480603	-0.283658
54	1	-0.953230	-0.270959	-0.046900
55	1	0.956920	3.046424	-2.254993

(+) Epi

1	6	-2.034438	-0.697990	1.247918
2	6	-1.259774	0.493911	0.592381
3	6	-2.003218	0.665175	-0.758227
4	6	-3.433946	1.183306	-0.542611
5	6	-4.176585	-0.036259	0.080623
6	6	-3.074988	-1.099249	0.190876
7	7	-2.290056	-0.803437	-1.101307
8	1	-1.451990	-1.352286	-1.203916
9	6	0.235807	0.301982	0.414421
10	6	1.078384	1.420087	0.412840
11	6	2.433818	1.272074	0.203460
12	6	2.915870	-0.012382	-0.007642
13	7	2.147455	-1.070960	-0.000678
14	6	0.837877	-0.929656	0.213879
15	1	-2.525455	-0.399555	2.162315
16	1	-1.411613	1.394827	1.170487
17	1	-1.436919	1.148441	-1.535782
18	1	-3.433189	2.047988	0.104320
19	1	-3.883999	1.484612	-1.479328
20	1	-4.596992	0.179065	1.051381
21	1	-4.986549	-0.373222	-0.553166
22	1	-3.407446	-2.122906	0.223249
23	1	0.676818	2.400820	0.584075
24	1	3.097914	2.109832	0.204955

25	1	0.297741	-1.854068	0.240635
26	1	-2.844084	-0.916377	-1.934579
27	1	-1.382083	-1.520395	1.498234
28	17	4.657248	-0.263059	-0.300381

(+) Epi + Amide

1	6	1.175433	2.791865	-1.333859
2	6	0.281266	2.549269	-0.145378
3	8	0.047025	1.390432	0.263418
4	6	2.652486	3.051514	-0.940496
5	6	3.304845	1.940498	-0.172871
6	6	3.772988	2.006737	1.096947
7	6	3.594531	0.599562	-0.643910
8	7	4.327556	0.797205	1.466511
9	6	4.236348	-0.077967	0.409473
10	6	3.391816	-0.070189	-1.857548
11	6	4.671583	-1.397039	0.287527
12	6	3.813736	-1.380761	-1.982527
13	6	4.451411	-2.037618	-0.915578
14	7	-0.238681	3.611322	0.463323
15	6	-1.141789	3.517030	1.608197
16	6	-0.776050	-1.935110	1.812657
17	6	-1.580117	-2.292797	0.519199
18	6	-0.467485	-2.275437	-0.570383
19	6	0.365393	-3.562850	-0.454675
20	6	1.096887	-3.377496	0.910411
21	6	0.682298	-1.958422	1.342125
22	7	0.533298	-1.278701	-0.008764
23	1	0.232366	-0.299382	0.056525
24	6	-2.793973	-1.433662	0.228315
25	6	-3.980416	-2.022174	-0.213401
26	6	-5.086728	-1.240454	-0.492932
27	6	-4.961513	0.126132	-0.316634
28	7	-3.860507	0.700766	0.099811
29	6	-2.793589	-0.054561	0.373236
30	1	1.108735	1.921329	-1.969363
31	1	2.711516	3.967775	-0.363171
32	1	3.198501	3.236839	-1.860018
33	1	3.764748	2.830696	1.775718
34	1	4.793695	0.623833	2.322483
35	1	2.943647	0.430250	-2.694816
36	1	5.182112	-1.892008	1.091242
37	1	3.690267	-1.894584	-2.916258
38	1	4.795794	-3.044747	-1.049504
39	1	-2.094415	3.098423	1.315139
40	1	-1.299574	4.509160	2.000466
41	1	-0.704809	2.900204	2.378686
42	1	-0.948676	-2.645981	2.607575
43	1	-1.929595	-3.313972	0.591497
44	1	-0.794587	-2.007579	-1.560144
45	1	-0.258579	-4.443461	-0.485477
46	1	1.074307	-3.636331	-1.269654
47	1	0.793655	-4.106389	1.648092
48	1	2.168472	-3.452695	0.794559

49	1	1.381771	-1.435767	1.970926
50	1	-4.046255	-3.086830	-0.333430
51	1	-6.007612	-1.667241	-0.828392
52	1	-1.934263	0.484997	0.701858
53	1	1.396198	-1.268919	-0.531708
54	1	-0.059178	4.513043	0.089592
55	1	0.810237	3.638146	-1.903731
56	1	-1.038492	-0.955576	2.185911
57	17	-6.358471	1.191926	-0.665028

(+) Epi + Ester

1	6	1.105543	2.713065	-1.444972
2	6	0.232095	2.459800	-0.258205
3	8	-0.039829	1.340919	0.177884
4	6	2.553856	3.077905	-1.030441
5	6	3.232795	2.035859	-0.191130
6	6	3.670526	2.189592	1.082527
7	6	3.569186	0.677862	-0.575037
8	7	4.248960	1.020044	1.534479
9	6	4.204131	0.081651	0.530582
10	6	3.415727	-0.066932	-1.751735
11	6	4.677437	-1.229512	0.496861
12	6	3.877503	-1.369883	-1.789976
13	6	4.506155	-1.944838	-0.671348
14	8	-0.217435	3.562999	0.310669
15	6	-1.143644	3.535838	1.438158
16	6	-0.773128	-2.203504	1.741051
17	6	-1.601811	-2.421578	0.431968
18	6	-0.501777	-2.344687	-0.666749
19	6	0.329730	-3.637836	-0.650007
20	6	1.103504	-3.537208	0.700025
21	6	0.677775	-2.163122	1.247317
22	7	0.503680	-1.384524	-0.049212
23	1	0.184267	-0.422007	0.071376
24	6	-2.779646	-1.490034	0.219960
25	6	-3.940264	-1.961511	-0.398842
26	6	-5.008540	-1.112841	-0.621203
27	6	-4.874645	0.202621	-0.210461
28	7	-3.799614	0.665562	0.376865
29	6	-2.771424	-0.157179	0.598442
30	1	1.101921	1.823633	-2.057197
31	1	2.535322	4.020257	-0.497035
32	1	3.114179	3.249385	-1.943574
33	1	3.625097	3.052345	1.709765
34	1	4.698698	0.910014	2.409561
35	1	2.980551	0.372900	-2.628997
36	1	5.181754	-1.661625	1.339745
37	1	3.797945	-1.938094	-2.696818
38	1	4.884918	-2.946364	-0.738776
39	1	-2.057546	3.047341	1.141769
40	1	-1.316722	4.569568	1.669901
41	1	-0.687399	3.023970	2.269462
42	1	-0.922627	-3.003616	2.451193
43	1	-1.990439	-3.431006	0.419582

44	1	-0.839047	-2.009522	-1.632364
45	1	-0.300004	-4.512693	-0.712718
46	1	1.012002	-3.667187	-1.489949
47	1	0.839954	-4.323647	1.392090
48	1	2.171886	-3.579757	0.544856
49	1	1.381804	-1.682802	1.904230
50	1	-4.014736	-2.989048	-0.700598
51	1	-5.908700	-1.451039	-1.088149
52	1	-1.933837	0.297658	1.074615
53	1	1.361657	-1.329209	-0.579134
54	1	0.684320	3.528439	-2.015835
55	1	-1.037182	-1.280851	2.236190
56	17	-6.219270	1.352522	-0.479838

(-) Epi

1	6	1.251069	0.515279	-0.510780
2	6	1.910293	1.186578	0.751372
3	6	3.043482	0.227408	1.136100
4	6	4.169413	0.256651	0.094328
5	6	3.519915	-0.411416	-1.153218
6	6	2.098885	-0.761529	-0.681940
7	7	2.381962	-1.118335	0.789096
8	1	1.546816	-1.296234	1.323485
9	6	-0.243698	0.303147	-0.348963
10	6	-1.096318	1.413960	-0.375691
11	6	-2.455211	1.251267	-0.212310
12	6	-2.931658	-0.039458	-0.020076
13	7	-2.152933	-1.088494	0.000766
14	6	-0.836660	-0.932724	-0.165350
15	1	1.407095	1.135403	-1.383392
16	1	2.300368	2.171382	0.540243
17	1	1.191842	1.284734	1.553811
18	1	3.352329	0.258840	2.167197
19	1	4.488941	1.271223	-0.091088
20	1	5.035447	-0.295822	0.434605
21	1	3.486755	0.242622	-2.011598
22	1	4.058027	-1.302029	-1.450184
23	1	1.631822	-1.575678	-1.206643
24	1	-0.698945	2.399171	-0.530846
25	1	-3.127649	2.082079	-0.234255
26	1	-0.285268	-1.850475	-0.159817
27	1	3.001552	-1.905213	0.894973
28	17	-4.680554	-0.307482	0.208450

(-) Epi + Amide

1	6	-2.312078	-2.569977	-1.752684
2	6	-0.823997	-2.331239	-1.717673
3	8	-0.359067	-1.176182	-1.588889
4	6	-2.877222	-3.022534	-0.381189
5	6	-2.599653	-2.074718	0.747521
6	6	-1.814817	-2.313736	1.826679
7	6	-3.124752	-0.733610	0.915554
8	7	-1.798400	-1.212873	2.658843
9	6	-2.600137	-0.233527	2.122086

10	6	-3.998255	0.068749	0.170260
11	6	-2.913232	1.040208	2.595612
12	6	-4.308185	1.334888	0.631151
13	6	-3.768642	1.814848	1.837933
14	7	-0.028504	-3.392462	-1.812197
15	6	1.431504	-3.316118	-1.767757
16	6	1.345565	2.390701	-0.729186
17	6	0.894322	2.143978	-2.207775
18	6	-0.636730	2.189740	-2.124779
19	6	-1.132037	3.586130	-1.708479
20	6	-0.762733	3.657534	-0.194642
21	6	-0.007903	2.337160	0.035174
22	7	-0.845741	1.413856	-0.835780
23	1	-0.548596	0.439581	-0.929476
24	6	2.425683	1.472575	-0.196402
25	6	3.533948	1.999754	0.467901
26	6	4.507439	1.159912	0.979265
27	6	4.328468	-0.200482	0.806473
28	7	3.299153	-0.718512	0.182655
29	6	2.363242	0.092168	-0.317985
30	1	-2.778632	-1.646338	-2.060763
31	1	-2.476270	-3.999371	-0.133944
32	1	-3.947291	-3.156231	-0.502366
33	1	-1.263318	-3.192657	2.077660
34	1	-1.337457	-1.169520	3.533922
35	1	-4.448691	-0.298805	-0.732236
36	1	-2.524026	1.399020	3.529040
37	1	-4.999203	1.947731	0.085167
38	1	-4.049789	2.789253	2.187818
39	1	1.799881	-2.663982	-2.546048
40	1	1.827047	-4.306668	-1.926771
41	1	1.777605	-2.948997	-0.813618
42	1	1.718981	3.401577	-0.637320
43	1	1.279907	2.894323	-2.882492
44	1	1.221282	1.177141	-2.563815
45	1	-1.153376	1.732391	-2.950753
46	1	-0.650198	4.355947	-2.293634
47	1	-2.198981	3.685087	-1.858108
48	1	-0.148663	4.511089	0.050658
49	1	-1.649174	3.702896	0.423466
50	1	0.043428	1.991502	1.052310
51	1	3.642773	3.061215	0.583771
52	1	5.367787	1.539151	1.487921
53	1	1.562474	-0.402768	-0.822280
54	1	-1.809179	1.387962	-0.537010
55	1	-0.440008	-4.286534	-1.940333
56	1	-2.549192	-3.320464	-2.498085
57	17	5.548181	-1.341092	1.455968

(-) Epi + Ester

1	6	-2.360309	-2.429270	-1.903391
2	6	-0.886736	-2.224119	-1.752187
3	8	-0.347641	-1.118973	-1.667458
4	6	-2.997613	-2.953216	-0.591844

5	6	-2.715226	-2.091659	0.604176
6	6	-1.991053	-2.444302	1.695153
7	6	-3.154748	-0.728221	0.832154
8	7	-1.933081	-1.395072	2.590291
9	6	-2.644886	-0.333984	2.083762
10	6	-3.945415	0.172394	0.106867
11	6	-2.890258	0.929524	2.620366
12	6	-4.189170	1.428667	0.631359
13	6	-3.664261	1.802008	1.881756
14	8	-0.202578	-3.350528	-1.693060
15	6	1.250526	-3.383418	-1.540141
16	6	1.444082	2.453390	-0.626975
17	6	0.982539	2.323771	-2.117393
18	6	-0.548508	2.364821	-2.025068
19	6	-1.038670	3.733781	-1.523073
20	6	-0.647869	3.719675	-0.013420
21	6	0.092472	2.380319	0.138554
22	7	-0.757019	1.514035	-0.780984
23	1	-0.459500	0.551407	-0.922769
24	6	2.493416	1.465271	-0.160139
25	6	3.570974	1.896990	0.615650
26	6	4.508830	0.990383	1.075709
27	6	4.327069	-0.339918	0.741503
28	7	3.327799	-0.767383	0.009726
29	6	2.428862	0.109829	-0.444203
30	1	-2.799579	-1.486426	-2.192641
31	1	-2.635589	-3.956633	-0.408568
32	1	-4.066626	-3.032020	-0.758599
33	1	-1.512386	-3.373404	1.912829
34	1	-1.509224	-1.434264	3.484193
35	1	-4.387963	-0.113678	-0.828399
36	1	-2.514417	1.207237	3.586330
37	1	-4.822283	2.114669	0.102152
38	1	-3.895992	2.770747	2.280404
39	1	1.711199	-2.868433	-2.366595
40	1	1.498303	-4.427671	-1.552229
41	1	1.535198	-2.926223	-0.608387
42	1	1.847085	3.443667	-0.464157
43	1	1.363076	3.126407	-2.731973
44	1	1.310893	1.391917	-2.555654
45	1	-1.070609	1.959145	-2.874372
46	1	-0.566188	4.536007	-2.070886
47	1	-2.107729	3.838765	-1.651509
48	1	-0.018229	4.550676	0.267283
49	1	-1.524838	3.744841	0.618794
50	1	0.141431	1.977290	1.134639
51	1	3.683160	2.936646	0.857499
52	1	5.345295	1.296836	1.666612
53	1	1.653386	-0.314825	-1.042116
54	1	-1.720570	1.474515	-0.481163
55	1	-2.531728	-3.147612	-2.693870
56	17	5.496942	-1.564761	1.321705

References

(1) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A., *Proc. Natl. Acad. Sci. USA* **1998,** 95, 12088-12093.

(2) Beene, D. L.; Brandt, G. S.; Zhong, W.; Zacharias, N. M.; Lester, H. A.; Dougherty, D. A., *Biochem.* **2002,** 41, 10262-10269.

(3) Miyazawa, A.; Fujiyoshi, Y.; Stowell, M.; Unwin, N., *J. Mol. Biol.* **1999,** 288, 765-786.

(4) Schapira, M.; Abagyan, R.; Totrov, M., *BMC Struct. Biol.* **2002,** 2, 1-14.

(5) Le Novere, N.; Grutter, T.; Changeux, J. P., *Proc. Natl. Acad. Sci. USA* **2002,** 99, 3210-3215.

(6) Celie, P. H. N.; van Rossum-Fikkert, S. E.; van Dijk, W. J.; Brejc, K.; Smit, A. B.; Sixma, T. K., *Neuron* **2004,** 41, 907-914.

(7) Akk, G.; Auerbach, A., *Br. J. Pharmacol.* **1999,** 128, 1467-1476.

(8) Gerzanich, V.; Peng, X.; Wang, F.; Wells, G.; Anand, R.; Fletcher, S.; Lindstrom, J., *Mol. Pharmacol.* **1995,** 48, 774-782.

(9) Prince, R. J.; Sine, S. M., *Biophys. J.* **1998,** 75, 1817-1827.

(10) Badio, B.; Daly, J. W., *Mol. Pharmacol.* **1994,** 45, 563-569.

(11) Dukat, M.; Glennon, R. A., *Cell Mol. Neurobiol.* **2003,** 23, 365-378.

(12) Dougherty, D. A., *Curr. Opin. Chem. Biol.* **2000,** 4, 645-652.

(13) Kearney, P. C.; Nowak, M. W.; Zhong, W.; Silverman, S. K.; Lester, H. A.; Dougherty, D. A., *Mol. Pharmacol.* **1996,** 50, 1401-1412.

(14) Beene, D. L.; Dougherty, D. A.; Lester, H. A., *Curr. Opin. Neurobiol.* **2003,** 13, 264-270.

(15) Koh, J. T.; Cornish, V. W.; Schultz, P. G., *Biochem.* **1997,** 36, 11314-11322.

(16) Elmore, D. E.; Dougherty, D. A., *J. Org. Chem.* **2000,** 65, 742-747 Vistoli, G.; Pedretti, A.; Villa, L.; Testa, B., *J. Am. Chem. Soc.* **2002,** 124, 7472-7480 Partington, P.; Feeney, J.; Burgen, A. S., *Mol. Pharmacol.* **1972,** 8, 269-277.

(17) Siu, F. M.; Ma, N. L.; Tsang, C. W., *Chem. Eur. J.* **2004,** 10, 1966-1976.

(18) Graton, J.; Berthelot, M.; Gal, J. F.; Laurence, C.; Lebreton, J.; Le Questel, J. Y.; Maria, P. C.; Richard, R., *J. Org. Chem.* **2003,** 68, 8208-8221 Graton, J.; van Mourik, T.; Price, S. L., *J. Am. Chem. Soc.* **2003,** 125, 5988-5997.

(19) Spande, T. F.; Garraffo, H. M.; Edwards, M. W.; Yeh, H. J. C.; Pannell, L.; Daly, J. W., *J. Am. Chem. Soc.* **1992,** 114, 3475-3478.

(20) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J., *Chem. Phys. Lett.* **1996,** 255, 327-335.

(21) Cannizzaro, C. E.; Houk, K. N., *J. Am. Chem. Soc.* **2002,** 124, 7163-7169.

(22) Petersson, E. J.; Choi, A.; Dahan, D. S.; Lester, H. A.; Dougherty, D. A., *J. Am. Chem. Soc.* **2002,** 124, 12662-12663.

(23) Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van Der Oost, J.; Smit, A. B.; Sixma, T. K., *Nature* **2001,** 411, 269-276.

(24) Spang, J. E.; Bertrand, S.; Westera, G.; Patt, J. T.; Schubiger, P. A.; Bertrand, D., *Chem. Biol.* **2000,** 7, 545-555.

(25) Amer Chemical Soc Cashin, A. L.; Petersson, E. J.; Lester, H. A.; Dougherty, D. A., *J. Am. Chem. Soc*. **2005,** 127, 350-356.

(26) Boys, S. F.; Bernardi, F., *Mol. Phys*. **1970,** 19, 553-557.

(27) Foresman, J. B.; Frisch, E., *Exploring Chemistry with Electronic Structure Methods*; Gaussian, Inc.: Pittsburgh, PA, 1996.

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; J. A. Montgomery, J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; *et. al. Gaussian 98 (Revision A.9)*, Gaussian, Inc.: Pittsburgh PA, 1998.

(29) Flükiger, P.; Lüthi, H. P.; Portmann, S.; Weber, J. *MOLEKEL 4.0*, Swiss Center for Scientific Computing: Manno, Switzerland, 2000.

Section 1: Chapter 5

Computational Modeling of Nicotine and Carbamoyl Choline Binding to the α7 Nicotinic Acetylcholine Receptor

ab initio Modeling of Trp 149 Fluorination Effects on Nicotine Binding

Unnatural amino acid mutagenesis studies of the muscle-type nicotinic acetylcholine receptor (nAChR) show that a cation-π interaction with tryptophan 149 of the α subunit is essential to determining acetylcholine (ACh) potency, but that nicotine (Nic) potency is not regulated by the cation-π interaction.[1] Further studies have shown that a hydrogen bond to the backbone carbonyl of Trp 149 is important to Nic binding, and that epibatidine (Epi), a high affinity Nic analog, relys on both the hydrogen bond and the cation-π interaction in its binding.[2] The reason that we dismissed the importance of the cation-π interaction in Nic binding was that our measurements of the changes in the Nic EC_{50} with Trp fluorination did not produce a linear relationship with calculated cation-π binding energies. Our discussion of the shape of the Nic F-Trp curve interpreted the nonlinearity as a consequence of a change in the mechanism of Nic binding, which meant that the cation-π interaction no longer played a significant role in binding Nic once the interaction with Trp 149 had been weakened through fluorination. Figure 1 shows a hypothetical cartoon depicting this scenario.

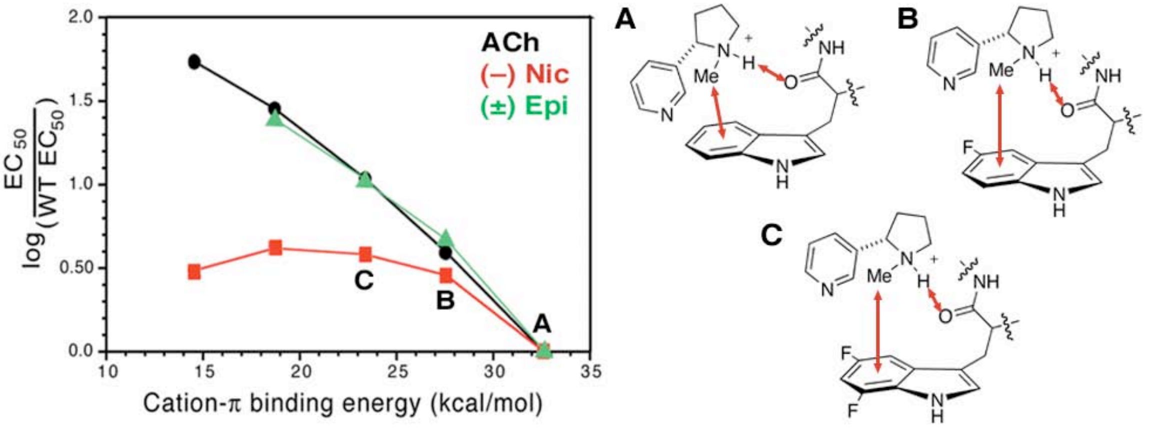


Figure 1. Change in Nic Binding Mechanism in Binding to F-Trp. Left: Fluorination plot showing ACh, Nic, and Epi data described in Chapters 2-4. Right: Hypothetical depiction of alteration in Nic binding mechanism.

The measure of cation-π binding in our previous fluorination plots, the energy of Na^+ binding to an indole, represents an idealized interaction.[3] If the actual binding geometry of the ligand differs from the alignment in the sodium calculations, directly above the six-membered ring of Trp, then fluorination may have different effects. Calculations performed on the actual ligand, nicotine, in a sufficiently realistic model of the binding site environment should yield a linear relationship with the observed changes in potency. These more elaborate calculations should take into account the other elements of the binding site that permit this change in binding mechanism. To this end, we sought to develop more complex models of ligand binding employing the actual ligand and a representation of the nAChR binding site.

We began with the indole complexes used in Chapter 4 to explore the effects of ester conversion of the Trp149/Thr150 backbone amide on ligand binding. We performed *in silico* fluorination of the indole moieties in these complexes, mirroring the introduction of fluorinated tryptophans into the nAChR. In each case, the geometry was optimized and the binding of the ligand was calculated at the HF/6-31G** level. The results of these calculations are plotted in Figure 2. The logarithm of the ratios of the unnatural mutant EC_{50}s to the wild type EC_{50} are again plotted against the calculated binding energy.

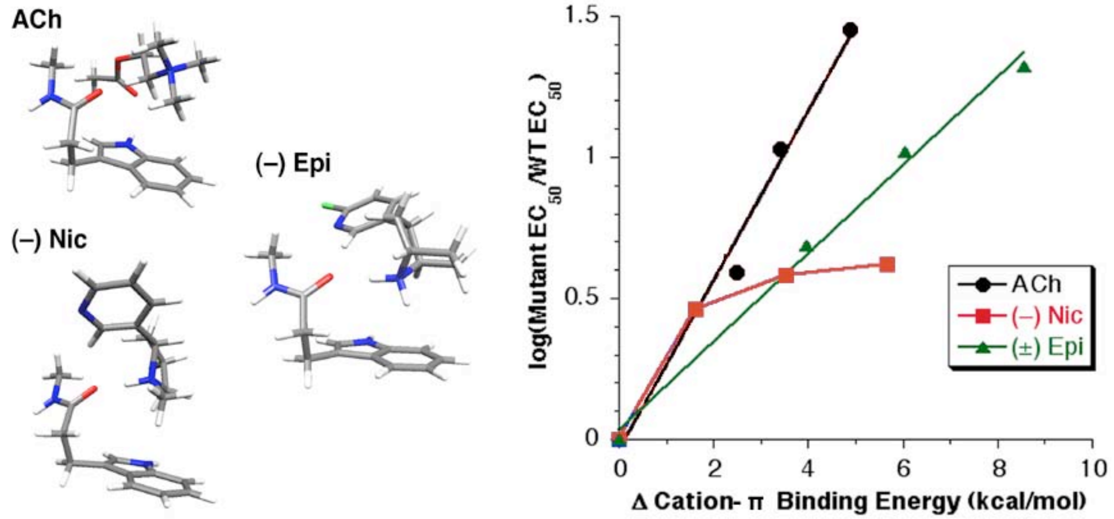


Figure 2. Fluorination Plot with Cation-π Complexes. Changes in Cation-π Binding Energies calculated with the ligand bound complexes at left and their fluorinated indole equivalents. Energies are from BSSE-corrected HF/6-31G** level calculations. D Cation-π Binding Energy computed by subtracting the binding energy of the ligand to Trp from the binding energy of the ligand to F_n-Trp.

In this case, the difference in binding energy between the parent indole and the fluorinated derivative is given. As in the simple Na^+ calculations, a linear relationship is observed between the calculated and observed binding of ACh and (±) Epi, but a "threshholding" behavior is observed for Nic. These calculations still predict a linear change in the Nic binding energy with fluorination. We take this to be an indication that we still have not achieved a sufficiently realistic representation of the binding site environment to describe the change in the binding mechanism of Nic. We also take the fact that the slopes of the ACh and Epi plots are not the same to be an indication of the inadequency of these models; the shift in their EC_{50}s is equivalent, thus we would expect that the calculated shift in their binding energy would be equivalent. (This assumes that the effect on gating is the same among the different unnatural mutant receptors.)

Full Models of the nAChR Binding Domain

We therefore sought to generate a computational model which took into account the full binding site. Computational models of acetylcholine-binding protein (AChBP), the soluble protein homologous to the nAChR binding domain, can be used as a model of the binding site. However, there is good reason to believe that this will still not give an accurate representation of ligand binding for the muscle nAChR. Among AChBP, the muscle receptor and the α3β4 and α7 neuronal receptors, the binding site residues of the "aromatic box" are universally conserved.[4, 5] Yet the relative affinities of ACh, Nic, and Epi vary over orders of magnitude among these four classes of proteins. (Table 1) For example, although AChBP binds Nic more tightly than ACh, the Nic EC_{50} for the muscle receptor is much higher than the ACh EC_{50}. Clearly, a model of the binding pocket that simply uses the box residues as they are positioned in the AChBP structure will not give an accurate representation of the muscle binding pocket.

Table 1. "Box" Residue Conservation and Ligand Affinity Variation

	AChBP	**Muscle**	**α3β4**	**α7**
	Y89	αY93	αY90	αY91
	W143	αW149	αW146	αW147
	Y185	αY190	αY187	αY186
	Y190	αY198	αY194	αY193
	W54	γW55/δW57	βW59	αW53
	Ligand Affinity/ACh Affinity			
Nic	0.023[a]	~ 100[b]	0.7[a]	0.1[a]
Epi	0.0003[a]	0.5[b]	10^{-5} [a]	0.01[a]

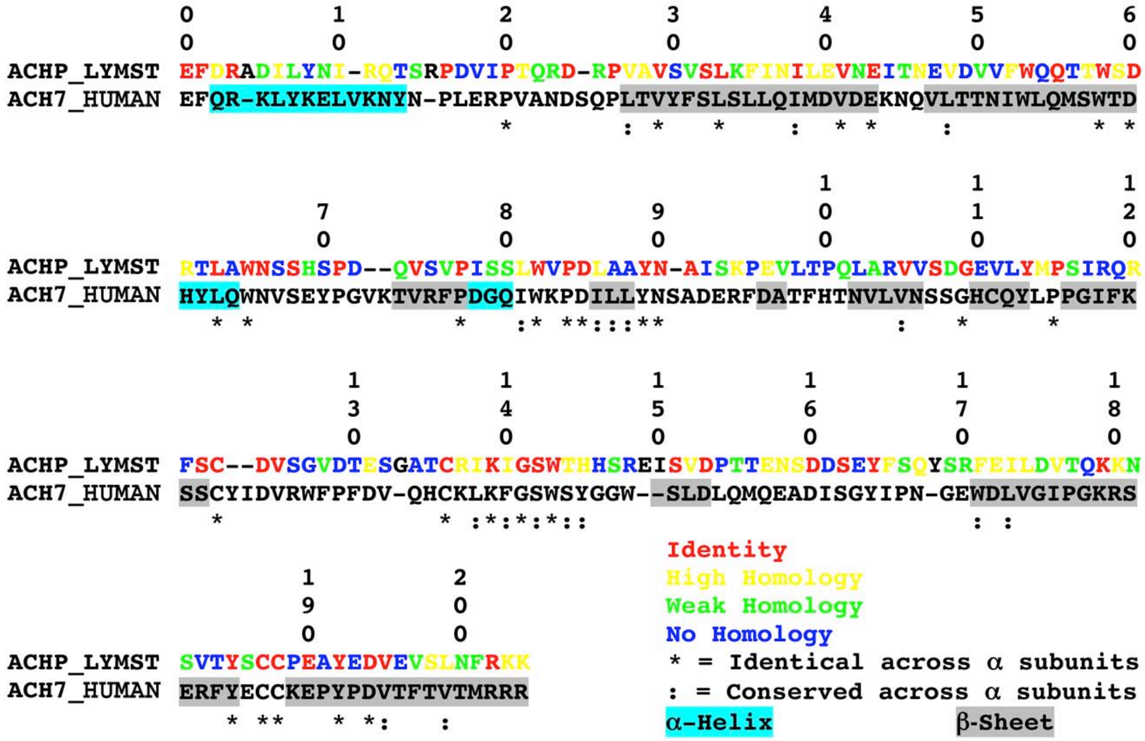
Table 1. Binding Site Residue Conservation and Ligand Affinity Variation. Sequence alignments based on *Lymnea stagnalis* AChBP, mouse muscle nAChR, mouse α3β4, and human α7 sequences. Nic and Epi relative affinities based on a mixture of (a) binding data,[6-8] and (b) EC_{50}s from functional studies.[2, 9]

The subtly of the binding site presents an intriguing problem for the physical chemistry of molecular recognition. Although the box residues are conserved, the rest of the proteins' structures obviously differ. It must be that the residues surrounding the box alter the shape subtly so that ligand affinities differ. This is analogous to the second solvation of a molecule in solution: although these residues do not contact the ligand directly, they have a substantial effect on its binding affinity. Therefore, to accurately represent the nAChR binding pocket, we must generate a model that includes the influence that these residues have on the box residues.

To generate such a model, we employed a mixture of homology modeling, molecular dynamics simulations (MD), and mixed method quantum mechanical/molecular mechanical (QMMM) docking and binding energy calculations. These computational methods are all essential to developing a high level computational model of the binding site. Homology modeling involves the threading of the primary amino acid sequence of the chosen nAChR onto the coordinates of one of the AChBP structures, followed by minimizations of this initial geometry. However, static minimizations do not give good representations of a dynamic protein like the nAChR, so we perform dynamics runs to allow the binding site to sample some of its relevant conformations. These MD runs are performed on structures with and without a ligand in the binding site. Finally, QMMM calculations provide for the treatment of subtle interactions like the cation-π interaction with *ab initio* theory while still permitting the representation of the whole protein. In combination, this approach represents the highest level modeling of the nAChR binding site currently possible.

Homology Modeling of the α7 nAChR

Modeling of the human α7 nAChR ligand binding domain began by aligning its sequence with the sequence of the *Lymnea stagnala* AChBP crystallized by Sixma and coworkers.[10] The following sequence alignment was generated using the T-Coffee website, followed by some adjustments.



```
            0         1          2          3         4         5         6
            0         0          0          0         0         0         0
ACHP_LYMST  EFDRADILYNI-RQTSRPDVIPTQRD-RPVAVSVSLKFINILEVNEITNEVDVVFWQQTTWSD
ACH7_HUMAN  EFQR-KLYKELVKNYN-PLERPVANDSQPLTVYFSLSLLQIMDVDEKNQVLTTNIWLQMSWTD
                                 *       : *   *    :  * *    :         * *

                                                      1         1         1
                     7           8         9          0         1         2
                     0           0         0          0         0         0
ACHP_LYMST  RTLAWNSSHSPD--QVSVPISSLWVPDLAAYN-AISKPEVLTPQLARVVSDGEVLYMPSIRQR
ACH7_HUMAN  HYLQWNVSEYPGVKTVRFPDGQIWKPDILLYNSADERFDATFHTNVLVNSSGHCQYLPPGIFK
              * *             *   :* **:::**               :   *     *

                       1         1         1         1         1         1
                       3         4         5         6         7         8
                       0         0         0         0         0         0
ACHP_LYMST  FSC--DVSGVDTESGATCRIKIGSWTHHSREISVDPTTENSDDSEYFSQYSRFEILDVTQKKN
ACH7_HUMAN  SSCYIDVRWFPFDV-QHCKLKFGSWSYGGW--SLDLQMQEADISGYIPN-GEWDLVGIPGKRS
              *              * :*:*:*::                           : :

                    1         2
                    9         0
                    0         0
ACHP_LYMST  SVTYSCCPEAYEDVEVSLNFRKK
ACH7_HUMAN  ERFYECCKEPYPDVTFTVTMRRR
               * **    * *:   :
```

Identity
High Homology
Weak Homology
No Homology
* = Identical across α subunits
: = Conserved across α subunits
α-Helix β-Sheet

Figure 3. Sequence alignment of the human α7 nAChR and *Lymnea stagnala* AChBP.

Comparisons to previous sequence alignments by McCammon, Sine, and Changeux show that these sequence alignments agree closely, except in four areas.[5, 11-14] The first, comprising residues 0–26 of AChBP, contains prolines 16, 20, and 26. It was decided that alignment with all three prolines should be maintained because this generated a favorable alignment with ACh binding proteins from two other species, *Aplysia* and *Bolinus*.[15] For the second region, AChBP 91-106, we used McCammon's alignment because it was based on functional data from the muscle nAChR collected by Sine.[16] Using Lys scanning mutagenesis, they established the register of the alignment with AChBP in this β-sheet region. The alignment of the Cys-loop (AChBP 123-136) shown below was chosen because the DV motif shown is conserved through many members of the nAChR family.[5] However, the loop refinement protocols used below alter the structure of this loop, so the alignment of this region is of little consequence. Finally, it is important to note that the alignment chosen for α7 with the AChBP 149-168 region, while it shows high similarity and identity, is different than the alignment used by Sine in an experiment in which AChBP was coupled to the transmembrane regions of the 5-HT$_{3A}$ receptor (5-HT$_3$R).[11] Again, these differences are probably immaterial once the loop geometry is minimized.

This alignment was then used to generate an initial homology model using Prime, within the Schrödinger suite of programs.[17] A single subunit of the Nic-bound AChBP crystal structure (Subunit A of 1UW6) was used as a template, without Nic or crystallographic waters present. Sidechain movement was permitted during the build step. The resulting structure was exported as a PDB file. This single subunit was used to build a symmetric pentamer by aligning it successively with the A, B, C, D, and E subunits of the AChBP structure using Swiss PdbViewer.[18] This pentameric α7 structure was returned to Prime where the sidechain prediction and loop refinement algorithms were used. Sidechain conformations were predicted for the whole pentamer. Residues that are identical among all nAChR α subunits as well as those that are highly conserved (i.e. homologous) among all nAChR α subunits were held fixed (Residues designated by "*" and ":", respectively, in Fig. 3). Loop refinement was performed only on the A subunit and on the B and E subunit loops which contacted the A subunit. This minimized computational expense as only one subunit would be taken on to molecular mechanics minimizations in GROMACS.[19]

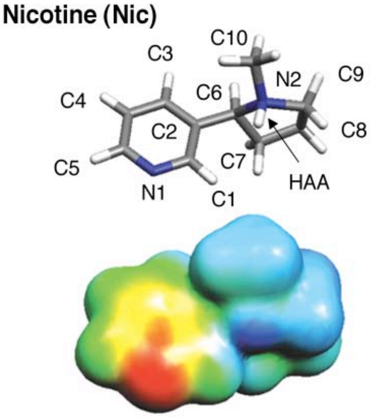
This nAChR α 7 monomer, with its sidechains and loops optimized, was again used to generate a symmetric pentamer by successive alignment with the subunits of AChBP in Swiss PdbViewer. This pentamer was then converted to GROMACS format for molecular mechanics minmizations. It was placed into a periodic box with 7 Å gaps between the protein and the box edge. Explicit solvation was added with SPC water molecules.[19] Sodium and chloride ions were placed in the box at a molarity of 150 mM, with an excess of sodium ions to neutralize the – 20 charge of the protein. Three minimization steps were then performed, with a gradual release of computational restraints on the protein atoms. Identical (*'ed) residues were held fixed throughout these minimization steps.

The A subunit from this minimized α7 pentamer was used to build another symmetric pentamer. The AChBP model used was simply taken from subunits A, B, C, D, and E of the 1UW6 (PDB code) crystal structure. Six models were generated for MD simulations: AChBP with Nic, carbamoyl choline (CCh), or no ligand and α7 with Nic, CCh, or no ligand.

Incorporation of Ligands into Computational Models

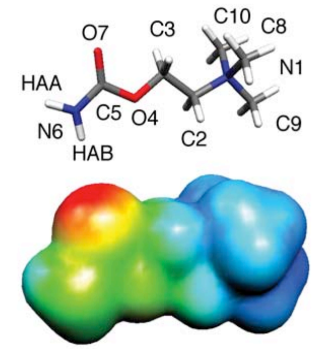
We inserted the ligand, Nic or CCh, into each of the 5 binding pockets. This was done by aligning the box residues of the A binding site of the AChBP structure with Nic (1UW6) or the C binding site of the AChBP structure with CCh (1UV6). Nic and CCh Gromacs parameters were generated using the ProDRG website (http://davapc1.bioch.dundee.ac.uk/programs/prodrg/).[20] However, the ProDRG charges were deemed unacceptable, because they placed almost all of the positive charge on the nitrogen atoms of the ammoniums (Fig. 4, ProDRG Charge).

Nicotine (Nic)



Atom Type	Atom #	ProDRG Charge	HF Charge
CH3	C10	0.075	0.210
NL	N2	0.653	-0.070
H	HAA	0.053	0.270
CH2	C9	0.110	0.180
CH2	C8	0.109	0.110
CH2	C7	0.000	0.050
CH1	C6	0.000	0.150
CB	C2	0.086	-0.250
CR61	C3	0.034	0.260
CR61	C4	0.034	-0.240
CR61	C5	0.034	0.580
NR6	N1	-0.188	-0.730
CR61	C1	0.000	0.480

Carbamoyl Choline (CCh)



Atom Type	Atom #	ProDRG Charge	HF Charge	Modified Charge
CH3	C8	0.073	0.270	0.270
NL	N1	0.782	-0.090	-0.090
CH3	C9	0.072	0.270	0.270
CH3	C10	0.073	0.270	0.270
CH2	C2	0.104	0.150	0.150
CS2	C3	0.263	0.420	0.330
OS	O4	-0.138	-0.610	-0.500
C	C5	0.414	1.240	0.840
O	O7	-0.643	-0.720	-0.600
NT	N6	0.000	-1.200	-0.760
H	HAB	0.000	0.500	0.410
H	HAA	0.000	0.500	0.410

Figure 4. Agonist Charge Parameters. Nic and CCh Gromacs parameters from ProDRG and HF/6-31G** calculations. HF/6-31G** electrostatic surfaces correspond to an energy range of + 10 to + 130 kcal/mol, where blue is highly positive and red is less positive.

ChelpG charges from HF/6-31G** calculations were used instead (Fig. 4, HF Charge). However, in the case of CCh, these partial charges created large intramolecular forces between carbamoyl protons and oxygens that the LINCS bond constraint algorithm (in the GROMACS MD engine) was unable to resolve. These charges were attenuated to permit the molecular dynamics simulations to proceed (Fig. 4, Modified Charge).

Initial Molecular Dynamics Simulations

The apo proteins and the Nic-bound proteins were again placed in periodic boxes with solvation and counterions as above. A series of seven minimization steps was performed to prepare the proteins for molecular dynamics simulations:

Minimization 1: Identical residues frozen, protein and nicotine strongly restrained.
Minimization 2: Identical residues frozen, protein backbone strongly restrained.
Minimization 3: Identical residues frozen, protein backbone weakly restrained.
Minimization 4: No residues frozen, identical residues strongly restrained.
Minimization 5: No residues frozen, identical residues weakly restrained.
Minimization 6: All non-hydrogen atoms strongly restrained.
Minimization 7: Completely unrestrained.

These minimized structures were then subjected to MD simulations with the GROMACS forcefield. The MD runs were begun at 0 K, and warmed up to 310 K with a linear annealing function over the first 25 ps. The protein (and drug, if applicable) was strongly restrained during the warm-up phase and then the restraints were relaxed over the next 125 ps. After this point the simulations were allowed to proceed unrestrained. The total system energies began to level out after about 1 ns, indicative of system equilibration. (Fig. 5)

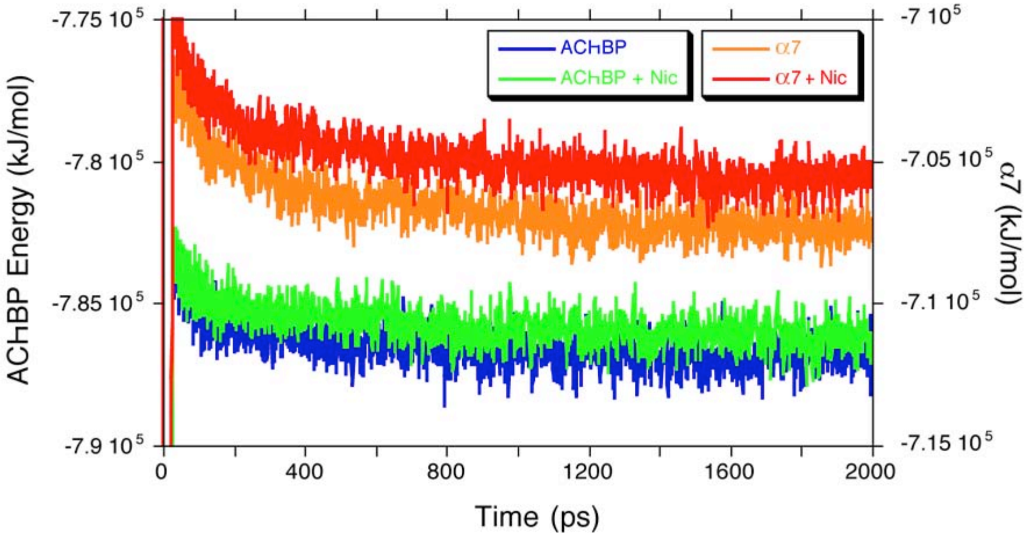


Figure 5. System Energy of First MD Run. Computed as described in Materials and Methods Section.

Plotting the RMSD of the protein atoms (following alignment of the identical, "*'ed", residues of the structure at each timepoint to the structure at t = 0) as a function of time showed us that the proteins had more or less equilibrated structurally as well. (Fig. 6) The AChBP structures had deviated by about 2.7 Å from their positions at time 0, (essentially their positions in the crystal structure). This is to be expected. Since the crystal structure resolution was 2.8 Å, we should not be surprised by this amount of fluctuation.[10] The α7 structures deviated more from their initial structrues, which is also not surprising. The structure must differ somewhat from the AChBP structure, and we would not expect all of that difference to be accounted for by minimization steps. When we compare the structure of the liganded to the unliganded structure, we see that the RMSD of the protein as a whole does not depend substantially on the presence of Nic.

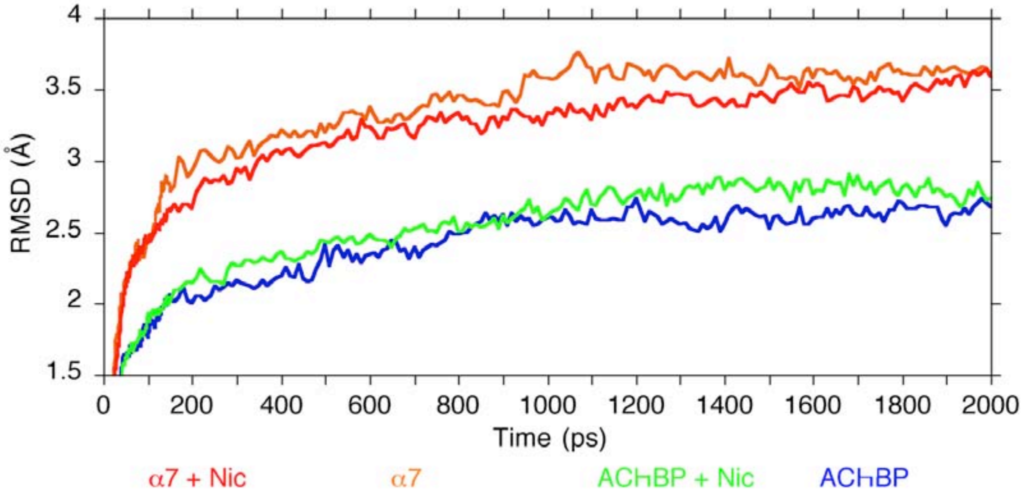


Figure 6. Protein RMSD Relative to Starting Structure for First MD Run. Computed as described in Materials and Methods Section.

An examination of the box residues (designated in Table 1) tells a different story. (Fig. 7) For AChBP, the RMSD of the box residues relative to their starting positions, like the RMSD of the protein as a whole, is independent of the presence of Nic. However, for α7, there is a 0.5 Å difference between the RMSD of the box residues for the apo protein and the Nic-bound protein. This indicates that for AChBP, the presence of the ligand has little effect on the shape of the box, but that it may help to hold the α7 box together. This is consistent with an evolutionary perspective on the roles of the two proteins. AChBP, whose role is to buffer the amount of ACh in the snail synapse and to "soak up" nAChR-targeted toxins, needs only bind the ligand, it does not need to gate.[6] α7, on the other hand, must gate, therefore it could have an inherently more flexible binding pocket. In the case of AChBP, drug binding occurs through a "lock-in-key" mechanism, whereas for α7, drug binding could be a more cooperative process, an induced-fit mechanism.

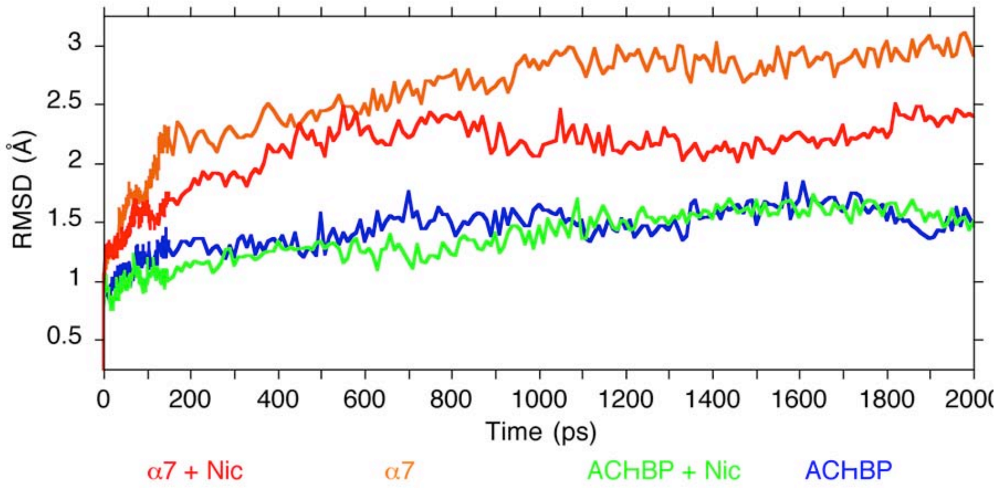


Figure 7. Box RMSD Relative to Starting Structure for First MD Run. Computed as described in Materials and Methods Section.

Snapshots of the binding site at t = 0 and t = 2 ns illustrate this point nicely. (Fig. 8) The binding site residues were carefully held in place throughout the pre-MD minimizations, so they start out in the same relative geometry, the geometry of the AChBP crystal structure. After 2ns, the AChBP box structures still resemble the initial structure, with or without Nic. The Nic-bound α7 box also resembles the box of the crystal structure, but the structure of the box from the apo α7 run has changed dramatically, and no longer seems ready to bind a ligand. Only snapshots of Box C (the C/D subunit interface) are shown, but these are generally representative. By plotting the RMSDs of the box residues from each subunit interface, we can see that the box is slightly more stable in the presence of Nic for AChBP as well. (Fig. 9, boxes labeled according to the Trp 143/147 subunit) However, this effect is small compared to 1) the difference in box stability between α7 and AChBP, and 2) the difference in box stability between unliganded α7 and α7 with Nic.

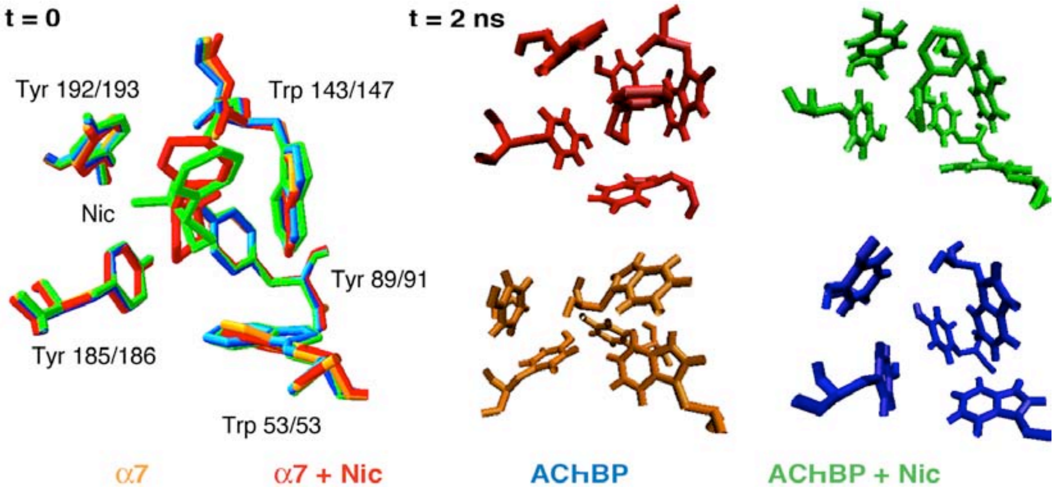


Figure 8. Snapshots of Box Residues from First MD Run. All images depict C/D subunit interface.

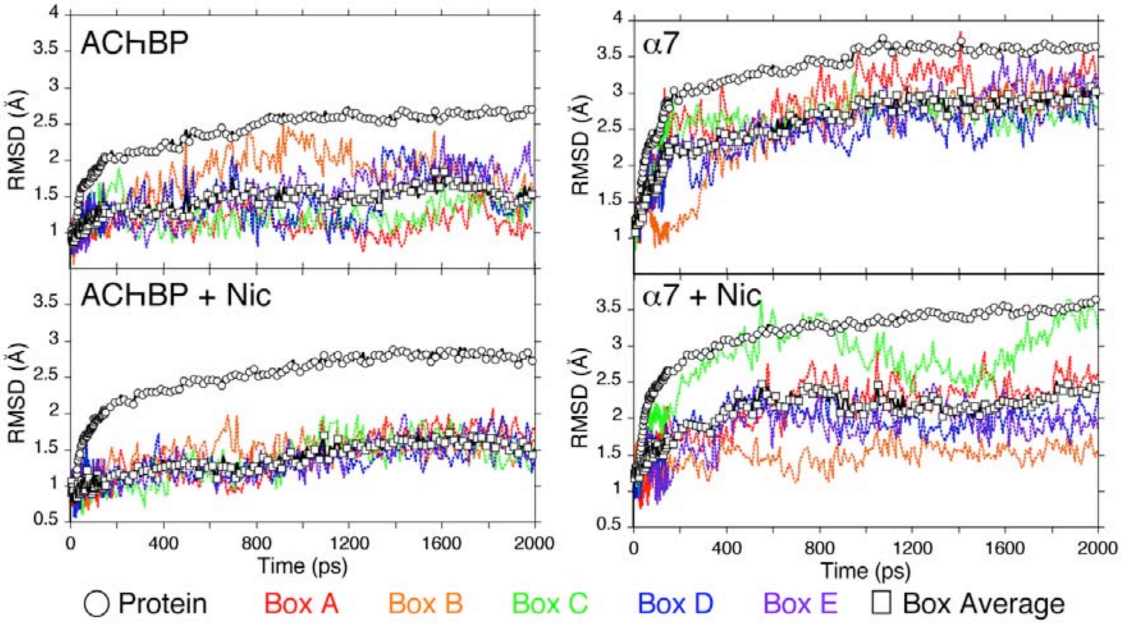


Figure 9. Protein and Individual Box RMSD Relative to Starting Structure. Computed as described in Materials and Methods Section.

An analysis of the overall structure of the proteins revealed substantial changes to the structures of the α7 pentamers relative to the AChBP crystal structure. While the AChBP pentamers had retained a small pore throughout the MD simulation, the "pore" of the α7 pentamers had become almost completely occluded. This was due to the inward movement of the β4-β5 loop (residues 91 – 101 of AChBP, residues 93 – 104 of α7, see Fig. 3) We felt that this was "unphysiological" in that it did not seem possible to achieve ion flux through such a small vestibule (which provides access to the transmembrane pore).

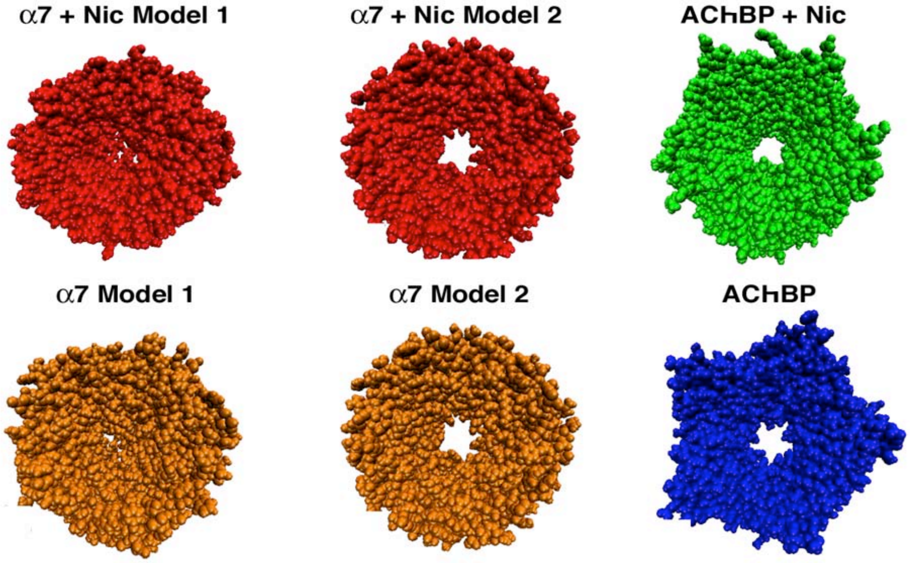


Figure 10. Space-filling Model Images of the Full Pentamers of α7 and AChBP.

Second Generation Molecular Dynamics Simulations

This pore shrinkage might serve as an indictment of the validity of our observations concerning box stability. We therefore generated a second model for MD simulations that we will call Model 2 (calling the original α7 model, Model 1). (Fig. 10) Model 2 was identical to the pre-minimization state of Model 1 except that the backbone coordinates of the β4-β5 loop had been returned to their positions in the AChBP crystal structure (See Materials and Methods for details). We were also concerned that the differences might come from the way in which we had built our α7 model 1 pentamer, by aligning one subunit successively to the A, B, C, D, and E subunits of the 1UW6 crystal structure. As a control experiment, we built a new AChBP pentamer in a similar fashion. Rather than simply use one of the pentamers from one of the crystal structures, we generated a symmetric pentamer by alignment of the A subunit from the 1UW6 structure with the B, C, D, and E subunits. Model 2 and the new AChBP were minimized and subjected to MD simulations as above. Many of the features of the Model 1 simulations were recapitulated. For example, large differences in the changes of the RMSD of the whole protein were again observed. (Fig. 11) AChBP changed less than α7. However, in this case, we found that the box residues of α7 were no more stable relative to their initial positions with Nic than without it. In fact, the box at the D/E interface was more stable without Nic. An examination of snapshots of the binding sites supports the numerical data. Box D (the D/E subunit interface) has come apart sufficiently that Nic is diffusing out of it after 2 ns. (Fig. 12) The 2 ns timepoint for Box D is shown from the Model 1 MD simulation, showing the box holding together.

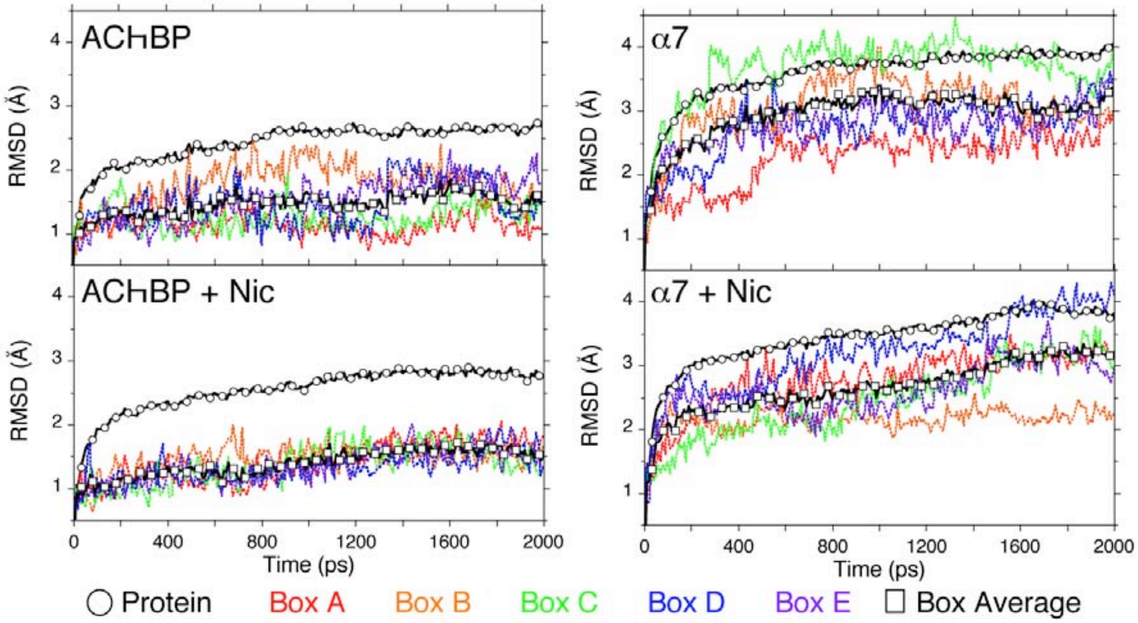


Figure 11. Protein and Individual Box RMSD Relative to Starting Structure of Model 2 MD run. Computed as described in Materials and Methods Section.

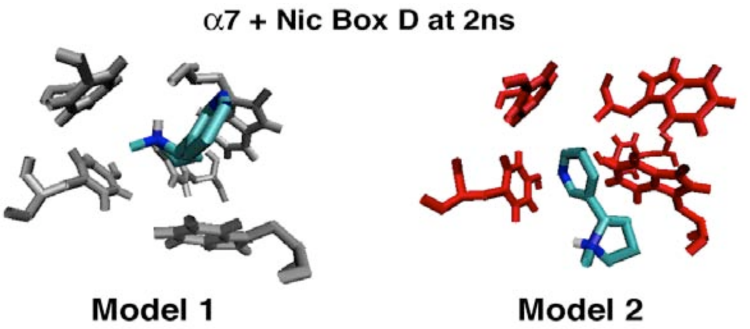


Figure 12. Comparison of Box D Stability from MD Simulations with α7 Model 1 and α7 Model 2.

Final Molecular Dynamics Simulations

The difference in the effect of Nic on box stability in the α7 Model 2 MD simulations compared to the α7 Model 1 simulations made us concerned that our efforts to minimize the α7 structure while holding the β4-β5 loop in the same place that it occupies in the AChBP structures had destabilized the box. We therefore started a third set of MD simulations using the α7 Model 2 structure following the third minimization (in which identical residues were held fixed and the backbone was weakly restrained, see above). These MD runs were initiated as before, and the α7 structures behaved similarly to those in Model 1. Since we felt that the position of the β4-β5 loop was now reasonable and that the box stability seemed reasonable (see below), we allowed these simulations to continue for 5 ns. The AChBP MD runs were conducted as before, and we included simulations with carbamoyl choline (CCh), the ACh analog with which Sixma and coworkers had cocrystallized AChBP (PDB code 1UV6).[21] Again, the total system energy indicated that the proteins had largely equilibrated after 1 ns. (Fig. 13)

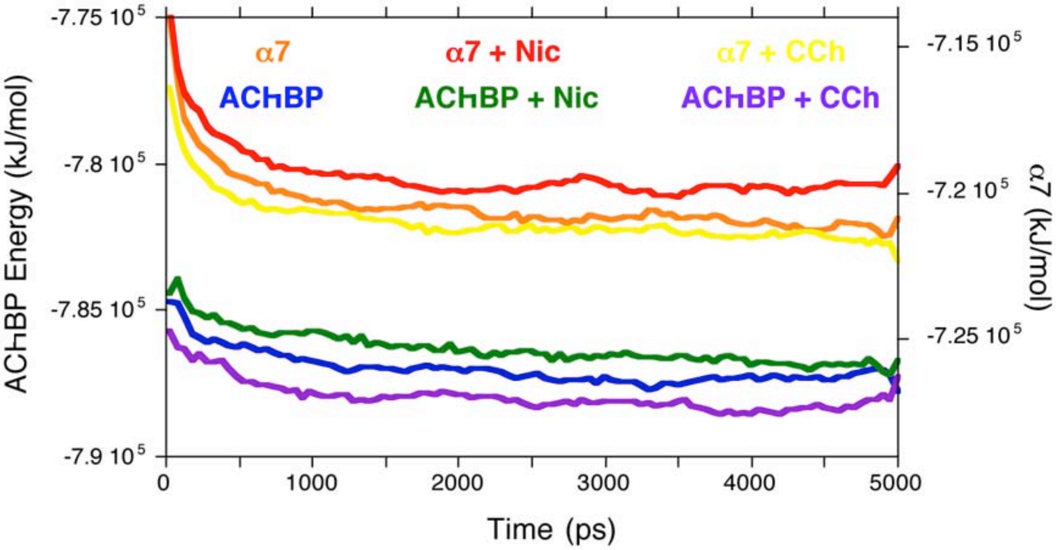


Figure 13. System Energy of Third MD Run. Computed as described in Materials and Methods Section.

As before, the RMSD from time 0 of the protein as a whole seemed indifferent to the presence of ligand, and large differences were observed between the group of AChBP simulations and the group of α7 simulations. (Fig. 14)

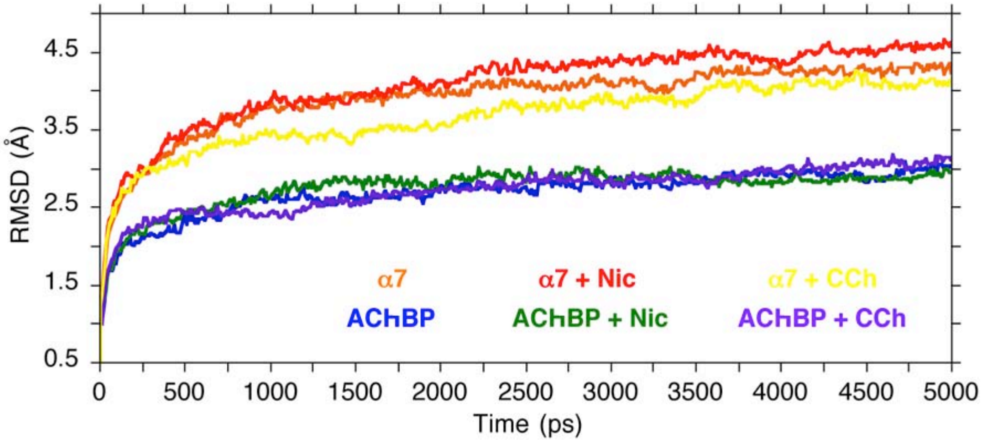
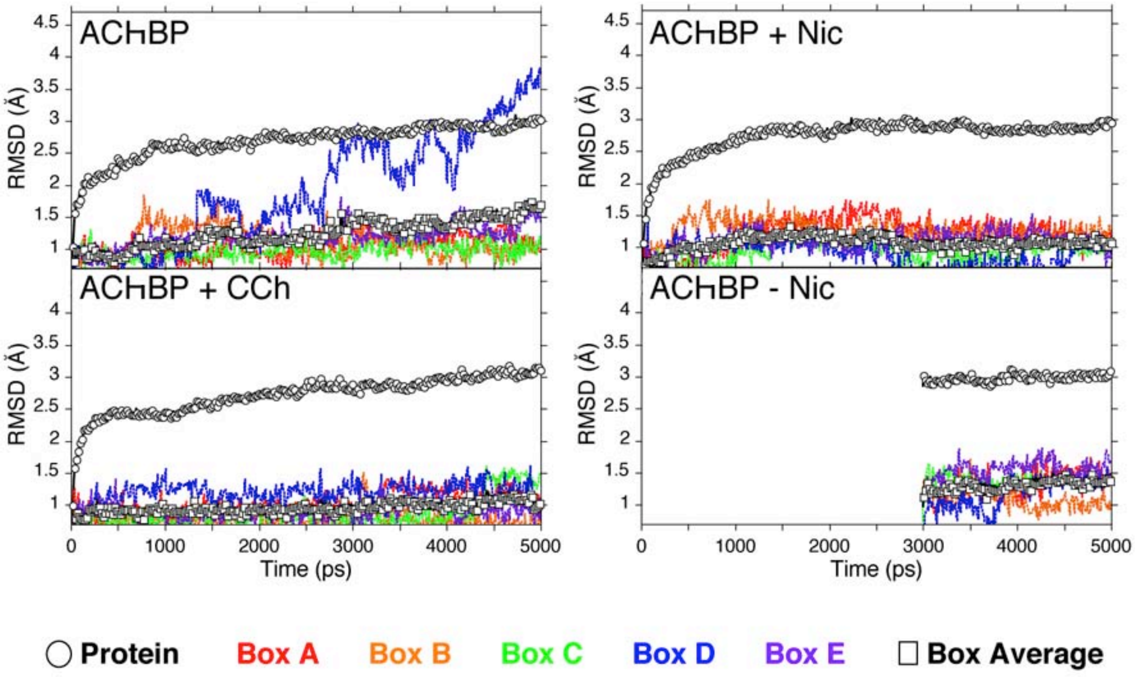
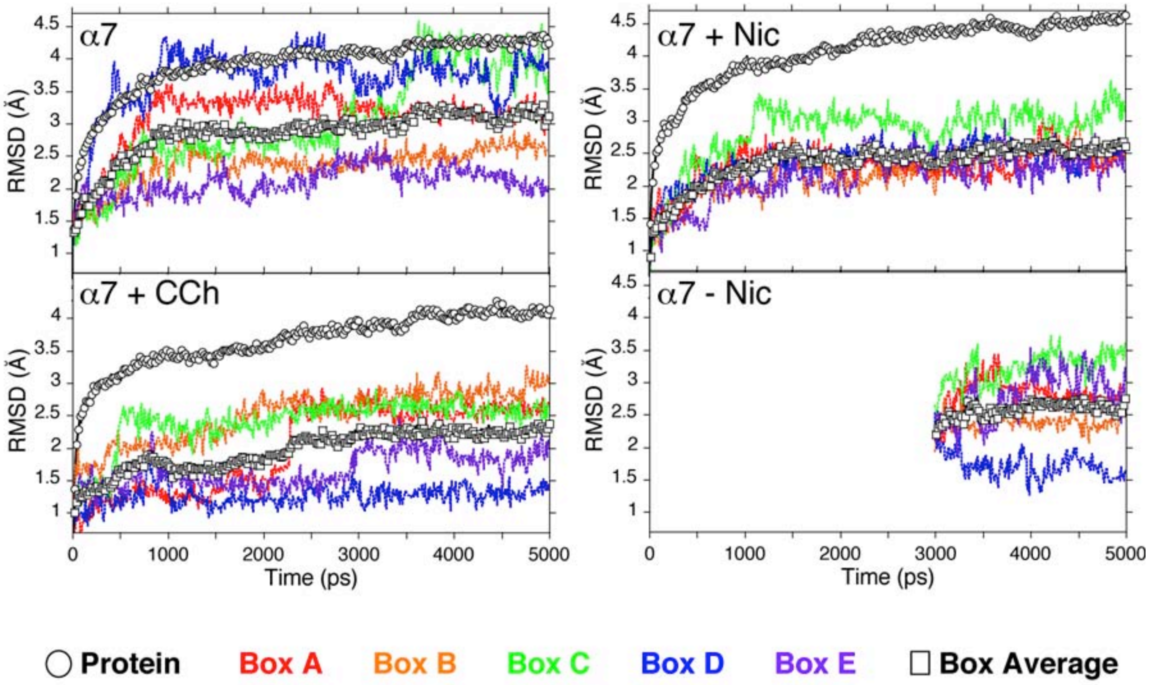


Figure 14. Protein RMSD Relative to Starting Structure for Third MD Run. Computed as described in Materials and Methods Section.

The RMSD of the box residues again showed a dependence on the presence of ligand in the case of α7, but not for AChBP. As one can see from Figure 15 on the following page, the RMSD of the box residues of α7 with Nic get as high 4 Å in the case of boxes C and D in the apo α7 simulation, but they get no higher than 3.2 Å in the liganded simulations (Box C, α7 + Nic). The RMSDs of all but one of the AChBP boxes (AChBP box D) are around 1 Å. In fact, inspection of the structure of this box (See AChBP Box D Clusters, Supporting Information) shows that this change is entirely due to a downward "flip" of only one sidechain, Tyr 89. Simulations in which Nic was removed from the structure of an MD run with Nic after 3 ns (α7 – Nic and AChBP – Nic) further demonstrate that Nic stabilizes the box for α7. The RMSDs of three of the boxes increase for the α7 run, but removal of Nic has little effect on the AChBP box. Admittedly, one of the α7 boxes does appear to get more stable. However, this really reflects the imperfection of RMSD as a measure of box stability, for an inspection of images of the box does not bear this out (See Supporting Information). Figure 16 presents images of Box A from each of the simulations to illustrate the opening of the box seen for apo α7 and the retention of box shape seen in all other cases. These snapshots were taken after 2 and 5 ns, respectively. In all of the snapshots of liganded simulations, Nic and CCh retain Trp 143 (AChBP)/Trp 147 (α7) as a primary site of interaction, in keeping with the crystal structures and our functional data on the muscle receptor (Trp 149).[1, 2] This is true in all of the simulations, although, in α7 + Nic Box D, this interaction is primarily through the pyridine ring, counter to our expectations. (See Supporting Information for images representative of each box.)



AChBP Box Residues



α7 Box Residues

Figure 15. Box Residues RMSD Relative to Starting Structure for Third MD Run. Computed as described in Materials and Methods Section.

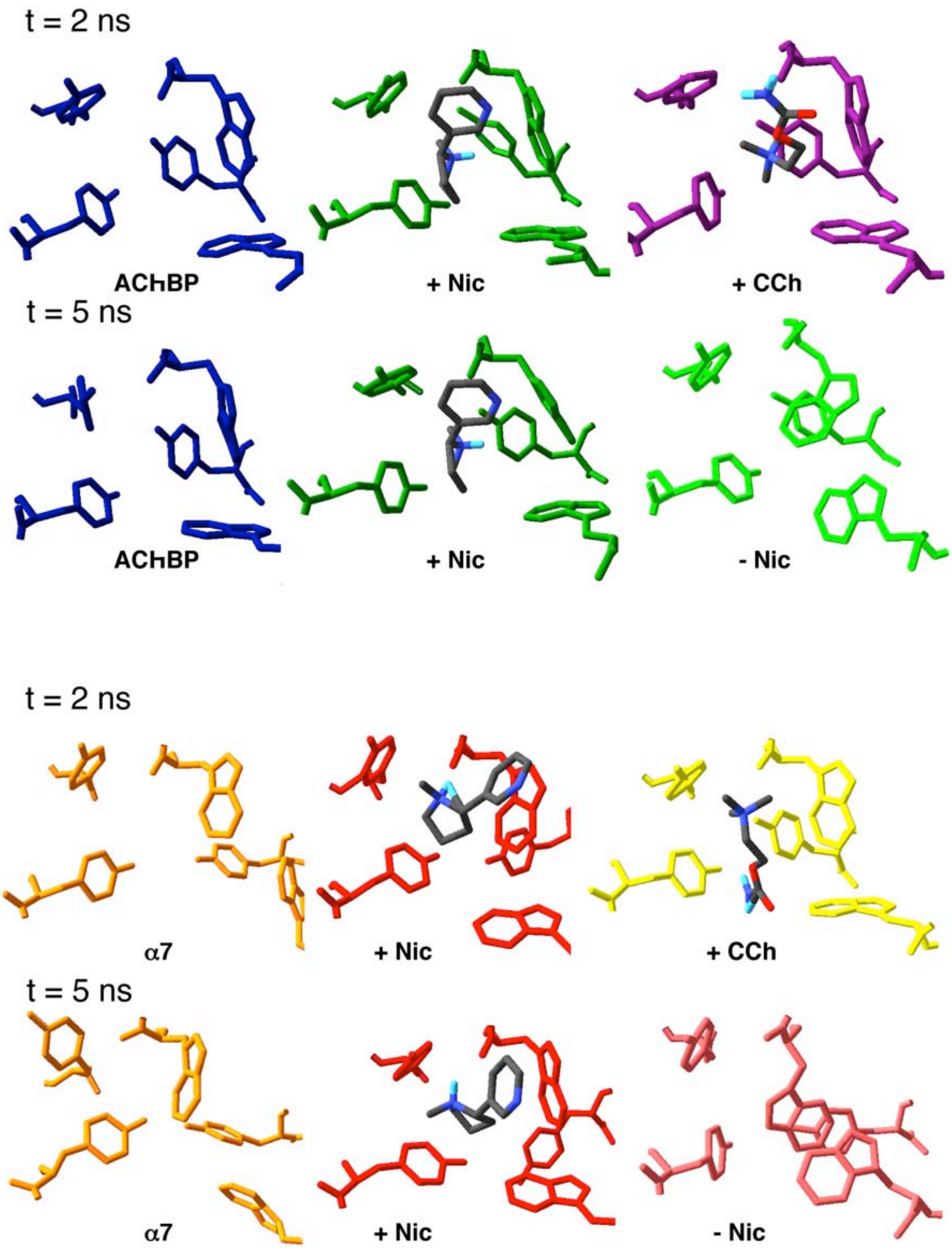


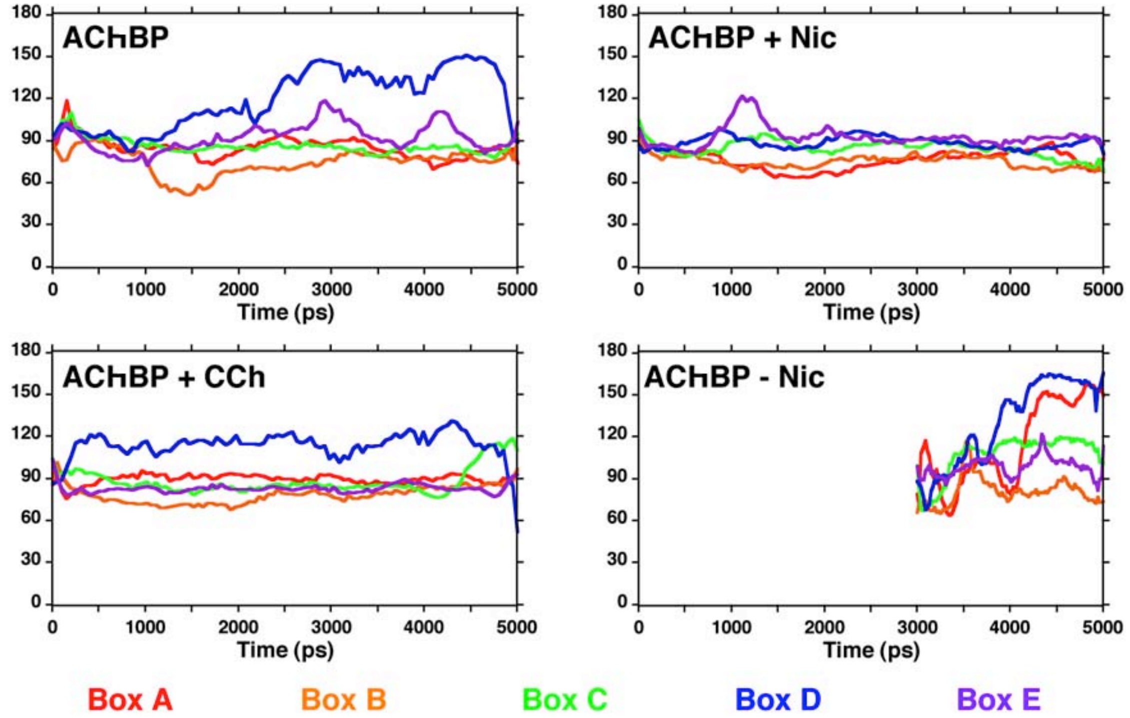
Figure 16. Snapshots of Box Residues from Third MD Run. All images depict Box A (A/B interface).

MD Summary and Comparison to Experiment and Other Simulations

Several other groups have generated homology models of the extracellular domains of various nAChRs.[5, 12, 13, 22] McCammon, Taylor, and coworkers have also subjected a model of the α7 binding domain and AChBP to dynamics simulations.[12-14] Although they focused on changes in the overall shape of the receptor rather than specific changes in the binding site amino acids, they too found that the binding site tended to deviate from the AChBP box structure.[12, 13] They characterized this by changes in the distances between the α carbons of several residues in the binding site.[12] They found that computational docking of ACh before and after the 10 ns of MD yielded different binding orientations.[12] The ACh ammonium tended to be targeted to Trp 147 prior to MD and to Trp 53 after MD. Their AChBP simulation was performed for 45 ns and corroborates much of what we observe, albeit with a different interpretation. They simulated the AChBP with no ligand or with one ACh molecule in each binding site (a total of five ACh). They found that there was an agonist-dependent organization of the binding pocket for AChBP that they characterized through the motion of the Cys 187/Cys188 loop and the plane angle between the sidechains of Trp 143 and Trp 54 (on the adjacent subunit).[14] They found that in the presence of ACh, this angle would remain at about 95°, but that without a ligand present, the angle would flicker up to 150° for some binding sites and remain there for others. While our simulations are substantially shorter, one can see the same behavior in our AChBP simulations. (Fig. 17) McCammon and Taylor corroborate this ACh-dependent alignment of the two Trps with fluorescence quenching experiments performed on AChBP. They find that the Trp fluorescence lifetimes change in the presence or absence of ACh.[14]

These results may seem contradictory to ours, but an examination of the same angle parameter in our simulations shows that the angle changes that they attribute to ligand-dependent organization of the AChBP box are small compared to the changes we observe with α7. We see an even more severe change in this angle, to near 0° or 180°, both angles indicative of the two Trp sidechains becoming parallel. We calculated the plane angle between the sidechain of Trp 143/Trp 147 (AChBP/α7) and each of the other aromatic sidechains in the box. (Figs. 17 –20) In many cases, the angle data paints a picture of a much more unstable box structure than the RMSD data. Some large fluctuations, such as the periodic change in the angle with Tyr 89/Tyr 91 (AChBP/α7), correspond to the spinning of the aromatic sidechain about the C_α-C_β bond. Of these angles, the angle formed with Tyr 185/Tyr 186 (AChBP/α7), may be a reasonable predictor of box stability, in that there is a greater deviation in the angle for the apo structures than for their respective liganded structures and the α7 angle deviation is greater than that for AChBP.

AChBP: Trp 143 / Trp 53 Plane Angle (°)



α7: Trp 147 / Trp 53 Plane Angle (°)

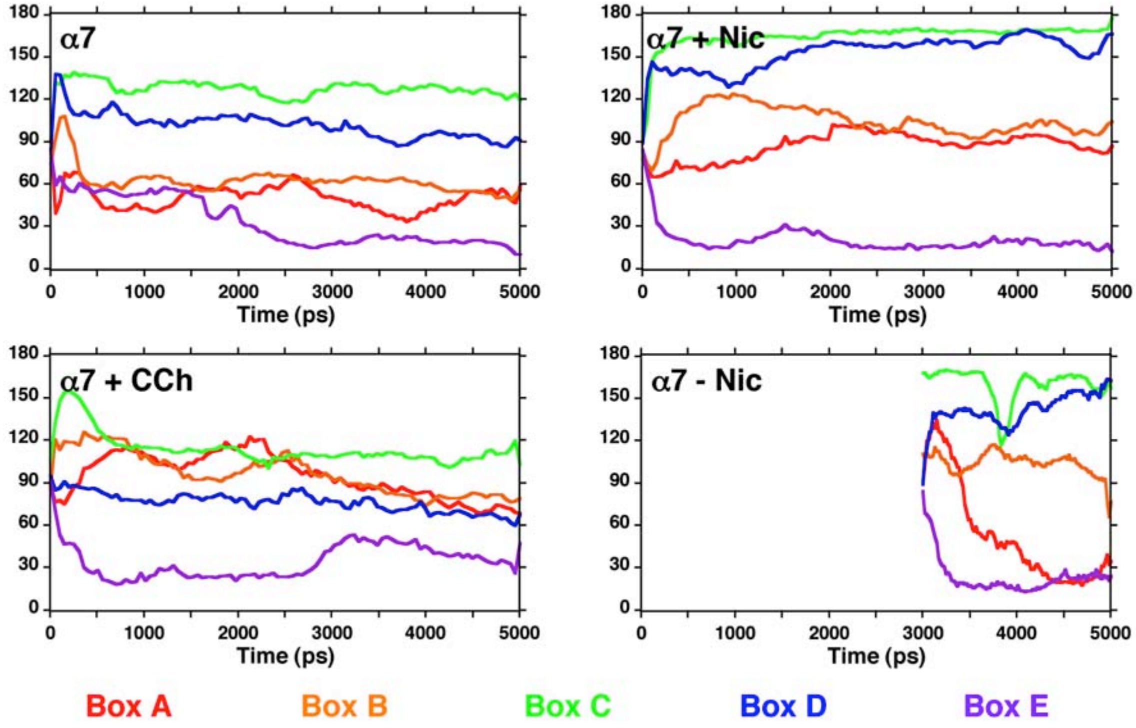
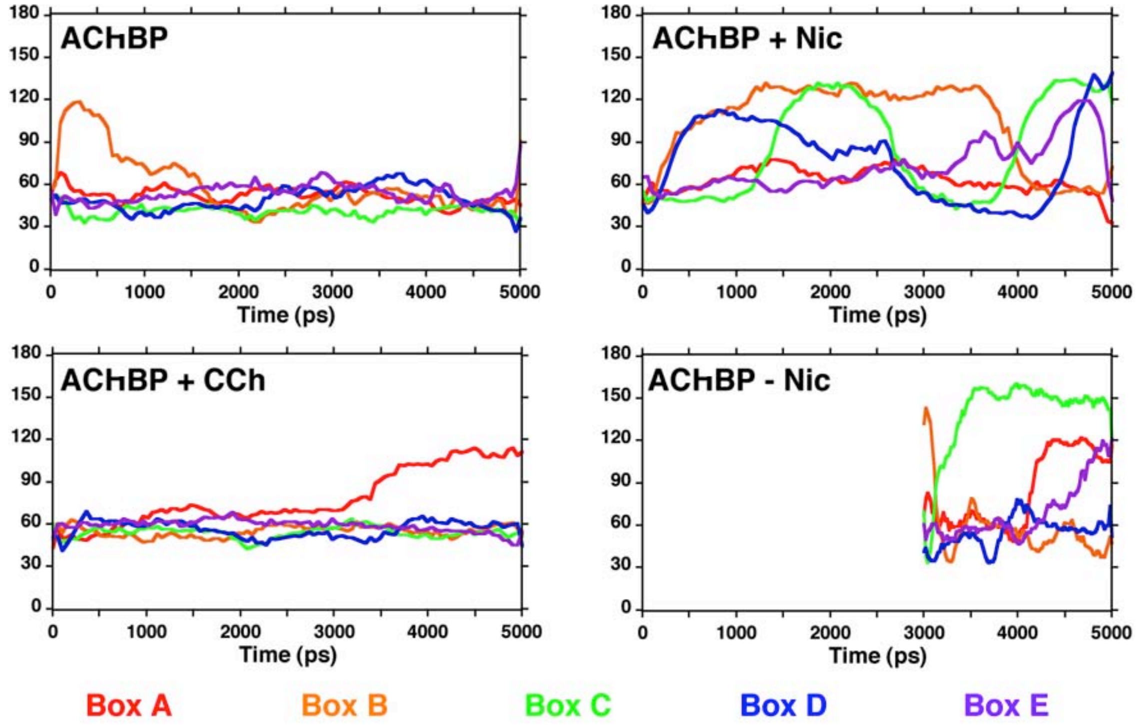


Figure 17. Sidechain Plane Angles from Third MD Run. Calculated as described in Materials and Methods.

AChBP: Trp 143 / Tyr 89 Plane Angle (°)



α7: Trp 147 / Tyr 91 Plane Angle (°)

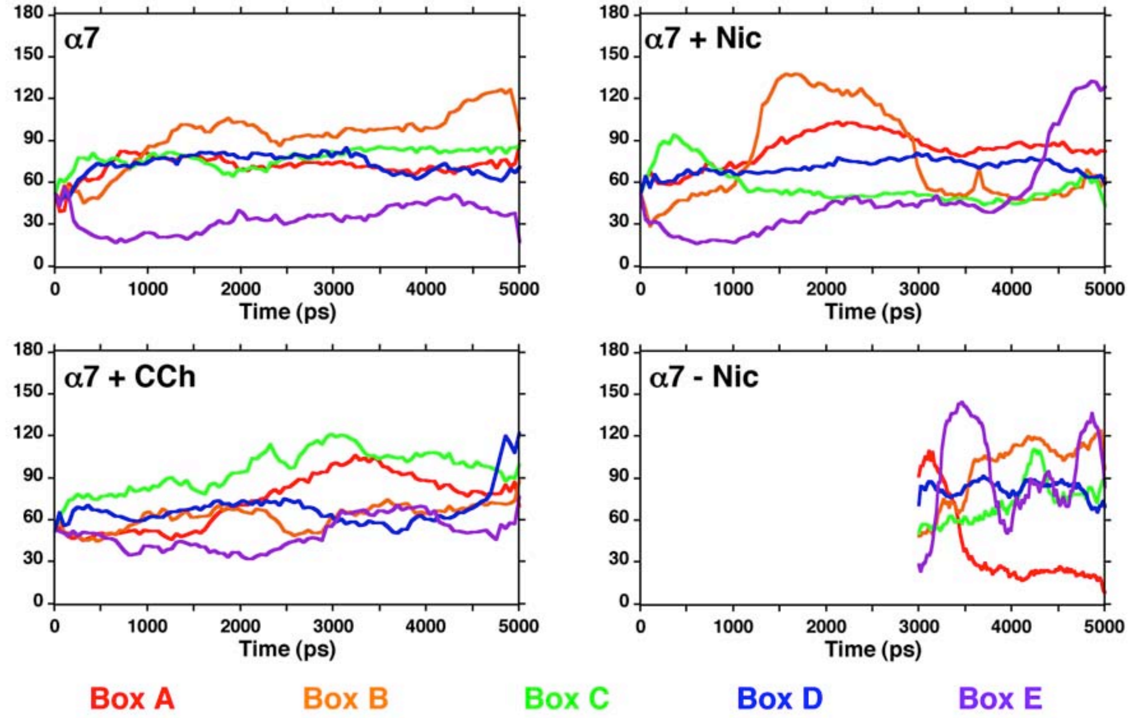
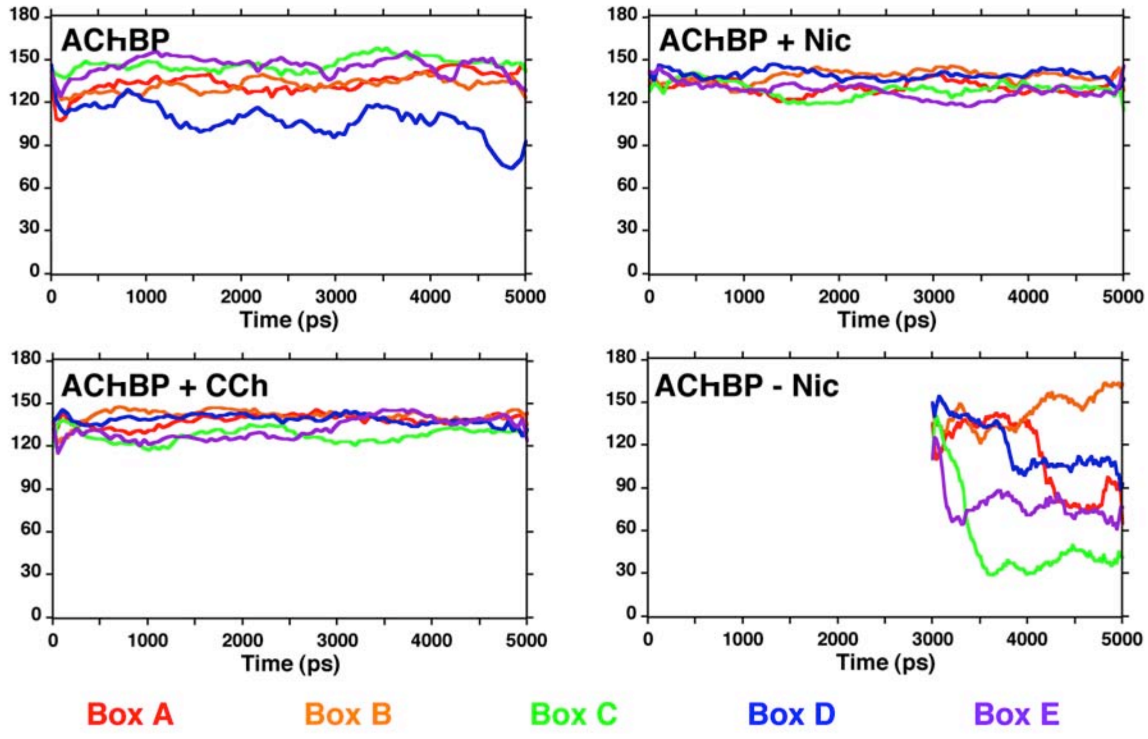


Figure 18. Sidechain Plane Angles from Third MD Run. Calculated as described in Materials and Methods.

AChBP: Trp 143 / Tyr 185 Plane Angle (°)



α7: Trp 147 / Tyr 186 Plane Angle (°)

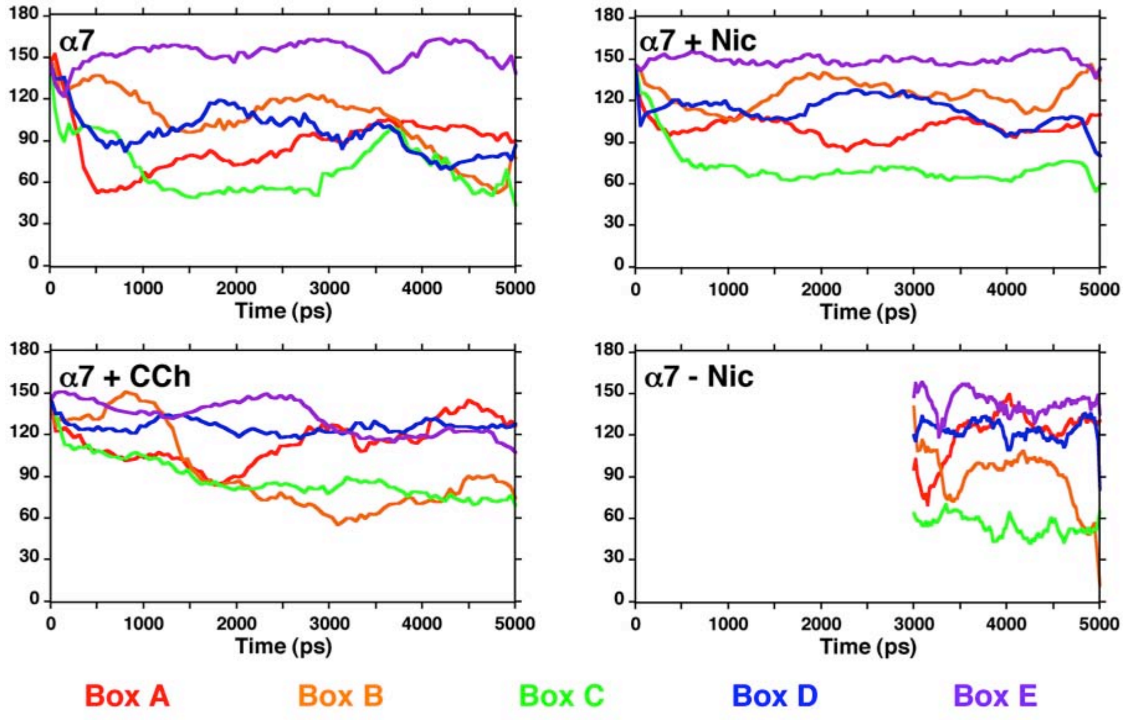
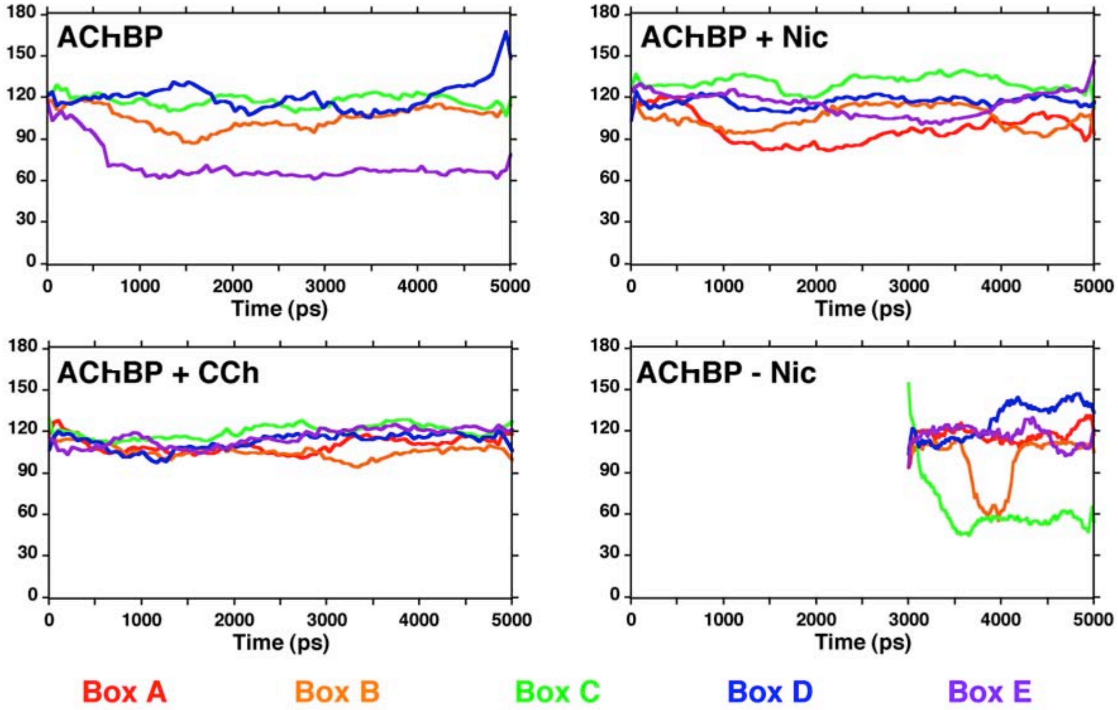


Figure 19. Sidechain Plane Angles from Third MD Run. Calculated as described in Materials and Methods.

AChBP: Trp 143 / Tyr 190 Plane Angle (°)



α7: Trp 147 / Tyr 193 Plane Angle (°)

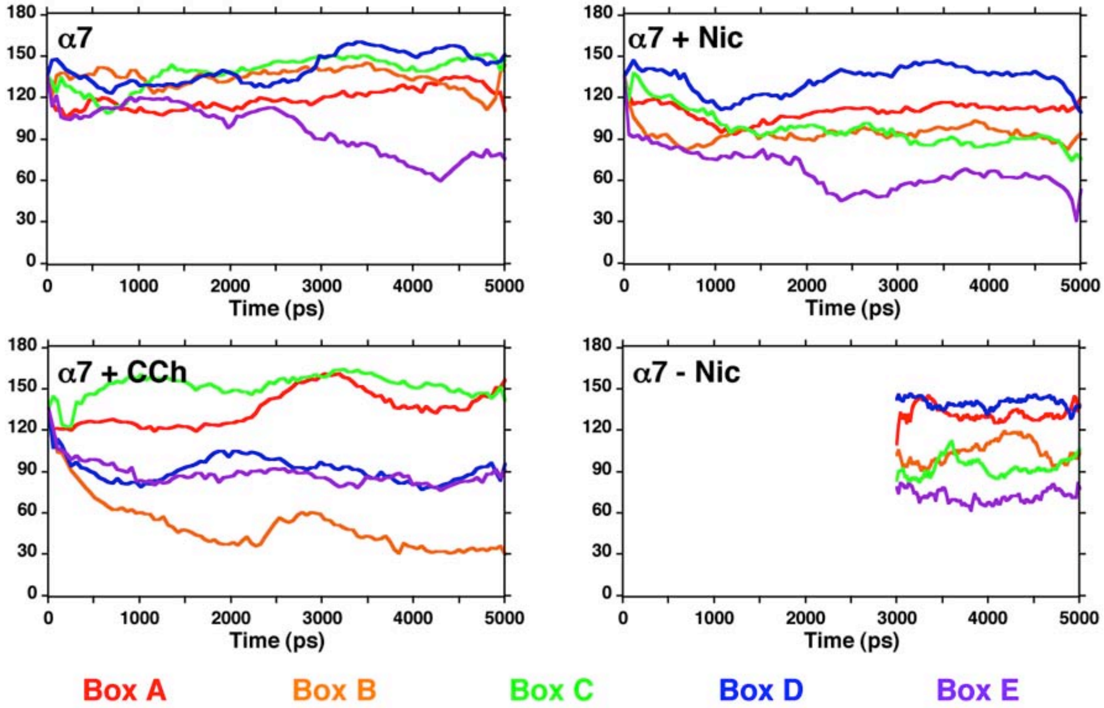


Figure 20. Sidechain Plane Angles from Third MD Run. Calculated as described in Materials and Methods.

Box RMSD still seems like the best measure of stability. If we align the box residues for a liganded simulation with the box residues from the corresponding box in an unliganded simulation, we can calculate the difference in the box shape depending on ligand occupancy. (Table 2) Rather than use the full trajectories, we used a clustering algorithm in the GROMACS analysis suite to identify clusters of structures within the fifth nanosecond of our MD simulations by comparing the RMSD of the box residues of that structure (and ligand, if applicable) with all other structures in the trajectory. The structure with the lowest RMSD to all other structures in that cluster was selected as a mean structure. Typically, each binding site generated three or four clusters, one or two of which represented most of the trajectory. (See Materials and Methods and Supporting Information for a more detailed description of this process and images of the clusters.) These representative structures were used to calculate differences between the liganded and unliganded structures in Table 2 and will be used for all of the following analysis and QMMM calculations. Clearly, the shape of the box for AChBP depends much less on the presence of ligand than for α7.

	RMSD Relative to Corresponding Box from Unliganded Simulation				
	CCh	Nic		CCh	Nic
AChBP Box A	0.85 Å	0.88 Å	α7 Box A	3.52 Å	2.87 Å
AChBP Box B	0.71 Å	0.77 Å	α7 Box A	3.32 Å	3.47 Å
AChBP Box C	0.54 Å	0.77 Å	α7 Box A	3.61 Å	3.67 Å
AChBP Box D	2.61 Å	3.05 Å	α7 Box A	3.53 Å	3.10 Å
AChBP Box E	1.21 Å	0.95 Å	α7 Box A	2.81 Å	2.08 Å
AChBP Average	1.18 Å	1.28 Å	α7 Average	3.36 Å	3.04 Å

Table 2. α7 and AChBP Box RMSD Relative to Unliganded Simulation. Box residues from representative snapshots (See Supporting Information) for each box aligned with corresponding box from unliganded simulation and RMSD calculated with Swiss PDBViewer.

Of the experimental data available for comparison, there are of course AChBP binding studies,[6] and some studies of soluble versions of the extracellular domain of α7, produced in attempts to crystallize the protein prior to the discovery of AChBP.[23] However, the most relevant experimental data to which we can compare our calculated structures is the structural information that is available for the “actual” systems. The first AChBP structure, released in 2001 was ostensibly unliganded, but further refinement showed that it did contain molecules of the cationic buffer HEPES in three of its binding sites and a molecule of ammonium sulfate in another of its binding sites.[10, 21] These are extremely low affinity “agonists”, so it seems reasonable to continue to consider this an unliganged state. In 2004, the Sixma lab released the structures of AChBP with Nic and CCh bound.[21] These structures are remarkable in their similarity. (Fig. 21) The RMSD between the box residues of the liganded and unliganded structures is about 0.3 Å, with the main difference being a slight movement of Tyr185 and Tyr 192 toward Trp 143.

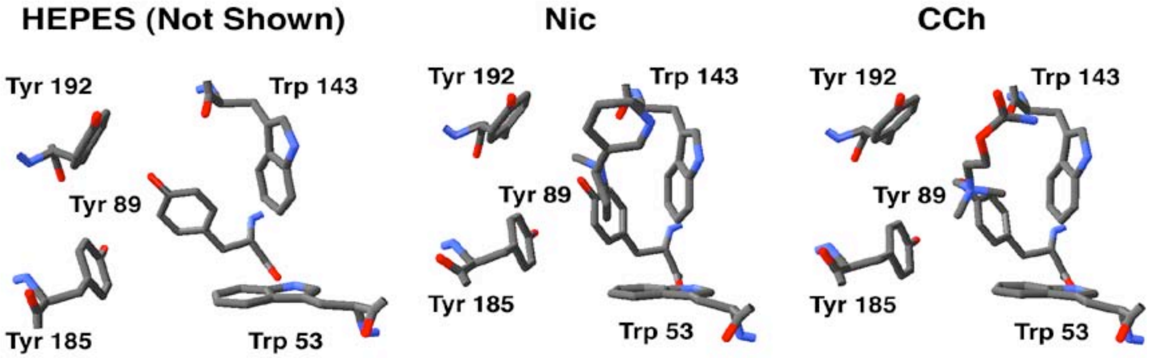


Figure 21. Images of AChBP Box Residues from an Crystallographic Studies by Sixma. (PDB codes 1IB9, 1UW6, 1UV6, left to right)

The similarity between the ostensibly unliganded AChBP structure and its liganded cousins is consistent with the impression of AChBP formed from our simulations. The binding site of AChBP is preorganized to bind nAChR agonists and antagonists. Several lines of research have indicated that AChBP's structure may bear the most resemblance to the desensitized state of the nAChR, the state with the highest ligand affinity. Functional studies in our own lab have shown that some of the contacts (such as hydrogen bonds) observed in AChBP only seem to be present in the desensitized state of the muscle nAChR. In the process of generating a gateable version of AChBP by attaching it to the transmembrane domains of the 5-HT$_3$R, the Sine lab found that they had to replace several of the AChBP loops with loops of the 5-HT$_3$R.[11] This caused a drop in ACh and Nic affinity, consistent with the idea that they had disrupted the ligand-binding template of AChBP in order to make it gate.[11] Thus, it seems that gateable receptors may necessarily be more flexible than the AChBP structures would imply.

Images of the *Torpedo electroplax* nAChR from cryo-electron microscopy studies by Unwin support this image of a flexible receptor.[24] This receptor is highly homologous to the muscle receptor that we study functionally. Figure 22 shows an image of its two ligand binding sites (most functional receptors bind only two agonist molecules, studies of solubilized α7 extracellular domains have observed the binding of five ACh molecules, but the gateable receptor is only believed to bind two molecules).[8] One can see that the binding pocket is more open, not the well-formed box of AChBP. The parallel orientation of the two Trps, observed by both ourselves and McCammon, is also present. Since any deviation form the accepted crystallographic template is always treated as suspect in homology modeling, the visual similarity between these images and some of our structures is comforting. However, this is not born out statistically. A comparison of the Unwin structure to some of our unliganded representative structures shows that it bears the most similarity to Box E, but that the RMSD between these two structures is still 3.0 Å.

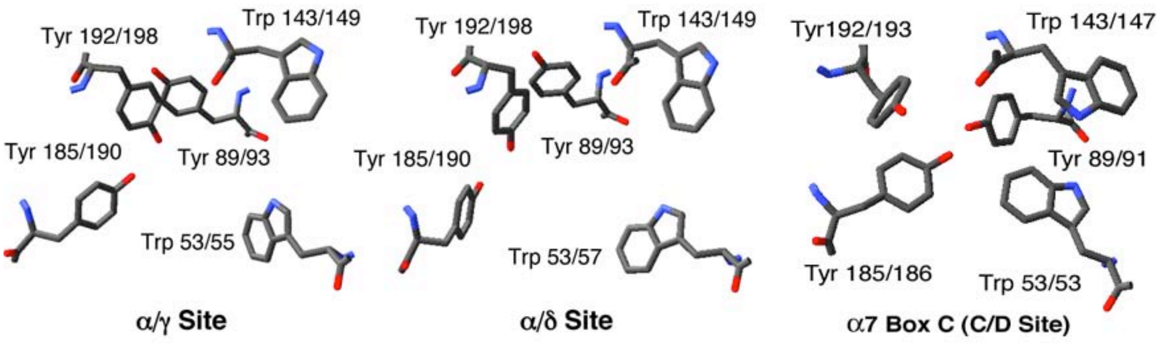


Figure 22. Images of the Muscle nAChR Box Residues from an Unliganded Cryoelectron Microscopy Image by Unwin. (PDB code 2BG9)

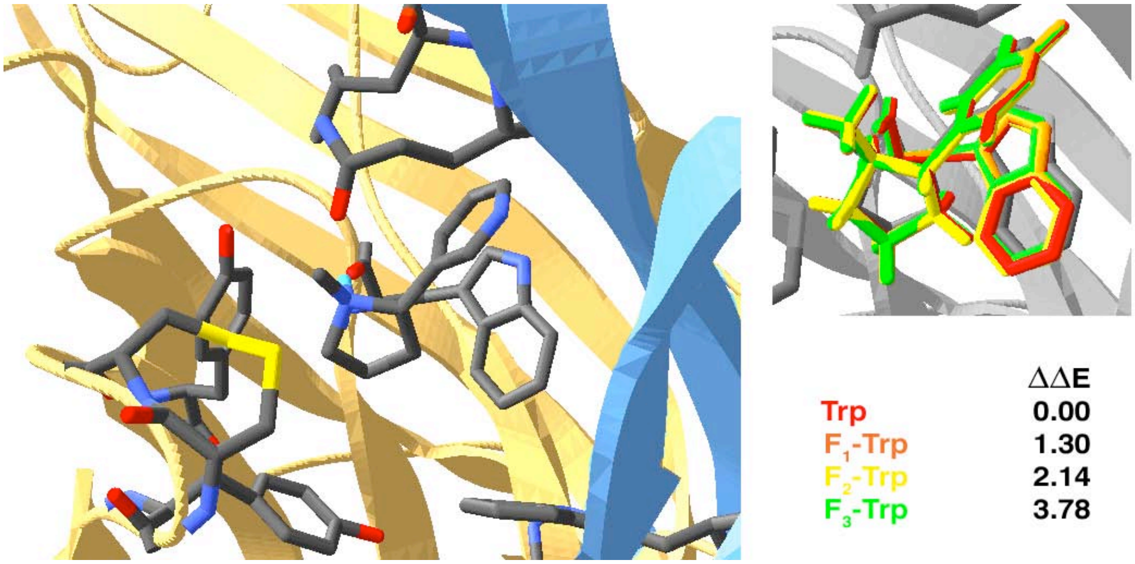
The above comparisons to experimetntal data and analogous simulations by acknowledged experts in the field gave us confidence in taking our representative structures on to do QMMM calculations of ligand binding affinities.

QMMM Calculations

For an initial test of our ability to perform QMMM calculations on ligand binding to the nAChR, two structures were selected from the representative structures generated for analysis of the MD runs. The structure representative of Nic binding to α7 Box A and the structure representative of CCh binding to α7 Box D were pared down to the protein dimer which formed the binding interface of interest with the ligand. These dimers were prepared for QMMM calculations by neutralizing most of the charge outside of a ~ 10 Å sphere around Trp 147 and the ligand. Then QMMM minimizations were performed in which only the ligand and Trp 147 were treated with quantum theory (HF/lacvp*) and allowed to move. The rest of the protein was frozen and treated with OPLS 2001, a force field commonly used in molecular mechanics simulations of proteins.[25] We minimized the geometries of, and calculated energies for, the ligand-bound structures of each binding site dimer with Trp, F_1-Trp, F_2-Trp, or F_3-Trp at position 147 of subunit A or D. We also minimized and calculated energies for the apo structures by deleting the ligand from the structure. We calculated the change in energy for each fluorinated Trp relative to the wild type residue with the following formula, shown for F_1-Trp in the α7 A/B binding site with Nic as an example:

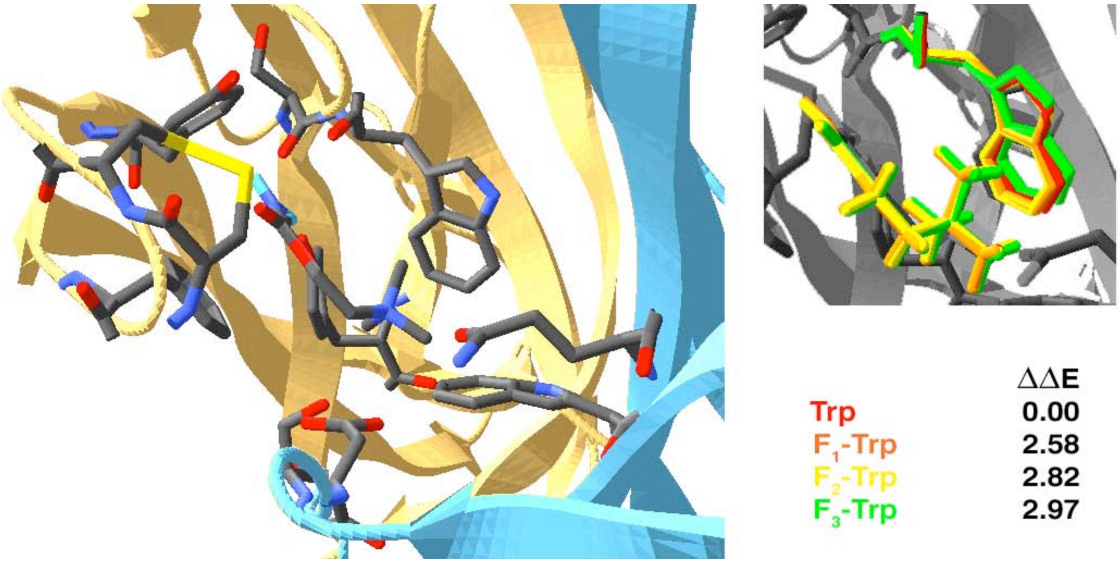
$$\Delta\Delta E_{\text{F1-Trp}} = [E_{\alpha7(\text{A-F1-Trp147/B})+\text{Nic}} - E_{\alpha7(\text{A-F1-Trp147/B})}] - [E_{\alpha7(\text{A-Trp147/B})+\text{Nic}} - E_{\alpha7(\text{A-Trp147/B})}]$$

These calculated energies are equivalent to the changes in EC_{50} that we observe in our fluorination plots. The results of these two sets of calculations are summarized in Figures 23 and 24.



	ΔΔE
Trp	0.00
F_1-Trp	1.30
F_2-Trp	2.14
F_3-Trp	3.78

Figure 23. QMMM Calculations of the Effects of Fluorination on Nic Binding at the A/B Interface of the α7 Pentamer from the Third MD Run. A subunit is shown in yellow, B subunit in blue. Residues within 5 Å of the ammonium group of Nic are shown. Left: Image of the binding site from the initial structure taken from the MD trajectory. Top Right: Overlayed structures of Nic and Trp 147 (or F_n-Trp mutant) shown colored as listed in bottom right. Bottom Right: Changes in Nic binding energy upon Trp fluorination. Energies in kcal/mol.



	ΔΔE
Trp	0.00
F_1-Trp	2.58
F_2-Trp	2.82
F_3-Trp	2.97

Figure 24. QMMM Calculations of the Effects of Fluorination on CCh Binding at the D/E Interface of the α7 Pentamer from the Third MD Run. D subunit is shown in yellow, E subunit in blue. Residues within 5 Å of the ammonium group or carbonyl oxygen of CCh are shown. Left: Image of the binding site from the initial structure taken from the MD trajectory. Top Right: Overlayed structures of CCh and Trp 147 (or F_n-Trp mutant) shown colored as listed in bottom right. Bottom Right: Changes in CCh binding energy upon Trp fluorination. Energies in kcal/mol.

The two ligands show very different responses to fluorination of Trp 147. For Nic, each successive fluorination produces an at least 0.8 kcal mol decrease in binding energy, consistent with a cation-π interaction. Although the increments are not precisely equal, they are comparable in magnitude. For CCh, on the other hand, the change in CCh binding energy between Trp and F_1-Trp is very large, but the changes upon incorporation of F_2-Trp and F_3-Trp are only marginally larger. This is not consistent with a cation-π interaction and is similar to the changes we observed experimentally in the Nic fluorination plots.

The binding geometry of Nic is somewhat unexpected in that it is the C7 and C8 methylene units that contact the Trp ring (See Figure 4 for atom designations), rather than the groups α to the pyrrolidine nitrogen, which we usually think of as more cationic. An examination of the partial charges on these atoms in Figure 4 (charges summed into heavy atoms) shows that those groups carry significant partial charges of 0.05 and 0.11 respectively, although these are admittedly smaller than the charges on the *N*-methyl group (0.27). To confirm that the effects on Nic binding were really the result of perturbing a cation-π interaction, we performed the analogous calculations on a neutral, deprotonated nicotine. (These were single point calculations.) The results are shown in Table 3. They indicate that for neutral nicotine, binding to the fluorinated tryptophans is actually more energetically favorable than binding to Trp itself (possibly due to basis set superposition effects).

Table 3. Trp Fluorination Effects on Neutral Nic Binding

	ΔΔE (kcal/mol)
Trp	0.00
F_1-Trp	- 0.21
F_2-Trp	- 1.83
F_3-Trp	- 2.35

These computational models of Trp fluorination are not consistent with the definition of cholinergic agonists as having a cation-π interaction and nicotinic agonists as lacking a cation-π interaction. However, they are consistent with a model of the effects of Trp fluorination that depends on the affinity of the agonist. At the human α7 nAChR, ACh is a ten-fold weaker agonist than Nic, and CCh is typically a ten-fold weaker agonist than ACh at neuronal nAChRs, although a CCh binding affinity value does not seem to be available in the literature for the α7 receptor.[26] Replacing Trp with F_1-Trp has an initial effect on the binding, but subsequent fluorination has little effect. In keeping with our "change in binding mechanism" explanation for a "thresholding" behavior in the fluorination plot, a greater repositioning of the Trp sidechain can be seen for the CCh calculations than for Nic. This change is still fairly small. The distance of the CCh ammonium nitrogen to the centroid of the six-membered indole ring has only increased from 4.7 Å for Trp to 4.8 Å for F_3-Trp. Studies of tetramethylammonium binding to benzene have shown that small changes in the orientation of the alkylammonium cation can have large energetic effects.[28] However, this does not seem to be the source of the apparent lack of a cation-π interaction in the QMMM calculations. To ascertain the inherent effects of the position of the CCh ammonium center on its binding affinity absent the protein, we removed the CCh and Trp moiety form each QMMM calculation and performed a gas-phase single point energy calculation on this geometry.

Table 4. Energies Calculated from Geometries Extracted from QMMM Calculations

	E(Trp)*	E(CCh)*	E(Trp+CCh)*	Difference*	Δ ΔE (kcal/mol)
Trp	-662.339287	-494.398024	-1156.760119	-0.022808	
F_1-Trp	-761.202049	-494.397788	-1255.619460	-0.019623	1.999
F_2-Trp	-860.049125	-494.397843	-1354.464937	-0.017969	3.036
F_3-Trp	-958.878190	-494.398268	-1453.292058	-0.015600	4.523

* HF/6-31G* energies in hartrees.

These calculations indicate that the inherent changes in the energies of the CCh binding geometries are consistent with a cation-π interaction and that the nonlinearity in the QMMM calculations is an effect of the protein. This effect must arise from the stability of the Trp moiety in the binding site because the energy of CCh does not factor into the QMMM calculations. To fully represent the binding site, we may need to allow movement of sidechains other than just those being treated with QM. These initial calculations represent the simplest QMMM strategy, it is not surprising that we would need to do more complicated calculations to really give the QMMM methodology a chance to work.

Conclusions and Future Work

In the preliminary experiments described here, we have shown that we can generate a homology model of a nicotinic receptor extracellular domain and subject it to MD simulations followed by QMMM calculations of the effects of unnatural mutations on binding affinities. Our MD simulations show an opening of the α7 binding pocket that is some ways consistent with Unwin's electron microscopy images of the Torpedo receptor. We are confident in our ability to model the receptor in dynamics simulations because of the similarity observed between our AChBP data and the computational and experimental work of McCammon and Taylor. While there may be some interest in comparing our results to known α7 nAChR agonist binding data, the primary goal of this line of research is to generate a homology model of the mouse muscle nAChR extracellular domain for which we have abundant unnatural mutagenesis data. One place to start would be to follow the protocol outlined here to generate a model of the muscle receptor using AChBP as a template. Another starting point would be to simply use Unwin's Torpedo nAChR structure either without modeling, or as a template for making a homology model of the mouse muscle nAChR. Both of these may be equally useful and valid approaches, one ostensibly starting from the desensitized state (AChBP) and one from the unliganded state (Unwin) of the receptor. Unfortunately, we have little data to compare them to, but the calculations performed on $\alpha7$ indicate that relatively subtle effects can lead to a deviation from the incremental change in cation-π binding observed in all of our small molecule calculations. Even those performed on the more elaborate Trp complexes still predicted a stepwise change in binding affinity for all agonists. We hope that employing MD and QMMM methodology with a model of the muscle receptor will help to unravel the mystery of what happens to nicotine in a receptor with a fluorinated Trp 149.

Materials and Methods

ab initio Small Molecule Calculations

Acetylcholine, (–) nicotine, (+) epibatidine, (–) epibatidine, 3-(1H-Indol-3-yl)-N-methyl-propionamide, 3-(5-Fluoro-1H-indol-3-yl)-N-methyl-propionamide, 3-(5, 7-Difluoro-1H-indol-3-yl)-N-methyl-propionamide, 3-(5, 6, 7-Trifluoro-1H-indol-3-yl)-N-methyl-propionamide, 3-(4, 5, 6, 7-Tetrafluoro-1H-indol-3-yl)-N-methyl-propionamide, the bound complexes shown in Figure 1, and their fluorinated equivalents were optimized at the HF/6-31G level of theory. The starting coordinates of were taken from the geometries given in Chapter 4. The optimized geometries were fully characterized as minima by frequency analysis. Energies were calculated at the HF/6-31G level. Basis set superposition error (BSSE) corrections were determined in the gas phase at the HF/6-31G level, using the counterpoise correction method of Boys and Bernardi.[29] Zero point energy (ZPE) corrections were included by scaling the ZPE correction given in the HF/6-31G level frequency calculation by the factor of 0.9135 given by Foresman and Frisch.[30] All calculations were carried out with the Gaussian 98 program.[31] Binding energies were determined by comparing the BSSE- and ZPE-corrected energies of the separately optimized ligand and tryptophan analog to the energy of the complex. The following optimized geometries are included as .log files on the hard drive that accompanies this thesis. The tables on the four pages following the file names show the primary data that are summarized in the fluorination plot in Figure 1.

File Naming Format: All files involving the indole moiety are denoted a7_Trp147 because the original Trp orientation came from Changeux and coworker's docked α7 nAChR homology model (See Chapter 4). Files denoted "631g" represent the HF/6-31G optimized geometry corresponding to the uncorrected energies in the following tables. CP1 refers to the Table entry "CPC_Dummy_Trp_w_Lig" in the energy tables, the couterpoise correction calculation in which the Trp moiety is replaced by dummy atoms and the ligand is left intact. CP2 refers to the Table entry "CPC_Dummy_Lig_w_Trp" in the energy tables, the couterpoise correction calculation in which the ligand is replaced by dummy atoms and the Trp moiety is left intact. CP3 refers to the Table entry "CPC_No_Trp_w_Lig" in the energy tables, the couterpoise correction calculation in which the Trp moiety is removed entirely and the ligand is left intact. CP4 refers to the Table entry "CPC_No_Lig_w_Trp" in the energy tables, the couterpoise correction calculation in which the ligand is removed entirely and the Trp moiety is left intact. "clepb" refers to (–) epibatidine and "clepb2" refers to (+) epibatidine.

Gaussian Log Files for Trp Fluorination Calculations

a7_F2Trp147_n_ach_631g.log
a7_F2Trp147_n_ach_cp1.log
a7_F2Trp147_n_ach_cp2.log
a7_F2Trp147_n_ach_cp3.log
a7_F2Trp147_n_ach_cp4.log
a7_F2Trp147_n_clepb2_631g.log
a7_F2Trp147_n_clepb2_cp1.log
a7_F2Trp147_n_clepb2_cp2.log
a7_F2Trp147_n_clepb2_cp3.log
a7_F2Trp147_n_clepb2_cp4.log
a7_F2Trp147_n_clepb_631g.log
a7_F2Trp147_n_clepb_cp1.log
a7_F2Trp147_n_clepb_cp2.log
a7_F2Trp147_n_clepb_cp3.log
a7_F2Trp147_n_clepb_cp4.log
a7_F2Trp147_n_nic_631g.log
a7_F2Trp147_n_nic_cp1.log
a7_F2Trp147_n_nic_cp2.log
a7_F2Trp147_n_nic_cp3.log
a7_F2Trp147_n_nic_cp4.log
a7_F2Trp_n_nic_631g.log
a7_F3Trp147_n_ach_631g.log
a7_F3Trp147_n_ach_cp1.log
a7_F3Trp147_n_ach_cp2.log
a7_F3Trp147_n_ach_cp3.log
a7_F3Trp147_n_ach_cp4.log
a7_F3Trp147_n_clepb2_631g.log
a7_F3Trp147_n_clepb2_cp1.log
a7_F3Trp147_n_clepb2_cp2.log
a7_F3Trp147_n_clepb2_cp3.log
a7_F3Trp147_n_clepb2_cp4.log
a7_F3Trp147_n_clepb_631g.log
a7_F3Trp147_n_clepb_cp1.log
a7_F3Trp147_n_clepb_cp2.log
a7_F3Trp147_n_clepb_cp3.log
a7_F3Trp147_n_clepb_cp4.log
a7_F3Trp147_n_nic_631g.log
a7_F3Trp147_n_nic_cp1.log
a7_F3Trp147_n_nic_cp2.log
a7_F3Trp147_n_nic_cp3.log
a7_F3Trp147_n_nic_cp4.log
a7_FTrp147_n_ach_631g.log
a7_FTrp147_n_ach_cp1.log
a7_FTrp147_n_ach_cp2.log
a7_FTrp147_n_ach_cp3.log
a7_FTrp147_n_ach_cp4.log
a7_FTrp147_n_clepb2_631g.log
a7_FTrp147_n_clepb2_cp1.log
a7_FTrp147_n_clepb2_cp2.log
a7_FTrp147_n_clepb2_cp3.log
a7_FTrp147_n_clepb2_cp4.log
a7_FTrp147_n_clepb_631g.log
a7_FTrp147_n_clepb_cp1.log
a7_FTrp147_n_clepb_cp2.log
a7_FTrp147_n_clepb_cp3.log
a7_FTrp147_n_clepb_cp4.log
a7_FTrp147_n_nic_631g.log
a7_FTrp147_n_nic_cp1.log
a7_FTrp147_n_nic_cp2.log
a7_FTrp147_n_nic_cp3.log
a7_FTrp147_n_nic_cp4.log
a7_Trp147_n_ach_631g.log
a7_Trp147_n_ach_cp1.log
a7_Trp147_n_ach_cp2.log
a7_Trp147_n_ach_cp3.log
a7_Trp147_n_ach_cp4.log
a7_Trp147_n_clep2_cp1.log
a7_Trp147_n_clep2_cp2.log
a7_Trp147_n_clep2_cp3.log
a7_Trp147_n_clep2_cp4.log
a7_Trp147_n_clepb2_631g.log
a7_Trp147_n_clepb2_cp1.log
a7_Trp147_n_clepb2_cp2.log
a7_Trp147_n_clepb2_cp3.log
a7_Trp147_n_clepb2_cp4.log
a7_Trp147_n_clepb_631g.log
a7_Trp147_n_clepb_cp1.log
a7_Trp147_n_clepb_cp2.log
a7_Trp147_n_clepb_cp3.log
a7_Trp147_n_clepb_cp4.log
a7_Trp147_n_nic_631g.log
a7_Trp147_n_nic_cp1.log
a7_Trp147_n_nic_cp2.log
a7_Trp147_n_nic_cp3.log
a7_Trp147_n_nic_cp4.log
ach_631g.log
clepb2_631g.log
clepb_631g.log
F2Trp147_n_631g.log
F3Trp147_n_631g.log
FTrp147_n_631g.log
nic_631g.log
Trp147_n_631g.log

Acetylcholine Cation-π Binding Energies

Geometry	**Trp***	**(kcal/mol)**	**F-Trp***	**(kcal/mol)**	**F_2-Trp***	**(kcal/mol)**	**F_3-Trp***	**(kcal/mol)**
Ligand	-478.1380		-478.1380		-478.1380		-478.1380	
Ligand_ZPE_Correction	0.2528		0.2528		0.2528		0.2528	
Lig/Trp	-1124.2726		-1223.0936		-1321.9115		-1420.7139	
Lig/Trp_ZPE_Correction	0.5180		0.5089		0.4999		0.4911	
Trp	-646.0922		-744.9165		-843.7375		-942.5425	
Trp_ZPE_Correction	0.2626		0.2537		0.2448		0.2360	
Binding_E		-26.61		-24.54		-22.59		-20.96
Binding_E_ZPE_Correction		-25.11		-23.13		-21.26		-19.67
CPC_Dummy_Trp_w_Lig	-478.1365		-478.1361		-478.1358		-478.1356	
CPC_Dummy_Lig_w_Trp	-646.0941		-744.9187		-843.7398		-942.5447	
CPC_No_Trp_w_Lig	-478.1351		-478.1351		-478.1346		-478.1344	
CPC_No_Lig_w_Trp	-646.0900	-3.45	-744.9134	-3.95	-843.7362	-3.01	-942.5413	-2.89
CPC_Binding_E		-21.66		-19.17		-18.25		-16.78
ΔΔ_Binding_E		0.00		-2.49		-3.41		-4.88
EC_{50}		1.20		4.70		13.00		34.00
log(Mutant_EC_{50}/WT_EC_{50})		0.00		0.59		1.03		1.45

* Energies in Hartrees.
ZPE = Zero Point Energy
CPC = Counterpoise Correction

(–) Nicotine Cation-π Binding Energies

Geometry	Trp*	(kcal/mol)	F-Trp*	(kcal/mol)	F_2-Trp*	(kcal/mol)	F_3-Trp*	(kcal/mol)
Ligand	-495.8964		-495.8964		-495.8964		-495.8964	
Ligand_ZPE_Correction	0.2611		0.2611		0.2611		0.2611	
Lig/Trp	-1142.0438		-1240.8656		-1339.6829		-1438.4848	
Lig/Trp_ZPE_Correction	0.5262		0.5171		0.5081		0.4993	
Trp	-646.0922		-744.9165		-843.7375		-942.5425	
Trp_ZPE_Correction	0.2626		0.2537		0.2448		0.2360	
Binding_E		-34.64		-33.07		-30.75		-28.80
Binding_E_ZPE_Correction		-33.20		-31.71		-29.49		-27.54
CPC_Dummy_Trp_w_Lig	-495.8958		-495.8957		-495.8958		-495.8959	
CPC_Dummy_Lig_w_Trp	-646.0938		-744.9173		-843.7384		-942.5432	
CPC_No_Trp_w_Lig	-495.8950		-495.8951		-495.8952		-495.8952	
CPC_No_Lig_w_Trp	-646.0901	-2.82	-744.9132	-2.95	-843.7348	-2.64	-942.5394	-2.82
CPC_Binding_E		-30.38		-28.76		-26.86		-24.72
ΔΔ_Binding_E		0.00		-1.62		-3.52		-5.66
EC_{50}		45.00		130.00		172.00		188.00
log(Mutant_EC_{50}/WT_EC_{50})		0.00		0.46		0.58		0.62

* Energies in Hartrees.
ZPE = Zero Point Energy
CPC = Counterpoise Correction

(–) Epibatidine Cation-π Binding Energies

Geometry	**Trp***	**(kcal/mol)**	**F-Trp***	**(kcal/mol)**	**F_2-Trp***	**(kcal/mol)**	**F_3-Trp***	**(kcal/mol)**
Ligand	-992.6185		-992.6185		-992.6185		-992.6185	
Ligand_ZPE_Correction	0.2597		0.2597		0.2597		0.2597	
Lig/Trp	-1638.7737		-1737.5925		-1836.4097		-1935.2112	
Lig/Trp_ZPE_Correction	0.5253		0.5159		0.5069		0.4980	
Trp	-646.0922		-744.9165		-843.7375		-942.5425	
Trp_ZPE_Correction	0.2626		0.2537		0.2448		0.2360	
Binding_E		-39.53		-36.08		-33.70		-31.52
Binding_E_ZPE_Correction		-37.81		-34.61		-32.33		-30.20
CPC_Dummy_Trp_w_Lig	-992.6157		-992.6172		-992.6173		-992.6176	
CPC_Dummy_Lig_w_Trp	-646.0924		-744.9172		-843.7382		-942.5431	
CPC_No_Trp_w_Lig	-992.6156		-992.6159		-992.6160		-992.6162	
CPC_No_Lig_w_Trp	-646.0888	-2.32	-744.9130	-3.45	-843.7343	-3.26	-942.5392	-3.33
CPC_Binding_E		-35.49		-31.16		-29.07		-26.87
ΔΔ_Binding_E		0.00		-4.33		-6.42		-8.62
EC_{50}		0.72		3.50		7.50		15.00
log(Mutant_EC_{50}/WT_EC_{50})		0.00		0.69		1.02		1.32

* Energies in Hartrees.

ZPE = Zero Point Energy

CPC = Counterpoise Correction

(+) Epibatidine Cation-π Binding Energies

Geometry	Trp*	(kcal/mol)	F-Trp*	(kcal/mol)	F_2-Trp*	(kcal/mol)	F_3-Trp*	(kcal/mol)
Ligand	-992.6185		-992.6185		-992.6185		-992.6185	
Ligand_ZPE_Correction	0.2597		0.2597		0.2597		0.2597	
Lig/Trp	-1638.7722		-1737.5925		-1836.4097		-1935.2112	
Lig/Trp_ZPE_Correction	0.5249		0.5160		0.5070		0.4986	
Trp	-646.0922		-744.9165		-843.7375		-942.5425	
Trp_ZPE_Correction	0.2626		0.2537		0.2448		0.2360	
Binding_E		-38.59		-36.08		-33.70		-31.52
Binding_E_ZPE_Correction		-37.10		-34.59		-32.29		-29.87
CPC_Dummy_Trp_w_Lig	-992.6165		-992.6177		-992.6177		-992.6176	
CPC_Dummy_Lig_w_Trp	-646.0922		-744.9170		-843.7379		-942.5407	
CPC_No_Trp_w_Lig	-992.6160		-992.6164		-992.6164		-992.6164	
CPC_No_Lig_w_Trp	-646.0890	-2.32	-744.9129	-3.39	-843.7341	-3.20	-942.5362	-3.58
CPC_Binding_E		-34.78		-31.20		-29.09		-26.30
ΔΔ_Binding_E		0.00		-3.57		-5.68		-8.48
EC_{50}		0.72		3.50		7.50		15.00
log(Mutant_EC_{50}/WT_EC_{50})		0.00		0.69		1.02		1.32

* Energies in Hartrees.
ZPE = Zero Point Energy
CPC = Counterpoise Correction

α7 Homology Modeling Protocol (a7_model3a-i)

Two previous attempts were made to model the α7 receptor extracellular domain using AChBP as a template. Both previous models were generated in a similar fashion, but in the first attempt, our sequence alignment was different than that given in Figure 3, and in the second attempt, we did not freeze or restrain the identical and homologous residues as we do here. Thus, this version is designated as a7_model3 below, although it is referred to as Model 1 in text. % indicates beginning of command line for Gromacs commands. The files listed are found in the a7_model3 folder on the accompanying hard drive.

Used Swiss PDB Viewer 3.7 (Swiss PDB) to create a file with the A subunit of the AChBP structure with Nic (1UW6) without Nic or water.

achbp_nic_template.pdb

Obtained the human α7 nAChR sequence from the NCBI database.

a7.txt and seq.txt

Built the initial α7 homology model with Prime in QSite with the alignment shown in Figure 3 by uploading achbp_nic_template.pdb and a7.txt. This required some manual adjustment of the alignment assigned by QSite. Extended model to ends of template, otherwise default build settings. Allowed sidechain movement during build. Exported this as a pdb file.

a7_model3a_initial.pdb

Disulfide cysteines renamed as Cyx by Prime, polar Hs added twice. Renamed to Cys and deleted extra Hs. Saved in Swiss PDB.

a7_model3b_CCfix.pdb

Built a pentamer in Swiss PDB by aligning identical (*) residues in a7_model3b to identical residues in A, B, C, D, and E subunits of achbp_nic_whole.pdb, the first pentamer in 1UW6.

a7_model3c_whole.pdb

Brought a7_model3c back to Prime. Predicted sidechains on all residues in all subunits except the identical and highly homologous residues denoted * and : in alignment. Exported structure.

a7_model3d_scref1.pdb

Performed loop refinement in Prime on the following loops:

Residues	Subunit	Residues	Subunit
24-26	A, B	117-128	A, B, E
38-42	A, E	133-135	A
44-50	A, E	152-162	A, E
68-70	A	181-195	A, B
90-98	A, B	229-232	A

Loop refinement crashed on first loop of B subunit (after all of A). Exported structure.

a7_model3e_loopsref1.pdb

Saved only A subunit in Swiss PDB.

a7_model3e_single.pdb

Built a pentamer in Swiss PDB by alignment to subunits of achbp_nic_whole as above.

a7_model3f_whole.pdb

Imported this file into Gromacs v. 3.1.4. Generated Gromacs and topology files: a7_model3f.gro, a7_equil.top, a7_equil_A(B,C,D,E).itp, a7_equil_A(B,C,D,E)_pr.itp

```
%pdb2gmx -f a7_mode3f_whole.pdb -o a7_model3f.gro -p a7_equil.top -i a7_equil_pr.itp
```

Generated rectangular box with dimensions: 9.49600   9.36000   6.83900

```
%editconf -f a7_model3f.gro -o a7_box.gro -d 0.7
```

Changed parameters to generate hexagonal box: 9.496   9.36   6.839 0 0 0 0 0 3.9485

Added waters to hexagonal box.

```
%genbox -cp a7_box.gro -cs spc216.gro -o a7_box_h2o.gro -p a7_equil.top
```

Generated mdp file for genion run.

a7_equil.mdp

Generated start (.tpr) file for genion, looked for charge in output, -20.

```
%grompp -f a7_equil.mdp -c a7_box_h2o.gro -p a7_equil.top -o a7_equil.tpr
```

Made index file from a7_box_h2o.gro.

```
%make_ndx -f a7_box_h2o.gro -o a7_equil.ndx
```

Generated ions. 68 ions = 150 mM for ~ 9nm box, more Na to compensate for -20 charge.

```
%genion -s a7_equil.tpr -o a7_box_h2o_ion.gro -n a7_equil.ndx -g genion.log -np 68 -nn 48
```

Picked SOL index group to be replaced.

Edited a7_equil.top with:

```
; Include ion topologies
#include “ions.itp”
```

and changed SOL 16472-116 = 16356, added Na 68, Cl 48.

a7_model3g - Generated mdp file for minimization run 1. define = BKBN, freeze_grps = high_homology.

a7_min1.mdp

Made a new index file from a7_box_h2o_ion.gro. Added an index group of the homologous (:) and identical (*) residues for the A subunit. Combined these to form high_homology index group.

```
%make_ndx -f  a7_box_h2o_ion.gro -o a7_equil.ndx
```

Definitions of identical residues for index group in a7_index

Generated position restraints for backbone.

```
%genpr -f a7_box_h2o_ion.gro -n a7_equil.ndx -o res_bkbn.itp -fc 1000 1000 1000
```

Added restraint statement to a7_equil_A(B,C,D,E).itp files:

```
#ifdef BKBN
#include "res_bkbn.itp"
#endif
```

Generated start file for minimization 1 (a7_model3g).

```
%grompp -f a7_min1.mdp -c a7_box_h2o_ion.gro  -n a7_equil.ndx -p a7_equil.top  -o a7_min1.tpr
```

Minimization run 1.

```
%mdrun -s a7_min1.tpr -o a7_min1.trr -x a7_min1.xtc -c a7_min1.gro -e a7_min1.edr -g a7_min1.log &
```

a7_min1 (a7_model3g) did not converge after 5000 steps. See a7_min1.log.

a7_model3h - Generated mdp file for minimization run 2. define = BKBN, freeze_grps = high_homology.

```
a7_min2.mdp
```

Generated position restraints for backbone.

```
%genpr -f a7_min1.gro -n a7_equil.ndx -o res_bkbn.itp -fc 500 500 500
```

Generated start file for minimization 2 (a7_model3h).

```
%grompp -f a7_min2.mdp -c a7_min1.gro  -n a7_equil.ndx -p a7_equil.top  -o a7_min2.tpr
```

Minimization run 2 (a7_model3h).

```
%mdrun -s a7_min2.tpr -o a7_min2.trr -x a7_min2.xtc -c a7_min2.gro -e a7_min2.edr -g a7_min2.log &
```

a7 _min2 (a7_model3h) did not converge after 5000 steps. See a7_min2.log.

a7_model3i - Generated mdp file for minimization run 3. define = BKBN, freeze_grps = identity.

```
a7_min3.mdp
```

Generated position restraints for backbone.

```
%genpr -f a7_min1.gro -n a7_equil.ndx -o res_bkbn.itp -fc 500 500 500
```

Generated start file for minimization 2 (a7_model3g).

```
%grompp -f a7_min3.mdp -c a7_min2.gro  -n a7_equil.ndx -p a7_equil.top  -o a7_min3.tpr
```

Minimization run 3 (a7_model3i).

```
%mdrun -s a7_min3.tpr -o a7_min3.trr -x a7_min3.xtc -c a7_min3.gro -e a7_min3.edr -g a7_min3.log &
```

a7_min3 (a7_model3i) converged after 4 steps. See a7_min3.log

Exported protein to pdb file.

```
editconf  -f a7_min3.gro -n a7_equil.ndx –o a7_model3i_min3.pdb
```

a7_model3j – This was the first set of minimizations and first MD run with model3. An identical protocol was run with a7_model3k, which differs from a7_model3j only in that no Nic is present. The binding site waters (two per site) were also included in a7_model3k. Therefore, the steps pertaining to Nic are not present and the charge balance with counterions uses 68 Na^+ ions. With these exceptions, all steps described here were performed on the nicotine-less a7_model3k with “model3k” substituted into the commands in place of “model3j.”

Constructed symmetric pentamer from A subunit of a7_model3i in Swiss PDB.

a7_model3j.pdb

Placed two binding site water molecules into model3j: A 1090, B 1045, C 1045, C 1052, C 1081, D 1042, E 1039, E 1041, E 1063, A 1048.

a7_model3j_nic.pdb

Saved nicotine molecules to 5nic.pdb

5nic.pdb

Used PRODRG to produce nicotine coordinates in Gromacs format and nic+1.itp file. Generated a different set of Gromacs coordinates for each nicotine molecule to preserve orientation relative to pentamer. Saved as nic+1_gromacs.doc. Charges in nic+1.itp replaced with charges from HF/6-31G** ChelpG calculations.

davapc1.bioch.dundee.ac.uk/programs/prodrg/prodrg.html

nic+1_gromacs.doc and nic+1.itp

Returned a7_model3j.pdb to Gromacs. Generated a7_model3j.gro.

```
%pdb2gmx –f a7_mode3j.pdb –o a7_model3j.gro –p a7_model3j.top –i a7_model3j_pr.itp
```

Added nicotine and water coordinates to a7_model3j.gro by hand, renumbering residues and atoms.

a7_model3j_nic.gro

Added to a7_model3j.top:

```
#Include “nic+1.itp”
```

Generated rectangular box with dimensions: 9.35200 9.14800 7.94400

```
%editconf  -f a7_model3j.gro –o a7_model3j_box.gro  -d 0.7
```

Attempted to change parameters to hexagonal box, triclinic box too skewed, kept rectangular.

Added waters to box.

```
%genbox -cp a7_model3j_box.gro -cs spc216.gro -o a7_model3j_box_h2o.gro -p
a7_model3j.top
```

Generated mdp file for genion run: a7_model3j.mdp

Generated start (.tpr) file for genion, looked for charge in output, -15.

```
%grompp –f a7_model3j.mdp –c a7_model3j_box_h2o.gro –p a7_model3j.top –o a7_model3j.tpr
```

Made index file from a7_model3j_box_h2o.gro.

```
%make_ndx –f a7_model3j_box_h2o.gro –o a7_model3j.ndx
```

Generated ions. 63 ions=150 mM for ~ 9nm box, more Na to compensate for –15 charge. Added fewer Na because of cationic Nic.

```
%genion –s a7_model3j.tpr –o a7_model3j_box_h2o_ion.gro –n a7_model3j.ndx –g genion.log
–np 63 –nn 48
```

Picked SOL index group to be replaced.

Edited a7_model3j.top with:

```
; Include ion topologies
#include “ions.itp”
```

and changed SOL 16818-111 = 16707, added Na 63, Cl 48.

a7_model3j_min1 – identical residues frozen, protein and Nic restrained at fc = 1000.

Generated mdp file for minimization run. define = BACKBONE, freeze_grps = identity.

```
a7_model3j_min1.mdp
```

Made a new index file from a7_model3j_box_h2o_ion.gro. Added an index group of identical (*) residues for a7_model3j. Added an index group of Protein + NCT.

```
%make_ndx –f a7_model3j_box_h2o_ion.gro –o a7_model3j.ndx
```

Generated position restraints for protein and nicotine.

```
%genpr –f a7_model3j_box_h2o_ion.gro –n a7_model3j.ndx –o res_bkbn.itp –fc 1000 1000 1000
```

Added restraint statement to a7_model3j_A(B,C,D,E).itp files:

```
#ifdef BACKBONE
#include "res_bkbn.itp"
#endif
```

Generated start file for minimization.

```
%grompp –f a7_model3j_min1.mdp –c a7_model3j_box_h2o_ion.gro –n a7_model3j.ndx –p
a7_model3j.top -o a7_model3j_min1.tpr
```

Minimization run.

```
%mdrun –s a7_model3j_min1.tpr –o a7_model3j_min1.trr –x a7_model3j_min1.xtc –c
a7_model3j_min1.gro –e a7_model3j_min1.edr –g a7_model3j_min1.log &
```

a7_model3j_min1 did not converge after 5000 steps. See a7_model3j_min1.log.

a7_model3j_min2 – identical residues frozen, backbone restrained at fc = 1000.

Generated position restraints for backbone.

```
%genpr –f a7_model3j_min1.gro –n a7_model3j.ndx –o res_bkbn.itp –fc 1000 1000 1000
```

Generated mdp file for minimzation run. define = BACKBONE, freeze_grps = identity.

```
a7_model3j_min2.mdp
```

```
%grompp –f a7_model3j_min2.mdp –c a7_model3j_min1.gro –n a7_model3j.ndx –p a7_model3j.top -o a7_model3j_min2.tpr
%mdrun –s a7_model3j_min2.tpr –o a7_model3j_min2.trr –x a7_model3j_min2.xtc –c a7_model3j_min2.gro –e a7_model3j_min2.edr –g a7_model3j_min2.log &
```

a7_model3j_min2 did not converge after 5000 steps. See a7_model3j_min2.log.

a7_model3j_min3 – identical residues frozen, backbone restrained at fc = 500.

Generated position restraints for backbone.

```
%genpr –f a7_model3j_min2.gro –n a7_model3j.ndx –o res_bkbn.itp –fc 500 500 500
```

Generated mdp file for minimzation run. define = BACKBONE, freeze_grps = identity.

```
a7_model3j_min3.mdp
%grompp –f a7_model3j_min3.mdp –c a7_model3j_min2.gro –n a7_model3j.ndx –p a7_model3j.top -o a7_model3j_min3.tpr
%mdrun –s a7_model3j_min3.tpr –o a7_model3j_min3.trr –x a7_model3j_min3.xtc –c a7_model3j_min3.gro –e a7_model3j_min3.edr –g a7_model3j_min3.log &
```

a7_model3j_min3 did not converge after 5000 steps. See a7_model3j_min3.log.

a7_model3j_min4 – identical residues restrained at fc = 1000.

Generated position restraints for identical residues.

```
%genpr –f a7_model3j_min3.gro –n a7_model3j.ndx –o res_bkbn.itp –fc 1000 1000 1000
```

Generated mdp file for minimzation run. define = BACKBONE.

```
a7_model3j_min4.mdp
%grompp –f a7_model3j_min4.mdp –c a7_model3j_min3.gro –p a7_model3j.top -o a7_model3j_min4.tpr
%mdrun –s a7_model3j_min4.tpr –o a7_model3j_min4.trr –x a7_model3j_min4.xtc –c a7_model3j_min4.gro –e a7_model3j_min4.edr –g a7_model3j_min4.log &
```

a7_model3j_min4 converged after 2610 steps. See a7_model3j_min4.log.

a7_model3j_min5 – identical residues restrained at fc = 500.

Generated position restraints for identical residues.

```
%genpr –f a7_model3j_min4.gro –n a7_model3j.ndx –o res_bkbn.itp –fc 500 500 500
```

Generated mdp file for minimzation run. define = BACKBONE.

```
a7_model3j_min5.mdp
%grompp –f a7_model3j_min5.mdp –c a7_model3j_min4.gro –p a7_model3j.top -o a7_model3j_min5.tpr
%mdrun –s a7_model3j_min5.tpr –o a7_model3j_min5.trr –x a7_model3j_min5.xtc –c a7_model3j_min5.gro –e a7_model3j_min5.edr –g a7_model3j_min5.log &
```

a7_model3j_min5 converged after 1093 steps. See a7_model3j_min5.log.

a7_model3j_min6 – Protein –H restrained at fc = 1000.

Generated position restraints for identical residues.

```
%genpr –f a7_model3j_min5.gro –n a7_model3j.ndx –o res_bkbn.itp –fc 1000 1000 1000
```

Generated mdp file for minimzation run. define = BACKBONE.

```
a7_model3j_min6.mdp
%grompp –f a7_model3j_min6.mdp –c a7_model3j_min5.gro –p a7_model3j.top -o a7_model3j_min6.tpr
%mdrun –s a7_model3j_min6.tpr –o a7_model3j_min6.trr –x a7_model3j_min6.xtc –c a7_model3j_min6.gro –e a7_model3j_min6.edr –g a7_model3j_min6.log &
```

a7_model3j_min6 converged after 2260 steps. See a7_model3j_min6.log.

a7_model3j_min7 – unrestrained.

Generated mdp file for minimzation run. define = .

```
a7_model3j_min7.mdp
%grompp –f a7_model3j_min7.mdp –c a7_model3j_min6.gro –p a7_model3j.top -o a7_model3j_min7.tpr
%mdrun –s a7_model3j_min7.tpr –o a7_model3j_min7.trr –x a7_model3j_min7.xtc –c a7_model3j_min7.gro –e a7_model3j_min7.edr –g a7_model3j_min7.log &
```

a7_model3j_min7 converged after 323 steps. See a7_model3j_min7.log.

a7_model3k_min1 – Identical residues frozen, protein and Nic restrained at fc = 1000. Did not converge after 5000 steps. See a7_model3k_min1.log.

a7_model3k_min2 – Identical residues frozen, backbone restrained at fc = 1000. Did not converge after 5000 steps. See a7_model3k_min2.log.

a7_model3k_min3 – Identical residues frozen, backbone restrained at fc = 500. Converged after 1220 steps. See a7_model3k_min3.log.

a7_model3k_min4 – Identical residues restrained at fc = 1000. Converged after 434 steps. See a7_model3k_min4.log.

a7_model3k_min5 – Identical residues restrained at fc = 500. Converged after 79 steps. See a7_model3k_min5.log.

a7_model3k_min6 – Protein –H restrained at fc = 1000. Converged after 14 steps. See a7_model3k_min6.log.

a7_model3k_min7 – Unrestrained. Converged after 59 steps. See a7_model3k_min7.log.

Both a7_model3j_min7.gro and a7_model3k_min7.gro were taken on to MD runs as described below. These runs were abandoned because of concerns about the position of residues 93 – 104.

a7_model4 – As noted in the text, the pore vestibule of a7_model3j and k was occluded by an inward motion of the residues in the β4-β5 loop (relative to their positions in the initial alignment), so we generated a fourth generation model in which these residues were moved to positions where their backbone would better align with AChBP. This is referred to as Model 2 in text. Nicotine was added and this was minimized in a procedure identical to that for a7_model3j. Thus, this protocol is abbreviated and is given only so that the reader can understand the origin of the names of the included files.

Replaced residues 93-104 in a7_model3i_min3.pdb with coordinates from a7_model3a_initial.pdb.

a7_model4a_single.pdb

Constructed symmetric pentamer in Swiss PDB.

a7_model4b_whole.pdb

Placed two binding site water molecules into model4b (same waters used in model3j).

a7_model4b_h2o.pdb

Moved a7_model4b_h2o.pdb to Gromacs. Generated.gro,.top, and A, B, C, D, E.itp files

```
%pdb2gmx –f a7_model4b_h2o.pdb –o a7_nic.gro –p a7_nic.top –i a7_nic.itp
```

Added nicotine coordinates to a7_nic.gro by hand, using nic+1_gromacs.doc. Renumbered # of atoms at top of file. Renumbered Nic atom and molecule numbers appropriately. Added #Include "nic+1.itp" to a7_nic.top.

a7_nic.gro

Generated rectangular box with dimensions: 9.50300 9.38200 7.93000

```
%editconf –f a7_nic.gro –o a7_nic_box.gro –d 0.7
```

Added waters to box.

```
%genbox –cp a7_nic_box.gro –cs spc216.gro –o a7_nic_box_h2o.gro –p a7_nic.top
```

Generated mdp file for genion run: a7_nic.mdp

Generated start (.tpr) file for genion, looked for charge in output, -15.

```
%grompp –f a7_nic.mdp –c a7_nic_box_h2o.gro –p a7_nic.top –o a7_nic.tpr
```

Made index file from a7_nic_box_h2o.gro.

```
%make_ndx –f a7_nic_box_h2o.gro –o a7_nic.ndx
```

Generated ions. 63 ions=150 mM for ~ 9nm box, more Na to compensate for –15 charge.

```
%genion –s a7_nic.tpr –o a7_nic_box_h2o_ion.gro –n a7_nic.ndx –g genion.log –np 63 –nn 48
```

Picked SOL index group to be replaced. Edited a7_nic.top and changed SOL 17234-111 = 17123, added Na 63, Cl 48.

a7_nic_min1 – Identical residues frozen, protein and Nic restrained at fc = 1000. Did not converge after 5000 steps. See a7_nic_min1.log.

a7_nic_min2 – Identical residues frozen, backbone restrained at fc = 1000. Converged after 3019 steps. See a7_nic_min2.log.
a7_nic_min3 – Identical residues frozen, backbone restrained at fc = 500. Converged after 90 steps. See a7_nic_min3.log.
a7_nic_min4 – Identical residues restrained at fc = 1000. Converged after 1520 steps. See a7_nic_min4.log.
a7_nic_min5 – Identical residues restrained at fc = 500. Converged after 375 steps. See a7_nic_min5.log.
a7_nic_min6 – Protein –H restrained at fc = 1000. Converged after 67 steps. See a7_nic_min6.log.
a7_nic_min7 – Unrestrained. Did not converge after 5000 steps. See a7_nic_min7.log.

Carbamoyl choline was added in a manner analogous to the Nic insertion. Since the 1UV6 structure does not have CCh in every binding site, the C binding site was aligned with each of the binding sites of a7_model4b_whole.pdb to generate a five-fold symmetric pdb file with five CChs. No binding site waters were included with the CCh files. As with Nic, the CCh ligands were inserted by hand into the a7_cch.gro file. The other procedures related to a7_nic were applied to a7_cch, except that once the MD runs were begun, the CCh runs, both a7 and achbp (below) tended to crash. This was due to large intramolecular forces due to the relatively large partial charges on the carbamoyl end of CCh. Thus these were reduced until the runs performed successfully, and the charges in Figure 3 were used in cch.itp.
a7_cch_min1 – Identical residues frozen, protein and Nic restrained at fc = 1000. Did not converge after 5000 steps. See a7_cch_min1.log.
a7_cch_min2 – Identical residues frozen, backbone restrained at fc = 1000. Did not converge after 5000 steps. See a7_cch_min2.log.
a7_cch_min3 – Identical residues frozen, backbone restrained at fc = 500. Did not converge after 500 steps. See a7_cch_min3.log. Only three rounds of minimization, se below.

a7_min1 – Identical residues frozen, protein and Nic restrained at fc = 1000. Did not converge after 5000 steps. See a7_min1.log.
a7_min2 – Identical residues frozen, backbone restrained at fc = 1000. Converged after 2315 steps. See a7_min2.log.
a7_min3 – Identical residues frozen, backbone restrained at fc = 500. Converged after 11 steps. See a7_min3.log.

a7_min4 – Identical residues restrained at fc = 1000. Converged after 610 steps. See a7_min4.log.
a7_min5 – Identical residues restrained at fc = 500. Converged after 15 steps. See a7_min5.log.
a7_min6 – Protein –H restrained at fc = 1000. Converged after 15 steps. See a7_min6.log.
a7_min7 – Unrestrained. Did not converege after 5000 steps. See a7_min7.log.

achbp_nic_min1 – Identical residues frozen, protein and Nic restrained at fc = 1000. Did not converge after 5000 steps. See achbp_nic_min1.log.
achbp_nic_min2 – Identical residues frozen, backbone restrained at fc = 1000. Did not converge after 5000 steps. See achbp_nic_min2.log.
achbp_nic_min3 – Identical residues frozen, backbone restrained at fc = 500. Did not converge after 5000 steps. See achbp_nic_min3.log.
achbp_nic_min4 – Identical residues restrained at fc = 1000. Did not converge after 5000 steps. See achbp_nic_min4.log.
achbp_nic_min5 – Identical residues restrained at fc = 500. Did not converge after 5000 steps. See achbp_nic_min5.log.
achbp_nic_min6 – Protein –H restrained at fc = 1000. Converged after 3900 steps. See achbp_nic_min6.log.
achbp_nic_min7 – Unrestrained. Converged after 1062 steps. See achbp_nic_min7.log.

achbp_cch_min1 – Identical residues frozen, protein and Nic restrained at fc = 1000. Did not converge after 5000 steps. See achbp_cch_min1.log.
achbp_cch_min2 – Identical residues frozen, backbone restrained at fc = 1000. Did not converge after 5000 steps. See achbp_cch_min2.log.
achbp_cch_min3 – Identical residues frozen, backbone restrained at fc = 500. Did not converge after 5000 steps. See achbp_cch_min3.log.
achbp_cch_min4 – Identical residues restrained at fc = 1000. Converged after 2538 steps. See achbp_cch_min4.log.
achbp_cch_min5 – Identical residues restrained at fc = 500. Converged after 358 steps. See achbp_cch_min5.log.
achbp_cch_min6 – Protein –H restrained at fc = 1000. Converged after 250 steps. See achbp_cch_min6.log.
achbp_cch_min7 – Unrestrained. Converged after 101 steps. See achbp_cch_min7.log.

achbp_min1 – Identical residues frozen, protein and Nic restrained at fc = 1000. Did not converge after 5000 steps. See achbp_min1.log.
achbp_min2 – Identical residues frozen, backbone restrained at fc = 1000. Did not converge after 5000 steps. See achbp_min2.log.
achbp_min3 – Identical residues frozen, backbone restrained at fc = 500. Did not converge after 5000 steps. See achbp_min3.log.
achbp_min4 – Identical residues restrained at fc = 1000. Converged after 2987 steps. See achbp_min4.log.
achbp_min5 – Identical residues restrained at fc = 500. Converged after 150 steps. See achbp_min5.log.
achbp_min6 – Protein –H restrained at fc = 1000. Converged after 281 steps. See achbp_min6.log.
achbp_min7 – Unrestrained. Converged after 134 steps. See achbp_min7.log.

The min7 files from these minimizations were taken on to MD, with names carried over from these minimization runs (i.e. a7_model3j_md1a uses a7_model3j_min7.gro as input). Since rapid collapse of the "aromatic box" (defined in text) was observed for the a7_nic runs, a7_min3.gro and a7_nic_min3.gro were used as input files to initiate the a73 and a7_nic3 MD runs. For consistency, the a7_CCh run was only subjected to three rounds of minimization.

General MD Protocol

The following protocols were used for all MD simulations. The α7 simulation is described here, but all steps and filenames are applicable to the other MD runs through the simple replacement of "a7_nic" with one of the following: a7_model3j, a7_model3k, a7, a7_nic, a73, a7_nic3, a7_nonic, a7_cch, achbp, achbp_nic, achbp_nonic, or achbp_cch. For reference to the description of these runs in text: a7_model3j refers to the first MD run with Nic. a7_model_3k refers to the first MD run without Nic. a7 refers to the second MD run, and a7_nic to the second MD run with Nic. a73 and a7_nic3 refer to the third MD run. Only the second MD run for AChBP is saved and is listed as achbp and achbp_nic.
a7_nic_md1a - 50 ps. Annealing from 0 to 310 K over first 25 ps, held at 310 K for second 25 ps. Protein and ligand restrained with fc = 1000.
Generated a7_nic_md1a.mdp file for minimization run. See below.
Generated position restraints for protein and ligand (or just protein if applicable).

```
%genpr  -f a7_nic_min7.gro  -n a7_nic.ndx  -o a7_nic_res_bkbn.itp  -fc 1000 1000 1000
```

Generated start file for MD run.

```
%grompp -f a7_nic_md1a.mdp -c a7_nic_min7.gro -p a7_nic.top -o a7_nic_md1a.tpr
```

MD run started.

```
%mdrun -s a7_nic_md1a.tpr -o a7_nic_md1a.trr -c a7_nic_md1a.gro -e a7_nic_md1a.edr -g
a7_nic_md1a.log &
```

a7_nic_md1b - 50 ps. Backbone restrained with fc = 1000.

Generated a7_nic_md1b.mdp file for minimization run. See below.

Generated position restraints for protein backbone.

```
%genpr -f a7_nic_md1a.gro -n a7_nic.ndx -o a7_nic_res_bkbn.itp -fc 1000 1000 1000
```

Generated start file for MD run.

```
%grompp -f a7_nic_md1b.mdp -c a7_nic_md1a.gro -t a7_nic_md1a.trr -p a7_nic.top -o
a7_nic_md1b.tpr
```

MD run started.

```
%mdrun -s a7_nic_md1b.tpr -o a7_nic_md1b.trr -c a7_nic_md1b.gro -e a7_nic_md1b.edr -g
a7_nic_md1b.log &
```

a7_nic_md1c - 50 ps. Backbone restrained with fc = 500.

Generated a7_nic_md1c.mdp file for minimization run, identical to a7_nic_md1b.mdp except for start time, tinit = 100.0.

Generated position restraints for protein backbone.

```
%genpr -f a7_nic_md1b.gro -n a7_nic.ndx -o a7_nic_res_bkbn.itp -fc 500 500 500
```

Generated start file for MD run.

```
%grompp -f a7_nic_md1c.mdp -c a7_nic_md1b.gro -t a7_nic_md1b.trr -p a7_nic.top -o
a7_nic_md1c.tpr
```

MD run started.

```
%mdrun -s a7_nic_md1c.tpr -o a7_nic_md1c.trr -c a7_nic_md1c.gro -e a7_nic_md1c.edr -g
a7_nic_md1c.log &
```

a7_nic_md1 - 850 ps.

Generated a7_nic_md1.mdp file for minimization run, identical to a7_nic_md1c.mdp except that tinit = 150.0, nsteps = 425000 and “define = -dBACKBONE” has been deleted.

Generated start file for MD run.

```
%grompp -f a7_nic_md1.mdp -c a7_nic_md1c.gro -t a7_nic_md1c.trr -p a7_nic.top -o
a7_nic_md1.tpr
```

MD run started.

```
%mdrun -s a7_nic_md1.tpr -o a7_nic_md1.trr -c a7_nic_md1.gro -e a7_nic_md1.edr -g
a7_nic_md1.log &
```

All subsequent runs such as a7_nic_md2 are started in a fashion similar to a7_nic_md1.

nonic - Runs in which Nic was removed from the binding site of an MD run which had used Nic were initiated as follows (a7 given as an example):

Dumped protein coordinates (specified as index group) from 3500 ps snapshot from run with Nic to use as starting point.

```
%trjconv  -f a7_nic_md3.trr  -s a7_nic_md3.tpr –o a7_nonic_md3.gro –n a7_nic.ndx –dump 3500
```

Imported .itp and .top files from a7 run for use with a7_nonic. Deleted SOL and ions from .top file. After this, a7_nonic treated exactly like original a7 setup; box generated, waters added, and ions added (68 Na^+) to prepare for minimization and MD.

a7_nonic_min1 – Protein restrained at fc = 1000. Did not converge after 5000 steps. See a7_nonic_min1.log.

This was then taken on to a7_nonic_md1a and further MD as above.

The primary analysis tools and the selected options within these commands are listed here:

System energies were obtained as a function of time using the command g_energy with standard options.

RMSD values were calculated using g_rms. They were obtained for the protein by aligning the atoms in the "identical" index group in the index file (e.g. a7_nic.ndx) and calculating the RMSD of the protein or the "identical" group. RMSDs of box residues were calculated by aligning the index group for the given box (i.e. index group "box_a" in a7_nic.ndx) and then calculating the RMSD of "box_a." this was done for each box and these values were averaged in Microsoft Excel.

Sidechain plane angles were calculated using g_sgangle. Index groups were defined for each sidechain. Trp sidechains were defined by the CG, CZ3, and NE atoms. Tyr sidechains were defined by the CG, CE1, and CE2 atoms. g_sgangle was used with the option – "noone" to calculate interplane angles between two residues.

Clusters of structures were generated according to the RMSD of box residues using g_cluster. A RMSD cutoff of 0.07 nm was used and the gromos algorithm was used. If no ligand was present, the "box_a" index file (for example) was used. If ligand was present, a new index group was generated with the corresponding ligand by first generating an index group for the ligand "NCT1" and then combining this with "box_a" to form "box_a_NCT1." the log files from these clustering steps are included on the hard drive with the following naming convention: "a7_nic3_Acluster.log" is the file for Box A. The associated .pdb files are named according to the size of the clusters with 1 being largest. The .pdb file corresponds to the trajectory timepoint listed as the mean structure in the .log file.

QMMM Minimizations

All QMMM calculations were conducted within the Schrödinger suite of programs.[17, 32] Structures generated from the MD trajectories were saved in Swiss PdbViewer as files with the two subunits that contained the binding site of interest (for example, a7_nic3_Acluster1.pdb was saved as a new file with only the A and B subunits and NCT1).[18] This file was imported into Schrödinger and atom types in the ligand were corrected to the appropriate Macromodel atom types. The protein and ligand were then given “Hydrogen Treatment” with the default settings (OPLS 2001 compatible). The ligand was then selected in Glide and the Protein Preparation procedure was applied to neutralize charge outside an 8-12 Å sphere around the ligand and Trp 147. Since this did not neutralize the overall system, we manually protonated carboxylate residues and lysine residues to achieve a charge of + 1 with ligand present or 0 without ligand. The prepared protein was then subjected to QMMM minimizations in QSite with non-bonded cut-offs turned off, all atoms except the Trp of interest and/or ligand frozen, and a HF/lacvp* lvel of theory for the quantum portion. Only the Trp (147 for a7_AB_nic.pdb, and 775 for a7_DE_cch.pdb) and the ligand were treated with quantum theory, and the entire residue was treated. Changes in binding energies were calculated by subtracting the energy of the minimized protein with ligand from the energy of the minimized protein without ligand. The differences between this energy for the Trp-containing protein and the mutant proteins was taken to calculate the ΔΔE value that we report. Fluorination was carried out by changing Macromodel atom types in the “Build” menu. Neutralization of Nic was carried out by deleting the H on the pyrrolidine N, and converting this to a neutral N in Macromodel atom typing.

References

(1) Beene, D. L.; Brandt, G. S.; Zhong, W.; Zacharias, N. M.; Lester, H. A.; Dougherty, D. A., *Biochem.* **2002,** 41, 10262-10269 Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A., *Proc. Natl. Acad. Sci. USA* **1998,** 95, 12088-12093.

(2) Amer Chemical Soc Cashin, A. L.; Petersson, E. J.; Lester, H. A.; Dougherty, D. A., *J. Am. Chem. Soc.* **2005,** 127, 350-356.

(3) Ma, J. C.; Dougherty, D. A., *Chem. Rev.* **1997,** 97, 1303-1324.

(4) Corringer, P. J.; Le Novere, N.; Changeux, J. P., *Annu. Rev. Pharmacol. Toxicol.* **2000,** 40, 431-458.

(5) Le Novere, N.; Grutter, T.; Changeux, J. P., *Proc. Natl. Acad. Sci. USA* **2002,** 99, 3210-3215.

(6) Smit, A. B.; Syed, N. I.; Schaap, D.; van Minnen, J.; Klumperman, J.; Kits, K. S.; Lodder, H.; van der Schors, R. C.; van Elk, R.; Sorgedrager, B.; Brejc, K.; Sixma, T. K.; Geraerts, W. P., *Nature* **2001,** 411, 261-268.

(7) Spang, J. E.; Bertrand, S.; Westera, G.; Patt, J. T.; Schubiger, P. A.; Bertrand, D., *Chem. Biol.* **2000,** 7, 545-555.

(8) Corringer, P. J.; Bertrand, S.; Bohler, S.; Edelstein, S. J.; Changeux, J. P.; Bertrand, D., *J. Neurosci.* **1998,** 18, 648-657.

(9) Beene, D. L.; Dougherty, D. A.; Lester, H. A., *Curr. Opin. Neurobiol.* **2003,** 13, 264-270.

(10) Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van Der Oost, J.; Smit, A. B.; Sixma, T. K., *Nature* **2001,** 411, 269-276.

(11) Bouzat, C.; Gumilar, F.; Spitzmaul, G.; Wang, H. L.; Rayes, D.; Hansen, S. B.; Taylor, P.; Sine, S. M., *Nature* **2004,** 430, 896-900.

(12) Henchman, R. H.; Wang, H. L.; Sine, S. M.; Taylor, P.; McCammon, J. A., *Biophys. J.* **2003,** 85, 3007-3018.

(13) Henchman, R. H.; Wang, H. L.; Sine, S. M.; Taylor, P.; McCammon, J. A., *Biophys. J.* **2005,** 88, 2564-2576.

(14) Gao, F.; Bren, N.; Burghardt, T. P.; Hansen, S.; Henchman, R. H.; Taylor, P.; McCammon, J. A.; Sine, S. M., *J. Biol. Chem.* **2005,** 280, 8443-8451.

(15) Hansen, J. L.; Theilade, J.; Haunso, S.; Sheikh, S. P., *J. Biol. Chem.* **2004,** 279, 24108-24115.

(16) Sine, S. M.; Wang, H. L.; Bren, N., *J. Biol. Chem.* **2002,** 277, 29210-29223.

(17) Schrodinger, I. *Prime*, Portland, OR.

(18) Guex, N.; Peitsch, M. C., *Electrophoresis* **1997,** 18, 2714-2723.

(19) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; R., v. D.; Berendsen, H. J. C., *Gromacs User Manual Version 3.2*; 2004.

(20) Schuettelkopf, A. W.; F., v. A. D. M., *Acta Crystal.* **2004,** D60, 1355-1363.

(21) Celie, P. H. N.; van Rossum-Fikkert, S. E.; van Dijk, W. J.; Brejc, K.; Smit, A. B.; Sixma, T. K., *Neuron* **2004,** 41, 907-914.

(22) Schapira, M.; Abagyan, R.; Totrov, M., *BMC Struct. Biol.* **2002,** 2, 1-14.

(23) Wells, G. B.; Anand, R.; Wang, F.; Lindstrom, J., *J. Biol. Chem.* **1998,** 273, 964-973 West, A. P., Jr.; Bjorkman, P. J.; Dougherty, D. A.; Lester, H. A., *J. Biol. Chem.* **1997,** 272, 25468-25473.

(24) Unwin, N., *J. Mol. Biol.* **2005,** 346, 967-989.

(25) Damm, W.; Halgren, T. A.; Murphy, R. B.; Smondyrev, A. M.; Friesner, R. A.; Jorgensen, W. L., *Abstr. Pap. Am. Chem. Soc.* **2002,** 224, U471-U471 Halgren, T. A.; Murphy, R. B.; Jorgensen, W. L.; Friesner, R. A., *Abstr. Pap. Am. Chem. Soc.* **2000,** 220, U277-U277.

(26) Jensen, A. A.; Mikkelsen, I.; Frolund, B.; Brauner-Osborne, H.; Falch, E.; Krogsgaard-Larsen, P., *Mol. Pharmacol.* **2003,** 64, 865-875 Xiao, Y.; Baydyuk, M.; Wang, H. P.; Davis, H. E.; Kellar, K. J., *Bioorg. Med. Chem. Lett.* **2004,** 14, 1845-1848 Chavez-Noriega, L. E.; Crona, J. H.; Washburn, M. S.; Urrutia, A.; Elliott, K. J.; Johnson, E. C., *J. Pharmacol. Exp. Ther.* **1997,** 280, 346-356.

(27) Gallivan, J. P.; Dougherty, D. A., *J. Am. Chem. Soc.* **2000,** 122, 870-874.

(28) Felder, C.; Jiang, H.-L.; Zhu, W.-L.; Chen, K.-X.; Silman, I.; Botti, S. A.; Sussman, J. L., *J. Phys. Chem.* **2001,** 105, 1326-1333.

(29) Boys, S. F.; Bernardi, F., *Mol. Phys.* **1970,** 19, 553-557.

(30) Foresman, J. B.; Frisch, E., *Exploring Chemistry with Electronic Structure Methods*. ed.; Gaussian, Inc.: Pittsburgh, PA, 1996.

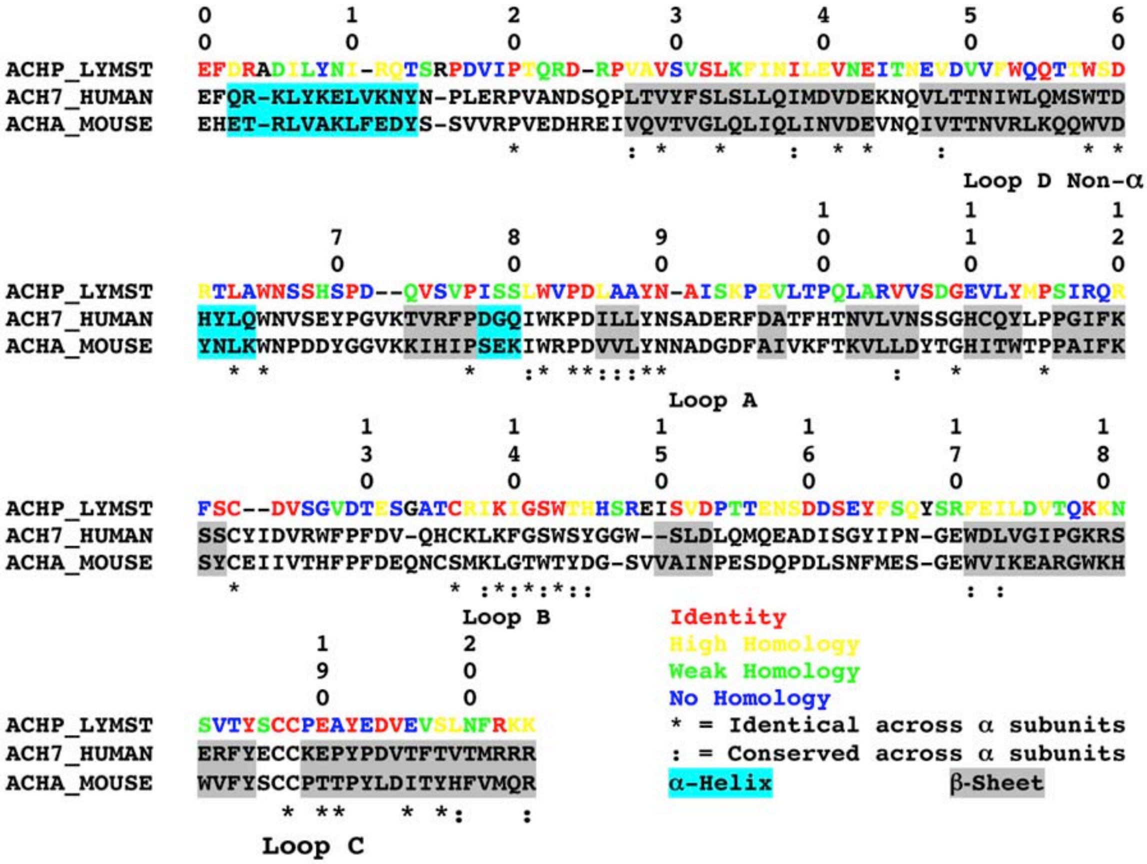
(31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; J. A. Montgomery, J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; *et. al. Gaussian 98 (Revision A.9)*, Gaussian, Inc.: Pittsburgh PA, 1998.

(32) Schrodinger, I. *QSite*, Portland, OR Schrodinger, I. *Glide*, Portland, OR.

Supporting Information

Sequence Alignment

The following alignment of the *Lymnea stygmala* AChBP, human α7 nAChR, and mouse muscle nAChR amino acid sequence is provided to assist in understanding the loop designations in Chapter 2 as well as the comparisons of the three receptors throughout Section 1.

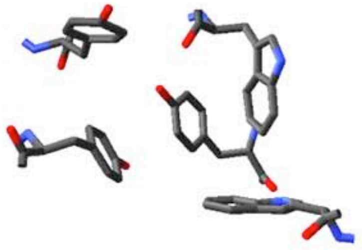


```
            0         1          2          3         4         5         6
            0         0          0          0         0         0         0
ACHP_LYMST  EFDRADILYNI-RQTSRPDVIPTQRD-RPVAVSVSLKFINILEVNEITNEVDVVFWQQTTWSD
ACH7_HUMAN  EFQR-KLYKELVKNYN-PLERPVANDSQPLTVYFSLSLLQIMDVDEKNQVLTTNIWLQMSWTD
ACHA_MOUSE  EHET-RLVAKLFEDYS-SVVRPVEDHREIVQVTVGLQLIQLINVDEVNQIVTTNVRLKQQWVD
                                 *       : *   *    :  * *    :         * *
                                                                Loop D Non-α
                                                      1         1         1
                     7           8         9          0         1         2
                     0           0         0          0         0         0
ACHP_LYMST  RTLAWNSSHSPD--QVSVPISSLWVPDLAAYN-AISKPEVLTPQLARVVSDGEVLYMPSIRQR
ACH7_HUMAN  HYLQWNVSEYPGVKTVRFPDGQIWKPDILLYNSADERFDATFHTNVLVNSSGHCQYLPPGIFK
ACHA_MOUSE  YNLKWNPDDYGGVKKIHIPSEKIWRPDVVLYNNADGDFAIVKFTKVLLDYTGHITWTPPAIFK
              * *             *   :* **:::**               :   *     *
                                            Loop A
                       1         1         1         1         1         1
                       3         4         5         6         7         8
                       0         0         0         0         0         0
ACHP_LYMST  FSC--DVSGVDTESGATCRIKIGSWTHHSREISVDPTTENSDDSEYFSQYSRFEILDVTQKKN
ACH7_HUMAN  SSCYIDVRWFPFDV-QHCKLKFGSWSYGGW--SLDLQMQEADISGYIPN-GEWDLVGIPGKRS
ACHA_MOUSE  SYCEIIVTHFPFDEQNCSMKLGTWTYDG-SVVAINPESDQPDLSNFMES-GEWVIKEARGWKH
              *              * :*:*:*::                         : :
                              Loop B        Identity
                    1         2             High Homology
                    9         0             Weak Homology
                    0         0             No Homology
ACHP_LYMST  SVTYSCCPEAYEDVEVSLNFRKK         * = Identical across α subunits
ACH7_HUMAN  ERFYECCKEPYPDVTFTVTMRRR         : = Conserved across α subunits
ACHA_MOUSE  WVFYSCCPTTPYLDITYHFVMQR         α-Helix            β-Sheet
                  * **    * *:    :
                Loop C
```

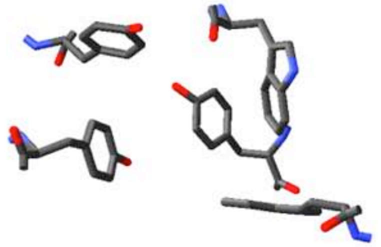
MD Clusters

The following twenty pages contain images of the "box" residues of AChBP or α7 from the various dynamics simulations described in Chapter 5. The mean structure from any cluster comprising greater than ten component structures is shown, taken from the last 1.5 ns of the 5 ns trajectory in the "+ CCh" cases, the last 1 ns of the 5 ns trajectory in the apo and "+ Nic" cases and the last 1850 ps of the "- Nic" trajectories. These clusters were generated with g_cluster in the GROMACs aanalysis suite of programs as described in the Materials and Methods section. The number of structures in each cluster is listed below that cluster. One structure is saved every 10 ns in the trajectory files.

AChBP Box A Clusters

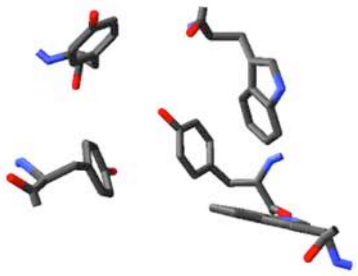


77 Structures



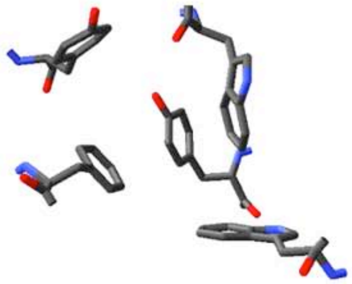
14 Structures

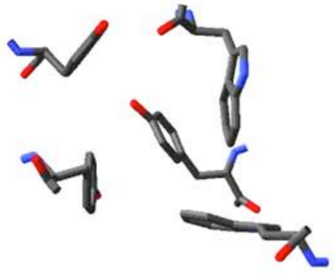
AChBP Box B Clusters



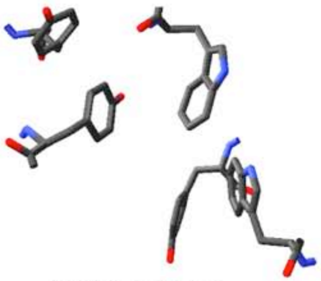
88 Structures

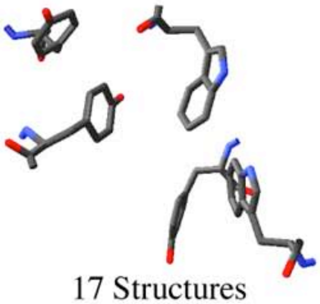
AChBP Box C Clusters

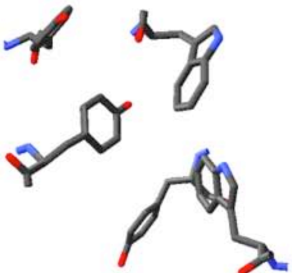

80 Structures

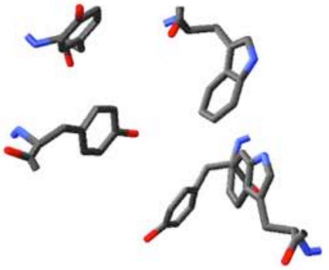

17 Structures

AChBP Box D Clusters

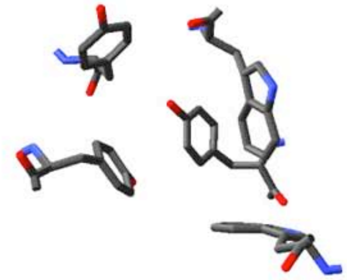

17 Structures


17 Structures

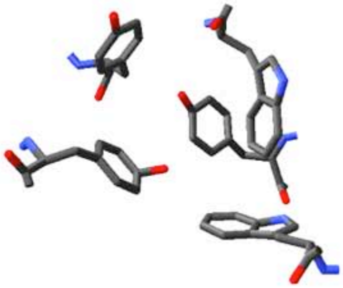

15 Structures


12 Structures

ACнBP Box E Clusters

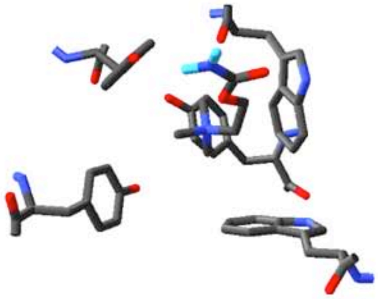


65 Structures

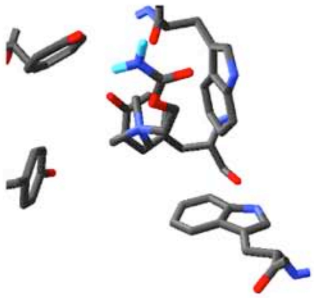


24 Structures

ACнBP + CCh Box A Clusters

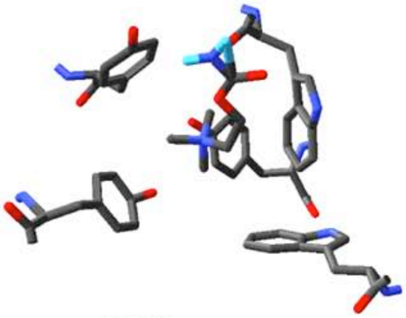


28 Structures

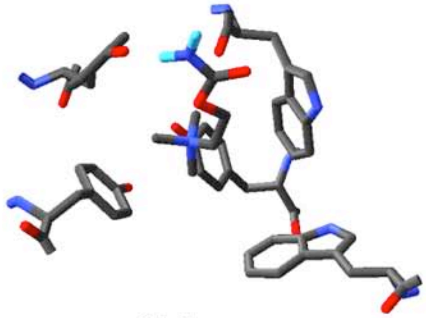


123 Structures

AChBP + CCh Box B Clusters

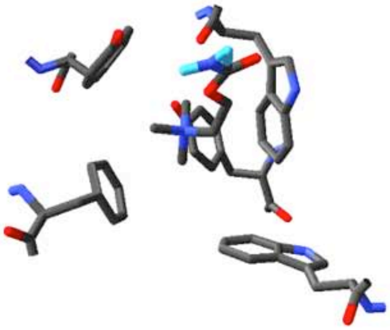


125 Structures



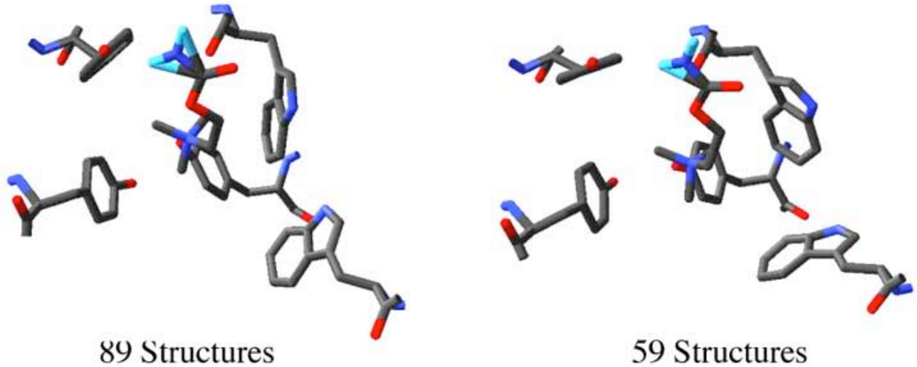
25 Structures

AChBP + CCh Box C Clusters



150 Structures

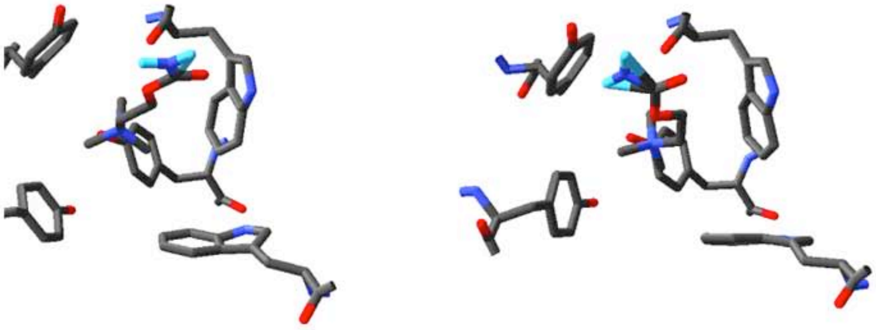
AChBP + CCh Box D Clusters



89 Structures

59 Structures

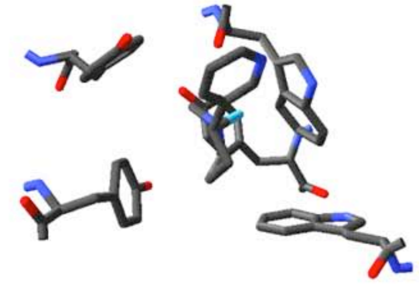
AChBP + CCh Box E Clusters



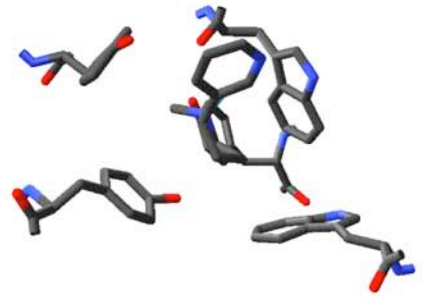
28 Structures

122 Structures

AChBP + Nic Box A Clusters

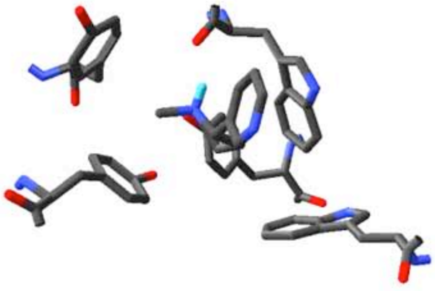


47 Structures

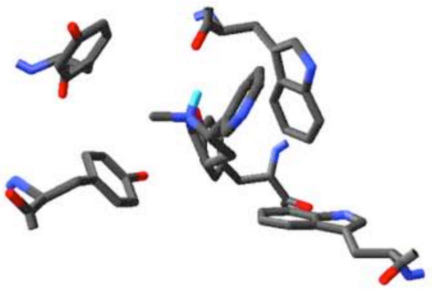


29 Structures

AChBP + Nic Box B Clusters

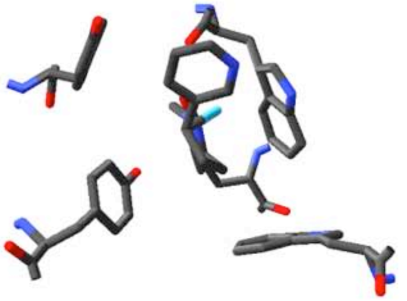


66 Structures

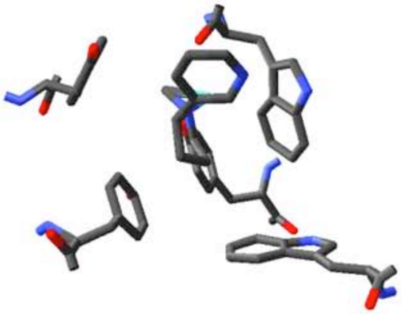


16 Structures

ACħBP + Nic Box C Clusters

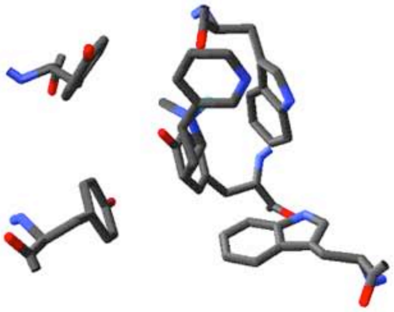


42 Structures

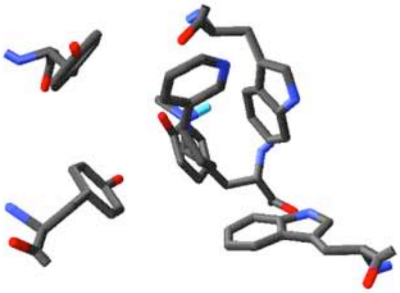


39 Structures

ACħBP + Nic Box D Clusters

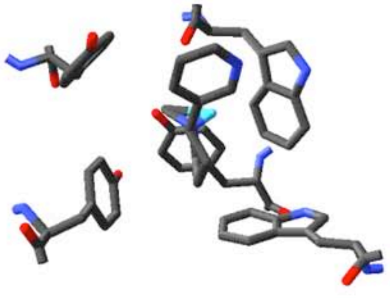


61 Structures

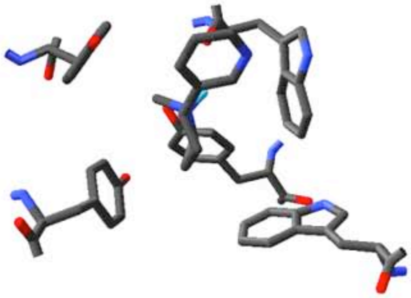


27 Structures

AChBP + Nic Box E Clusters

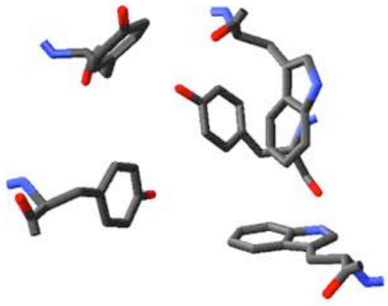


47 Structures



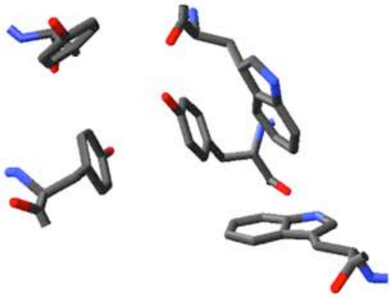
30 Structures

AChBP - Nic Box A Clusters



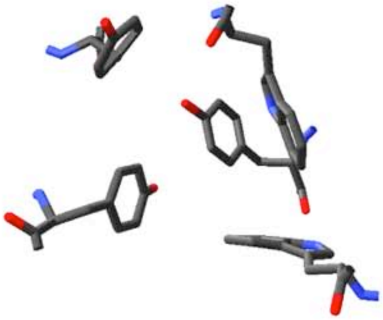
185 Structures

AChBP - Nic Box B Clusters



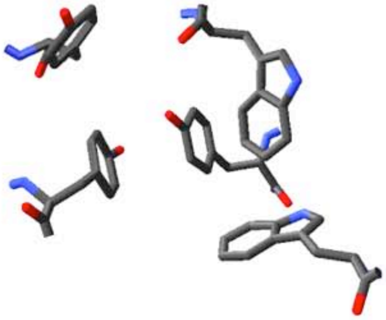
184 Structures

AChBP - Nic Box C Clusters



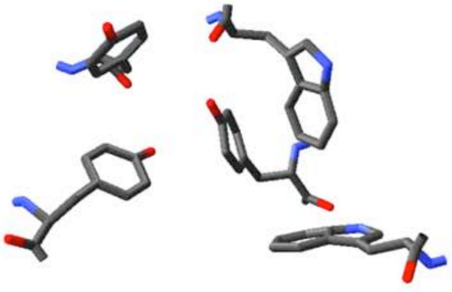
167 Structures

AChBP - Nic Box D Clusters

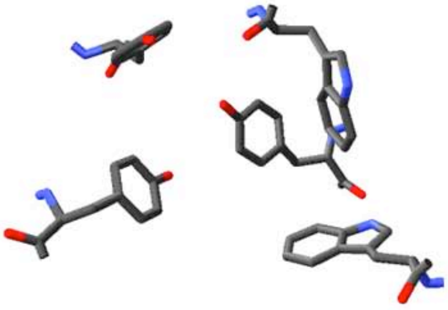


182 Structures

AChBP - Nic Box E Clusters

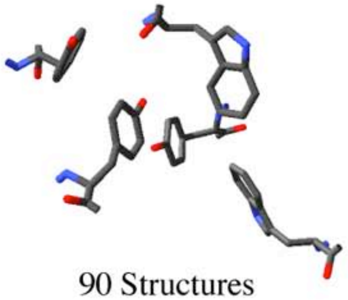


155 Structures

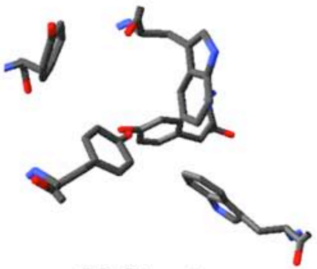


129 Structures

α7 Box A Clusters

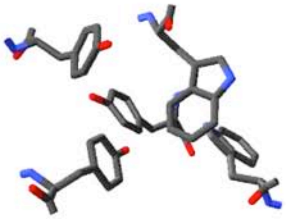


90 Structures

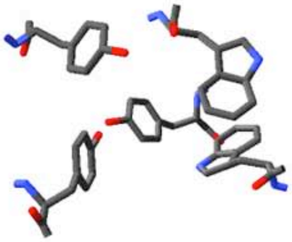


33 Structures

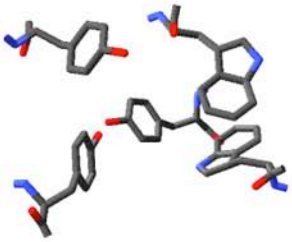
α7 Box B Clusters



63 Structures

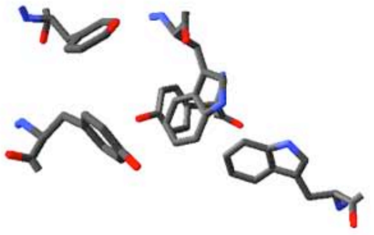


38 Structures

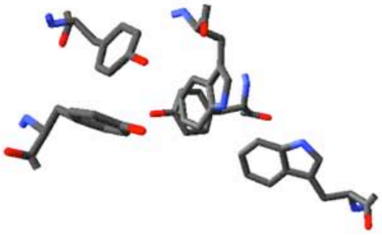


25 Structures

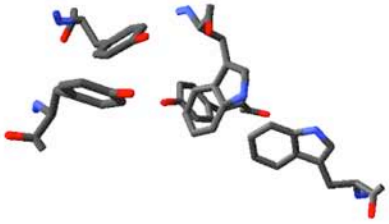
α7 Box C Clusters



35 Structures

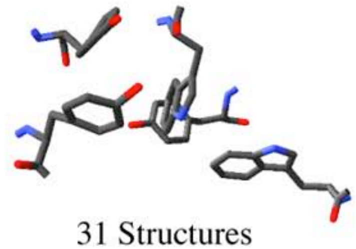


20 Structures

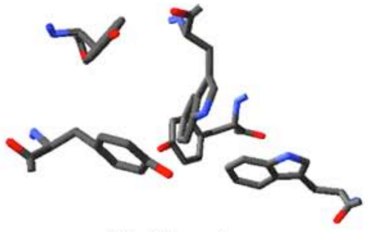


14 Structures

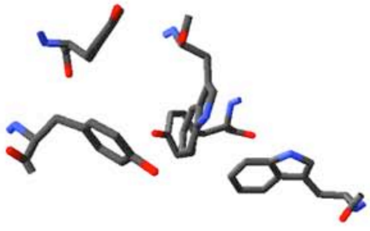
α7 Box D Clusters



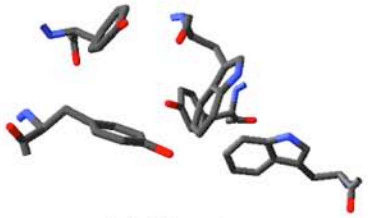
31 Structures



24 Structures

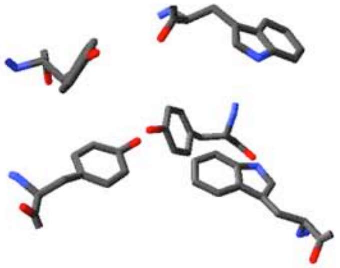


16 Structures

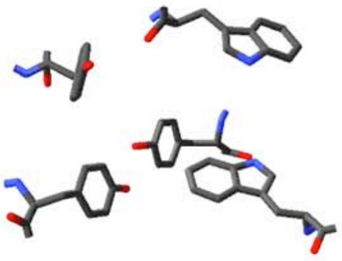


14 Structures

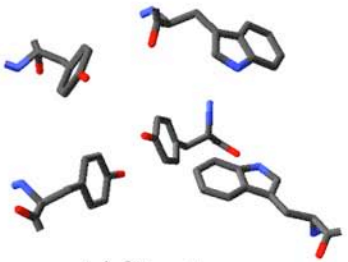
α7 Box E Clusters



81 Structures

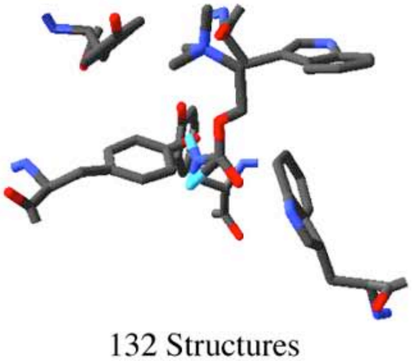


35 Structures



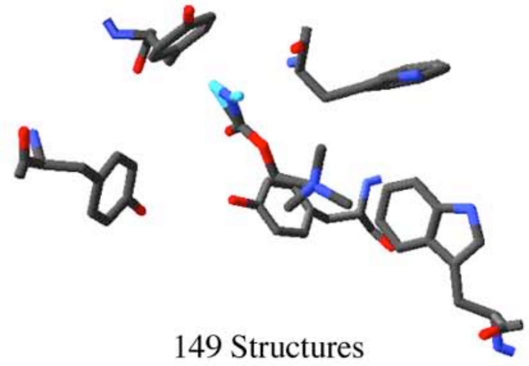
14 Structures

α7 + CCh Box A Clusters



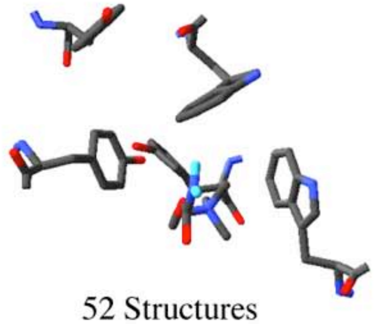
132 Structures

α7 + CCh Box B Clusters

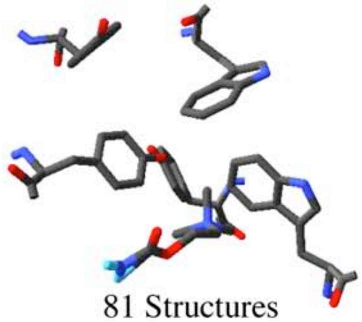


149 Structures

α7 + CCh Box C Clusters

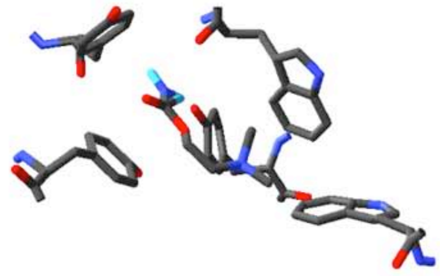


52 Structures

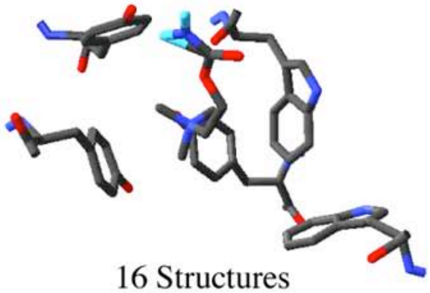


81 Structures

α7 + CCh Box D Clusters

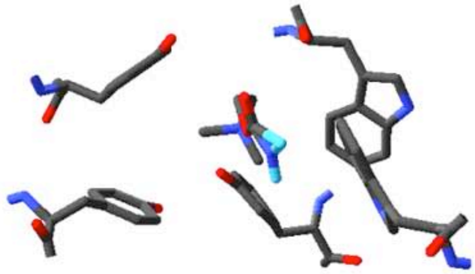


118 Structures



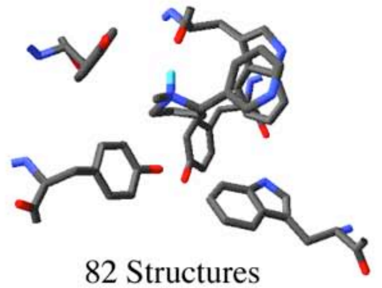
16 Structures

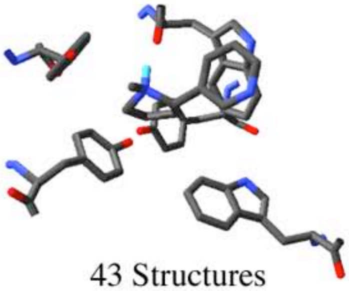
α7 + CCh Box E Clusters

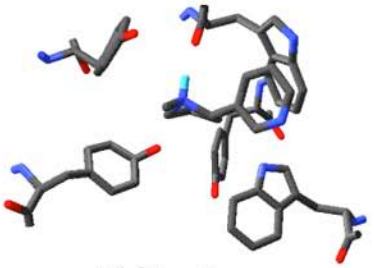


138 Structures

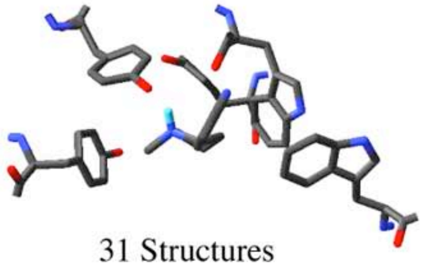
α7 + Nic Box A Clusters


82 Structures

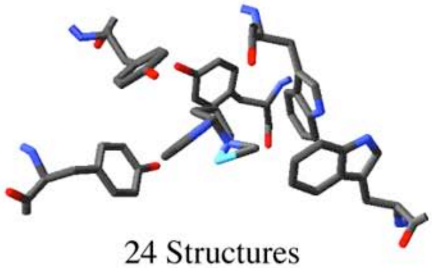

43 Structures


15 Structures

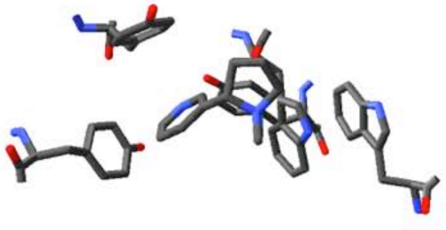
α7 + Nic Box B Clusters


31 Structures

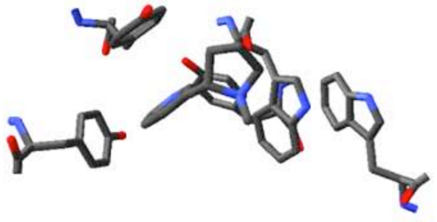

26 Structures


24 Structures

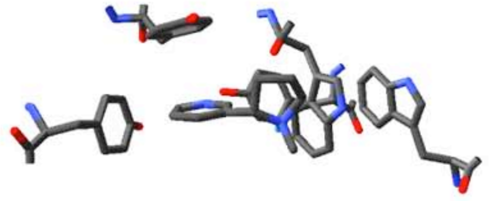
$\alpha 7$ + Nic Box C Clusters



80 Structures

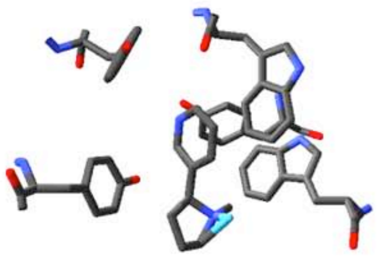


24 Structures

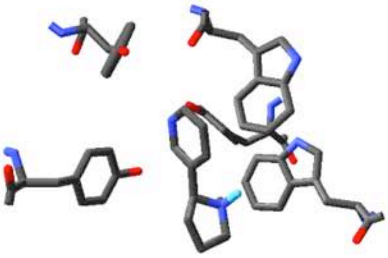


17 Structures

$\alpha 7$ + Nic Box D Clusters

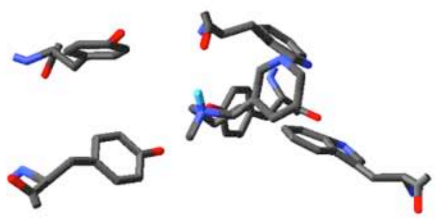


129 Structures

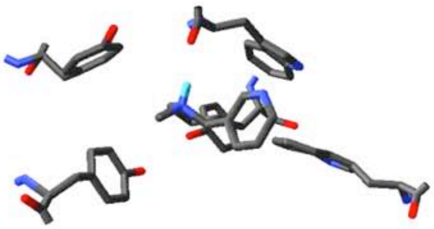


15 Structures

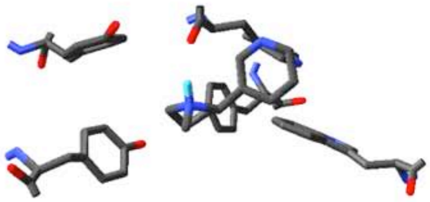
α7 + Nic Box E Clusters



36 Structures

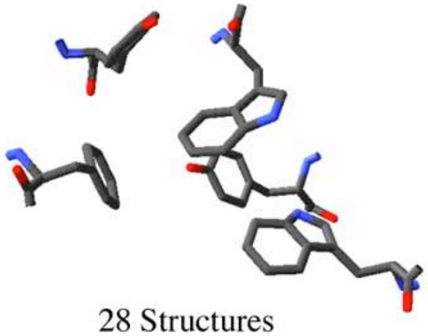


34 Structures

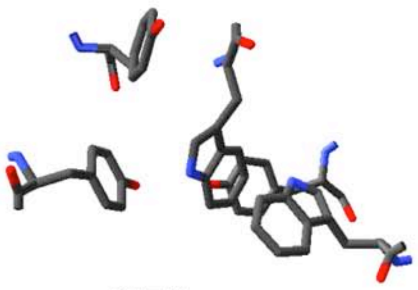


18 Structures

α7 - Nic Box A Clusters

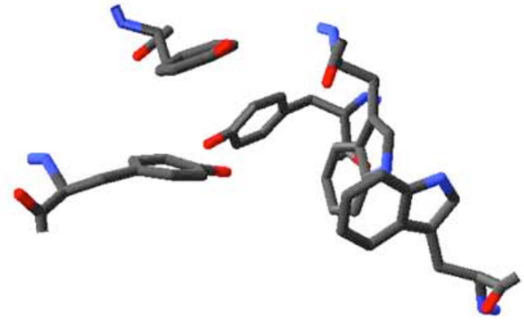


28 Structures



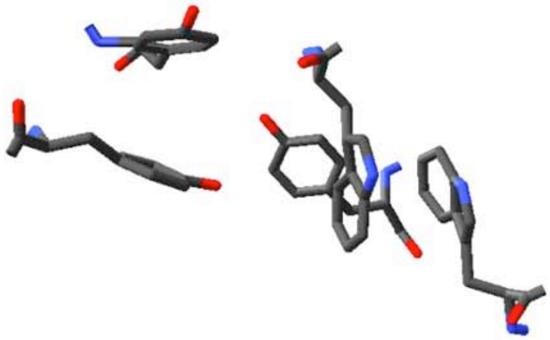
157 Structures

α7 - Nic Box B Clusters



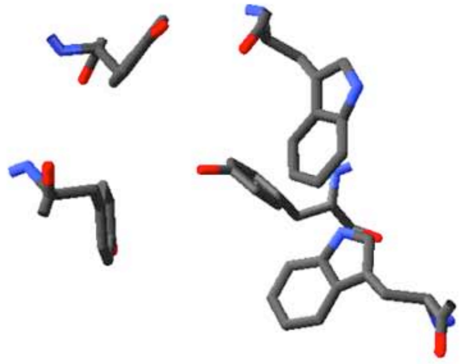
145 Structures

α7 - Nic Box C Clusters



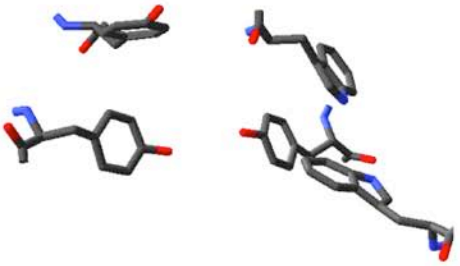
171 Structures

α7 - Nic Box D Clusters

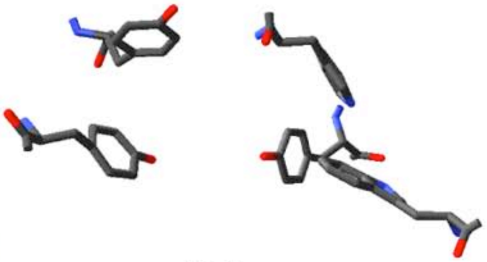


177 Structures

α7 - Nic Box E Clusters



78 Structures



86 Structures