Computational modelling and molecular dynamics simulations of ligand-gated ion channels

Shiva Amiri

Structural Bioinformatics and Computational Biochemistry unit,
Department of Biochemistry and
St. Hugh's College, Oxford

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

October 2006

Computational modelling and molecular dynamics simulations of ligand-gated ion channels

Shiva Amiri
Department of Biochemistry and St. Hugh's College, Oxford
Submitted for the degree of Doctor of Philosophy, Michaelmas 2006

Abstract

The Cys-loop *ligand-gated* super family of *ion channels* and related proteins have been studied using computational methods. Ligand-gated ion channels (LGIC) are pentameric, neurotransmitter-gated, ion selective receptors which play a key role in synaptic transmission. The nicotinic acetylcholine receptor (nAChR) is the archetypal member of the LGIC family. Found at neuronal-neuronal synapses and neuronal-muscular junctions, it is implicated in a variety of diseases and it plays a key role in nicotine addiction.

Computational techniques are used to investigate different properties of LGICs, such as ion selectivity, the location of the gate, the action of agonists, and the behaviour of the binding pocket. The aim was to apply existing computational methods and develop novel computational methodology to put forth hypotheses relating to both the general function and smaller scale behaviour of this class of membrane proteins.

Binding pocket dynamics of AChBP, a high resolution crystal structure homologue of the ligand binding domain of the nAChR, shows increased stability of the binding site in the presence of a ligand. Discreet zones of persistent water molecules were identified, which are involved in bridging ligands to residues in the binding site and may play a structural role. Ensemble docking on ligand-bound and ligand-free molecular dynamics (MD) trajectories illustrates the optimization of the binding site to a bound ligand.

To overcome the lack of complete structures of LGICs, a method was developed to generate model structures of membrane proteins by combining their separate domains. A model of the α7 nAChR was generated from homology models of its extracellular and transmembrane domains and the structure was used to analyze various properties such as pore profiles, electrostatics and conformational dynamics. *In silico* mutagenesis, together with MD and 'ensemble' docking, on the α7 nAChR investigated the role of key residues in the binding site. Residues were identified with computational techniques and verified with experimental work. This illustrates the utility of combining computational and experimental approaches.

During this work, the electron-microscopy structure of the *Torpedo marmorata* AChR became available. It agreed well with the previously modelled α7 nAChR model. The *Torpedo* AChR structure was used to make models of other LGICs. Coarse-grain MD allowed the identification of residues in the TM domain interacting with the lipid-bilayer. Born energy profiles through LGIC pores reveal that the EC domain plays a key role in ion selectivity.

Acknowledgements

This work and the past three years in Oxford would not have been possible without the encouragement, love and support of so many people. First of course are maman and baba who have been incredible parents, to them I dedicate this thesis. They have both lead by example by stressing the importance of education and independence. Baba always fostering science, knowledge, and invention and Maman for keeping me together and not letting a single day go by without calling me; I could not have done this without her. Shahru for her encouragement and her belief in me. My entire family has been supportive all along, those in Toronto, Iran, and the UK. To amoo Iraj for taking care of me and letting me talk to him about anything at the many confusing times. And amoo Bahram and Samine that have been so great in keeping the home-sickness away. Amoo Amin for his hospitality and zest for life. Also Faraz for coming and spending a few months in the UK which kept me company and laughing. Amoo Fred, khale Azam, khale Forough, aghayeh Parhami, Farid and Sam for their continuous love.

This leads me to why I'm here. Thanks goes to Gareth from my previous lab in Toronto, a former post-doc in Biochemistry at Oxford. One day he said 'you should apply to Oxford', so I did, without any expectation. A few weeks later, one morning in the computer lab when I was scrambling to get my assignment submitted, I got an email from Mark. I can't believe three years has gone by. I think your time in a D. Phil is hugely affected by your supervisor and your attachment to your lab and project. Mark has been supportive, encouraging, and he fosters a relaxed yet productive working environment, all with a great sense of humour. Also thanks to Phil for all of his help and advice on the various projects.

I originally came to do a Masters, but I decided to stay two more years and complete a D. Phil, a big part of this decision was the Sansomites, old and new members of Mark's group. I couldn't write an acknowledgement without thanking Alex who was always willing to help and was an inspiration. Oli, who was the first person I met when he came to 'collect' me from the Biochemistry reception; he turned out to be a great colleague, a great friend, and a great climber. This leads me to some of the other climbers, Jeff, who kept the Canadian sense of humour alive in the lab, Jen, and John. I can't imagine the past three years without John, he has been supportive, helpful and kind. He has also read my thesis, and has been there for me through so much. Loredana, inside and outside the lab, has been a best friend, my experience in the lab would not have been the same without her. Zara has been an amazing friend, who encouraged me at the right times and always full of good advice. Kaihsu for his help and for bringing strange topics to the table. Jorge, thanks for the fun times and for your faith in me when I needed it. Thanks to Tony for keeping me laughing; coffee wouldn't be the same without Rob, Syma, Emi, Pete, and Tim. Also to Katherine who encouraged me to join the feisty and winning Wolfson's/St. Antony's Women's football team.

Some of the friends I met in my time at Oxford have become my best friends. Hammad, who has been a remarkable friend and has saved me from every doubt, struggle, and worry. There is nothing we haven't discussed, he makes me want to do more and be better, every time I see him. To Orkideh who is one of the reasons I'm so glad I came to Oxford, I gained another sister. Some of my best memories are with her and I'm so glad

to have her in my life. Also Zuzanna, Zahra, and Shakirah for providing the best conversations a girl could ask for and for your passion to save the world, also Zunnoor for his friendship and help on a few key thesis issues. And last but definitely not least are my old friends, who have kept me close to home. In particular I have to thank Mahsa and Safura for their encouragement, loyalty, sense of humour, and friendship in every sense of the word. Also thanks to Mojdeh, Lisa, Amy, Teresa, Rosalynn, and Amanda for their faith and constant encouragement.

1 Introduction 7

1.1 The cell membrane and membrane proteins 7
1.2 Problems with structural studies of membrane proteins 9
1.3 Ion channels 12
1.3.1 Ligand-gated ion channels (LGIC) 13
1.3.1.1 Selectivity 14
1.3.1.2 The gate and gating 15
1.3.1.3 Ligand binding site 17
1.3.1.4 Coupling agonist binding to gating 19
1.3.2 Nicotinic acetylcholine receptors 21
1.3.2.1 The α7 nAChR 22
1.4 The AChBP as a model 23
1.5 Computational studies of LGICs 25

2 Methods 28

2.1 Homology modelling 28
2.2 ZAlign 29
2.3 Molecular dynamics simulations 31
2.3.1 Periodic boundaries 37
2.3.2 Long-range electrostatic interactions and PME 37
2.3.3 Temperature and pressure coupling 38
2.3.4 Constraints 40
2.3.5 Energy minimization 41
2.3.6 Analysis of protein fluctuations 41
2.4 Conformational sampling methods 42
2.4.1 Gaussian network models (GNM) 42
2.4.2 CONCOORD 44
2.4.3 Coarse-grained molecular dynamics simulations 45
2.5 Essential dynamics 47
2.5.1 Principle component analysis (PCA) 47
2.5.2 Grouping of eigenvectors 48
2.6 AutoDock 49
2.7 Poisson-Boltzmann calculations with APBS 50

3 Simulation studies of AChBP with nicotine, carbamylcholine, and HEPES 53

3.1 Introduction 53
3.2 Methods 56
3.2.1 Preparing the ligand: generating topologies for nicotine, carbamylcholine, and HEPES 57

3.2.2 Preparing the macromolecule 59
3.2.3 Analysis of MD trajectories 60
3.3 Results 61
3.3.1 Global behaviour of AChBP 61
3.3.2 Large scale analysis 65
3.3.2.1 Structural flexibility and convergence 65
3.3.2.2 Principle component analysis (PCA) 69
3.3.2.3 Eigenvector-averaging 73
3.3.2.4 Gaussian network models (GNM) 74
3.3.3 Small scale analysis: the binding pocket 77
3.3.3.1 Rigidity of the binding pocket 77
3.3.3.2 Ligand flexibility 83
3.3.3.3 Water in the binding pocket 87
3.4 Discussion 101
3.4.1 Overall conformational dynamics 101
3.4.2 Binding pocket flexibility 102
3.4.3 Ligand flexibility 103
3.4.4 Water in the binding pocket 104

4 Docking studies of AChBP: an MD ensemble approach 108

4.1 Introduction 108
4.2 Methods 109
4.2.1 Docking 109
4.2.2 Molecular dynamics (MD) simulations 110
4.2.3 Automated docking along an MD trajectory 110
4.3 Results 111
4.3.1 Docking of nicotine, carbamylcholine, and acetylcholine onto AChBP 111
4.3.2 Docking ligands onto MD trajectories 113
4.4 Discussion 123

5 Generating the structure of the α7 nicotinic acetylcholine receptor 127

5.1 Introduction 127
5.2 Methods 129
5.2.1 Homology modelling of the EC and TM domains 130
5.2.2 The alignment of multiple domains of proteins – ZAlign 132
5.2.3 Stereochemical checks and energy minimization 136
5.2.4 Electrostatic calculations 137
5.3 Results: using the α7 nAChR model for structural studies 138
5.3.1 Conformational sampling studies 138
5.3.1.1 Gaussian network models (GNM) 138
5.3.1.2 CONCOORD 141

5.3.2 Electrostatics, energetics, and pore-profiles 144
5.3.2.1 Pore properties and Poisson-Boltzmann calculations (work carried out with Kaihsu Tai) 144
5.3.3 Comparison of the α7 nAChR model with the 4 Å *Torpedo marmorata* structure 148
5.4 Discussion 154
5.4.1 Conformational sampling 155
5.4.2 Electrostatics 155

6 A computational study of α7 nAChRs: directing experiment 159

6.1 Introduction 159
6.2 Methods 161
6.2.1 Homology modelling and *in silico* mutagenesis 161
6.2.2 Docking 163
6.2.3 Molecular dynamics simulations 163
6.2.4 Automated docking along an MD trajectory 164
6.2.5 Electrophysiology on recombinant wild type and mutant α7 receptors* 164
6.3 Results 165
6.3.1 α7 nAChR and α7 nAChR mutants 165
6.3.1.1 Docking studies against the a7 nAChR and mutants 166
6.3.1.2 MD and ensemble docking of α7 nAChR and mutants 169
6.3.1.3 Leu118 site-directed mutagenesis 173
6.3.2 ACR-16 175
6.3.2.1 Generating the full structure 176
6.3.2.2 Homology modelling of the EC domain 177
6.3.2.3 Docking of nicotine, acetylcholine, and Levamisole onto ACR-16 179
6.3.3 UNC63-UNC63, UNC63-UNC29, and UNC63-LEV1 modelling and docking 185
6.3.3.1 Docking of Levamisole, nicotine, and acetylcholine onto UNC63-UNC63, UNC63-UNC29, and UNC63-LEV1 185
6.4 Discussion 187
6.4.1 Loop E and L118 of α7 nAChR 187
6.4.2 ACR-16 189
6.4.3 UNC63-UNC63, UNC63-UNC29, and UNC63-LEV1 190

7 Structure, energetics, and coarse-grain molecular dynamics simulations of ligand-gated ion channels 193

7.1 Introduction 193
7.2 Methods 195
7.2.1 Homology modelling 195
7.2.2 APBS 196
7.2.3 Coarse-gained MD 196
7.3 Results 197

7.3.1 Homology modelling of LGICs 197
7.3.2 Pore profile analysis of the LGICs 199
7.3.3 APBS studies of LGICs 202
7.3.4 Coarse-grain MD of α7 nAChR and other LGICs 208
7.4 Discussion 215
7.4.1 Pore profiles 216
7.4.2 APBS 216
7.4.3 Coarse-grained MD simulations 217

8 Conclusions and possible future directions 220

8.1 Selectivity 220
8.2 The gate and gating 221
8.3 Ligand binding site 223
8.3.1 Binding pocket dynamics 223
8.3.2 Water 224
8.3.3 Mutants in the binding site 225
8.3.4 Ligand binding and flexibility 226
8.4 Coupling ligand binding to channel gating 227
8.4.1 Cyclo-*N*meric motion 228
8.4.2 CG MD of LGICs 230
8.5 Summary 231

Appendix I Programs and scripts 234

I.1 Programs 234
I.1.1 ZAlign 234
I.1.2 EigenGrouping 247
I.2 Scripts 255
I.2.1 automateDocks 255
I.2.2 analyzeAutomateDocks 259
I.2.3 analyzeHbond 262

Appendix II Publications 266

Table of abbreviations

$5HT_3$	5-hydroxytryptamine
ACh	Acetylcholine
AChBP	Acetylcholine binding protein
ACR-16	Acetylcholine receptor – 16 (gene name)
APBS	Adaptive Poisson-Boltzmann Solver
CCE	Name for AChBP + Carbamylcholine simulation
CFTR	Cystic Fibrosis transmembrane conductance regulator
CG	Coarse-grained
CONCOORD	Constraints to coordinates
DN-IMI	Desnitro-Imidacloprid
EC	Extracellular
EM	Electron microscopy
EPE	Name for AChBP + HEPES simulation
GABA	Gamma-aminobutyric acid
GNM	Gaussian network model
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hERG	Human ether-a-go-go related gene
IC	Intracellular
IMI	Imidacloprid
K_v	Voltage gated potassium (K^+) channel
Kir	Inward rectifier potassium (K^+) channel
LEV-1	Levamisole resistant – 1 (gene name)
LGIC	Ligand-gated ion channel
MD	Molecular dynamics
MscS	Mechanosensitive channel of small conductance
MSF	Mean square fluctuation
nAChR	Nicotinic acetylcholine receptor
NCT	Name for AChBP + nicotine simulation
NMA	Normal modes analysis
PB	Poisson-Boltzmann
PCA	Principal component analysis
PDB	Protein data bank
PMF	Potential of mean force
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
TM	Transmembrane
UNC-63,29	Uncoordinated – 63,29 (gene name)
VMD	Visual Molecular Dynamics

1 Introduction

1.1 The cell membrane and membrane proteins

The 3 nm-thick, phospholipid bilayer, making up eukaryotic cell membranes is a highly effective barrier, separating the inner contents of each cell from its outer environment. Amino acids, nucleotides, phosphates, water and ions cannot freely cross the tight barrier that a membrane provides. Thus specialized proteins are required to facilitate the influx and efflux of molecules across the lipid bilayer (Figure 1.1). Membrane embedded proteins carry out diverse functions including transport across the membrane, cell recognition, adhesion, signalling, and cell regulation.

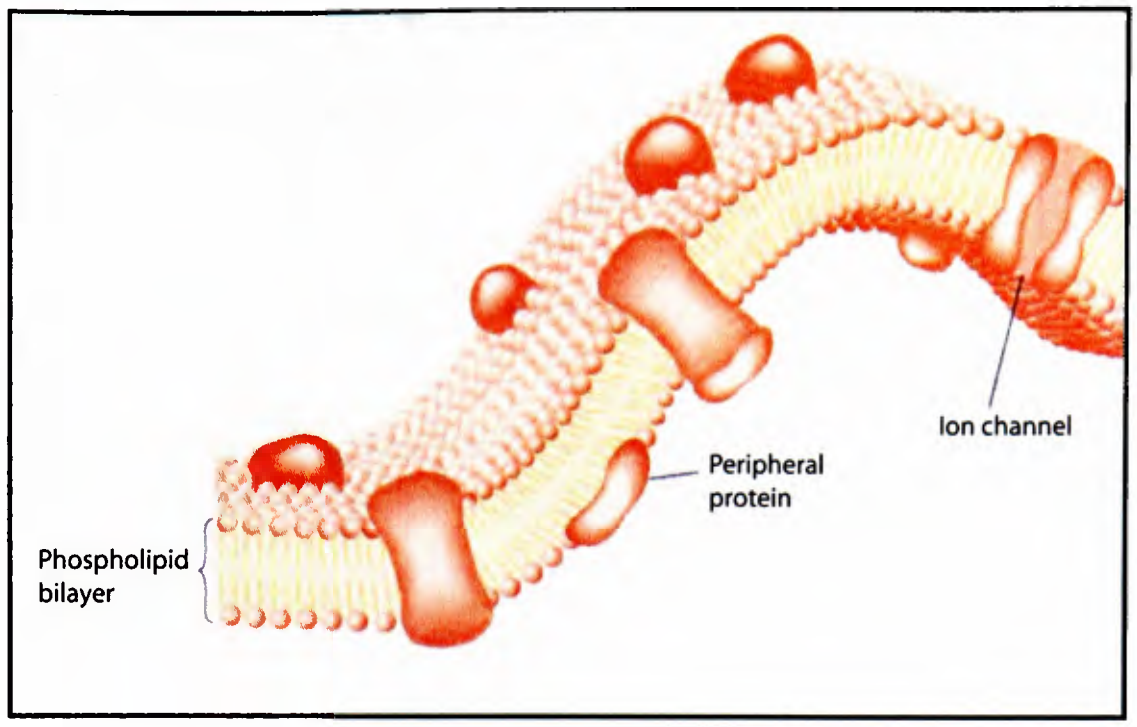


Figure 1.1 A phospholipid bilayer with several integral membrane proteins (modified from Nelson, *Comparative Neurophysiology* [1]).

Analysis of the genome suggests that 20-30% of genes code for membrane proteins [2-4] and these proteins makes up approximately 50% of drug targets [5]. Local anaesthetics, sedatives, anti-anxiety agents, anti-diabetic drugs and even antiviral therapies interact with ion channels in the cell membrane [6]. Even with such high abundance of membrane proteins and their pharmaceutical importance, there are still only 116 unique membrane protein structures in the Protein Data Bank (White, S.H. – website http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html). Most of these structures are obtained from bacterial proteins due to the relative ease (compared to eukaryotic systems) of over-expression and purification [7]. Fortunately, membrane protein structure determination is improving [8], with exponential growth of structure availability in the past few years, although still the numbers are much lower than those for globular proteins (~39000 structures available in the Protein Data Bank).

The first structure of a membrane protein, determined by electron diffraction, was that of bacteriorhodopsin from *Halobacterium salinarium* [9] and is archetypal of an inner membrane protein structure. This structure has since been determined by X-ray

crystallography at high resolution at various intermediate states [10, 11]. Structurally, the transmembrane spanning segment of membrane proteins is α-helical, while outer membrane proteins of prokaryotes are typically β-barrels. Table 1.1 is a list of some major families of integral membrane proteins.

Table 1.1 A summary of some major families of membrane proteins

Family	Function	Examples
Channels	allow flow of molecules down their (electro)chemical gradient	ion channels e.g. the nicotinic acetylcholine receptor, aquaporins
Pumps and transporters	couple energy release from one process to drive an unfavourable process (i.e. ion flow against electrochemical gradient)	ABC transporters, MFS e.g. LacY
Energy transducers	convert one form of energy into another. i.e. light energy into chemical energy	bacteriorhodopsin, members of the electron transport chain
Receptors	signal transduction across the membrane	G-coupled receptors, receptor-tyrosine kinases

1.2 Problems with structural studies of membrane proteins

The challenge in studying membrane proteins is the difficulty in attaining high resolution crystal structures. Often there is poor expression of the protein and, due to the nature of the lipid environment, it is difficult to obtain crystals of sufficient quality for structure determination. Membrane proteins are usually extracted and studied *in vitro* outside their native membrane environment. Problems also stem from the fact that mammalian membrane proteins are present at low levels in a given membrane and it is rare that a particular protein is a major constituent of the selected membrane.

To get around this problem, these membrane proteins are expressed in a bacterial host which brings a new set of challenges to the crystallization process. For mammalian membrane proteins post-translational modifications may be required that cannot be carried out in the bacterial host. During the crystallization process it is difficult to ensure that the stability and function of the protein are maintained. For example, the proteins can also form non-specific interactions with each other, causing the protein to precipitate. When crystals are obtained it is difficult to ensure that the protein is in its native state due to differences in its native membrane and the packing environment [12].

There are few structures available for the ligand-gated ion channels (LGIC) which are the focus of this thesis. There has been some success with cryo-electron microscopy (EM), where initially a 4 Å image of the transmembrane (TM) domain of the acetylcholine receptor from *Torpedo marmorata*, was obtained [13]. Approximately one million receptors were averaged, resulting in a 4 Å resolution structure of the TM domain of the acetylcholine receptor (AChR). This structure was used extensively for both experimental and simulation studies to understand the mechanisms of gating and permeation properties of LGICs [14, 15]. EM has also successfully determined the structure of an aquaporin at high (~ 1.9 Å) resolution, revealing details of interactions with lipids [16-18].

Recently a more complete structure of a nicotinic acetylcholine receptor (nAChR) was resolved by EM of the *Torpedo marmorata* ray postsynaptic membrane [19]. This 4 Å resolution structure is the first of a LGIC to include the TM, extracellular (EC), and intracellular (IC) domains (Figure 1.2). Although this structure is missing many residues in the IC domain, it is the most complete structure to date of a LGIC. This structure allows for the modelling of other receptors in the LGIC family and the

study of the interaction between the EC and TM domains, without having to model the interface between the two domains.

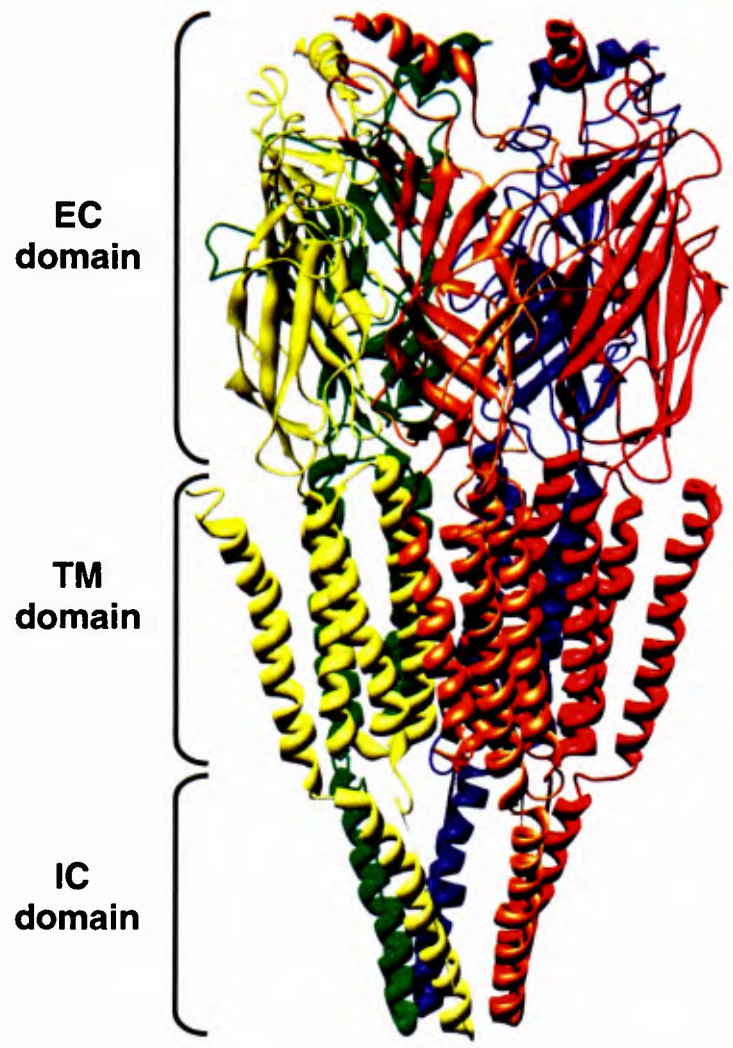


Figure 1.2 The 4 Å EM structure of the *Torpedo marmorata* AChR showing all three domains. This is the first complete structure of a LGIC.

Another significant advance in the study of LGICs was the release of a high resolution X-ray structure of AChBP from *Lymnaea stagnalis* [20] which resembles the EC domain of LGICs. This structure will be discussed in more detail in sections 1.3.1.3 and 1.4.

Given the difficulties in obtaining structures for membrane proteins, computational methods allow for the theoretical extrapolation of available structural information and the application of structural analysis to homologous proteins. A part of this thesis is focused on developing a method to combine structural data from different sources to construct and allow the study of more complete models of membrane proteins.

1.3 Ion channels

Ion channels are integral membrane proteins which allow the flow of ions across the membrane, down their electrochemical gradient. Ion channels are involved in nerve and muscle excitation, hormone secretion, cell proliferation, learning and memory, water and ion balance, regulation of blood pressure, fertilization, and cell death [14]. Figure 1.3 shows the nerve terminal, the neuromuscular junction, and a muscle fibre with various ion channels carrying out vital tasks between the muscle cell and the nerve cell. For example, the AChR, labelled 1 in Figure 1.3, is targeted by the neurotransmitter ACh in the neuromuscular junction. With the binding of the neurotransmitter, the channels allow the passage of sodium ions into the muscle cell, triggering depolarization and eventual muscle contraction.

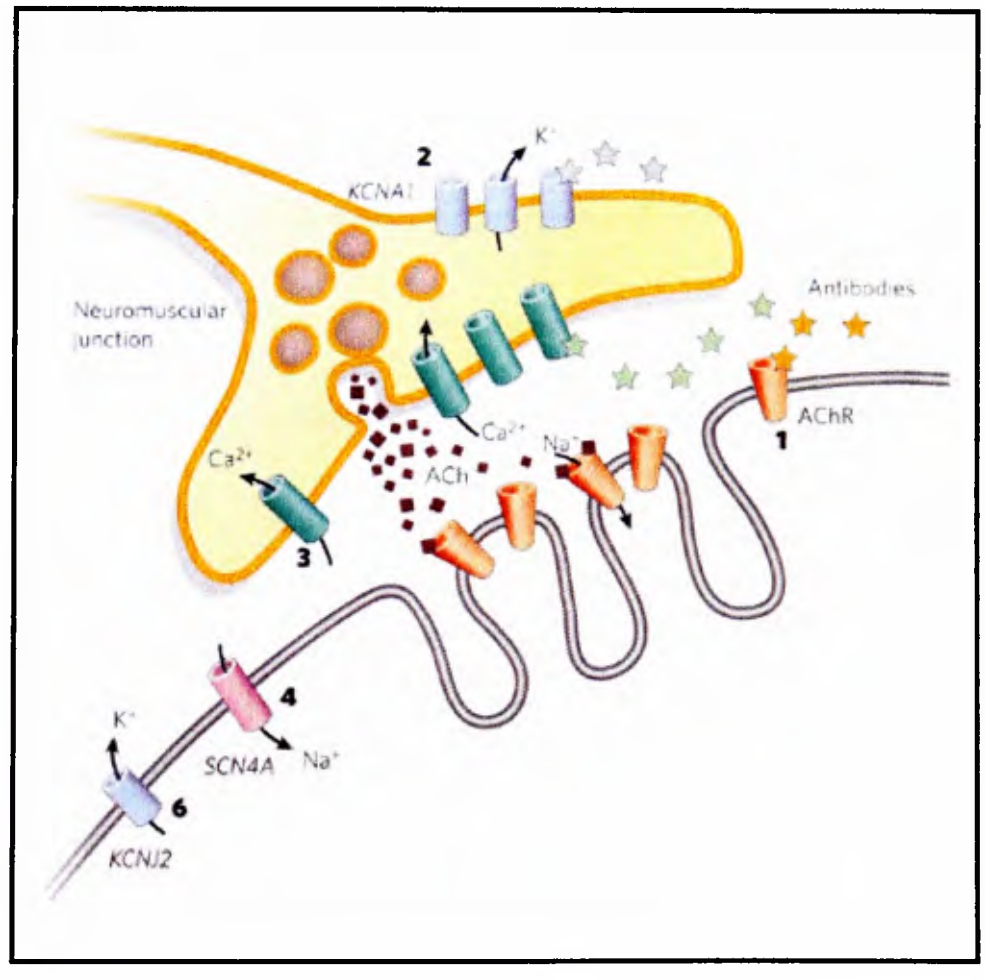


Figure 1.3 A nerve terminal, the neuromuscular junction and the muscle fibre together with various ion channels in the cell membrane. The normal function of these ion channels allows for correct muscle function while mutations in any of the labelled channels can lead to disease [14]. (Figure modified from Ashcroft [14]).

Ion channels are composed of one or more pore-forming domains, usually associated with other subunits. They are commonly tetrameric (K^+ channels) or pentameric (LGICs), however there are also a number of other compositions such as the dimeric chloride channels [21] or heptameric channels like the mechanosensitive channel of small conductance (MscS) channel [22]. The central pore through which ions move is commonly lined by one α-helix from each subunit. Voltage gated Ca^{2+} (Ca_V) and Na^+ (Na_V) channels are composed of a single subunit with four similar repeated domains while some K^+ channels are dimers with two repeated domains [14].

Pore-loop channels make up the largest family of ion channels and are evolutionarily related. They include the human ether-a-go-go (hERG) channels, inwardly rectifying K^+ (Kir) channels, and the glutamate receptor channels. Non-pore-loop channels are grouped into several classes with no apparent evolutionary relationships. These include the pentameric LGICs known as the Cys-loop receptors, the cystic fibrosis transmembrane conductance regulator (CFTR) channel (a member of the ABC transporter family) and chloride channels [14].

Recently high resolution structures of several ion channels have been obtained, including the Kir [23], K_v channels [24], the nAChR Cys-loop receptor [19], and chloride channels [25, 26]. These structures contribute greatly to the study of channel function.

1.3.1 Ligand-gated ion channels (LGIC)

LGICs are a class of membrane proteins whose functional state is modulated by the binding of a ligand. Upon binding of their ligand, conformational changes take place that open the ion pore and allow the passage of ions. Examples of LGICs are the nicotinic acetylcholine receptor (nAChR), glycine receptor, gamma-aminobutyric acid

($GABA_A$) receptors, and the 5-hydroxytryptamine(3) ($5\text{-}HT_3$) serotonin receptor [13, 27-29]. These are also known as "Cys-loop" receptors due to a pair of conserved cysteines in one of the ligand binding (EC) domain loops. The Cys-loop is considered important in the gating mechanism of LGICs [30-32]. From the LGICs, the acetylcholine and serotonin receptors have cation-selective pores, while the $GABA_A$ and glycine receptors are anion-selective. Each receptor type is composed of five subunits which can be of varying types (heteropentamers) or receptors can be composed of five identical subunits (homopentamers). For example in the nAChR, the five types of subunits are called α, β, γ, δ, ε and for a functional channel at least two of them must be of α type. The ligand binds at the interface between an α-subunit and an adjacent subunit (non-α type in heteromeric receptors, and α type in homomeric receptors). Subunits of α type are identified by a conserved pair of adjacent cysteines in loop C of the EC domain [33, 34]. In the TM domain, each subunit has four transmembrane (TM) α-helices (M1-M4), where the M2 helical segments line the pore and are considered important in the gating process [30, 35-37]. The extracelluar (EC) domain subunits each are composed of one α-helix, 10 stranded β-sheets, and two 3_{10}-helices [20, 38-41]. The intracellular domain is believed to be mainly α-helical [19], however the complete structure of this segment has not yet been resolved.

1.3.1.1 Selectivity

The ability of LGICs to allow for the passage of particular ions is attributed to the distribution of charged residues in the receptor and the size of the pore [42]. The LGICs discriminate between anions and cations, for example the AChRs permit flux of several species of cations but not anions. Rings of negatively charged residues facilitate the flux of cations for AChR and $5HT_3$. The anion-selective $GABA_A$ and

glycine receptors have rings of positive charge to allow the passage of anions. Experimental studies have been able to investigate the charge selectivity of LGICs by mutating key charged residues in the ion channels [43-48]. The focus of much of the charge selectivity studies has been on the M2 helical segment of the TM domain [48-51]. This helix, which lines the pore, contains rings of charged residues that are believed to influence ion conduction. Fewer studies have looked at the involvement of the EC domain in ion selectivity; this is investigated in chapter 7 of this thesis.

1.3.1.2 The gate and gating

Biological signals such as the binding of ligands, changes in membrane potential, changes in temperature, or mechanical stress regulate the opening and closing of ion channels. Ligand-gated ion channels are gated via the binding of a ligand. The gate of an ion channel acts as an energetic barrier to the movement of ions. The mechanism by which a neurotransmitter opens a channel gate has been the subject of several decades of study, with the greatest insights resulting from the study of the ACh skeletal muscle receptor [13, 19, 52-54]. Gating of the Cys-loop receptors without bound agonists occurs, but with low probability [55, 56]. The binding of a ligand dramatically increases the probability of channel opening and it is fully maximized with the binding of at least 2 ligands. Figure 1.4 shows a reaction scheme for LGIC gating, where **R** denotes the closed state and **R*** is the open state [57]. **A** represents the number of agonists and with increasing agonists the gating equilibrium constant $\boldsymbol{\theta}$ increases. The increase is thought to be around 10 million fold with the binding of 2 agonist molecules. The dissociation constants $\boldsymbol{K_1}$ and $\boldsymbol{K_2}$ in the closed state are in the micromolar range, although for the open state, $\boldsymbol{K_1}^*$ and $\boldsymbol{K_2}^*$ are estimated to be in the nanomolar range. Thus, there is a much greater affinity for ligands in the open state. This predicts that the conformation of the binding site

changes considerably between the open and closed state. As the opening rate can be quicker than the dissociation rate, several re-openings can occur with one ACh occupancy [57]. Binding site conformation is discussed in chapters 3 and 4.

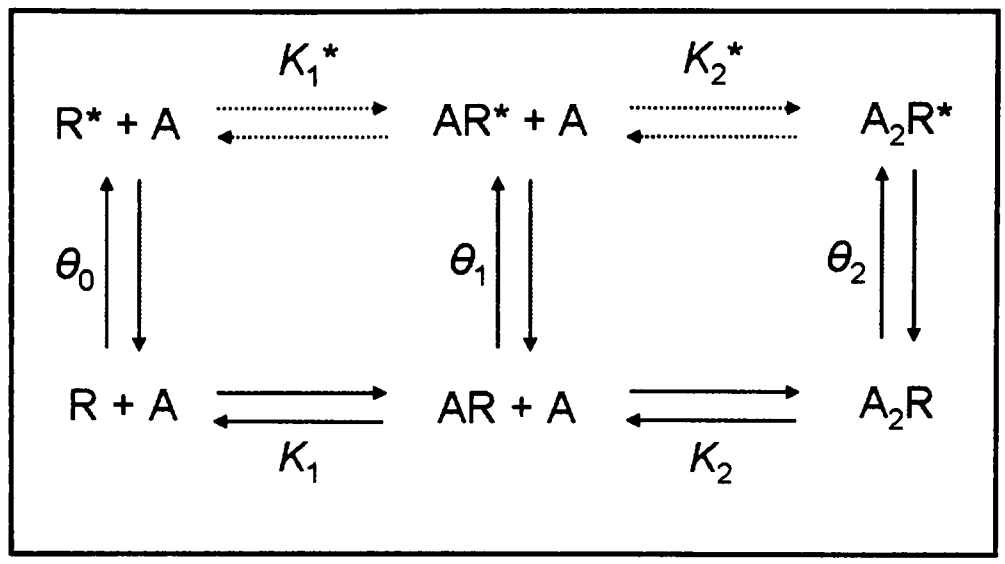


Figure 1.4 The transition states between the closed (R) and open (R*) states of Cys-loop receptors. A, represents a ligand and θ is the gating equilibrium constant which increases dramatically with the binding of the 2nd agonist. K_x and K_x* represent the dissociation constants of closing and opening respectively. (Figure modified from Sine and Engel [57])

In the nAChR a gating mechanism has been proposed where interactions across the EC and TM domains cause rotations of the pore lining helices which initiate rotational movements associated with channel opening [19, 58]. However the resolution of the structure is not sufficient to confirm the process. More recent insight into gating suggests that a cis-trans isomerization of a proline residue opens the gate of the $5HT_3$ receptor [59]. The gate of the LGICs and the concept of a gate in general has been a widely studied topic in ion channel research. Several studies have alluded to the possibility of a 'hydrophobic' gate in the transmembrane domain of LGICs [13, 45, 60]. This stretch of hydrophobic residues along the pore of these receptors is well conserved and serves as an efficient barrier to ion permeation in the closed state, while a small increase in radius and/or polarity of this region could open the pore [61, 62]. Substituted cysteine accessibility method (SCAM) studies on the gate of the

$5HT_3$ receptor suggest that the gate is located at the centre of the M2 helix [63]. However other studies suggest that the gate is lower down in the TM domain [33, 45, 64-67]. Although rotation of the M2 pore-lining helix is believed to be involved in gating [13, 19], some recent studies suggest [68] that there is minimal rotation of this helix.

The question of gating has been approached from several experimental directions, such as φ-value analysis based on rate-equilibrium free energy relationship (REFER) of mutants [69]. The value φ is a measure of the extent of the conformational change at the transition state [70]. Analysis of φ-values at different sites in the AChR suggests that the opening of the ion channel occurs as a conformational 'wave' via a sequence of domain shifts [68, 70-73] and is not dominated by one swift motion.

1.3.1.3 Ligand binding site

The binding site is buried at the interface of a 'principal' subunit and an adjacent 'complementary' subunit. In the muscle AChR the principal subunit is of α type and the complementary subunit is of δ or ε subunits. Depending on subunit composition (number of α-subunits), there can be up to five binding sites within a LGIC. Mutational and site-directed labelling studies of ligand binding have been the main avenues for studying the actions of a ligand on the ion channels in the LGIC family. The crystallographic structure of the acetylcholine binding protein (AChBP) [20, 38, 74-78] from *Lymnaea stagnalis* was the first atomic resolution structure with structural and sequence homology to the EC domain of LGICs. This allowed for the application of computational and experimental approaches for studying ligand behaviour at atomistic detail. Some of the X-ray structures of AChBP have been resolved with different ligands and their interactions with residues in the binding pocket can be explored via atomistic molecular dynamics and docking studies. The

functional studies had revealed that the binding site is enriched with hydrophobic and highly conserved aromatic residues and these were later confirmed with the structural determination of AChBP [79-81]. Figure 1.5 shows two subunits of AChBP with functionally significant loops and important residues in the binding site labelled.

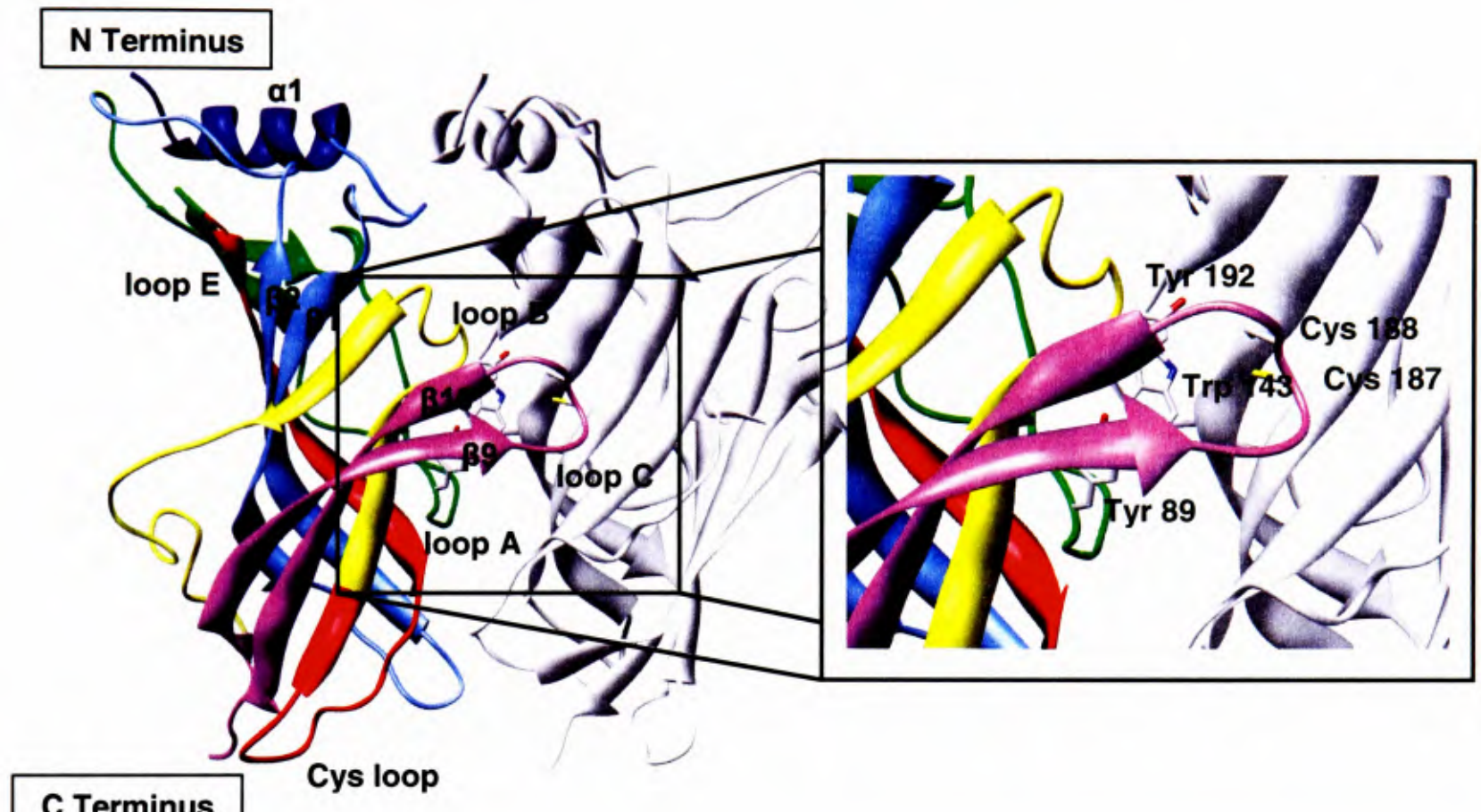


Figure 1.5 On the left, two subunits of AChBP are shown with some of the loops and β-sheets of the AChBP labelled. The short helix at the top of the structure is in dark blue, loop A is in green, loop B is coloured yellow, loop C is violet which is composed of β9 and β10. β1 and β2 sheets which lead down to the β1/ β2 loop and are in close proximity of the TM domain in LGICs are in light blue. Loop E which leads down to the Cys loop and is highly conserved amongst the Cys-loop family of LGICs is coloured red. On the right hand side, a few key residues in the binding pocket of AChBP are shown.

Ligands are believed to interact via hydrogen-bonding, van der Waals and cation-π interactions [79, 82] with residues in the binding pocket. Trp 143 from the principal subunit (Figure 1.5) is well conserved and makes the bulk of the cation-π interactions with the ligand. Hydrogen bonds are present between polar moieties of ligands and main chain and side chain elements of residues in the binding site, sometimes these hydrogen bonds take place via bridging water molecules [77]. The existence of

bridging water molecules in the binding site has also been reported for other proteins [83, 84] and will be investigated in this thesis (Chapter 3).

The structure of the binding site is different in the presence and absence of ligands. Recent molecular dynamics (MD) and tryptophan fluorescence data have suggested variable positions for loop C: a 'capped' state in the presence of a ligand and an 'uncapped' state in the absence of one [85]. Ligand-free structural data of the *Torpedo* nAChR [19] and the *Aplysia* AChBP [78] have confirmed uncapped conformations of loop C. The structure of the binding pocket, the protein-ligand interactions, and the role of water in the binding site are further investigated in chapter 3. The computational identification of key residues and side-chain conformations in the binding pocket in conjunction with experimental mutational testing is discussed in chapter 6.

1.3.1.4 Coupling agonist binding to gating

The X-ray structure of AChBP and the electro-microscopy structure of the TM domain of *Torpedo marmorata* [13] allowed for the study and better understanding of ligand binding in the EC domain and the key residues involved in forming a gate in the TM domain respectively. However many questions remain on the mechanism by which the binding of ligand signal is transmitted approximately 50 Å from the binding site to the TM pore, within tens of microseconds [86-89].

Mutational studies of key residues in the EC and TM domains have demonstrated impairment in gating [80, 90-100]. But it is the critical communication between the two domains that needs to be understood and recently there have been advances in this area. The interface between the β-sheet EC domain and the α-helical TM domain of LGICs has been the subject of multiple studies. Communication between the two domains is expected to involve the β1-β2, Cys, β8-β9-β10 loops of the EC and the

M2-M3 linker region of the TM domain [30, 71, 100]. A chimeric receptor was generated where AChBP was substituted for the EC domain of the $5HT_3$ receptor [101]. This construct was able to bind ACh, however it failed to convert ligand binding into channel gating. Once residues from $5HT_3$ were substituted into the AChBP β1-β2, Cys, β8-β9 loops, the structural compatibility with the TM domain was recovered and successful channel gating was achieved. However this result has not been reproduceable by other groups [102].

With the structure determination of the complete nAChR [19] at 4 Å resolution, the contact between the EC and TM domains could be studied with more accuracy (Figure 1.6). Analysis of this structure has lead to hypotheses linking ligand binding to channel gating. Arg 209 from β10 (extended from loop C) is believed to make a salt-bridge with Glu 45 of the β1-β2 linker. This electrostatic contact could be transmitted through the channel via Pro 272 on the M2-M3 linker of the TM domain.

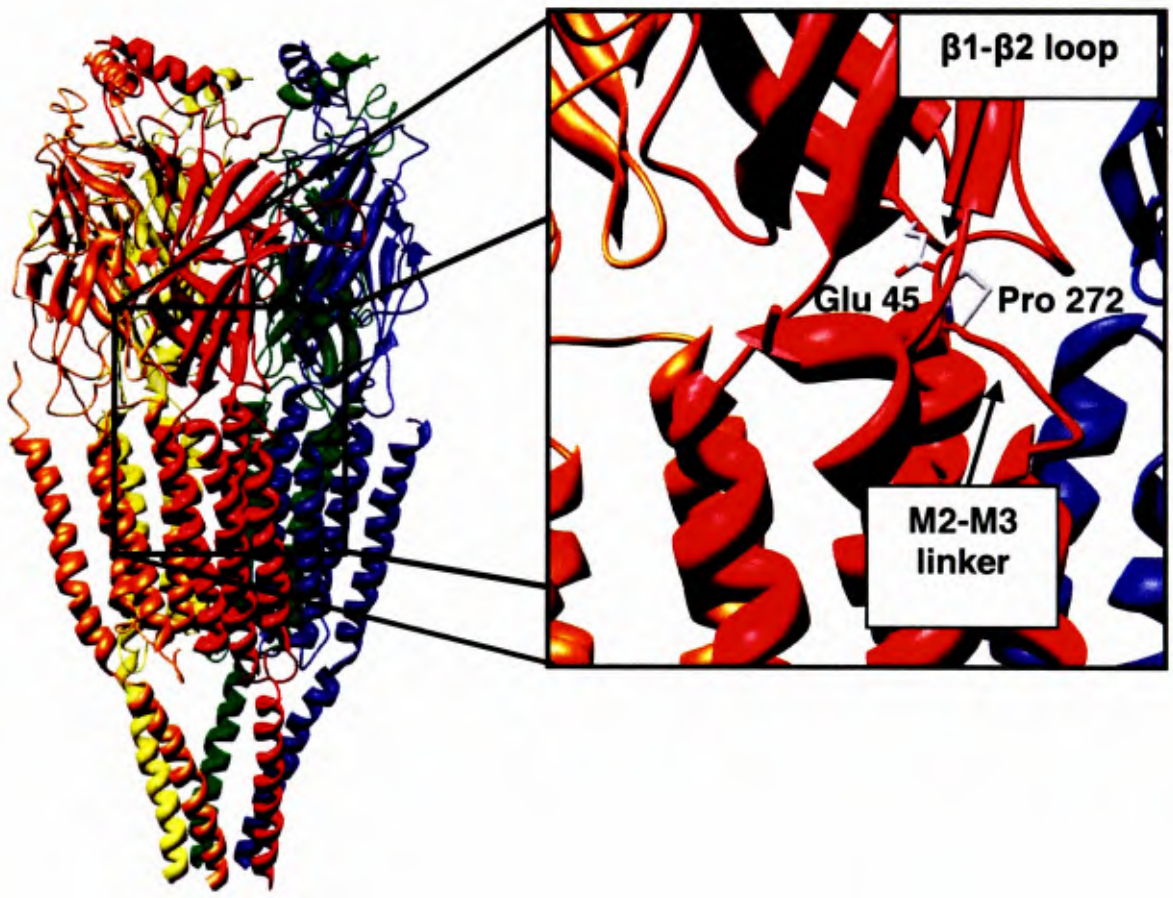


Figure 1.6 The 4 Å resolution structure of the *Torpedo marmorata* nAChR [19] (left hand side) is the first complete structure of a LGIC. A close-up of the EC-TM domain interface (right hand side) shows the M2-M3 linker of the TM domain and the β1-β2 loop of the EC domain.

Mutational studies confirm that altering any of the mentioned residues significantly affects the gating of the receptor [96]. Thermodynamic mutant cycle analysis (MCA) have been used to study the interdependence of residues in proteins by measuring the gating equilibrium constant θ with single, double, or triple mutations of key residues [103]. MCA studies confirm the interdependence of Arg 209 and Glu 45 and provide evidence for physical interactions between the two domains by correlating proximity and interdependence of residues.

More recent insight into the coupling of ligand binding to channel gating has been obtained from unnatural amino acid substitution for Pro 272 in the M2-M3 linker region [59] of the $5HT_3$ receptor. EC_{50} analysis (the concentration of agonist that generates 50% of the maximum response) for channel gating showed reduced potency by amino acid substitutions at Pro 272 that preferred a *trans* conformation. It is thought that the *cis* conformation of Pro 272 may bend the M2-M3 loop, which can alter the conformation of M2 and result in an open channel.

1.3.2 Nicotinic acetylcholine receptors

A substantial proportion of this work focuses on the nicotinic acetylcholine receptors (nAChRs) which are cation-selective channels allowing the passage of Na^+, K^+, and, less commonly, Ca^{2+} ions, and are activated by the neurotransmitter acetylcholine (ACh). They mediate fast synaptic transmission at neuronal-neuronal synapses and at neuromuscular junctions [33, 104]. They are involved in Alzheimer's disease [105], myasthenic syndromes [106], epilepsy [107], nicotine addiction [27, 31], and are targets for insecticides [108]. The nAChR proteins are pentamers composed of identical or highly homologous subunits each containing 6 loops (A-F) that make up the ACh binding site [109]. Homo-pentameric nAChRs are composed of either of α7,

α8, or α9 subunits [27], and hetero-pentameric nAChRs are composed of various combinations of α (α1 (muscle), α2-α10 (neuronal)), β subunits (β2-β4), δ (muscle) , ε (muscle), and γ (muscle) [27, 31, 34].

Neuronal nicotinic receptor subunits (α2-10; β2-4), contribute to a range of receptor subtypes depending on subunit composition [110], and are important targets for new drugs. Some of the drugs being developed are analgesics and drugs for ameliorating symptoms of Alzheimer's disease [111, 112], as well as drugs used to treat Parkinson's disease and schizophrenia [113]. Mutations in neuronal α4 and β2 nAChR subunits result in autosomal nocturnal frontal lobe epilepsy [107, 114, 115]. Also the underlying cause of several diseases are autoantibodies directed against nAChRs; these include myasthenia gravis (muscle nAChRs) [116], Rasmussen encephalitis (α7) [117] and autonomic neuropathy (α3) [118]. As well as human diseases discussed here, the nAChRs of insects are the targets for Imidacloprid, a neonicotinoid insecticide used extensively worldwide [108]. Nematode nAChRs are important targets for anthelmintic drugs which act against parasites, these drugs include Levamisole, Pyrantel and Morantel [119]. Docking studies of these compounds are carried out to investigate the protein-ligand interactions and subunit composition of the nAChRs in collaboration with experimental colleagues (Chapter 6).

1.3.2.1 The α7 nAChR

The α7 nAChR is a neuronal homopentameric cation conducting channel found in the central and peripheral nervous systems [28, 34, 120, 121]. It is involved in cognitive function, neurological disorders, and nicotine addiction [31, 33] and thus represents an important drug target. It shows some functional differences from muscle nAChR, in particular a relatively high permeability to Ca^{2+} ions [122]. The functions of wild-

type as well as many mutants of the α7 nAChR have been well characterized by electrophysiology in heterologous expression systems [109, 123] and thus it provides a good system to study as it allows comparison with experimental work.

Several α7-like receptors are also investigated in this work, namely the ACR-16 [124, 125] receptor, which is insensitive to the anthelmintic Levamisole, and also some Levamisole-sensitive subunits are investigated, although the subunit composition of these receptors is not yet clear.

The fivefold symmetry of the homopentameric nAChRs provides considerable advantages from a modelling and simulation perspective, especially since the template (AChBP) is also a homopentamer. Of the homopentameric nAChRs, the chicken α7 receptor has been the most intensively studied, both experimentally [109, 126, 127] and computationally [128-131].

1.4 The AChBP as a model

The acetylcholine binding protein (AChBP) is the first X-ray structure with high structural and sequence homology to the EC domain of LGICs. This 210 amino acid protein forms a stable homopentamer and it is most closely related to the α-subunits of nAChRs [20, 38, 77]. The AChBP binds known nAChR agonists and competitive antagonists such as nicotine, acetylcholine, and α-bungarotoxin. Most of the conserved residues in nAChRs, including those involved in ligand binding, are present in AChBP [132]. Binding sites in AChBP are located at subunit interfaces as in nAChRs and other LGICs. AChBP has sequence similarity of 15%-20% with $GABA_A$, glycine, and serotonin $5HT_3$ receptors, and 20%-27% sequence similarity to nAChRs [40, 41]. More recently, higher resolution crystal structures of AChBP have been solved with various ligands bound. These confirm earlier studies of the

important residues required for ligand binding [20, 38, 75, 77, 78] and are the subject of both experimental and computational studies.

Figure 1.7 Side-view (left) and top-view (right) figures of the AChBP with nicotine bound [77]. There is high structural and functional similarity with receptors in the LGIC family.

In light of the homology between the AChBP and ligand binding domain of LGICs, in particular the nAChRs (i.e. ~27% sequence identity with α7 nAChR), AChBP and homology models based on this structure provide a useful tool to investigate ligand binding, conformational dynamics, and the general behaviour of the LGICs. Several docking studies have been carried out on homology models of LGICs based on AChBP. For example, Le Novère and colleagues have built a three-dimensional model of the N-terminal domain of a homopentameric chicken α7 nAChR based on AChBP. They used this model to analyze the docking of ACh, epibatidine and nicotine [128], and were able to suggest plausible modes of binding for these ligands. In another modeling study, the authors were able to relate the pattern of binding affinities to different receptor isotypes [133].

1.5 Computational studies of LGICs

Computational methods facilitate extrapolation from molecular models of a structure to interpretation of physiological function and provide an opportunity to extend experimental data. Much of the earlier computational work in the LGIC area was focussed on the transmembrane domain and more specifically on the pore-lining M2 helix [35, 134-136]. The determination of the structure of the AChBP was a significant breakthrough for the computational study of LGICs and thus a number of investigators have used this as the basis of modelling and simulation studies of the EC domain of the nAChR and its interactions with various ligands [128].

Molecular modelling and simulations can provide further insight into mechanisms of channel function. A number of simulation studies of the TM domain components of the nAChR have been carried out. In order to explore the dynamics of TM helices, simulations on individual M2 helices, as well as M2 helix bundles have been carried out in a variety of environments [35, 134-137]. The structure of the AChBP has been used for a number of modelling studies of the EC domain of nAChR [128] and of other LGICs [138], largely focussing on receptor-ligand interactions. Insights into the symmetry breaking patterns in subunit dynamics of the α7 EC domain, which may be related to channel gating, were gained via MD simulations [129]. Simulations of the intact α7 nAChR have also been carried out by combining homology models for the different domains [130, 131]. To understand the electrostatic contribution to the energetics of cation permeation, studies of permeation through open vs. closed models of the acetylcholine receptor have been reported [139, 140].

Due to the difficulty in cystallizing membrane proteins, and the resulting membrane proteins fragments, an approach is developed in this thesis to construct LGICs from their different domains. Computational methods allow one to take partial

crystallographic, NMR, or microscopic data, along with other experimental evidence to hypothesize a plausible complete structure. The method developed here has been used for the α7 nAChR, $GABA_A$, glycine, and $5HT_3$ receptors. The computational docking studies carried out here identified residues which play an important role in protein function, such as controlling ligand access to the agonist-binding site (chapter 6). Based on these results, experimental studies have been carried confirming the importance of these residues. This demonstrates the benefits of combining computation with experiment and *vice versa*.

2 Methods

Various computational methods were used to study the nAChR and related proteins in the ligand gated ion channel (LGIC) family of membrane proteins. These methods range from atomistic molecular dynamics simulations to non-atomistic coarse-grain simulations, and other conformational sampling methods. Computational tools such as docking and *in silico* mutations are also used to explore ligand binding. An important part of this project has been developing methods to study membrane proteins. These are discussed along with other tools used to generate and study structures of membrane proteins.

2.1 Homology modelling

When the 3D structure of a protein has not been resolved with x-ray crystallography, NMR, or electron microscopy, we can use homology modelling to generate its structure based on a known (homologous) template structure. The sequence of the protein with unknown structure is aligned with the sequence of a homologous protein

of known structure, and the combined sequence and structural knowledge is used to generate a model structure. Modeller (versions 6.2 to 8.0) http://salilab.org/modeller/ [141-143] was used to carry out homology modelling in this thesis. Newer versions of Modeller were used as they were made available for further homology modelling. Modeller is used for homology modelling of three-dimensional protein structures given the PDB coordinates of the template and the sequence of the protein of interest to be generated based on the template's structure. Modeller uses a sequence alignment of the template protein and protein to be modelled to generate the new structure. For each model generated by Modeller, alignments were obtained from MultAlin [144]. The algorithm implements structure modelling by satisfying spatial restraints extracted from the template structure. Further restraints can be derived from multiple sources such as NMR experiments, rules of secondary structure packing, cross-linking experiments, fluorescence spectroscopy, site-directed mutagenesis, residue-residue and atom-atom potentials of mean force, image reconstruction in electron microscopy, etc. [141-143]. These restraints operate on distances, angles, dihedral angles, and other spatial features.

2.2 ZAlign

For proteins whose full structure is difficult to obtain, such as membrane proteins, partial structures or homologues of individual domains can be used [20, 29, 58, 145]. The problem then becomes how to re-assemble these different parts into the full-length structure. I wrote a program to take the different domains of a protein and combine them by aligning their central axis and optimizing interface distances. This is done for proteins with equal number of subunits in each of the domains to be combined.

In order to combine the domains, the n-fold rotational axis of each domain of the protein is aligned using a calculated central vector for each domain. One domain is held static while the other domain rotates and translates in the space specified in order to determine the optimal model. The parameter space for alignment of the separate domains must be determined by the user: x° rotation about the z-axis and $-a \leq z \leq +a$ translations on the z-axis (a full 360° rotation is possible, however for homo-meric proteins, a full rotation is not necessary, i.e. for a homo-pentamer, a 72° rotation will suffice). The two domains are brought close together, and rotations and corresponding translations are carried out. User defined rotations and translations, typically 1° rotations and 1 Å translations, around and on the pore (z)-axis respectively are used to generate the data. However, these can be altered to achieve the desired accuracy. At each step (rotation and respective translation) a score is calculated based on the following criteria:

i: The number of bad contacts, ***B***, between atoms of the different domains. A 'bad contact' is when the interatomic distance is less than 1.5 Å.

ii: The distance, ***T***, between the C-terminus of one domain and the N terminus of the other domain. An approximate van der Waals collision distance of 4.0 Å is used. This criterion was used for the nAChR and can be used when combining domains of the same protein i.e. joining the Cα atoms of termini residues - however when combining a protein complex, this value may have to adjusted or omitted i.e. combining AcrB (inner membrane protein) and TolC (outer membrane protein).

iii: Additional criteria, $\boldsymbol{A_{i...n}}$, defined by the user depending on the protein of interest. There may be specific distances or contacts between certain residues which have been shown to be important. These measurements can be incorporated into the scoring function. The overall scoring function $\mathbf{S} = w_B\boldsymbol{B} + w_T\boldsymbol{T} + (w_A\boldsymbol{A_{i...n}})$, is a linear

combination of the separate scores with the individual weights adjusted to give equal weight to each criteria.

Contour plots are generated, one for each scoring criteria, and one for each linear combination (linear combinations can be made from any two or more criteria). The best model corresponds to the lowest value of S. The scoring system can be extended by the user, however, the bad contacts, ***B***, criteria is universal for all proteins.

2.3 Molecular dynamics simulations

Molecular dynamics (MD) solves Newton's equations of motion for N interacting atoms of a system [146, 147]. The position of every atom is recorded with respect to time and the interactions between atoms are defined by a force-field [146]. Running MD on proteins allows us to visualize and study what happens to a protein on a very small time-scale. With today's computational capabilities, we can run simulations up to hundreds of nanoseconds [146]. This time-scale is still inadequate for observing some physiological phenomena such as channel gating, but is a vast improvement on what could be achieved a few short years ago. Different types of protein motions of different time-scales can be observed [148]. Table 2.1 describes the different types of protein motions and related time-scales.

Table 2.1

Type of Motion	Time scale	Functionality
Local motions – atomic fluctuations i.e. side chain motions	Femtoseconds (fs) to picoseconds (ps), 10^{-15} – 10^{-12} s	Ligand docking studies
Medium scale motions – rigid body motion, loop motions	Picoseconds (ps) to microseconds (μs), 10^{-12} – 10^{-6} s	Active site conformational changes, binding specificity
Large scale motions – domain and subunit motions	Microseconds (μs) to milliseconds (ms), 10^{-6} – 10^{-3} s	Allosteric transitions (i.e. channel gating), hinge-bending motions
Global motions – Folding/unfolding, subunit association	Millisecond (ms) to hours, 10^{-3} – 10^{4} s	Protein function

The data established can be interpolated, and used to hypothesize on the larger scale motions using various techniques, including principle component analysis (PCA). Molecular dynamics provides a useful complement for experimental structural biology and is used here, along with various other methods, to study the structural variability and behaviour of proteins.

The principal approximation made in MD is the Born-Oppenheimer approximation [146]. This is based on the difference in size between nuclei and electrons. Since the mass of electrons are much lower than those of the masses of nuclei, electrons rapidly adjust their position to any change in the nuclear position. This enables the electronic and nuclear terms to be separated and thus the potential energy of a molecule in its ground electronic state can be considered as a function of the nuclear coordinates only [146] . This approximation enables us to simplify the description of the dynamics of atoms, and it also allows us to represent the interactions and repulsions among atoms as functions depending on the nuclei position only. In this assumption the classical Newton's laws of motion are used.

Newton's equation of motion for a system of N atoms [146] :

$$m_i a_i = F_i \longrightarrow m_i \frac{\partial^2 r_i}{\partial t^2} = F_i, i = 1...N \qquad \textbf{2.1}$$

This equation describes the motion of a particle of mass m_i along a coordinate r_i with force F_i. The force at time t, $F(t)$, is calculated for all pairs of particles of the system at position $r(t)$. From the force $F(t)$, the acceleration $a(t)$ is computed and by knowing the velocity $v(t)$ and acceleration $a(t)$, the new position $r(t+\Delta t)$ and the new velocity $v(t+\Delta t)$ are obtained. Δt is the time interval called the time step and the force is assumed to be constant during the time step. The initial velocities are obtained from Maxwell-Boltzmann, Gaussian, or uniform distributions. The *Verlet* [149] and *leap-frog* [150] algorithms are most commonly used to integrate the equations of motions for molecular dynamics simulations. These assume the positions, velocities and accelerations can be approximated as Taylor series expansions. Positions are updated using the forces $F(t)$.

The force on each atom is the negative derivative of a potential function $V(r_1, r_2, ..., r_N)$:

$$F_i = -\frac{\partial V}{\partial r_i} \qquad \textbf{2.2}$$

The Hamiltonian (H) describes the energy function composed of the kinetic energy term (K) and the potential energy term (V), also known as the interaction function term [146, 148].

$$H(\rho, r) = K(\rho) + V(r) \qquad \textbf{2.3}$$

K depends on atomic masses, (ρ represents the momentum), and V is the interaction function composed of bonded and non-bonded interactions (r is the atomic coordinates).

The force-field describes the total potential energy of the system [146, 148]. The potential energy function of the system is broken down into the various energy types of the system i.e. bonded and non-bonded interactions (Equation 2.4). It also determines the parameters used for force-constants and atomic charges. These parameters are based on experimental data such as x-ray crystallography, NMR, Infra-Red spectroscopy, and quantum mechanical calculations, to name a few [146].

$$V(r) = V^{bonded}(r) + V^{non-bonded}(r) \qquad \textbf{2.4}$$

The non-bonded interactions are those between atoms that are separated by three bonds (1,4 interaction) or atoms from different molecules.

$$V^{non-bonded} = \text{van-der-Waals + electrostatic interactions} \qquad \textbf{2.5}$$

Bonded interactions are those between atoms directly bound to each other in the same molecule.

$$V^{bonded}(r) = V^{bond}(r) + V^{angle}(r) + V^{improperDihedral}(r) + V^{dihedral}(r) \qquad \textbf{2.6}$$

In equation 2.7, b_i is the bond length with b_{io} being the reference value with $k_{i,b}$ as the force constant.

$$\sum V_{bonds} = \sum_{i=1}^{N_{bonds}} k_{i,b}\left(b_i - b_{i,0}\right)^2 \qquad \textbf{2.7}$$

In equation 2.8, θ_i is the bond angle value and θ_{i0} is the reference angle with $k_{i,\theta}$ as the force constant.

$$\sum V_{angles} = \sum_{i=1}^{N_{angles}} k_{i,\theta}(\theta_i - \theta_{i,0})^2 \qquad \textbf{2.8}$$

The dihedral term, which represents the energy term for rotating about the central bond between four consecutively bonded atoms (Figure 2.1), has force constant k, n is the periodicity or multiplicity, and δ is the phase. Φ represents the angle between two planes.

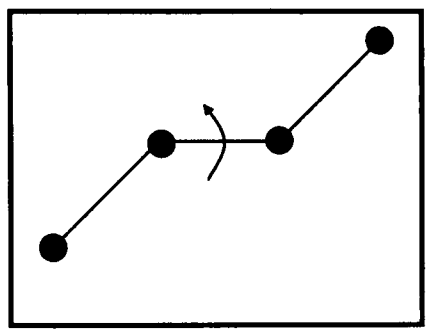


Figure 2.1 A schematic for a dihedral angle showing rotation about the central bond.

$$\sum V_{dihedrals} = \sum_{i=1}^{N_{dihedrals}} k_{i,\phi}\left[1+\cos\left(n_i\phi_i - \delta_i\right)\right] \qquad \textbf{2.9}$$

In order to keep the planar groups planar and to prevent molecules flipping over to their mirror images, improper dihedrals are used. Ψ represents the torsion angle between four atoms of a molecule.

$$\sum V_{impDihedrals} = \sum_{i=1}^{N_{impDihedrals}} k_{i,\psi}(\psi_i - \psi_{i,0})^2 \qquad \textbf{2.10}$$

For the non-bonded interactions, the electrostatic term is calculated using a Coulombic potential (equation 2.11). q_i and q_j are the charges on the atoms i and j, r_{ij}

is the separation distance on the atoms i and j, and ε_0 is the relative permittivity of free space.

$$\sum V_{electrostatics} = \sum^{pairs(ij)} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \textbf{2.11}$$

Van-der-Waals interactions describe the dispersive and exchange-repulsive interactions between atoms and are modeled using the Lennard-Jones function (equation 2.12). The collision diameter is σ (the energy is zero at this separation) and the well depth is ε.

$$\sum V_{Van-der-Waals} = \sum^{pairs(ij)} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad \textbf{2.12}$$

There are various MD packages with different force-fields. Some of the more widely used MD force-fields are AMBER [151], CHARMM [152], GROMOS [147, 153], and OPLS [154]. Each can offer certain advantages depending on the system being simulated. Where MD studies have been carried out in this work, the GROMOS96 force-field in the GROMACS MD package has been used [147]. This is a united atom force-field where there are no explicit non-polar hydrogens. These hydrogens are characterized with the carbon atom to which they are associated and thus these carbon atoms are given correspondingly larger van der Waals radii. This process reduces the computational cost of the system by decreasing the number of degrees of freedom, without sacrificing accuracy to any great extent.

2.3.1 Periodic boundaries

In order to mimic a bulk fluid environment such that all particles experience forces not restrained by boundaries, periodic boundary conditions are used [146, 147]. This enables a simulation to be carried out with a small number of particles with characteristics of a much larger and fluid environment. This also enables 'macroscopic' properties to be calculated from simulations using a relatively small number of particles. In this method, the simulation box in 3-dimensional space is surrounded by 26 boxes (much like a 3x3x3 cube, each unit representing a periodic image, minus the original square in the middle). This ensures that when a particle leaves the box during a simulation, it is replaced by an image particle that enters from the opposite side and thus the number of particles in the central box remain constant [146]. Other shapes can also be used; however in the studies carried out in this work, the boxes used were cubic.

2.3.2 Long-range electrostatic interactions and PME

The calculation of non-bonded interactions (electrostatic and van der Waals) is the most time-consuming part of a MD simulation [146]. In principle non-bonded interactions should be computed between every pair of particles in the system including their images; however this is inefficient and arguably unnecessary. For this reason, the calculation can be approximated using a cut-off. It is assumed that the largest contribution to the potential and forces comes from the near neighbours of the particle and a truncation can be made. The interactions between pairs of particles further than a spherical cut-off value are set to zero. For van der Waals interactions, the potential becomes insignificant beyond ~ 10 Å in the Lennard-Jones potential, thus these interactions are considered to be short range.

On the other hand, electrostatic interactions are long range interactions and their effects are felt at a considerably greater distance than the cut-off commonly used for short-range interactions [146, 147]. A variety of methods have been developed to calculate long-range forces, however the most commonly used is the Ewald summation method [155]; a variant of this method is employed for the simulations in this thesis. This method can be employed for calculating the full electrostatic energy of a unit cell. In the Ewald sum method, each point charge in the system is surrounded by a charge distribution of equal magnitude and opposite sign; this distribution is conveniently taken to be Gaussian and acts like an ionic atmosphere to screen the interactions between neighbouring charges. Now the screened interactions are short-ranged and the total screened potential is calculated by summing over all the particles in the central box and all the images in real space. Then another charge distribution of opposite sign is added to the first added Gaussian distribution and this cancelling distribution is summed in reciprocal space.

A fast implementation of the Ewald summation is the Particle Mesh Ewald (PME) method [156, 157]. In this method Fast Fourier Transformations are used to approximate the reciprocal space term of the Ewald sum. Here charges are interpolated to grid points. This procedure greatly reduces the computational cost of the reciprocal space sum.

2.3.3 Temperature and pressure coupling

It is necessary to control the temperature and pressure of a system due to drift as a result of integration errors, force truncation, and heating due to frictional forces. Both the weak coupling scheme of Berendsen [158] or the extended ensemble of Nosé-Hoover [159, 160] is available for use in the GROMACS package. The Berendsen

algorithm uses weak coupling to an external heat bath with first-order kinetics given temperature T_0. This means that a temperature deviation decays exponentially with the time constant τ.

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau_T} \qquad \textbf{2.13}$$

This system is efficient for relaxing a system to the target temperature and thus is the method of choice during the equilibration phase of an MD simulation. However, once the system has reached equilibrium, it is important to use a more accurate thermodynamic ensemble. The method introduced by Nosé, and later modified by Hoover, introduces a thermal reservoir and a friction term in the equations of motion. The friction parameter or 'heat bath' variable is a dynamic quantity with its own equations of motion. This is the method of choice after the equilibration stage, although usually the differences are negligible.

Pressure Coupling:

The Berendsen algorithm [158] and the extended Parrinello-Rahman [161] are supported by GROMACS for pressure coupling. The coordinates and box vectors are rescaled at every step with a matrix μ. This matrix has the effect of a first-order kinetics relaxation of the pressure towards a reference pressure P_0.

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P} \qquad \textbf{2.14}$$

The Berendsen scaling in GROMACS can be done isotropically or semi-isotropically. Isotropic scaling factor is equal for all three directions i.e. for water which is an isotropic system. Semi-isotropic pressure coupling is where the *x*/*y* directions are scaled independently from the *z* direction i.e. applicable for lipid bilayer; and in

anisotropic coupling, the scaling factor is calculated independently for each of the three axes [146].

When we are interested in calculating the thermodynamic properties of a system, the fluctuations in pressure and volume are important and therefore constant-pressure simulations using the Parrinello-Rahman method are employed. This method is similar to the Nosé-Hoover temperature coupling scheme. In the Parrinello-Rahman method, the equations of motion for the particles are changed [146] such that the box vectors, represented by a matrix, obey the equations of motion. It is recommended that Berendsen pressure coupling is used to reach the target pressure and this is followed with Parrinello-Rahman pressure coupling to maintain the correct pressure of the system. Usually Nosé-Hoover temperature coupling is used together with the Parrinello-Rahman pressure coupling.

2.3.4 Constraints

Constraints are requirements that bonds and angles are forced to meet throughout the simulation without affecting the other internal degrees of freedom. If molecules are allowed flexibility, it will result in conformational changes. The two most common method for applying constraints in MD are the SHAKE [162] and LINCS [163] algorithms.

LINCS is a more recent non-iterative constraint algorithm implemented in GROMACS, which is faster and more stable than SHAKE, which is iterative. In LINCS bonds are reset to their correct lengths after an unconstrained period of the simulation. Constraints are used to increase the time-step of a simulation. The time-step is dependent on the high frequency motions (e.g. bond vibrations), which are of

less interest than lower frequency motions of a protein; with constraint methods, the high frequency bond motions are ignored.

2.3.5 Energy minimization

An important step in setting up a simulation is minimizing the energy of the system. Energy minimization removes atomic clashes and bad geometry that may cause simulation instability. A system which is energy minimized corresponds to a stable arrangement of atoms. We obtain a low-energy configuration of the system by finding the *nearest local minimum* of the potential energy landscape. Two of the more common energy minimization algorithms are the steepest descent and conjugate gradient methods, both based on the first derivative of the potential energy function [146].

In the steepest descent method, a line-search is used to locate the minimum value in the function along a line through the multi-dimensional space and the current step is not dependent on the previous step. This method is good for relieving the highest energy features of an initial configuration and thus is useful when the system is far away from its energy minimum. The conjugate gradient method takes the previous step into account and refines its direction of movement down the gradient. This method performs better when the system is near the energy minimum [146]. The steepest descent method is used for the energy minimizations carried out in this work as it is ideal for removing bad contacts in high energy configurations.

2.3.6 Analysis of protein fluctuations

The overall dynamics of a protein can be assessed by the mean-square fluctuation (MSF) where a time window w is defined:

$$MSF_w = N^{-1}S^{-1}\sum_{i=1}^{N}\sum_{j=1}^{S}\left(X_{ij} - \langle X_i \rangle\right)^2 \qquad \textbf{2.15}$$

N is the number of atoms in the system, S is the number of snapshots in the sampling window, X_{ij} is the positional vector of atom i in the snapshot j and $\langle X_i \rangle$ is the average of the positional vector over all the snapshots in the window.

To assess the extent of conformational sampling, 'block analysis' of the MSF can be carried out for different lengths of time window w. Therefore, a time-averaged MSF is defined as:

$$MSF = W^{-1}\sum_{w=1}^{W} MSF_w \qquad \textbf{2.16}$$

W is the number of windows with a window length L and a simulation time T is $W=TL^{-1}$.

2.4 Conformational sampling methods

2.4.1 Gaussian network models (GNM)

GNM [164-166] (code obtained from Dr. Robert L. Jernigan's Laboratory website (Iowa State University) http://ribosome.bb.iastate.edu/software.html) allows the identification of the flexible regions of protein. It considers each residue as a 'ball' and distances between residues as 'springs'. There is a spring from a residue to all surrounding residues within a cut-off distance. The greater the number of neighbouring residues, the less freedom of motion that a residue is allowed. GNM was tested on a range of proteins including the HIV-1 reverse transcriptase [167] and influenza virus hemagglutinin [168]. Global motions and surprisingly finer interpretations have been seen with this method.

GNM produces distance matrices and generates theoretical B-values [167]. In equation 2.17, ΔR_{ij} represents the fluctuation in the separation $R_{ij}=|R_j-R_i|$ between the ith and jth Cα atoms in the protein. The fluctuations are assumed to obey a Gaussian distribution.

$$W(\Delta R_{ij}) = (\gamma^* / \pi)^{3/2} \exp(-\gamma^* \Delta R_{ij}^{\ 2}) \qquad \textbf{2.17}$$

In equation 2.18, $\{\Delta R\}$ represents the N-dimensional column vector formed by the fluctuations $\{\Delta R_1, \Delta R_2, \ldots, \Delta R_N\}$ of the Cα atoms. K is a constant, Γ is a symmetric matrix known as the Kirchoff matrix [169].

$$Z_N = K \exp(-\{\Delta R^T\}\Gamma\{\Delta R\}) \qquad \textbf{2.18}$$

Equation 2.19 outlines the elements of Γ which is the stiffness matrix where the mean square fluctuations of the Cα atoms are evaluated. r_c is the cut-off distance defining the range of non-bonded contacts. γ* is the normalization constant which is the counterpart of the single parameter in the Hookean potential [170]. Hooke's law is commonly encountered as the spring equation relating the force exerted by a spring to the distance it is stretched. The potential of this force is a parabolic function with the equilibrium distance resting at the middle of the parabola (lowest energy). The summation Γ_{ii} is performed on all off-diagonal elements on the ith column or row and represents how many other residues are connected to residue i.

$$\Gamma_{ij} = \begin{cases} -\gamma^* \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 \text{ if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{i \neq j} \Gamma_{ij} \text{ if } i = j \end{cases} \qquad \textbf{2.19}$$

The mean square fluctuations of the Cα atoms are evaluated from the diagonal elements of Γ^{-1} using $\langle \Delta R_k^2 \rangle = [\Gamma^{-1}]_{kk}$. The isotropic temperature (B) values are related to the mean square fluctuation and can be found using:

$$B_k = \frac{8\pi^2}{3}\langle \Delta R_k^2 \rangle = \frac{8\pi^2}{3}[\Gamma^{-1}]_{kk} \qquad \textbf{2.20}$$

2.4.2 CONCOORD

This method generates random structures by satisfying distance constraints extracted from the input structure [171]. CONCOORD has been tested on various proteins and the distance restrictions have been refined and show compatibility with protein structures. The differences between the generated structures are used to look at the structural variability of the protein. Furthermore, essential dynamics of the randomly generated structures can be carried out.

Given an initial structure, hundreds of new structures are generated within given cut-offs to allow for the conformational sampling of the protein. There are two separate steps in the program; in the first step all pair-wise interatomic distances of the given structure are measured. Different types of interactions are considered based on their distance. The values of the distance criteria were obtained from distance fluctuations in MD simulations of the B1 IgG-binding domain of streptococcal protein G [171]. Once all the distances for all the pairs of atoms are defined, the second step generates structures that fulfill the constraints defined in the first step.

Structures are each generated by starting from random coordinates and iteratively correcting those positions that violate the distance constraints. If the algorithm does not converge, meaning that not all distance constraints are satisfied, the structure is rejected and a new set of random coordinates is used. The structural 'space' is more

efficiently sampled than in MD where each structure is correlated with the previous. There is no correlation between any two structures that are generated as all of them start with random coordinates i.e. time-dependent structural ensemble.

This method is fast compared to MD and the information can be used to gain insight into the prominent modes of motion for the protein being studied. Validation of this method has been carried out using normal modes analysis (NMA). In NMA a diagonalization of Hessian matrix is carried out whose elements correspond to the second derivates of the potential energy with respect to the coordinates. It was found that these eigenvectors correspond very closely to those generated with MD simulations [171].

2.4.3 Coarse-grained molecular dynamics simulations

In an approach to increase the time scale and system size of simulations, coarse-grained (CG) models treat groups of atoms as single particles. This type of coarse-grained MD uses one CG particle to represent four heavy atoms (not H) [172] (see Figure 2.2). The parametrization of amino acids was based on methods derived by Marrink and colleagues [173, 174]. Soft harmonic terms maintain bond lengths and angles and CG particles interact via Lennard-Jones and Coulombic potentials (only the charged particle types). Only four CG particle types are used: "polar" (P), "mixed/apolar" (N), "hydrophobic/apolar" (C), and "charged", (Q) groups along with subtypes for the N and Q particles which reflect hydrogen bonding capacities by fine-tuning of Lennard-Jones interactions. The secondary and tertiary elements are restrained by a network of harmonic potentials i.e. elastic network, between Cα particles within a typical cut-off of 7 Å. Non-bonded interactions are described using Lennard-Jones potentials. The assignment of appropriate particle types is based on

partial charges and hydrogen bonding potentials of the amino-acid atoms. Therefore, for each amino acid residue, a single backbone particle of type N (mixed polar/apolar) is assigned, in addition, between zero and two side chain particles are also designated to each residue. The backbone particles subtypes (N) depend on the H-bonds within the backbone of the starting atomistic structure. i.e. N0 (no hydrogen bonding), Nd (hydrogen bonding donor), Na (hydrogen bonding acceptor), or Nda (hydrogen bonding donor and acceptor). The assignment of side chain particles is described in Table 2.2.

Table 2.2 CG amino acid side chain particle assignments

Residue	Side chain particle type	Explanation
Ala, Ile, Leu, Pro, Val	C	Hydrophobic
Phe	C+C	Large hydrophobic
Met, Cys	N0	Mixed polar/apolar and no hydrogen bonding with water, but sulphur containing group has a strong dipole.
Asn, Gln	Nda	Hydrogen bonding donor and acceptor
Ser/Thr	P	Polar residues
Tyr/Trp	C+Nd	Hydrophobic and hydrogen bond donor capacity
His	C+Nda	Hydrophobic ring and hydrogen bonding acceptor and/or donor
Asp/Glu	Qa	Charged with hydrogen-bonding acceptor capacity
Lys/Arg	C+Qd	Charged with hydrogen-bonding donor capacity * Arg has been changed to N+Qd in the newer implementation

Bonds and angles are treated with harmonic potentials as describe in Marrink et al., 2004. The mean distance between Cα atoms in protein at equilibrium is 0.38 nm with a force constant of 1250 kJ mol^{-1} nm^{-2}. Equilibrium bond angles are 90° for α-helical segnments, 130° for β-strands and 120° for random coils with a force constant of 25 (random coil) or 35 (secondary structure elements) kJ mol^{-1} rad^{-2} used as bond angle potentials. The CG structures are obtained from all-atom files, parametrized according to the CG model, energy minimized, and simulated.

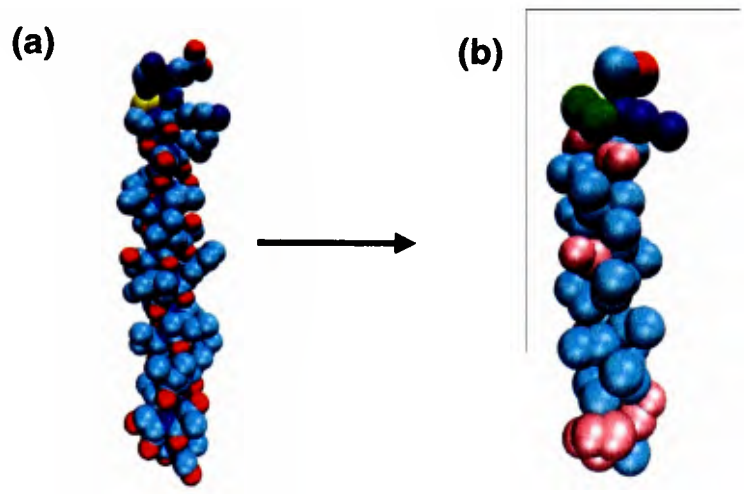


Figure 2.2 The M2 helix from the α subunit of the *Torpedo marmorata* structure [13] is shown in atomistic form (a) and coarse-grained form (b). In (a), the atoms are coloured using the CPK convention. In (b), the particles are coloured as outlined in Bond *et al.* [175]: green = "mixed polar/apolar"; cyan = "hydrophobic"; red/blue = " positive/negative charged"; and pink = "polar". The number of atoms is reduced approximately four times in the coarse-grained representation.

2.5 Essential dynamics

2.5.1 Principle component analysis (PCA)

Principle Component Analysis (PCA) is a method used to reduce the dimensionality of data and to identify patterns in multidimensional data sets [176]. In cases of MD analysis, the major modes of motion are characterized using PCA. This is useful, as it is difficult to pick out patterns in trajectories of protein simulations to interpret the important motions of the protein [177].

For a dataset, here an MD trajectory of atomic coordinates in the form of *x*,*y*,*z*, the mean is subtracted from each of the data dimensions [176].

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \textbf{2.21}$$

Then the covariance matrix is calculated.

$$\mathrm{cov}(X,Y) = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{(n-1)} \qquad \textbf{2.22}$$

The covariance is always measured between two dimensions so if we have a dataset with more that 2 dimensions, we will have more than one covariance [176]. Therefore for an *n*-dimensional data set, we have:

$$\frac{n!}{(n-2)!*2} \qquad \textbf{2.23}$$

different covariance values. We use the eigenvectors and eigenvalues of the covariance matrix:

$$C^{mxn} = \left(C_{i,j}, C_{i,j} = \mathrm{cov}(Dim_i, Dim_j)\right) \qquad \textbf{2.24}$$

which tell us how the data sets are related to each other. The eigenvector with the *highest* value is the principle component of the data set. The eigenvectors are ordered from highest to lowest which gives the components in their order of significance or the most important modes of motion of a protein [176].

2.5.2 Grouping of eigenvectors

Principle component analysis from MD simulations is often very crowded with information. The server DYNAMITE [178] analyzes MD trajectories (or

CONCOORD data) and generates covariance plots or porcupine depictions of the first few modes of motion (principle components). These representations are also usually very complicated and hard to comprehend, especially for large proteins such as the nAChR and related LGICs. At the same time, understanding the major modes of motion of these proteins is of great importance because it would shed light on how a large protein of such arrangement functions. It is useful to be able to look at the motion of certain larger sections of the protein to see how they move with respect to each other. This is helpful in understanding how pentameric proteins function in general, and may be applied to hexamers, heptamers, etc.

I have written a program (EigenGrouping) to take covariance information from MD simulations and group the resultant eigenvectors according to the specifications of the user. Here the PCA data is grouped such that one can see the major modes of motion in different parts of a protein, such as different domains.

2.6 AutoDock

Autodock [157] is used to dock small molecules onto proteins. It is useful in studying drug candidates and looking at possible binding modes of a ligand. Autodock 3.0.5 was used to carry out all the docking studies in this thesis. Autodock consists of three separate programs: Autotors, Autogrid, and Autodock, which carry out the docking in steps. Step 1 (Autotors) defines the rotatable bonds of the ligand. Step 2 (Autogrid) calculates the three-dimensional grid of interaction energy based on the receptor's coordinates. The user can define the grid spacing in the 3D interaction grid. Grid spacing defines points on a grid map, centered on a selected region of the macromolecule, which store the potential energy of a 'probe' atom to all the atoms in

the macromolecule [179]. Step 3 (Autodock) carries out the actual docking based on the grid parameters and bond information established in the first two steps.

The free energy scoring function is based on the AMBER force-field [151], regression analysis, and a large set of diverse protein-ligand complexes with known inhibition constants. Simulated Annealing (SA) is the original search method employed from the earliest version of Autodock, however Autodock 3.0.5 allows for the genetic Algorithm (GA) method which is more efficient than SA.

The user defines a rectangular box around the receptor, specifies the rotatable bonds, and a starting configuration, and Autodock generates the docks.

A GUI has been made to facilitate preparing and running Autodock jobs. AutodockTools (ADT) http://www.scripps.edu/mb/olson/doc/autodock/tools.html allows the user to visualize the grids, and run the three steps via the GUI rather than from command line. It also allows for analysis of the output once Autodock has finished running. For all Autodock runs in this thesis, the macromolecule files were prepared using ADT, the ligands were also prepared using ADT, however, ligand charges were obtained from Spartan [180]. 50 Genetic Algorithm (GA) runs were carried out for each dock, with 100x100x100 grid points and 0.375 Å grid spacing (lower for acetylcholine – see chapter 4). Complete details about the docks carried out are in chapter 4 of this thesis.

2.7 Poisson-Boltzmann calculations with APBS

The *potential of mean force* (PMF) is the free energy required to position a particle, such as a water molecule or an ion, at position $\mathbf{x}_1$ relative to a reference state. Carrying out a PMF calculation is highly CPU intensive, therefore to save time and computational resources an estimate of this value can be obtained from a static

structure using *Poisson-Boltzmann* calculations. Here only the 'fixed' or static charge of the protein interacting with an ion or another protein is calculated as interactions are reduced to electrostatic contributions only. The solvent is treated as a continuum to reduce the size of the calculation and averaging effects are incorporated. Water is a continuous medium with a dielectric constant $\varepsilon_w \approx 78.5$ and proteins are given $2 \leq \varepsilon_p \geq 16$.

The Born energy is the energy it takes to move an ion of a certain radius from vacuum to water. The Poisson equation is concerned with finding the electrostatic potential for a set of fixed charges i.e. the protein. Mobile ions are treated as continuous distributions as in the Debye-Huckel theory [181]. Their electric potential is derived from the Boltzmann distribution, and the total charge $\boldsymbol{\rho} := \boldsymbol{\rho}_f + \boldsymbol{\rho}_+ + \boldsymbol{\rho}_-$ ($\boldsymbol{\rho}_f$ is the charge of fixed atoms, $\boldsymbol{\rho}_+$ is the contribution of the mobile positive ions, and $\boldsymbol{\rho}_-$ represents the charge of the negative ions) gives us the *Poisson-Boltzmann* equation.

The program Adaptive Poisson-Boltzmann Solver (APBS) [182] is used to solve the Poisson-Boltzmann equation by discretizing the problem on a grid. Sample points along the pore axis at which to place the ion were derived from the program HOLE [183]. PDB2PQR [184] is then used to assign partial charges and radii for the atoms based on the CHARMM22 force-field. The energy of a cation or anion placed at a sample point on the pore centre line identified by HOLE for each data point (sample points are usually 0.5 Å or 1 Å apart along the *z*-axis) is calculated.

3 Simulation studies of AChBP with nicotine, carbamylcholine, and HEPES

3.1 Introduction

AChBP is the first LGIC ligand binding domain homologue to have its structure determinted by X-ray crystallography. It has since been the subject of many studies, including being co-crystallized with various ligands and from different species. Running MD simulations of the AChBP allows us to study to the structural dynamics of this protein and possibly gain insight into the dynamics of the nAChR and other LGICs. It is also a useful complement to the modelling work done on the α7 nAChR (chapter 5) and α7 nAChR EC domain simulations (chapter 6).

There are advantages and disadvantages of using AChBP to study the function of the nAChR and other LGICs. The advantages include the structural similarity, the high sequence identities (20-26% [38, 77]), the atomic resolution of the structure, and the availability of ligand bound structures with several different ligands [20, 38, 75-78,

185] which are also ligands of the nAChR. Thus, ligand-protein interactions may be well characterized using AChBP as a template to study LGICs.

Obvious disadvantages are the fact that the AChBP is not a membrane protein or an ion channel. The AChBP lacks the TM and IC domains of the LGICs and thus the structural changes resulting from ligand binding which translate to channel gating cannot be observed using the AChBP. However, currently this is the best structure available to investigate the dynamics of LGICs. Using MD we can gain insight into possible modes of motion and thus mechanisms of function of LGICs. It may be most useful for finer-focussed MD studies such as the ligand-protein interactions of the binding pocket which may be easily transferable between the AChBP and LGICs.

In this chapter I will discuss 6 simulations, 3 of which have ligands bound, 3 are the same structures in a ligand-free state (Table 3.1).

Table 3.1 Summary of simulations.

Simulation	Number of Ligands	Number of atoms	Number of Water Molecules	PDB code
NCT	5 nicotine molecules	64011	17782	1UW6
NCT-apo	0 nicotine molecules	63507	17618	1UW6
CCE	2 carbamylcholine molecules (between subunits C/D, and D/E)	61151	16873	1UV6
CCE-apo	0 carbamylcholine molecules	59332	16267	1UV6
EPE	4 HEPES molecules (one missing from subunits B/C - occupied with SO_4)	39560	9621	1UX2
EPE-apo	0 HEPES molecules	65094	18163	1UX2

Here, the structural changes which take place in the presence and absence of ligands on both large scale and small scale motions are studied. The simulations carried out are all 10 ns long. This is clearly not long enough to observe the complete conformational dynamics of the protein, but we can use the MD data to look at motions beginning to take place, we can extrapolate longer range motions and also look at the interactions taking place at an atomic level. Tools such as PCA can be used to define the major modes of movement in the protein given a trajectory. Current computational resources only allow us to simulate proteins on the nanosecond time-scale at the atomic level. Large proteins such as the AChBP are more CPU intensive, thus each ns calculated on two processors took approximately 5 days to complete, taking 50 days for a 10 ns simulation. Motion analysis techniques were used to studying differences in the presence and absence of ligands. Since MD cannot completely address longer-term motions, Gaussian network models (GNM), CONCOORD, PCA, and coarse-graining of eigenvectors were used to look at the large scale motions (see Methods chapter).

The MD data was used to focus on the smaller-scale motions such as the protein-ligand interactions taking place in the binding pocket which are well characterized. Distances and dihedrals of residues in the binding pocket were measured along the trajectory to observe changes in local structure in the presence and absence of ligands. The behaviour of nicotine and carbamylcholine in the binding pocket was studied. During the analysis of the binding pocket, it became apparent that there exists a strong network of water molecules involved in bridging between the ligand to the macromolecule. A closer look was taken at the dynamics of water in the binding pocket of the AChBP in the presence and absence of ligands. The results from the small-scale analysis shows 1) A more stable structure of the binding pocket is

maintained in the presence of a ligand 2) Waters exist in discrete pockets inside the binding pocket that may be useful in the design of ligands. These water molecules act as bridges between ligand and protein; they also create a network of hydrogen bonds that has structural implications and 3) Ligands exhibit varying degrees of mobility within the binding pocket, nicotine seems to prefer one mode of binding and shows little flexibility, while carbamylcholine shows some flexibility of its chain and specifically the quaternary nitrogen CH_3 groups. Modes of binding seem relatively stringent, with some degree of flexibility dependent on the structure of the ligand.

3.2 Methods

The six protein molecular dynamics (MD) simulations carried out in this chapter corresponding to both ligand-bound and ligand-free states of AChBP are listed in Table 3.1 (simulation names from this table will be used for the remainder of this chapter). Crystal structures of AChBP (see Table 3.1) were used with their crystal waters included. Ligand-free structures were generated by removing the ligands but retaining any crystal waters within the binding pocket. The simulation with nicotine had 5 nicotine molecules bound, one to each binding site. The simulation with carbamylcholine had 2 molecules bound to the binding pockets located between subunits C and D and subunits D and E (the three remaining binding pockets were ligand-free). The simulation with HEPES bound had 4 HEPES molecules, one missing from binding pocket in the interface of subunits B and C.

Molecular dynamics simulations were carried out with GROMACS v 3.1.4 (www.gromacs.org) [186-188] and the GROMOS96 [147] force-field parameters were used. Simulations were done under constant temperature, pressure and number of particles. The Berendsen method [158] was used for temperature and pressure

coupling during equilibration, whereas the Nosé-Hoover method [159, 160] was used for temperature coupling along with the Parinello-Rahman method [161] for pressure coupling in production runs and the thermostat was held at 310 K. Long range electrostatic interactions were evaluated with PME [156, 157, 189, 190] and LINCS [163] was used as the constraint algorithm for bond lengths. Integration time steps of 2 fs were used for all runs.

3.2.1 Preparing the ligand: generating topologies for nicotine, carbamylcholine, and HEPES

In order to run the MD simulations of AChBP with the three ligands in the crystal structure, topology files for the ligands had to be created. The initial structures of the ligands were those from the crystal structures of AChBP [77]. Nicotine was taken from 1UW6, carbamylcholine from 1UV6, and HEPES from 1UX2.

It is believed that the pyrrolidine nitrogen in nicotine, carbamylcholine and HEPES is a positively charged quaternary N group [20, 191, 192], which is the case in acetylcholine, the endogenous ligand of nAChR. The crystal structures do not contain information on the protonation state of the ligand, therefore, before carrying out charge calculations to generate the topology, InsightII (Accelrys, San Diego, CA) was used to protonate the ligands. Once the correct ligand structure was generated, PRODRG [193] (http://davapc1.bioch.dundee.ac.uk/programs/prodrg/) was used to assign atom types and initial backbone of the topology file for each ligand. However, PRODRG at that time did not produce GROMOS96 topology files, which was the force-field of choice for these simulations. Refinements had to be made so that the atom-types, bond-types, and dihedrals matched that of the GROMOS96 force-field. Spartan [180] was used to assign atomic charges to the ligands. Although computationally costly and relatively slow, this is the most accurate way of assigning

charges to molecules because the charges are calculated from first principles; they are not a generalization or estimation, and not based on data from other structures. Spartan adds on all hydrogen atoms and produces charges for each separately. The charges for the non-polar hydrogens have to be combined with the root atom of that group as GROMOS96 is a united-atom force-field and does not account for non-polar hydrogens separately. Charges for nicotine, carbamylcholine, and HEPES were calculated with 6-31G* basis sets using Spartan, and the charges replaced the charge estimates suggested by PRODRG.

Once the topology file is generated, the ligand is energy minimized and simulated in a box of water for 1 ns. This short simulation allows for the study of the behaviour of the ligand and ensures that the bonds and angles of the ligand meets with the expected stereochemistry of the molecule [191, 194-196].

The topologies generated were stable in both the energy minimization and production runs. Thus the ligand (Figure 3.1) is combined with the protein to carry out the MD runs of AChBP with ligands.

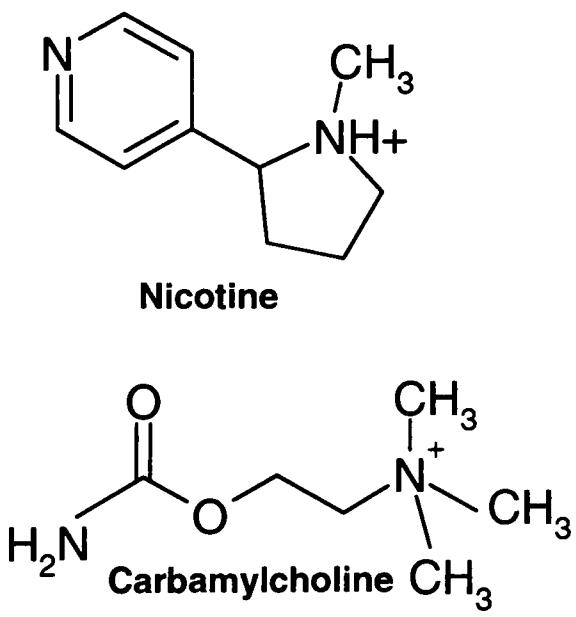


Figure 3.1 The ligands used for simulations in this chapter, nicotine, carbamylcholine and HEPES, are shown. Topology files are available on CD.

3.2.2 Preparing the macromolecule

For the ligand-bound simulations, the ligands were positioned according to their respective crystal structures. The ligands were removed from the crystal structure or the ligand-free simulations (during the first few hundred ps of simulation, water molecules fill the cavity where the ligand was held). Each system was energy-minimized using < 1000 steps of steepest descent method until convergence. Energy minimization was carried out with position restraints, then molecular dynamics with position restraints for 100 ps, followed by energy minimization without position restraints, then equilibration for 1 ns, and finally the production run was carried out for 10 ns for each simulation.

The 2.2 Å crystal structure of AChBP (PDB: 1UW6) has one nicotine molecule per binding site. Each of the five nicotine molecules were protonated and included in the simulation in the position of the crystallographic nicotine molecules. All crystal waters were included in the simulation and counter-ions were added to neutralize the system. This simulation was later extended for another 10 ns, however, the analysis presented here involves the first 10 ns only.

The 2.5 Å crystal structure of AChBP, with carbamylcholine bound, has two pentameric AChBP molecules in the deposited PDB file (1UV6). One of these has two carbamylcholine molecules bound and the other has one. I chose the AChBP with two ligands bound for the simulation as it is believed that the open (activated) state of the Cys-loop receptors is associated with two bound agonists [197, 198]. Similar to the 1UW6 simulation, all crystal waters were kept for the simulation. For the CCE-apo simulation the crystal structure was stripped of the two carbamylcholine molecules to obtain a ligand-free state. This allows us to compare the difference in behaviour in the presence and absence of carbamylcholine.

Of the three structures used in this study, the crystal structure of AChBP with HEPES bound has been resolved at the highest resolution (2.1 Å). There are 4 HEPES molecules bound to the binding pockets of the protein, the binding pocket between subunits B and C has an ammonium sulphate molecule instead of a HEPES molecule, the ammonium sulphate is removed for the MD simulation. HEPES is present in the crystallization buffer and it is not an agonist or antagonist of AChBP or any of the LGICs. Thus the HEPES bound AChBP is believed to be structurally similar to the desensitized conformation of LGICs [128], however, this is not yet confirmed. The HEPES molecules are bound in two different orientations, both observed in previous studies [20, 40, 77].

3.2.3 Analysis of MD trajectories

The GROMACS suite of programs was used for analysis of the MD trajectories. Protein-ligand interactions were visualised with LigPlot [199] and VMD [200]. Distance and dihedral calculations were carried out with VMD. Persistent water molecules were identified with an in-house program, WatRes, developed by John Holyoake. The water density calculations are carried out using Gromacs 3.2.1 program g_ri3D developed by Oliver Beckstein and the related picture and several others are generated with Chimera [201]. Principal component analysis (PCA) was carried out using Gromacs programs, in particular g_covar and g_anaeig. Covariance and porcupine plots were generated with Dynamite/DynaTraj [178]. Gaussian network models [164, 166, 167] and a program I wrote for the simplification of eigenvectors 'EigenGrouping', are used as coarse-grained approaches for studying the large-scale motions of AChBP. Ligand diagrams are drawn with ISIS/Draw (http://www.ch.cam.ac.uk/cil/SGTL/MDL/ISISdraw.html).

3.3 Results

3.3.1 Global behaviour of AChBP

All simulations are stable after an initial rise in the first ~2.0-2.5 ns of their root mean square deviation (RMSD) values. For the duration of the simulations, all ligands remain inside their respective binding pockets. Overall, the subunits of the protein exhibit asymmetrical movement and there is no obvious single mode of motion that can be discerned visually. Such asymmetry has also been reported for simulations of a model of the α7 nAChR EC domain pentamer [129]. Table 3.2 displays the average Cα RMSD values, calculated for the final nanosecond of each simulation discussed in this chapter.

Table 3.2 Average (final ns of simulation) RMSD values for each simulation

Simulation name	Cα RMSD (nm) – averaged over the final ns of simulation
NCT	0.268
NCT-apo	0.266
CCE	0.247
CCE-apo	0.274
EPE	0.234
EPE-apo	0.247

The RMSD graphs for each simulation with ligands bound and their corresponding ligand-free simulations are shown below (Figure 3.2 and Figure 3.3). Overall, the structures are stable with little deviation from the original structure with RMSD values as expected. The RMSD graph for NCT and apo-NCT is shown in Figure 3.2 (a); Figure 3.2 (b) displays the RMSD values for each subunit. There is not a significant difference between the ligand-bound and ligand-free RMSD values for these two simulations, which is likely due to the short time of the simulation.

In Figure 3.2 (b), all subunits are stable and fluctuate only slightly from the original structure. The overall RMSD is slightly higher than for the RMSD of each separate subunit. This is because the overall RMSD also takes into account the deviations of the individual subunits with respect to each other.

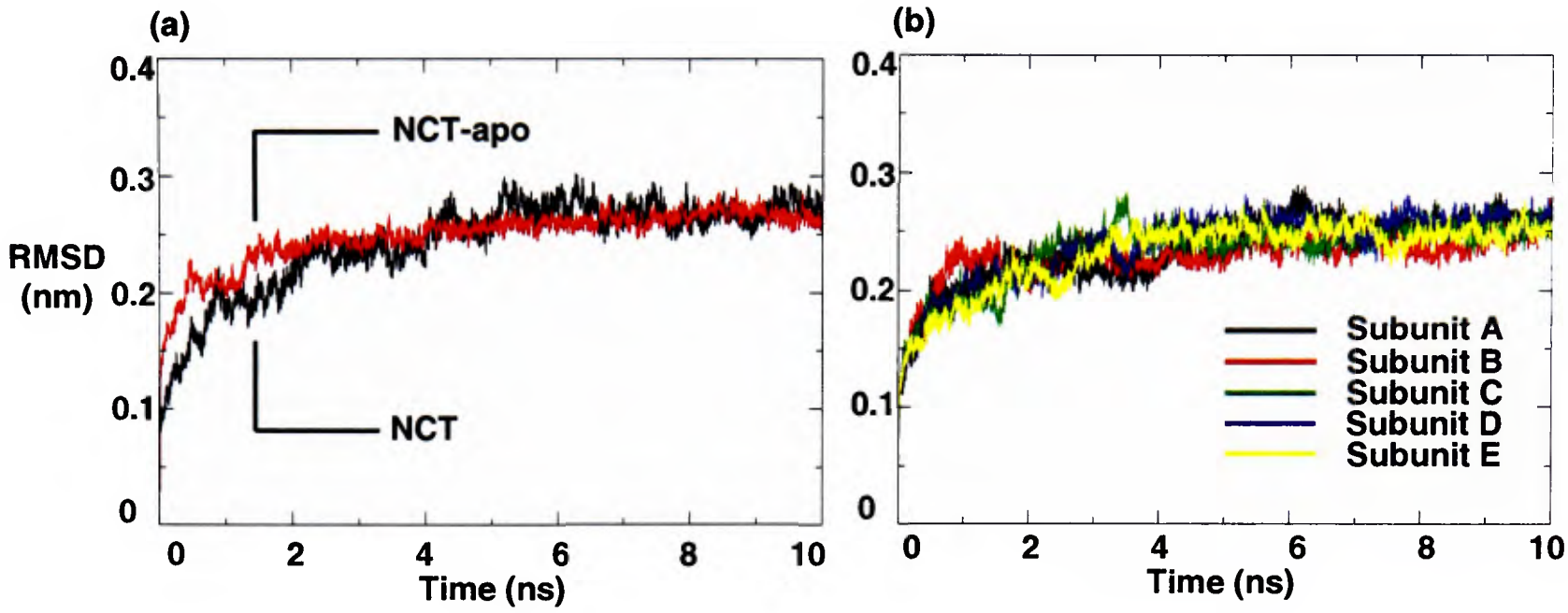


Figure 3.2 (a)The Cα RMSD of NCT bound (black) NCT-apo (red). (b) RMSD plot shown for each separate subunit of the NCT simulation.

The RMSD graphs for CCE and EPE simulations with ligands bound and their corresponding ligand-free simulations, are shown in Figure 3.3 (a) and (b). The EPE and CCE simulations are also stable with slight deviations from the original structures.

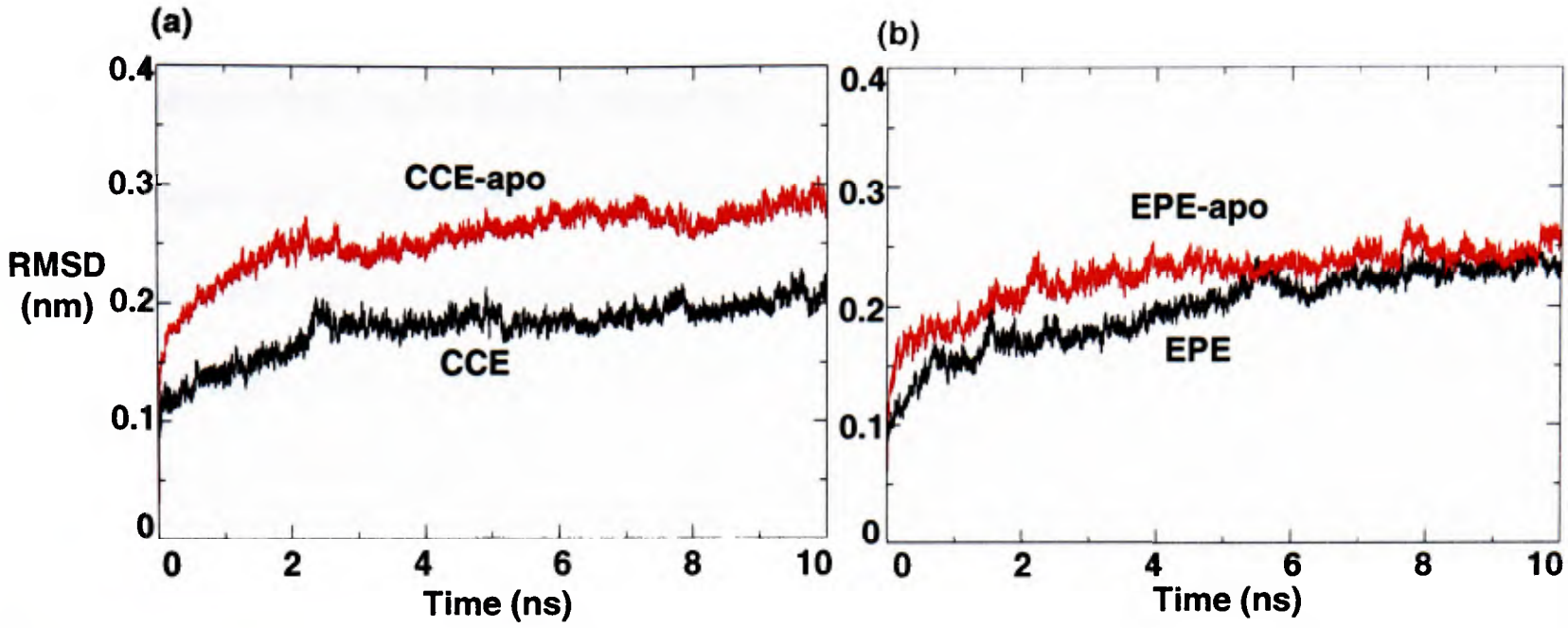


Figure 3.3 (a) The Cα RMSD of CCE (black) and CCE-apo (red). (b) The RMSD of EPE (black) and EPE-Apo (red). The structures remain stable after an initial relaxing phase of ~2.5 ns.

The HEPES bound and nicotine bound simulations have similar RMSD values for the ligand-free and ligand-bound states, while CCE has significantly lower RMSD values than its corresponding ligand-free simulation. This could be attributed to the number of ligands in the crystal structure such that the fewer ligands in CCE (two carbamylcholine molecules) results in a larger RMSD in the ligand-free state compared to four or five ligands in EPE and NCT respectively. The higher number of bound ligands in these crystal structures could have stabilized the protein, resulting in lower RMSD values and the short duration of these simulations does not allow for the structural reconfiguration to the 'biological' ligand-free state. For example, HEPES is not expected to activate, i.e. result in major structural changes in the protein.

Figure 3.4 shows the average displacement values for the residues from CCE and CCE-apo. These values indicate the displacement of each Cα atom from the reference X-ray structure. Only subunits A and B are shown for both simulations. The displacement values were averaged for the five subunits. The highest displacement in the ligand bound state is at the termini, specifically the N-terminal alpha-helices at the

top of the protein and the C-terminal loops. The tip of loop C (see Figure 3.4) also shows high displacement values from its original structure, but not as much as the ligand-free state where the β-strand structure of this loop is not maintained and seems to have very high displacement values. This is interesting because only two of five subunits have a ligand in CCE but all subunits in the ligand-bound simulation have intact loop C structures, thus the presence of a bound ligand affects the structure of the entire protein complex. Loop C has higher displacement values for the ligand-free state in all three sets of simulations. CCE-apo has higher displacement for most regions, the β-strands have moved more than the ligand-bound state where the structure has lower displacement at the core, and the higher movement is at the termini and the loop regions.

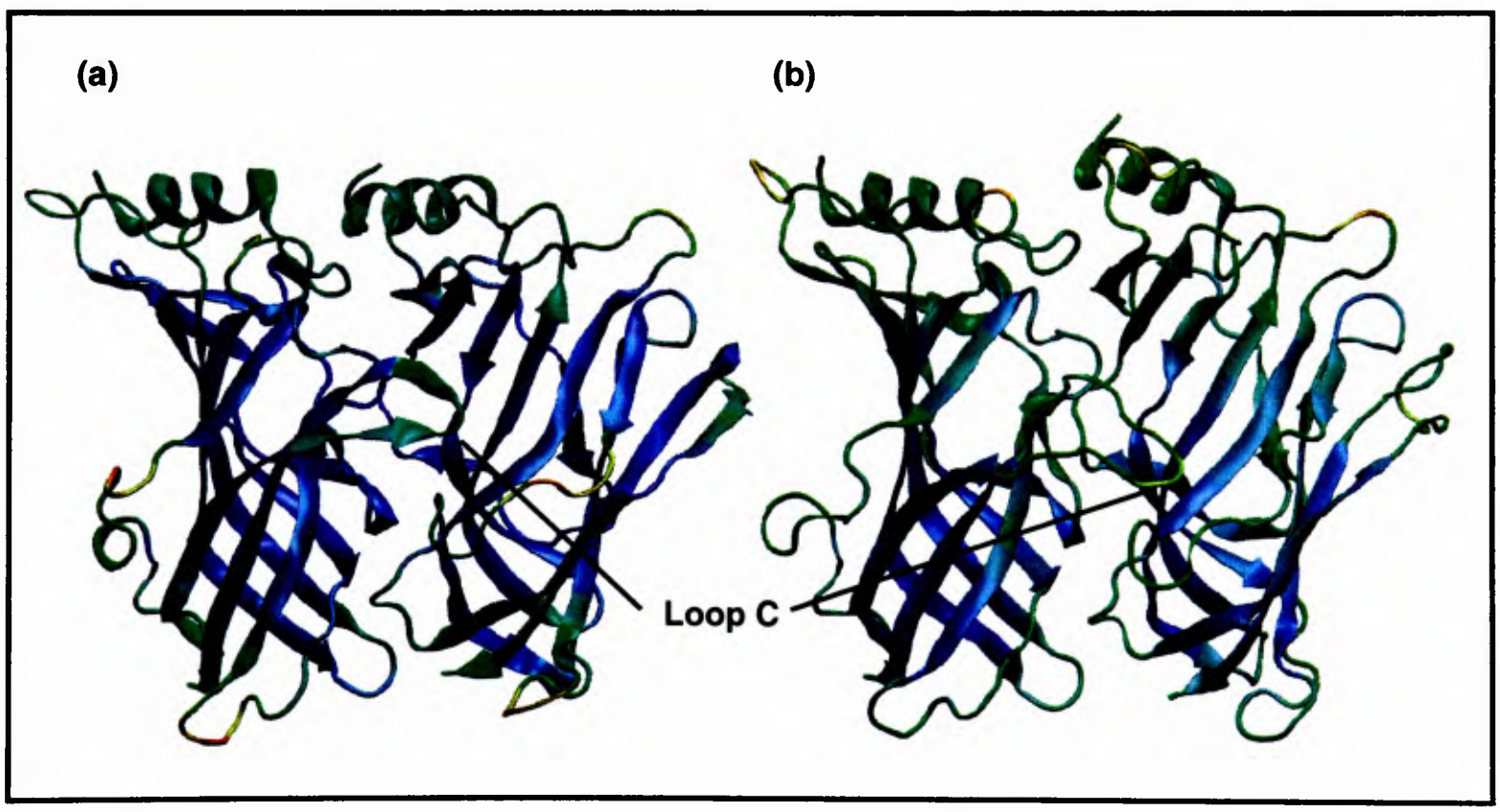


Figure 3.4 (a) AChBP with carbamylcholine bound, coloured with average Cα atom displacement values from the starting structure, averaged over all 5 subunits of CCE (Red - high displacement, blue - low displacement). (b) CCE-apo coloured with average Cα atom displacement values, averaged over all 5 subunits. The ligand-free state has higher average displacement in the core β-sheet segments (all of loop C). The ligand-bound state has higher average displacement at both termini regions and at the tip of loop C, whereas the core regions have very low displacement values.

The difference in displacement between the ligand-bound and ligand-free states is not as clear in the NCT and EPE set of simulations. The loops have higher displacement in the ligand-free states i.e. loop C, but the higher displacement in the core regions of CCE-apo is not as apparent in the other two apo simulations.

3.3.2 Large scale analysis

3.3.2.1 Structural flexibility and convergence

MD data for the large scale motions of AChBP displayed asymmetry among the different subunits. Although AChBP is a homo-pentamer with five-fold rotational symmetry, the subunits do not have identical dynamics. The movement was more pronounced in two of the five subunits. Most of the movement is seen at the termini for the both the ligand-bound and ligand-free simulations, while the central β-sheet regions are generally more rigid in the ligand-bound states, specifically in the binding pocket regions. The ligands kept their location and orientation throughout the 10 ns simulation. None of the ligands rotated during the simulation, indicating the strong interactions and favourable position from the starting structure.

RMSF plots of NCT against CCE, and each against their ligand-free states (NCT-apo, CCE-apo) are shown in Figure 3.5 (a) NCT (black) and CCE (red) RMSF values averaged along the five subunits. NCT has slightly higher RMSF values for both the core segments and the loops. (b) RMSF plot of NCT-apo (black) and CCE-apo (red), both ligand-free simulations show very similar RMSF values. and Figure 3.6. NCT and CCE have very similar RMSF values when plotted against each other (Figure 3.5 (a)). NCT-apo and CCE-apo also display similar RMSF values (Figure 3.5 (b)). The ligand bound forms display lower RMSFs for the core secondary structure regions of the protein but higher RMSF for some of the loop segments such as the Cys loop.

The Cys loop is believed to play a key role in the gating of Cys-loop receptors, it is highly conserved with two cysteines approximately 14 residues apart; AChBP shares this feature with the Cys-loop receptors. Other segments with high RMSF values for all simulations are the β1-β2 loop, the top of loop A, and loops B and C.

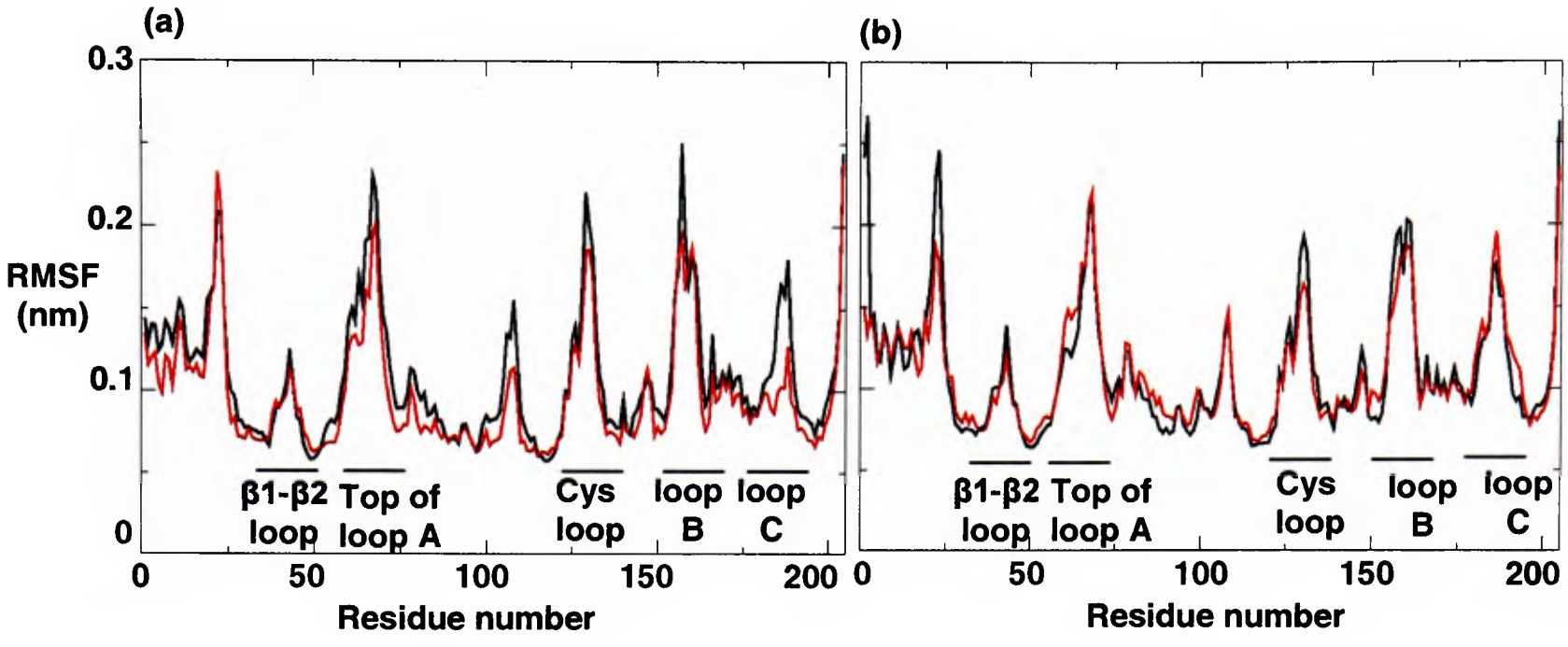


Figure 3.5 (a) NCT (black) and CCE (red) RMSF values averaged along the five subunits. NCT has slightly higher RMSF values for both the core segments and the loops. (b) RMSF plot of NCT-apo (black) and CCE-apo (red), both ligand-free simulations show very similar RMSF values.

NCT RMSF values plotted against NCT-apo are very similar (Figure 3.6 (a)). This suggests that 10 ns may not be sufficient to observe major structural changes for this particular protein. CCE and CCE-apo also exhibit similar RMSF values (Figure 3.6 (b)), however, the ligand-bound state has lower RMSF values except for the Cys loop, this is also the case for NCT vs NCT-apo. Most of the core segments of NCT and CCE have lower RMSF values than their ligand-free states, however some of the functionally noted loops such as the Cys loop, loop B, and loop C (in NCT only) have greater RMSF values than the ligand-free simulations. This is further evidence of the probable functional relevance of these loops and their altered behaviour and higher mobility in the presence of a ligand. Thus, AChBP's core is less mobile in NCT and CCE while some of the loops have increased fluctuations.

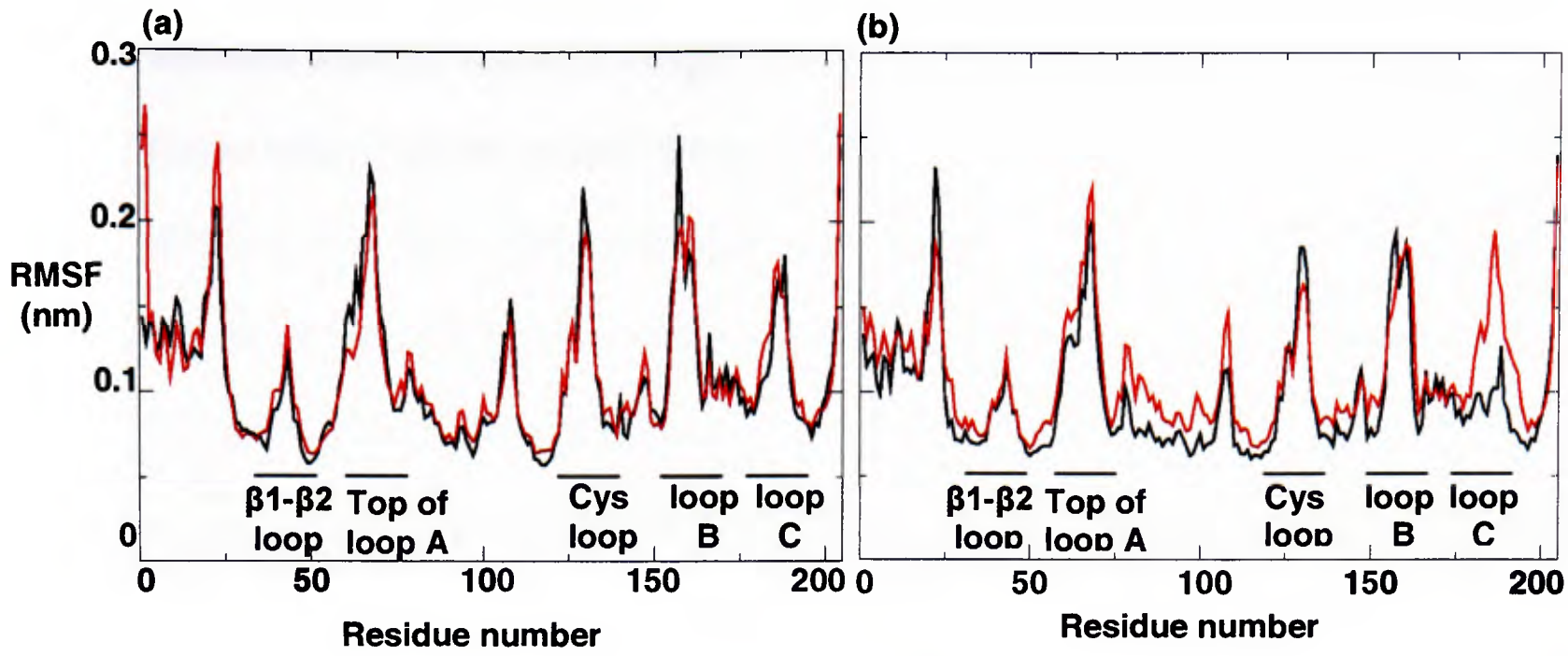


Figure 3.6 (a) NCT (black) and NCT-apo (red) RMSF values averaged along the five subunits. The values are very similar for the ligand-bound and ligand-free simulations, however, most sections of the core have higher RMSF values in NCT-apo. (b) CCE (black) and CCE-apo (red) RMSF values averaged along the five subunits. The ligand-bound simulation has lower RMSF values in most segments of the protein.

The overall dynamics of the protein can be assessed by the *block-analysis* of the mean square fluctuation (MSF) values of the Cα atoms (see Methods chapter). The final 7.5 ns of each simulation with time windows of 1.0, 2.0, 4.0, and 7.5 ns (Figure 3.7) [202] are used for the block-analysis. This calculates the average fluctuation for a given window ('block') of the simulation which illustrates whether the protein fluctuation changes in each block or if it stays in the same conformation in the given window. In Figure 3.7, the simulations without ligands display higher mean square fluctuations than the ligand-bound simulations, perhaps indicating some increased structural stability resulting from the presence of the ligand. The non-zero slopes of the lines indicate that no simulation has converged RMSF values. CCE has the lowest mean square fluctuation, followed by NCT, then CCE-apo and next is EPE. Since HEPES is not a high affinity agonist for the AChBP, it is not certain that the protein would respond as it does to nicotine, acetylcholine or any other high affinity ligand. It seems that the presence or absence of HEPES does not result in substantial differences in the dynamics of the AChBP, which agrees with experimental data suggesting that

unbound subunits resemble HEPES bound subunits [41, 77]. EPE and EPE-apo (cyan) merge together towards the end of the simulation, but NCT and CCE are quite different than their ligand-free states for the duration of the simulation. In all simulations, the MSF values continue to increase which indicates that the protein is still going through structural changes. The actual gating of the homologous nAChR happens on a ms time-scale, therefore it is not expected that a 10 ns simulation would show the full range of motion for the AChBP.

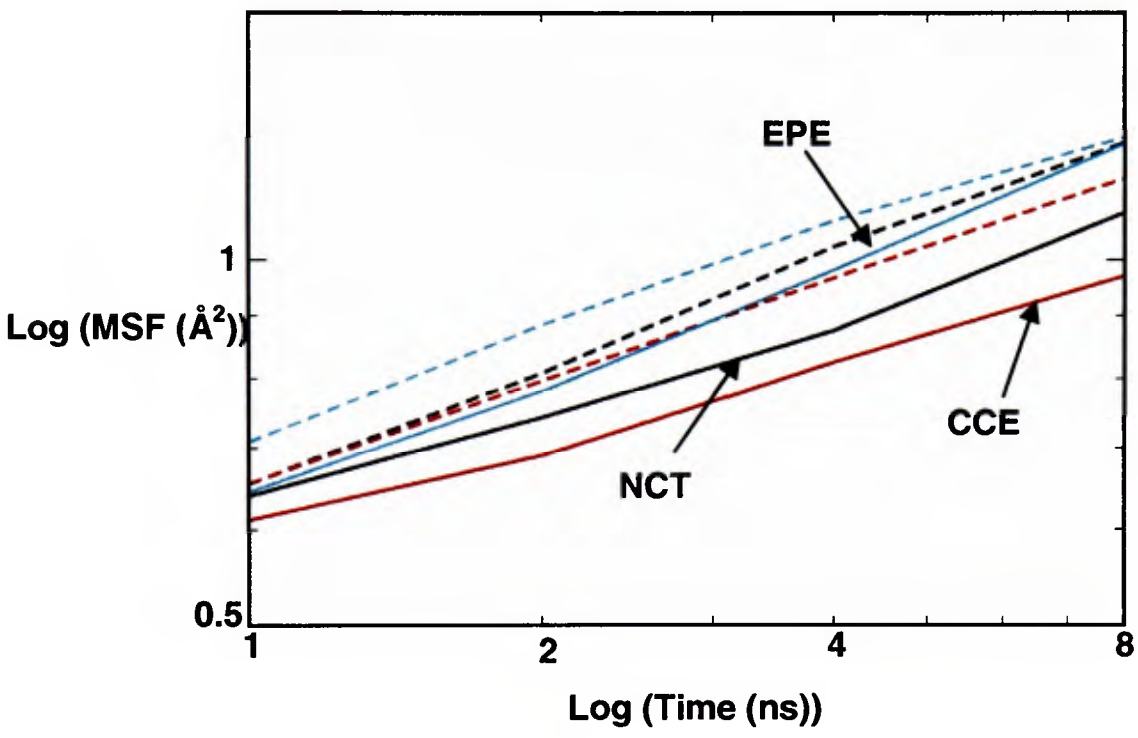


Figure 3.7 The mean square fluctuations (MSF) of the six simulations of AChBP are shown. Three simulations with ligands (CCE:solid red, NCT:black, and EPE:cyan lines) have lower mean square fluctuations than their corresponding ligand-free simulations (dashed red, black, and cyan lines).

The radius of gyration in Figure 3.8 is calculated for the three sets of simulations to observe possible distortions and changes in relative subunit arrangement. This value describes how the area of a cross-section is distributed about a given axis. AChBP with nicotine (NCT) and HEPES (EPE) shows similar values for the ligand-bound and ligand-free states of the simulation but with carbamylcholine, the ligand-bound state has a higher radius of gyration than that of the ligand-free simulation. In all cases, the ligand-bound simulations have a smooth and slight decrease in their radius of gyration during the length of the simulation. For NCT-apo there is a jump at ~6 ns and CCE-apo is showing a fluctuating decline in Rg values. An initial and considerable rise in the radius of gyration for EPE-apo in the first few frames of the simulation, (Figure

3.8 (b)) is the result of the outward movement of two loops near the alpha-helix at the top of the protein. The gyration values for EPE-apo continue to fluctuate throughout the simulation. NCT and CCE have very similar values for their ligand-bound states with a stable decrease in Rg values.

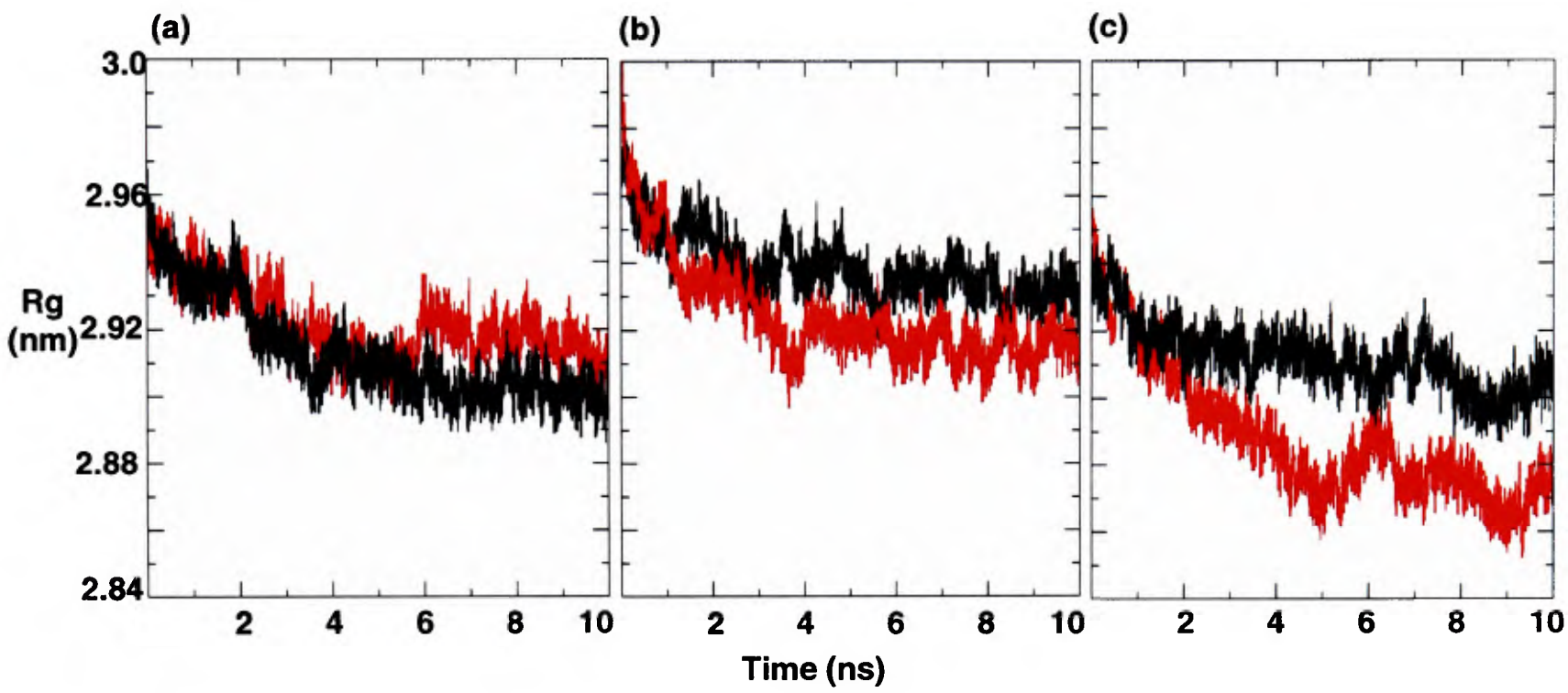


Figure 3.8 Radius of gyration of for AChBP simulations. (a) NCT (black) and NCT-apo (red). (b) EPE (black) and EPE-Apo (red). (c) CCE (black) and CCE-apo (red). Ligand-bound simulations show lower fluctuation and more consistency in their radii of gyration.

3.3.2.2 Principle component analysis (PCA)

Principle component analysis (PCA) is a statistical tool based on the diagonalization of the covariance matrix of the Cα coordinates of a protein (see Methods chapter). It is particularly useful when patterns are difficult to find, such as the major modes of motion of large proteins from MD trajectories. The projections of a trajectory on the eigenvectors of its covariance matrix are the principle components and describe the major motions. Figure 3.9 shows covariance and porcupine plots from AChBP simulations generated by DynaTraj [178, 203]. In porcupine plots, each 'spike' represents the element of the eigenvector associated with a Cα atom. The porcupine plot (Figure 3.9 (b)) shows that there is greater movement in two of the five subunits.

This is the case for all three simulations with ligands bound, even in NCT where all five subunits were occupied with nicotine. This asymmetry is also observed to a smaller extent in the non-liganded simulations which agrees with previous studies of the ligand free α7 nAChR simulations [129] (see also chapter 8).

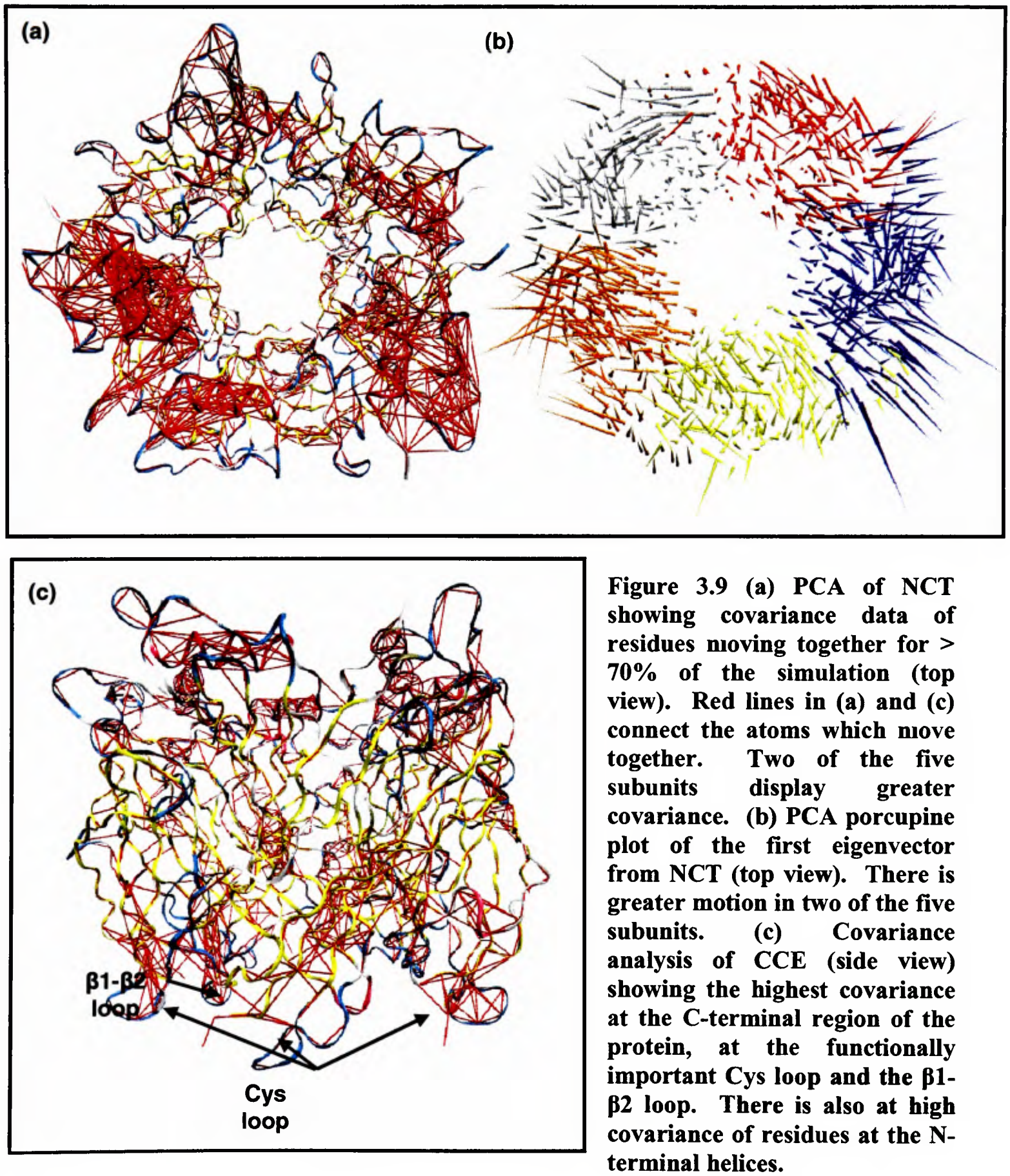


Figure 3.9 (a) PCA of NCT showing covariance data of residues moving together for > 70% of the simulation (top view). Red lines in (a) and (c) connect the atoms which move together. Two of the five subunits display greater covariance. (b) PCA porcupine plot of the first eigenvector from NCT (top view). There is greater motion in two of the five subunits. (c) Covariance analysis of CCE (side view) showing the highest covariance at the C-terminal region of the protein, at the functionally important Cys loop and the β1-β2 loop. There is also at high covariance of residues at the N-terminal helices.

PCA of the large-scale motions from ligand-bound simulations reveal a breathing motion of the protein, this has also been recently observed for a model of the α7 nAChR [130]. The breathing motion in that study was illustrated using radius of gyration calculations and could be part of the opening and closing motions of the ion channel. The projection from the first eigenvector of the simulations carried out here shows this breathing motion of the protein (see supplementary CD). The covariance analysis (Figure 3.9 (a) and (c)) identifies the residue pairs which move together during the simulation (covariance of > 70%). The highest covariance is seen at the N-terminal and C-terminal segments. In particular the functionally important Cys loop and β1-β2 loop have significantly higher covariance than the rest of the protein (Figure 3.9 (c)). High covariance is also seen at the N-terminal helices and loops, perhaps suggesting that the top segment of the protein has a key functional role and is more involved in the gating mechanism than previously thought. Intra-subunit motion is also apparent in all of the six simulations, regardless of the presence of a ligand. Each subunit displaying a circular pattern of 'spikes' running between the two ends of a subunit.

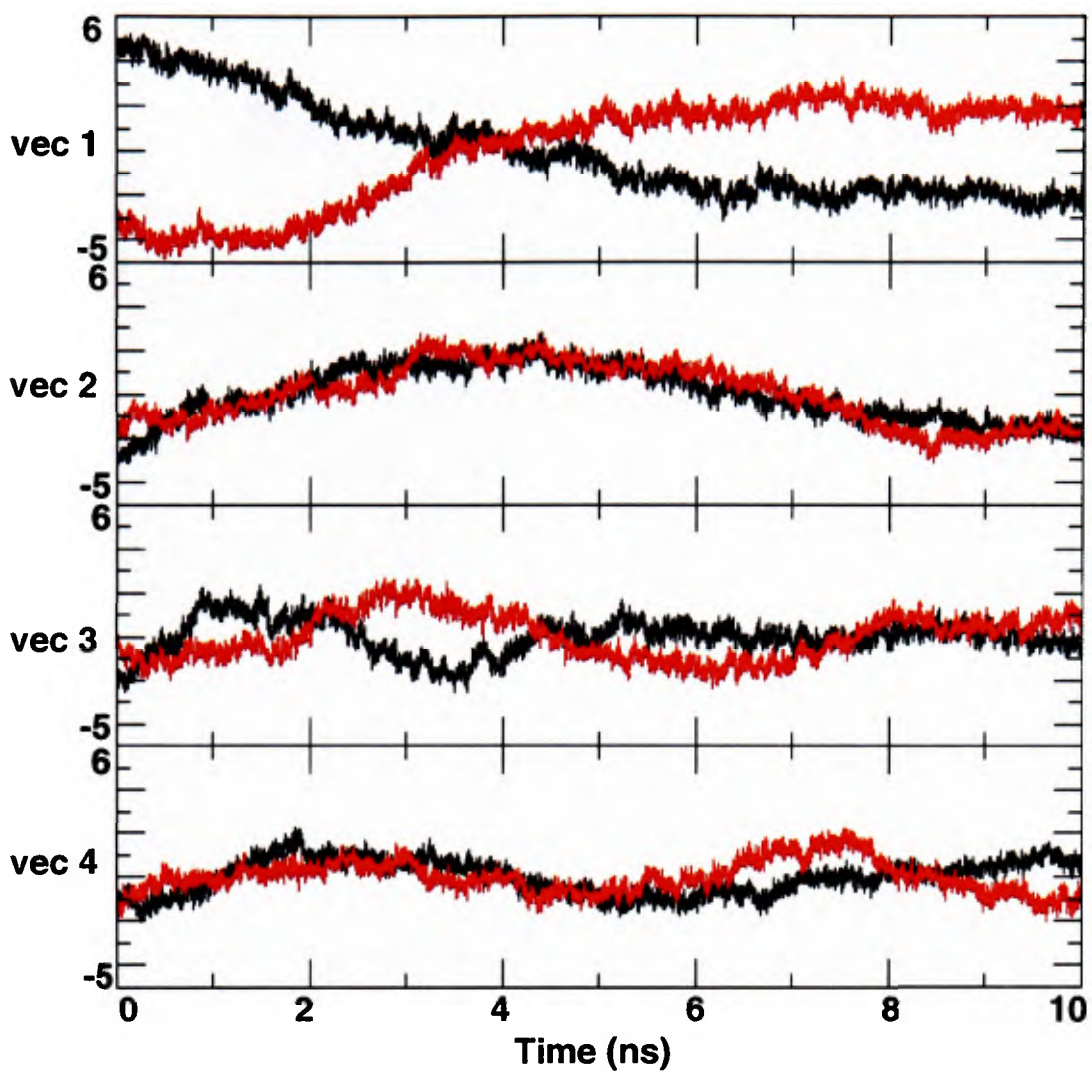


Figure 3.10 Projection on eigenvectors in nm for NCT (black) and NCT-apo (red).

The first 4 eigenvectors are shown in Figure 3.10, for the simulation of AChBP with nicotine (NCT) and AChBP without nicotine (NCT-apo). Eigenvectors 2 and 4 show identical values and thus similar motions. The first eigenvector for both simulations show similar magnitude but reciprocal values. The major mode of motion in the NCT-apo involves most loop regions and core segments with no particular symmetry of motion between the subunits (Figure 3.11 (a)), the same observations are made in CCE-apo. The ligand-bound simulations show the greatest motion at the termini segments, specifically the Cys loop and the N-terminal helices (Figure 3.11 (b)). More symmetry is apparent in the ligand-bound simulations where the major motions are more uniform across subunits.

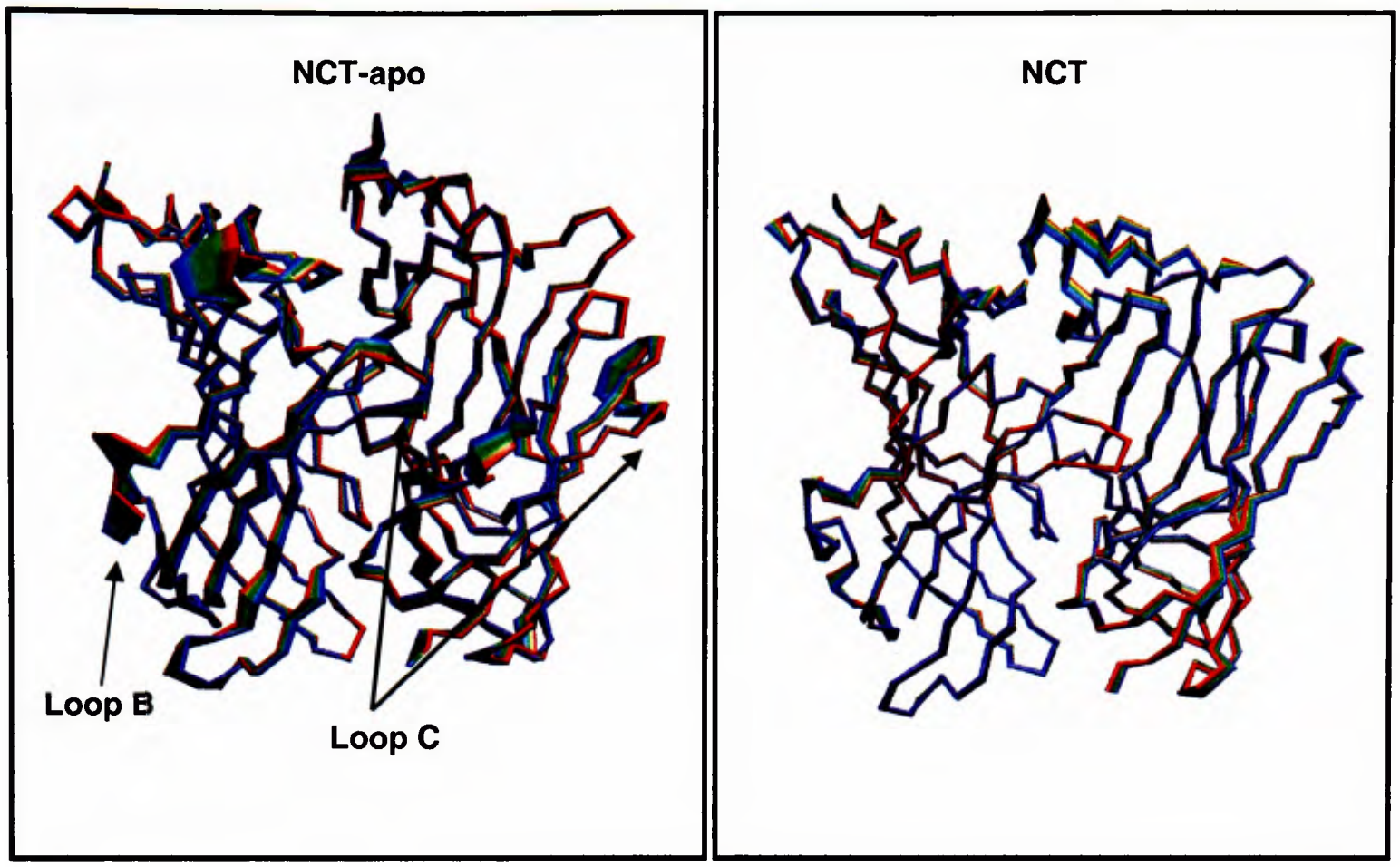


Figure 3.11 (a) Eigenvector 1 from subunits A and E from NCT-apo. (b) The first eigenvector from NCT is shown on the left of the same subunits. In the apo simulation, there is more movement in both the core regions and the outer loops.

3.3.2.3 Eigenvector-averaging

It is difficult to infer the large-scale major modes of movement from MD trajectories, especially trajectories of large proteins such as the AChBP. Although PCA simplifies MD data and allows the identification of the major modes of motion from a trajectory, it is still a challenge to study the resulting information for large systems such as that of the AChBP and other large proteins. It is of interest to observe how the movement of individual subunits are correlated with their neighbouring subunits, how movement within a single subunit is divided, and the overall interaction of subunits. If we simplify the eigenvector data from a trajectory, we may be able to better visualize the dynamics of the different segments of a protein and understand how separate domains and subunits relate to each other. For example porcupine plots, where each cone represents an element of the eigenvector associated with each Cα atom of the protein, are too crowded and the number of spikes make it difficult to visualize the dynamics

of the different segments of the protein. Therefore, I developed a program, *EigenGrouping*, to take the output eigenvector data from the Gromacs covariance analysis programs and average the x,y, and z values related to each Cα atom given a simplification factor (Figure 3.12). The user decides the factor of simplification depending on the features they are looking for. One can simplify the porcupine plot by a factor of 2 or 3 to facilitate the discernment of the data; or simplification can be done with a much higher factor, perhaps to get a few 'spikes' (eigenvectors) per domain, or subunit. In Figure 3.12 the eigenvector averaging for the first eigenvector from NCT allows the simpler illustration of the greater motions of two of the five subunits in the simulation.

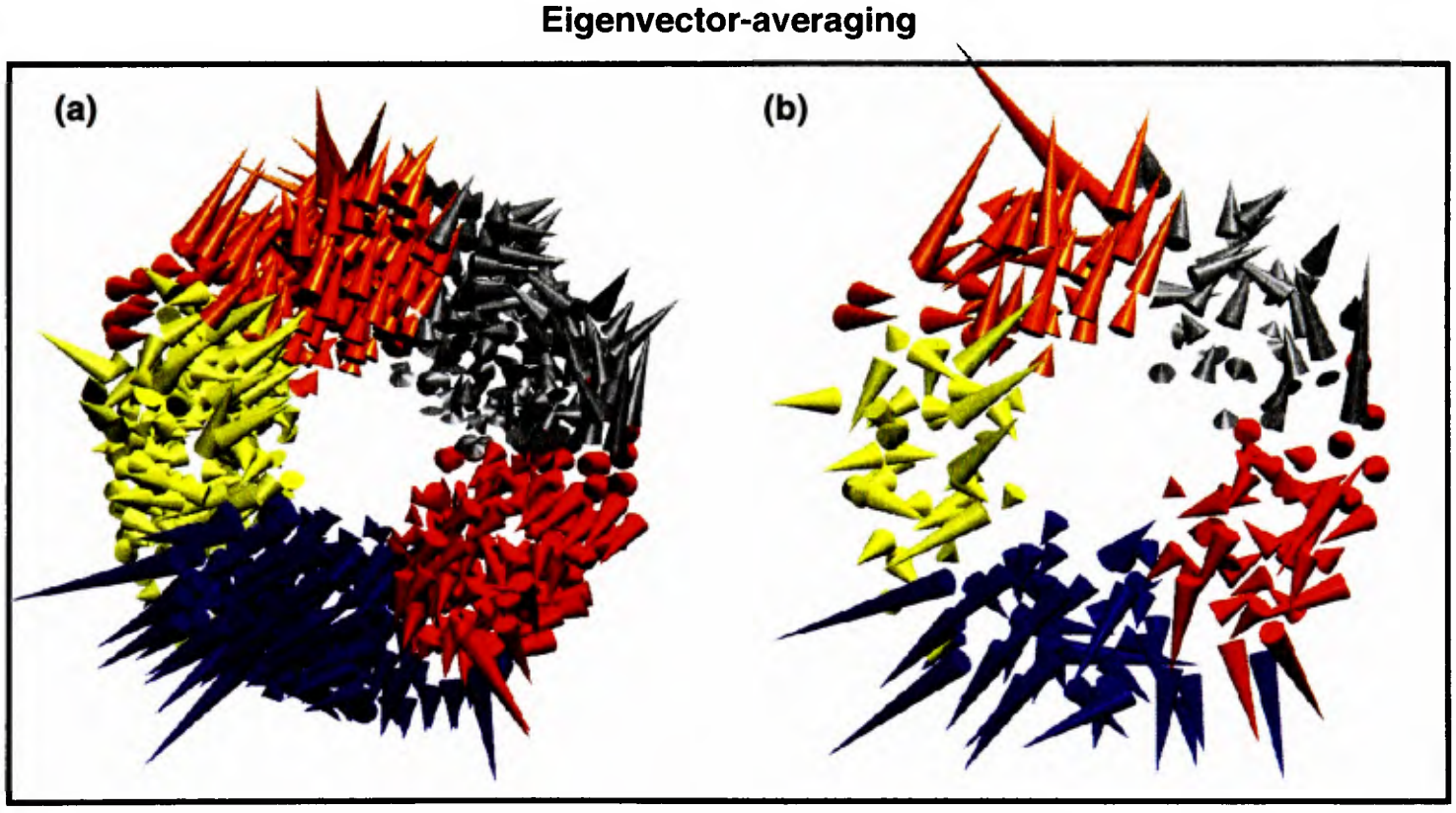


Figure 3.12 (a) Porcupine plot of the first eigenvector from NCT, showing one 'spike' per Cα atom. (b) The simplified (averaging of 'spikes') porcupine plot of the first eigenvector from NCT. A simplification factor of 5 was used.

3.3.2.4 Gaussian network models (GNM)

It is of interest to attempt to predict the large scale motions of AChBP in order to understand the inter- and intra-subunit relationships and possibly the motions

involved in the conformational transition between a closed and an open state of the nAChR. In principle, this might be achieved by long time-scale atomistic molecular dynamics simulations via MD analysis methods such as PCA. However, the absence of the IC domain from the model may result in conformational instability in such simulations in the absence of restraints (Hung and Sansom, unpublished results), and also extended timescales (>> 20 ns) would be required to see any significant motions. Therefore a more 'coarse-grained' approach, which omits atomistic detail but has been shown to allow prediction of overall patterns of mobility in a number of proteins [164, 165, 204, 205], is used in this section. GNM (See Methods) was used to predict the equivalent of crystallographic B-values (i.e. residue-by-residue fluctuations) for AChBP. Such models have been shown to give good predictions of experimental B-values for a number of proteins [164, 165, 204, 205].

GNM model calculations for the first and final frames of the three sets of simulations were carried out. From visual inspection of the calculated relative mobility mapped onto the structures from NCT (Figure 3.13), it can be seen that there are number of "hotspots" i.e. regions for which the predicted mobility is higher than for the remainder of the structure. These correspond to loop C in the EC domain, in particular the residues at the tip of this loop, as was also seen in the average displacement figures of CCE and CCE-apo. It is interesting that the linker region (where EC comes close to the TM in nAChR) is of high predicted mobility; in the ligand-bound case, the β1-β2 loop has higher mobility than that of the ligand-free structure. This loop is believed to come into close promixity of the TM domain and may be part of the mechanism responsible for transmitting ligand-induced conformational change from the EC to the TM domain in the nAChR [13, 57]. The higher mobility of loop C, Cys

loop and β1-β2 loop is also seen in the final frame of CCE, compared to the first frame.

Other hot-spots are at the upper (i.e. extracellular) segments and the loops near the TM domain and were also observed with PCA data. However, one should be cautious in over-interpreting such results in terms of the possible function of the protein in question and relating this to proteins in the LGIC family of receptors.

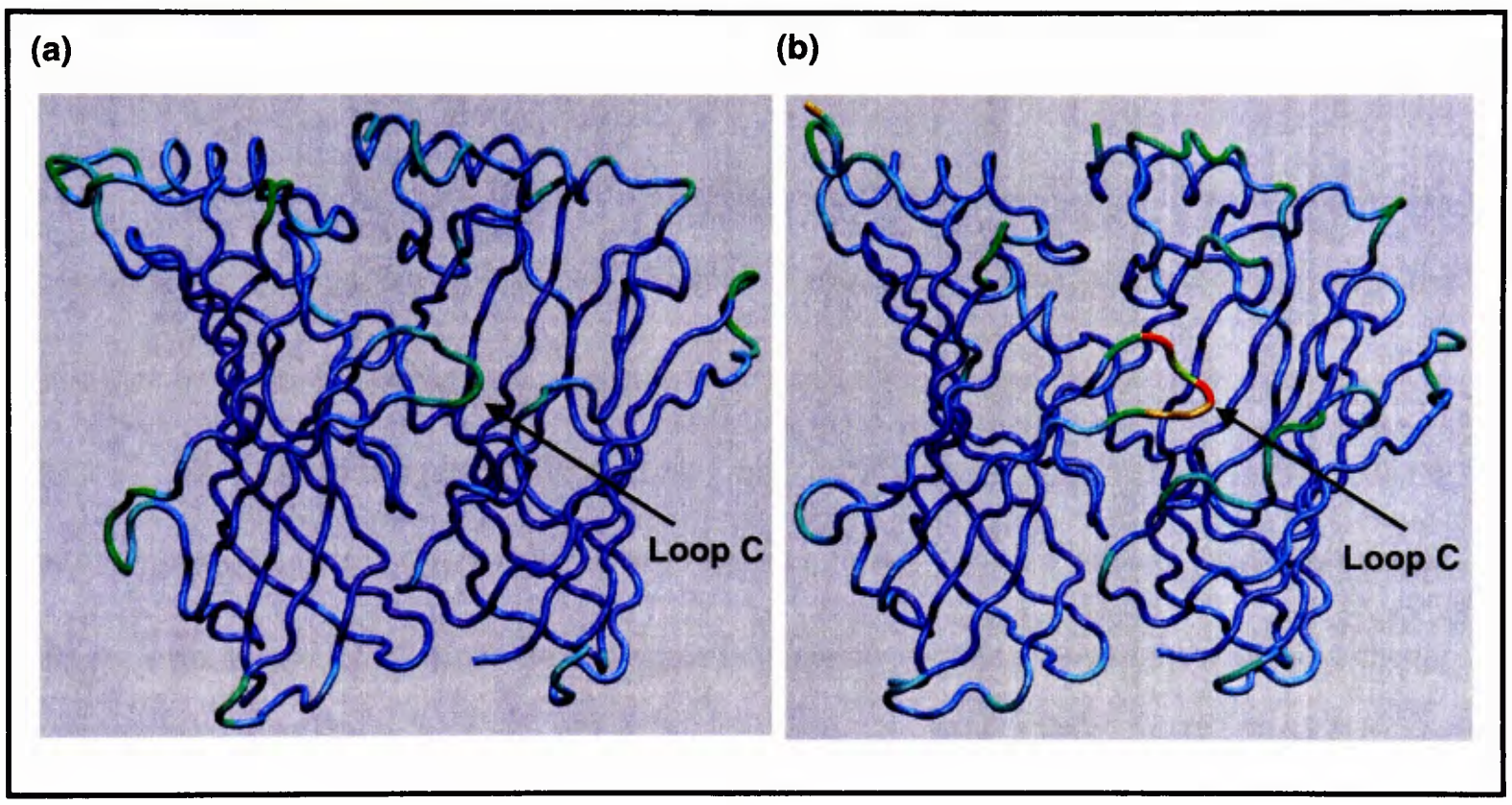


Figure 3.13 The first two subunits of the (a) first frame and (b) last frame of the NCT simulation are shown coloured with GNM theoretical B-values. The first frame is less flexible especially in loop C, and the top segments of the protein. On average, the Cys loop and the β1-β2 loop are also less flexible at the start of the simulation.

For CCE-apo and NCT-apo the first and final frames do not show major changes in mobility. The only obvious difference is the slightly higher mobility in the N-terminal segments of the protein. Also apparent in the ligand-free simulations is that the core regions of the protein have slightly higher mobility values in the final frame in comparison to the initial frame, indicating a loss of structural integrity with the loss

of the ligand. This behaviour is also reported in the PCA analysis of the simulations, as well as average displacement calculations (Figure 3.4).

3.3.3 Small scale analysis: the binding pocket

3.3.3.1 Rigidity of the binding pocket

As it is unclear how agonist binding leads to structural changes and channel gating in the homologous nACh receptor, the structural behaviour of the binding pocket of AChBP was investigated. The dynamics of residues, the ligand-protein interactions, the role of water molecules, and the network of hydrogen bonding in the binding pocket are explored in this section. Distances were measured between the ligands and residues in the binding pocket, and also between different residues of the binding pocket, to look for structural stability differences in the presence and absence of a ligand. Side chain dihedrals of residues in the binding site were calculated to look at the differences in the flexibility of residues when there is a ligand present and when there is no ligand. Dihedrals of the ligands themselves were also measured in order to observe the internal movement of the ligands inside the binding pocket. This could shed light on the modes of binding for each ligand.

Much of the binding site is buried at the interface of two adjacent subunits, one subunit makes up the 'principal' side and the other the 'complementary' subunit (Figure 3.14). Loop C from the 'principal' side, covers the ligand binding pocket and is an integral part of the binding region of the protein (Figure 3.14). This loop of LGICs has been the subject of several experimental and computational studies as it may play a key role in the binding of ligands [74, 77, 206]. This loop is believed to move in and cover the ligand once one is bound, while in the absence of a ligand its structure is less likely to be constrained.

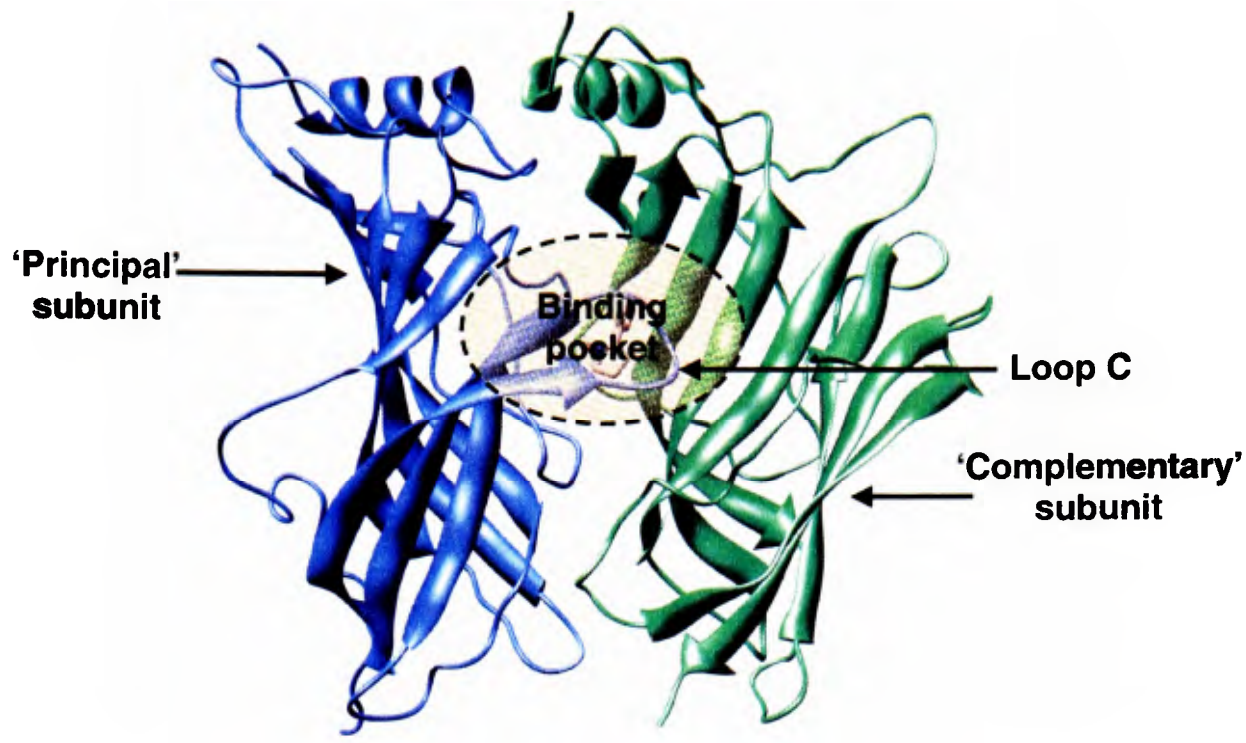


Figure 3.14 Two subunits of AChBP shown with the 'principal' and 'complementary' subunits labeled. The binding pocket lies on the interface of the two subunits and is highlighted within the dashed circle. Loop C makes up part of the 'principal' subunit for each binding pocket.

The Cα RMSD values for loop C of each simulation were calculated (Figure 3.15). The NCT-apo and CCE-apo simulations show higher RMSD values. RMSF calculations were also carried out for loop C of all simulations. Again, higher RMSF values are reported for loop C of NCT-apo and CCE-apo, suggesting less structural integrity of this loop in the absence of a ligand. Overall, there were lower RMSD and RMSF values of the binding pocket region, including loops A, B, and C in the ligand-bound simulations (Figure 3.15).

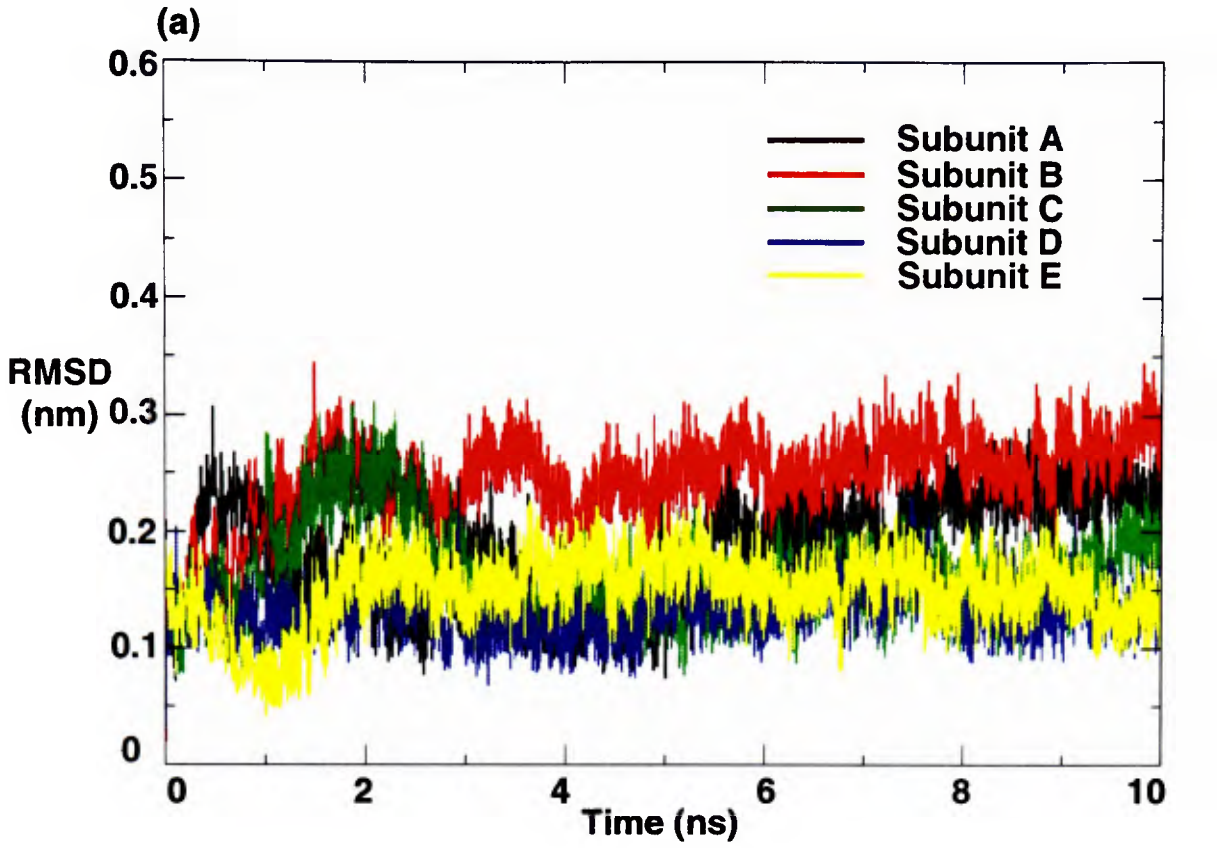


Figure 3.15 (a) Cα RMSD of loop C (residues Asn 182 – Asp 195) with respect to its principal and complementary subunits for the simulation of AChBP with carbamylcholine bound (CCE). (b) Cα RMSD of loop C for the simulation of AChBP without carbamylcholine (CCE-apo). Loop C has greater root mean square deviations in the absence of a ligand.

PCA analysis with DynaTraj [178] was also used to study the conformational dynamics of the binding pocket. In NCT, the residues in the binding pockets were more stable compared to the same residues in NCT-apo. Loop C was particularly

mobile in the absence of a ligand as observed in subunits of both the NCT-apo and CCE-apo simulations compared to simulations with ligands bound. Simulations with bound ligands exhibited smaller fluctuations suggesting that the ligand brings about lower structural mobility to the binding pocket.

To study the structural difference of atoms in the binding pocket which interact with ligands, in the presence and absence of a ligand, RMSD calculations of several atoms (Met114 N, Trp143 O, Trp143 HE1, Thr145 O, and Tyr192 HH) from all 5 subunits have been carried out (Figure 3.16 (b) and (c)). Residues involved in a great number of interactions with the ligand are indicated on Figure 3.16 (a) and Figure 3.21. These come into close proximity of the ligand and previous studies on nAChRs have indicated their importance in receptor activation [15, 33]. More recent crystallographic studies have also highlighted the role of these residues in their interactions with ligands [74, 76, 77, 185].

In the NCT and CCE simulations, RMSD values of these atoms are lower in comparison to the respective atoms in NCT-apo and CCE-apo simulations. Furthermore, there is greater variability of RMSD values for ligand-free states. This represents more structural variability for these atoms in the ligand-free simulations and illustrates their increased stability in the presence of a ligand. Thus it appears that the ligand brings structural integrity to the binding site of the protein, which is in agreement with the large-scale analysis in the earlier sections.

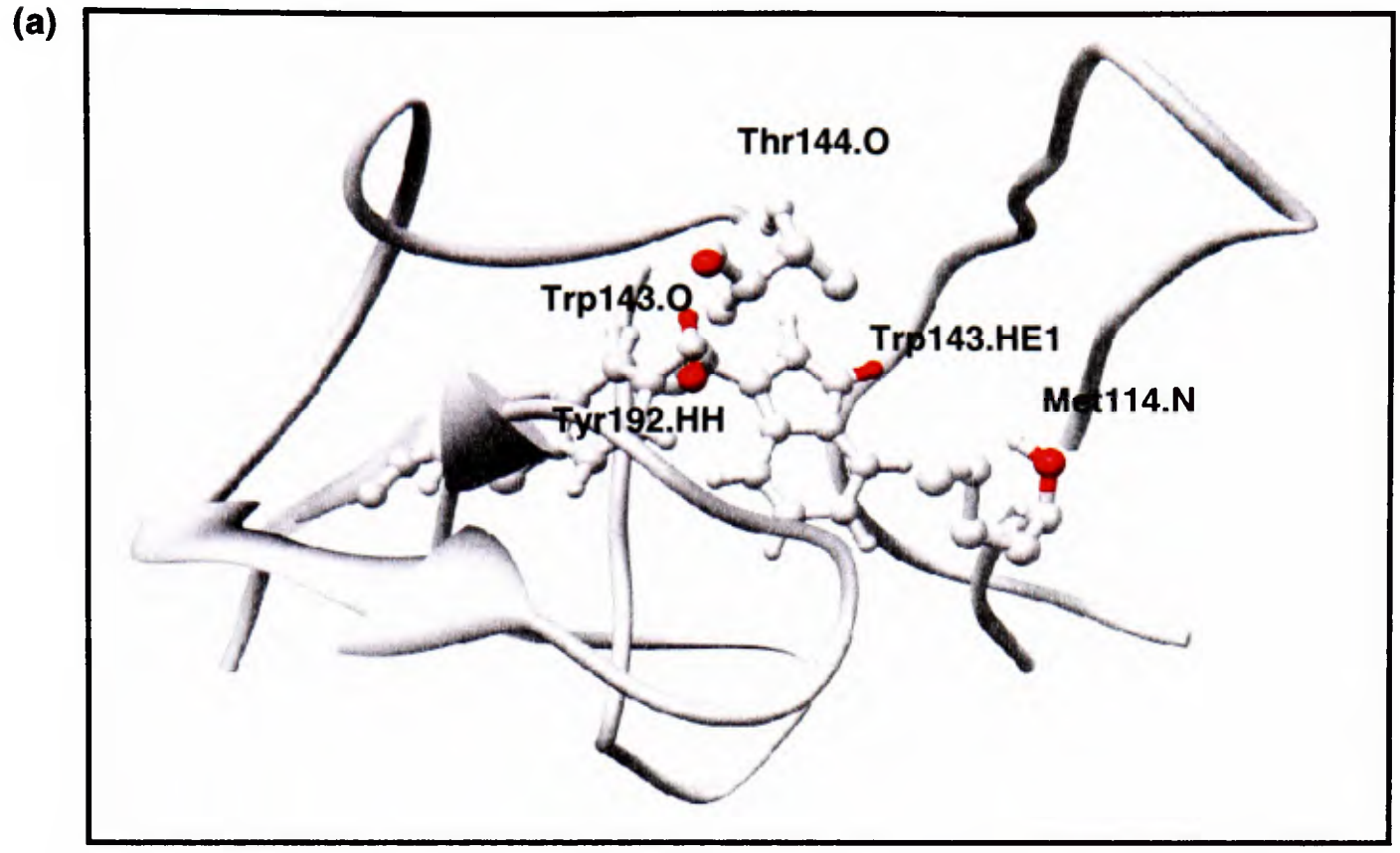


Figure 3.16 (a) Atoms interacting with the ligand in the binding pocket are highlighted and RMSD calculations are carried out on these atoms in the presence and absence of ligands. (b) RMSD of atoms shown in (a) from NCT and NCT-apo. These atoms have smaller RMSD values in the presence of a ligand. (c) RMSD of atoms shown in (a) from CCE and CCE-apo. This comparison also shows the higher stability of the atoms interacting with a ligand in the ligand-bound simulation.

Dihedral data was obtained for residues in the binding site which interact with the ligand (Met114, Leu102, Trp143, Thr144, Ser142, Gln55). Most residues exhibited lower conformational sampling in the presence of a ligand, but in the ligand-free

simulations, the range of dihedral angles for the same residues were considerably higher.

As well as calculating Cα RMSD values and dihedral angle calculations of the binding pocket residues, the distances between residues and between the ligand and neighbouring residues of the binding pocket were measured to observe how the shape of the binding pocket changes over the course of 10 ns (Figure 3.17). Cys187 (loop C), Thr144 (β8), Trp143 (β8), Met114 (loop E), Leu102 (loop E), and Gln55 (loop G) were specifically considered due to their interactions with the ligand. The ligand-free simulations (NCT-apo and CCE-apo) and subunits without a ligand in the CCE simulation (subunits A, B, and E) exhibit larger fluctuations in these key distances compared to those with nicotine or carbamylcholine bound. For example, the distances between Met114 and Thr144, considering all 5 subunits, for NCT ranges from 0.75 to 0.9 nm, but a much larger range of 0.65 to 1.12 nm exists for the NCT-apo simulation. The distance between Cys187 of loop C and Thr144 on loop B has a range of 0.98 to 1.65 nm, with very small fluctuations in distance for 4 of 5 subunits in NCT (Figure 3.17). For NCT-apo, the range is much higher at 0.63 to 1.68 nm and there are considerable fluctuations in the distance, showing that loop C is not being held in place as in the NCT simulation.

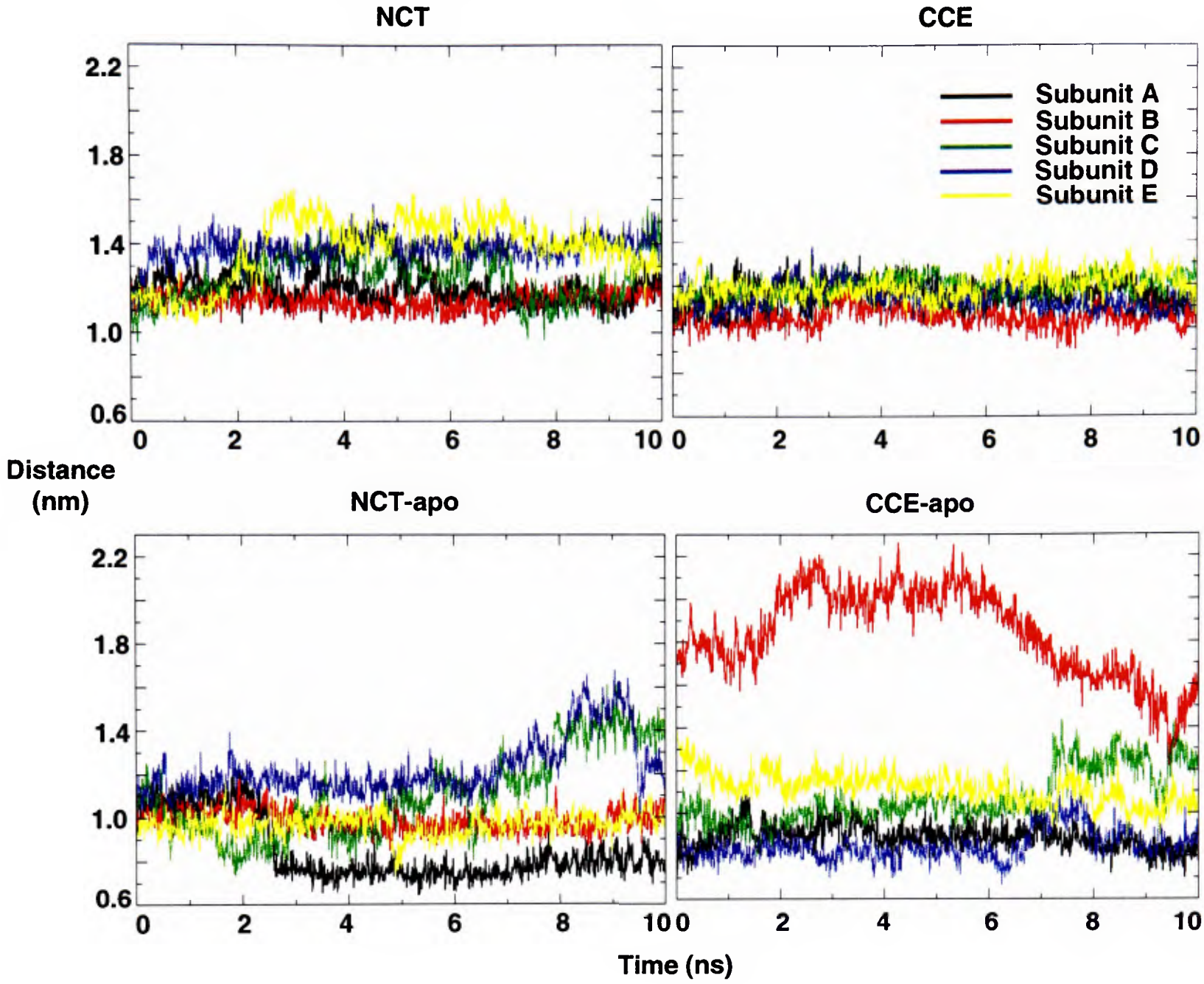


Figure 3.17 Distance between Cys187 on loop C to Thr144 on loop B of the binding pocket of AChBP for the length of the simulation for NCT, CCE, NCT-apo, and CCE-apo. Disances for the ligand-free simulations fluctuate more than those for the ligand-bound simulations.

3.3.3.2 Ligand flexibility

The flexibility of nicotine and carbamylcholine within the binding site were studied using RMSD calculations and dihedral data. The ligands for both NCT and CCE simulations, maintained their position in the binding pocket for the duration of the simulation. Although both ligands held their place and general orientation in the binding pocket, nicotine was more rigid than carbamylcholine. Analysis of the dihedral angles of the ligands shows a stable conformation for nicotine, consistent with only one binding mode. The only conformational change in nicotine is the tilting

of the 6' ring. There is very little conformational flexibility observed from the trajectory and dihedral data (Figure 3.18 (a)). Nicotine has a very high affinity for AChBP compared to carbamylcholine and this is partly due to the greater number of contacts between nicotine and residues in the binding pocket [77, 207]. Docking studies (chapter 4) also illustrate that nicotine prefers one mode of binding. This has also been suggested by experimental studies [77, 78, 130, 132, 207, 208].

The dihedral data for carbamylcholine shows more flexibility of the chain (Figure 3.18 (b)). Although, the molecule does not flip around in the binding pocket and maintains its orientation, there are rotations at both ends of the ligand. This is apparent by visual study of the trajectory and also validated by dihedral data. The main movement in this molecule is from the rotation of the quaternary nitrogen, this is also observed in solid state NMR data of acetylcholine binding [209-211].

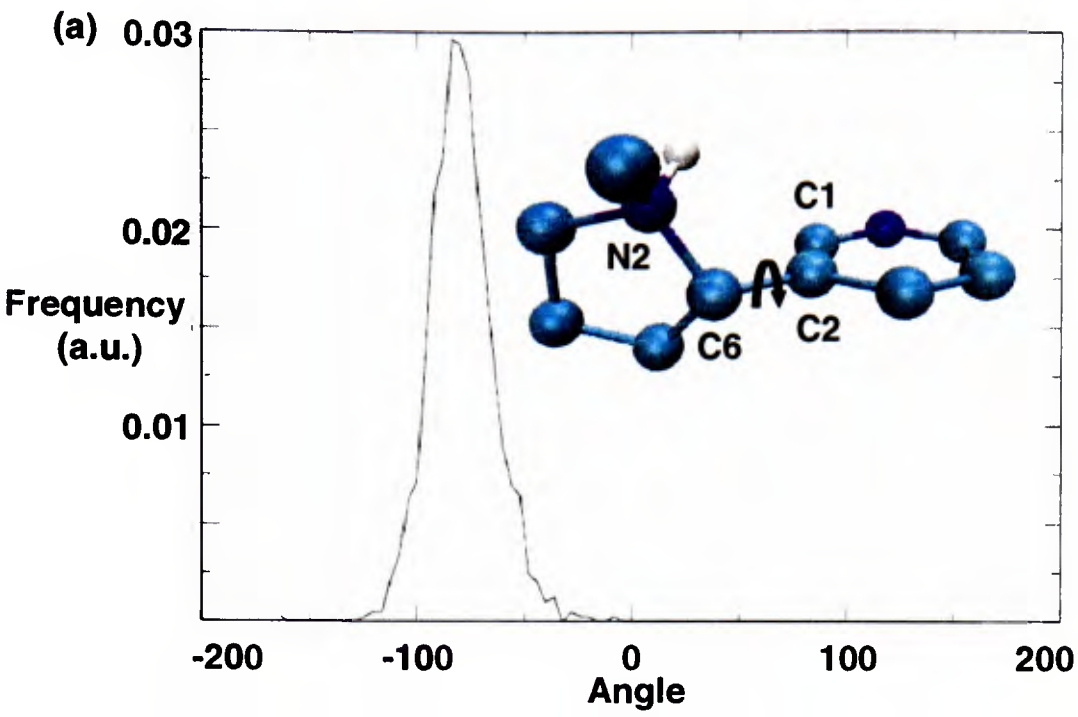


Figure 3.18 Dihedral angle data for (a) nicotine (C1,C2,C6,N2) and (b) carbamylcholine (C2,C3,N1,C8). There is greater spatial exploration of carbamylcholine's methyl groups due to the rotations of the quaternary nitrogen group. Nicotine's main conformational change is the flipping of its 6' ring.

RMSD calculations of each ligand show that carbamylcholine is very stable and the only movement causing a slight deviation from the original crystal structure is the twisting of its chain at both ends. With nicotine, the higher RMSD values are the result of the 6' ring twisting in 4 of the 5 nicotine molecules. A break-down of individual nicotine RMSDs is shown in Figure 3.20.

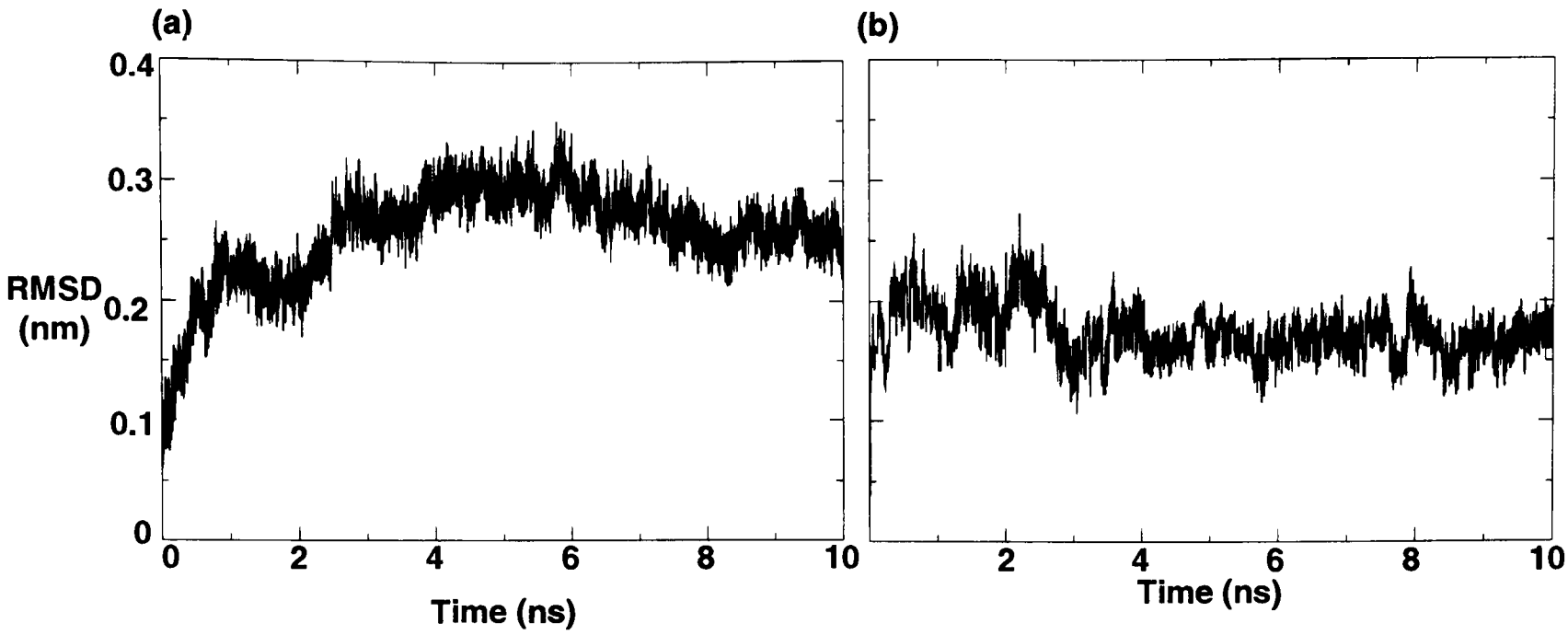


Figure 3.19 (a) The RMSD of 5 nicotine molecules from NCT, one to each binding pocket. The nicotine's are relatively rigid in the binding pocket, 4 of the 5 show a flip of the 6' ring and this is the cause of the deviation. (b) RMSD of 2 carbamylcholine molecules from CCE. There is very little deviation from the original structure.

The nicotine molecules move very little from their positions in the binding pocket.

The RMSD plots of individual nicotine molecules are different; in some subunits there are very low RMSD values because the 6' ring is not flipping due to hydrogen bonding patterns with surrounding residues (Figure 3.20). The chain flexibility of carbamylcholine and the resulting RMSD is higher for the first ~2.7 ns and then relaxes. This ligand also did not have much displacement from its original position in the X-ray structure (< 0.15 nm).

Figure 3.20 shows the RMSD of the five nicotine molecules inside the binding sites of the protein. The nicotine molecules did not show much movement throughout the simulation indicating the strong interactions which exist between the ligand and the surrounding residues in the binding pocket of the protein.

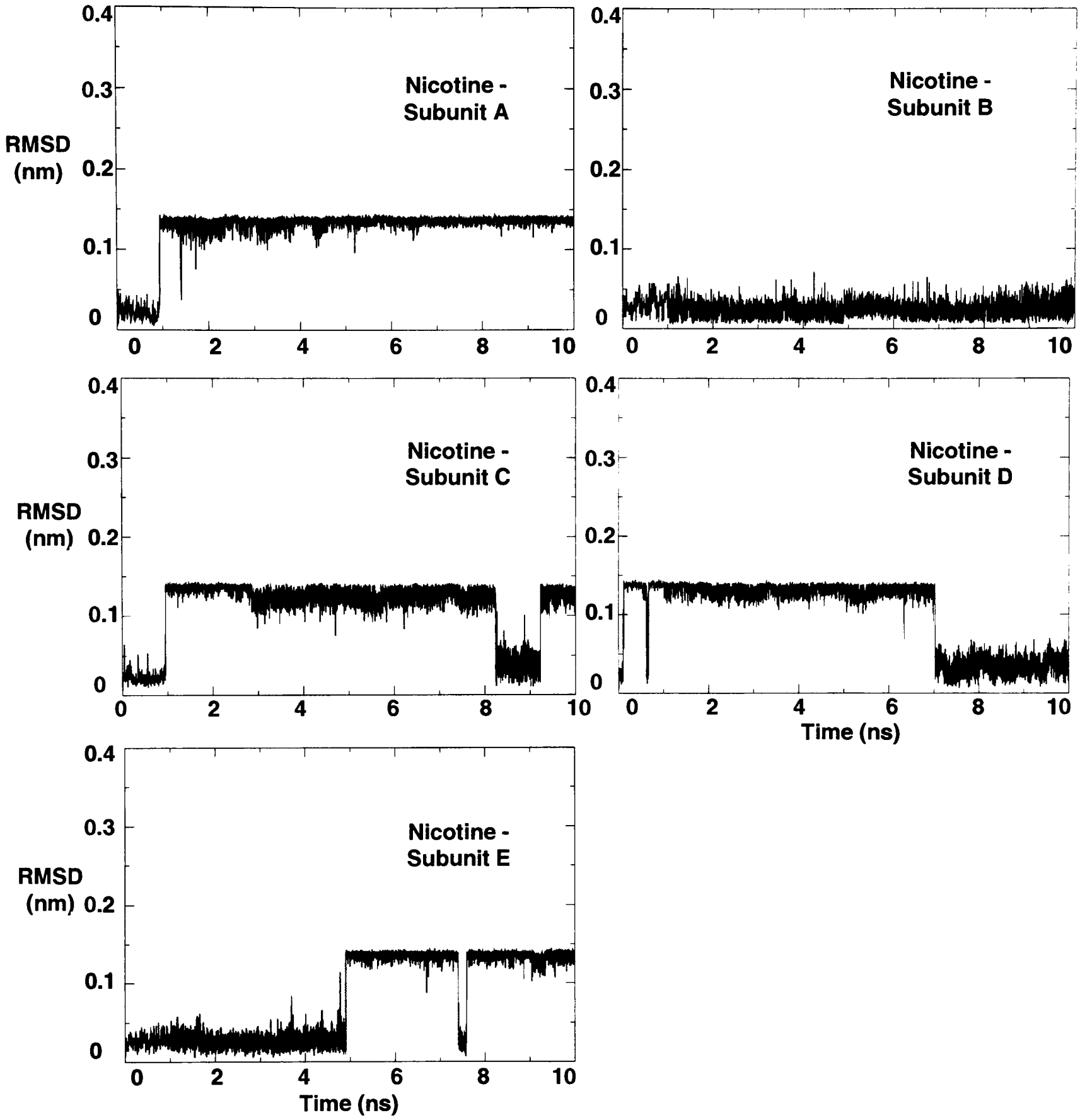


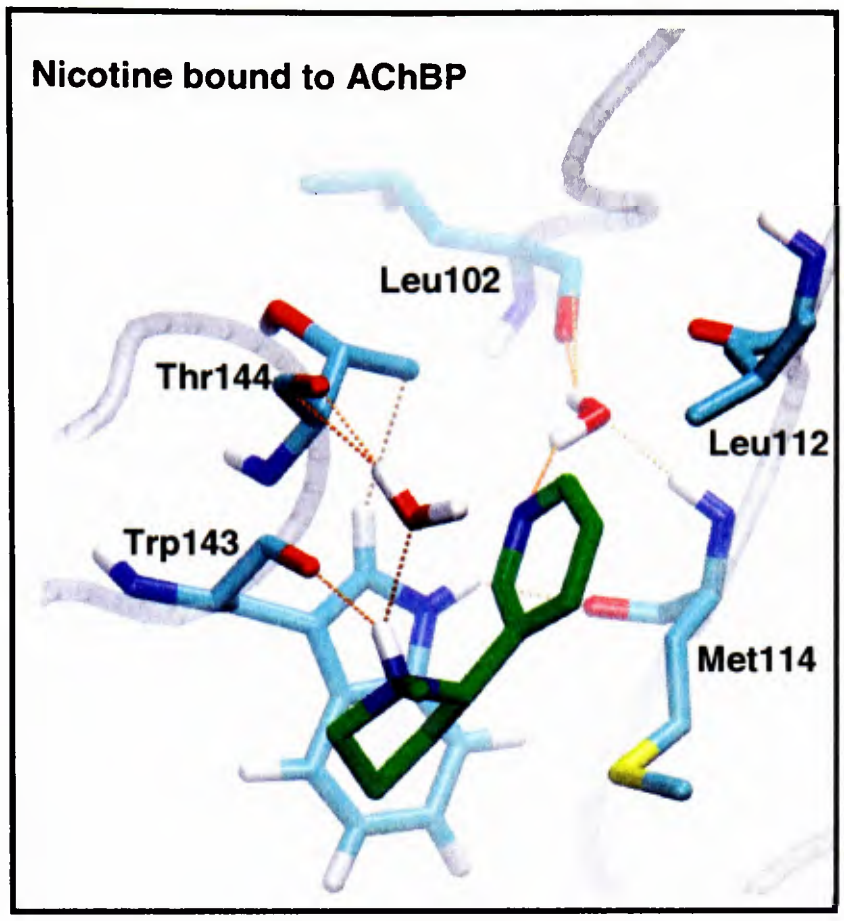
Figure 3.20 Nicotine remains in the binding pocket in the same position as in the crystal structure for the 10 ns duration of the simulation. The jump in RMSD values for 4 of the 5 nicotine molecules is the flip of the 6' ring.

3.3.3.3 Water in the binding pocket

Ligand-protein interactions which lead to structural changes in the protein that ultimately open and close the channel are not well understood. Heavy involvement of water molecules in the binding pocket and in the binding of the ligand to the protein

throughout the simulations with bound ligands was observed. Other studies have alluded to the role of water in agonist binding in other LGICs such as the glutamate receptor [83, 84, 212-215], and mass spectrometry and thermodynamics studies have revealed the role of water molecules in complexes which are formed between SH2 domains and phosphopeptides [214]. The work by Chung and colleagues suggests that water molecules appear to make a favourable contribution to the free energy of binding and that they contribute to the specificity of the interaction [216-219]. Thus, the role of water both in ligand-protein interactions and its structural role in the binding pocket were further investigated. Figure 3.21 shows interactions of nicotine and carbamylcholine with surrounding residues in the binding pocket through bridging water molecules. Figure 3.21 (a) shows a snapshot at ~5 ns of nicotine (at the interface of subunits B and C) interacting with Trp143 and Thr144 through the same bridging water, while Leu102, Met114, and Leu112 are also interacting with nicotine through a bridging water molecule. Averaged across all 5 subunits, water in this location is present for 92% of the duration of the simulation and water in the former location is present for 45% of the duration. Figure 3.21 (b) shows carbamylcholine at ~6 ns interacting with surrounding residues through 2 water molecules. One water molecule is bridging Trp143 to carbamylcholine and the other is bridging Leu102 and Met114 to carbamylcholine. Averaged across the 2 subunits with carbamylcholine bound, a water molecule bridging Trp143 is present for 79.5% of the simulation, and the other, bridging carbamylcholine to Leu102 and Met114, is sustained for 92.5% of the time.

(a)



(b)

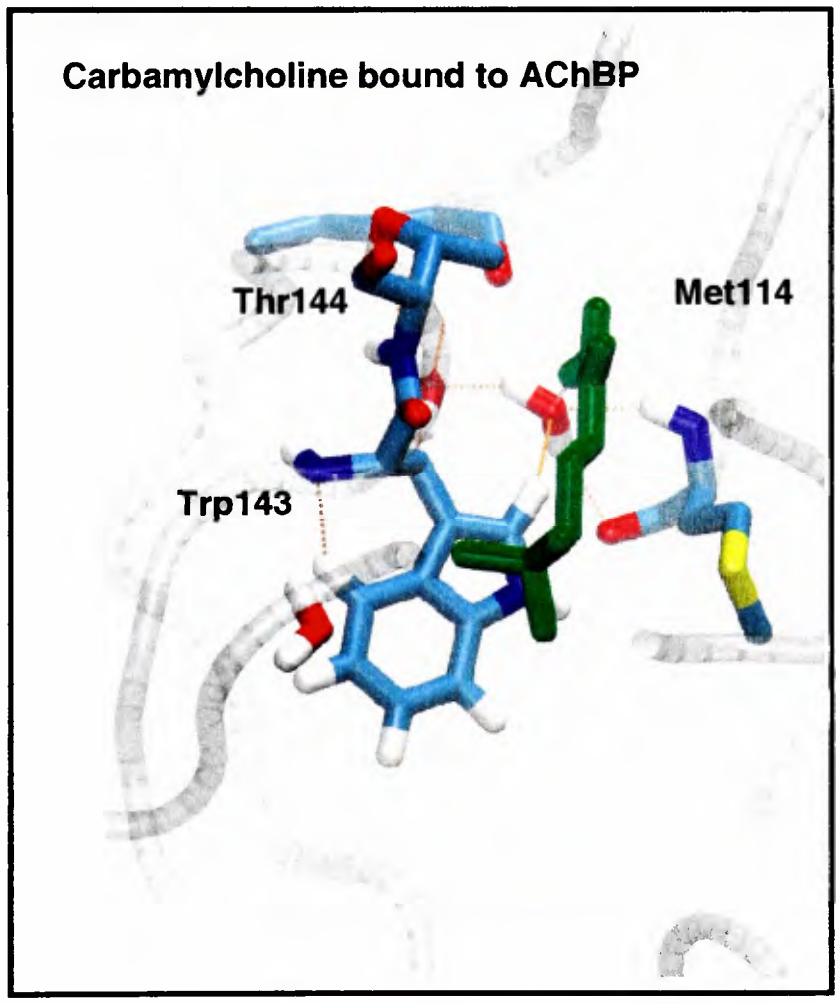


Figure 3.21 (a) Nicotine (NCT at ~ 5 ns) and (b) carbamylcholine (CCE at ~ 6 ns) interacting with Trp143 and Thr144 through a bridging water molecule and to Leu102 (NCT only) and Met114 via another water molecule.

Hydrogen bond analysis (using the Gromacs g_hbond program) between the protein and the ligands nicotine and carbamylcholine are shown (Figure 3.22 (a) and (c)).

Hydrogen bonds are also calculated between the protein, this time including the water molecules, and the ligand (Figure 3.22 (b) and (d)). For both nicotine and carbamylcholine, the number of hydrogen bonds to the ligand is higher when water molecules hydrogen bonded to the protein are accounted for.

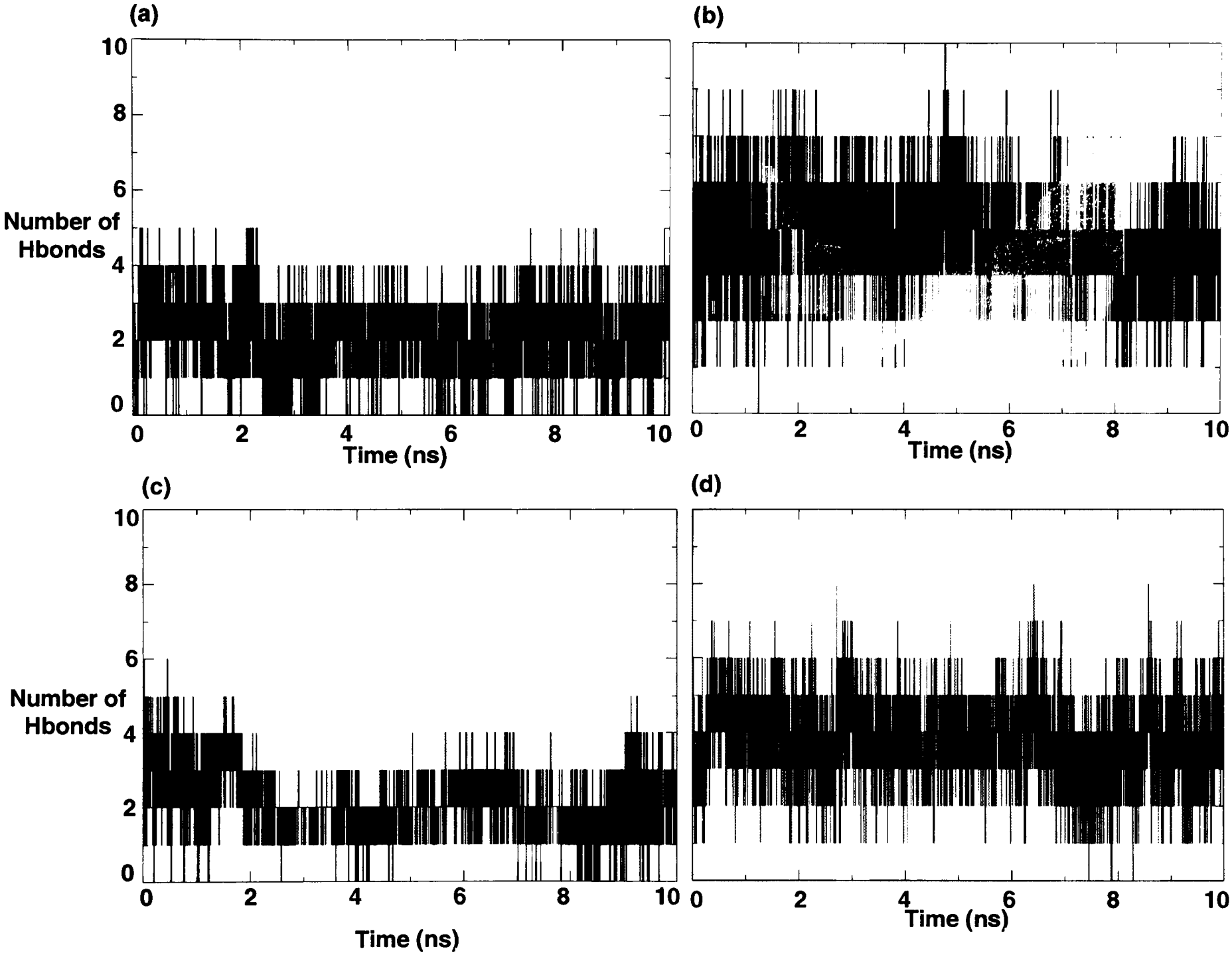


Figure 3.22 (a) NCT - and (c) CCE - The number of hydrogen bonds in each frame between the ligand and the protein. (b) NCT - and (d) CCE - The number of hydrogen bonds between the ligand and the protein but accounting for solvent insertion (bridging water molecules). When water molecules are accounted for, the number of hydrogen bonds to ligand is higher.

Ligplot was used to examine the interactions between the ligands and surrounding residues. 1 ns snapshots from each simulation were taken to observe the interactions

between ligand and protein mediated by water. For most subunits water plays a significant role in the ligand's interaction with surrounding residues in the form of a bridging molecule between the ligand and protein. Some of these bridging waters are seen throughout the simulation. Ligplot shows abundant hydrogen bonding from both nitrogen atoms of nicotine, and the oxygen atom of carbamylcholine to surrounding residues, which are present for the entire duration of the simulation. Figure 3.23 shows nicotine and carbamylcholine interacting with surrounding residues via water molecules. The number and duration of these bridging waters hints at their important role in the interaction of the protein with the ligand. In the NCT and CCE simulations, a water molecule bridging the ligand and Leu102 or Met114 on loop E is present for more than 92% of the time, averaged for the subunits with ligands bound. Often the same water molecule bridges both of these residues to the ligand. Also commonly seen are water molecules bridging the ligand with Trp143 or Thr144 or both at the same time.

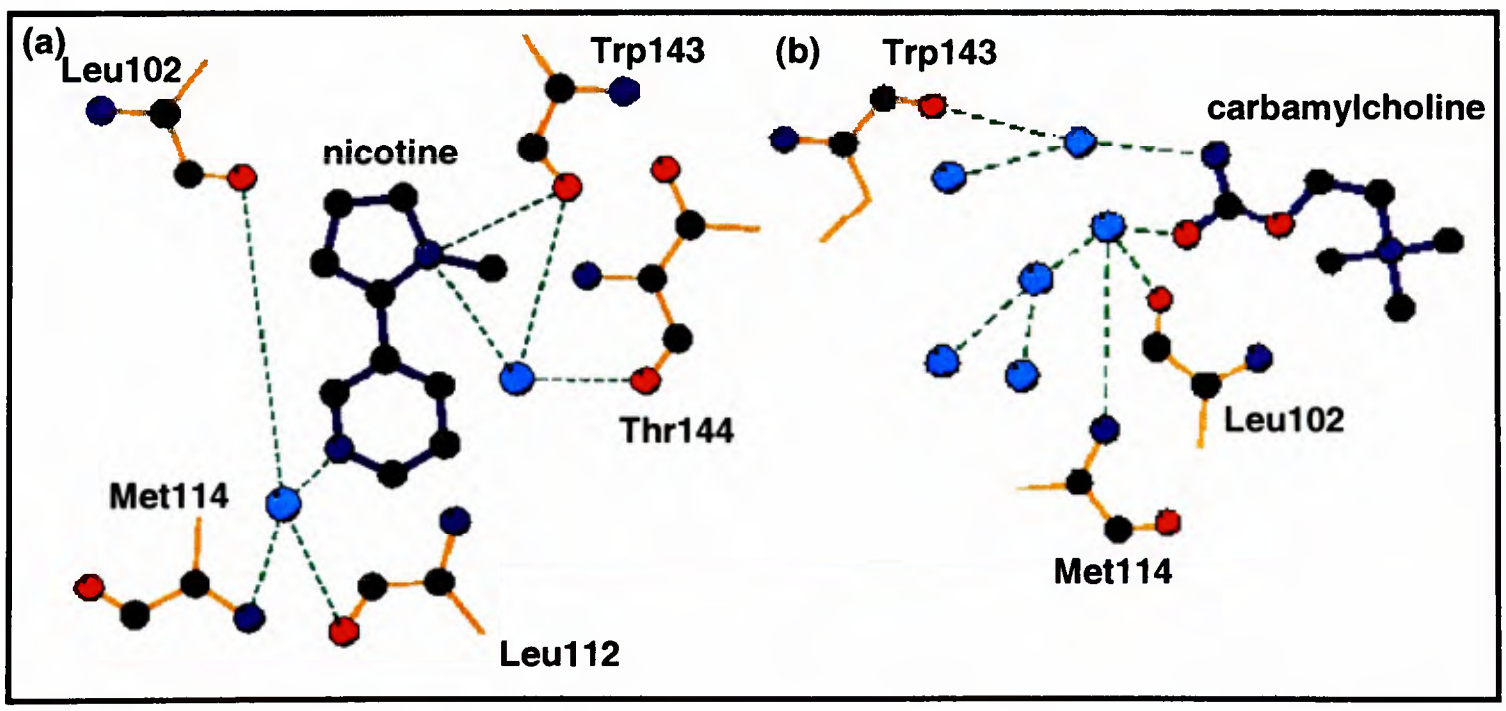


Figure 3.23 Ligplot schematics of (a) nicotine and (b) carbamylcholine bound to surrounding residues via bridging water molecules.

Bridging waters also exist between the cysteines on loop C and the ligands. As pointed out earlier, the cysteines on loop C from α-subunits of nicotinic receptors are

thought to be functionally important as they are evolutionarily highly conserved among the LGICs. Also recent MD data has shown that loop C moves in and covers the binding pocket in the presence of a ligand [130, 206] and thus the hydrogen bonding with the ligand, observed here, could play an important role in determining the behaviour of loop C. Other bridging waters commonly exist between the ligand and Gln55, and also with Tyr165. Tyr165 has its side chains facing upward in the binding site, which enables hydrogen bonding with the ligand. Table 3.3 (NCT) and Table 3.4 (CCE) list the top 10 water molecules, based on number of occurrences, bridging the ligand to a surrounding residue via a water molecule. An interesting observation which arose from this analysis, is that two of the five nicotine molecules (NCT 18814 and NCT 18817) have the greatest number of hydrogen bonds to neighbouring residues via bridging water molecules. Previous studies of nAChR have suggested that only 2 of the 5 subunits are required for the action of the nAChR [13, 58] and thus the bulk of the interactions with the ligand may also take place in two subunits which is observed here. PCA from the large-scale analysis sections of this chapter also showed asymmetry between the five subunits, with two of the five subunits having greater movement during the simulation of AChBP (Figure 3.9). The two binding sites with the mentioned nicotine molecules are on the interface of subunits B/C and E/A.

Table 3.3 Top ten water molecules bridging nicotine to a residue in the binding pocket in NCT (Chain identifier indicated in brackets)

Water molecule	Residue 1	Residue 2	Number of occurences
SOL1256	Met527 (B)	NCT 18814	1870
SOL7253	Met527 (B)	NCT 18814	1342
SOL1078	Met115 (E)	NCT 18817	1056
SOL1078*	Trp968 (E)	NCT 18817	620
SOL5732*	NCT 18813	Ser143 (A)	308
SOL13378	NCT 18817	Trp968 (E)	188
SOL1256*	Trp350 (B)	NCT 18814	147
SOL7253*	Trp350 (B)	NCT 18814	130
SOL17829	Tyr165 (E)	NCT 18817	100
SOL6834	NCT 18816	Trp762 (D)	79

Note: The water molecules with *'s indicate that the same water is now bridging the ligand to a different residue in the binding pocket (may not be displayed, only top ten shown).

Table 3.4 Top ten water molecules bridging carbamylcholine to residues in the binding pocket in CCE (Chain identifier indicated in brackets)

Water molecule	Residue 1	Residue 2	Number of Occurences
SOL9297	Met729 (C)	CCE 1026	2003
SOL14171*	CCE 1027	Trp758 (D)	1387
SOL14171	CCE 1027	Tyr807 (D)	1386
SOL1428	CCE 1026	Tyr602 (C)	580
SOL6702	CCE 1027	Trp758 (D)	505
SOL1428**	CCE 1026	Trp553 (C)	429
SOL1428*	CCE 1026	Thr554 (C)	403
SOL6702*	CCE 1027	Tyr807 (D)	377
SOL11808	Met934 (D)	CCE 1027	354
SOL16192*	CCE 1027	Try807(D)	315

Note: The water molecules with *'s indicate that the same water is now bridging the ligand to a different residue in the binding pocket (may not be displayed, only top ten shown).

Now that water molecules bridging the ligand to the protein along the trajectory have been identified, a more detailed look at these water molecules is necessary to determine the percentage of time they occupy these locations in the binding pocket. Analysis of 'persistent' waters (those present for $\geq$ 40% of the simulation within a cut-off of 0.1 nm) in the binding pocket revealed five distinct zones where a water molecule was preferentially found over the course of 10 ns. The persistent water molecules were identified by calculating the standard deviation of their x, y, and z

coordinates for a given window of the simulation using an in-house program. There were other areas where waters were commonly found in the binding pocket, but here the 5 most commonly occupied zones and those occupied for the longest stretches of time are described (see Table 3.5) (see supplementary CD). For the CCE simulation with only 2 ligands bound in adjacent subunits, the greatest number of static waters were found in the ligand-bound subunits throughout the simulation. For the binding pockets that do not have a bound ligand, water molecules were observed in these zones, but were more mobile and stayed for a shorter amount of time than waters in the binding sites with ligands bound.

Table 3.5 Average percentage of time each zone is occupied by at least one water molecule.

ZONE	Average for NCT (5 subunits)	Average for CCE (2 subunits)
1	92	92.5
2	45	79.5
3	40	89.5
4	60	76
5	55	50

Figure 3.24 (a) is a snapshot of the final frame of NCT showing the five zones that represent the location of persistent waters in the binding pocket. Zone 1 lies within loop E between Leu102 and Met114. This location is occupied with at least one water molecule > 90% of the time in both the NCT and CCE simulations. Notably, waters in this position have also been observed in crystallographic studies [77]. Water molecules can be seen 'competing' with each other for this location, usually if one water molecule leaves the site, another will quickly replace it (Figure 3.26), suggesting a strong functional role for water in that location of the binding pocket. Zone 2 is the area near the Trp143/Thr144 of loop B (β8 sheet) where there is a

bridging water connecting the ligand with either one or both residues in both the NCT and CCE simulations. Zone 3 describes the low β5 sheet (loop A) of the binding pocket with residues Tyr90 and Ile93 that are believed to be involved in ligand binding from previous experimental studies [76]. In several nanoseconds of the ligand-bound simulations, this location is occupied with one or more water molecules, however this is seldom the case in the ligand-free simulations. Zone 4 is at the base of loop C near Glu194 and Tyr192. Tyr192 is heavily involved in hydrogen-bonding to the ligand. Waters in this zone may play a role in keeping the integrity of loop C and obstructing its movement in the presence of a ligand. Zone 5 is at the tip of loop C where there are two adjacent cysteines forming a disulfide bridge. These cysteines have been the focus of many studies and are highly conserved in the α-subtype of nAChRs, which make up the 'principal' subunit of every binding site in the nAChR [65, 220, 221]. It is believed that the cysteines on loop C interact with the ligand in the binding pocket of nAChR and loop C has recently been reported to 'close off' the binding pocket in the presence of a ligand [130, 131]. As stated above, RMSD plots show lower structural deviations for loop C in the presence of a ligand. There are several bridging waters in position between the cysteines on loop C and the ligands; this key interaction keeps loop C closed in and near the ligand when one is bound. Figure 3.24 (b) shows an average water density plot for one of the binding pockets from the NCT simulation. There is higher density of water in the defined zones of the binding pocket in this simulation compared to other locations of the protein and to the corresponding ligand-free simulation. In comparison to bulk water, the density of water molecules in the binding pocket is much greater, and the densities are in discreet zones, which indicate that the position of these waters is highly conserved. The highest densities of water molecules are in the zones identified in this work.

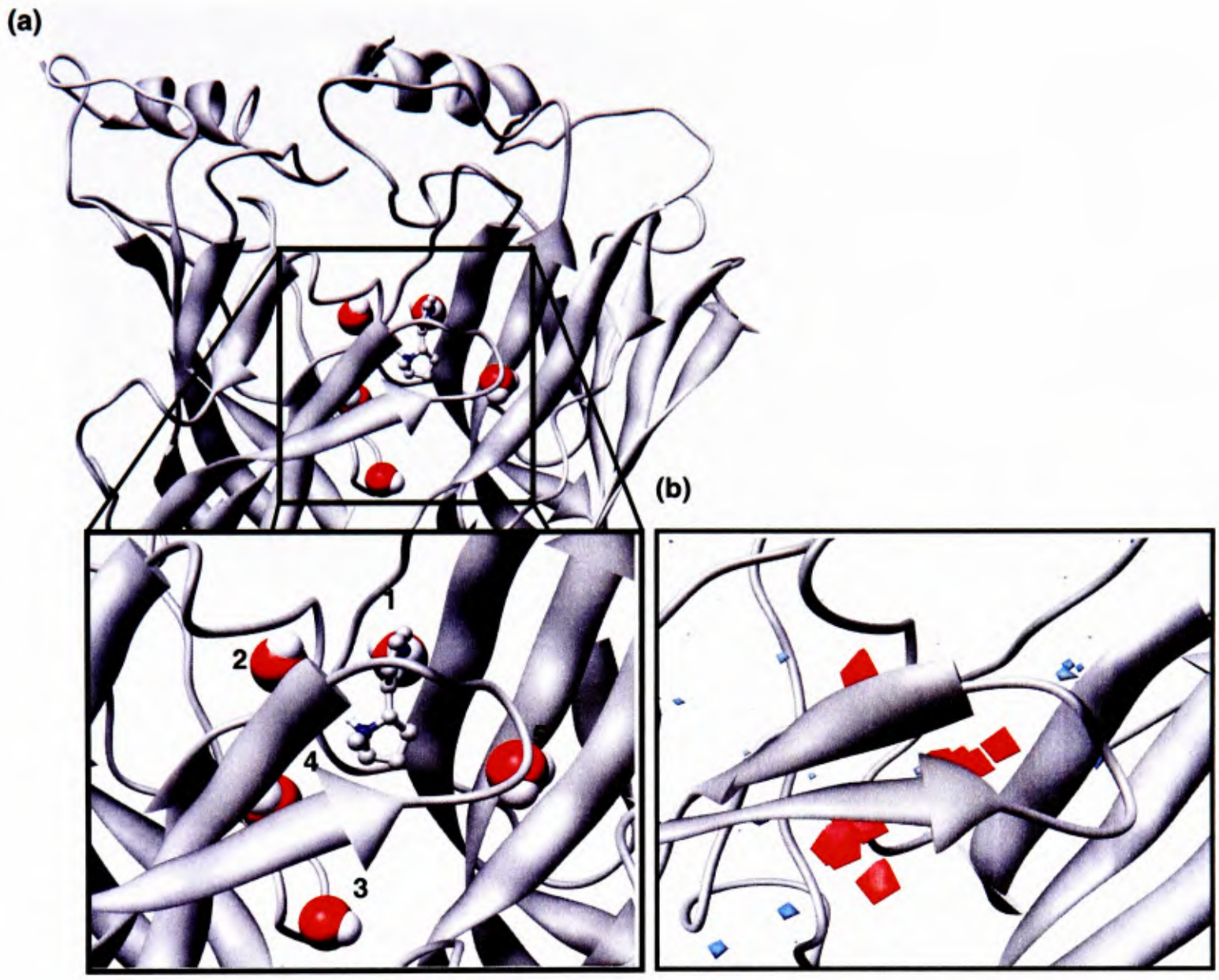


Figure 3.24 (a) A final frame snapshot of NCT and a close-up figure showing water molecules in the 5 zones identified as being occupied with water molecules for ≥ 40% of the time of the simulation. (b) An average water density plot of one of the subunits from NCT, showing the highest density of water molecules in the zones identified in the binding pocket.

Averaged water density calculations for CCE and CCE-apo are reported in Figure 3.25. In the ligand-bound simulation, the distribution of water pockets in the binding site is symmetrical and more pronounced. Interestingly, there is a high density of water molecules in every binding pocket, even though only two binding sites have a bound ligand. This agrees with our previous results suggesting that the presence of only 2 ligands is enough to create structural changes in the whole protein structure, with implications for gating.

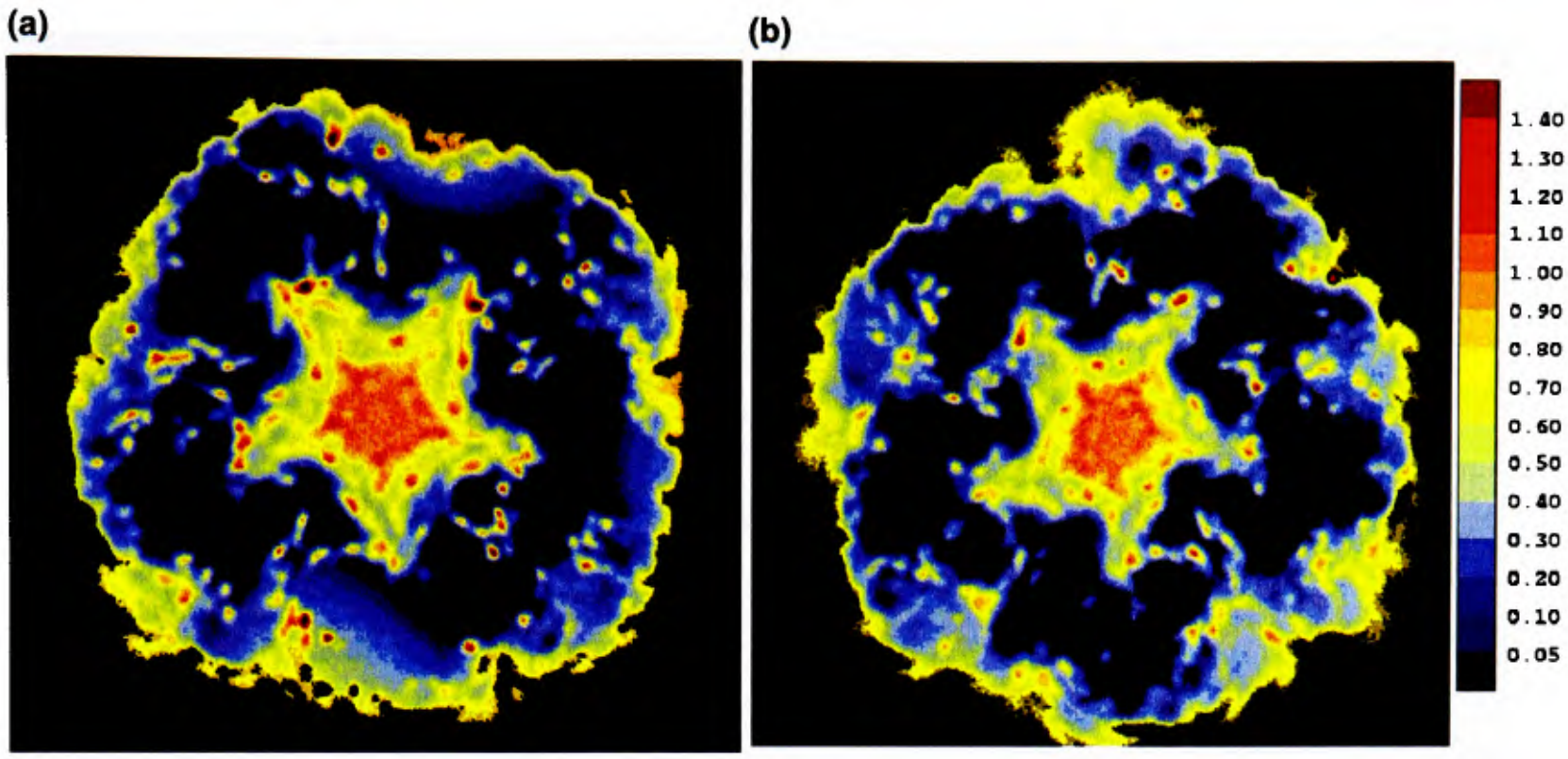


Figure 3.25 Average water density plots for CCE (a) and CCE-apo (b). Higher density of water molecules is found in the ligand bound (CCE) simulation. The clusters of high water density are symmetrical and higher in number than the respective ligand-free simulation. The scale represents the relative water density relative to bulk water density i.e. 1.0 is bulk water. Only water density near the protein is shown.

Table 3.6 shows a breakdown of the water molecules resident in the identified zones for each subunit of NCT and CCE. For CCE, the majority of water residence is in subunits C and D, where a ligand was present, further confirming the presence of persistent water molecules in the presence of a ligand. These figures are the output from an in-house program, WatRes.exe, which identifies waters 'resident' for a specified duration.

Table 3.6 The percentage of time water molecules are resident in the identified zones in (a) NCT and (b) CCE

(a)

Zone	Subunit	Water	% of time
1	A	1164	35
	A	1120	100
	B	1256	60
	C	1327	100
	C	1285	100
	D	1397	100
	D	1399	100
	E	1078	100
	E	1421	20
2	A	1167	25
	A	4290	20
	B	2598	100
	D	1029	100
3	B	7382	100
	B	6730	74
	D	7254	100
4	A	1036	44
	A	1042	98
	C	1031	100
	C	1032	100
	D	1074/1091	100
	D	1077	23
5	A	1044	75
	B	1055	100
	D	1050	100
	D	1076	5

(b)

Zone	Subunit	Water	% of time
1	A	1041	100
	B	1067/1069	100
	B	1034	87
	C	1062	85
	C	1065	85
	D	1075/1083	94
	D	1059	100
	D	1076	72
	E	1048	72
	E	1049	73
2	C	1061	82
	D	1079	73
	D	1086	64
	D	1058	77
	E	1080/1047	53
3	C	1064	100
	D	1071	79
	E	1087	77
4	A	1036	45
	A	1042	97
	C	1031	100
	C	1032	100
	D	1078	50
	D	1074	52
	E	1054	78
5	B	1055	100
	D	1076	15
	D	1050	100

The consequence on the structure if a water molecule in one of the identified zones moves out was examined. Figure 3.26 shows the disruption of interactions of the ligand with surrounding residues as a water molecule moves out of zone 1. Here three water molecules in zone 1 of one of the subunits are followed for the duration of the simulation. The distance between the O atom of each water molecule to the N atom of Leu102 is measured. Two water molecules are positioned in zone 1 between

residues Leu102 and Met114 on loop E (zone 1). A third water molecule enters and displaces one of the two waters and replaces it in the exact position. This water then remains in that position for most of the simulation. Upon leaving (just before 9 ns), the β-sheets of loop E collapse closer together and the nature of interactions with the ligand is altered. The hydrogen bonds from the bridging water molecule to the ligand are no longer present and the ligand makes fewer contacts with its neighbouring residues. The network of hydrogen bonds from the bridging waters plays an important role in the binding of the ligand to the protein, and it also participates in structural changes that may be involved in the function of the protein. This stresses the important function of water molecules in the binding pocket both structurally and in ligand-protein interactions.

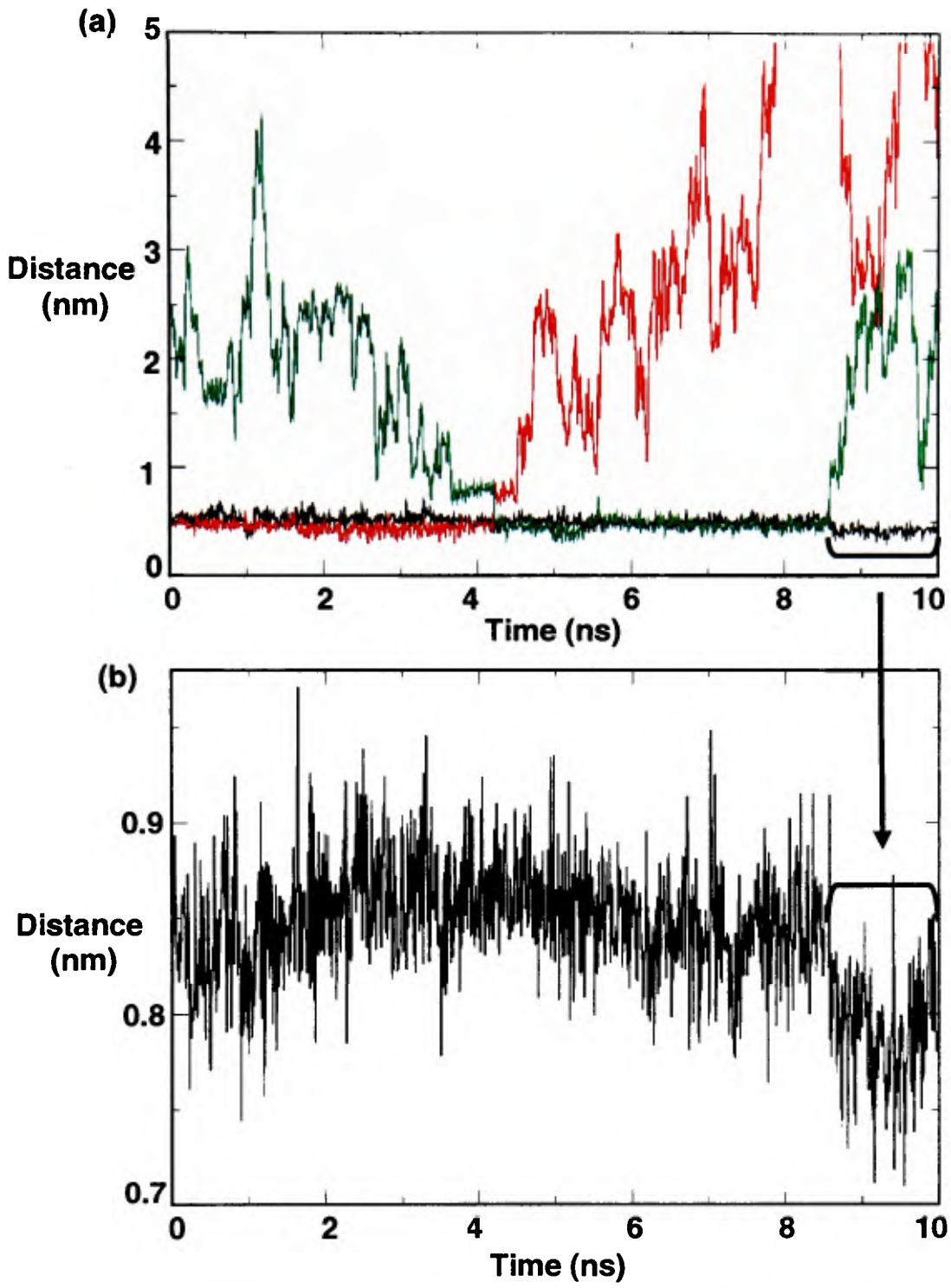


Figure 3.26 Following the trajectory of three water molecules in Zone 1 of NCT by calculating their distance from the N atom of Leu102 to the O atom of each water molecule. (a) A water molecule (green) comes in and knocks out one of the two water molecules in zone 1, replacing its position. This water remains until ~8.6 ns. (b) When it leaves the distance between the loops collapses together.

Figure 3.27 illustrates the break-up of interactions upon the exit of water molecules. From NCT, a ligplot snapshot shows interactions between nicotine and surrounding residues via bridging water molecules at ~4 ns and then again at ~9 ns. Once these water molecules are gone, the interactions with the ligand are significantly reduced.

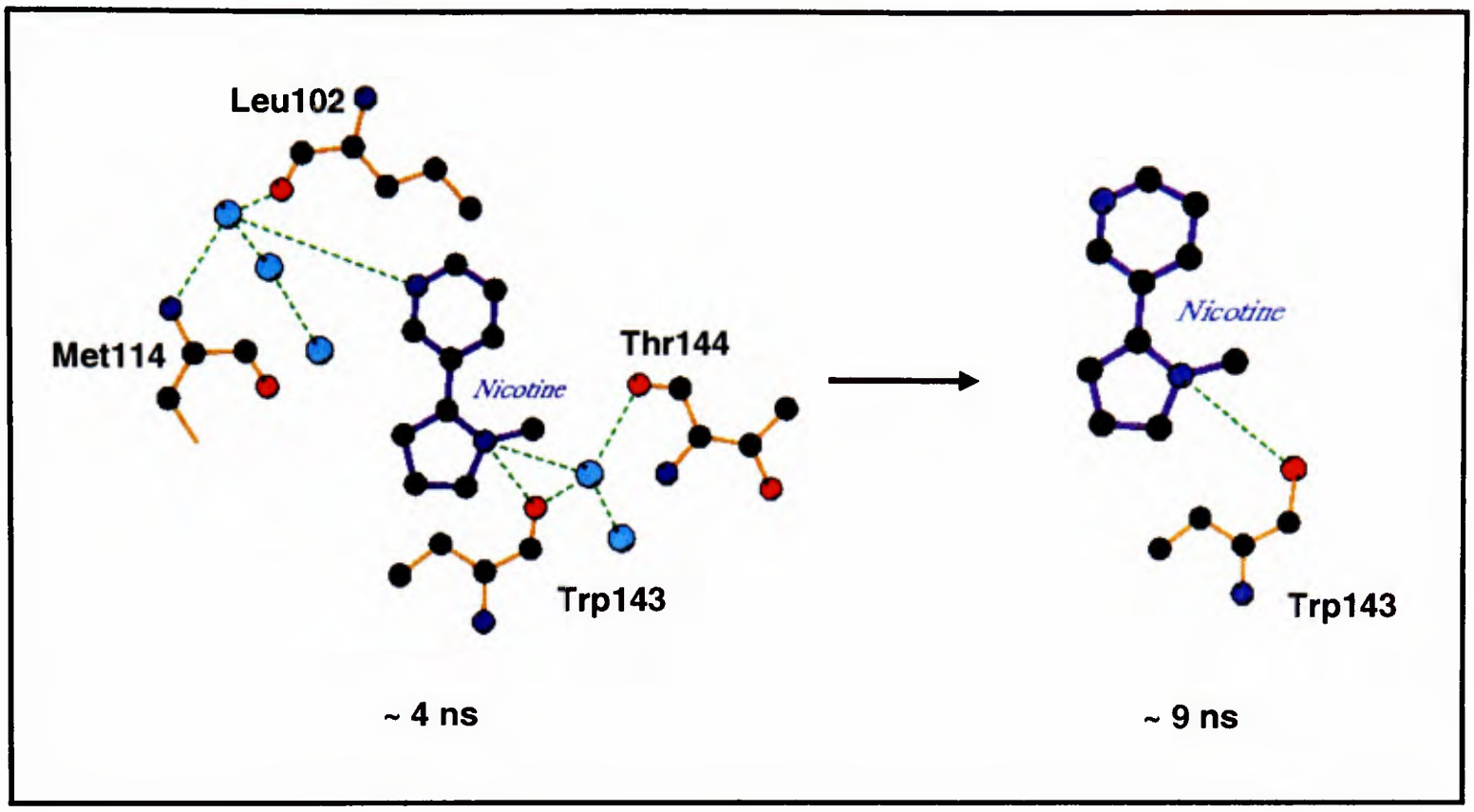


Figure 3.27 (a) Ligplot diagrams of nicotine interacting with residues in the binding pocket via bridging waters in zones 1 and 2 at ~ 4ns. (b) The interactions with the ligand are reduced when the water molecules leave the zones at 9 ns.

It appears that the network of hydrogen bonds has a two-fold function: one, water molecules keep the structural integrity of the binding pocket, and second, it increases the number of interactions with the ligand while in the binding site.

3.4 Discussion

3.4.1 Overall conformational dynamics

Six molecular dynamics simulations were carried out; three with ligands (NCT, CCE, and EPE), and the other three were their ligand-free states (NCT-apo, CCE-apo, and EPE-apo). In general the ligand-bound states were more conformationally stable with lower MSF values. Average displacement figures show higher overall displacement in ligand-free simulations, whereas the ligand-bound simulations have higher displacements in their termini. Interestingly, overall RMSD calculations between

NCT and NCT-apo do not show a large difference between the two states, although RMSD values of functionally relevant segments, such as the binding pocket, are lower in the ligand-bound state. This could be attributed to the starting crystal structure having five bound ligands versus only two for CCE, and the longer time needed to achieve an 'apo' state.

GNM shows changes between the first and final frames of the ligand-bound simulations, in particular in functionally important loops, such as loop C, where higher flexibility of the loop is observed at the final state. With computational tools such as PCA and the GROMACS suite of programs, both global conformational changes and more local conformational changes in the binding site can be studied. PCA data shows that there is asymmetry in the motion of individual subunits, even in NCT a homo-pentamer with a ligand bound to each of the five binding sites. This asymmetry has been observed in other MD work [129, 130].

3.4.2 Binding pocket flexibility

One of the main objectives of this work was to take a more detailed look at the binding pocket. The structural changes that take place in the presence and absence of ligands and the role of water molecules in the binding site both as structural components and as a bridge between the ligand and the protein are studied.

Binding pocket flexibility was studied using distance calculations between residues in the binding pocket, dihedral angle behaviour calculations of binding pocket residues, and RMSD calculations of atoms and protein segments involved in ligand binding. The distance calculations between residues in the binding pocket show greater fluctuating distances for the ligand-free (NCT-apo and CCE-apo) simulations. Distance calculations between the ligand and surrounding residues (NCT and CCE

simulations only) show stable distances throughout the simulation, illustrating the limited movement of the binding pocket residues. Our distance calculations show that there is less flexibility and movement of loop C in the presence of a ligand. The distances between residues on loop C and residues inside the binding pocket are highly variable in the ligand-free state compared to more stable distances in the ligand-bound simulations. Earlier studies also show that in the event of ligand binding, loop C comes in and 'caps' the binding site and exhibits less mobility [130, 131].

Dihedral angle data of side chains of residues in the binding pocket (Met114, Ser102, Leu102) shows less sampling of conformational space when a ligand is bound to the binding site, the side chains show more movement in the ligand-free simulations.

RMSDs of atoms interacting with the ligands are considerably lower and less variable in the presence of a ligand. These atoms are more mobile in the absence of a ligand, stressing the structural disorder, thus it appears that the ligand brings structural integrity to the binding pocket.

3.4.3 Ligand flexibility

The behaviour of nicotine and carbamylcholine in the binding pocket was studied to better understand how the conformation of the ligand and ligand-protein interactions, may lead to structural changes in the ion channel. Both ligands did not move away from the binding pocket for the duration of the simulation. Carbamylcholine has more flexibility, specifically at the ends of its choline chain, this also agrees with previous experimental and computational studies [77, 78, 130, 132, 207-211]. Nicotine was much less flexible overall, exhibiting only one mode of binding, with a slight tilt of the 6' ring which agrees with our docking studies and published

experimental data [77, 207]. Nicotine interacts with surrounding residues through both N atoms in most frames of the NCT simulation, these hydrogen bonds on the two separate rings of nicotine, keep the molecule rigid when bound to the channel. CCE interacts through its O and N atoms which are at the same end of the molecule, leaving the other end more mobile.

3.4.4 Water in the binding pocket

Nicotine and carbamylcholine interact with surrounding residues via bridging waters and this commonly results in a network of hydrogen bonds that extend beyond the initial ligand-residue bond and adds rigidity to the residues in the binding site. Several zones of 'persistent' waters have been identified in the binding pocket of the AChBP. During the ligand-bound simulations (NCT and CCE), these zones have a high average density of water molecules and waters in these zones commonly act as a bridge between the ligand and the protein. A single water molecule can act as a bridge to several residues at one time. Persistent waters are not as common in NCT-apo and CCE-apo and they are not maintained for the same duration as in the ligand-bound simulations. Certain zones are occupied for greater than 92% of the time on average in the NCT and CCE simulations. In ligand-bound simulations, if a water molecule leaves one of the defined zones it is quickly replaced by another, and at times water molecules 'compete' for particular zones in the binding pocket. This phenomenon seldom happens in the NCT-apo and CCE-apo simulations, where water lost from one of the zones is not replaced and water densities in the described zones are lower and display asymmetrical distribution.

The hydrogen bonding network in the binding pocket is not only important in ligand-protein interactions, but may also play a structural role. Loop C residues such as

Cys187 are commonly hydrogen-bonded to the ligand through water molecules (zones 4 and 5). Therefore the position and flexibility of loop C is affected by the hydrogen bonding through the water molecules and is an example of how the structure of the binding site is contributed to by these waters. Loops A and B are also held more rigidly in the presence of water, which are more abundant and persistent in the NCT and CCE simulations. These bonding patterns have structural implications for the binding site which are likely to be important in the homologous nAChR proteins. The high density of water in zone 1, and the resulting dense network of hydrogen bonds, keeps the loops of the 'complementary' subunit, such as loop E, in a structurally stable state. Zone 1, where water molecules commonly 'compete' for a position, is also a short distance away from the highly conserved and functionally important Cys loop which comes into close proximity of the TM domain of LGICs.

The observations made regarding persistent waters in the binding pocket, such as their residence in certain zones for long stretches of the simulation, their maintenance of the binding pocket structure through hydrogen bonding networks, and their role in bridging loops in the binding site (i.e. loop C) and the ligand, illustrates their crucial role in the structure and function of the protein.

Due to the short length of the simulations (10 ns) the longer term behaviour of the binding pocket of the AChBP and the possible mechanism by which structural changes are transmitted to the TM domain of the nAChR during gating (millisecond time-scale) can only be inferred. However, with molecular dynamics data the structural differences of binding pockets with and without ligands, and the dual role of water molecules, both at keeping a rigid binding site and at acting as a bridge from the ligand to residues in the binding site, have been illustrated. Further studies can be

carried out to see how docking of ligands is affected in a ligand-free (less rigid) binding pocket and in the presence and absence of bridging waters molecule.

The work in this chapter shows the structural instability that results with the removal of a ligand in ligand-free MD simulations. The ligand maintains a more stable and symmetrical global conformation of the protein. Small scale analysis conducted on the binding pocket show that the ligand maintains its position and conformation in the binding site and water molecules play a key role in ligand-protein interactions; the network of hydrogen bonding, amplified in the presence of a ligand may also have structural implications.

4 Docking studies of AChBP: an MD ensemble approach

4.1 Introduction

The AChBP simulations discussed in the previous chapter were used to carry out docking studies with several ligands of pharmacological interest. Nicotine, acetylcholine, carbamylcholine were docked onto the AChBP crystal structures and followed by docking onto successive frames of MD simulation trajectories of AChBP. The aim was to better understand the ligand-protein interactions at a structural level, to see if docking of ligands onto MD trajectories is improved or hindered as the protein goes through structural changes, and to observe the binding modes for each of the ligands. It is of interest to see how the starting crystal structure and the subsequent simulation can influence the docking of ligands and if a particular ligand-bound simulation is more accommodating to that ligand when docking back onto the structure. Also, the study of whether and how docking improves or deteriorates along a trajectory is important in understanding the ligand-protein interactions and the

structural state for optimal binding. A few other studies have used the combined methodology of molecular dynamics and docking, one which discusses the inclusion of protein flexibility via MD to improve ligand docking [222]. The others are studies of dihydrofolate reductase which use MD on different docked conformations in order to identify plausible modes of binding [223, 224].

4.2 Methods

4.2.1 Docking

AChBP was used as a docking target for nicotine, acetylcholine and carbamylcholine (Figure 4.1). All of the docking was carried out with the program Autodock [179] (see Methods chapter for Autodock parameters used). Ligand atom types had their charge calculated as follows. The charges on ACh were assigned according to Segall *et al.* [225]. Charges for the other ligands were calculated with the 6-31G* basis set using Spartan [180] (see complementary CD for charges). All docking results were visualized with VMD [200] and UCSF Chimera [201]. A script was written to carry out docking onto multiple frames of the AChBP simulations (see section 4.2.3).

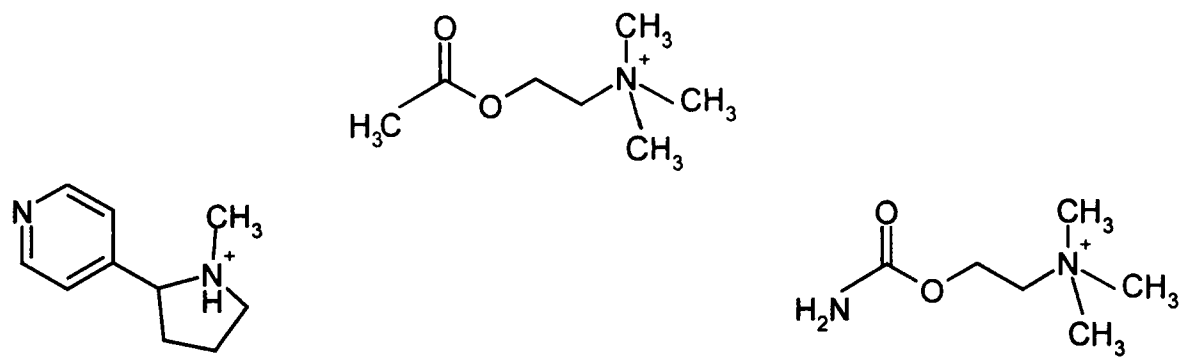


Figure 4.1 The three ligands used for the docking studies in this chapter.

4.2.2 Molecular dynamics (MD) simulations

Molecular dynamics simulations were carried out as described in chapter 2. For docking studies onto the AChBP trajectories (see section 4.2.3), the simulations from chapter 3 were used (see Table 3.1). The simulation names in this table will be used for the remainder of the chapter.

4.2.3 Automated docking along an MD trajectory

In order to dock ligands onto the AChBP MD ensemble of structures, successive frames of each trajectory were used to dock ligands back into the structure. 100 frames were used from each trajectory, 1 every 100 ps for a 10 ns simulation (see Figure 4.2). This should provide enough sampling as major conformational changes of side-chains generally do not occur faster than 100 ps. Taking 100 frames instead of 10000 also saves on computational resources and time. Each system (100 frames) takes approximately 24 h on a single processor desktop computer. The docking script uses Autodock to dock the ligand onto each of the 100 frames of the trajectory. 50 genetic algorithm (GA) docks were performed for each frame and were ranked according to their energies. Ligands with high affinity for the binding site are expected to have one or a few favourable modes of binding. Thus the 50 docks should fall into one or a small number of docking clusters, all clusters are ranked by their energies (lowest to highest). The analysis script records the top dock (lowest ranked energy) and also the number of docks out of 50 in the top ranked cluster for each saved frame of the simulation. A flow-chart describing this process is shown in Figure 4.2.

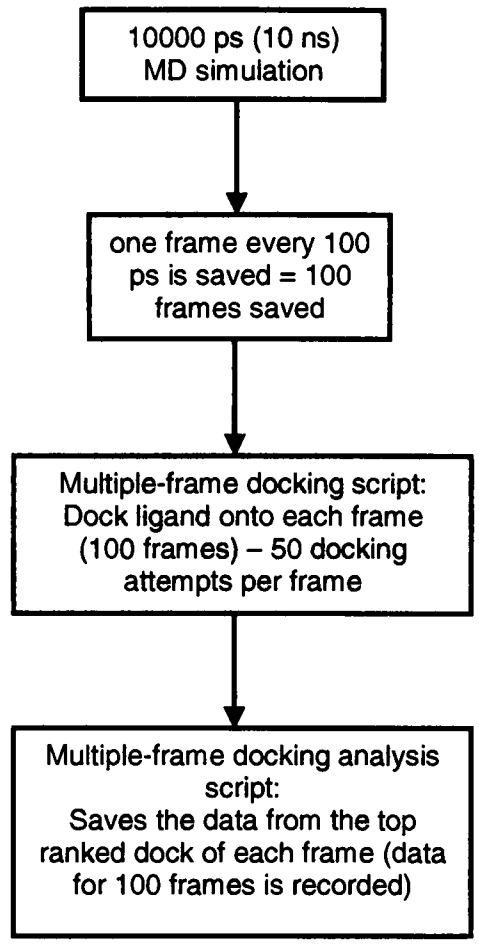


Figure 4.2 A flow-chart showing the docking process onto 100 frames of each simulation starting with a 10 ns simulation. One frame is saved every 100 ps. A docking script is then used to dock the ligand onto each saved frame. An analysis script records the top ranked dock for each frame.

4.3 Results

4.3.1 Docking of nicotine, carbamylcholine, and acetylcholine onto AChBP

The crystal structures of AChBP [77] with nicotine bound (PDB code 1UW6) and carbamylcholine bound (PDB code 1UV6) were used as controls; nicotine, carbamylcholine and acetylcholine were docked back onto the structures. This provides assurance in the docking methodology for docking other ligands onto the same structures and also for docking onto the MD trajectories of these structures. Nicotine (Figure 4.3 (a) and (b)), carbamylcholine (Figure 4.3 (c) and (d)), and acetylcholine docked onto the structures in the same modes proposed by the crystal structures (assuming acetylcholine docks in the same mode as carbamylcholine [77]). Carbamylcholine and acetylcholine needed finer grained parameters (smaller grid

spacing), to dock in the correct mode. In both structures, nicotine bound most favourably, with the lowest docking energies of the three ligands, followed by acetylcholine and finally carbamylcholine. This matches the binding affinities of the ligands for the protein [77].

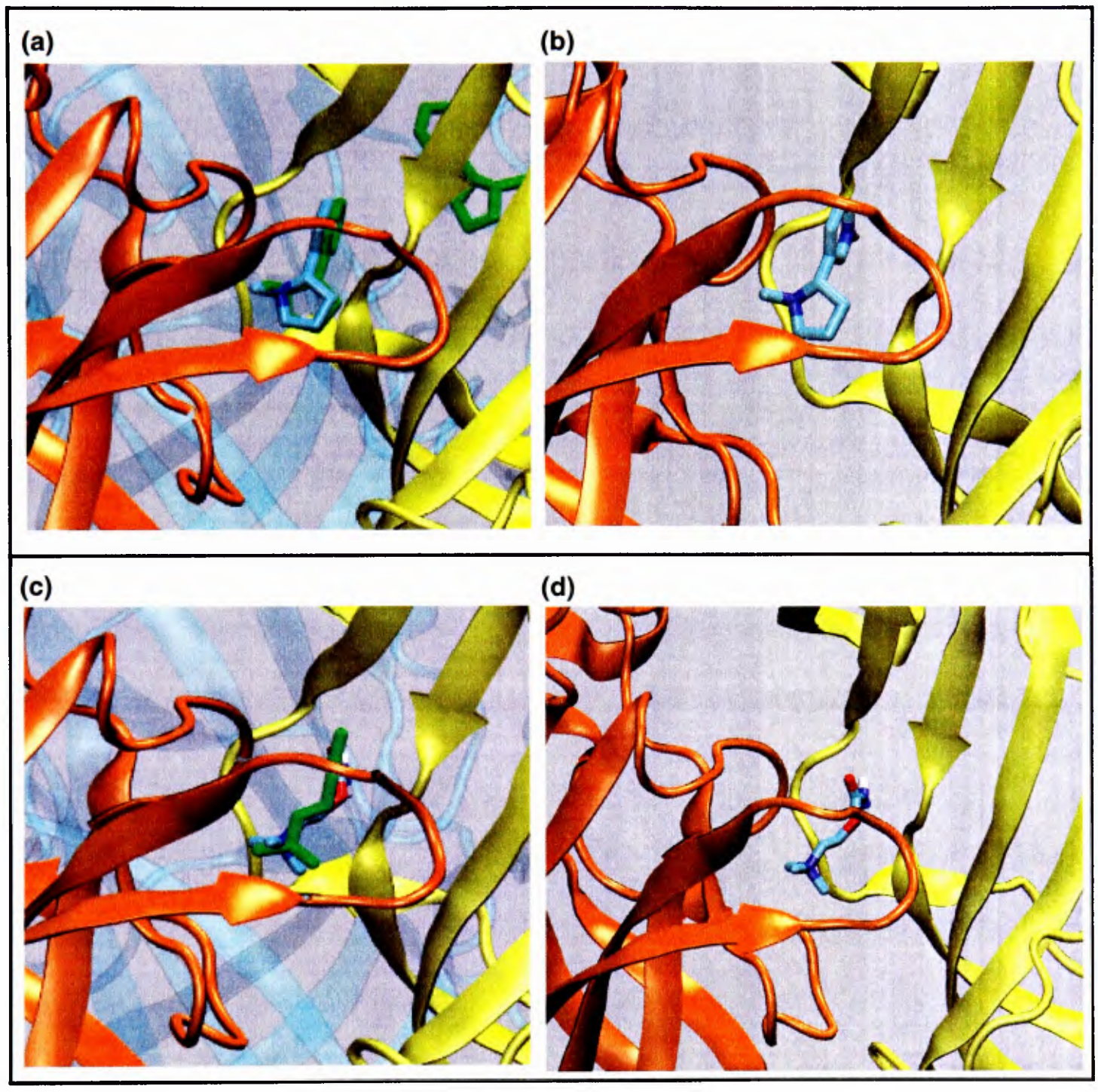


Figure 4.3 (a) An overlay of docked nicotine against the nicotine (green) in the crystal structure of AChBP (PBD code 1UW6) – note: there are 5 nicotine molecules in the X-ray structure (the other green nicotine is in the adjacent subunit). (b) Nicotine docked back onto AChBP. (c) An overlay of docked carbamylcholine onto the carbamylcholine (green) in the crystal structure of AChBP with carbamylcholine bound (PBD code 1UV6). (d) Carbamylcholine dock onto AChBP.

4.3.2 Docking ligands onto MD trajectories

Nicotine, acetylcholine and carbamylcholine were docked onto successive frames of the four simulations listed in Table 3.1. These simulations are discussed in detail in chapter 3. NCT was run with nicotine in each of its five binding sites, NCT-apo had the nicotine molecules taken out for a ligand-free simulation. CCE was run with two carbamylcholine molecules in binding pockets at the interface of subunits C/D and D/E and one of the ligand-bound subunits was used for the docking studies.

Two general trends have been observed with the docking of nicotine, acetylcholine, and carbamylcholine along the trajectories of AChBP MD simulations: one is the improved docking of ligands across the trajectories of ligand-bound simulations (and deterioration on ligand-free simulations) and second, ligands docked most favourably (lowest energies, correct binding mode, and good consensus amongst docking attempts) onto the simulation of AChBP with the same or similar ligand bound. For example, nicotine docks back most favourably onto the simulation of AChBP with nicotine bound (PDB code 1UW6).

Nicotine, carbamylcholine, and acetylcholine docks on NCT frames are compared to docks onto NCT-apo frames to observe the difference when docking a ligand back onto a ligand-free *versus* a ligand-bound simulation (Figure 4.4). All three ligands show improved docking on the ligand-bound NCT simulation while the docking declines on the ligand-free NCT-apo. The most striking difference is observed for nicotine, which docks very well onto NCT such that the docking energies continue to improve along the trajectory. However, for NCT-apo the docks do not improve and nicotine ends up out of the binding site towards the end of this simulation (Figure 4.4 (a)). The same trend is true for both carbamylcholine in Figure 4.4 (b) and

acetylcholine in Figure 4.4 (c), where the docks improve on NCT and deteriorate on NCT-apo; however the improvement of energies here is much more subtle.

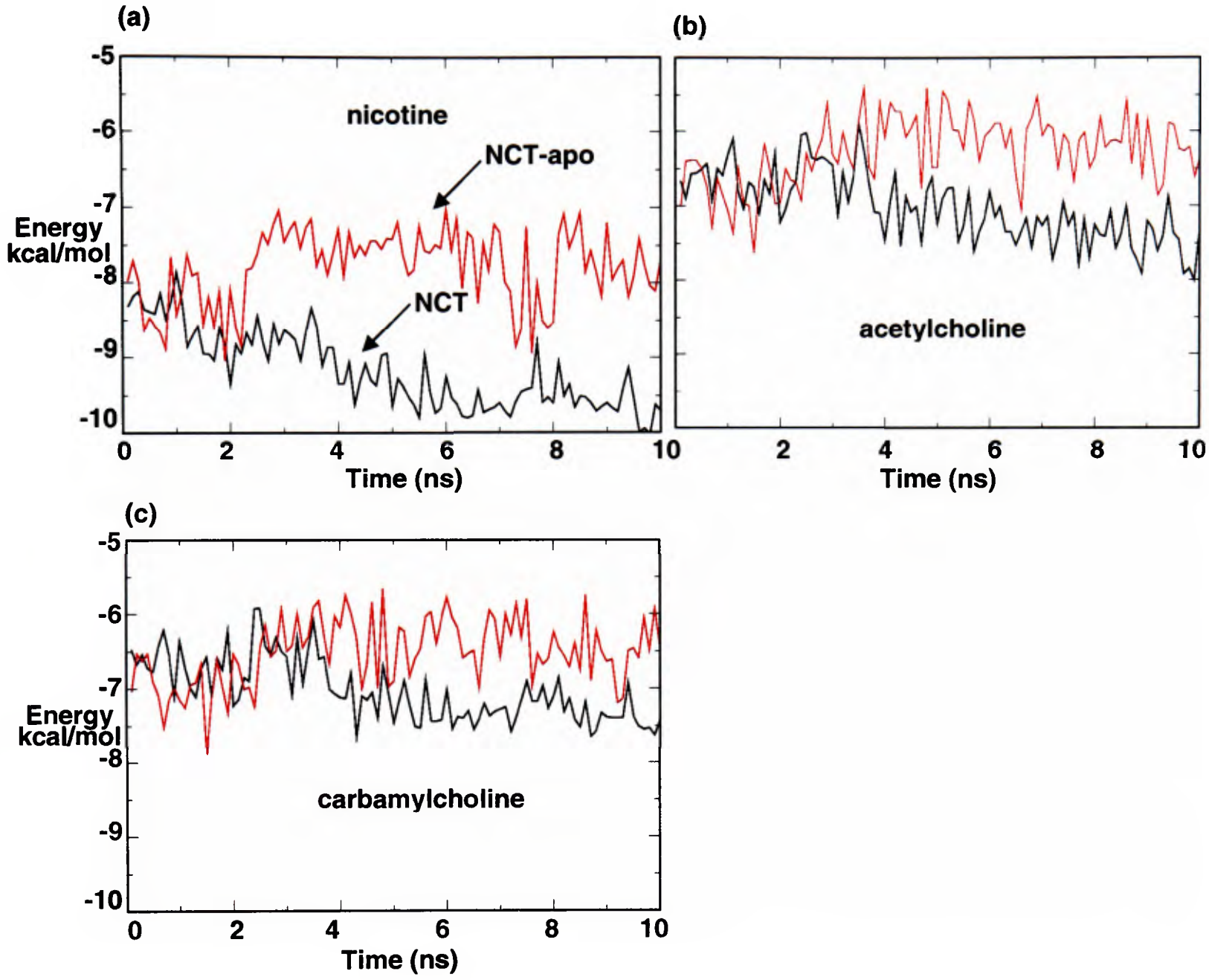


Figure 4.4 (a) Nicotine docks along NCT (black) and NCT-apo (red). (b) NCT (black) and NCT-apo (red) with carbamylcholine docks along the trajectory. (c) Acetylcholine docks long the NCT (black) and NCT-apo (red) trajectories. All three ligands show a difference of docking energies between NCT and NCT-apo. The docking does not improve along the ligand-free trajectories for any of the ligands.

The same trend is illustrated for CCE and CCE-apo simulations. Acetylcholine and carbamylcholine docked very well onto CCE frames and the docking improves along the trajectory. The docks of these same ligands did not improve along CCE-apo, and again the docks become slightly worse with time. Figure 4.5 shows that nicotine, carbamylcholine, and acetylcholine dock much more favourably onto CCE frames than on CCE-apo, with a large difference for nicotine, followed by acetylcholine, and

finally carbamylcholine. The difference between CCE and CCE-apo for carbamylcholine is larger than for carbamylcholine docks on NCT and NCT-apo, illustrating the preferred binding of carbamylcholine back onto the structure which was run with carbamylcholine bound in the MD simulation, as was the case with nicotine.

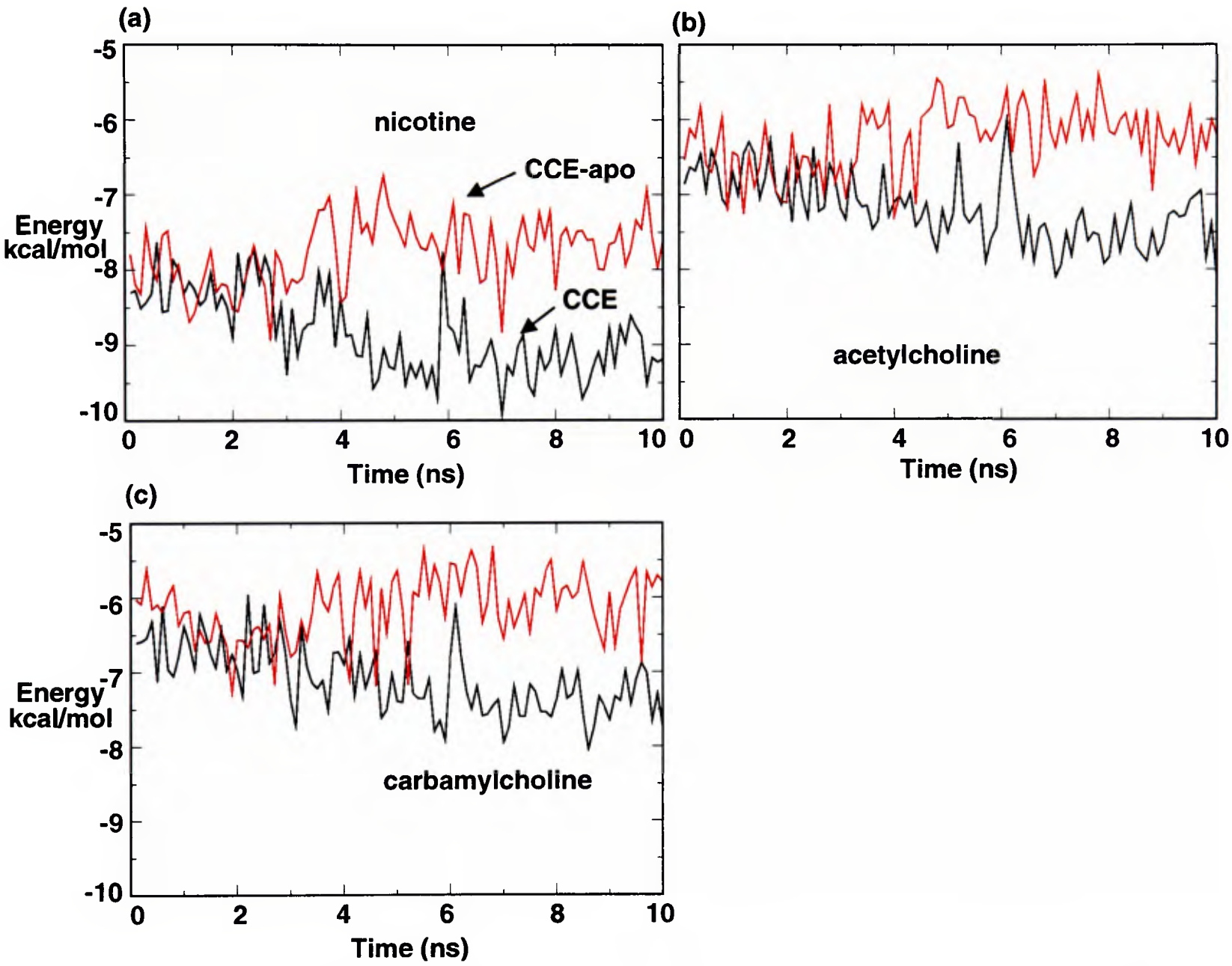


Figure 4.5 (a) CCE (black) and CCE-apo (red) with nicotine bound along the trajectory. (b) CCE (black) and CCE-apo (red) with acetylcholine docked along the trajectory. (c) CCE (black) and CCE-apo (red) with carbamylcholine docked along the trajectory. All docks on ligand-bound simulations become more favourable with time, this is not the case in ligand-free frames.

Thus the improved binding of ligands onto ligand-bound simulations and the favoured docking back of a ligand onto a simulation run with the same ligand is demonstrated. This is indicative of a different liganded structural state versus a ligand-free state of

the binding pocket and perhaps a 'favoured' structural state for each ligand. Also, this improvement shows that the binding site of ligand-bound simulations had optimized to the ligand during the simulation.

Figure 4.6 compares the energies for each ligand in the ligand-bound simulations. In NCT and CCE, the docking energy of all three ligands becomes more favourable along the 10 ns trajectory. Docking energies became progressively lower with time, which is due to the changing structure of the binding pocket during the ligand-bound simulations. Nicotine had the lowest docking energies in the NCT simulation, it also had the lowest energy of all three ligands for CCE, however, these energies where not as low as the docking energies for NCT. The slightly more favourable docking of nicotine in the NCT trajectory, the quicker decrease in energy levels, and the lower over-all and final energies, suggests that there is structural adaptation of the binding pocket to the particular ligand during this simulation.

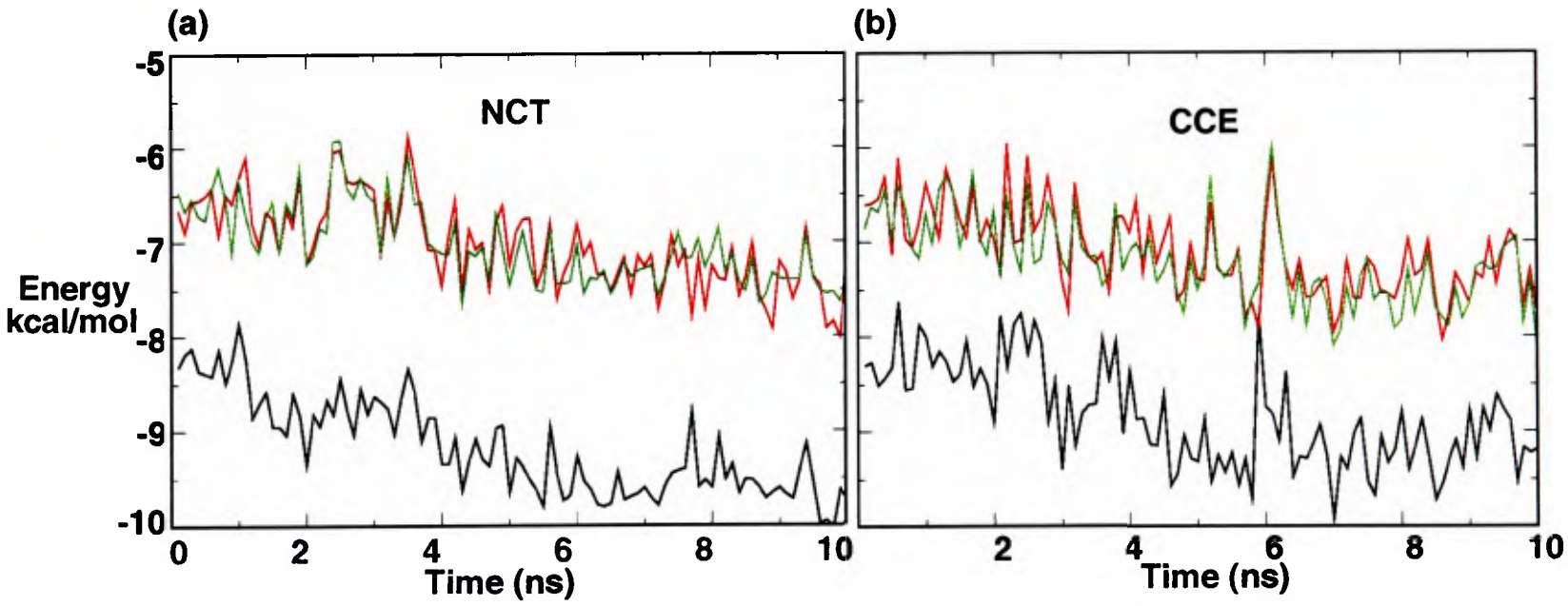


Figure 4.6 (a) Docking onto successive frames of the NCT simulation with nicotine (black), carbamylcholine (red) and acetylcholine (green). (b) Docking onto frames of the CCE simulation with nicotine (black), carbamylcholine (red) and acetylcholine (green). In both (a) and (b) nicotine docks with the lowest energy, and in NCT, nicotine has lower docking energies than for CCE. Carbamylcholine and acetylcholine docks improve along the trajectory.

The docking energies for the ligands are also compared for the ligand-free simulations (Figure 4.7). There was no improvement of docks onto the NCT-apo and CCE-apo

trajectories. Indeed, the docks worsen with each frame of the ligand-free simulations, and by the end of these simulations ligands cannot dock into the binding site. The binding pocket of the apo simulations, in particular loop C, collapse in, and obstruct the ligand from positioning itself in the pocket. RMSD calculations and distances of residues in the binding pocket show that there is more structure deviation of binding pocket residues when no ligand is bound (chapter 3) (Amiri *et al.*, PEDS – submitted). Thus the different structure of the apo binding pocket which results after only a few ns of simulation does not favour ligand binding. It seems that in all cases it takes ~2.5 ns for the structure of the binding pocket to change in response to the environment of the binding pocket, i.e. whether or not a ligand is bound.

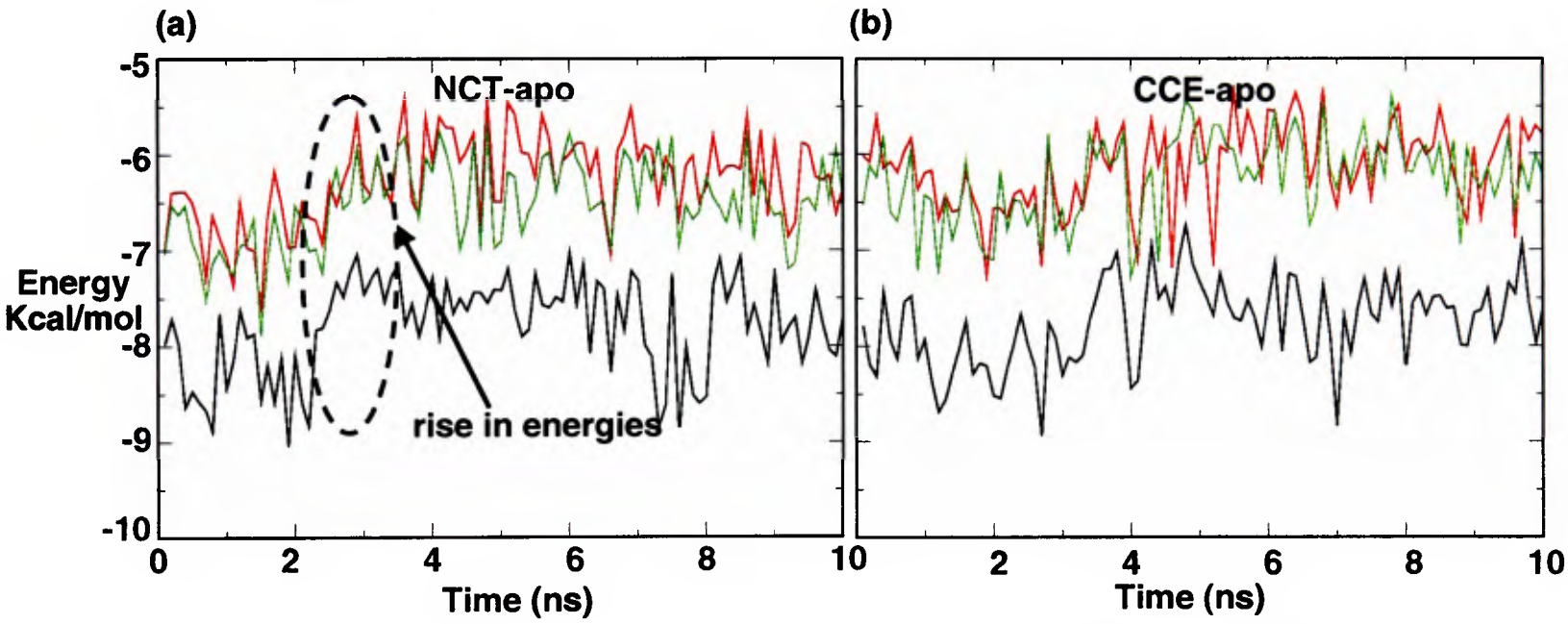


Figure 4.7 (a) Docking onto successive frames of the NCT-apo simulation with nicotine (black), carbamylcholine (red) and acetylcholine (green). (b) Docking onto frames of the CCE-apo simulation with nicotine (black), carbamylcholine (red) and acetylcholine (green). In both (a) and (b) the docking energies are higher than the ligand bound simulations and they do not improve along the trajectory.

An overlay of AChBP from NCT-apo at 2.4 ns and 3.0 ns, i.e. before and after a jump in energies (circled area on Figure 4.7 (a)), reveals a significant difference between the position of loop C in the two frames. At ~ 2.4 ns there is a sharp increase in energy and the docking energies are not stable or on the decline as in the ligand-bound simulations. Figure 4.8 shows the comparison of the NCT-apo structures at 2.4

and 3.0 ns. It can be seen that loop C comes closer into the binding pocket, thus changing the shape of the binding site and perhaps obstructing the binding of ligands. Due to the absence of a ligand, the binding pocket of the apo simulations change shape within the first few nanoseconds and thus are no longer capable of accommodating ligand binding.

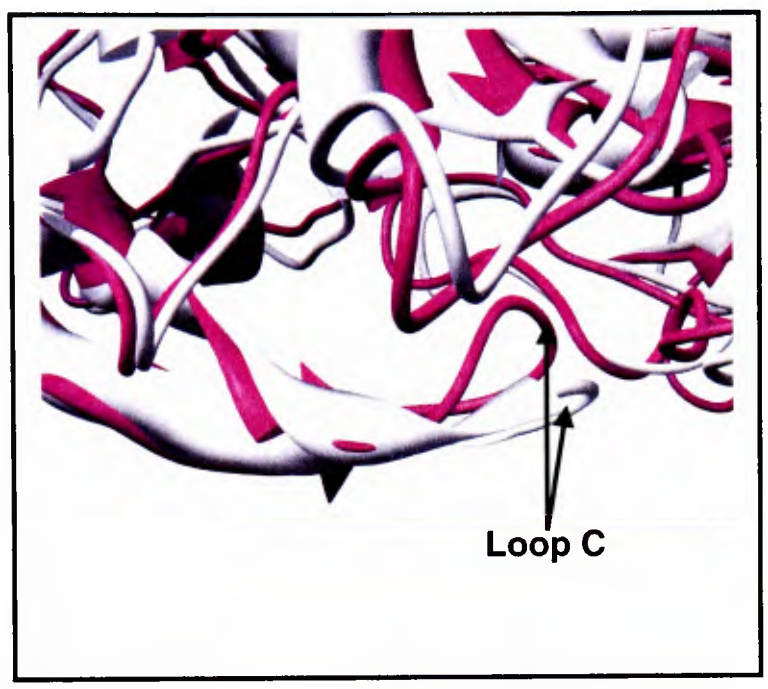


Figure 4.8 A top view overlay of the binding pocket from subunits B/C from NCT-apo, illustrates the changing position of loop C from 2.4 ns (grey) to 3.0 ns (pink). Due to the absence of a ligand in the simulation, the loop seems to collapse into the binding pocket. At ~2.4 ns the docking energies begin to rise and there is no consistent decline of docking energies as in the ligand-bound simulations.

As well as recording the energies for the top ranked dock of each frame, the number of docks in the top ranked cluster, out of 50, was also recorded. Thus for each frame:

- The ligand is docked onto the frame 50 times (user-defined)
- The docked conformations are clustered by structural similarity (RMSD)
- The mode(s) of binding (clusters) are ranked by increasing energy
- The number of docks in the lowest energy cluster (top ranked dock) is extracted

The size of each cluster is another indication of the favourability of a particular binding mode; the larger the cluster, the more likely that a particular binding mode is energetically favourable. For example, nicotine usually has very few possible predicted modes of binding and thus there are only one or very few clusters dividing the 50 docks, indicating consensus amongst the 50 dock attempts. It seems that there

is only one main mode of binding for this particular ligand, however acetylcholine and carbamylcholine usually have several clusters with similar binding modes. The small size and rotatable bonds of acetylcholine and carbamylcholine may make it more difficult for docking programs to distinguish one mode, if indeed there is only one binding mode. Fortunately, there is an X-ray structure of AChBP with carbamylcholine bound which provides one mode of binding, although others could be possible.

For NCT, nicotine has a very high number of docks in its top ranked cluster; the number of docks in this cluster are close to 50 in some frames (100% consistency) (Figure 4.9 (a)). Thus there is agreement for one mode of binding as most of the 50 docking attempts of each frame fall into one cluster. Interestingly, the number in the top cluster is decreasing slightly, even though docking is becoming energetically more favourable.

There are more clusters, but, a relatively high number of docks are in the top ranked cluster for acetylcholine and carbamylcholine docks on NCT (Figure 4.9 (a)). These are in agreement with the binding mode of the X-ray structure of carbamylcholine. Significantly, in NCT-apo (Figure 4.9 (b)) all the top ranked clusters have fewer docks and this result is particularly noticeable for nicotine, whose numbers completely deteriorate in the apo state. The number of docks in the highest ranked cluster is also lower for acetylcholine and carbamylcholine and the modes of binding do not match those of the X-ray structure in many frames, with some docks completely outside of the binding pocket.

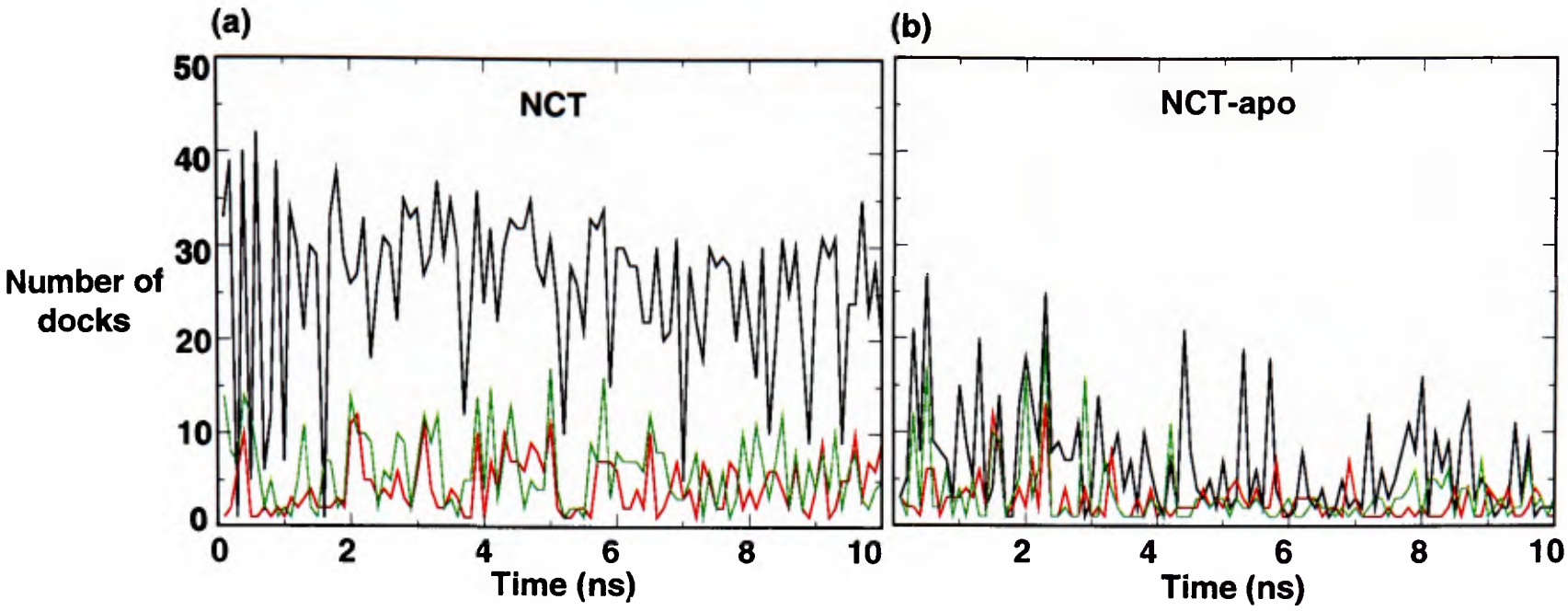


Figure 4.9 (a) NCT with nicotine (black), carbamylcholine (red), and acetylcholine (green) docked along the trajectory. (b) NCT-apo with nicotine (black), carbamylcholine (red), and acetylcholine (green) docking along the trajectory. Nicotine has very high numbers in its top ranked dock for each frame compared to that of the ligand-free simulation.

Acetylcholine and carbamylcholine have higher numbers of docks in the top clusters for each frame in the CCE simulation (Figure 4.10) than for the NCT simulation. Thus the binding pocket of AChBP which had carbamylcholine bound during the simulation is more accommodating to acetylcholine and carbamylcholine. Nicotine docks with less success onto CCE, only slightly better than acetylcholine. However, on NCT nicotine docks had significantly higher numbers in their top clusters compared to acetylcholine. Thus the MD simulation of AChBP which was run with nicotine in the binding pocket (NCT) is structurally favoured by nicotine when the ligand is re-docked onto frames of the trajectory, and the same trend holds for carbamylcholine and acetylcholine which favour the binding pocket of CCE.

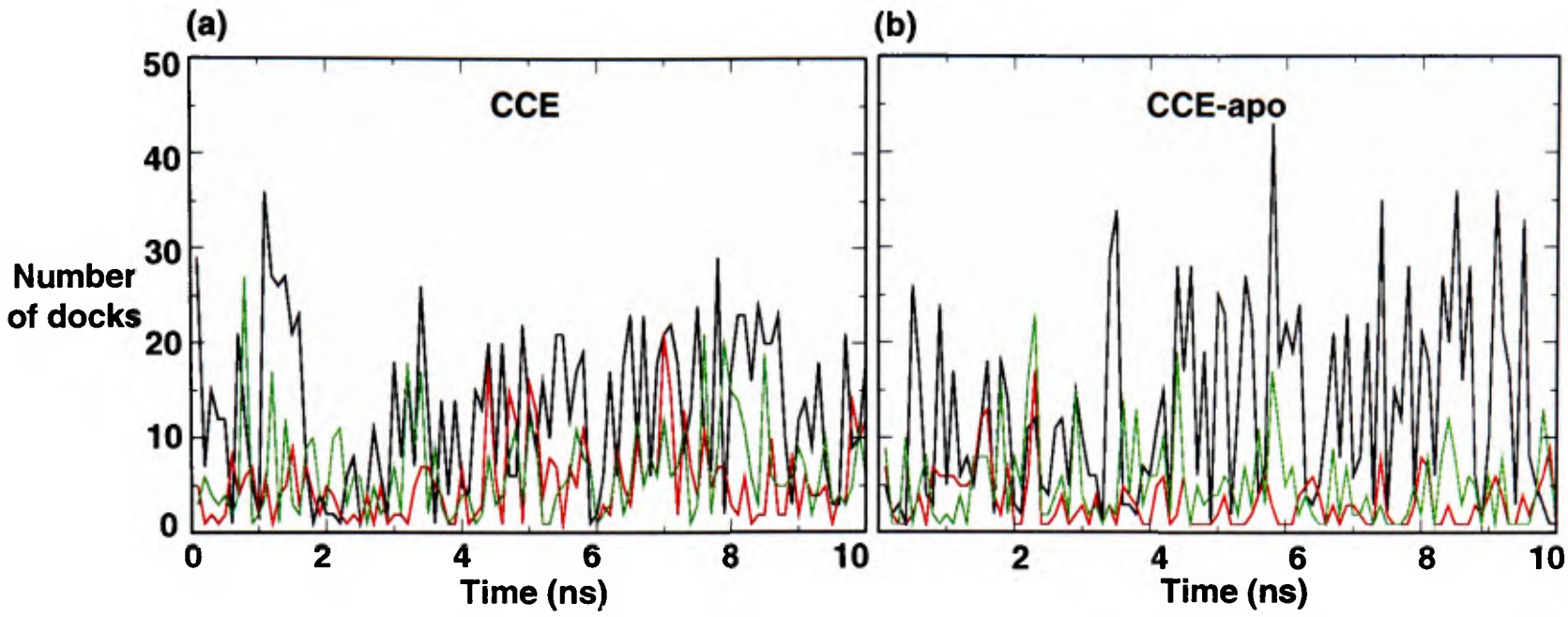


Figure 4.10 (a) CCE with nicotine (black), carbamylcholine (red) and acetylcholine (green) docked along the trajectory. (b) CCE-apo with nicotine (black), carbamylcholine (red) and acetylcholine (green) docked along the trajectory. Acetylcholine and carbamylcholine have larger lower energy clusters for CCE, however nicotine clusters are higher in the ligand-free simulation.

A comparison of the number of docks in the top ranked cluster of each frame between ligand-bound and ligand-free simulations is shown in Figure 4.11 and Figure 4.12. There is a large difference for nicotine docking onto NCT and NCT-apo frames. Also, both acetylcholine and carbamylcholine have higher numbers of top ranked docks on the ligand-bound simulations, but the difference is not as striking as it is for nicotine. As expected, acetylcholine and carbamylcholine have more docks in their top cluster for the NCT and CCE simulations rather than for NCT-apo and CCE-apo, and the docks improve with time. However, surprisingly, nicotine has a higher number of top ranked docks on CCE-apo, rather than CCE (Figure 4.10). This strange behaviour of nicotine, having more docks in the top cluster for CCE-apo than CCE, is not shared with the energy pattern, as nicotine docks with lower energies with improvements along the ligand-bound trajectory (Figure 4.5 (a)). Also, further analysis into why the top cluster numbers are high for nicotine binding in CCE-apo,

shows that nicotine is not binding in the correct mode and it is low in the binding pocket.

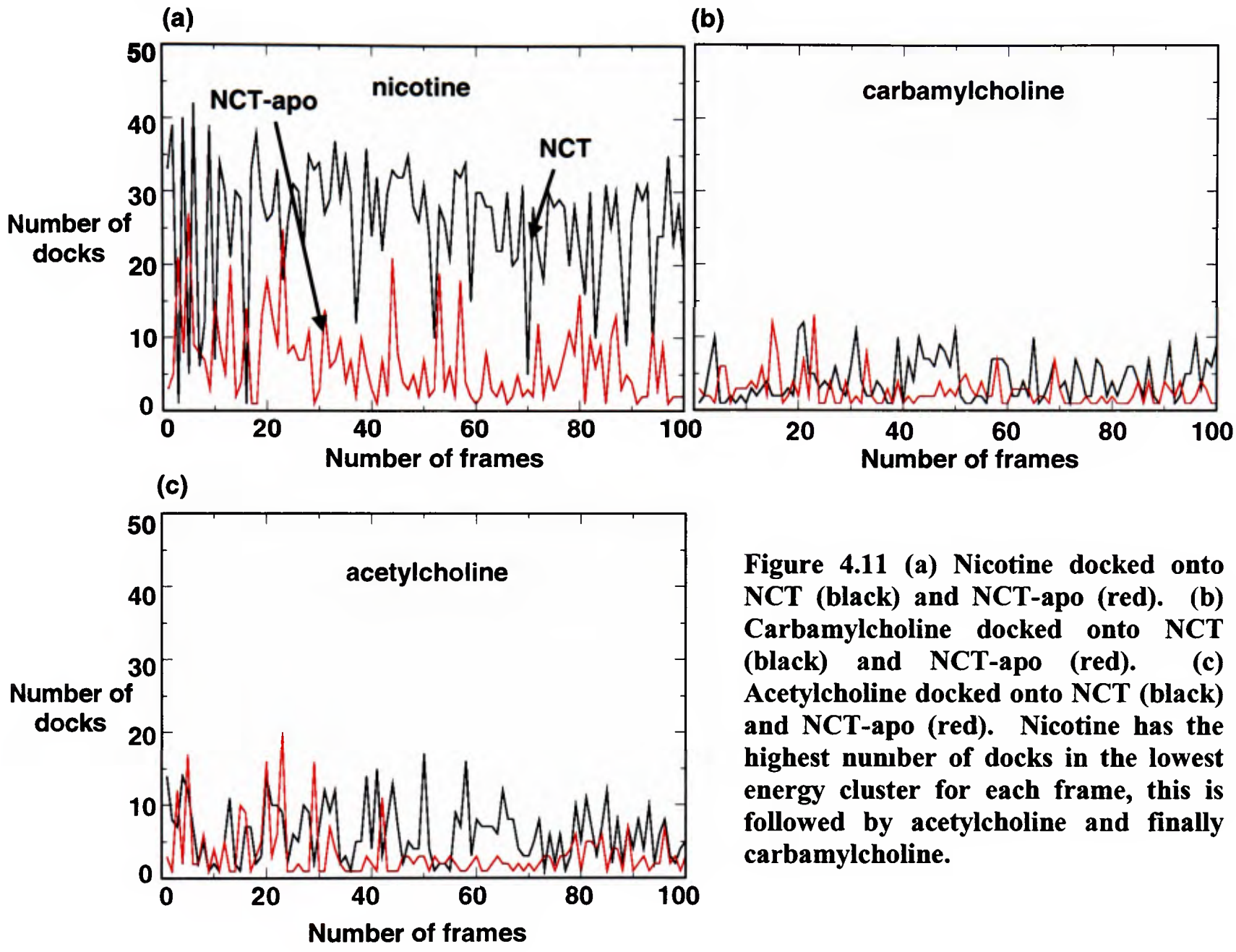


Figure 4.11 (a) Nicotine docked onto NCT (black) and NCT-apo (red). (b) Carbamylcholine docked onto NCT (black) and NCT-apo (red). (c) Acetylcholine docked onto NCT (black) and NCT-apo (red). Nicotine has the highest number of docks in the lowest energy cluster for each frame, this is followed by acetylcholine and finally carbamylcholine.

This case exposes a problem with this type of docking, in particular 'high-throughput' docking where trying to look at the results for each frame is extremely time-consuming due to the huge amount of data generated (100 x 50 x number of simulations). In ensemble docking, one can only look at the top dock or a few of the top docks for each frame, however sometimes the best dock may be in a lower ranked cluster, and thus one may miss important docking results. Energies are very close between the ranks and thus looking only at the top dock may not be enough. Also the top docks are not always the best mode of binding or those matching the X-ray structure. Thus, these analysis methods are not perfect and need improvement,

especially when one wants to observe more detail for each snapshot taken for the simulation. However, nicotine docking behaviour may also support the hypothesis that the binding pocket has structurally conformed to carbamylcholine in CCE and nicotine cannot as easily dock into this state.

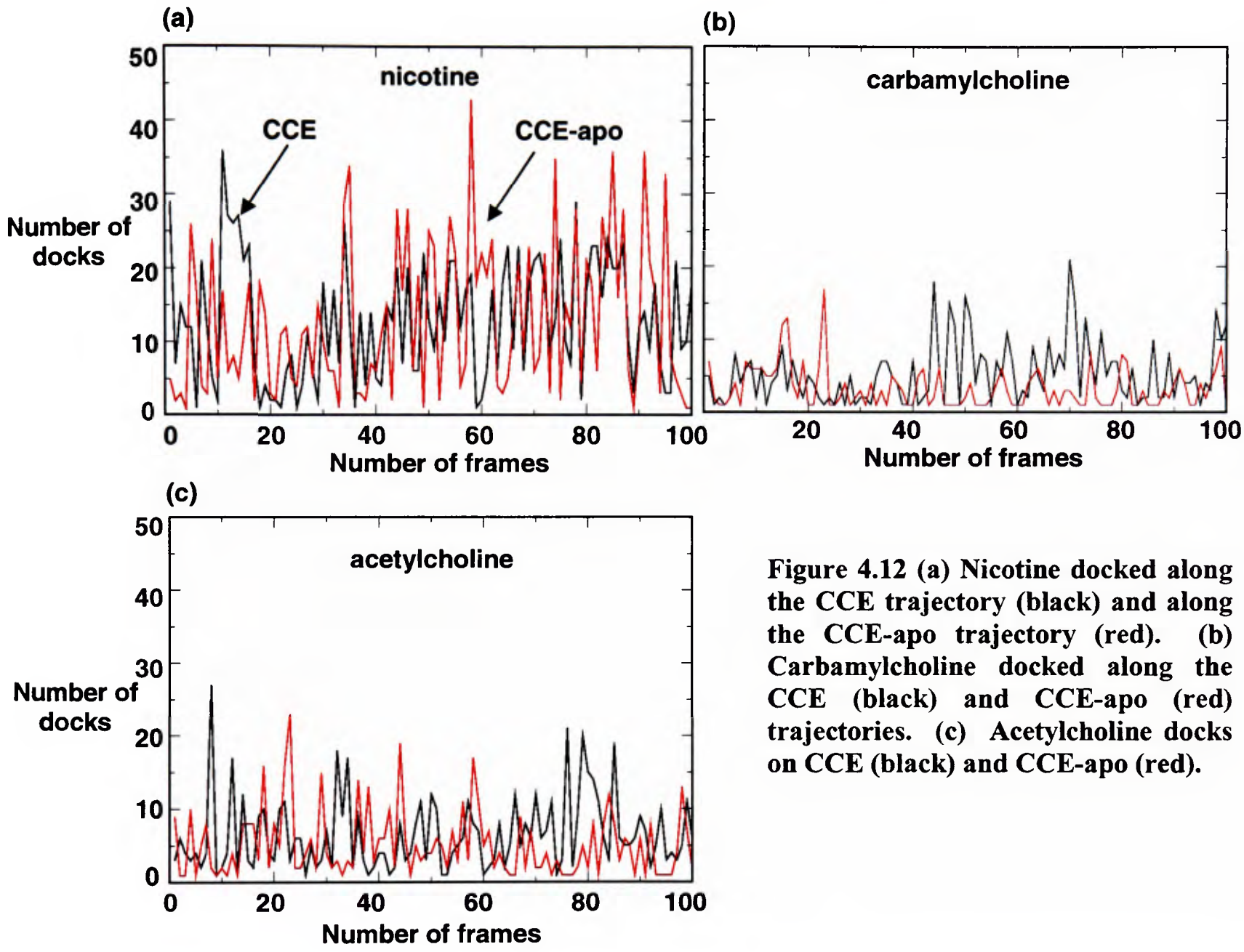


Figure 4.12 (a) Nicotine docked along the CCE trajectory (black) and along the CCE-apo trajectory (red). (b) Carbamylcholine docked along the CCE (black) and CCE-apo (red) trajectories. (c) Acetylcholine docks on CCE (black) and CCE-apo (red).

4.4 Discussion

Nicotine, acetylcholine, and carbamylcholine were docked both onto the X-ray structure of AChBP and onto multiple frames of ligand-bound and ligand-free simulations of AChBP. Docking of all ligands showed consistent improvement in terms of energies, along the ligand-bound simulations, while docking deteriorated against ligand-free simulations. This indicates that a ligand changes the structure of

the binding pocket with time such that docking a ligand back onto the final frames of ligand-bound simulations is most favourable.

Nicotine docked best onto the simulation which was run with nicotine bound (NCT) and had the lowest energy of docks on CCE as well. Carbamylcholine and acetylcholine had similar docking energies in both simulations, with acetylcholine having slightly better docks. This behaviour suggests that the binding pocket conforms to a particular ligand and thus docking back the same ligand is more favourable than docking a different ligand. Since the binding pocket becomes more accommodating to a particular ligand, docking is dependent on the structure of the binding pocket.

By carrying out docking on the MD simulation trajectories, ligand docks were studied on the slightly different conformations of the binding site from the successive frames of the simulations. Thus it was possible to distinguish the structurally favourable binding pockets for ligand docking. Other studies which have also used the combined method of MD and docking to 1 – account for protein flexibility [222] and 2 – to identify correct modes of binding using several initial docked states [223, 224] have reported successful results in comparison to using and trusting only one docking mode.

With respect to the previous chapter, here it is observed that the structure of the binding pocket is optimized to the bound ligand and thus docking back the same ligand is energetically favourable with improvements along the trajectory. The shape of the binding pocket is changing during the simulation and this change results in the decrease of docking energies in the ligand-bound simulations. Also the structure of the binding pocket in the ligand-free simulations is not accommodating to ligand

docking, thus there is a distinct difference in structure of the binding pocket in the presence and absence of a ligand as was reported in the previous chapter.

However, it was shown here that this type of 'high-throughput' docking may result in some important data being overlooked as it is very time-consuming to go through each dock for every frame of the simulation. It is also difficult to distinguish a good dock from a bad dock by the docking energy assignments alone, as a dock with low energy can lie outside the binding site and be in an incorrect binding mode, thus visual inspection of every dock is important yet impractical with this method. These docking studies could be further improved with better docking software that allows for more obvious recognition of the correct binding mode.

5 Generating the structure of the α7 nicotinic acetylcholine receptor

5.1 Introduction

Computational methods allow us to take incomplete structural data from various sources and combine them to generate a more complete structure. Currently, there are no high resolution X-ray structures available for a nAChR. However, there are structures of separate components which enable us to 'reconstruct' the structure of the nAChR at atomic resolution. Recently several high resolution X-ray structures have become available for a water-soluble homologue of the EC domain, the acetylcholine binding protein (AChBP) from the snail *Lymnaea stagnalis* and *Aplysia californica* [20, 74, 77]. At the time this work was carried out, only one structure was available for AChBP from *Lymnaea stagnalis* [20] and this is the structure which was used to model the EC domain of the α7 nAChR. Unwin and colleagues [13, 29, 60, 226, 227] have generated a 4 Å resolution image of the TM domain of the nAChR from the electric ray *Torpedo marmorata* using cryo-electron microscopy (EM) (Protein Data

Bank (PDB) code 1OED). Furthermore, a chimeric protein formed by fusing the AChBP with the TM domain of a nAChR homologue (the 5-HT_3 receptor) has shown to form ligand-gated channels [101]. Since the structure of AChBP has been solved, several groups have used this structure to carry out modeling and ligand-receptor studies of LGICs [30, 31, 39, 128, 228].

Molecular dynamics simulations of a homology model of α7 nAChR extracellular (EC) domain based on AChBP have also been carried out showing asymmetrical patterns related to channel gating and structural changes in the presence of a ligand [129]. Fewer studies have focused on models of an intact nAChR. Recently however, an MD simulation has been carried out on a α7 nAChR model [131]. Preliminary studies on the electrostatics and energetics of cation permeation through open and closed models of the acetylcholine receptor have been reported [139, 140]. The AChBP structure used here is believed to be structurally very close to the desensitized state of the nAChR receptor, it is still unclear whether it closely resembles the closed state of an LGIC.

Thus there were molecular structures of the TM domain of a nAChR based on EM data, and an X-ray structure of a homologue of the EC domain. The intracellular (IC) domain of the protein (formed by the region of polypeptide chain between M3 and M4) is missing due to the absence of a template structure for this region during the time of this work.

With models for the TM and EC domains, the remaining problem was how to reassemble these fragments into a model of a (more or less) complete nAChR. Here the process by which the structure of the α7 nAChR is created and used for structural studies is described. This general method is applicable to other proteins and protein complexes.

The recent 4 Å resolution structure of the full nAChR from *Torpedo mamorata* (Unwin, 2005) has resolved of 80% of the 2335 residues and includes much of the IC domain. However, there are still many missing residues mostly located in the M3/M4 IC loop, loops 7 and 8 of the non-α subunits, and the C termini of δ (12 residues) and γ subunits (17 residues). Also several of the loop regions are poorly defined and thus prevent detailed interpretation [19]. This IC region may have important consequences with respect to future calculations of ion conductance and selectivity [44]. The 4 Å nAChR structure is compared to the constructed α7 nAChR model later in this chapter.

5.2 Methods

An outline of the method to generate an optimal model of an intact α7 nAChR, and examine aspects of its predicted dynamic behaviour, is given in Figure 5.1. Briefly, homology models of the α7 TM and EC domains were constructed independently of one another, using the *Torpedo* TM domain (PDB code 1OED) and snail AChBP (PDB code 1I9B) structures respectively as templates. The resultant models of the α7 TM and α7 EC domains were then 'docked and joined' by the procedure described in more detail below. The stereochemistry of the final model was then adjusted by energy minimization then stereochemical verification was carried out with procheck [229]. The optimal structure was used for coarse-grain motion analysis, electrostatics, and pore radius calculations. Gaussian network models (GNM) [164-167, 204, 230] and CONCOORD were used for the coarse-grain conformational sampling analysis, Poisson-Boltzmann calculations for the electrostatics calculations (described in [62], and HOLE [231] for the pore-profile studies.

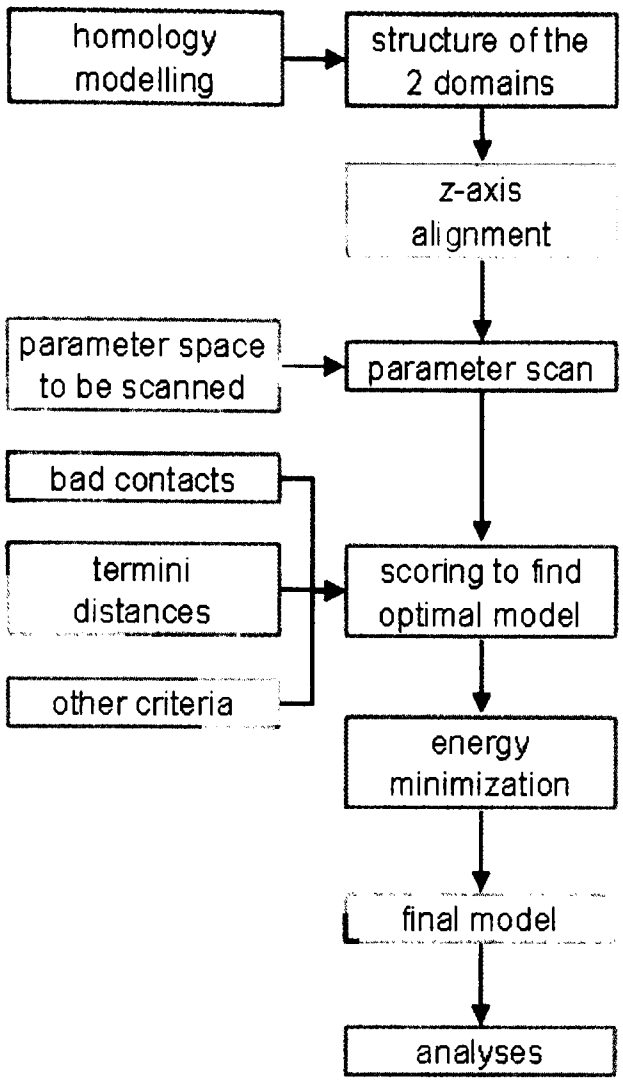


Figure 5.1 The process of generating a homology model from the sum of two or more structures and using this model to carry out structural studies. ZAlign is used to combine 2 or more homology models by exploring the parameter space defined for the structure. An optimal model is chosen and energy minimized, it is then used for further analyses.

5.2.1 Homology modelling of the EC and TM domains

For the structure prediction of α7 nAChR, the first step was homology modelling of the separate domains of the receptor using available crystallographic and microscopy data. Both the EC and TM domains were modeled as homopentamers, each domain consisting of five α subunits. Given a structure template and a sequence of choice, Modeller produced 100 possible structures. For the TM domain, the chicken α7 TM domain sequence was modelled onto the *Torpedo marmorata* TM domain structure (PDB code 1OED) [13]. For the EC domain, the chicken α7 sequence was modelled onto the 2.7 Å AChBP structure (PDB code 1I9B) [20]. The chicken α7 sequence is homopentameric like the AChBP, which has 27 % sequence identity to the α7 nAChRs. Although the 2.7 Å AChBP structure was used initially, more recent models were generated using the newer structure at higher resolutions [77].

Modeller 6v2 [141-143] was used to generate homology models of the chicken α7 nAChR sequence for each domain. The top 10 models for each domain were ranked according to their Modeller energies. In order to choose the best of the top 10 structures, Procheck [186] was run to check the stereochemical quality of each model. The model with the best overall Procheck score was selected and used in subsequent structural studies. This same procedure was carried out for both the TM and EC domains. The sequence alignment used is shown in Figure 5.2. The sequence numbering scheme used in the remainder of this chapter corresponds to that of the chicken sequence in the alignment shown. Residues 263 – 290 correspond to the M2 helix, which also corresponds to 1′ to 28′ using the 'prime' numbering scheme for M2 of Auerbach and others [15, 51, 232]. Residue D289 (alignment) = D27′, situated at the C-terminus of M2, corresponds to D5* in the 'star' numbering scheme used by Lester *et al.* [15] for the M2-M3 linker. Residue D5* is proposed by Unwin and colleagues [13] to interact with a residue at the tip of the β1-β2 loop of the EC domain which could play a role in transmitting conformational changes from the EC domain to the TM domain. In the chicken α7 model, an equivalent β1-β2 loop to M2-M3 loop interaction between residues K68 and D289 (alignment numbering) is proposed.

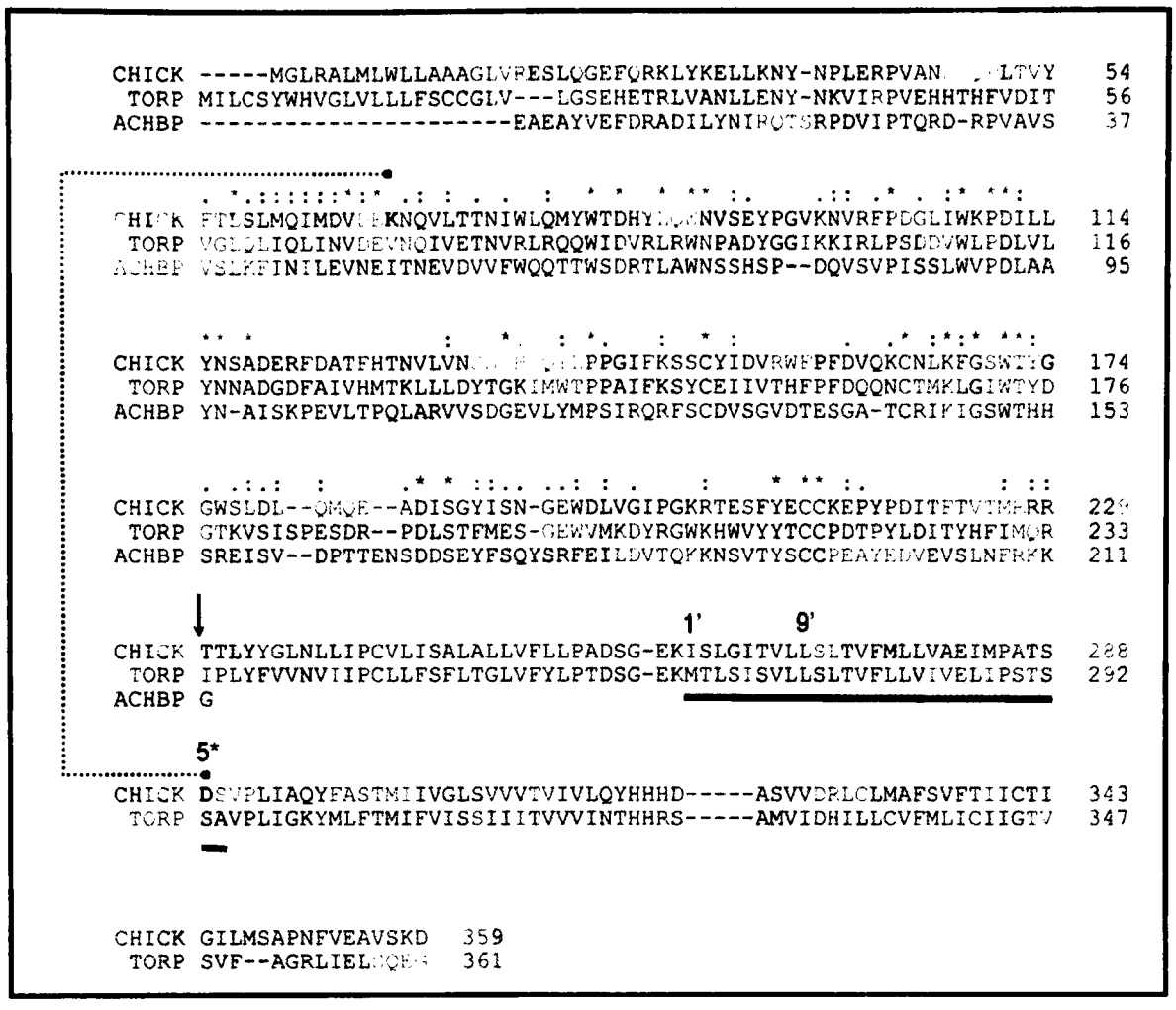


Figure 5.2 The *Torpedo marmorata* nAChR α subunit sequence (for the TM domain), and the *Lymnaea stagnalis* (snail) AChBP sequence (for the EC domain), were used and aligned with the chicken α7 nAChR sequence. Figure taken from [233].

After generating homology models for each domain, the goal was to combine both domains in a restraint-based method to construct a model of the α7 nAChR which includes both the EC and TM domains.

5.2.2 The alignment of multiple domains of proteins – ZAlign

In order to combine the two domains, the 5-fold rotational axes of the TM and EC domains of the receptor were aligned. The bottom domain was held static while the top domain was rotated and translated in the space specified in order to determine the optimal model. The parameter space for alignment of the separate domains must be determined by the user.

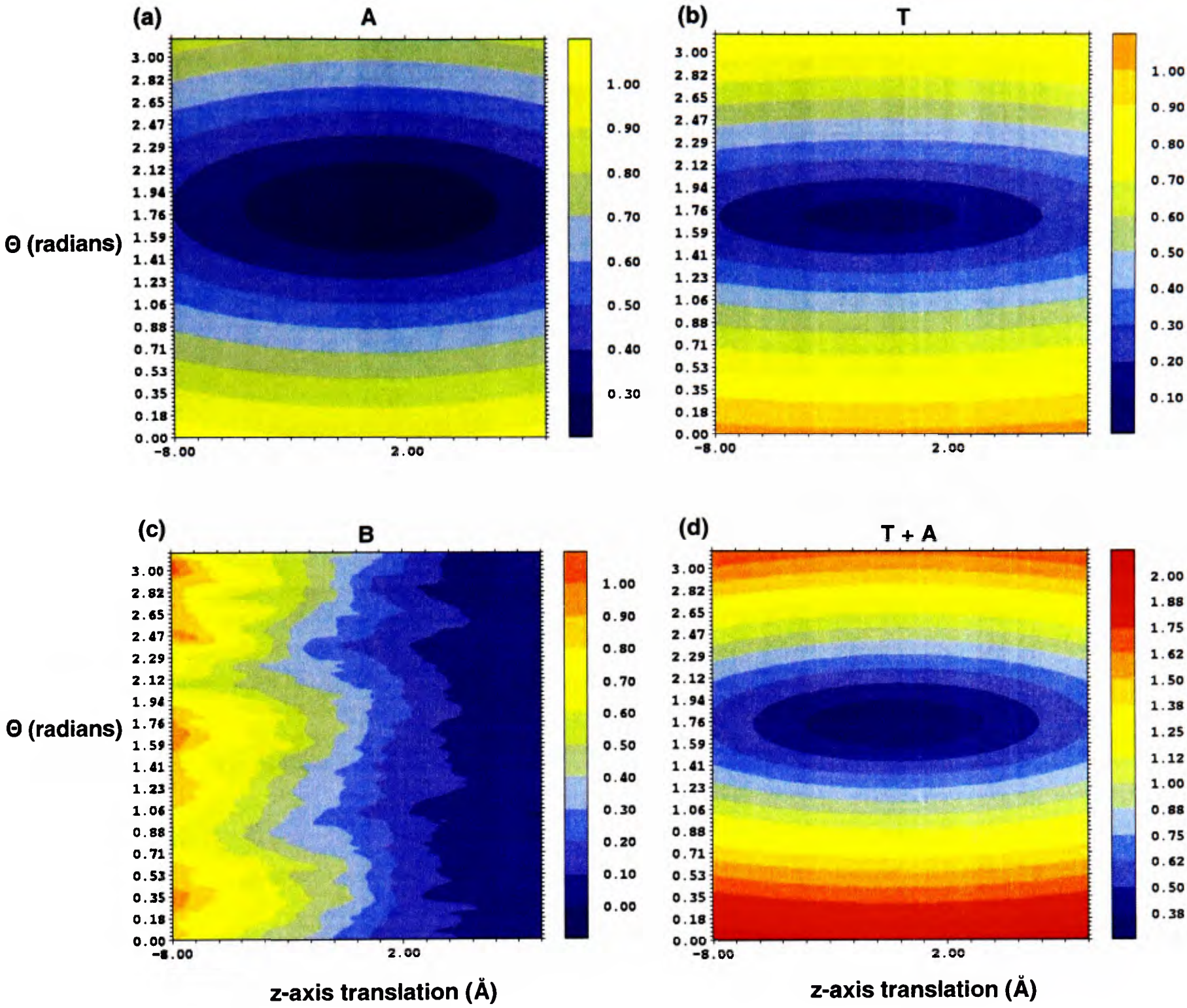
In this case, a full 360° degree rotation about the z-axis and $-8 \leq z \geq +8$ Å translations on the z-axis were completed. Because the protein is a homopentamer, the program was also run with 72° degree rotation. This served as a check against the results of the complete 360° rotation, which produced matching results as expected. The two domains were brought close together, and rotations and corresponding translations were carried out. In this example rotations of 1° and translations of 1 Å around and on the pore (z)-axis respectively were carried out. At each step a score was calculated based on various criteria. For the α7 nAChR model the scoring was based on three criteria:

i: The number of bad contacts, ***B***, between atoms of the EC and TM domains. A 'bad contact' is when the distance between atoms of the separate domains is less than 1.5 Å.

ii: The distance, ***T***, between the C-terminus of the EC domain and the N terminus of the TM domain.

iii: An additional criteria, ***A***, which is the distance between Cα atoms of K68 and D289 which correspond to the proposed interaction between a Val at the end of the EC β1-β2 loop and a Ser on the M2-M3 linker region of the TM domain. This interaction is thought to play a major role in linking conformational transitions in the EC and TM domains. The overall scoring function becomes: $\mathbf{S} = w_B\boldsymbol{B} + w_T\boldsymbol{T} + w_A\boldsymbol{A}$, a linear combination of the three separate scores with all weights giving equal value to each criterion.

Various contour plots were generated, one for each scoring criteria, and one for three linear combinations were produced (Figure 5.3). The best model corresponds to the lowest value of S.


(a)
A
(b)
T
(c)
B
(d)
T + A
Θ (radians)
Θ (radians)
3.00
2.82
2.65
2.47
2.29
2.12
1.94
1.76
1.59
1.41
1.23
1.06
0.88
0.71
0.53
0.35
0.18
0.00
-8.00
2.00
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
2.00
1.88
1.75
1.62
1.50
1.38
1.25
1.12
1.00
0.88
0.75
0.62
0.50
0.38
z-axis translation (Å)
z-axis translation (Å)

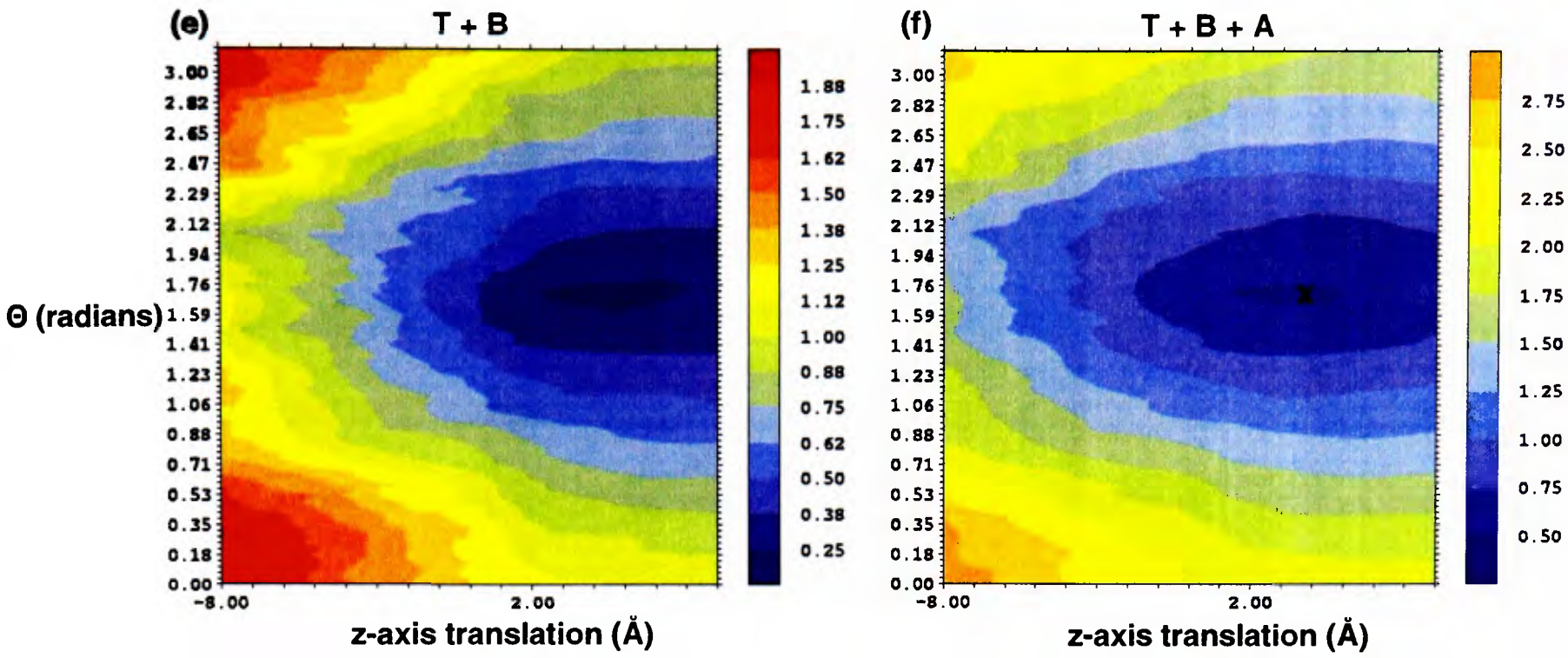


Figure 5.3 Contour plots of the three scoring functions and their linear combinations. Colour range: red represents the highest distances and highest number of bad contacts while dark blue represents the smallest distance and the least number of bad contacts. (a) A: Cα distance between K68 and D289. (b) Termini distance between the C-terminus of the EC domain and the N-terminus of the TM domain. (c) Bad contacts between atoms in the EC and TM domain. (d) Linear combination of A and T distances. (e) Linear combination of T and B. (f) A linear combination of all three scoring criteria, the darkest region is where all three scored are optimized and where the model was chosen from.

The alignment method allows the generation of a model consistent with the input restraints, which include chain connectivity, minimization of stereochemically unfavourable contacts between the EC and TM domains, and placement of the EC domain β1-β2 loop close to the M2-M3 loop in the TM domain. From the figures shown (Figure 5.3), the criteria used to generate the model yield a single, well defined minimum on the scoring function contour plot, and thus a model can be unambiguously assigned. In the generated model, the EC domain interacts closely with the TM domain. The M4 helices are packed rather more loosely on the surface of the M1-M3 TM helix bundle.

5.2.3 Stereochemical checks and energy minimization

Once the optimal model was chosen according to the scoring function from the ZAlign program, Procheck was used to carry out stereochemical checks on the model. This program allows assessment of model quality on the basis of bad contacts, residues lying outside expected regions of Ramachandran space. The generated structure was given a core score of 91.1% which is relatively high value for a model. The structure was then energy minimized using GROMACS 3.1.4 (see Methods chapter) and the energy-minimised model of the combined EC and TM domains was used to investigate dynamic and energetic properties of this model in relation to nAChR function.

The model generated here is assumed to correspond to a closed conformation of the nAChR, as the AChBP crystallized with HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid) is believed to be structurally similar to a closed state [77] and the TM domain structure from *Torpedo marmorata* is also believed to be in a closed state as the receptor was imaged in the absence of a ligand. In order to understand how the structure relates to the function, it is of interest to attempt to predict the large-scale motions of the protein and relate them to the plausible conformational transition between the closed and open state of the receptor. This could be achieved by long time-scale atomistic molecular dynamics simulations. This method has been successfully employed for e.g. the EC domain of the α7 nAChR [129] and more recently on the model of the α7 nAChR [131], but the absence of the IC domain from the model may result in conformational instability in the absence of restraints [137] and also extended timescales (>> 20 ns) would be required to see any significant motions. It is important to note that at this stage there are several levels of error, due to the fact that the two domains are modelled on structures from two different species,

one is homopentameric, and the other a heteropentamer. Structural errors could have occurred during the homology modelling of the separate domains, especially because the TM domain is based on a relatively low resolution template. There may also be errors during the combining of the domains, as all possible criteria for the joining of two domains may not have been exhausted.

Given the possibility of error in the model, the value of performing molecular dynamics simulations at an atomistic level is questionable. Therefore, more 'coarse-grained' approaches were used, which omit atomistic detail but have been shown to predict overall patterns of mobility in a number of proteins (coarse-grained MD simulation of α7 nAChR and other LGICs will be discussed in chapter 7). The Gaussian network model (GNM) [164-166] was employed along with CONCOORD [171, 234]. Pore profile analysis using HOLE and electrostatics calculations using Adaptive Poisson-Boltzmann Solver (APBS) [182] were also carried out (see Methods chapter for details).

5.2.4 Electrostatic calculations

The Born energy of a Na^+ ion placed at successive points along the pore axis was estimated using Poisson-Boltzmann (PB) calculations as described in Beckstein and Sansom [62] (see Methods chapter). Sample points along the pore axis at which to place the ion were derived from the program HOLE [183] which also yielded the pore radius profile of the nAChR model. The energy minimized model of the α7 nAChR was prepared for the PB calculations: PDB2PQR [184] was then used to assign partial charges and radii for the atoms which generated a net charge of -45 *e* on the α7. To simulate the effects of the lipid bilayer, the nAChR structure was embedded in a low-dielectric slab with thickness 30 Å.

The PB calculation box with dimensions 96 × 96 × 210 $Å^3$, contained the slab-embedded nAChR structure (the pore being oriented along the *z*-axis) and one cation. The cation was placed at sample points, each 1 Å apart along the *z*-axis. For inserting the cation at a sample point with coordinate *z*, the PB potential of mean force (PMF) was calculated as: $\Delta G_{PB}(z) = G_{AChR+ION}(z) - G_{ION} - G_{AChR}$. The charge on the cation was +1 *e* with radius 1.68 Å, equivalent to the Born radius of sodium [235]. A dielectric constant of 2.0 was used for the nAChR and the membrane-mimetic slab and for the solvent (water) the value was 78.5. The effective radius for a water molecule, 1.4 Å, was used as the radius of a solvent probe sphere.

5.3 Results: using the α7 nAChR model for structural studies

5.3.1 Conformational sampling studies

5.3.1.1 Gaussian network models (GNM)

GNM allows the identification of flexible regions of proteins (see Methods chapter). It produces distance matrices and generates theoretical B-values [164-166]. The theoretical B values were put in the PDB file using a script in order to colour the structure based on the flexibility information derived from GNM. Visual Molecular Dynamics (VMD) [200] was used to visualize the structure, coloured by the theoretical B-values, and thus identified the flexibility of different regions of the protein. GNM was run on the full model as well as the separate domains. In Figure 5.4 (a), the GNM calculation for the homo-pentameric α7 model is mapped onto the structure of the model. It can be seen that there are number of "hotspots" i.e. regions for which the predicted mobility is higher than for the remainder of the structure. These correspond to two loops in the EC domain, those around residues E184 and

C212 (alignment numbering), and the upper (i.e. extracellular) ends of TM helices M1, M3 and especially M4 (see Figure 5.4 (b)).

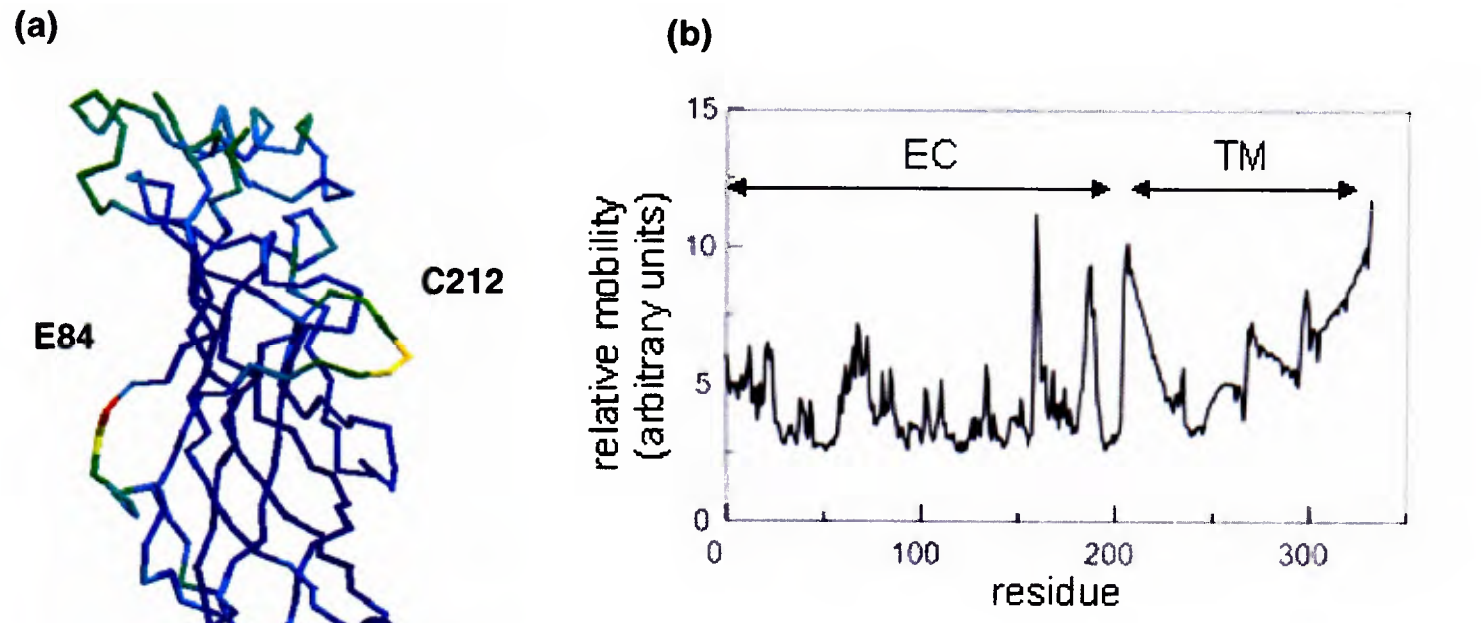


Figure 5.4 (a) One subunit of the α7 nAChR model coloured by the theoretical B-values generated by GNM. Blue = least mobility; red = the highest mobility regions. Loop C has great flexibility along with some other regions in the binding pocket. The top of the receptor is also relatively flexible and the TM domain is generally more flexible than the inner β-sheets of the EC domain. The β1-β2 loop of the EC domain comes close to the M2-M3 linker region of the TM domain through K68-D269 is also shown. (b) The B-values from GNM are shown, averaged along all 5 subunits. These values indicate the relative mobility of the corresponding residues. Figure modified from Amiri *et al.* [233].

The two high mobility loops in the EC domain correspond to two binding sites, one (residues 213-220) for the neurotransmitter acetylcholine and its analogues (this site is also a high affinity bungarotoxin, BgTx, binding site), and the other (residues 183-186) a low affinity BgTx binding site [28, 77, 221]. Loop C, which covers the binding pocket, has been the focus of many studies and is believed to 'cap' binding site upon ligand binding; it is held loosely in the absence of a ligand [130]. GNM has characterized this loop as highly mobile. Interestingly, the region of the polypeptide

linking the EC domain to the TM domain is predicted to be quite flexible. There is a gradient in the predicted mobility such that the EC ends of the TM helices are on average of higher mobility than the IC ends. Also, there is higher predicted mobility for the outer helices, especially M4, than the inner helices. This is perhaps expected as M4 stands farther apart than the other three helices and is not bound to the rest due to the missing residues in the IC domain. At the top segment of the TM domain, near the EC domain, the alpha helices are farther apart than the bottom section where the gate is situated, and the radius of the pore is small in this segment.

One should be careful not to over-interpret such results in terms of possible gating models for the nAChR. However, the gradient in mobilities of the TM helices suggests a model in which both inter- and intra-helical motions contribute to the gating of the channel. It is important to note that both atomistic simulations [137] and φ-value analysis of the kinetics of mutant nAChRs [69, 70] are consistent with a model where the upper half of the M2 helix moves relative to the lower (N-terminal) half. Another recent study suggests that rotations of the M2 are minimal [68]. It is interesting that the EC-M1 linker has higher predicted mobility whereas the mobility in the vicinity of the suggested contact between the β1-β2 loop of the EC domain and the M2-M3 loop has lower predicted mobility. The β1-β2 to M2-M3 linker contact in the α7 model seems to be mediated by a salt bridge (K68 to D289; see Figure 5.4). This may be an important part of the gating mechanism responsible for transmitting ligand-induced conformational change from the EC to the TM domain. The link between the two segments has recently been investigated experimentally [59] and is believed to be the main interaction between the EC and TM domains which leads to the gating of the channel.

5.3.1.2 CONCOORD

CONCOORD [171] uses distance restrictions to generate protein conformations around a given structure (see Methods chapter for details). CONCOORD was used to produce 500 structures, with the predicted model as a starting structure. Using DYNAMITE [178, 203], PCA is carried out on the CONCOORD structures, resulting in a set of eigenvectors which identify the more dominant motions of the protein [234]. Porcupine plots show the more significant motions of the protein and can identify the protein segments which move together. CONCOORD was run on each domain separately and also on the whole protein to study the changes that take place when the full receptor is analyzed rather than its separate domains.

Separate domains:

The eigenvector plot of the TM domain, shows an increase in the magnitude of motion moving towards the top of the domain, and the outer helices have more movement than the inner helices (Figure 5.5 (a)). Also, two subunits show greater movement, indicating a possible asymmetry in motion between subunits, as expected, since the TM domain was based on the heteropentameric TM domain from *Torpedo marmorata* [13]. There is also asymmetry between the subunits of the EC domain Figure 5.5 (b)). Although it is a homopentameric structure, some have eigenvectors of greater magnitude. This agrees with recent MD simulation studies where asymmetry was observed in the EC domain of the α7 nAChR, with greater movement in two of five subunits [129]. This is in agreement with atomistic MD studies of AChBP presented in chapter 3 where there is greater asymmetry in the ligand-free simulations.

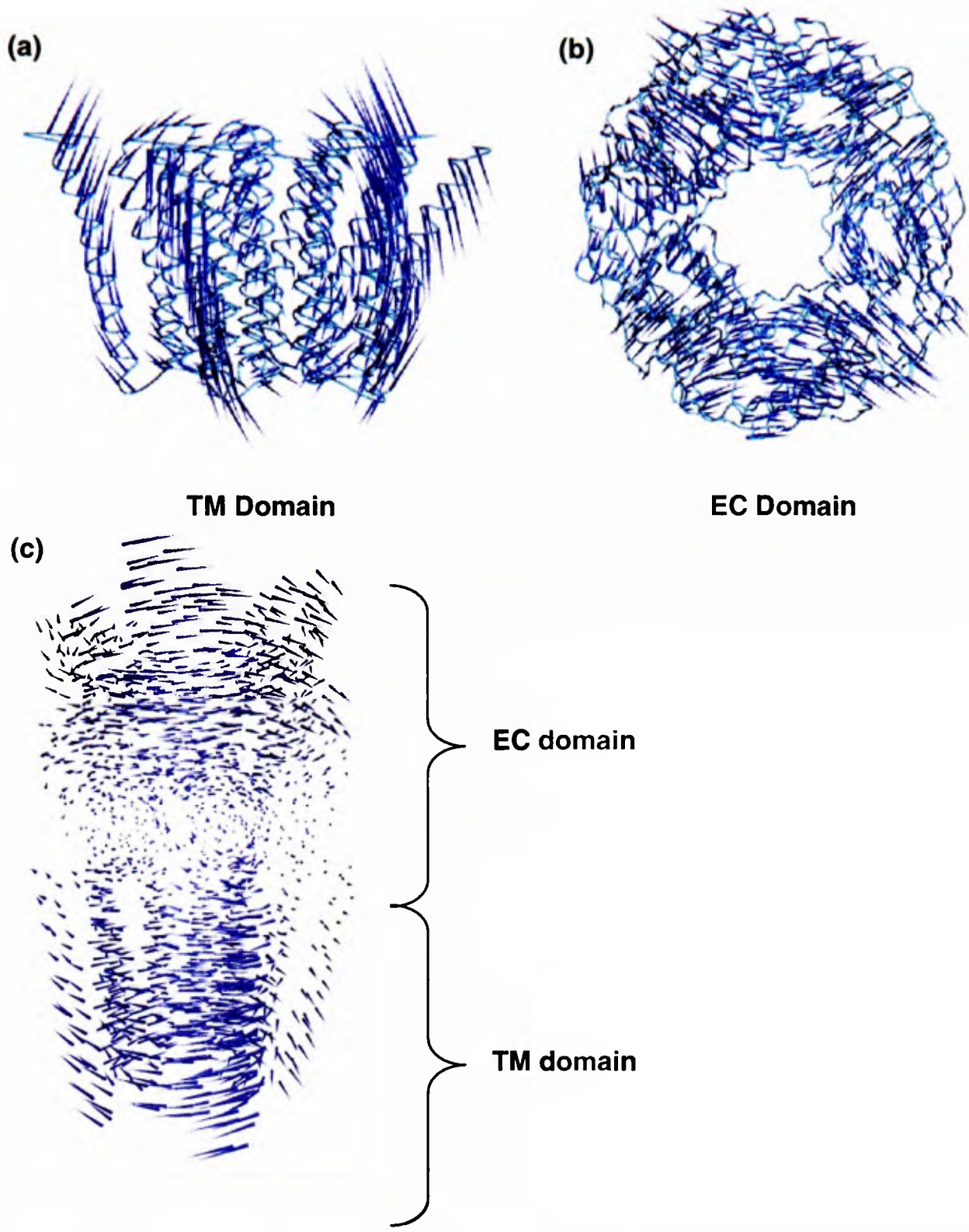


Figure 5.5 The porcupine plots have an x number of spikes, each spike representing the element of the eigenvector associated with each Cα atom of the protein; the larger the spike, the larger the magnitude of motion of the corresponding Cα atom. (a) α7 nAChR TM domain eigenvector plot (first eigenvector) showing greater freedom of motion for outer alpha helices and asymmetry of motion (side view). (b) The α7 nAChR EC domain eigenvector plot (first eigenvector) showing asymmetric motion of subunits and intra-subunit motion (top view). (c) Full nAChR eigenvector plot (first eigenvector), showing EC domain spikes rotating to one side and TM domain spikes rotating in the opposite direction. This rotation may be involved in the opening of the pore. The second eigenvector (not-shown) displays intra-subunit motion. A combination of these motions could result in the conformational changes involved in the gating mechanism.

Entire protein:

The CONCOORD results for the combined domain are significantly different than that of the separate domains. The most prominent motion identified (first

eigenvector), shows the rotation of the EC domain and TM domain in opposite directions (Figure 5.5 (c)). EC domain eigen-elements (spikes) all rotate to one direction while the TM domain eigen-elements rotate to the opposite direction, suggesting a twisting motion that may be involved in the opening and closing of the pore. This behaviour has been noted in other more recent studies [236, 237] as well as in the MD analysis of AChBP from chapter 3. The degree of rotation is less in the middle of the protein, where the two domains join, and increases as you move away from the centre of the protein. The second eigenvector (not shown) shows intra-subunit movements, where each subunit undergoes twisting about itself. The eigenvectors show the dominant motions of the receptor independently of each other, thus a probable motion can be a combination of the motions identified by the eigenvectors. There may be intra-subunit rotations of both the EC domain subunits and the TM domain subunits, accompanied by a full body rotation in opposite directions, which may result in the opening of the channel.

Covariance plots can also be generated using CONCOORD's output. They show which parts of the protein are closely related and could move together during the motion of the receptor. Many residues are covaried in each subunit and there are also strong links between subunits in the EC domain (Figure 5.6), which suggest that these parts could move together. There are numerous interactions between residues within each subunit in the EC domain (red areas), therefore there may be intra-subunit motions which are transmitted to the TM domain, as was also shown in the procupine plots. The EC domain subunits are closely linked at the top region of the domain, thus there may also be some degree of rigid-body rotation of the entire domain as a result of the intra-subunit motions, and also indicated by the eigenvector plots.

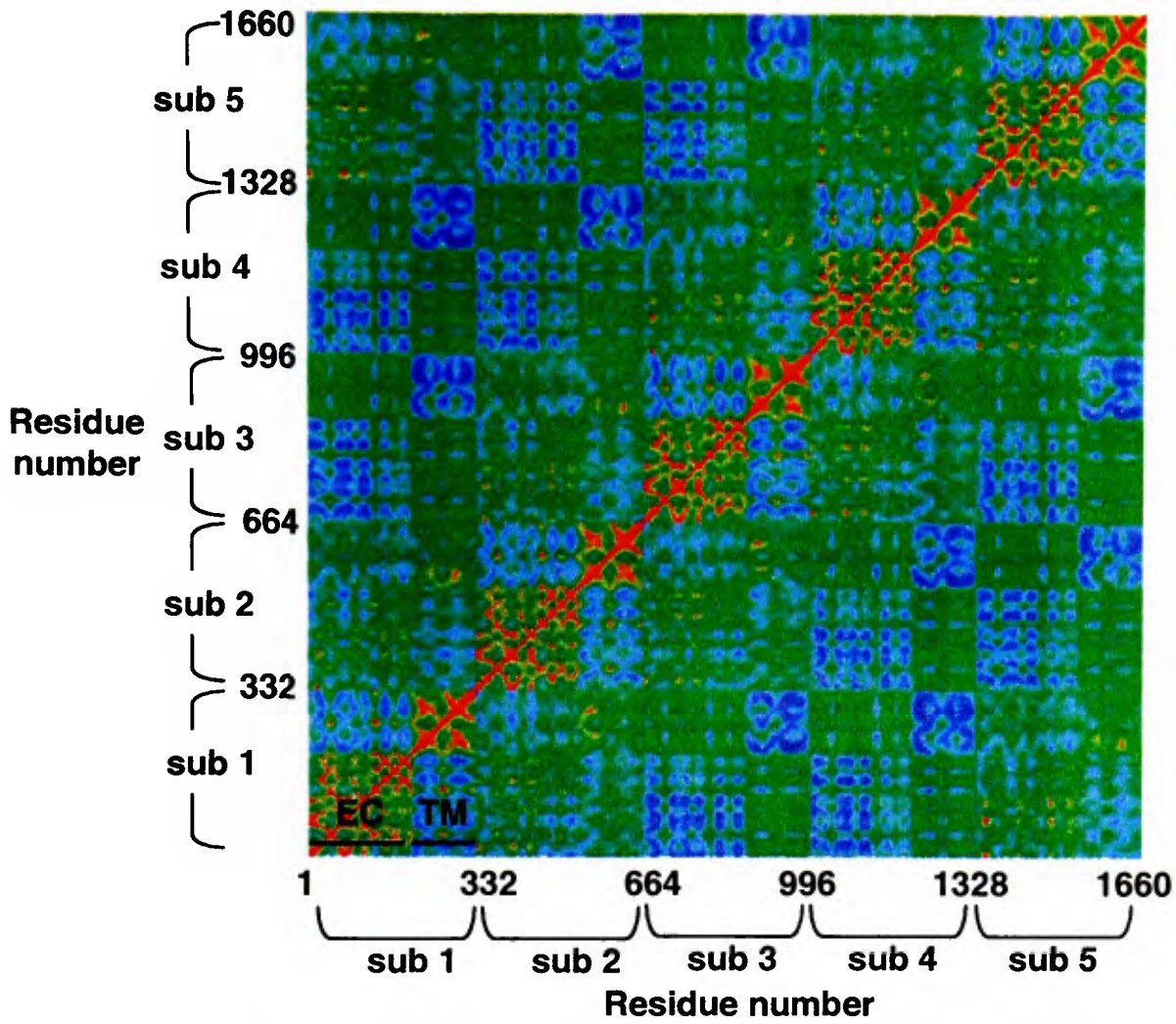


Figure 5.6 Covariance contour plot of the α7 nAChR model. This plot is generated using the 500 structures from CONCOORD. It shows which residues and subunits move together. The scores are normalized between -1 and 1. Red represents the highest covariance (1) (segments moving together), and blue represents negative covariance (-1) (segments moving away from each other). All five subunits move together and there is also intra-subunit motion, where residues in a subunit experience uniform motion. Each subunit is somewhat covaried with its adjacent subunit. There are red areas which illustrate how regions of the EC domain are heavily covaried with TM segments (red dots in blue TM regions).

5.3.2 Electrostatics, energetics, and pore-profiles

5.3.2.1 Pore properties and Poisson-Boltzmann calculations (work carried out with Kaihsu Tai)

The properties of the pore through the centre of the α7 nAChR model have been studied. This model corresponds to a *closed* state of the channel: however, the pore properties are still informative. The pore radius profile was analysed using the program HOLE [183]. A number of constrictions along the length of the pore, which spans from $z \sim -25$ Å (at the intracellular end) to $z \sim +80$ Å at the extracellular mouth are revealed from the resultant pore-lining surface (Figure 5.7 (a)) and pore radius profile (Figure 5.7 (c)). The radius drops to ~2 Å in the vicinity of the ring of Thr

residues at 6′ and to ~3 Å in the hypothesized location of the gate of the nAChR at Leu 9′ and Val 13′. These make up the main constriction of the pore in the TM domain. On the M2 of each subunit, the latter region thus consists of two conserved rings of hydrophobic residues. In the EC region at $z \sim +45$ Å, there is a less pronounced constriction (radius ~5 Å). It is useful to recall that the radius of a solvated Na^+ ion, allowing for just a single solvation shell, is ~4 Å.

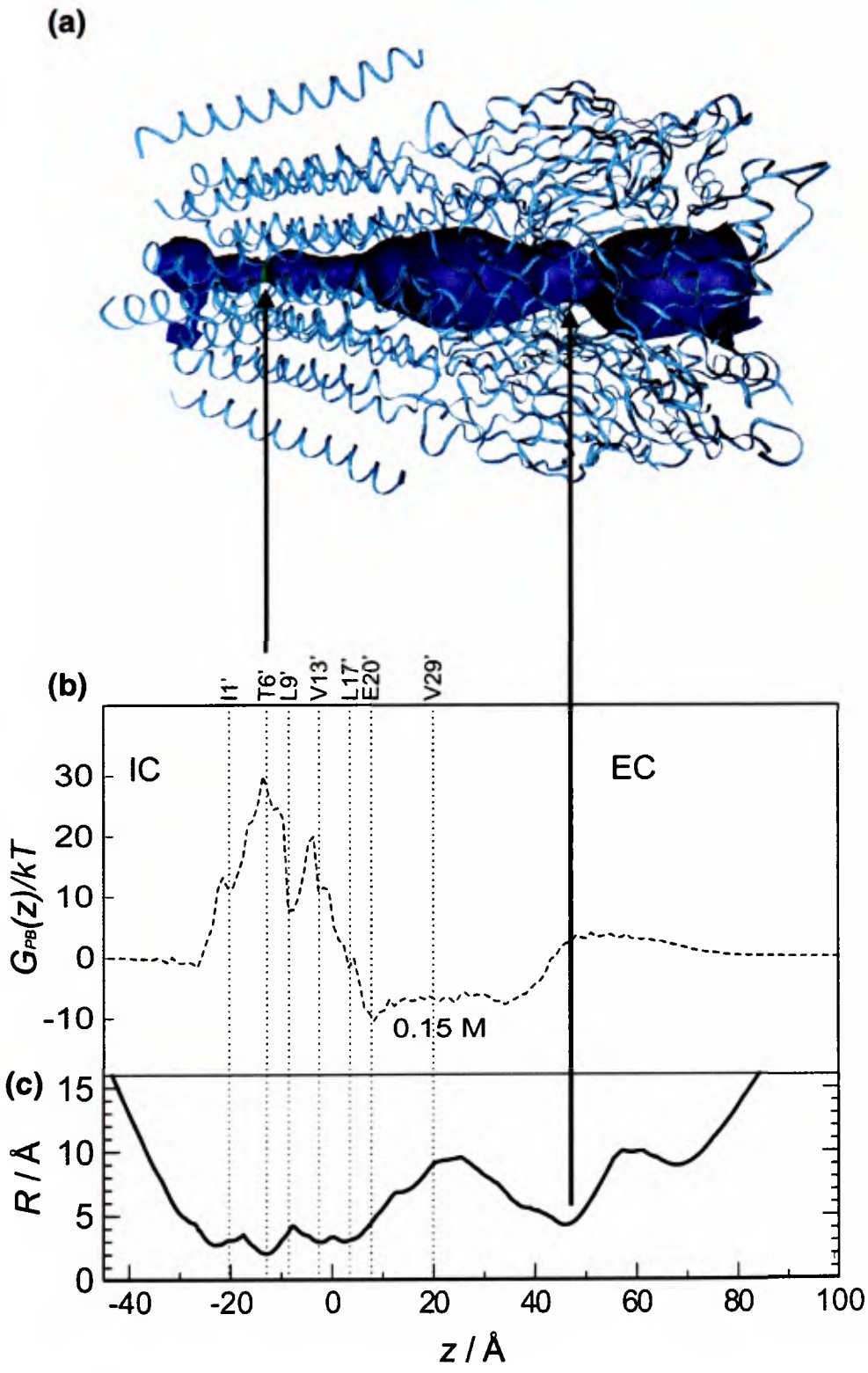


Figure 5.7 (a) Model of the α7 nAChR rotated so that the pore (z-axis) coincides with the graphs in (b) and (c). The pore lining surface in blue is calculated using HOLE [183]. The most constricted region (TM domain) is in green and corresponds to the position T6' and the highest PB energy. (b) The PB profile of a monovalent cation (Na^+) through the intact α7 nAChR model (black broken line) at 0.15 M. (c) Pore radius profile calculated for the intact α7 nAChR model using HOLE. Figure modified from Amiri *et al.* [233].

In addition to the pore radius profile, the energetics of a Na^+ ion as it is translated along the pore axis can be estimated. The TM domain analysis is covered in detail in

Amiri *et al.* [233] which includes PMF calculations and umbrella sampling methods (see Methods chapter); although computationally expensive, these methods could also be extended to the EC domain in future work. The Poisson-Boltzmann (PB) equation was used to estimate the Born energy (see Methods chapter) of a Na^+ ion along the length of the whole channel (broken line in Figure 5.7 (b)). The PB energy calculations were for an ionic strength corresponding to 0.15M (physiological ionic strength). There is a pronounced barrier to ion permeation in the constricted region of the M2 helix bundle with a barrier height of > 25 kT which is considerably greater than that for the PMFs of the atomistic simulations [233]. When the PB energy profile calculation was repeated for just the M2 helix bundle of the α7 model, this difference was less pronounced and it results in a barrier height of ~15 kT. This value is still greater than that from the atomistic simulations (~10 kT) which may be a consequence of intrinsic overestimation of barrier heights for narrow (< 3 Å) channels by PB calculations vs. full atomistic PMFs [62].

An upper limit (g_{MAX}) can be obtained on the conductance of the *closed* state of the α7 channel based on the free energy profile for the M2 bundle using a method developed for the gramicidin channel [238]. This yields an estimate of g_{MAX} = 0.02 pS which is well below the detection limit of patch clamp measurements of single channel currents, and thus would correspond to a closed state of the channel, which is in accordance to the proposed state of the model.

Figure 5.7 shows a broad energy well (depth ca. -2 kT) for Na^+ in the lower part of the vestibule (from z ~ +10 Å to +40 Å). However to examine the possible role of the vestibule in more detail, a range of conditions must be considered.

The vestibule generates a significant well for Na^+ ions at a physiological ionic strength (0.15 M), and thus may be expected to play a role in the overall ion

selectivity of the α7 nAChR. To confirm this, the PB energy profile at 0.15 M of a Cl^- ion (grey broken line in Figure 5.8) was calculated. This reveals a considerable (~ 20 *kT*) barrier to anion permeation presented by the vestibule. Thus, the calculations suggest that the extracellular vestibule *may* contribute to the ion selectivity of the channel [239]. This question is further investigated with Born profile calculations of full LGIC models in chapter 7.

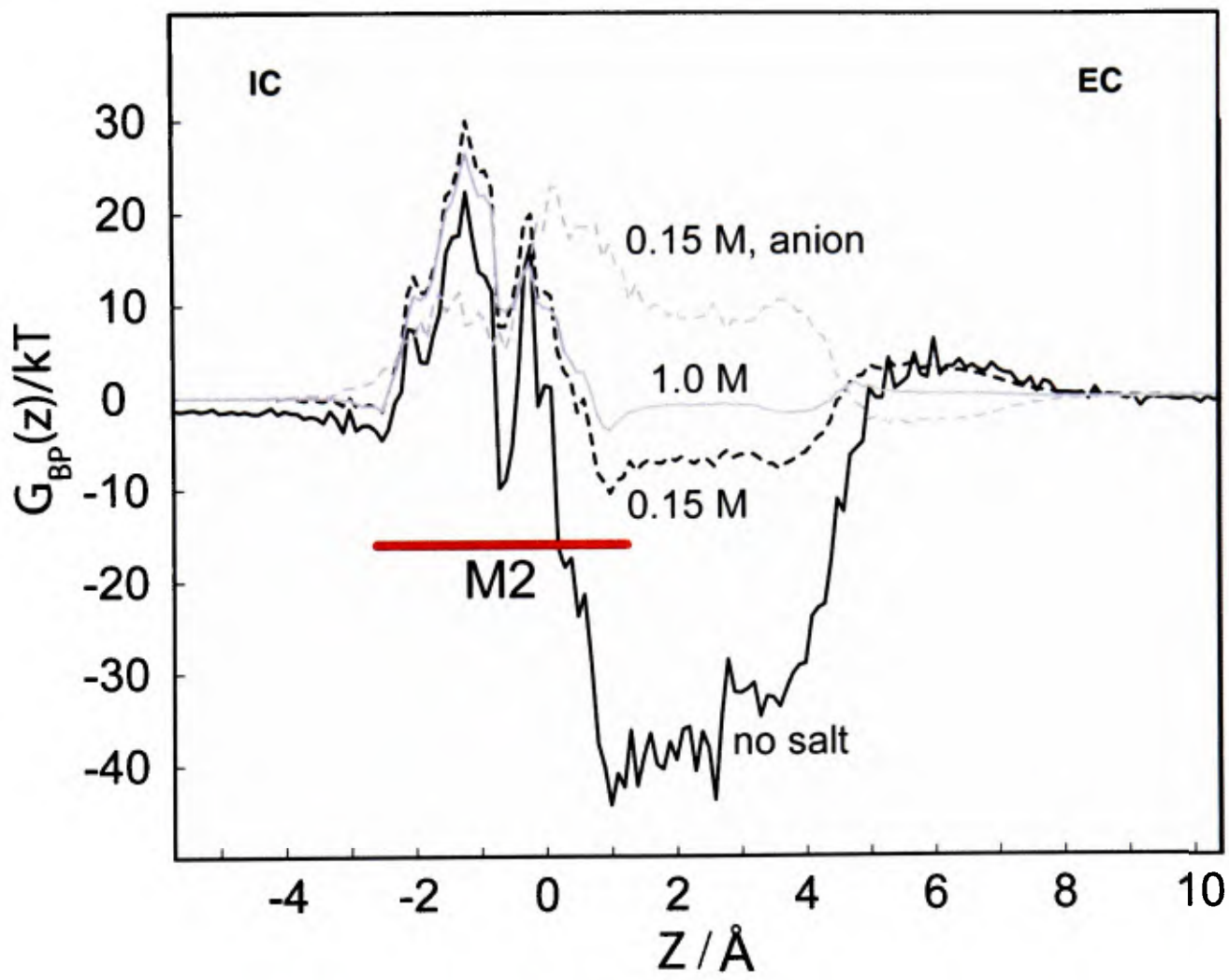


Figure 5.8 The Poisson-Boltzmann (PB) profile of Na+ through the α7 nAChR channel at several ionic strengths. Solid black line = zero ionic strength, broken black line = 0.15 M NaCl, solid grey line = 1.0 M NaCl. The PB profile of Cl- is also shown in the presence of 0.15 M NaCl (broken grey line). Figure modified from Amiri *et al.*, 2005.

5.3.3 Comparison of the α7 nAChR model with the 4 Å *Torpedo marmorata* structure

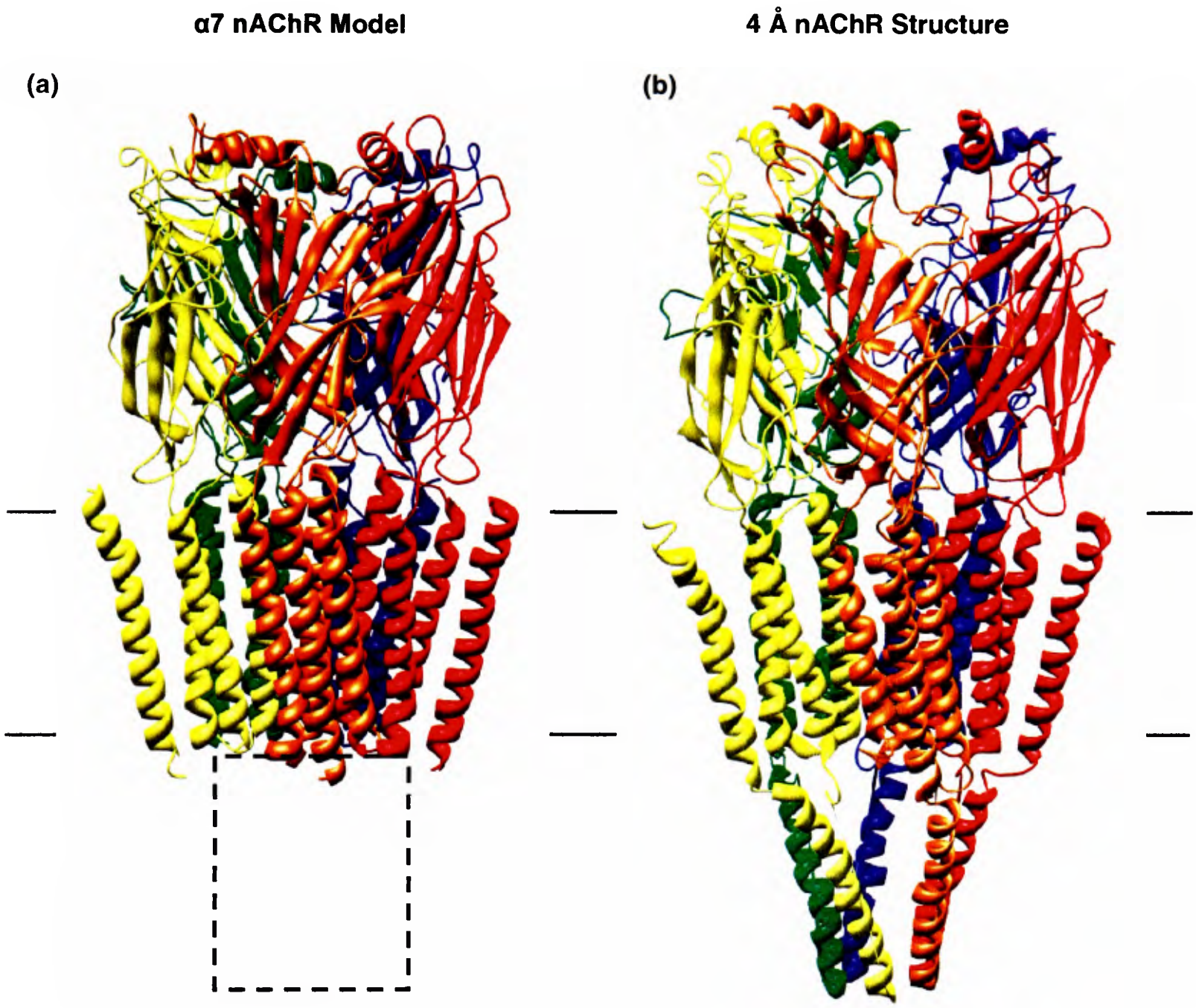
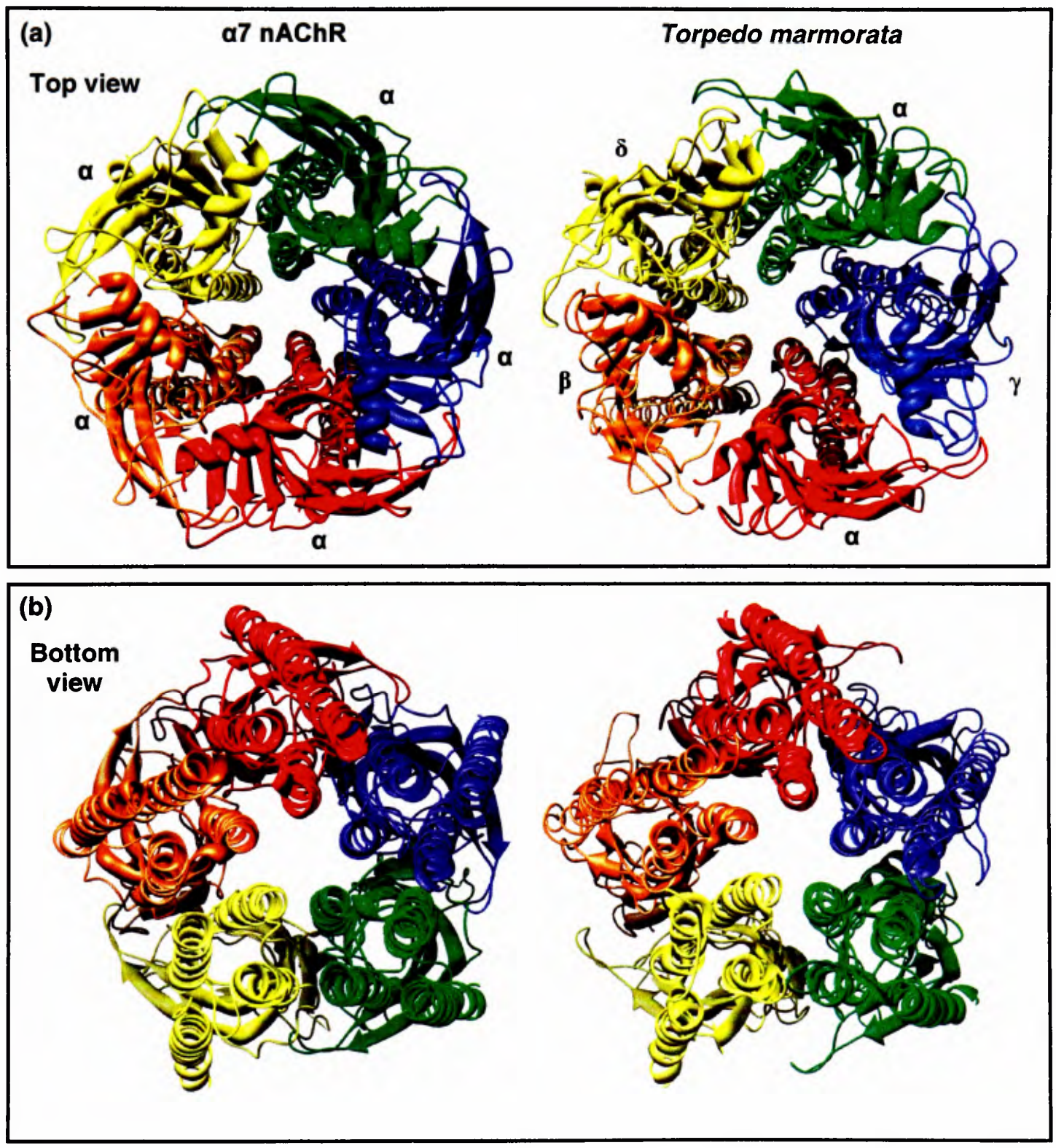


Figure 5.9 (a) The model of the α7 nAChR generated by the method described in this chapter. The structure is the combination of a homology model based on the AChBP structure (EC domain) [41, 77] and the *Torpedo marmorata* [13] TM domain structure. The IC domain is missing for the α7 nAChR model, indicated by the dashed rectangle.

Recently, a full structure of the *Torpedo marmorata* nAChR at 4 Å has been made available [19], though with missing residues in the intracellular (IC) domain (and a few other smaller segments). This is a heteropentameric nAChR from the neuromuscular junctions of the *Torpedo marmorata* (Figure 5.9 (b)). The structure has been used to compare with the α7 nAChR generated with the method described in

this chapter (Figure 5.9 (a)), although a comparison of whole structures is not completely valid as they are different in their subunit composition, one being heteromeric, and other homomeric. Comparison of various sections of the α7 nAChR and the *Torpedo marmorata* structure could allow us to observe the structural differences and similarities between the two channels and verify certain structural qualities such as the inter-domain interface.



Fignre 5.10 (a) The top view and (b) bottom view of the α7 nAChR model (left) compared to the 4 Å *Torpedo marmorata* structure (right).

The 4 Å nAChR structure is a heteropentameric ion channel (Figure 5.9 (b)), formed from a ring of homologous subunits: α, γ, α, β, δ, arranged around a central axis. The model of the α7 nAChR (Figure 5.9 (a)) is a homopentameric channel composed of all α7 subunits. The EC domain is modelled on the homopentameric AChBP while the TM is modelled on the heteropentameric *Torpedo marmorata* nAChR. Calculating a Cα RMSD value of the two whole structures would not be very helpful as there are differences in the number of residues, and the subunits of the *Torpedo marmorata* are not identical as they are in the α7 nAChR, so undoubtedly there will be substantial differences in the two structures. However, what is of importance is the structure of the important regions of the proteins. Also, the 4 Å nAChR is of relatively low resolution and it is missing some structural data that is available in the EC domain of this model of the α7 nAChR because it was modelled on a high resolution crystal structure.

Profit [240] (Martin, http://www.bioinf.org.uk/software/profit/) was used to fit the Cα atoms of residues from the two structures in order to obtain their RMSD values. Several sections were chosen for an RMSD fit from an α-subunit of *Torpedo marmorata* and one of the α-subunits of the α7 nAChR. There are two tables below (Table 5.1 and Table 5.2), one showing the comparison between α-subunits of the different proteins and one comparing an α-subunit of the α7 nAChR with a β-subunit of the *Torpedo marmorata* structure. α-subunits are the major participant in ligand binding and they have the highly conserved adjacent cysteines on loop C which are believed to be functionally important, especially in ligand binding. One expects that the α-subunits of these two different proteins should be structurally more similar than the comparison between an α-subunit and a β-subunit.

Table 5.1 RMSD of different regions comparing the α-subunits of the α7 nAChR model with an α-subunit of the full *Torpedo marmorata* structure.

Region	α7 nAChR numbering	Torpedo marmorata numbering	RMSD (Å)
Top of EC domain	1 - 157	2 to 158	3.8
Loop C Region (middle of EC domain)	158 - 192	161 - 195	2.6
Termini region of EC domain	193 - 207	197 - 211	1.7
Loop β1-β2	40 - 50	42 - 52	0.4
M2-M3 linker	264 - 268	269 - 273	0.7
CYS loop	126 - 140	128 - 142	3.1
Loop C	182 - 196	186 - 200	1.4
M2	236 - 264	241 - 269	0.6
EC + TM domain (figure)	1 - 331	1-437 (extra residues taken out)	3.7

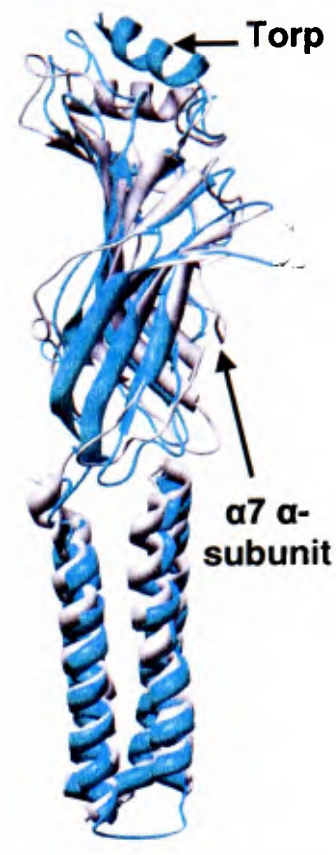
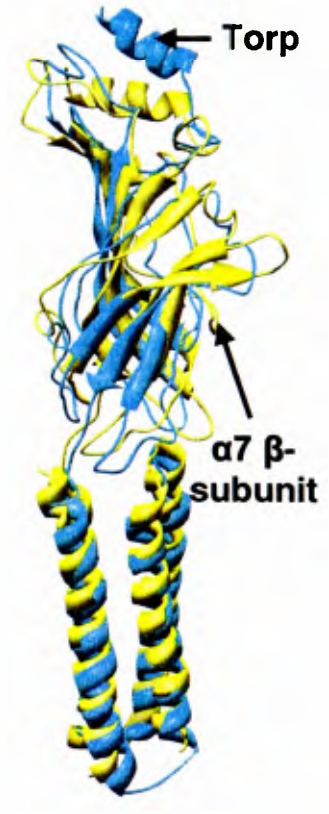


Table 5.2 RMSD of different regions comparing the α-subunit of the α7 nAChR model with a β-subunit of the *Torpedo marmorata* structure.

Region	α7 nAChR numbering	Torpedo marmorata numbering	RMSD (Å)
Top of EC domain	333 - 490	343 - 500	3.7
Loop C Region (middle of EC domain)	491 - 525	503 - 537	3.5
Termini region of EC domain	526 - 540	536 - 550	1.8
Loop β1-β2	372 - 382	382 - 392	0.7
M2-M3 linker	596 - 600	606 - 610	1.6
CYS loop	458 - 472	468 - 482	3.0
Loop C	514 - 528	527 - 541	4.0
M2	568 - 596	578 - 606	0.7
EC + TM domain (figure)	1 - 331	1-437 (extra residues taken out)	3.4



As expected, the top regions of the EC domains have a higher RMS value when fitted together. The small α-helix on the top of each subunit is on a slant in the *Torpedo marmorata* and flat in the AChBP structure. Also the differences between the homologous subunits of the *Torpedo marmorata* structure and the identical subunits

of the α7 nAChR generate a relatively high RMSD value when fit together. The EC domain is split into three sections and then the functionally important regions are also fitted separately.

The residues near and on loop C from both structures are more similar structurally. This loop is believed to have great functional importance. Recent studies show that loop C covers a bound ligand and is more flexible in the absence of one [130], also discussed in chapter 3. The small differences could be the result of loop C being in different states at the time of capture. The α7 nAChR is modelled on the AChBP which had HEPES bound at the time of crystallization, but the *Torpedo marmorata* images were taken without ligands bound so it is expected that their loop Cs are slightly different.

Interestingly, the termini regions have the smallest RMSD values. Lining up the EC and TM domains was the main challenge in generating the model and thus having the termini regions agree with that of the *Torpedo marmorata* structure is encouraging. The β1-β2 loop is believed to come into close proximity of the M2-M3 linker region of the TM domain. The interaction between the two segments may transmit the motion of the EC domain to the TM domain of the receptor [59] and has a low RMSD in the α7 model, indicating that this region has been accurately modelled.

For further comparison, RMSD fits were carried out between the β-subunit of the *Torpedo marmorata* structure and one of the α-subunits of the α7 nAChR. In this case, most of the RMSD values are greater than the α/α comparison, this may be attributed to the lower sequence identity. The top regions of the EC domain have slightly less RMSD values than the α/α fit, however, all other sections, including the functionally important loops, helices, and sheets, have higher RMSD values in the α/β fits.

The β1-β2 loop and M2-M3 linker have a much lower RMSD value for the α/α fit than the α/β fit. The α-subunits are where the ligands bind in the heteromeric structure, so it is interesting that in the α-subunit fits these important sections have higher structural similarity. The Cys loop has a relatively large RMSD value. This could be partly due to the loop's position on the outside of the receptor which gives it greater freedom of motion. Also, this loop in AChBP, which was the template structure for the EC domain, could be in a slightly different state because of the bound ligand in that structure. The M2 helix of the α/α fit has a small RMSD value as expected. This helix lines the pore of the ion channel and a small rotation of the M2 helix is thought to have a role in opening the gate and allowing for the passage of ions [35, 49, 61, 64, 68, 241, 242]. Again, the small RMSD value is encouraging as it validates the model by showing that the functionally important regions are very close to that of the full structure of the *Torpedo marmorata.*

A model of the α7 nAChR structure was generated using the 4 Å nAChR *Torpedo marmorata* structure [19]. The chicken α7 sequence was used to make the homology model and to provide another avenue of comparison to the model generated from two separate structures. The new α7 TM domain is very similar to the generated α7 model, with greater symmetry than the *Torpedo marmorata* structure's TM domain. The TM domain used to generate the original α7 model was based on the same *Torpedo marmorata* TM structure [13] (did not include the EC domain), thus the similarity is not surprising. The EC domain's β-sheets are not fully formed, with more loop regions than expected; this is because the template structure is of low resolution and the β-sheets are not as well defined as they are in higher resolution structures such as the AChBP [77].

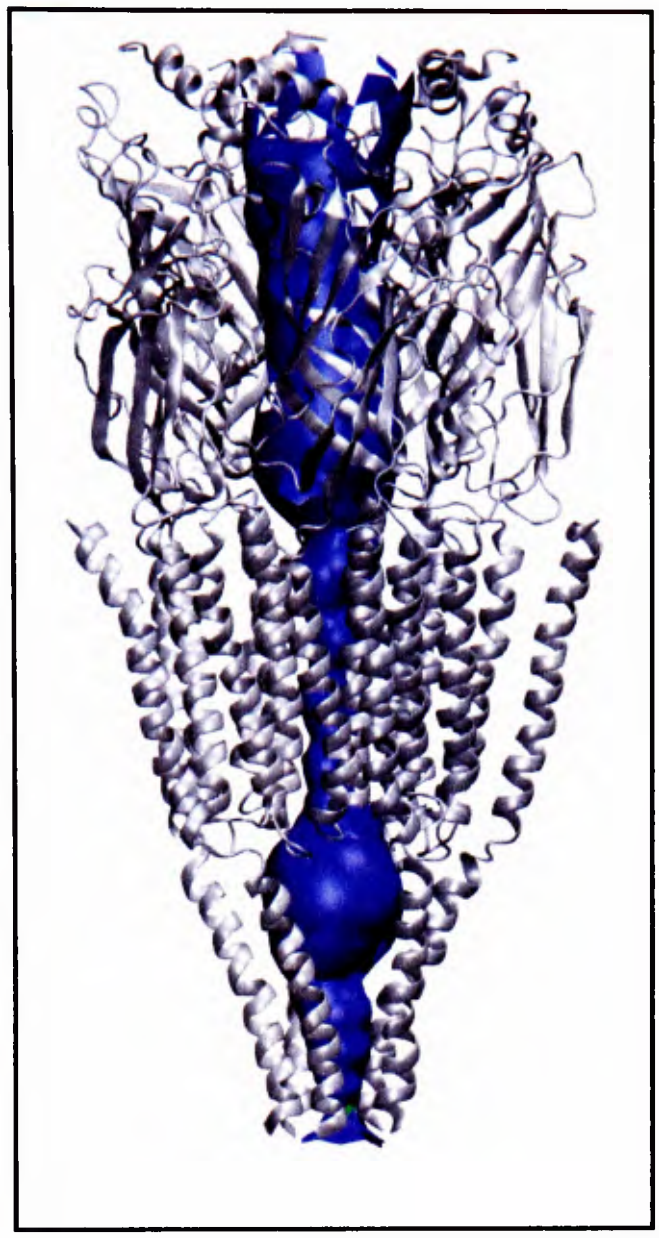


Figure 5.11 The α7 nAChR structure based on the 4 Å *Torpedo marmorata* nAChR structure [19]. The pore profile is shown and the biggest constriction lies in the IC domain.

5.4 Discussion

Although the constructed model of the α7 nAChR lacks the intracellular domain, calculations based on this model provide valuable insights into the structural basis of nAChR function. Also this model's EC domain is based on a higher resolution crystal structure, thus making it more accurate for carrying out structural studies on this domain. Note that it is believed that the AChBP structure corresponds to the desensitized state of the nAChR EC domain, however, this remains uncertain [15, 128]. The EM structure of the TM domain was obtained in the absence of agonist (acetylcholine), this corresponds to a closed state structure. Thus, the α7 model corresponds most closely to the receptor in a closed conformation.

5.4.1 Conformational sampling

The GNM results reveal several regions of elevated flexibility. The more flexible EC loops (183-186 and 213-220) correspond to those where ligands bind. This is suggestive of a model in which ligand binding may lead to a modulation of a conformational equilibrium [243] as a component of the conformational transition leading to channel opening. The GNM results also indicate that there is a flexibility gradient along the length of the TM domain, down the pore axis. This suggests that the extracellular half of the transmembrane domain may move relative to the intracellular half. From analysis of the effects of point mutations on the gating kinetics of the nAChR, a conformational wave that is propagated along the length of the molecule from the EC domain to the TM domain has been suggested [70, 72]. This agrees with the TM mobility gradient seen in the GNM results, if the wave is propagated along the TM helices from the EC to the IC end of the domain.

5.4.2 Electrostatics

A significant energetic barrier in the centre of the M2 helix bundle is revealed from PB energy calculations (atomistic MD based PMF calculations are discussed in Amiri *et. al.*, 2005). This is consistent with a 'hydrophobic gate' model, in which a relatively narrow hydrophobic region in the centre of the pore forms the barrier to ion permeation in a closed channel. As shown by calculations in model systems, a modest increase in radius and/or polarity of this region could open such a gate [61, 62, 244]. Free energy profile calculations on the TM domain of the *Torpedo* nAChR suggest a similar hydrophobic gate region in the centre of the pore [245] which is also consistent with the structural (EM) data [13, 60, 226] and with the single channel kinetic analysis of mutational data [70]. On the basis of accessibility studies of

cysteine mutations, it has been argued that the gate is at the intracellular end of the transmembrane pore rather than in the centre [33, 45, 65, 66]. This does not readily agree with the current calculations. However, it is important to note that application of a similar cysteine accessibility procedure to the $5HT_3$ receptor [63] has yielded results consistent with a gate located midway along M2, consistent with this study.

Other models of intact nAChR have been reported [139]. This study is the first to: (a) present an explicit and objective protocol for combining the EC and TM domains in one model; (b) provide a comparison of continuum electrostatic and atomistic (i.e. MD based) free energy analyses of the barrier(s) to cation permeation in a closed state nAChR; and (c) employ coarse-grain dynamics approaches to study the plausible motions of the full receptor.

It is important to critically assess the possible limitations of the approaches employed in this study. The main limitation is that the TM domain model is based upon relatively low resolution structural data. However, there is little one may do to remedy this as there is an absence of higher resolution structural data. In terms of the GNM and CONCOORD calculations of residue mobility, it should be remembered that the results are suggestive of motions that may be related to gating, but do not provide a detailed atomic resolution mechanism for channel gating. Encouragingly, the twisting motion observed with CONCOORD is supported by recent computational studies of nAChR gating [236, 237]. Also the observed twisting motion fits in with the recent study suggesting a reorientation and rigid body movement of M2 helix by the cis-trans isomerisation of a proline residue on the M2-M3 linker, opens the channel pore of the $5HT_3$ receptor [59].

Future computational studies are likely to benefit from improved structural data on which to base models, especially for the IC domain with its potential effects on ion

selectivity [44]. A model including the IC domain would also allow unrestrained MD simulations of the full structure of a nAChR in a lipid bilayer, which at present are rather challenging due to the likely effect of the absence of an IC domain model on structural stability during such simulations. Furthermore, once such a model is obtained, it will be useful to develop an open state model by using computational studies to integrate (low resolution) structural data and using indirect (e.g. mutational data) in an objective fashion. Such a model could be used to probe, for example, the high permeability of α7 nAChR to Ca^{2+} ions, relative to other nAChR species.

Comparison of the 4 Å nAChR [19] to the generated model has been encouraging. The key regions, which have been shown experimentally to play important roles in the gating mechanism, exhibit small RMSD values when fitted together. The 4 Å nAChR was used to generate a model of the full α7 nAChR. However, as before, this structure is of low resolution, compared with the AChBP structure which was used to model the EC domain of the α7 nAChR model.

The construction of α7 nAChR from the combination of its TM and EC domains was reported in this chapter. The model of the α7 nAChR was used to carry out various structural studies to explore the conformational flexibility, plausible modes of global motion, electrostatics, and pore characteristics. This method can be used to combine domains in other protein and protein complexes with symmetry between the domains, in particular of membrane proteins where there is a clear lack of structural data.

6 A computational study of α7 nAChRs: directing experiment

6.1 Introduction

Due to the sequence homology between AChBP and the extracellular (EC) domain of neuronal α7 nAChR, homology modelling, followed by ligand docking and MD, may provide a useful tool to investigate ligand binding. An α7 nAChR EC model and four α7 mutant EC models were generated based on the 2.2 Å AChBP structure with nicotine bound (PDB code 1UW6) [77]. Five corresponding simulations of the α7 nAChR (one wild-type and four mutants) were used to carry out MD ensemble docking studies (as described in chapter 4) with several ligands of pharmacological interest. Predictions from this study were tested and verified by experimental work with our collaborators Dr. David Sattelle in Human Anatomy and Genetics, University of Oxford. Electrophysiological studies, carried out by Shimomura and others in collaboration with Dr. Sattelle and colleagues, confirm the *in silico* acetylcholine binding predictions. Furthermore, these studies show that residue L118

from α7 nAChR is also a major determinant for the behaviour of the insecticide Imidacloprid (IMI) and its derivative Desnitro-Imidicloprid (DN-IMI). DN-IMI is a derivative of Imidacloprid lacking the nitro group (see Figure 6.2). The guanidine moiety of this compound is protonated at neutral pH.

The aim was to use computational methods to identify key residues involved in ligand binding and study the binding modes for each of the ligands. Also the intention was to better understand ligand-protein interactions from a structural perspective and study whether docking on MD trajectories improves or hinders the binding of ligands as the protein goes through structural changes. The functions of wild-type, as well as many mutants of the α7 nAChR, have been well characterized by electrophysiology in heterologous expression systems [109, 123]. This computational study shows how *in silico* methods can help identify a residue in α7 which plays an important role in the agonist-binding site and also affects the dynamics of the receptor.

Another homologue of nAChR, the ACR-16 (acetylcholine receptor) [124], which closely resembles the α7 nAChR [124], and three other α7 nAChR-like receptors: UNC63, UNC29, and LEV1 [246-248] are modelled and used for docking studies. Electrophysiological studies on the neuromuscular junction of *Caenorhabditis elegans* have shown that there are at least two nAChR subtypes present; one which is sensitive to the anthelmintic, Levamisole, while the other is unaffected [249]. Levamisole is used for eradicating nematode infestations in domestic animals [250]. It causes hypercontraction, paralysis, and death to wild-type *C. elegans* [246]. It is believed to be a full agonist of nematode muscle nAChR however experimental work shows that it does not show the same efficacy on mammalian muscle nAChR [251]. Recently, *ACR-16* has been shown to contribute to the Levamisole-insensitive receptor subtype

[252, 253] which may represent a novel target for parasitic nematode control [119] while UNC63, UNC29, and LEV1 are sensitive to the drug Levamisole.

Unc63, unc-29, and lev-1 are the genes which encode nicotinic acetylcholine receptor subunits in *C. elegans* [246]. UNC63 is a α-type subunit (possessing two adjacent cysteines) while UNC29 and LEV1 are non-α type [246]. Levamisole-sensitive receptors in the body wall muscle of nematodes require the functional expression of UNC63, UNC29, and LEV1 [246].

Here, computational modelling, MD, and ligand docking studies complement and direct experimental work on α7 nAChRs.

6.2 Methods

6.2.1 Homology modelling and *in silico* mutagenesis

For the chicken α7 nAChR and ACR-16, homology models were made using AChBP (PDB code 1UW6, which has nicotine bound) from *Lymnaea stagnalis* [77]. A pair wise sequence alignment was generated with MultAlin [144] with ~27 % sequence identity (Figure 6.1).

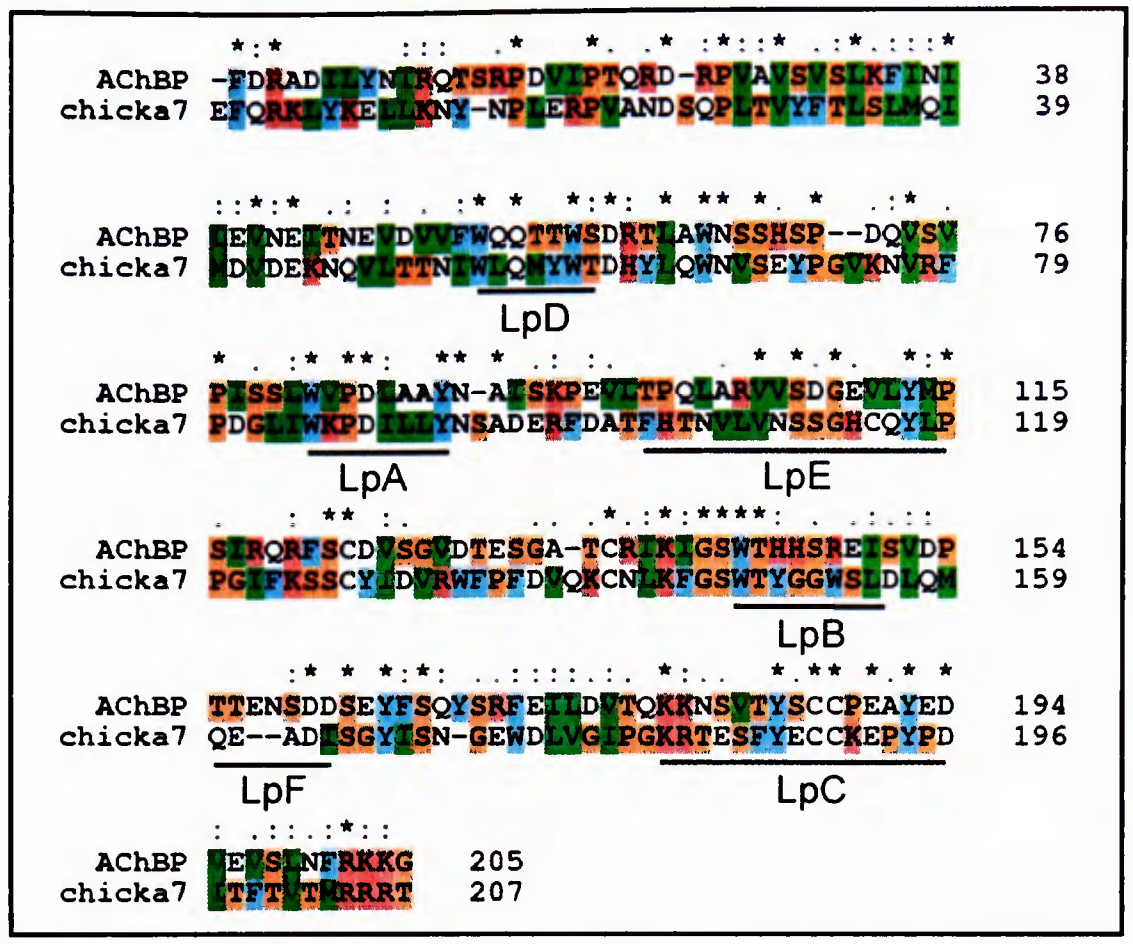


Figure 6.1 The sequence alignment used to make the chick α7 nAChR using AChBP as template. Loops A-F are labelled on the alignment. The colour scheme is as follows: blue - aromatic residues, green - hydrophobic residue, orange - polar residues, red - basic residues. Figure taken from Amiri *et al.*, Mol. Pharm., 2006 - submitted.

The alignments were used as input for the program MODELLER [142] to generate 100 pentameric models of the ligand binding domain. The quality of the 5 lowest energy structures was checked with the PROCHECKv3.5.4 program [186]. The model with the lowest percentage of residues in the disallowed region of the Ramachandran plot was selected for docking studies. Only one binding site from 2 adjacent subunits was used. *In silico* site-directed mutagenesis of the α7 nAChR model was performed using PyMOL [254, 255] to generate four mutant models. PyMOL performs energy minimization for the mutation and further energy minimization was carried out on each model using GROMACS v 3.1.4 (www.gromacs.org) [186-188] until convergence with the steepest descent method as described in chapter 2.

6.2.2 Docking

The α7 nAChR homology model and the four α7 nAChR mutant models: α7 L118D, α7 L118E, α7 L118K, and α7 L118R, were used to dock nicotine, acetylcholine (ACh), Imidacloprid (IMI) and desnitro(DN)-Imidacloprid (DN-IMI). The ACR-16, and combinations of UNC63, UNC29, and LEV1 subunits were used as docking targets for nicotine, ACh, and Levamisole. All of the docking studies were carried out with the program Autodock [179] (see Methods chapter). Autodock was run with default parameters except for custom atom types for the ligand and with charges calculated as below. The charges on ACh were assigned according to Segall *et al.* [225]. Charges for the other ligands were calculated with the 6-31G* basis set using Spartan (Wavefunction Inc. [180]) (charges are available on the supplementary CD). All docking results were visualized with VMD [200] and UCSF Chimera [201].

6.2.3 Molecular dynamics simulations

For docking studies to be carried out on the MD trajectories of α7 nAChR model and its 4 mutant models, five 10 ns simulations were run, one for each model (see Table 6.1). The simulation names in Table 6.1 will be used for the remainder of this chapter. The methodology for these simulations follows those described in chapter 3. All of these simulations were run without a ligand.

Table 6.1 Simulations of α7 nAChR and four mutant simulations of α7 nAChR.

Simulation name	Number of atoms	Approximate box size (nm)	Duration (ns)
α7	65169	9.2 x 9.1 x 8.4	10
L118D	67037	9.3 x 9.4 x 8.3	10
L118E	67039	9.3 x 9.4 x 8.3	10
L118K	67095	9.3 x 9.4 x 8.3	10
L118R	66464	9.5 x 9.5 x 8.0	10

6.2.4 Automated docking along an MD trajectory

Ligand docking onto the α7 nAChR and mutants was studied by the MD simulation ensemble docking approach described in chapter 4 (4.2.3). Successive frames of each trajectory were used to dock ligands back into the structure. 100 frames were used from each trajectory, 1 every 100 ps for a 10 ns simulation. An analysis script records the top dock (lowest ranked energy) and also the number of docks out of 50 in the top ranked cluster for each chosen frame of the simulation. (Note: the docking energies recorded can only be used comparatively between different docks as the values determined by the software are arbitrary and scaled).

6.2.5 Electrophysiology on recombinant wild type and mutant α7 receptors*

Experimental studies on the wild-type and mutant α7 nACh receptors were carried out by Dr. D. Sattelle and colleagues. *Xenopus laevis* oocytes were prepared as described in Shimomura *et al.* [256]. Dose-response data were obtained by challenging oocytes with increasing concentrations of agonist. Details of the experimental work can be found in Amiri *et al.,* Mol Pharm 2006 - submitted.

*This experimental work is carried out by M. Shimomura, M. Akamatsu, and K. Matsuda from the Graduate School of Agriculture, Kyoto University, in collaboration with D. Sattelle and A. Jones from the Department of Physiology, Anatomy and Genetics, University of Oxford.

6.3 Results

6.3.1 α7 nAChR and α7 nAChR mutants

The sequence alignment used to model the chicken α7 homopentameric $(\alpha 7)_5$ agonist-binding domain using the structure of the acetylcholine binding protein (AChBP) from *Lymnaea stagnalis,* is shown in Figure 6.1. The homology models and a preliminary docking study suggest that position L118 (Figure 6.2 (a)) may play a key role in determining agonist binding on α7. L118, which is situated in loop E, is close to the ligand binding site and it was found that it may consequently influence ligand-protein interactions in this region [109].

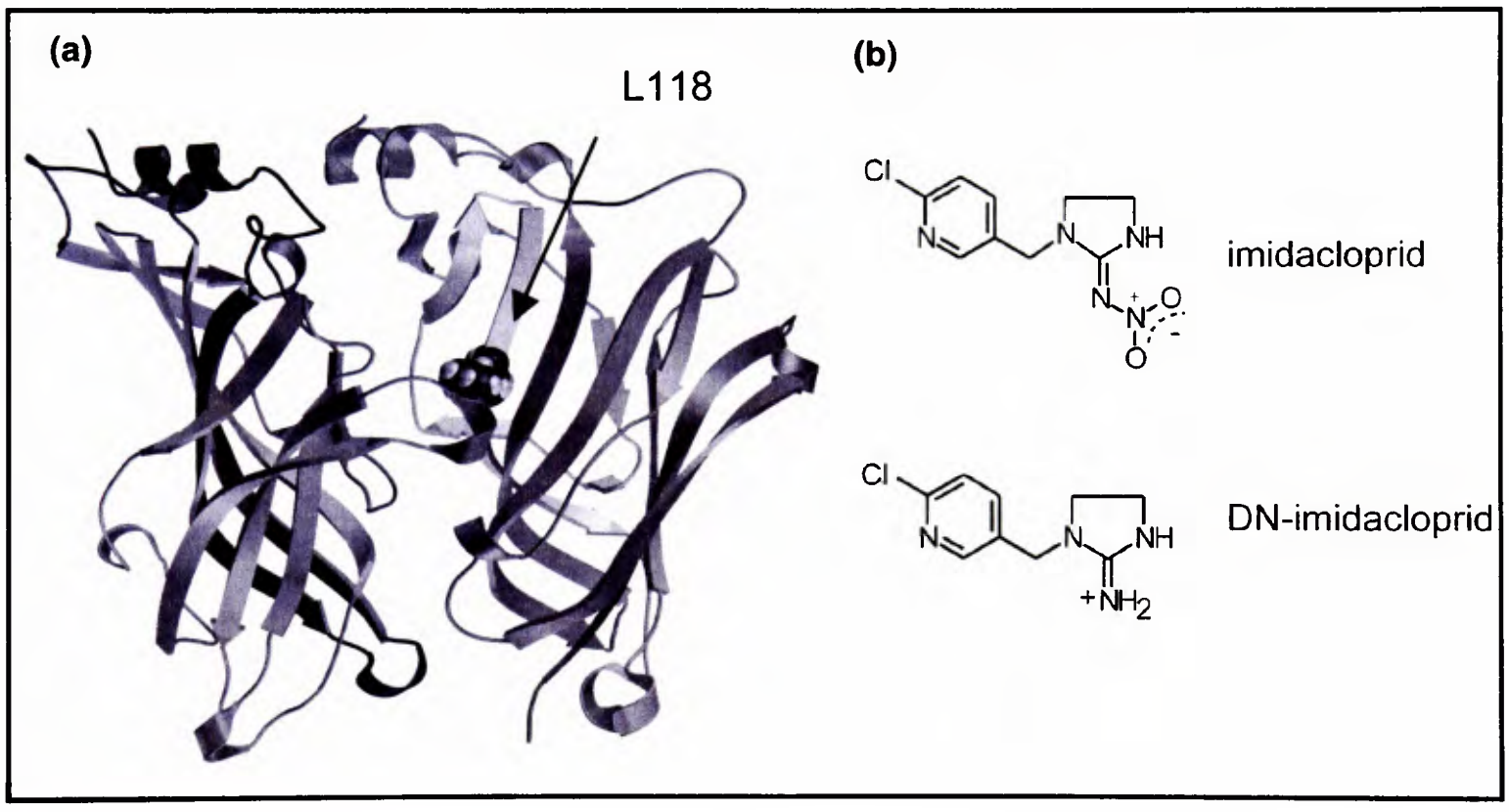


Figure 6.2 (a) Two subunits of the α7 nAChR are shown with L118 labelled in spacefill. (b) Imidacloprid and DN-Imidacloprid are shown (acetylcholine is shown in chapter 4). Figure from Amiri *et al.*, Mol. Pharm. 2006 - submitted.

In addition to docking studies of initial models (wild-type and mutant models), further docking studies carried out on successive frames of the five simulations allow for a dynamic study of ligand docking. The MD simulations illustrate the significant

influence of L118 on the overall dynamics of the receptor and the behaviour of the binding pocket.

6.3.1.1 Docking studies against the α7 nAChR and mutants

6.3.1.1.1 Wild-type α7 nAChR

Nicotine and acetylcholine, as well as IMI and DN-IMI, were docked onto the α7 nAChR homology model. The docking of nicotine and acetylcholine served as an important verification of the models since crystal structures of a nicotine and carbamylcholine (acetylcholine derivative) bound to the template structure are available [77]. Thus the binding modes of these ligands can be checked and compared to the crystal structures of AChBP. This is currently the only homologous, high resolution, ligand-bound X-ray structure available with residues in the binding site conserved across the LGIC family. If the binding modes do not match it is an indication that the model may have incorrect side-chain conformations. Figure 6.3 (a) shows the lowest energy dock of nicotine on α7 nAChR overlayed onto the X-ray structure of nicotine in AChBP (PDB code 1UW6). The lowest energy docking result with acetylcholine is shown in Figure 6.3 (b) and shows that the orientation of this dock is in very close agreement to that found crystallographically for carbamylcholine in AChBP (PDB code 1UV6)[77].

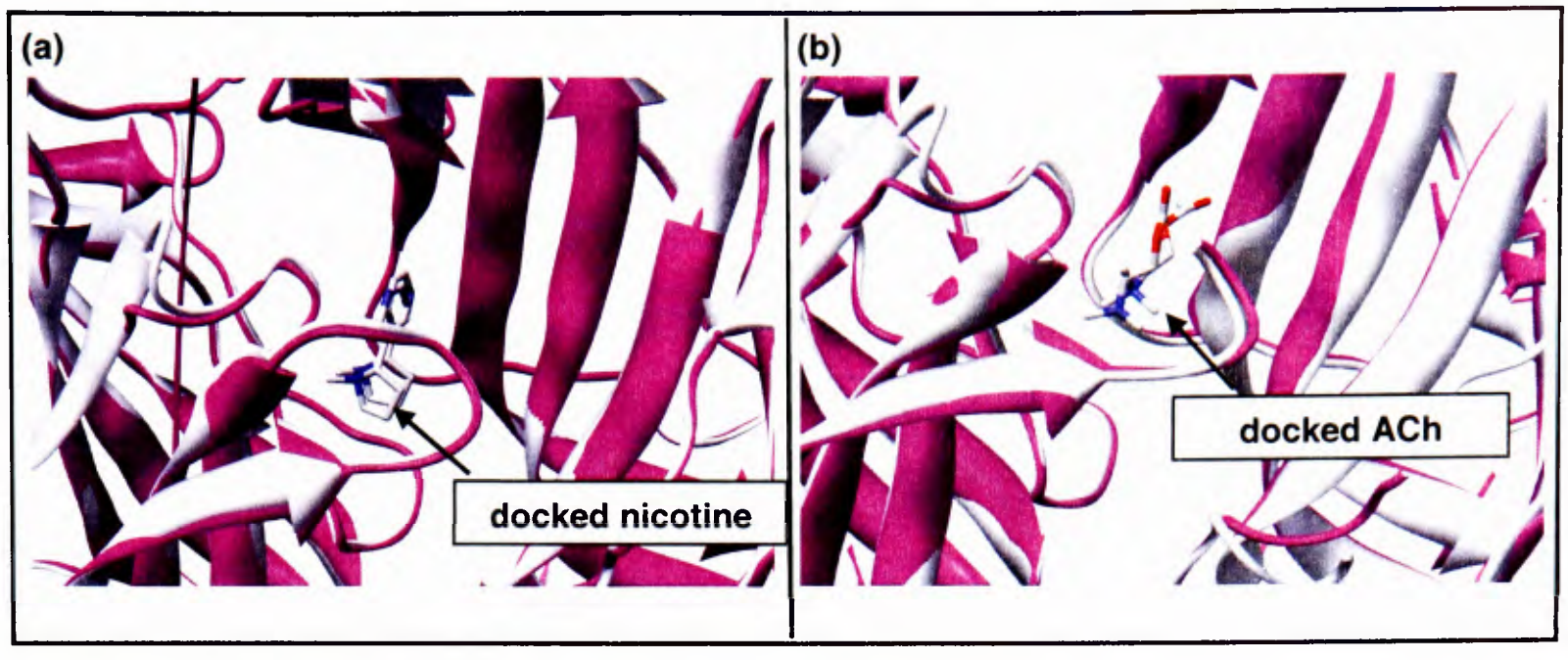


Figure 6.3 (a) Nicotine docked onto the α7 nAChR model (purple) overlayed onto the X-ray structure of AChBP with nicotine bound (silver). (b) Acetylcholine dock on the α7 nAChR model (purple) overlayed onto the X-ray structure of AChBP with carbamylcholine bound (silver).

There was one main mode of binding for nicotine with 45 out of 50 docks in the top ranked cluster. This indicates that nicotine docks highly favourably in that particular orientation and, since it matches closely with that of the crystal structure of AChBP with nicotine bound, it provides confidence in the model. The ACh docks spanned an energy range of 0.3 kcal/mol and were tightly clustered with 4 clusters in the binding pocket, which included one that matched the crystal structure. However, for IMI there were 13 binding modes and 7 of those were within an energy cut-off of 1.5 kcal/mol of the lowest energy dock (Figure 6.4 (a)). 6 out of 9 clusters for DN-IMI were within an energy cut-off of 1.5 kcal/mol of the lowest energy dock. These were comparable with the energies for IMI (Figure 6.4 (b)). As there were numerous docking solutions found for IMI and DN-IMI and there is currently no structural information to confirm the findings, there is a lack of confidence that the procedure could make reliable predictions for these compounds. Furthermore, it was reasoned that there might be substantial movement of residues in the binding pocket [85, 129, 130]. Also the binding pocket may have conformed to its bound ligand (discussed in

chapter 4). Thus given the size of IMI and DN-IMI, the pocket may have to undergo substantial movement to accommodate these molecules. Therefore it is best not to form precise predictions based on the docking results of IMI and DN-IMI.

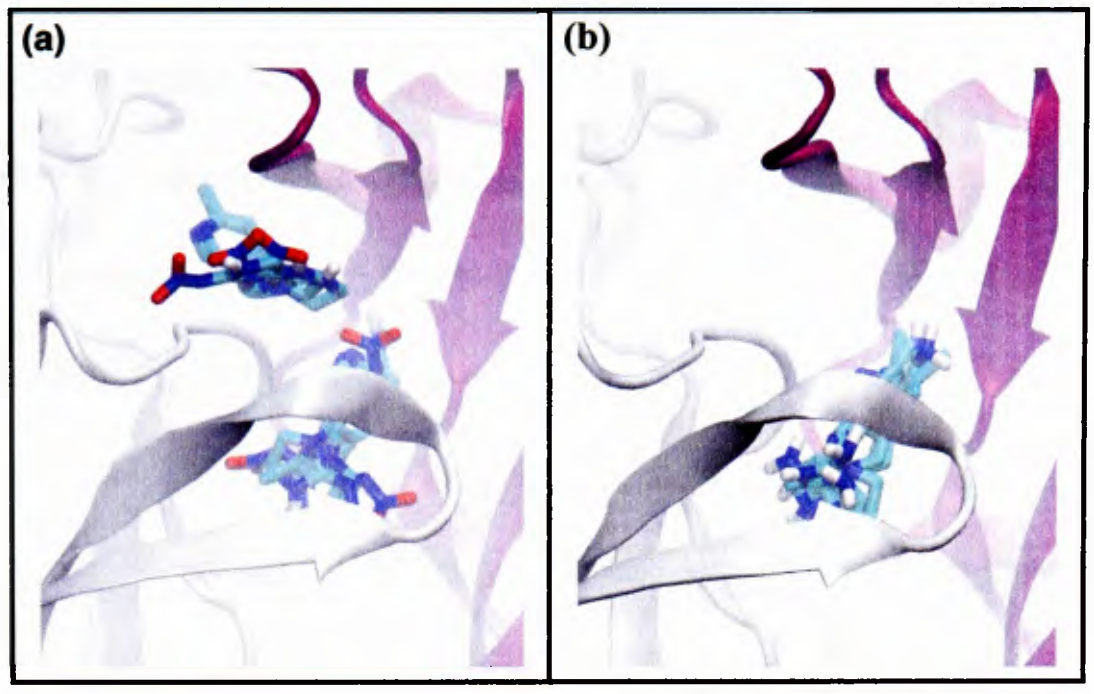


Figure 6.4 (a) Modes of binding for IMI in the wild-type α7 nAChR simulation. (b) DN-IMI docking clusters in the wild-type α7 nAChR simulation. In the absence of an X-ray structure, it is difficult to have confidence in the binding mode.

6.3.1.1.2 Mutant α7 nAChRs

As the comparison of the ACh docking with the crystal structure of carbamylcholine with AChBP is very similar, it provided confidence that the procedure could be used further to predict the interaction of ACh with the receptor mutants. Thus, the following *in silico* mutageneses were performed on this position: L118E, L118D, L118K and L118R. The results for ACh docking are from the top ranked cluster of each frame and are shown in Figure 6.5. It can be seen for the L118D mutation (Figure 6.5 (a)) and the L118E mutation (Figure 6.5 (b)) that the mode of docking is similar to wild-type shown in Figure 6.5. There is a slight shift in the position of the ACh towards what would be the surface of the membrane in the full length of the receptor. This presumably stems from the increased negative charge at L118 which pulls the quaternary nitrogen moiety downwards compared to wild-type. The docking

energies are close to that of the wild-type dock. When ACh was docked onto the L118K and L118R mutants, it was found that ACh would generally not dock into the binding pocket. On the first frame of the L118K simulation, 24% of the docks were in the binding pocket but in the wrong orientation; by the final frame, no ACh dock was inside the binding site. In the case of L118K, there were 3 clusters that positioned ACh near the binding pocket but in an inverted orientation. The effect was even more marked for the L118R mutation, where not one ACh dock was inside the binding pocket in any of the frames of the simulation. For the basic mutations docking energies for ACh were higher than those of the wild-type. These studies suggested that negatively charged mutations at this position would retain ACh binding, but positively charged mutations would abolish ACh binding.

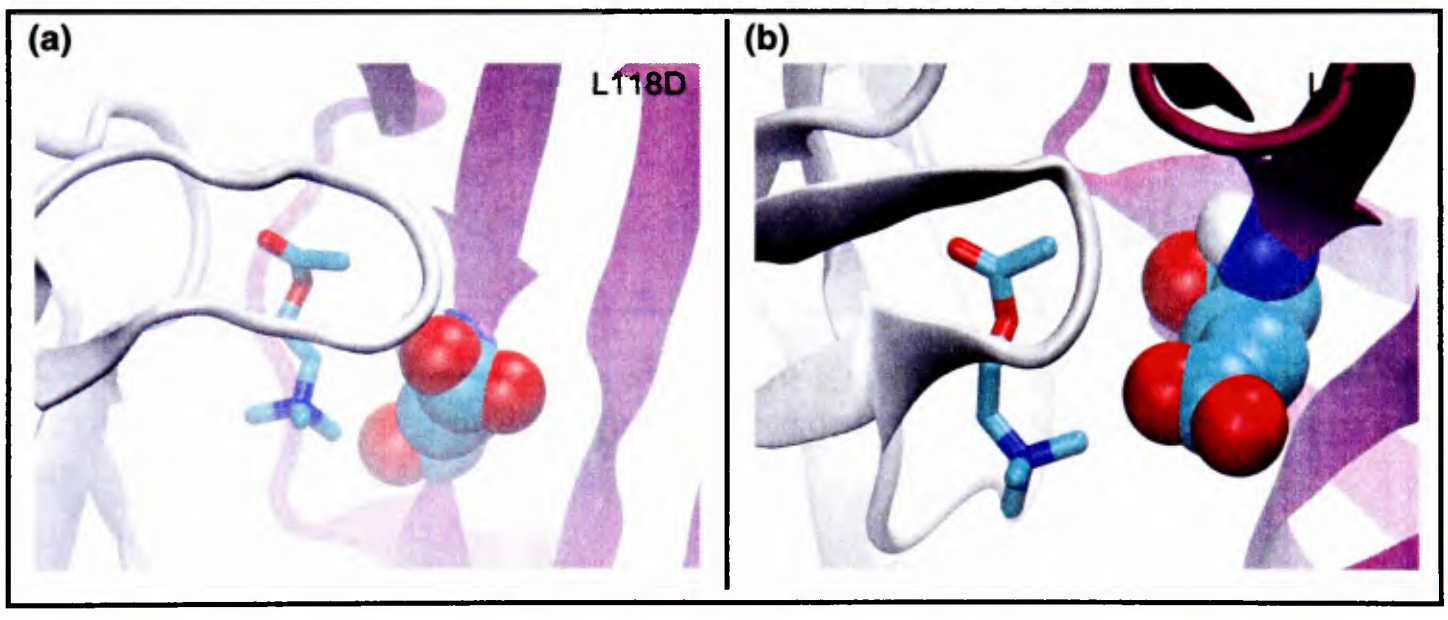


Figure 6.5 A close-up of acetylcholine docked onto L118D and L118E. Subunit A is in silver, subunit B shown in purple. (a) Acetylcholine docked onto L118D with D118 shown in spacefill. (b) ACh docked onto L118E with E118 shown in spacefill. The mode of binding is similar to that of the wild-type and the X-ray structure of carbamylcholine bound to AChBP.

6.3.1.2 MD and ensemble docking of α7 nAChR and mutants

Before embarking on experiments to confirm the above predictions, it was important to see how robust the homology models were to local fluctuations. It was also of interest to explore the effect of these mutations on the stability of the structure. Thus

10 ns molecular dynamics simulations for the wild-type and mutant α7 subunits were performed (see Table 6.1).

Simulations of the four mutant receptors showed noticeable intra-subunit structural deviation and inter-subunit structural rearrangemet (Figure 6.6). In all cases there was distortion and loss of five-fold symmetry. Thus the mutation of L118 in the binding pocket has a global effect on the receptor, also illustrated in the RMSF plots of the wilt-type against each mutant in Figure 6.7.

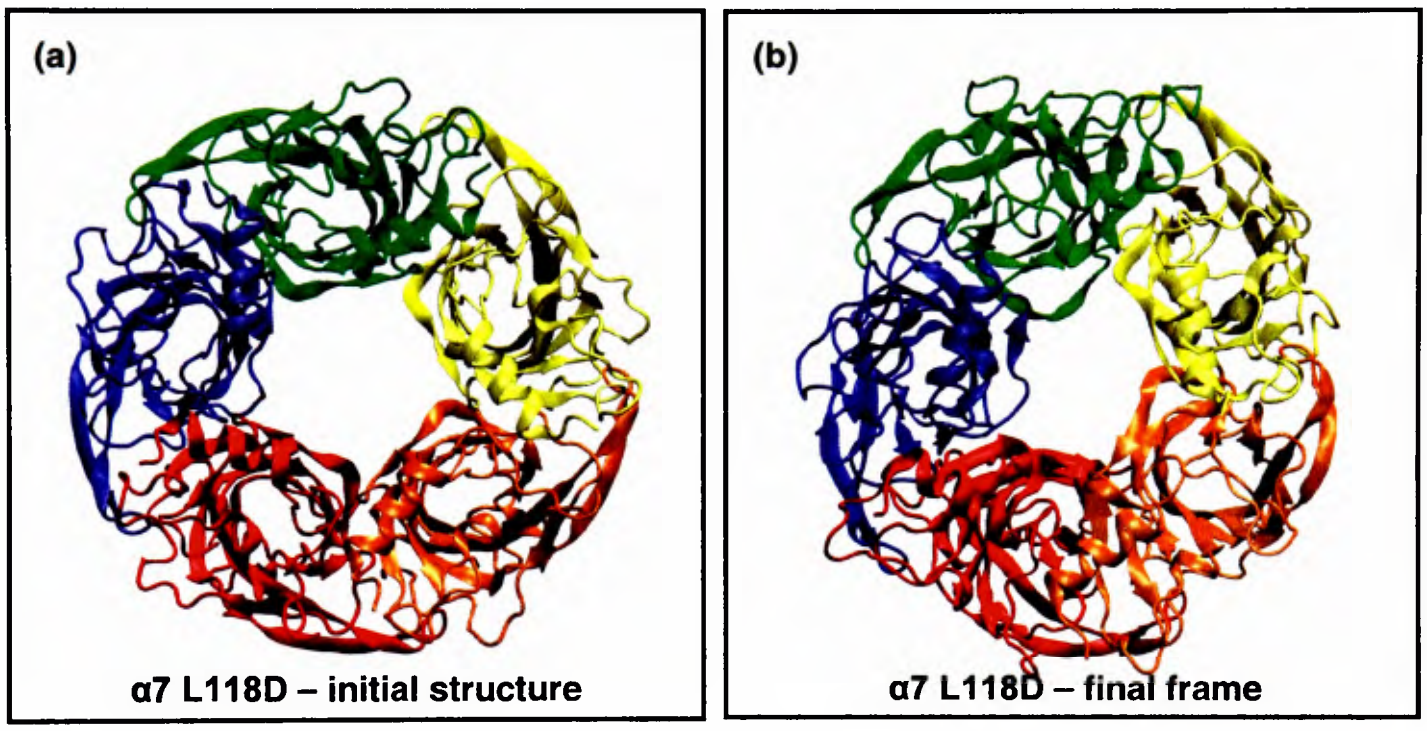


Figure 6.6 (a) The first frame of the α7 nAChR L118D simulation (top view). (b) The final frame of the L118D simulation (top view) showing distortions in the structure after 10 ns.

A common method for analyzing the relative stability of protein simulations is to analyze the root mean square fluctuations (RMSF) of the Cα atoms. Analysis of the Cα RMSF for these proteins revealed that the mutant structures showed a higher mean RMSF (Figure 6.7 (a)-(d)), indicating that this mutation has a large influence on the dynamics of the protein. In particular the regions corresponding to loop C (residues 180 to 196 for D, E, and R) and loop E (residues 112-119, the location of the mutation) appear to exhibit larger fluctuations compared to wild-type.

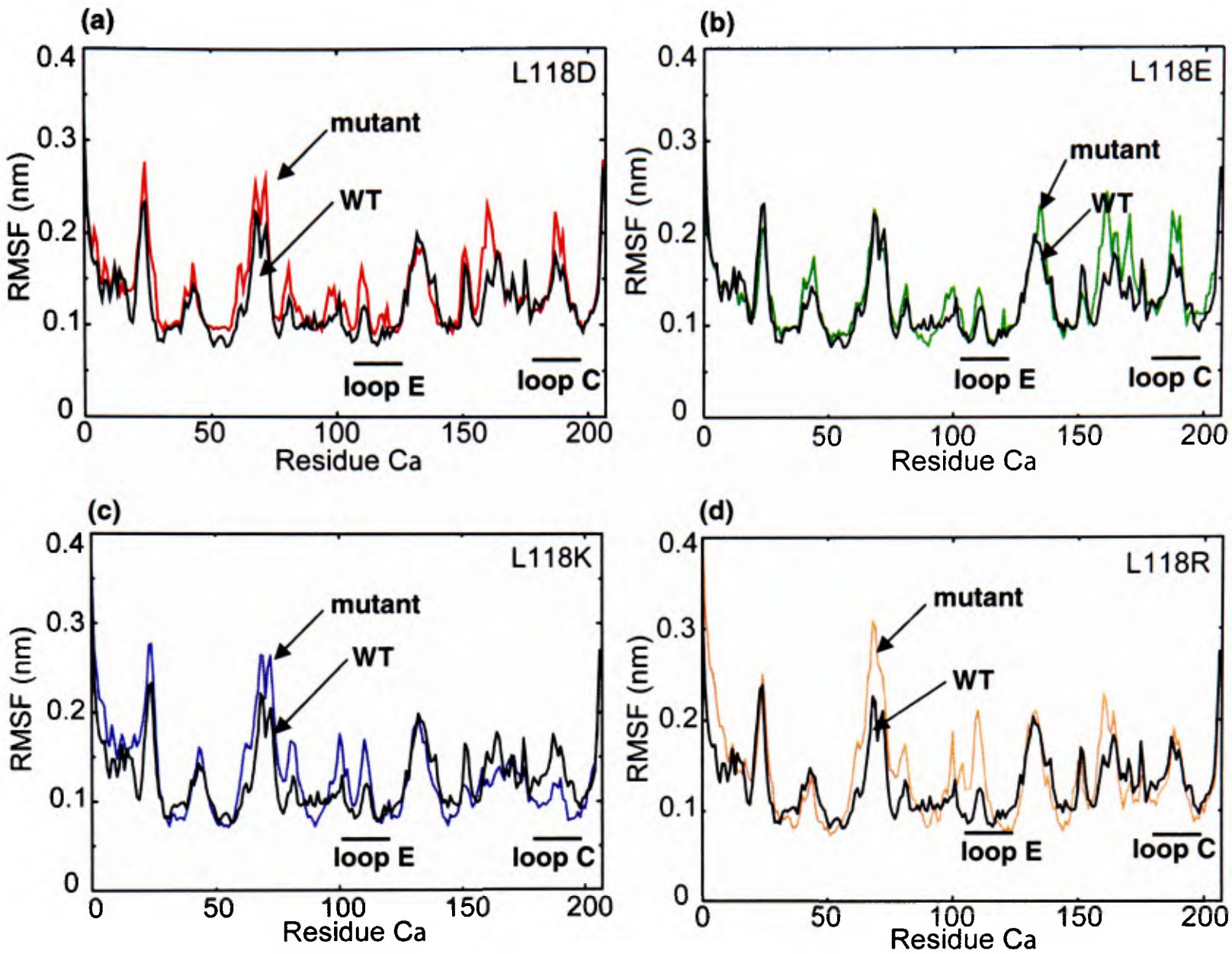


Figure 6.7 Four RMSF plots of the mutant simulations against the wild-type α7 nAChR simulation (shown in black in each case) , averaged over 5 subunits. (a) RMSF for L118D, (b) RMSF for L118E, (c) RMSF for L118K, and (d) RMSF for L118R. The mutant RMSF values are higher for most sections of the protein in each simulation. Fignre from Amiri *et. al.*, Mol. Pharm., 2006 - submitted.

As the result of an MD simulation is to produce a series of coordinates versus time, this presents a means to test the model and the sensitivity of the docking of ACh with respect to local fluctuations of the model. ACh and nicotine were docked back into 100 successive frames of each simulation to assess how well the binding pocket retains its shape and ability to accommodate the ligands and also to see how the ligand docking responds to local fluctuations. The results observed for the starting model were maintained for a portion of the simulation, until the binding pocket started to distort due to the absence of a ligand in the simulations (see chapter 4). Docking of

ACh to snapshots from the wild-type, L118D and L118E simulations resulted in docking solutions that were consistent with carbamylcholine bound in the X-ray structure of AChBP [77]. Nicotine docks on L118D were also successful, with improved energies and higher number of docks in lowest energy cluster for the first few ns of the simulation. Following this, in the 4-10 ns duration of the L118D and L118E simulations, the docking energies increased and ACh could not dock inside the binding site. Conversely, L118K and L118R mutant simulations failed to produce docking solutions with the ACh in the binding pocket at all, even in the first few nanoseconds when the binding site is still intact. Nicotine docks were also much worse for L118K and L118R; in most cases nicotine is not docked inside the binding pocket. Thus the results from the *in silico* study suggested that the negatively charged mutations (L118D and L118E) would retain the capability to bind ACh and nicotine, but the positively charged mutations (L118K and L118R) would not.

There are two main modes of binding for IMI for the frames of the L118D simulation. Towards the end of the wild-type simulation, IMI cannot enter the binding site as loop C has moved in as there is no ligand present during the MD simulation. This agrees with the AChBP docking results (chapter 4), where docking energies increase with time during the ligand-free simulations.

Although there are two major modes of docking for IMI, it has lower energies of docking and more docks in the binding pocket for L118D and L118E simulations compared to L118K and L118R. Its structure could be sterically unfavourable with the basic mutations, making it difficult to study this dock with the current computational capabilities. Also the lack of protein flexibility in AutoDock may be affecting the docking of IMI.

6.3.1.3 Leu118 site-directed mutagenesis

A summary of the functional data produced by Shimomura, M. and colleagues (details in Amiri *et al.*, Mol. Pharm 2006 - submitted) is presented here.

The effects of L118D, L118E, L118K and L118R mutations on the functional α7 nAChR expressed in *Xenopus laevis* oocytes show:

1 - Control experiments on wild-type α7 show that IMI is a partial agonist, consistent with earlier experiments [257].

2 - Responses of the α7 receptor to IMI were abolished following the L118D and L118E mutations and the EC_{50} values of ACh were increased by these two mutations.

3 - α7 L118D and L118E mutants still responded to ACh in complete contrast to the responses by L118K and L118R mutations which blocked the responses to ACh, but responses to Imidacloprid were observed.

4 - The maximum response to DN-IMI of the wild-type α7 nAChR was slightly greater than the response to ACh, resembling results of a previous study [258].

5 - The L118D and L118E mutations did not significantly affect the maximum response of α7 to DN-IMI but, the concentration-response curve of DN-IMI was shifted to higher concentrations by these mutations (Table 6.2). In contrast with the effects on the maximum response to Imidacloprid, the response to DN-IMI was abolished in the L118K and L118R mutations and therefore the pEC_{50} values for DN-IMI could not be determined for these two mutant receptors.

Table 6.2 pEC_{50}, Imax and Hill coefficient values for acetylcholine, Imidacloprid and desnitro-Imidacloprid on wild-type and mutant α7 receptors expressed in *Xenopus* oocytes. Table from Amiri *et al.*, Mol. Pharm., 2006 - submitted.

	Acetylcholine			Imidacloprid		
	Imax*	pEC_{50}	Hill	Imax	pEC_{50}	Hill
WT	1.03 ± 0.04	3.91 ± 0.05	1.4 ± 0.2	0.49 ± 0.03	3.63 ± 0.07	1.6 ± 0.4
L118D	1.07 ± 0.04	2.78 ± 0.04	1.3 ± 0.2	ND**	ND	ND
L118E	0.97 ± 0.02	3.12 ± 0.03	1.7 ± 0.2	ND	ND	ND
L118K	ND*	ND	ND	1.14 ± 0.10	2.94 ± 0.05	2.1 ± 0.4
L118R	ND	ND	ND	1.07 ± 0.05	3.00 ± 0.02	2.4 ± 0.4

	Desnitro-Imidacloprid		
	Imax	pEC_{50}	Hill
WT	1.21 ± 0.05	5.14 ± 0.07	1.2 ± 0.2
L118D	1.24 ± 0.11	4.21 ± 0.11	1.5 ± 0.5
L118E	1.37 ± 0.06	4.28 ± 0.08	1.6 ± 0.3
L118K	ND	ND	ND
L118R	ND	ND	ND

*Normalized maximum response.

**Not determined because the response to agonists was not detected or very small.

The concentration-response curves of ACh, Imidacloprid and DN-IMI for (a) the wild-type, (b) L118D, (c) L118E, (d) L118K and (e) L118R mutants are shown in Figure 6.8. In comparison with the experimental work, the *in silico* work showed that DN-IMI docks onto the wild-type and mutant forms of the receptor. In the wild-type, DN-IMI docked very well with 36/50 docks in the top cluster of docks which also matches that of one of the major modes of IMI docking. The docks improved particularly well for L118D and L118E mutants, much like ACh, while less of an improvement, or none at all, was observed for L118K and L118R. Although we do

not have X-ray data for the binding of this ligand and we have chosen not to focus on IMI docking, by using derivatives of a ligand we can gain insight into preferred mode(s) of binding and the DN-IMI results match those of the experimental data.

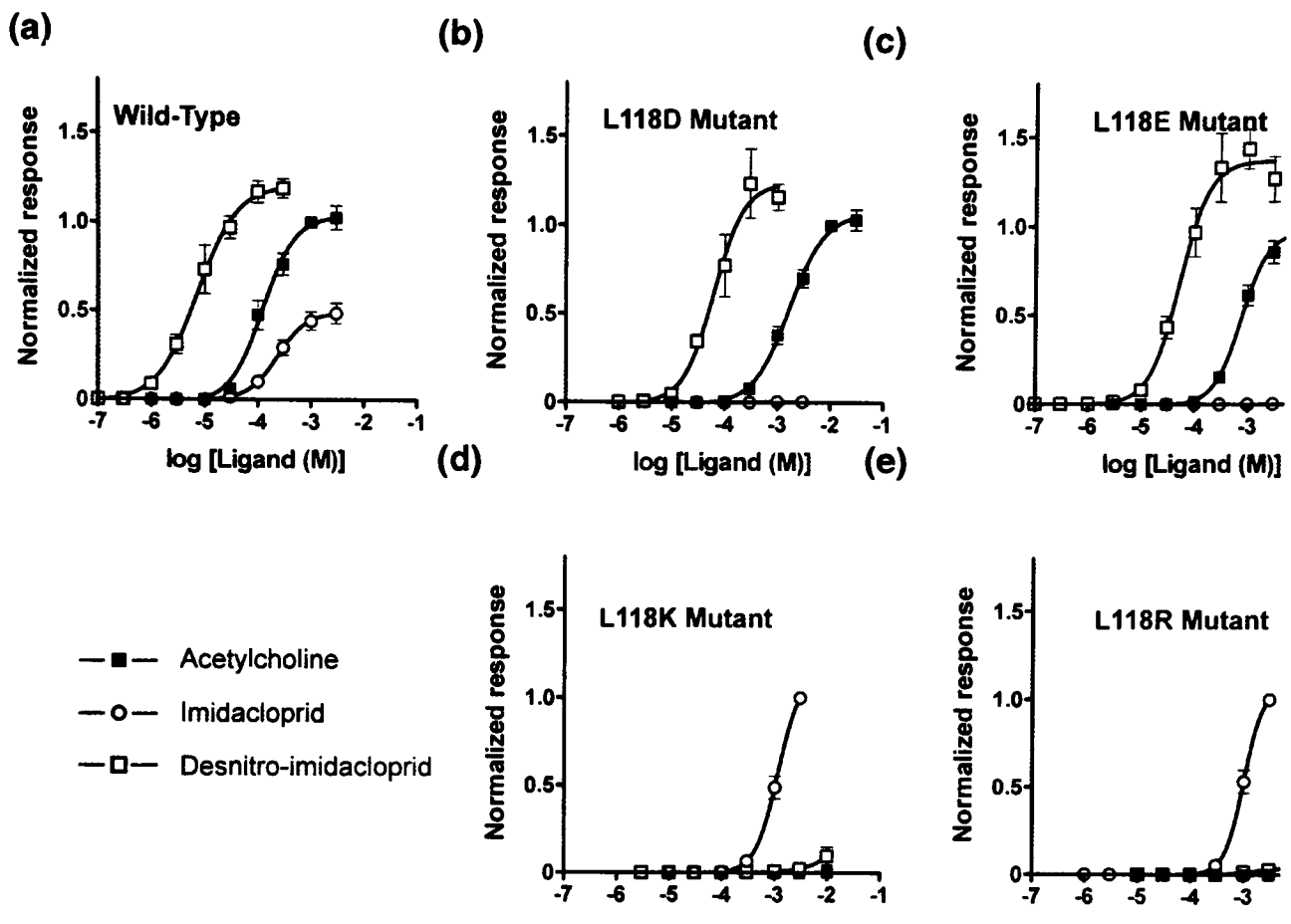


Figure 6.8 Normalized response curves plotted against ligand concentration (ACh, IMI, and DN-IMI) in the wild-type and four mutants of α7 nAChR. Figure from Amiri *et al.*, Mol. Pharm., 2006 - submitted.

6.3.2 ACR-16

Both the vertebrate α7 subunit [120] and ACR-16 [124] form functional homomeric ion channels when expressed in *Xenopus laevis* oocytes and they share high sequence homology in their N-terminal ligand binding domains (56% amino acid identity and 75% similarity). Here a molecular model of the ACR-16 ligand-binding domain, as well as the full ion channel, which represents the most complete three-dimensional model of the ACR-16 (using the method described in chapter 5) to date, is described. The complete model would allow for future MD studies of the channel and other more

coarse-grained investigation of conformational dynamics such as GNM or coarse-grained MD (see Methods chapter). Pore profiles and electrostatics calculations can also be carried out on this model as it has the TM domain as well as the EC domain. Docking studies of nicotine, acetylcholine and Levamisole were carried out on the EC domain based on the AChBP structure. Homology models of Levamisole-sensitive EC domains are also generated using UNC29, UNC63, and LEV1 subunits and subjected to docking studies.

6.3.2.1 Generating the full structure

The full structure of ACR-16 was obtained by two methods. First, homology modelling based on the 4 Å resolution nAChR [19]. Second, ZAlign was used to join the EC domain and TM domain of ACR-16 [233] (see chapter 2) using two different structures as templates (Figure 6.9). The EC domain was modelled on the AChBP X-ray structure [77] and the TM domain was modelled on the *Torpedo marmorata* TM domain EM structure [13]. This model has an EC domain of higher resolution and could be a better model to use when carrying out studies such as molecular dynamics and docking of ligands. The limitations of this model are that the TM domain is based on a heteromeric structure while the EC domain is based on the homo-pentameric AChBP.

Figure 6.9 The full structure of the ACR-16 receptor, generated from the combination of two domains using ZAlign as described in [233] and in chapter 5.

6.3.2.2 Homology modelling of the EC domain

For the docking studies, homology models of the homopentameric ACR-16 receptor were generated using three different templates. A first model was made using the crystal structure of the acetylcholine binding protein (AChBP) from *Lymnaea stagnalis* crystallized with nicotine bound (PDB code 1UW6) [77] at 2.2 Å resolution. This structure was chosen to study the docking of nicotine back onto the homology model of ACR-16, as a model based on a structure with nicotine bound would give the most meaningful results (see chapter 4). This model will be discussed in detail. The alignment used for the homology model of ACR-16 based on AChBP is shown in Figure 6.10.

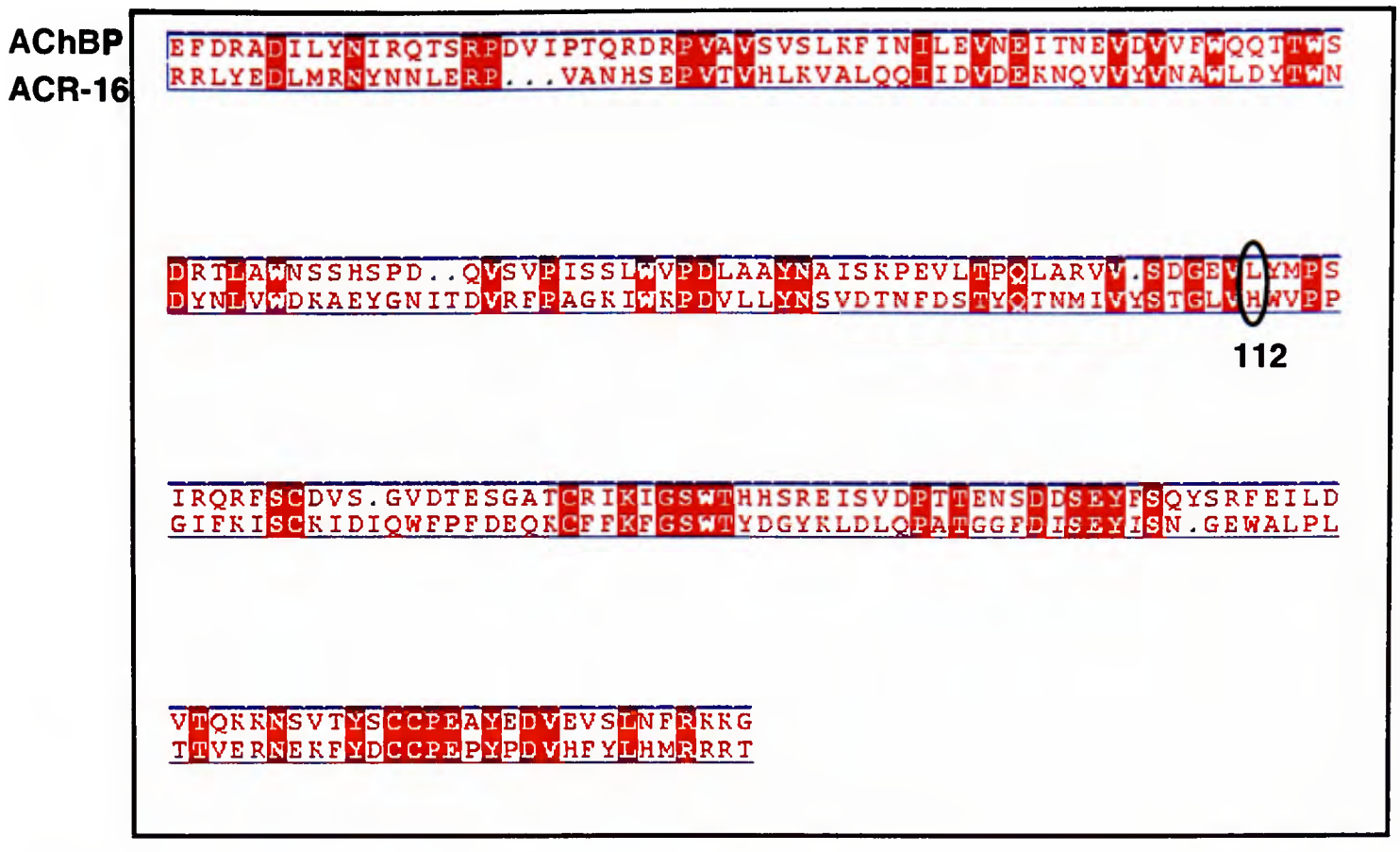


Figure 6.10 The alignment used for modelling the ACR-16 receptor using AChBP as a template (AChBP numbering). Matching residues between the two structures are highlighted in red. The histidine residue (H 112) in the binding pocket to be mutated is circled.

For comparison purposes, a second model of ACR-16 was based on the 4 Å resolution electro-microscopy structure of the *Torpedo marmorata* nAChR [19]. This is the first structure with both the EC domain and the TM domain together; it also includes the intracellular domain (IC), although several residues are missing. This structure is a hetero-pentamer composed of homologous subunits in the order α,γ,α,β,δ, while the ACR-16 is a homo-pentamer, however, this is the only available structure of a complete nAChR and the model would still be informative.

The third structure used to generate a model of the ACR-16 is a model of the α7 nAChR based on the AChBP. This is a model based on a model, however, because the α7 nAChR and ACR-16 have very high sequence identity, making this model and carrying out comparisons is important.

Here the focus will be on the homology model of ACR-16 based on AChBP, as this model was based on the highest resolution template and was the most successful

model when docking all of the ligands. Also during the homology modelling process a crucial residue in the binding pocket was identified by comparing the side-chain orientations of the two top-scoring models with AChBP. The first model (called 'incorrect-His' model), would not allow the correct docking of acetylcholine, however, the second model (called 'correct-His' model) docked the ligands in the correct mode as the side-chains in the binding pocket matched those of AChBP.

6.3.2.3 Docking of nicotine, acetylcholine, and Levamisole onto ACR-16

6.3.2.3.1 Nicotine binding

In the first model selected from homology modelling, the 'incorrect-HIS' model, nicotine bound with a flipped conformation. In the crystal structure the 6' ring is facing the 'complementary' side, however in the docking studies using the 'incorrect-HIS' model, nicotine is rotated around completely such that the 6' ring is near the 'principal' side of the binding pocket. In the 'correct-HIS' model, nicotine is positioned correctly, although the nitrogen on the 6' ring is flipped in the top mode of binding. This flipping of the ring is observed in molecular dynamics studies of AChBP with nicotine bound (Amiri *et al*, PEDS, 2006 – submitted) (see chapter 3). The 6' nitrogen is matching that of the X-ray structure in some of the other modes of binding in the new model. Figure 6.11 shows nicotine docked onto both the 'incorrect-HIS' and 'correct-HIS' models of ACR-16 and compared to the crystal structure of nicotine bound to AChBP [77].

Nicotine docked with the most favourable energies of the three ligands with docking energies spanning a range of 0.07 kcal/mol. 49/50 docks were in the same docking mode, closely resembling the binding mode of nicotine in the crystal structure of AChBP with nicotine bound [77].

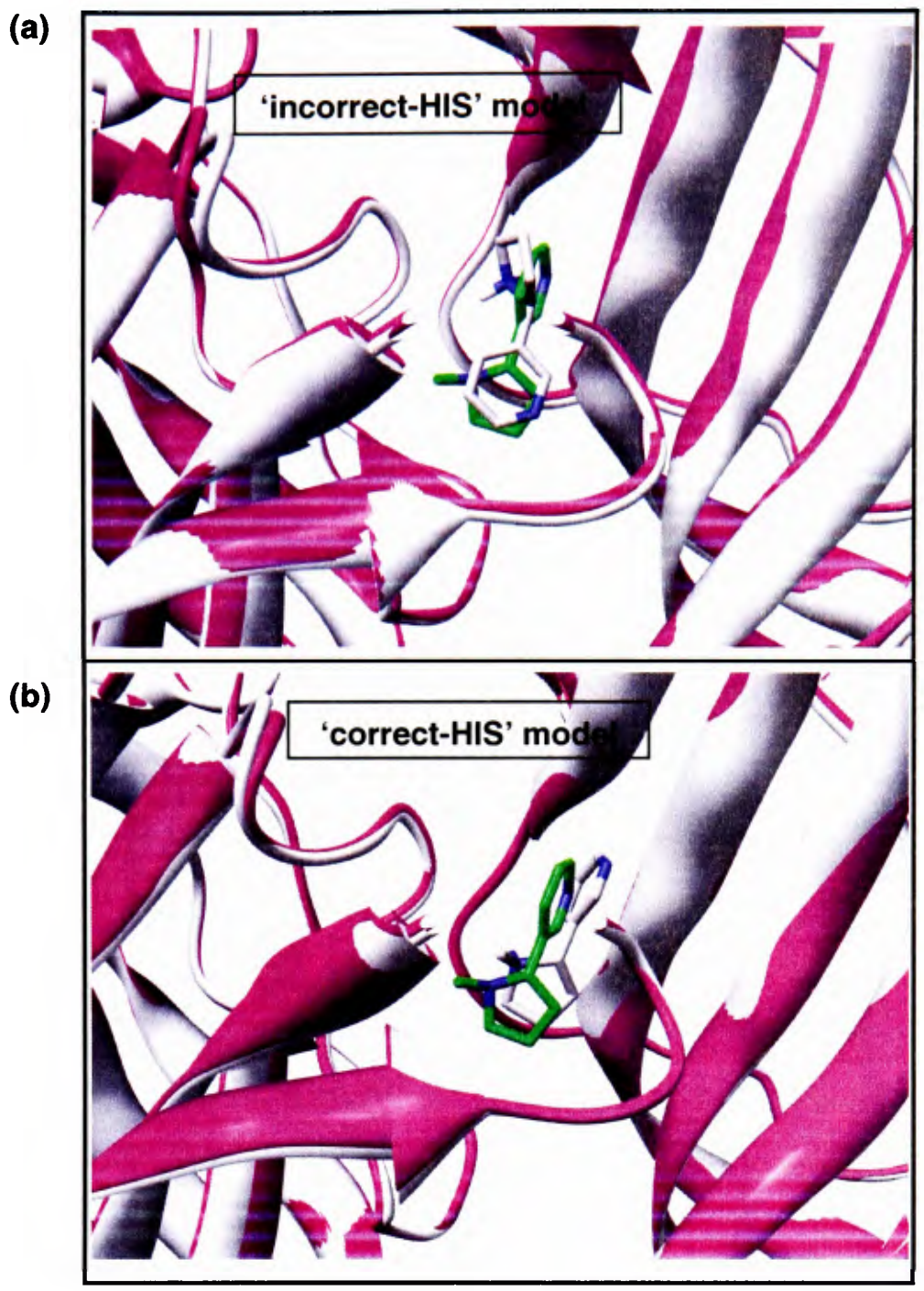


Figure 6.11 (a) ACR-16 homology model ('incorrect-His') based on 1UW6 fitted onto the 1UW6 crystal structure. (b) ACR-16 homology model ('correct-HIS') based on 1UW6 fitted onto the 1UW6 crystal structure. This model shows the correct mode of binding for nicotine and acetylcholine (not shown). The nicotine in the X-ray structure is coloured green in (a) and (b).

The difference between the 'incorrect-HIS' model and the 'correct-HIS' model is mainly the side-chain conformation of His112 (see Figure 6.12). Remarkably this difference made both acetylcholine and nicotine dock in completely wrong orientations. Thus, by comparing the side-chains of different homology models, a key residue in the binding affinity of ligands is identified. Further study of the importance of this residue is essential, thus our collaborators, Dr. David Sattelle and colleagues in the department of Physiology, Anatomy and Genetics at the University of Oxford, are

currently in the process of carrying out a mutation study on this residue in order to study the action of ligands in mutated states. From the preliminary *in silico* docking, the suggested mutations are Arg112 and Glu112 in order to observe both the effects of charge and steric hindrance.

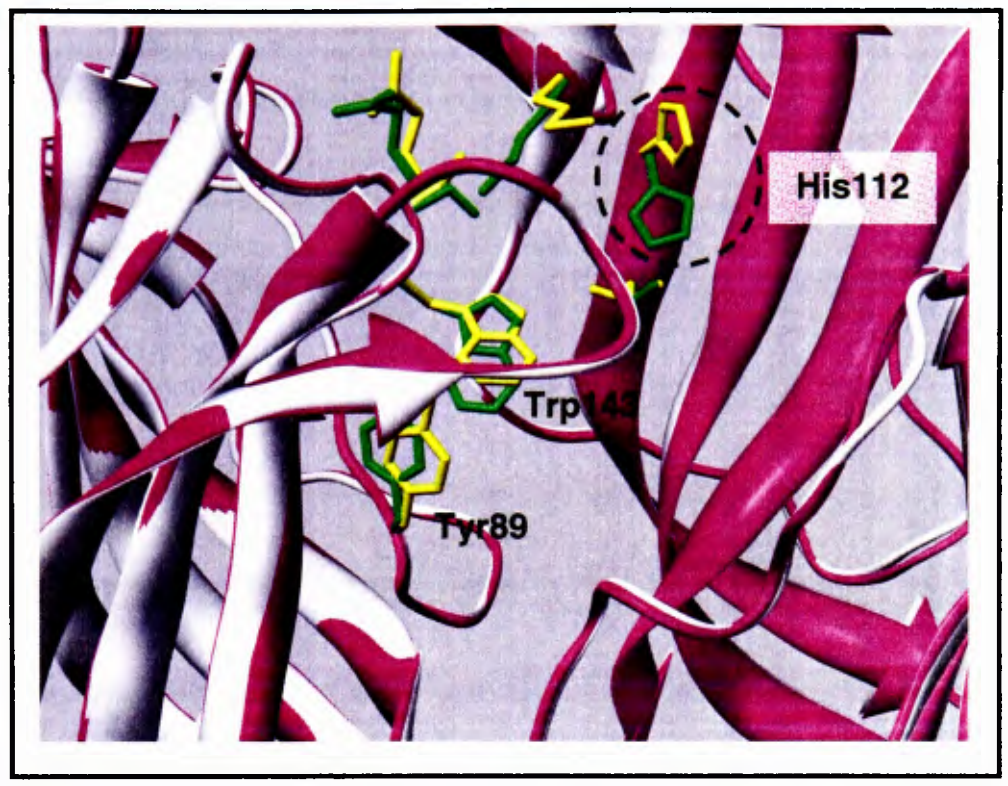


Figure 6.12 A comparison of the 'incorrect-HIS' model (side chains coloured yellow) for ACR-16 and the 'correct-HIS' model (side chains coloured green) showing the difference in side-chain conformation of His112 (AChBP numbering). This small difference completely alters (corrects) the mode of binding of the ligands.

6.3.2.3.2 Acetylcholine binding

Acetylcholine binds less favourably (more positive energies) than nicotine and usually the docks are split into several different clusters, suggesting different modes of binding or the lack of convergence of the docking algorithm. The docking energy and the clusters of low energy docks decrease by making the grid spacing smaller (see Methods chapter for details on grid spacing – section 2.6). The finer the grid spacing, the more points are defined within the grid and thus more detail in the interactions of the ligand with the protein. With smaller grid spacing, the acetylcholine docks more

accurately resemble that of carbamylcholine from the AChBP crystal structure (PDB code 1UV6) [77].

In the ACR16 homology model based on AChBP, 41/50 of the acetylcholine docks are in the same binding mode as carbamylcholine in the AChBP crystal structure (Figure 6.13) [77]. When acetylcholine is docked onto the actual crystal structure of AChBP, it binds in the same orientation as carbamylcholine as well. In the ACR16 model based on *Torpedo marmorata* nAChR, ACh is flipped around and does not bind as well. This can be attributed to the low resolution of the *Torpedo marmorata* structure and thus error in the side-chain conformations. The ACR-16 model based on the α7 nAChR homology model also does not bind ACh or nicotine as effectively as the ACR-16 model based on AChBP. This is due to the fact that this is a model based on a model and there could be an accumulation of modelling errors such as incorrect side-chain configurations. Therefore, good docking results were not expected. This was indeed the case and thus it enforces the importance of correct side-chain conformations.

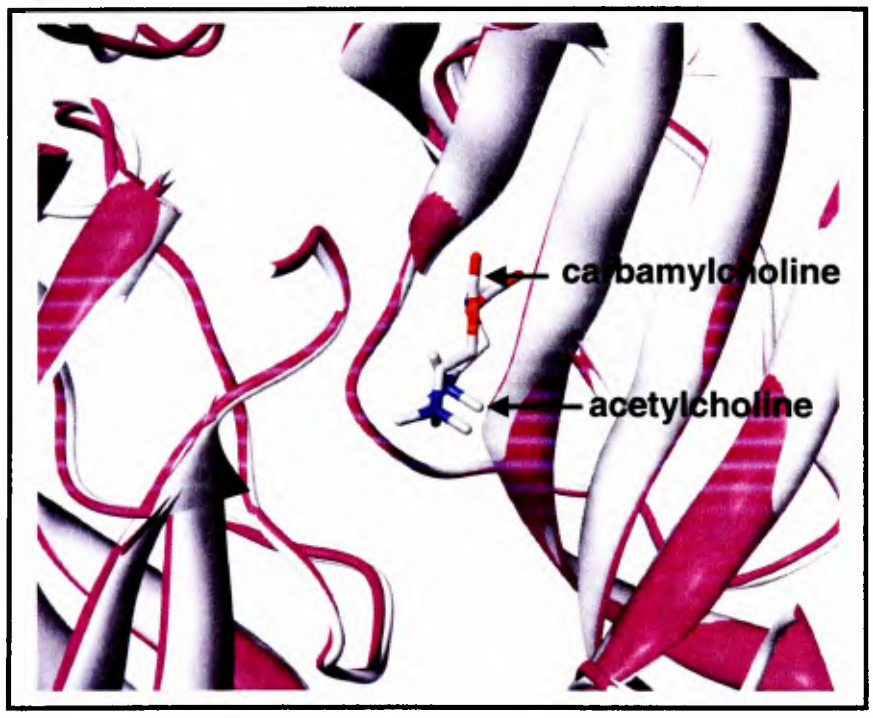


Figure 6.13 The docking of acetylcholine onto ACR-16 homology model based on the AChBP crystal structure with nicotine bound (1UW6). This dock is fitted onto the AChBP crystal structure with carbamylcholine bound.

The lowest energy binding mode for ACh (AChBP template) has the quaternary nitrogen facing upwards and the rest of the molecule is in a more horizontal

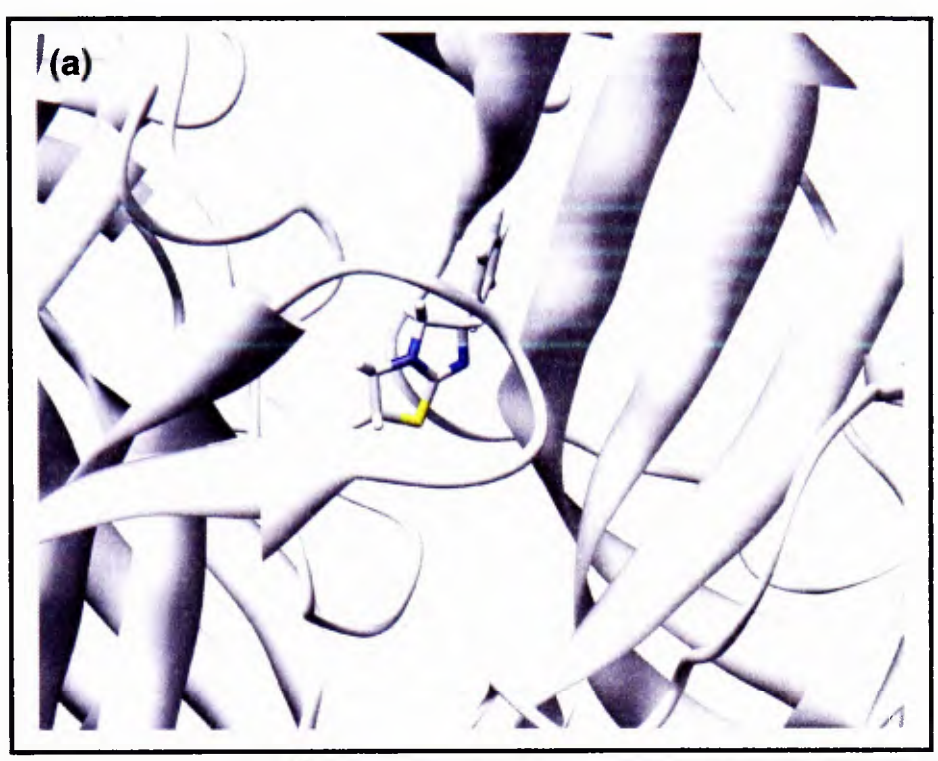
conformation compared to the binding mode of carbamylcholine in the AChBP crystal structure. Acetylcholine docks preferentially very close to loops D, E, and F in the homology models used here. As well as His112, which completely altered ligand binding in two slightly different conformations of its side-chain, a highly conserved Leu on loop D in nAChRs, which is Gln55 in AChBP, could be a candidate for mutational studies on ACR-16. Also Leu102 on loop E of AChBP which is replaced with Thr in the other sequences, (this leucine is heavily involved in main-chain hydrogen-bonding with the ligand via bridging water molecules in the AChBP observed in MD studies of AChBP with nicotine and carbamylcholine bound – see chapter 3). Thus further investigation of these residues may help the understanding of binding modes of ligands in this group of receptors and why ACR-16 is insensitive to Levamisole.

The lowest energy docks and the smallest number of docking modes for both nicotine and acetylcholine were achieved with the ACR-16 model based on AChBP with nicotine bound. The second lowest energy docking of nicotine was onto ACR-16 based on the *Torpedo marmorata* structure and finally least preferentially to the ACR-16 model based on the α7 nAChR homology model. This is not surprising as the AChBP structure has higher resolution than that of the *Torpedo marmorata* nAChR structure and the chance of error is reduced in the first model compared to the third model which is itself based on a model.

6.3.2.3.3 Levamisole Binding

Although it is believed that ACR-16 is a Levamisole-insensitive channel, *in silico* docking studies show that Levamisole binds with favourable energies which span an energy range of 1.12 kcal/mol (Figure 6.14 (a)). 46/50 docks were in the same conformation in the lowest energy dock cluster, suggesting one main mode of

docking. A comparison of the docking modes of nicotine, acetylcholine and Levamisole are shown in Figure 6.14 (b). The quaternary N atom is overlapping for the three ligands. These results are somewhat surprising as ACR-16 does not respond to Levamisole. However the events of binding and a ligand's potency to generate a response may be separate.



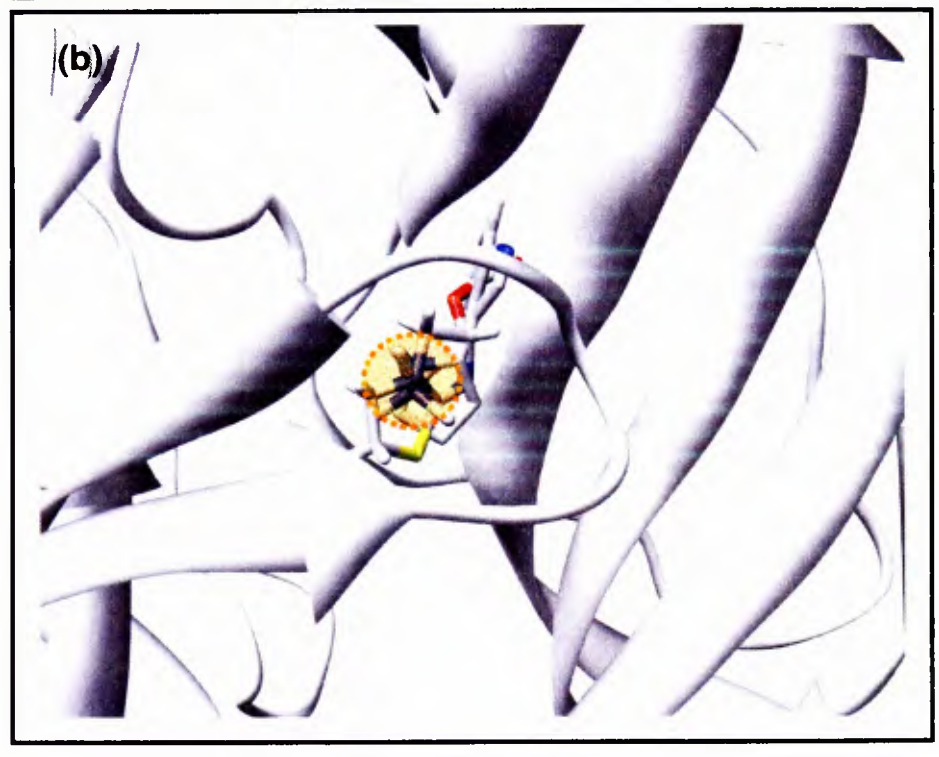


Figure 6.14 (a) Levamisole docked favourably onto ACR-16, although this receptor is believed to be Levamisole-insensitive. (b) Comparison of ACh, nicotine, and Levamisole docks. The pyrrolidine N from all three ligands lies in the same position.

6.3.3 UNC63-UNC63, UNC63-UNC29, and UNC63-LEV1 modelling and docking

Homology modelling of three combinations of these subunits is carried out with subsequent docking of nicotine, acetylcholine and Levamisole. It is not clear what the subunit compositions of these receptors may be, and thus various combinations are tested here. Again only two subunits were used for the docking studies. The three combinations used for docking are: UNC63-UNC63 (homo-meric combination), UNC63-UNC29, and UNC63-LEV1. To choose a good model, the side-chains were compared with that of the 'correct-HIS' model for ACR-16 which bound acetylcholine and nicotine in the correct mode. If the side-chains did not agree with the correct ACR-16 model, or the nicotine dock did not match the correct mode, the model was rejected as in the ACR-16 case (see 6.3.2.3).

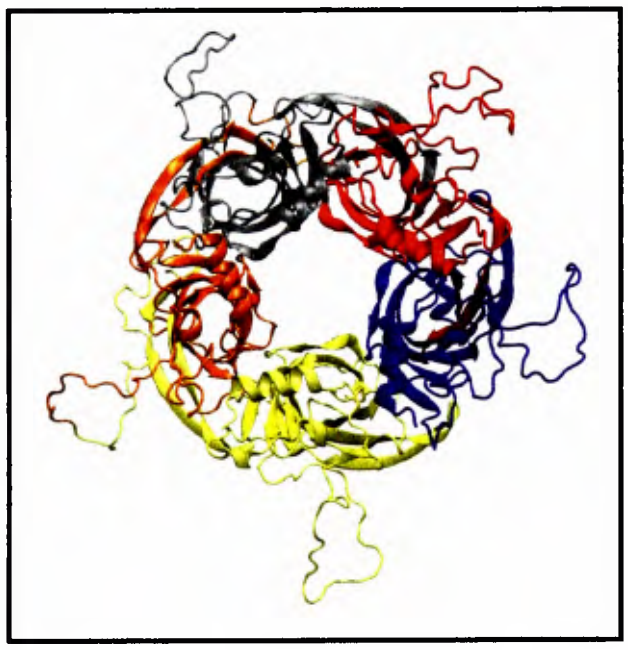


Figure 6.15 The homology model for UNC63-UNC63 shown (top view). These receptors have a longer loop B segment that extends outwards. UNC29 and LEV1 have similar structures.

6.3.3.1 Docking of Levamisole, nicotine, and acetylcholine onto UNC63-UNC63, UNC63-UNC29, and UNC63-LEV1

6.3.3.1.1 UNC63-UNC63

This homomeric combination of subunits bound Levamisole best with only 3 docking clusters spanning an energy range of 0.5 kcal/mol. This is followed by nicotine and finally ACh which both have a great number of clusters, but maintain low docking

energies. Nicotine's largest cluster had 31/50 docks and matched the mode of binding of nicotine in the crystal structure of AChBP (PDB code 1UW6). Levamisole had the same mode of binding as in ACR-16.

6.3.3.1.2 UNC63-UNC29

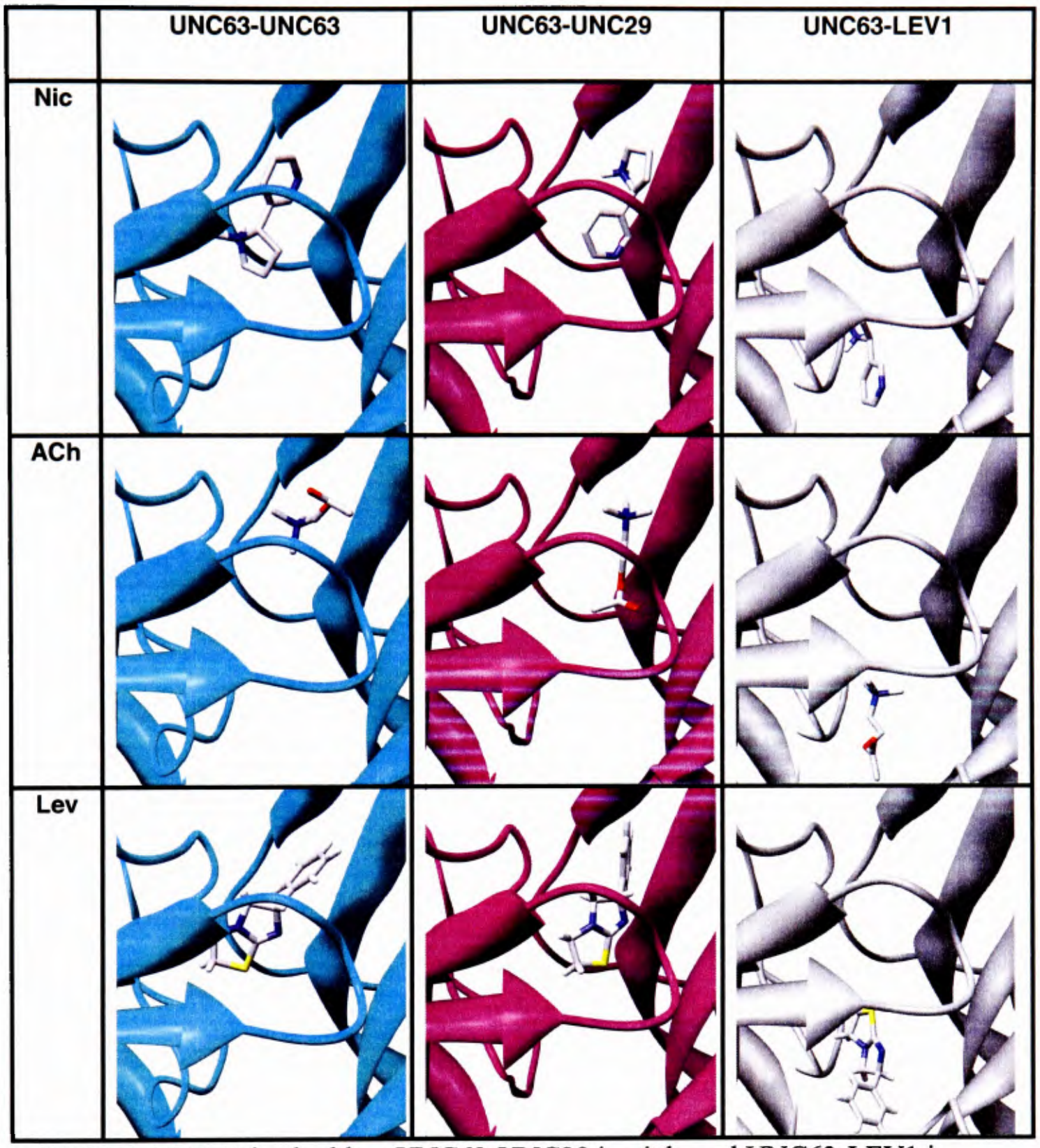
Nicotine and Levamisole docked very well onto this structure with only a few clusters each. Nicotine had slightly lower energies than Levamisole, however the 6' ring of nicotine is flipped around, not matching that of the X-ray structure as in the previous model. Acetylcholine also docked successfully, but the dock closest to the mode of binding of carbamylcholine was not in the top 4 cluster of docks.

6.3.3.1.3 UNC63-LEV1

This model resulted in the worst acetylcholine and nicotine docking of the three models, with none of the top-ranked docks inside the binding pocket. Levamisole docks were slightly better than those of nicotine and acetylcholine with fewer clusters and one cluster matched the binding mode of Levamisole in ACR-16. Nicotine and acetylcholine had a very high number of clusters suggesting various binding modes. In most cases nicotine did not bind inside the binding pocket, with only 13/50 docks matching the X-ray structure mode of binding.

A summary of nicotine, acetylcholine, and Levamisole docks onto the three models is presented in Table 6.3. The homomeric combination had the best docking results and this agrees with the experimental data that it forms functional homomeric channels [246].

Table 6.3 Summary of docks for the three Levamisole-sensitive subunit combinations: UNC63-UNC63, UNC63-UNC29, and UNC63-LEV1.



*UNC63-UNC63 in sky blue, UNC63-UNC29 in pink, and UNC63-LEV1 in grey.

6.4 Discussion

6.4.1 Loop E and L118 of α7 nAChR

The results indicate that L118 is influential on the binding properties and selectivity of agonists and also on the global dynamics of the receptor. The prediction of ACh-nAChR interactions for several mutants at position L118 of α7 is supported by the *in*

vitro experiments, suggesting that the model is capable of capturing the binding of ACh with homomeric α7 nAChRs. The experimental results show that substitutions of L118 in loop E by acidic or basic residues strikingly affect the maximum responses of α7 to ACh. A simple explanation for the effects on this mutation is that if L118 is mutated to a positive side-chain, the interaction with the charged quaternary nitrogen is strong enough to prevent binding to the pocket and thus abolishes binding and the channel fails to open.

If the receptor conforms to a bound ligand (during a simulation or captured bound in a crystal structure), then the model will suffer from being biased away from a conformation that would bind Imidacloprid. Also, recent simulation studies on AChBP [85] and on AChBP-based homology models of α7 [129] suggest that there can be substantial changes in the shape of the binding pocket, in particular the conformation of loop C (also chapter 3). Taking into consideration these limitations, here a potential model that can rationalize the data is proposed.

The importance of electrostatic forces for the interaction of IMI with nAChRs have been previously studied [259]. Imidacloprid possess a negatively charged nitro group whilst its desnitro derivative (DN-IMI) possesses a positive charge at the guanidine moiety. It is interesting to note that the effects of DN-IMI with respect to the mutations are quite similar to ACh. The experimental work shows that Imidacloprid itself is inactive against the L118D and L118E mutants, but is active against the L118K and L118R mutants, suggesting that the electrostatic interaction between the nitro group and this residue is critical in determining whether it will bind as an agonist or not.

However, electrostatic interactions are not the only forces determining agonist interactions with the α7 nAChR. The L118D and L118E mutations shifted the ACh

concentration-response curve to higher concentrations, suggesting that such mutations in loop E may lead indirectly to a conformational change in another region of the receptor playing a key role in interactions with agonists. This is also supported by the large fluctuations in the structure from the MD data illustrating the extended effect of this mutation on the entire structure. Nevertheless, electrostatic interactions can explain the changes in the concentration-response curves of the Imidacloprid derivative DM-IMI lacking the nitro group. Consistent with this, the L118K and L118R mutations abolished responses of the α7 nAChR to DN-IMI, whereas L118D and L118E mutations permitted the agonist actions of this ligand.

In conclusion, molecular modelling and site-directed mutagenesis has been used to predict that Leu118 of the α7 nAChR subunit contributes to agonist binding. Interestingly, Leu118 is not conserved across the nAChR family. Thus, the residue at this location may participate in determining subunit-specific responses to agonists. Also loop E is highlighted as playing a key role in the binding of the insecticide, Imidacloprid. These results, taken in combination with previous studies on loop C, loop D, and F [256, 260, 261] are enhancing the understanding of the binding to nAChRs of commercially important nicotinic agonists.

6.4.2 ACR-16

The ACR-16 model generated here, which includes both EC and TM domains (Figure 6.9), represents an important tool for identifying compounds that can potentially interact with the receptor. This, in conjunction with the use of the highly tractable genetic model *C. elegans*, allows for rapid screening for novel anthelmintics [119].

Nicotine and acetylcholine were docked onto two different models, based on AChBP, with different conformations of the His 112 side chain. One of the models would not bind any of the ligands in the correct conformation and thus presented another case for mutational studies of His112. A slightly different conformation of its side chain is unfavourable for ligand binding. Nicotine bound very favourably to the 'correct' model of ACR-16 with only one docking cluster and very favourable energies, approximately 3 kcal/mol lower than Levamisole docking energies. Although ACR-16 is a Levamisole-insensitive channel, Levamisole docks were also highly favourable, with only 3 docking clusters and relatively low docking energies. This result is unexpected and could suggest that a ligand may bind but not elicit a response in the ion channel. Nicotine and acetylcholine also showed favourable docking against ACR-16.

6.4.3 UNC63-UNC63, UNC63-UNC29, and UNC63-LEV1

Of the Levamisole-sensitive channels, the best docks were from the homomeric UNC63-UNC63 combination, in agreement with studies suggesting that this subunit is α-type and can also form functional homomeric channels [246]. This combination could potentially form functional homomeric Levamisole-sensitive receptors. All three combinations responded well to Levamisole and nicotine, although acetylcholine docks were less favourable. The Levamisole docks shared the same docking mode as the Levamisole-insensitive ACR-16. UNC63-UNC29 docks were less favourable (higher energies and more clusters) than UNC63-UNC63 docks but more favourable than the docks onto UNC63-LEV1. This may suggest that the latter combination may not be a functional form of a channel.

Both the ACR-16 study and the L118 study show how computational docking and modelling studies can identify key residues, provide hypotheses to be tested, and open avenues of experimental work. However it was also noted that docking software cannot always distinguish the best mode of binding, for example in the case of Imidacloprid. It is likely that the structure of the binding pocket conforms to the bound ligand and is not particularly accommodating to other ligands (see chapter 4).

7 Structure, energetics, and coarse-grain molecular dynamics simulations of ligand-gated ion channels

7.1 Introduction

It was a natural extension to the α7 nAChR modelling (chapter 5), to generate models of other LGICs. The LGICs have a high degree of global structural homology revealed from sequence similarities and a high degree of functional homology (see chapter 1). Computational studies of these channels would allow for the rationalization of functional differences (i.e. permeation properties) based on local structure and sequence differences. Using the method developed and discussed in chapter 5, models of the human $GABA_A$, glycine, and $5HT_{3A}$ (serotonin) receptors were generated by combining homology models for their TM and EC domains. Once a full structure of the *Torpedo marmorata* nAChR [19] was made available, which included residues in the IC domain, new models of the LGICs were generated. Although this structure has many missing residues, in particular between the M3 and

M4 segments of IC domain, and it is a heteromeric structure while those modelled are all homopentamers, it is the most complete structure of a LGIC to date. It is believed the IC domain has functional importance [19, 44, 98, 262, 263], especially for ion conduction, and thus full channel models would be more accurate for analysis. The goal here was to study how the differences in sequence were reflected in the permeation properties, pore properties, and dynamics that can be inferred from computational studies of ion channels. There have been recent advances in the study of the structure and function of Cys-loop receptors suggesting that the interaction of residues on the interface between the EC domain and TM domain results in the slight conformation changes required to open the channels ([15, 57], and recent work [59] suggests that it is a cis-trans isomerization of a proline residue on the same interface that is a key element to the gating mechanism. The question of gating has been approached from several experimental approaches, such as mutation studies [67, 264], φ-value analysis [69], and electrostatic interactions [100]. Here pore profiles and Poisson-Boltmann calculations of an ion placed at successive points along the pore of each model are carried out, similar to that carried out in chapter 5. Also it is of interest to study the conformational dynamics of these receptors and their interactions with the lipid bilayer using molecular dynamics. One previous atomistic MD study was carried out on a combined model of the α7 nAChR [131]. However as these structures are very large (~ 1800 residues) and they are homomeric homology models based on a (partly) heteromeric structure, the size, the missing residues, and possibility of error during homology modelling makes them unattractive candidates for atomistic MD. Also the time required and computational cost of carrying out atomistic MD simulations on such large systems is very high, and it would not allow for the observation of certain properties and events of interest, such as bilayer

assembly. Thus for simulating the dynamics of LGICs coarse-grain (CG) MD is the method of choice [172, 173, 175]. It allows the spontaneous formation of the lipid bilayer, as the lipid environment is crucial to the function of these receptors. The positioning of these lipids can help our understanding of the function and enable the study of residues interacting with the lipid bilayer. Once bilayer assembly has taken place, the LGICs can be studied computationally in their 'natural' environment with a more accurate placement of the bilayer [265].

7.2 Methods

7.2.1 Homology modelling

Two different methods were used to construct the LGICs in this chapter. The template structures used for the homology modelling of these receptors are different for the two methods. The first method, employed before the full *Torpedo marmorata* structure was resolved using electron microscopy [19], is the method discussed in chapter 5 where the TM and EC domains are modelled separately, and then combined using ZAlign. In this first method the transmembrane domain was modelled on the *Torpedo marmorata* TM domain structure (PDB code 1OED) [13] and the EC domain was modelled onto the 2.7 Å AChBP structure (PDB code 1I9B) [20]. However, once the 4 Å resolution full structure of the *Torpedo marmorata* nAChR became available [19] it was used to remodel all the LGICs. The models from this latter approach will be discussed in this chapter as they represent the most complete structure, which includes the IC domain, and also there is likely less error in the interface between the EC and TM domains by not having to combine them using ZAlign.

Modeller 6v2 [141-143] was used to generate homology models of the chicken α7 nAChR, $GABA_A$, glycine, and $5HT_3$ (serotonin) receptors. The top 10 models out of 100 for each protein were ranked according to their Modeller energies. In order to choose the best of the top 10 structures, Procheck [186] was run to check the stereochemical quality of each model. The model with the lowest percentage of residues in the disallowed region of the Ramachandran plot as scored by Procheck was chosen and used in subsequent structural studies.

7.2.2 APBS

The Poisson-Boltzmann (PB) equation was used to estimate the Born energy of a Na^+ and a Cl^- ion along the length of the whole channel (see chapter 2). The program Adaptive Poisson-Boltzmann Solver (APBS) [182] was used to solve the Poisson-Boltzmann equation. Sample points along the pore axis at which to place the ion were derived from the program HOLE [183], which also yielded the pore radius profile of the LGIC models (see Methods chapter for a detailed description of this method).

7.2.3 Coarse-gained MD

In order to increase the time-scale and save on computational resources for simulating these large LGIC systems, coarse-grained (CG) models that treat groups of atoms as single particles were used (for details see Methods chapter). The CG structures of the LGICs were obtained from all-atom files, parametrized according to the CG model, energy minimized, and simulated.

7.3 Results

7.3.1 Homology modelling of LGICs

The 4 Å resolution *Torpedo marmorata* nAChR structure [19] was used to model all the LGICs discussed in this chapter. Two chicken α7 receptors are discussed here, one is the model generated and discussed in chapter 5 (α7*) and the other was regenerated using the *Torpedo marmorata* nAChR structure (α7). The human homomeric α1-type $GABA_A$ and glycine receptors as well as the $5HT_{3A}$ (serotonin) receptor sequences were used for the respective models. In the EC domain, these receptors all share high sequence identity of the Cys loop which is marked by two cysteines separated by 13 residues [266] (see Figure 7.1), thus making up the 'Cys-loop' family of receptors. There is also high sequence conservation in the TM domain, in particular in the hydrophobic regions [266] (see Figure 7.1) that includes the hypothesized 'hydrophobic gate' of the channels [13, 60, 226]. In other regions, their sequence identity is relatively low, for example $5HT_{3A}$ and $GABA_A$ receptors share between 11 and 12 % sequence identity [267] with the glycine and $GABA_A$ alpha subunit homopentameric channels sharing the highest sequence identity of the group at 34.5% as determined by ClustalW identity scoring [268, 269]. In particular segments, such as the M4-IC linker region, it is difficult to have complete confidence in the structures generated, as there is a great deal of missing structural information, and substantial variation in the number of residues among the receptors. This makes the homology modelling process a challenge and thus the resulting models and any conclusions drawn from the computational study of these models must be treated with care.

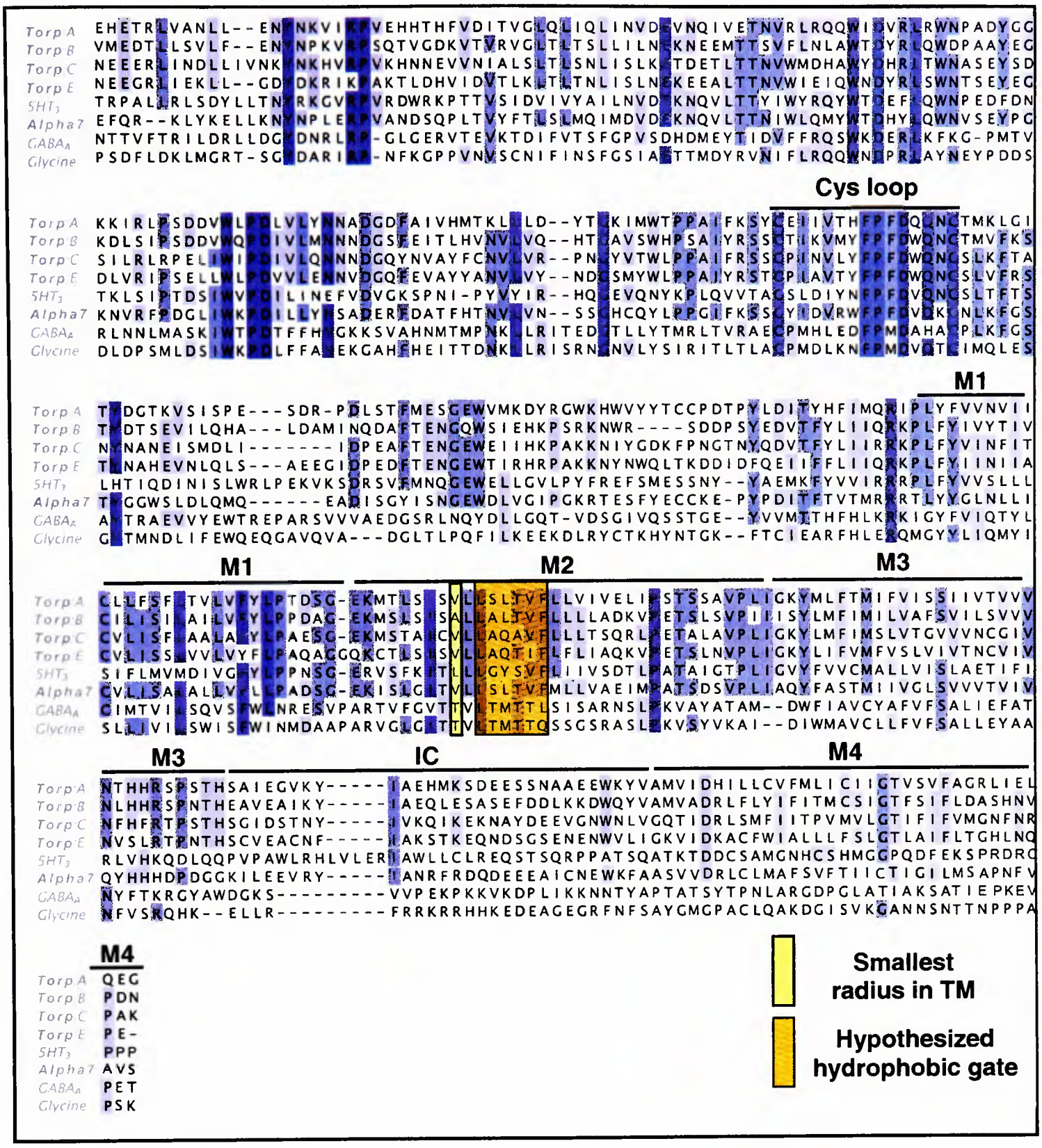


Figure 7.1 The multiple sequence alignment of the four different subunits of *Torpedo marmorata* nAChR, chicken α7 nAChR, human α1 $GABA_A$, human α1 glycine, and human $5HT_{3A}$ receptors are shown. The M1-M4 helical segments are annotated, also the structurally significant Cys loop is labelled, which is a common feature amongst the modelled receptors. MultAlin [144] generated the alignment and Jalview [270] was used to view the sequence alignment. The darkness of the colour blue indicates the degree of sequence homology amongst the sequences.

However until complete, high resolution, structures become available this type of modelling is a good approach in understanding the probable structural make-up of the proteins. Also the homology models can be used to study the structural, energetic, and dynamic properties of these receptors.

7.3.2 Pore profile analysis of the LGICs

Pore profile analysis was carried out on all of the LGIC models and the *Torpedo marmorata* nAChR EM structure [19]. Not surprisingly, the pore profiles for all of the LGICs are similar in radii for most regions (Figure 7.2). The main difference is seen for EC region of the α7* model which was generated by the method described in chapter 5. In this method, the EC domain was modelled on the AChBP structure [77]. The pore profile of this model suggests that the pore of the AChBP may be smaller than that of the *Torpedo marmorata* structure. All of the pore profiles for the LGIC models based on the *Torpedo* structure have very similar pore profiles with the IC domain being the most constricted segment with radii as low as 0.65 Å ($5HT_3$ and glycine receptors). An early study has suggested a possible pathway for the entry of ions into the cytoplasm, in which the ions exit the receptor through the sides of the IC domain [271]. The next area of greatest constriction is at the TM domain and there are also areas of larger constriction in the EC domain (see Figure 7.2). Representations of the pore surface for each of the modelled receptors are shown in Figure 7.3. All of the models display the smallest radii (red) in the IC region (except for the α7* model lacking the IC domain).

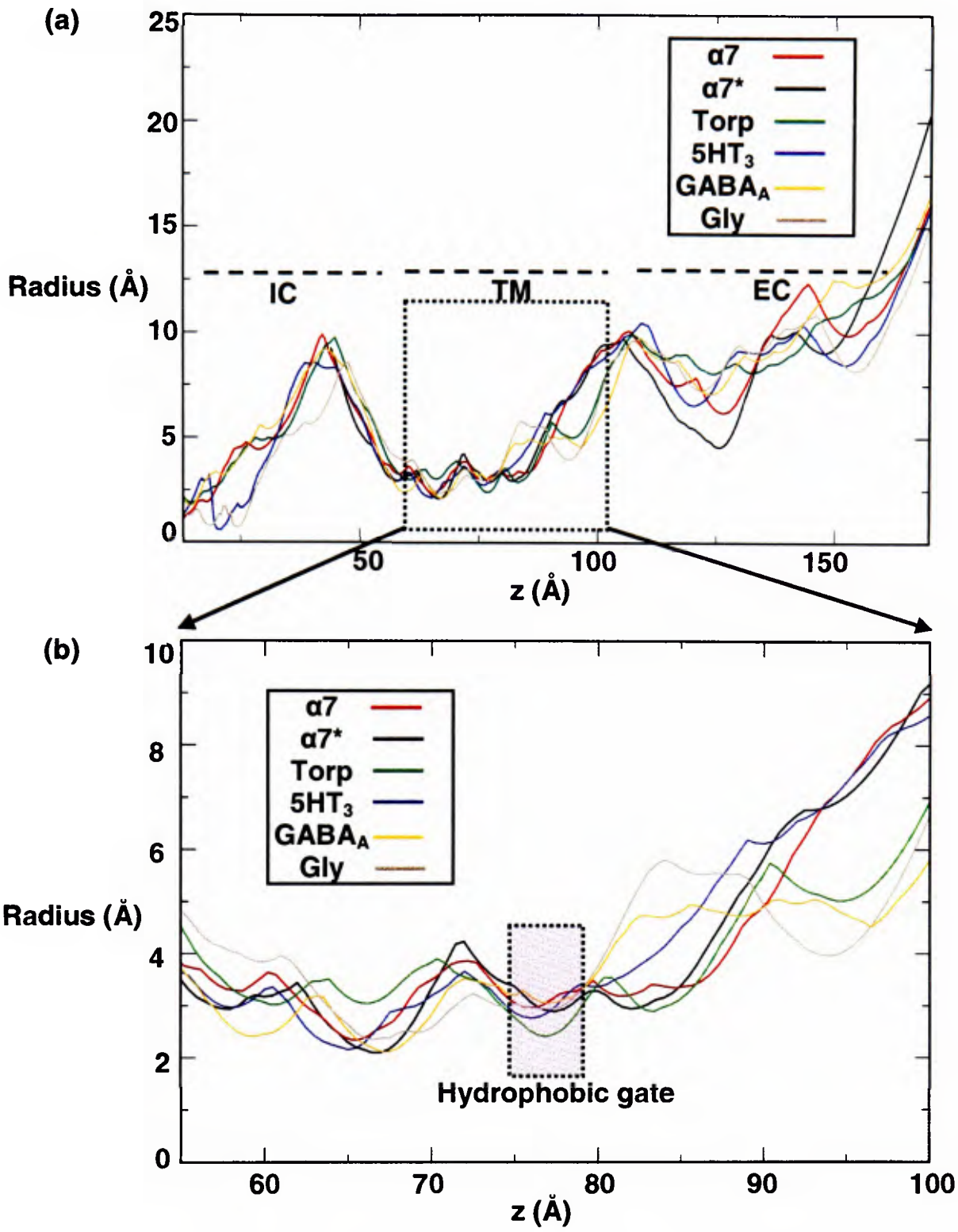


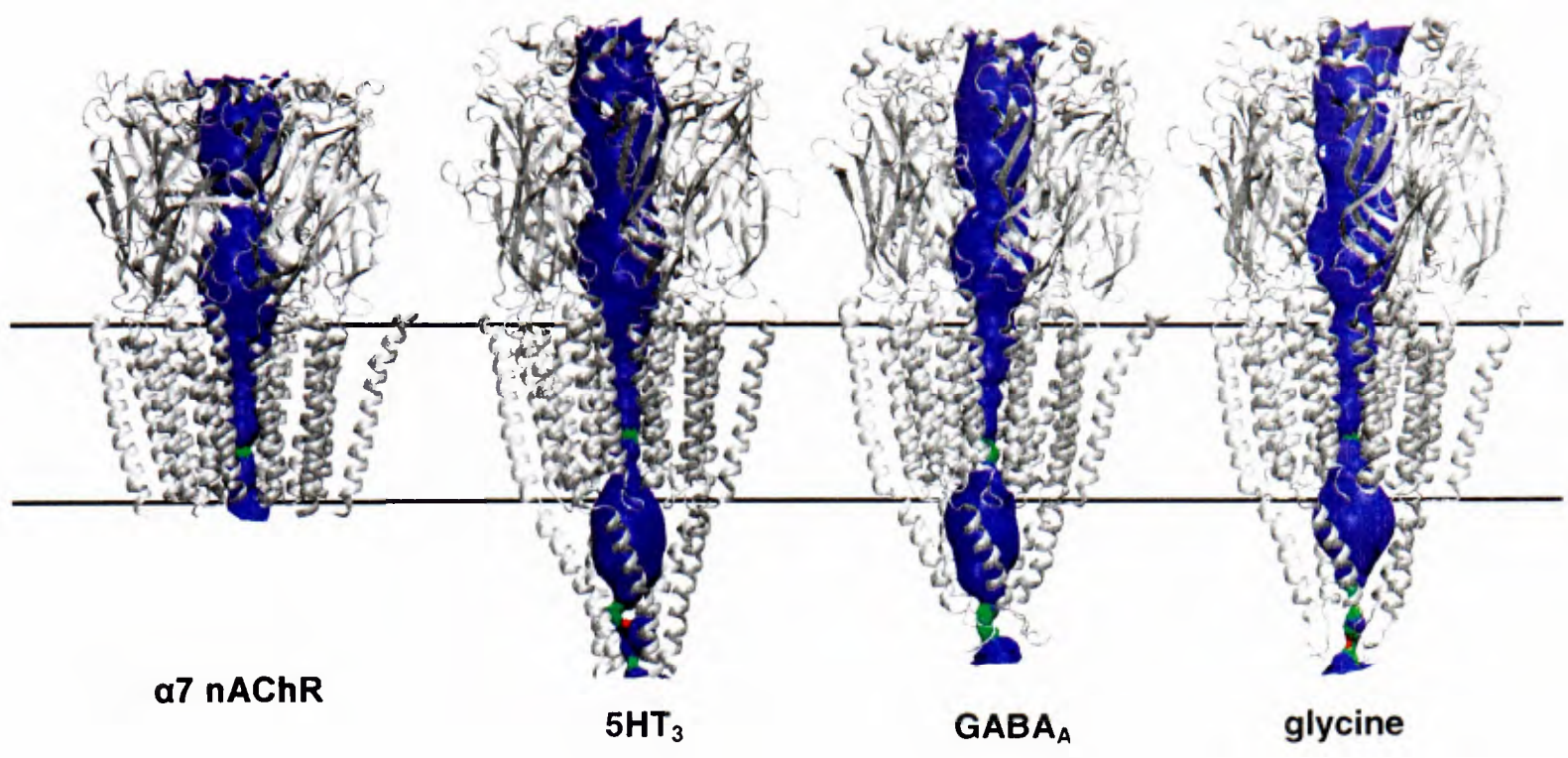
Figure 7.2 (a) Pore profiles of the LGICs showing both models of the α7 nAChR. All profiles show the highest restriction in the IC domain followed by the TM domain and then the EC domain. The α7* which was based on the AChBP has the smallest radius in the EC domain. (b) A close-up of the pore profile of the TM domain, showing the region of the hydrophobic gate, which is not necessarily the area of smallest radius. Above the hydrophobic gate, the anion-selective channels display larger radii than the cation-selective channels.

The LGICs show a highly constricted region of ~ 3 Å in the TM domain where there is believed to be a hydrophobic gate consistent with EM structural data [13, 60, 272] and this is also consistent with single channel kinetic analysis of mutational data [68, 70]. The smallest radius in the TM domain for α7* nAChR, $5HT_3$, $GABA_A$, and

glycine is found at the Thr marked in yellow on Figure 7.1 and it measures ~ 2.3 Å. Using the 'prime' numbering scheme for M2 of Lester and other [15, 51], this Thr lies at position 6' and is conserved across all of the receptors, except the *Torpedo marmorata* nAChR itself. The smallest radius in the TM domain of the *Torpedo marmorata* EM structure is more central, at the hydrophobic gate region, at residue Leu (9') where the radius at this position is ~ 3 Å. The hydrophobic gate is a stretch of residue between positions 9' to 13', as marked on the alignment on Figure 7.1. Interestingly, a number of mutation accessibility studies have argued that the gate is at the intracellular end of the TM domain pore rather than the centre [33, 45, 65-67].

There is a less pronounced constriction in the middle of each EC domain of ~ 5 Å for α7* and ~ 6.5-7.5 Å for the rest. It is important to note that the radius of a solvated Na^+ ion, allowing for just a single solvation shell, is ~ 4 Å.

Due to the number of missing residues in the IC domain, it is difficult to draw any conclusions for this segment of the proteins. However the pore profiles of the TM and EC domains suggest that the gate may be formed with a longer sequence of residues than previously thought. It may also point to the possibility that different receptors may have slightly different positions for their gate, a few angstroms apart.



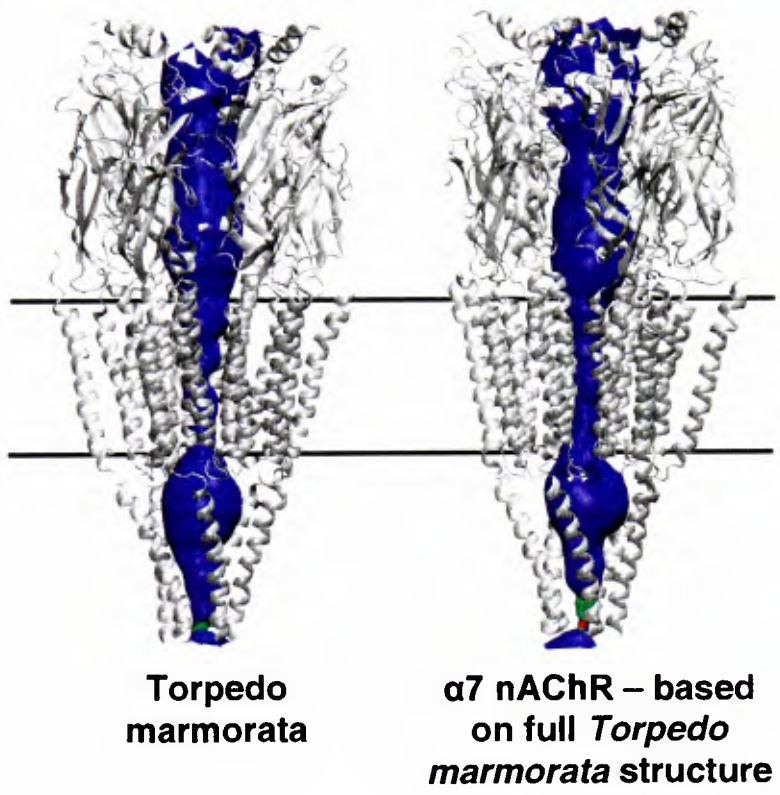


Figure 7.3 The pore surface representations are displayed for each LGIC modelled. Red = lowest radius (< 1.15 Å), green < 2.3 Å, blue = highest pore radius (> 2.3 Å). The areas of highest constriction are within the IC domain for each receptor. The TM domain also displays a high degree of constriction near or at the hydrophobic gate region, depending on the receptor.

7.3.3 APBS studies of LGICs

The Born energy of a Na^+ ion placed at successive points along the pore axis was estimated using Poisson-Boltzmann (PB) calculations, as described in [62] (see chapter 2). The models of the α7* nAChR, α7 nAChR, $5HT_3$, $GABA_A$, and glycine receptors were used to run PB calculations. The charge and box dimensions for each system are shown in Table 7.1. The Born energy of a Cl^- ion is also used for comparison purposes.

Table 7.1 The charge for each receptor as calculated by PDB2PQR

Receptor	Charge (e)
α7* nACh	-45
α7 nACh	-65
Torpedo marmorata nACh	-65
$5HT_3$	-10
$GABA_A$	+35
glycine	+10

* Each APBS run was carried out in box size of 140 x 140 x 240 $Å^3$

In each case, the pore was oriented along the *z*-axis. An ion was placed at the sample points on the pore centre line identified by HOLE. Sample points were each 0.5 Å apart along the *z*-axis. The charge on the Na^+ cation was +1 *e* with radius 1.680 Å, equivalent to the Born radius of sodium [235], and the charge on the Cl^- anion was -1 *e* with radius 1.937 Å equivalent to the Born radius of chloride [235]. In a cube with dimensions 10 × 10 × 10 $Å^3$, fine-grid focusing around the ion was used. The number of grid points was 97 × 97 × 193. The Born energy profiles for all of the proteins are shown in Figure 7.4.

In comparing the PB energies for the modelled ion channels, the highest energies are derived in the IC domain, followed by the TM domain and finally the EC domain (Figure 7.4). It has been suggested that the IC domain is involved in the lateral extrusion of ions [271]. Here, with the high Born energies and small radii of this segment, it does seem plausible that ions can escape through the sides, rather than the bottom of the IC domain alone. However, due to the missing residues in the IC domain and the lack of confidence in the alignment, it is ambitious to draw any conclusions from the APBS results of this section.

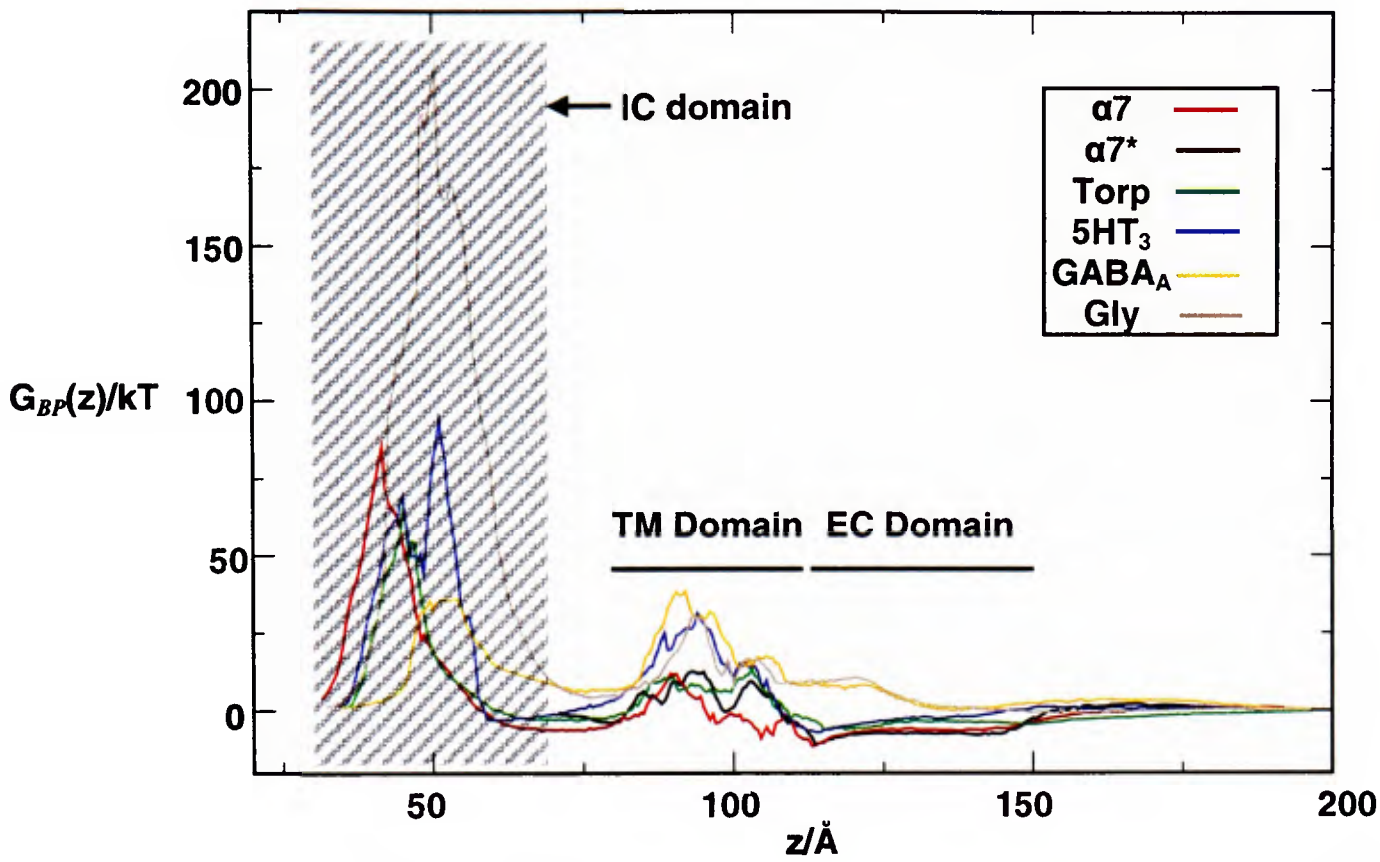


Figure 7.4 The Born energy profiles as calculated by the program APBS [182] are shown for the LGICs. The crossed region are the energies for the IC domain, however because this region has many missing residues and due to the lack of confidence in the alignment in this segment, it is best to consider the TM domain and EC domains only for this analysis. A close-up of the TM and EC domains is found in Figure 7.5.

There are very high Born energies in the IC domain, in particular for the glycine receptor. This is probably not only due to the very small pore size of that region, but also due to possible incorrect residue assignment and occlusion of the pore by loop segments. Thus, the discussion and observations for the APBS results will be limited to the TM and EC domains.

It is of interest to see whether there is a marked difference in the Born profile between the cation-selective nAChR and $5HT_3$ receptors and the anion-selective $GABA_A$ and glycine receptors, and to investigate which parts of the channel may have the greatest role in selectivity. Favourable energies are expected for the cation-selective receptors as the probe ion is Na^+.

The APBS results show that there is a distinct difference between the cation-selective and anion-selective ion channels (Figure 7.5). The glycine and $GABA_A$ receptors are showing higher Born energies in the EC domain (highlighted region) while the cation-

selective ion channels have lower (negative) Born energies. Interestingly, there are high permeation energies for all channels in the TM domain. Thus this suggests that the EC domain may be involved in ion selectivity. The other important difference is that the anion-selective channels do not have Born energies lower than 0, the energies are always above this value (unfavourable).

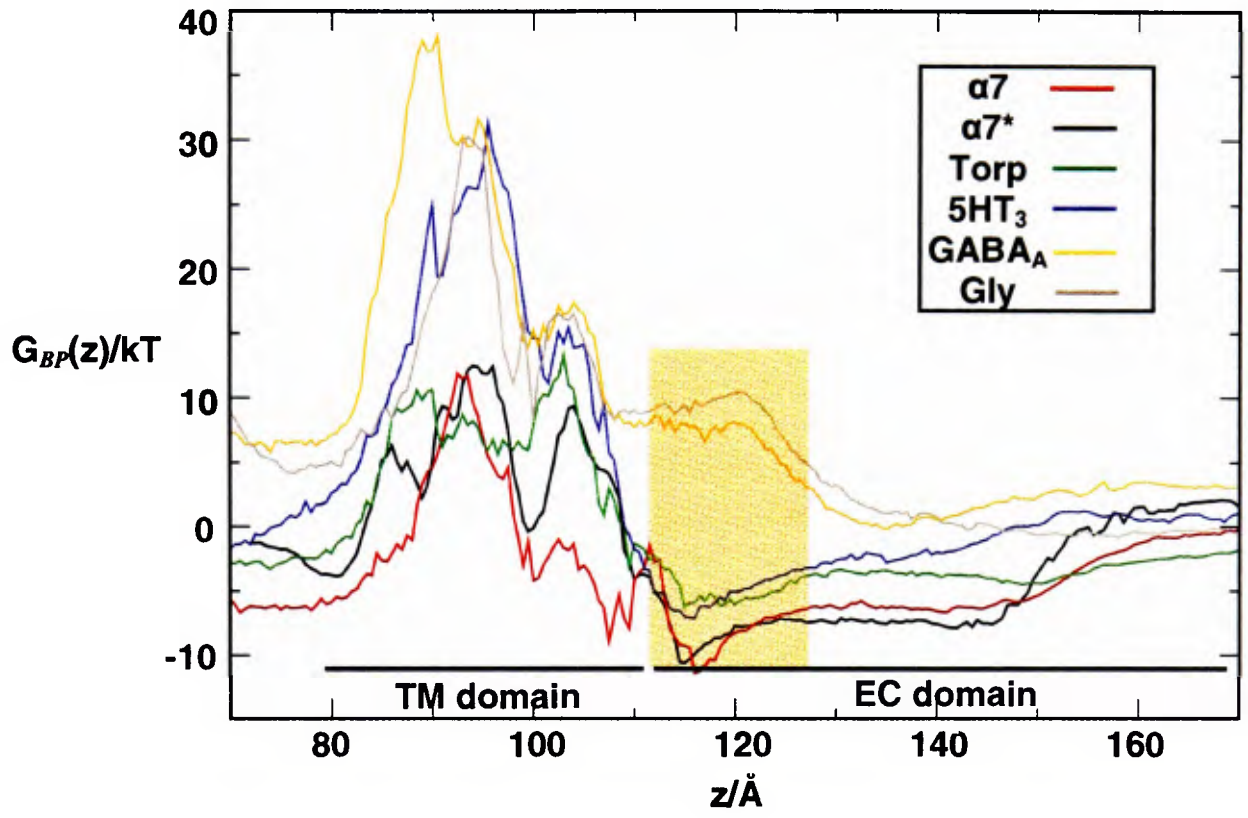


Figure 7.5 The Born energy profiles as calculated by the program APBS [182] are shown for the TM and EC domains only of the LGICs. The highlighted region shows the difference in energies between the cation-selective (nAChR and $5HT_3$ receptors) and anion-selective receptors ($GABA_A$ and glycine receptors).

The cation-selective α7*, α7 nAChR, and *Torpedo* EM structures are very close in their APBS results with almost identical maximum energy values and very close minimum energies. This group also has the lowest Born energy values of all the receptors, corresponding to the EC domain while the anion-selective receptors display unfavourable Born energies in this domain.

The anion-selective $GABA_A$ receptor has the highest Born energies in the TM domain of ~38 kT. This is followed by the glycine receptor and the serotonin receptors at ~30 kT, also, showing significantly higher PB results than the nACh receptors in this

region. There is a pronounced barrier to permeation in the constricted region in the lower segments of M2 helical bundle for all of the receptors discussed here. The highest Born energies in the TM domain are from the $5HT_3$, the $GABA_A$, and the glycine receptor, ~20+ kT greater than the others.

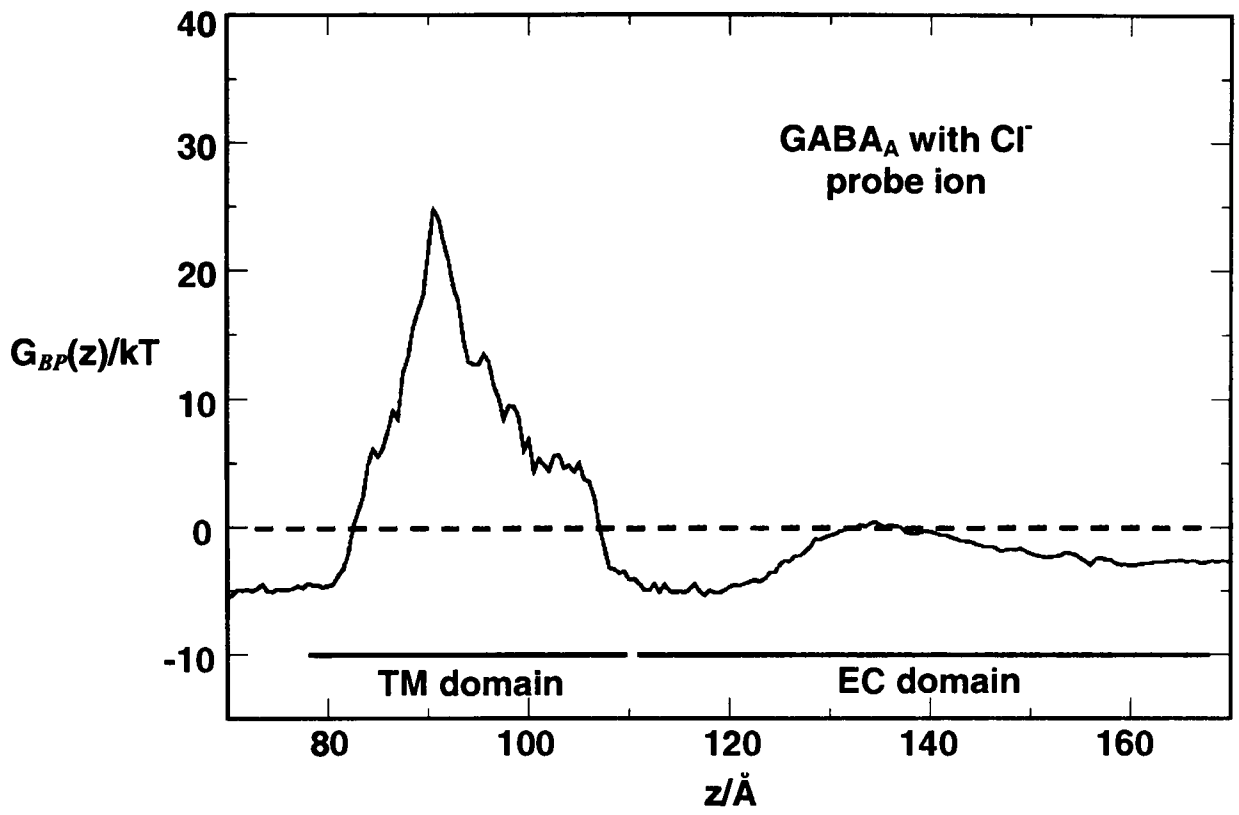


Figure 7.6 Born energy profile for the $GABA_A$ receptor using a Cl^- ion. Here there are negative (favourable) Born energies in the EC domain, permeation was unfavourable for a Na+ ion in this domain. There is still a barrier in the TM domain of the receptor, however it is not as high as with the Na^+ ion.

To further investigate the gate in the TM domain of LGICs and the role of the EC domain in selectivity, APBS was carried out with a Cl^- ion. Encouragingly, (Figure 7.6) there are favourable Born energies in the EC domain of the $GABA_A$ receptor with an energy of ~ -5 kT in the lower segment of the EC domain (just above the TM domain). This matches the Born energies of the cation-selective *Torpedo marmorata* AChR and the $5HT_3$ receptor in that region, and is slightly higher than the α7. There are still high Born energies in the TM domain, which demonstrate that regardless of channel selectivity and ion type there is a barrier to permeation in the TM domain.

Contrastingly, the EC domain displays distinctly different Born energies between the two channel types, thus indicating its involvement in selectivity.

In Figure 7.7, each protein is coloured according to its APBS value with Na^+ as the probe ion. For comparison purposes, all the values are scaled so that they share the same range.

It is evident that the high Born energies for $5HT_3$, $GABA_A$, and glycine receptors are at the area of smallest radius, and lower than the hydrophobic gate locus. In general, the TM domain has greater Born energy values than the EC domain. The more positive energies are in the lower TM region while the top portion of the TM domain has lower Born energies. The EC domain Born energies, which segregate the cation-selective and anion-selective channel types, display higher Born energies in the bottom segment of the EC domain, just above the TM domain, for the anion-selective receptors. Thus there is an energetic barrier for cations before reaching the TM domain. However with Na^+, for the cation-selective receptors, there is a stretch of low Born energy values from middle of the EC domain to the mid-section of the TM domain. The $GABA_A$ and glycine receptors have higher Born energies for a Na^+ ion across the EC domain, which suggests that the EC domain plays a role in ion selectivity.

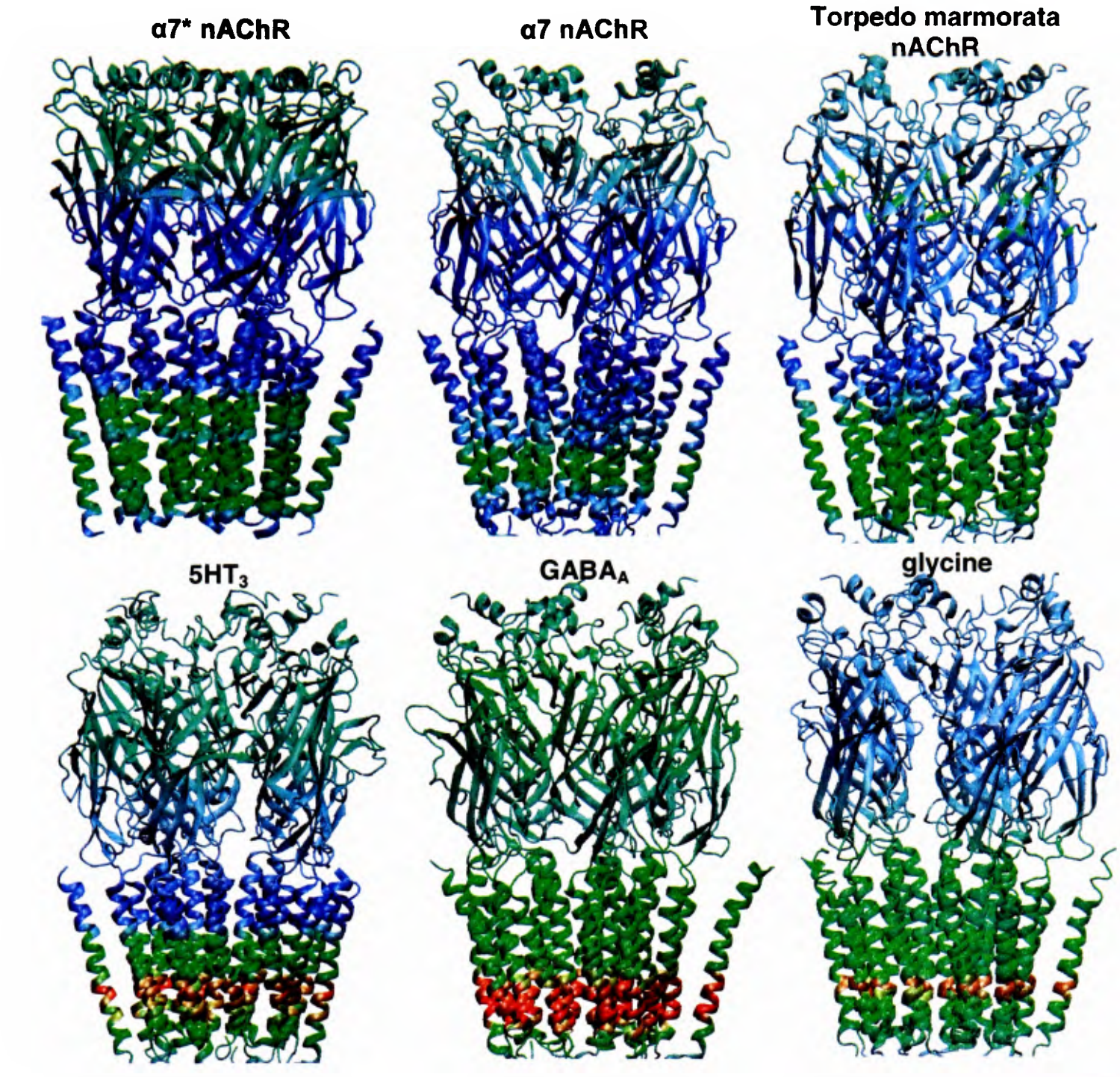


Figure 7.7 Each receptor is coloured by its Born energy values for Na^+ permeation. The colour range is normalized with respect to the Born energy range for all the channels. The highest Born energies are in the lower part of the TM domain, followed by the top segment of the EC domain.

7.3.4 Coarse-grain MD of α7 nAChR and other LGICs

Coarse-grain (CG) molecular dynamics simulations were carried out on the LGICs discussed in this chapter. The aim was to observe the self-assembly of lipids around each channel and to study the lipid-protein interactions in each case. As discussed before, atomistic simulations would not allow for the observation of this behaviour due to the very short time-scale that is feasible with atomistic MD. Also, the large

system size makes these proteins good candidates for coarse-grain simulations. The protein was solvated and lipids were placed randomly in the simulation box. Each system was energy minimized for >1000 steps using the steepest descent method. This was followed by 200 ns of simulation with position restraints with force constant of 1000 $kJmol^{-1}nm^{-1}$ on the Cα atoms of the protein, which allowed for the self-assembly of lipids around the protein. The bilayer formation took approximately 20 ns in each case (see Figure 7.8). The position restraints were then removed for >100 ns of simulation. During the lipid self-assembly phase for all of the systems simulated, it was found that a group of lipids aggregate at the top of the EC domain. This is an artefact of the lipid-assembly process which does not occur *in vivo* as there is no lipid self-assembly around a folded channel; the bilayer is preformed when the subunits are inserted into the membrane. Thus the analysis was restricted such that these 'non-biological' lipids were discounted from the results. Figure 7.8 shows the first 20 ns of α7 CG simulation with all the Cα atoms restrained. During this phase of the simulation, the lipids quickly aggregate and assemble a lipid bilayer around the receptor.

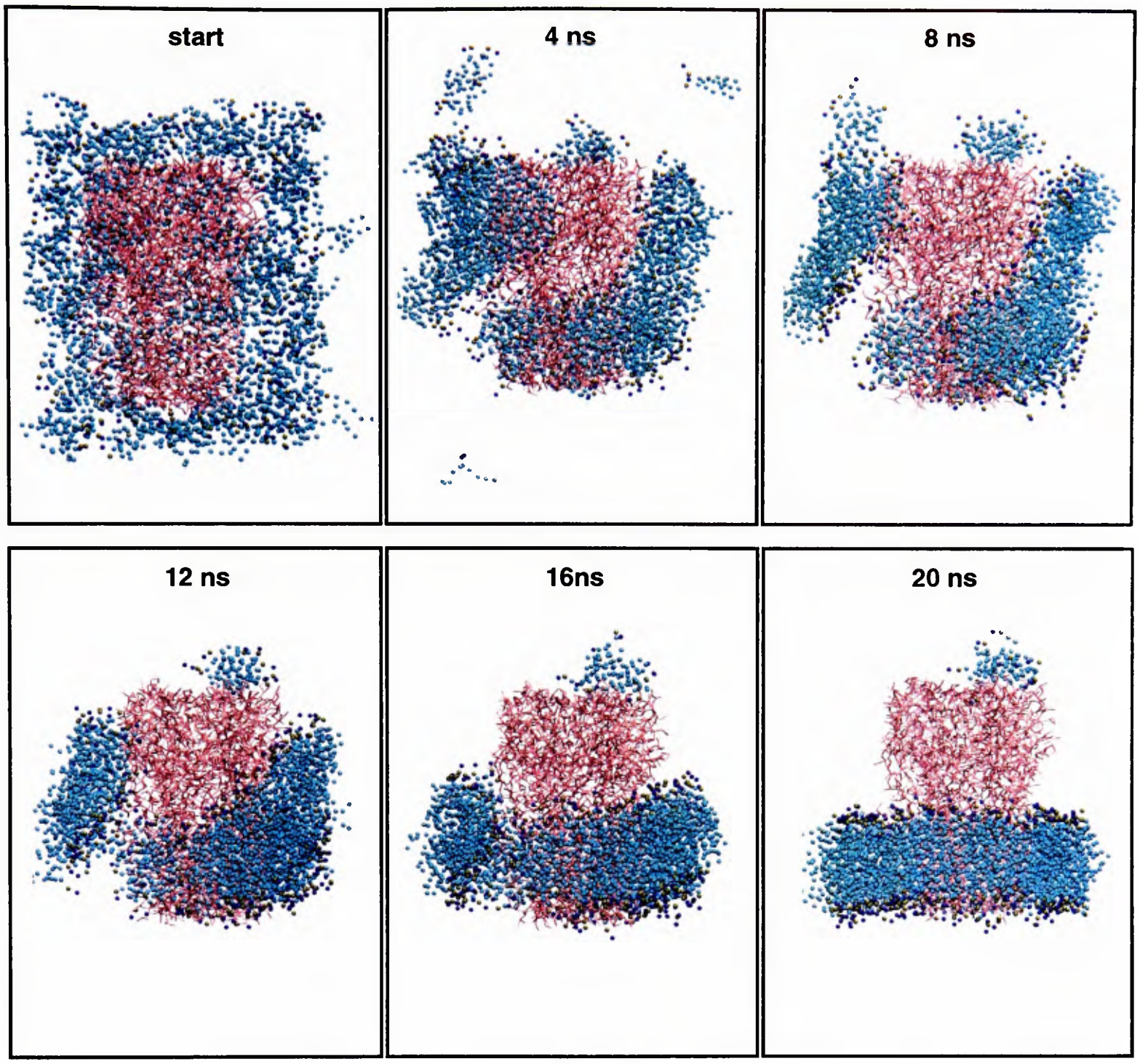


Figure 7.8 The first 20 ns of the α7 nAChR coarse-grained simulation showing bilayer assembly around the receptor. The bilayer is completely formed in the first 20 ns of the simulation. A group of lipids tend to aggregate at the top of the EC domain where there are several hydrophobic residues. This is an artefact of the simulation as the bilayer is pre-formed *in vivo*.

The small clump of lipids is also shown in Figure 7.8, this artefact is present for all of the CG simulations carried out (i.e. for $5HT_3$, $GABA_A$, glycine, and for the *Torpedo marmorata* nAChR). In addition to the micelle formation at the top of the EC domain, there has also been bicelle formation. These micelle and bicelle aggregates remain in the system and refrain from joining the lipid bilayer around the protein, thus few lipids remain to form a complete bilayer. The water to lipid ratio is a key element of successful lipid bilayer assembly around a protein in the CG simulations.

Therefore, when there is excessive micelle or bicelle formation, changing the water to lipid ratio may result in better bilayer assembly.

To compare the CG simulation residue fluctuations with an atomistic simulation of a homologous protein, RMSF calculations for the α7* nAChR (whose EC domain is based on the AChBP) and the simulation of AChBP with nicotine bound (no TM domain) were carried out. Figure 7.9 shows the RMSF calculations for one subunit of each system.

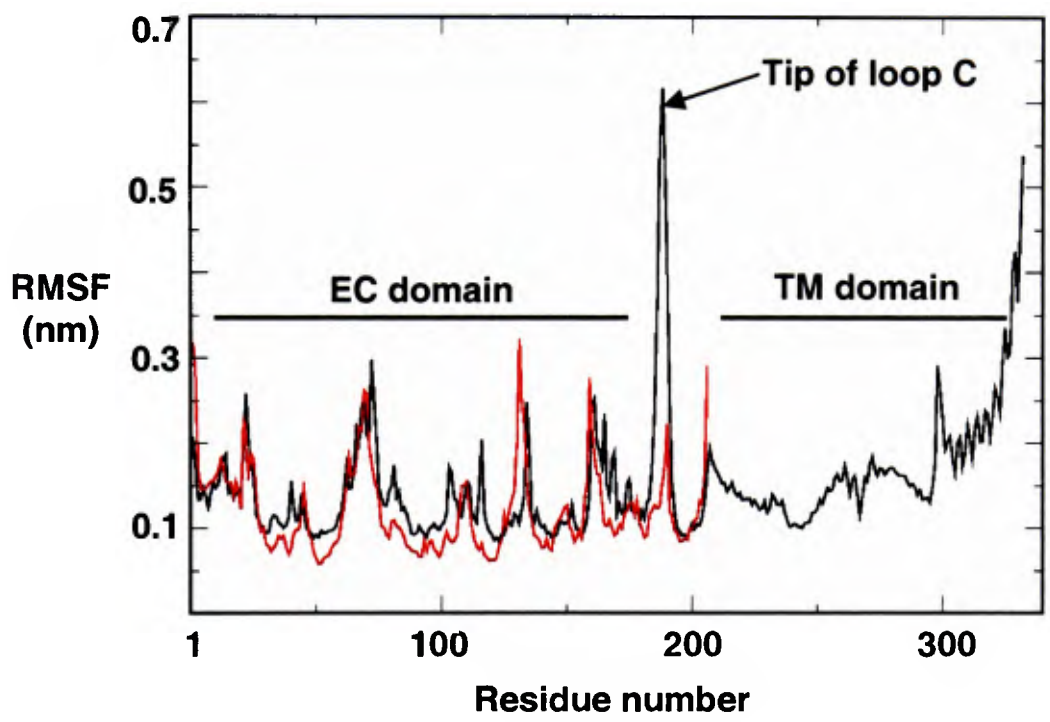


Figure 7.9 RMSF calculations for one subunit of the α7* nAChR CG simulation (black) and the AChBP atomistic simulation (red) show the very close values for the fluctuation of Cα atoms for these two structures. This demonstrates that the CG model is reproducing the same motions as the more detailed atomistic MD method. The large fluctuation of loop C from the CG simulation may be matched by the atomistic simulation if longer time-scales were possible. Also the presence of the bound nicotine in AChBP, restraints the movement of loop C (see chapters 3 and 4).

The RMSF fluctuations are very close for the CG and atomistic simulations, except for one atom in the CG calculation at the tip of loop C. Other sections are in good agreement with each other, thus illustrating that the protein in the CG simulation is behaving similarly to more accurate atomistic MD.

An in-house script was used to record all the residues interacting with lipids in each frame of the simulation. This allows the identification of residues interacting with lipids and also reveals the exact location of the bilayer. Lipid-protein interactions of the nAChR with its phospholipid bilayer have been shown to be important for its structural stability and function [273-277]. Thus we can identify exactly which residues are most involved in interacting with surrounding lipids during a CG MD simulation. This analysis is shown for the α7* nAChR model and the $GABA_A$ model representing a cation-selective and an anion-selective Cys-loop receptor. For the α7 nAChR the first 100 ns with position restraints on the protein are shown (Figure 7.10) in order to illustrate the rapid lipid aggregation and bilayer assembly around the receptor.

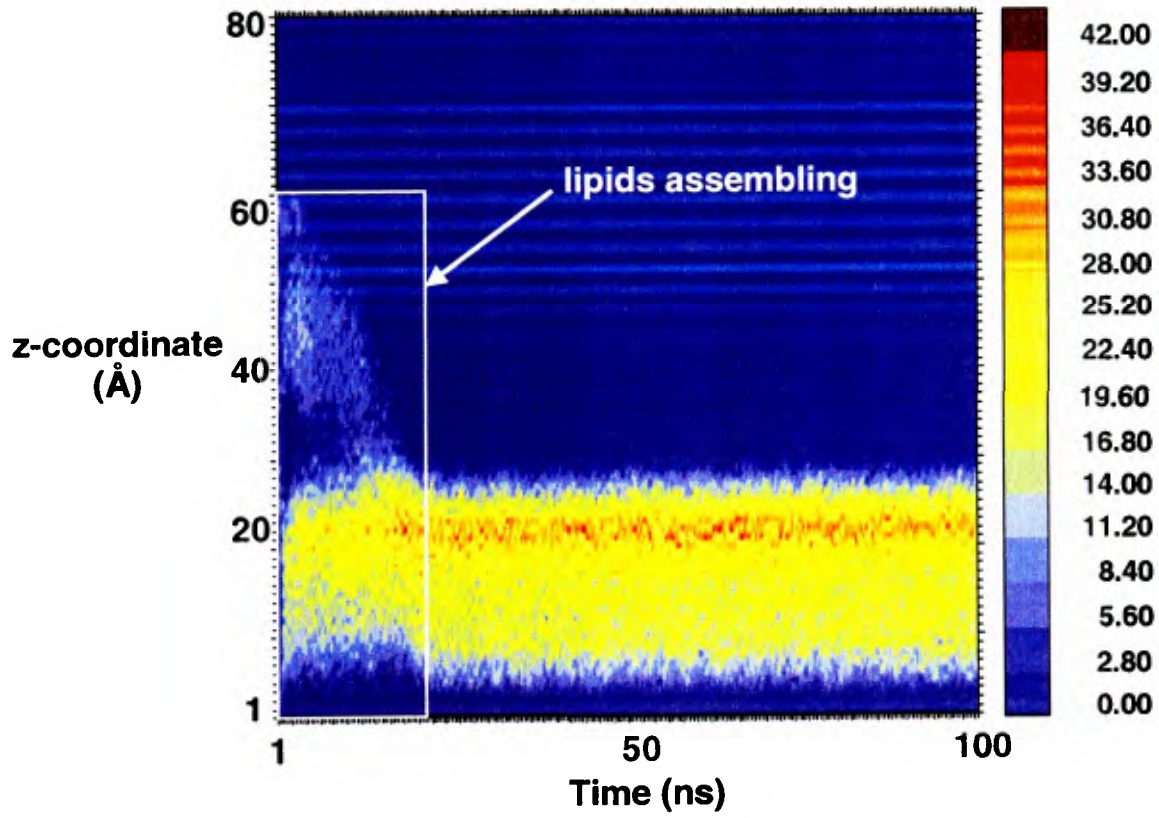


Figure 7.10 Lipid-protein interactions for the α7 nAChR CG simulation. The number of protein-lipid interactions is plotted for each frame (x-axis) across the z-coordinate (y-axis). In the first several nanoseconds of the simulation, the lipids quickly assembly around the TM domain and retain their position.

Figure 7.10 shows the lipids assembling to form a lipid bilayer around the α7 nAChR. In the first ~50 frames (20 ns), the lipids aggregate and the rest of the time they remain around the protein. In Figure 7.10, the most red region of the lipid band represents a small segment of the M4 which has the highest lipid interaction. The interactions are less intense towards the bottom ends of the lipid-protein interaction band as some green and blue starts to appear. α7* nAChR is coloured by the number of lipid interactions in Figure 7.11. It is evident that segments of the M4 make the most contacts with surrounding lipids. It is interesting that the middle section of M4 has the highest number of lipid-protein interactions, which is also the case for $GABA_A$, also shown (Figure 7.13). The M4 has been the focus of many experimental and MD studies which show that the M4 helical segment has the largest contacts with the lipid bilayer [278-280]. Mutational studies of M4 residues interacting with the lipid bilayer suggest the alteration of channel gating and kinetics [281-286].

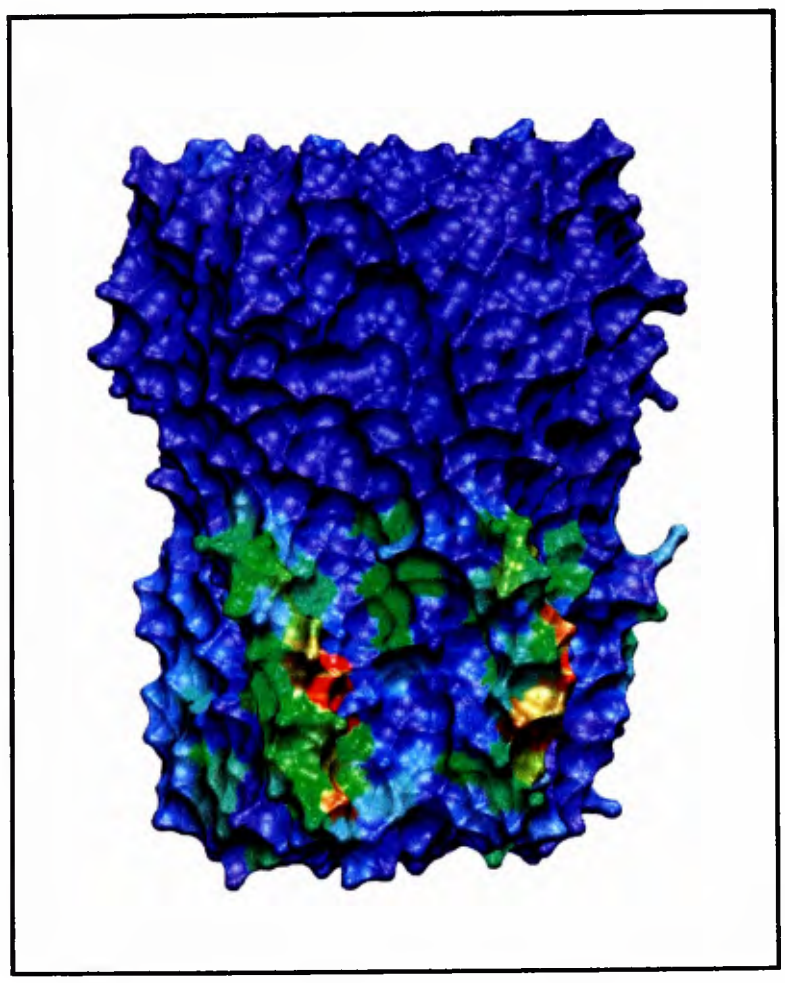


Figure 7.11 The α7 nAChR residues are coloured according to the number of interactions with lipids in the bilayer. Blue indicates the least interaction while red indicates high lipid-protein interactions. Note that the colour-scale used for the lipid-protein interaction plots (Figure 7.10) is not the same as those used to colour the protein. The protein is coloured according to the number of contacts each residue makes with lipids for a defined segment of the simulation.

The residues with the highest number of interactions with lipids are in the mid-section of the M4 helical stretch for all of the simulations. If we name the M4 stretch with x" (double-prime) numbering starting with AMV (*Torpedo marmorata* α subunit – see Figure 7.1), the Ile 17" and Cys 18" for α7* nAChR has the highest number of protein-lipid interactions and the corresponding Ala 17" and Thr 18" for $GABA_A$ show the highest interactions with lipids in the bilayer (Figure 7.13). Figure 7.12 shows the lipid-protein interactions plot for $GABA_A$. There are slightly more lipid-protein interactions towards the lower end of the M4 helix with lower interactions towards the top of the helix. There are also many lipid-protein interactions with M1 and M3, although the M4 has the highest number due to the fact that it is most exposed to lipids.

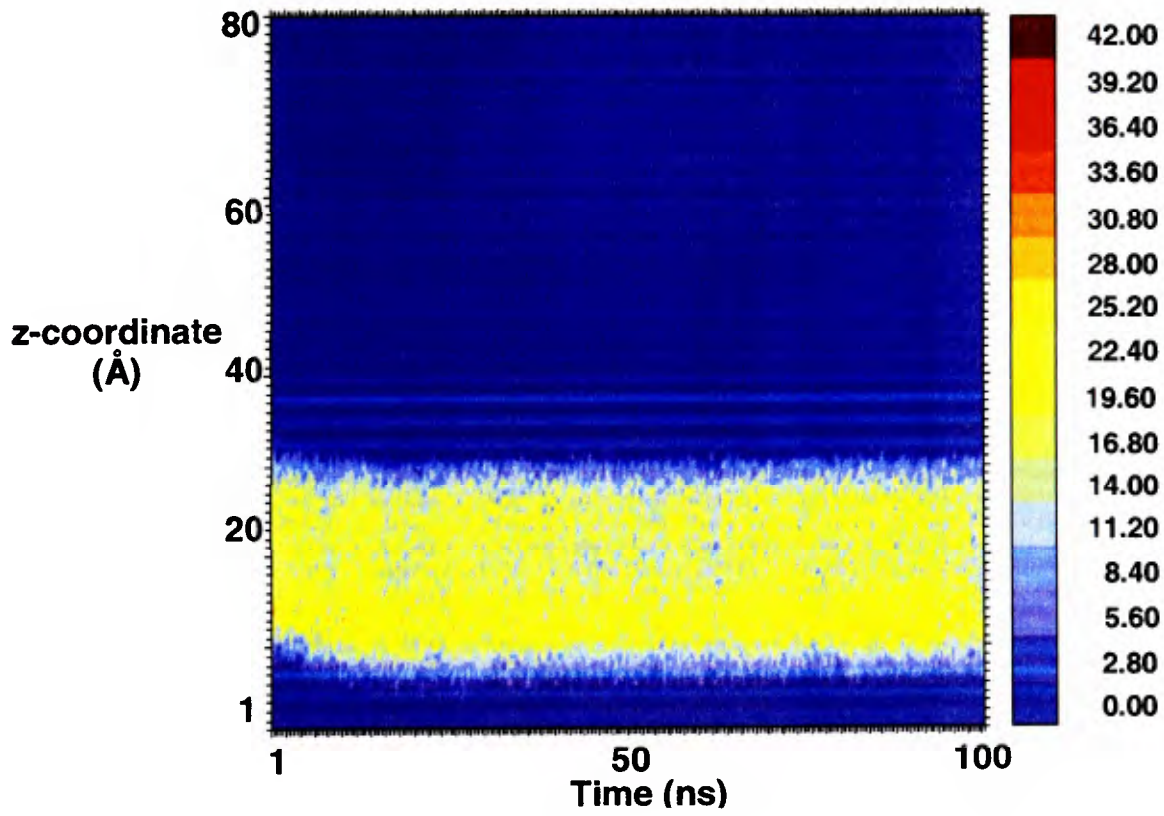


Figure 7.12 The $GABA_A$ receptors interactions with lipids are shown for a 100 ns stretch of the CG simulation. The lipids surround the TM domain.

The lipid-protein interactions are fewer towards the top segments of the M4 for the $GABA_A$ receptor. This slight difference in the location of the highest number of

interactions may be dependent on the sequence alignment, as the sequences are quite different in the M4 helical segment of the protein. The highest number of lipid-protein interactions for $GABA_A$ is towards the lower M4 helix.

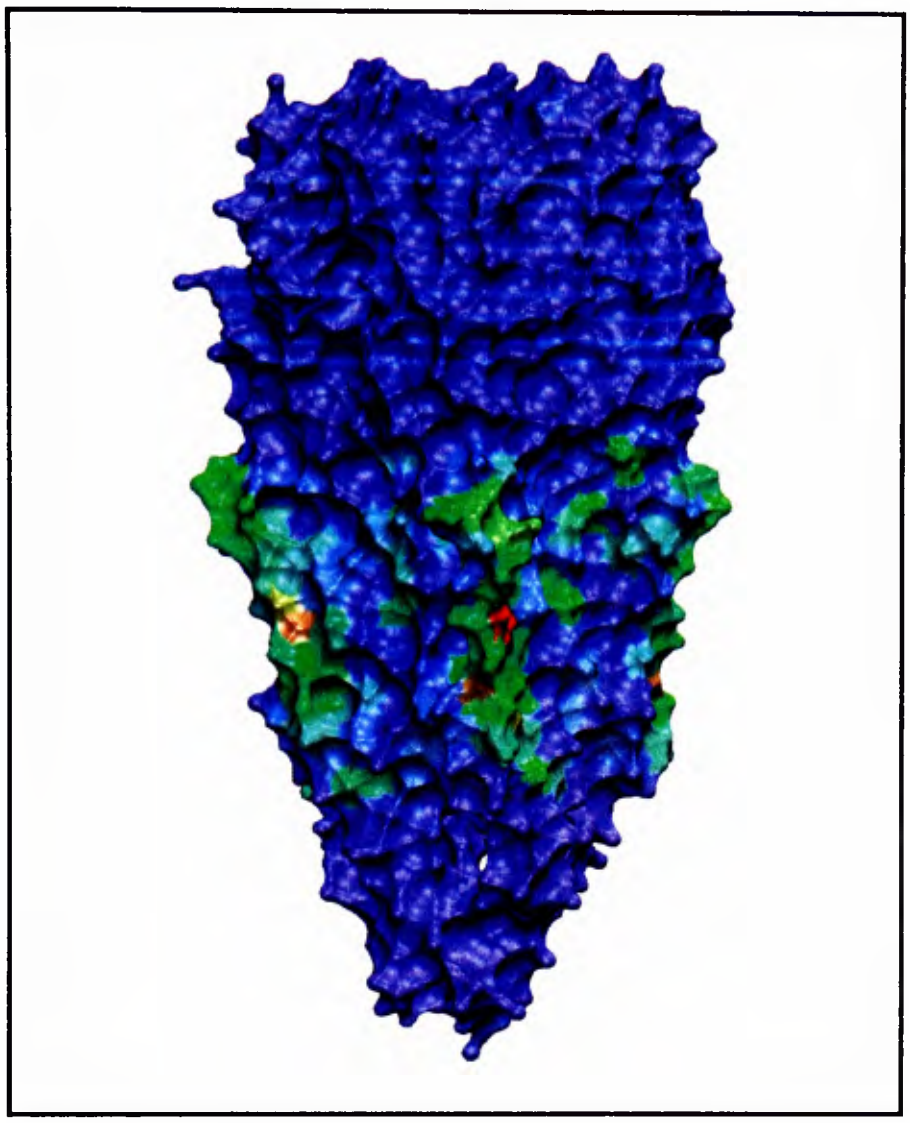


Figure 7.13 The $GABA_A$ receptor residues coloured by their number of interactions with surrounding lipids. As in the α7 nAChR, the lipids have aggregated around the TM domain. Blue indicates the lowest number of protein-lipid interactions while red indicates the highest number. As for the α7 nAChR, the highest number of lipid protein interactions are in the M4 segment.

7.4 Discussion

The homology modelling of the Cys-loop family of receptors allowed for the study of the electrostatics, pore profiles, and coarse-grain MD simulations of these systems. The *Torpedo marmorata* nAChR structure [19], which includes all three domains of the receptor, was used as the template to model four other Cys-loop receptors which share a highly conserved $CX_{13}C$ motif in the EC domain near the TM domain. These receptors share high structural and functional homology, although their sequence identity is not high amongst all in the group. The goal was to understand how the

sequence differences result in differences in selectivity and to gain insight into the function through the study of conformational dynamics, lipid-protein interactions, pore properties, and selectivity.

The template structure used for the homology modelling is missing several residues in the IC domain and the relatively low sequence identity (~10-35%) between some of the receptor types makes homology modelling a challenge, in particular in the M3-IC-M4 segment where the sequence identity is particularly low. Never the less, generating the homology models and carrying out analysis would lead to some answers about the plausible mechanism of function for this group of receptors.

7.4.1 Pore profiles

The pore profiles showed a very similar pattern of pore radii along the ion channel pore for each protein. The highest constriction was in the IC domain, this was followed by the TM domains, and finally the EC domain. In most cases, the constriction in the TM domain was just below the hypothesized 'hydrophobic gate', suggesting that there could be longer stretch of residues which constitute a physical 'gate' expanding beyond the hydrophobic girdle. Although the lack of structural data in the IC domain hampers attempts to form definitive conclusions on the pore size, in all of the modelled channels the highest restriction was at the base of this domain. It has been suggested that there may be lateral extrusion of ions from the IC domain [271], which seems plausible with the pore-profiles of the generated homology models.

7.4.2 APBS

Significant differences for ion conduction were observed between the cation-selective and anion-selective channels in the EC domain suggesting that this portion of the ion

channel pore is involved in selectivity. For Na^+, the anion-selective channels did not have Born energies lower than zero while the cation-selective receptors had negative Born energies (favourable). The segments in the TM domain with the highest Born energies corresponded mainly with the most constricted regions in the pore profile analysis and were not necessarily at the hydrophobic gate. In all receptors, the lower segments of the TM domain had showed the highest barrier to Na^+ permeation, with $GABA_A$, glycine and then the $5HT_3$ receptors having the highest Born energies. It is important to keep in mind that the template structure used to model these LGICs most closely represents a closed state of the receptor [19]. The Born profiles for the IC domain are ignored as there are many missing residues in this region and thus it is questionable to make a strong conclusion for this part of the protein. The high Born energy for this segment is partly due to the small pore radius and also the possible error in the original alignment. With Na^+, there are higher Born energies in the top segment of the EC domain for the cation-selective receptors and the lower part of the TM domain. Thus there is a section of lower Born energies mid-way down the EC domain to the middle section of the TM domain. Promisingly, using Cl^- as the probe ion showed favourable energies in the EC domain of anion-selective channels, further suggesting that this domain is involved in selectivity. With Cl^- there was still a large permeation barrier in the TM domain of the LGICs, indicating that there is a barrier in this domain, regardless of ion type.

7.4.3 Coarse-grained MD simulations

Lipid assembly was observed within the first 20 ns of CG MD simulation for all of the LGIC systems discussed. This allowed for the observation of the position of each receptor within the lipid-bilayer and the study of residues most involved in lipid-

protein interactions. Lipid-protein interactions with the nAChR have been shown to be important in the structural stability and function of this channel. Residues in the M4 helical segment of each receptor displayed the highest protein-lipid interactions in agreement with experimental data [273-277]. Analysis of the CG MD simulations demonstrated that these interactions were highest in the middle region of the M4 helix. The M4 has been the focus of many experimental and atomistic MD studies suggesting that it has the largest contact with lipids in the bilayer and mutations of lipid-facing residues alters channel function [279-281, 284-286].

The CG simulations can be improved by optimizing lipid-water ratios such that micelle and bicelle formation is minimized. Micelle formation can also be the result of a bad alignment in the TM domain. As discussed earlier it was difficult to have confidence in the alignment for the M3-IC-M4 segment, and because the M4 is crucial in bilayer assembly, it may be that the homology model needs to be optimized.

Although the global conformational dynamic properties of proteins (e.g. flexibility) from the CG MD simulations closely match those of the atomistic simulations, the details of the dynamic behaviour in CG simulations have not yet been assessed adequately. Thus conclusions on the dynamic behaviour leading to function cannot be drawn based on the current CG model, although the method holds promise and is being improved for the analysis of conformational dynamics for much longer time-scales. It can eventually be used to observe functionally important events such as gating.

8 Conclusions and possible future directions

The work described in this thesis addresses several aspects of ion channels, such as selectivity, ligand-protein interactions, small-scale and global dynamics, and methods development to facilitate the study of ion channels and other membrane proteins. Here, the different aspects of ion channels (as discussed in the introduction) are addressed, this time including the results and conclusions drawn from this thesis.

8.1 Selectivity

The ability of an ion channel to discriminate for a particular ion type was addressed with electrostatic studies using APBS of full channel models using the *Torpedo marmorata* structure [19]. Most studies thus far had focused on the TM and more recently the IC domains. In agreement with functional data, the highest Born energies were observed in the IC domain, followed by the TM domain. However, the IC domain in this study was largely ignored due to the lack of confidence in the correctness of the models in that region as there are many missing residues in the EM

structure on which these models were based [19]. The Born energies of both the cation-selective and anion-selective receptors were relatively high at the intracellular end of the TM domain with lower energies towards the top of this domain. It is important to note that the template structure used to model these channels is believed to be in a closed state. There were key differences in the Born profiles between the two receptor types in the EC domain thus suggesting that the EC domain may have an important role in selectivity. The Born energies are higher in the bottom half of the EC domain for the anion-selective channels. With Na+ as the probe ion, the anion-selective channels showed unfavourable Born energies, all above 0 for all the segments of the channels, while the cation-selective channels showed negative energies in the EC and some parts of the TM domain. For comparison purposes and to further investigate the permeation properties of the TM domain, Cl^- was also used as the probe ion. Here, the anion-selective channels showed favourable energies in the EC domain while there was still a barrier to permeation in the TM domain. This suggests that while there is a barrier to permeation in the TM domain of LGICs, regardless of the ion species, the EC domain has a role in selecting the ion for conductance. Potential of mean force (PMF) calculations (see chapter 2) of the ion pore for the EC domain of LGICs would be useful in providing a more realistic representation of the energetics of an ion at a particular point in the pore.

8.2 The gate and gating

The complete α7 nAChR was constructed and used for structural studies in chapter 5, in particular pore-profiles and electrostatic studies of the TM domain and also the whole model were carried out to gain insight into the gate of the receptor. The most constricted region in the TM domain is just below the hydrophobic girdle and there

were high Born energies at the region of smallest radius and that of the hydrophobic girdle.

In Chapter 7, the LGIC structures based on the complete *Torpedo* nAChR were used for pore-profiles and electrostatics calculations. Pore-profile analysis of the LGICs illustrated a consistent pattern amongst all of the receptors studied with the highest constriction in the IC domains of all of the receptors, followed by the TM domain and finally the EC domain. It has been suggested that the IC domain may be involved in the side-ways ejection of ions [271], which seems plausible with the pore-profile analysis carried out here. However as the modelling in the IC segment is speculative due to the lack of structural data, the focus will be on the radii of the TM domain segments. The pore profiles show the highest constriction in TM domain is not always at the hydrophobic girdle; it can be below the gate, thus suggesting that a longer segment of the TM domain may form the 'gate'. This is in agreement with the APBS studies that show high energies at and around the hydrophobic gate.

Results from the combined model of the α7 nAChR from chapter 5 and the new model based on the *Torpedo marmorata* structure from chapter 7 show the highest Born energies in the hydrophobic gate region in the TM domain. This is also the case for the other receptors under study in chapter 7 (ignoring the IC domain). The hydrophobic residues in the pore lining helices of the TM domain make it unfavourable for ions to pass through. Although the pore-size is small at the hydrophobic girdle (note models are based on a closed structure), it is still large enough for ions to pass.

8.3 Ligand binding site

8.3.1 Binding pocket dynamics

The ligand binding site was a major focus of this work; both the dynamics and the ligand-protein interactions were studied in detail. Conformational dynamics studies of the α7 nAChR in chapter 5 showed higher flexibility of loop C of the binding pocket compared to the rest of the structure. Atomistic MD studies of AChBP (chapter 3) showed that RMSD values for residues in the binding pocket in the ligand-free state were higher than the RMSD values for ligand-bound simulations. Inter-residue distance and side-chain dihedral calculations of binding pocket residues showed higher fluctuations for ligand-free simulations as well. The MD studies suggest that the binding pocket is more flexible when no ligand is bound. These were extended to look at ligand docking into the binding site along trajectories of 10 ns duration. Ligand binding deteriorated in the simulations which were run without a ligand, partly due to the higher flexibility of loop C and the distortion of the binding pocket in general. Thus the structure of the binding pocket is different in the presence and absence of a ligand. In chapter 4, the improvement of docks along a ligand-bound trajectory suggests that binding pocket conforms to the bound ligand and re-docking ligands is conformationally facilitated. Furthermore, docking back a ligand which matches the ligand in the simulation is more favourable, thus suggesting that the binding pocket conforms to the bound ligand and is accommodating to its re-binding or the binding of those ligands of similar structure, which suggests an induced-fit mechanism [287] for ligand binding .

8.3.2 Water

Persistent water molecules were observed in the binding pocket, which are involved in both ligand-protein interactions and in the structural integrity of binding site and the protein. Water molecules were found to bridge ligands to residues in the binding pocket; these waters have higher densities and were sustained for longer periods in ligand-bound simulations. All the zones identified were occupied with water molecules for ≥ 40% of the time (i.e. zone 1 (loop E) for ≥ 92% of the simulation's duration) and commonly water molecules competed for the identified zones in the binding pocket, further suggesting that water in those locations has functional significance.

In Figure 8.1, zone 1 is shown in loop E of AChBP. It is a small distance away from the functionally important Cys loop which comes into close proximity of the TM domain.

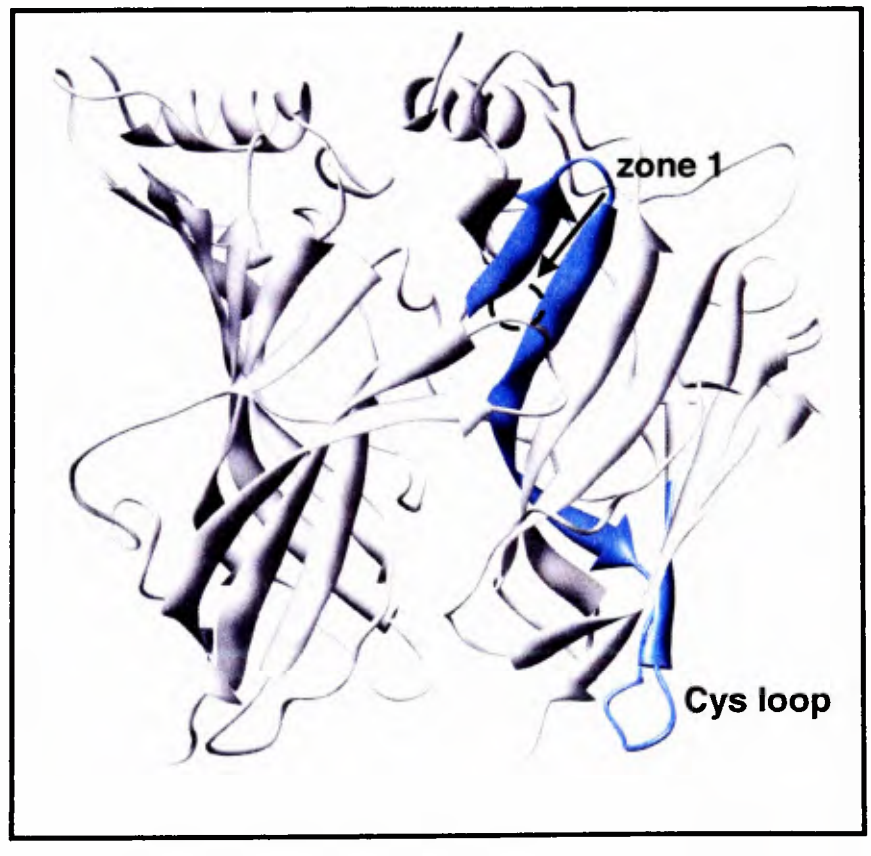


Figure 8.1 Two subunits of AChBP are shown with zone 1 (in loop E) labelled. This zone, which is occupied by water molecules > 92% of the duration of ligand-bound simulations (10 ns), is a short distance away from the functionally relevant Cys loop.

Waters are also involved in hydrogen bonding networks in the binding site that could be responsible for the structural rigidity of the protein in the ligand-bound state. The hydrogen network of these persistent water molecules may affect the structure of these loops and be involved in the gating mechanism. For example loop C is pulled closer into the binding pocket via H-bonds through bridging water molecules to the ligand.

8.3.3 Mutants in the binding site

The L118 mutant of α7 nAChR resulted in both global conformational change (intra and inter-subunit distortion) and changes in ligand binding. Acidic mutants to L118 of chicken α7 nAChR improved ACh binding and basic mutants resulted in the inability of ACh to bind correctly to the binding site. Here, a key residue in loop E is identified by computational methods (homology modelling, MD simulation, and docking) and the observations were confirmed by our experimental collaborators.

Another α7-like nAChR model, the ACR-16, was generated by the method described in chapter 5 which was the first full model of this receptor. The modelling process for ACR-16, based on AChBP, resulted in a significant observation regarding side-chain orientation of binding pocket residues. Two different models, with slightly different conformations of the His 112 side-chain in the binding pocket, produced very different docking results, one model being unable to reproduce the correct docking mode of both nicotine and acetylcholine. Thus this suggests that ligand docking is a very sensitive event, dependent on even the smallest difference in binding pocket residue side-chain conformations. His 112 is currently being investigated by our experimental collaborators. This observation also alludes to the challenges of homology modelling. It is important, yet difficult, to ensure the correct side-chain

orientations of the generated models. Using a structure of high resolution is helpful and perhaps testing a model using docking should be taken into consideration [288]. ACR-16, which is a Levamisole-insensitive ion channel, was used for docking studies with various ligands including Levamisole. Also Levamisole was docked onto Levamisole-sensitive α7-like models of unknown subunit composition (UNC-63, UNC-29, and LEV-1), this may also help in identifying possible subunit compositions. Surprisingly, Levamisole docked successfully onto the Levamisole-sensitive and insensitive channels, illustrating that a channel's affinity for a ligand is independent of the channels sensitivity to it which is the case with agonists and antagonists which can both bind, but the effect is very different. It was also observed that the homo-pentameric UNC-63/UNC-63 subunit composition is most successful in docking studies, which agrees with experimental observation and suggests it may be a functional in its homomeric form.

8.3.4 Ligand binding and flexibility

MD trajectories of AChBP with ligands bound (chapter 3) show nicotine has one binding mode while carbamylcholine has more flexibility of its chain, which may allow it to bind in slightly different conformations at the choline end. Dihedral data and RSMD values of the ligands show very slight tilting of nicotine's 6' ring in the binding pocket as well, while carbamylcholine's chain seems to rotate more freely. This agrees with solid-state NMR data showing rotations of the methyl groups of the choline chain [209-211].

Docking studies of nicotine, acetylcholine, and carbamylcholine onto AChBP (chapter 4) also confirm lower energies and preferential docking in one mode for nicotine. It was also shown that nicotine bound most favourably onto the ligand

bound simulations, followed by acetylcholine and finally carbamylcholine. The docked conformation of ligands leads insight into how a ligand interacts with its environment, and it can be used to construct other ligands and also help with drug design.

8.4 Coupling ligand binding to channel gating

The method for combining separate domains of proteins discussed in chapter 5 is a tool for studying the structure-function relationships of proteins such as the gating of receptors. This is not as effectively studied on protein fragments as the interface between domains is important in understanding the structural changes being transmitted from one domain to another. The model of the α7 nAChR (and other LGICs) were generated and used extensively for structural studies including pore-profiles, conformational dynamics using GNM and CONCOORD, and electrostatic studies.

The conformational dynamics studies in chapter 5 using CONCOORD on the α7 nAChR suggested a plausible twisting motion between the two domains which could be a part of the gating mechanism. Asymmetrical intra-subunit motions were also observed with CONCOORD suggesting that each subunit may respond slightly differently in the event of gating. The GNM data provided insight on the more structurally flexible regions of the protein or segments which may have the most structural freedom during gating, suggesting a wave of flexibility in the TM domain.

The global conformational dynamics of LGICs were also studied with atomistic simulations of AChBP. PCA studies of simulation data from AChBP confirm asymetrical subunit motions also observed for the α7 nAChR in previous studies

[129], also the first eigenvector from PCA analysis of AChBP shows a twist to closing motion of the protein.

Structural rigidity in the presence of ligands was demonstrated with lower MSF and RMSD values in ligand-bound simulations compared with ligand-free simulations, thus suggesting that the ligand brings structural integrity to the protein. Thus a more flexible structure is likely without a ligand, however ligand-binding results in a rigidity of the structure which also involves intra-subunit movements as well as an over-all, slight, twisting motion.

Theoretical extension for global dynamics of ion channels can be carried out by taking simple cyclo-*N*mers and analysing their motion with normal mode analysis.

8.4.1 Cyclo-*N*meric motion

The modes of motion of different segments, domains, and subunits are important in understanding the mechanism of movement of multi-meric proteins. If we could represent the motion of each subunit using one or a few eigenvalues, we may understand how the different subunits inter-relate. The motions of these large proteins can be described in simple terms and thus it was worth exploring the motion of simple cyclic compounds to gain insight into the motions of larger cyclic molecules.

The cyclic nature of AChBP and the LGICs is shared amongst many families of proteins. Proteins exist as higher-order complexes such as cyclo-tetramers, pentamers, hexamers, heptamer, etc. and their mechanisms of movement are likely dictated by their shape and architecture. Therefore it is useful to step back and look at the normal modes of motion for simpler cyclic molecules such as pentane, hexane and heptane.

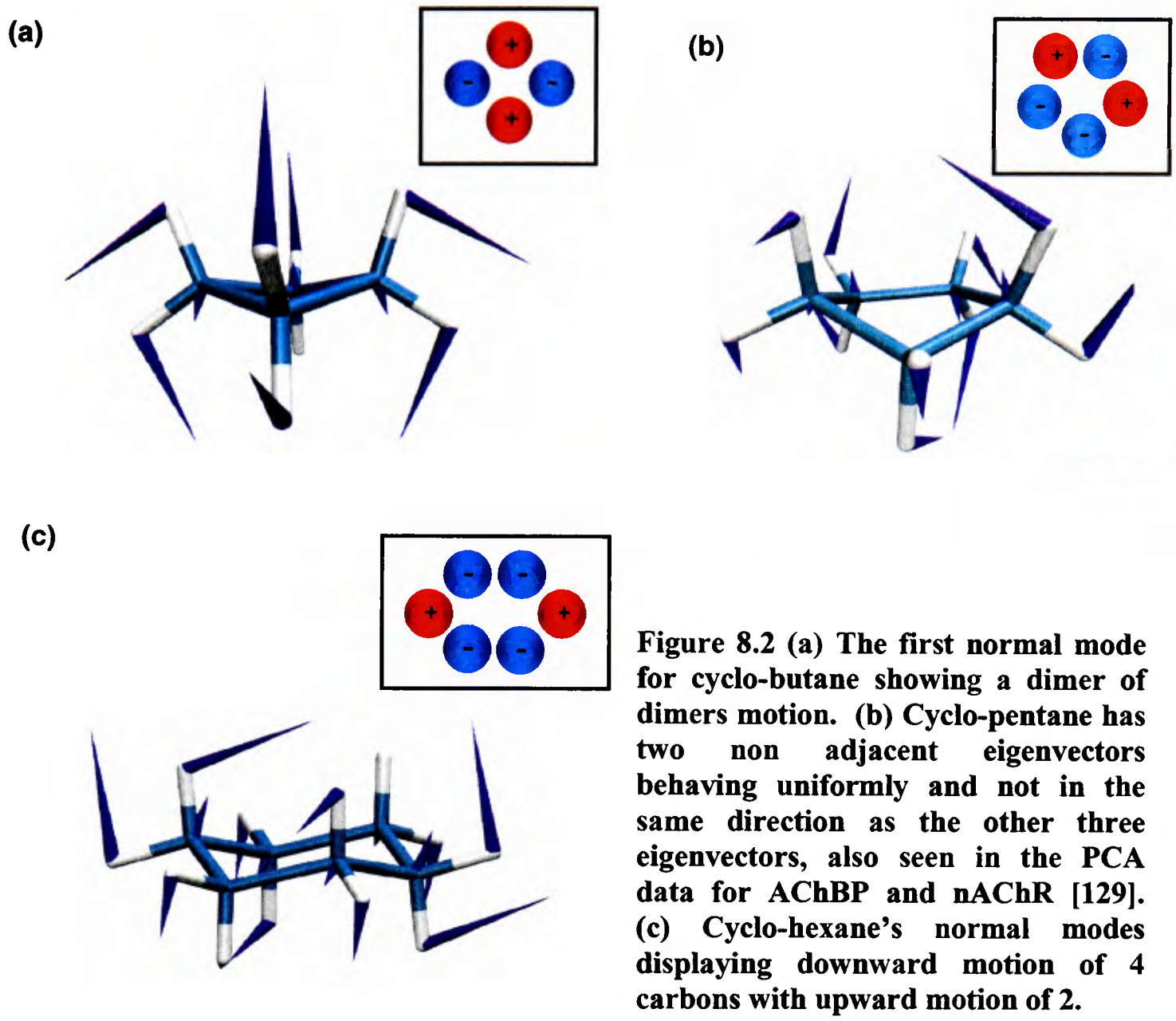


Figure 8.2 (a) The first normal mode for cyclo-butane showing a dimer of dimers motion. (b) Cyclo-pentane has two non adjacent eigenvectors behaving uniformly and not in the same direction as the other three eigenvectors, also seen in the PCA data for AChBP and nAChR [129]. (c) Cyclo-hexane's normal modes displaying downward motion of 4 carbons with upward motion of 2.

Normal mode analysis (NMA) for cyclic compounds was carried out using Spartan [180] (Figure 8.2). The first ten modes were analyzed; however only the first mode is discussed here as it represents the major motions for these compounds. The first normal mode for cyclo-butane corresponds to the dimer of dimers motion where neighbouring carbons have eigenvectors in opposing directions. Therefore in a protein of four domains, this corresponds to two domains moving in one direction and the other two in the opposite direction. The domains alternate in their up and down motions which is exhibited in many tetrameric proteins such as K^+-channels; this motion has been suggested in MD simulations of the ligand binding domain of inward rectifier K-channels [289]. For cyclo-pentane, two non-adjacent carbons have

eigenvectors in the same direction and magnitude (down, down, up, down, up) as was observed in the simulations of AChBP (chapter 3) and that of nAChR molecular dynamics studies [129] where two subunits showed asymmetrical patterns to their neighbouring subunits. The first mode of the cyclo-pentane is also the first eigenvector of the PCA data for the pentameric AChBP. Cyclo-hexane prefers the chair conformation of the ring structure, with eigenvectors showing a pattern of down, down, up, down, down, up. It is hypothesized that proteins of this nature hold this conformation, and also that of the 'boat', however the chair conformation is the first mode of motion for 6-C organic molecules and is the most energetically favourable state. For cyclo-heptane, we see an asymetrical pattern where the eigenvectors for each mode explore more directions than are observed in a simpler molecule such as the cyclo-propane. There is more asymmetry for molecules of 2n+1 C atoms versus those of an even number which show symmetry motion for their first modes. Although only the first modes are discussed here, the 2^{nd} and 3^{rd} normal modes of these simple cyclic compounds could also be useful in understanding the motion of cyclic proteins.

8.4.2 CG MD of LGICs

Although the conformational dynamics of Cα atoms from CG simulations match closely those of the atomistic simulations, making definitive predictions of global behaviour is not yet justified as the details of dynamic behaviour have not yet been assessed adequately. However this method holds great promise in analyzing the motion of these large proteins on much greater time-scales and thus being able to observe key functional events such as gating.

CG MD was used to study bilayer assembly around the LGICs. The bilayer formed rapidly in all cases and it allowed for the identification of residues interacting with the bilayer. In agreement with experimental and atomistic MD data, the M4 from the TM domain, has the most number of interactions with the lipid bilayer [279, 280]. Also, mutations in this helix result in the alteration of gating and kinetics of the ion channel [281, 283-286] which further illustrates the importance of lipid-protein interactions. Thus the lipid-assembly phase of the CG MD simulation is critical in obtaining a more realistic simulation environment. These studies can be extended by converting back to an atomistic model once bilayer assembly has completed which allows for atomistic MD simulations with a more realistic bilayer position.

8.5 Summary

Coarse-graining methods such as eigenvector averaging of AChBP MD data (chapter 3), conformational dynamics studies of α7 nAChR using CONCOORD and GNM (chapter 5), NMA of cyclic compounds, and PCA of AChBP have added to the over-all understand of gating motions. MD simulations of AChBP and ligand binding studies on MD frames have shown the increased structural mobility of a ligand-free protein, and that the binding site conforms to the bound ligand, favouring the re-docking of the same or similar ligands. This behaviour suggess an induced fit mechanism for ligand binding and is characteristic of the allosteric nature of these proteins.

Computational studies have been able to show how small changes in the binding pocket, such as a mutant residue, can affect not only the global behaviour but significantly alter ligand binding behaviour. The sensitivity of ligand-binding to even

the slightest differences in side-chain conformations is also demonstrated and confirms the subtlety and precision of the binding and gating process.

The electrostatic studies suggest that the EC domain plays a key role in ion selectivity while the TM domain provides an occlusion, namely the gate. From the electrostatics studies and pore-profile analysis, it can be concluded that the different segments of these receptors are involved in ion conduction and it is not solely due to the discrimination in one portion of the receptor. The involvement of the EC domain in ion selectivity was confirmed and the barrier to permeation in all channel types was further investigated using a Cl^- ion. In this case, Born energies were favourable in the EC domain of anion-selective receptors while the TM domain maintained a relatively high barrier to permeation.

Here, it is demonstrated that computational methods can be used to predict structural and behavioural changes in proteins, which can lead to or complement experimental study. The study of the different characteristics of LGICs such as the conformational sampling, electrostatics, and ligand binding behaviour has extended our understanding of the structural and functional properties of the ligand-gated super family of ion channels.

Appendix I Programs and scripts

I.1 Programs

I.1.1 ZAlign

```
##############################################################################
#  ZAlign.pl reads in two or more PDB files of separate domains of protein and
#  aligns the domains onto the z-axis.
#  It produces user-defined output statistics on rotations (theta) and corresponding
#  translations (z) to allow the user to choose the optimal model (opt(theta, z)) for
#  the protein.
#  The default criteria of output statistics are
#  1) termini distances between domains and
#  2) the number of bad contacts between domains
#  The user can also specify other distances to be calculated during the course of the
#  program.
#  precondition:  at least two PDB files must be provided by the user.
#  post-condtion:  the chosen (theta, z), is input for make_PDB.pl, which will make
#  the model with the optimal (theta, z) value.
#  Written by:  Shiva Amiri
##############################################################################


use strict;
use Math::Trig;
use POSIX qw(ceil floor);

use constant ANGLE => pi/90;
use constant PHI_Z_MAX => pi;
use constant RESTRAINT => 4.0;
use constant Z_RANGE => 10.0;
use constant DELTA_Z => 1;

my @static_temp;
my @static_temp2;
my @static_tokens;
my @static_coordinates;

my @mobile_temp;
my @mobile_temp2;
```

```
my @mobile_tokens;
my @mobile_coordinates;

my @static_file;
my @mobile_file;
my @new_stat_file;
my @new_mob_file;


my $pdb_outfile = "ACRB_TOL.pdb";
my $BC = "Bad_Contacts_ACRB_TOL.txt";
my $termini_score = "Termini_Score_ACRB_TOL.txt";


my @static_midpoint_1;
my @static_midpoint_2;

my @mobile_midpoint_1;
my @mobile_midpoint_2;

my @mobile_vector;
my @static_vector;

my $theta_s;
my $phi_s;

my $theta_m;
my $phi_m;




############################## termini distance variables##############################

my @static_rotate_array_A_t;
my @static_rotate_array_B_t;
my @static_rotate_array_C_t;

my @mobile_rotate_array_A_t;
my @mobile_rotate_array_B_t;
my @mobile_rotate_array_C_t;

my $static_residue_A_t = "1B";
my $static_residue_B_t = "509B";
my $static_residue_C_t = "382B";
my $static_residue_D_t = "255B";
my $static_residue_E_t = "128B";

my $mobile_residue_A_t = "206";
my $mobile_residue_B_t = "412";
my $mobile_residue_C_t = "618";
my $mobile_residue_D_t = "824";
my $mobile_residue_E_t = "1030";

my $distance_A_t;             # the distance between the c-alpha atoms of the target
residues of chain A
my $distance_B_t;             # the distance between the c-alpha atoms of the target
residues of chain B
my $distance_C_t;             # the distance between the c-alpha atoms of the target
residues of chain C
my $distance_D_t;             # the distance between the c-alpha atoms of the target
residues of chain D
my $distance_E_t;             # the distance between the c-alpha atoms of the target
residues of chain E

my $min_distance_A_t = 1000;  # the min distance between c-alpha atoms of target
residues in CHAIN A
my $min_distance_B_t = 1000;  # the min distance between c-alpha atoms of target
residues in CHAIN B
my $min_distance_C_t = 1000;  # the min distance between c-alpha atoms of target
residues in CHAIN C
my $min_distance_D_t = 1000;  # the min distance between c-alpha atoms of target
residues in CHAIN D
my $min_distance_E_t = 1000;  # the min distance between c-alpha atoms of target
residues in CHAIN E
```

```
my $x_min_A_t;          # the x-coord of the phi-optimized c-alpha of target mobile
residue - chain A
my $y_min_A_t;          # the y-coord of the phi-optimized c-alpha of target mobile
residue - chain A my $z_min_A_t;              # the z-coord of the phi-optimized c-
alpha of target mobile residue - chain A

my $x_min_B_t;          # the x-coord of the phi-optimized c-alpha of target mobile
residue - chain B my $y_min_B_t;              # the y-coord of the phi-optimized c-
alpha of target mobile residue - chain B my $z_min_B_t;              # the z-coord of
the phi-optimized c-alpha of target mobile residue - chain B

my $x_min_C_t;          # the x-coord of the phi-optimized c-alpha of target mobile
residue - chain C my $y_min_C_t;              # the y-coord of the phi-optimized c-
alpha of target mobile residue - chain C my $z_min_C_t;              # the z-coord of
the phi-optimized c-alpha of target mobile residue - chain C

my $x_min_D_t;          # the x-coord of the phi-optimized c-alpha of target mobile
residue - chain D my $y_min_D_t;              # the y-coord of the phi-optimized c-
alpha of target mobile residue - chain D my $z_min_D_t;              # the z-coord of
the phi-optimized c-alpha of target mobile residue - chain D

my $x_min_E_t;          # the x-coord of the phi-optimized c-alpha of target mobile
residue - chain E my $y_min_E_t;              # the y-coord of the phi-optimized c-
alpha of target mobile residue - chain E my $z_min_E_t;              # the z-coord of
the phi-optimized c-alpha of target mobile residue - chain E

my $min_average_t = 10000;

################################################################################

my @phi_optimized_coords;
my @phi_optimized_coords_t;
my @z_trans_coords;

my $z_r_t;
my $z_t_t;
my $z_r_u;
my $z_t_u;
my $z_r_b;
my $z_t_b;

my $bad_contacts = 0;
my $bad_ca_contacts = 0;


print "please enter static PDB filename:  \n";
my $static_infile = <STDIN>;
print "please enter mobile PDB filename:  \n";
my $mobile_infile = <STDIN>;

# the pdb file with the coordinates from the static protein (non-rotating)
open(STATIC_IN, "<$static_infile") || die "Cannot open $static_infile $!";

# the pdb file with the coordinates from the mobile protein (rotating)
open(MOBILE_IN, "<$mobile_infile") || die "Cannot open $mobile_infile $!";

# output pdb file of the aligned protein
open(PDB, ">$pdb_outfile") || die "Cannot open $pdb_outfile $!";

# output file with termini score values
open(TERMINI, ">$termini_score") || die "Cannot open $termini_score $!";

# output file with badcontacts
open(BADCONTACTS, ">$BC")  || die "Cannot open $BC $!";


print TERMINI "       Theta                    Z                TERMINI Distance
       \n\n";
print BADCONTACTS "       Theta                    Z                BAD CONTACTS
BAD c-alpha CONTACTS   \n\n";


format TERMINI=
  @#.############             @#.##          @####.############
  $z_r_t, $z_t_t, $distance_E_t
.
```

```
format BADCONTACTS=
  @#.############             @#.##          @####          @####
  $z_r_b, $z_t_b, $bad_contacts, $bad_ca_contacts
.

my $static_counter = 0;

        while (<STATIC_IN>)    {
                chomp ($_);
                @static_temp = "";
                @static_temp2 = "";
                @static_tokens = "";


                if (($_ =~ /^ATOM/) or ($_ =~ /^HETATOM/))    {
                        $static_counter++;
                        @static_tokens = split(' ', $_);      # split each line at
spaces
                        push(@static_temp, $static_tokens[6]); # store x-value into
@static_temp
                        push(@static_temp, $static_tokens[7]); # store y-value into
@static_temp
                        push(@static_temp, $static_tokens[8]); # store z-value into
@static_temp
                        push @static_coordinates, [@static_temp];

                        push(@static_temp2, $static_tokens[0]);
                        push(@static_temp2, $static_tokens[1]);
                        push(@static_temp2, $static_tokens[2]);
                        push(@static_temp2, $static_tokens[3]);
                        if (($static_counter > 7354) && ($static_counter <= 7639))  {
                                push(@static_temp2, "A");
                                push(@static_temp2, substr($static_tokens[4],1,100));
                                push(@static_temp2, $static_tokens[5]);
                                push(@static_temp2, $static_tokens[6]);
                                push(@static_temp2, $static_tokens[7]);
                                push(@static_temp2, $static_tokens[8]);
                                push(@static_temp2, $static_tokens[9]);
                                push(@static_temp2, $static_tokens[10]);
                        }
                        elsif (($static_counter > 14993) && ($static_counter <= 15278))
{
                                push(@static_temp2, "B");
                                push(@static_temp2, substr($static_tokens[4],1,100));
                                push(@static_temp2, $static_tokens[5]);
                                push(@static_temp2, $static_tokens[6]);
                                push(@static_temp2, $static_tokens[7]);
                                push(@static_temp2, $static_tokens[8]);
                                push(@static_temp2, $static_tokens[9]);
                                push(@static_temp2, $static_tokens[10]);
                        }
                        elsif (($static_counter > 22632) && ($static_counter <= 22917))
{
                                push(@static_temp2, "C");
                                push(@static_temp2, substr($static_tokens[4],1,100));
                                push(@static_temp2, $static_tokens[5]);
                                push(@static_temp2, $static_tokens[6]);
                                push(@static_temp2, $static_tokens[7]);
                                push(@static_temp2, $static_tokens[8]);
                                push(@static_temp2, $static_tokens[9]);
                                push(@static_temp2, $static_tokens[10]);
                        }
                        else {
                                push(@static_temp2, $static_tokens[4]);
                                push(@static_temp2, $static_tokens[5]);
                                push(@static_temp2, $static_tokens[6]);
                                push(@static_temp2, $static_tokens[7]);
                                push(@static_temp2, $static_tokens[8]);
                                push(@static_temp2, $static_tokens[9]);
                                push(@static_temp2, $static_tokens[10]);
                                push(@static_temp2, $static_tokens[11]);
                        }

                        push @static_file, [@static_temp2];                    }

}
```

```
        close(STATIC_IN);

        # Reading in the second pdb file and saving the coordinates into a
        # two-dimentional array
        while (<MOBILE_IN>)    {
                chomp ($_);
                @mobile_temp = "";
                @mobile_temp2 = "";
                @mobile_tokens = "";


                if (($_ =~ /^ATOM/) or ($_ =~ /^HETATOM/) or ($_ =~ /^TER/))        {
                        @mobile_tokens = split(' ', $_);      # split each line at
spaces
                        push(@mobile_temp, $mobile_tokens[6]); # store x-value into
@mobile_temp
                        push(@mobile_temp, $mobile_tokens[7]); # store y-value into
@mobile_temp
                        push(@mobile_temp, $mobile_tokens[8]); # store z-value into
@mobile_temp
                        push @mobile_coordinates, [@mobile_temp];

                        push(@mobile_temp2, $mobile_tokens[0]);
                        push(@mobile_temp2, $mobile_tokens[1]);
                        push(@mobile_temp2, $mobile_tokens[2]);
                        push(@mobile_temp2, $mobile_tokens[3]);
                        push(@mobile_temp2, $mobile_tokens[4]);
                        push(@mobile_temp2, $mobile_tokens[5]); # store x-value into
@mobile_temp
                        push(@mobile_temp2, $mobile_tokens[6]); # store y-value into
@mobile_temp
                        push(@mobile_temp2, $mobile_tokens[7]); # store z-value into
@mobile_temp
                        push(@mobile_temp2, $mobile_tokens[8]);
                        push(@mobile_temp2, $mobile_tokens[9]);
                        push(@mobile_temp2, $mobile_tokens[10]);
                        push(@mobile_temp2, $mobile_tokens[11]);
                        push @mobile_file, [@mobile_temp2];
                }

        }

        close(MOBILE_IN);


(@static_midpoint_1, @static_midpoint_2) = get_midpoints(\@static_file);
(@mobile_midpoint_1, @mobile_midpoint_2) = get_midpoints(\@mobile_file);



# gets two midpoints for each domain to be used to construct the midpoint vector.
# This subroutine finds the average of a slab from the top and bottom of each domain
sub get_midpoints()    {

        my ($dataset) = shift;
        my @coords = @{$dataset};
        my @temp;
        my @calc1;
        my @calc2;

        my @midpoint_1 = "";
        my @midpoint_2 = "";

        my $z_max = -1000;
        my $z_min = 1000;

        my $slab = 5;

        for my $i (0 .. $#coords)     {
                if ($coords[$i][8] > $z_max)  {
                        $z_max = $coords[$i][8];
                }
        }
        my $z_max_fl = floor($z_max);

        for my $i (0 .. $#coords)     {
```

```
                if  ($coords[$i][8] < $z_min) {
                        $z_min = $coords[$i][8];
                }
        }
        my $z_min_cl = ceil($z_min);

        for my $i (0 .. $#coords)     {

                if (($coords[$i][3] eq "CA") && ($coords[$i][8] > $z_min_cl) &&
($coords[$i][8] < ($z_min_cl + $slab)))      {
                        $temp[0] = $coords[$i][6];
                        $temp[1] = $coords[$i][7];
                        $temp[2] = $coords[$i][8];
                        push @calc1, [@temp];
                        @temp = "";
                }
        }
        my $calc1_counter = 0;
        for my $j (0..$#calc1) {
                $calc1_counter++;
                $midpoint_1[0] = $midpoint_1[0] + $calc1[$j][0];
                $midpoint_1[1] = $midpoint_1[1] + $calc1[$j][1];
                $midpoint_1[2] = $midpoint_1[2] + $calc1[$j][2];
        }

        $midpoint_1[0] = $midpoint_1[0]/$calc1_counter;
        $midpoint_1[1] = $midpoint_1[1]/$calc1_counter;
        $midpoint_1[2] = $midpoint_1[2]/$calc1_counter;

        #print "MIDPOINT 1 => X = $midpoint_1[0]             Y = $midpoint_1[1]
        Z = $midpoint_1[2]\n";

        ###################################################################

        for my $i (0 .. $#coords)     {
                if (($coords[$i][3] eq "CA") && ($coords[$i][8] > $z_max_f1 - $slab) &&
($coords[$i][8] < $z_max_f1)) {
                        $temp[0] = $coords[$i][6];
                        $temp[1] = $coords[$i][7];
                        $temp[2] = $coords[$i][8];
                        push @calc2, [@temp];
                        @temp = "";
                }
        }
        my $calc2_counter = 0;
        for my $j (0..$#calc2) {
                $calc2_counter++;
                $midpoint_2[0] = $midpoint_2[0] + $calc2[$j][0];
                $midpoint_2[1] = $midpoint_2[1] + $calc2[$j][1];
                $midpoint_2[2] = $midpoint_2[2] + $calc2[$j][2];
        }

        $midpoint_2[0] = $midpoint_2[0]/$calc2_counter;
        $midpoint_2[1] = $midpoint_2[1]/$calc2_counter;
        $midpoint_2[2] = $midpoint_2[2]/$calc2_counter;

        @calc1 = "";
        @calc2 = "";

        return(@midpoint_1, @midpoint_2);
}



build_vectors(0);
build_vectors(1);

# This subroutine uses the array containing midpoints of each domain to construct the
midpoint vector
sub build_vectors ()   {

        my $BOOL = shift;


        if ($BOOL == 1)        {
                print "building mobile vector\n";
```

```
                $mobile_vector[0] = $mobile_midpoint_1[0] - $mobile_midpoint_2[0];
                $mobile_vector[1] = $mobile_midpoint_1[1] - $mobile_midpoint_2[1];
                $mobile_vector[2] = $mobile_midpoint_1[2] - $mobile_midpoint_2[2];
                print "MOBILE vector => ($mobile_vector[0], $mobile_vector[1],
$mobile_vector[2])\n";
        }

        else {
                print "building static vector\n";
                $static_vector[0] = $static_midpoint_1[0] - $static_midpoint_2[0];
                $static_vector[1] = $static_midpoint_1[1] - $static_midpoint_2[1];
                $static_vector[2] = $static_midpoint_1[2] - $static_midpoint_2[2];
                print "STATIC vector => ($static_vector[0], $static_vector[1],
$static_vector[2])\n";
        }



}

($phi_s, $theta_s) = get_angles(\@static_vector);
($phi_m, $theta_m) = get_angles(\@mobile_vector);



# determines the angles of rotation for the midpoint vector of each subunit
sub get_angles ()      {

        my ($dataset) = shift;
        my @vector = @{$dataset};
        my @new_vec;
        my @final_vec;
        my $theta;
        my $phi;

        $new_vec[0] = $vector[0];
        $new_vec[1] = $vector[1];
        $new_vec[2] = 0;

        print "\n\n";
        $theta =
acos($new_vec[1]/(sqrt(($new_vec[0]*$new_vec[0])+($new_vec[1]*$new_vec[1])+($new_vec[2
]*$new_vec[2]))));
        print "THETA = $theta\n";

        $new_vec[0] = $vector[0]*cos($theta) - $vector[1]*sin($theta);
        $new_vec[1] = $vector[0]*sin($theta) + $vector[1]*cos($theta);
        $new_vec[2] = $vector[2];

        print "TEMP VECTOR => X = $new_vec[0]    Y = $new_vec[1]    Z = $new_vec[2]\n\n";

        $phi =
acos($new_vec[2]/(sqrt(($new_vec[0]*$new_vec[0])+($new_vec[1]*$new_vec[1])+($new_vec[2
]*$new_vec[2]))));
        print "PHI   = $phi\n";

        #zvector=T2*Vyz;
        $final_vec[0] = $new_vec[0];
        $final_vec[1] = $new_vec[1]*cos($phi) - $new_vec[2]*sin($phi);
        $final_vec[2] = $new_vec[1]*sin($phi) + $new_vec[2]*cos($phi);

        print "FINAL VECTOR => X = $final_vec[0]    Y = $final_vec[1]    Z =
$final_vec[2]\n\n";

        return ($phi, $theta);
}


@new_stat_file = rotate_domain(\@static_file, $theta_s, $phi_s, 0);
@new_mob_file = rotate_domain(\@mobile_file, $theta_m, $phi_m, 1);

#  This subroutine straightens and aligns each domain onto the z-axis given the
# rotation angles determined by get_angles().
```

```
sub rotate_domain {

        my ($dataset1) = shift;
        my @file = @{$dataset1};
        my $theta = shift;
        my $phi = shift;
        my $bool = shift;
        my @new_file;
        my @static_dist;
        my @midpoint;

        for my $j (0 .. $#file)       {
                for my $k (0 .. $#{$file[$j]})        {
                        $new_file[$j][$k] = $file[$j][$k];
                }
        }

        # ROTATION MATRIX
        for my $i (0 .. $#new_file)   {
                $new_file[$i][7] = $file[$i][7]*cos($theta) - $file[$i][8]*sin($theta);
                $new_file[$i][8] = $file[$i][7]*cos($phi)*sin($theta) +
$file[$i][8]*cos($phi)*cos($theta) - $file[$i][9]*sin($phi);
                $new_file[$i][9] = $file[$i][7]*sin($phi)*sin($theta) +
$file[$i][8]*sin($phi)*cos($theta) + $file[$i][9]*cos($phi);
        }

        if ($bool == 0){
                @midpoint =  &get_static_midpoints(\@new_file);
                # TRANSLATION ONTO Z_AXIS
                for my $i (0 .. $#new_file)   {
                        $new_file[$i][7] = $new_file[$i][7] - $midpoint[0];
                        $new_file[$i][8] = $new_file[$i][8] - $midpoint[1];
                }
        }
        else {
                @midpoint =  &get_mobile_midpoints(\@new_file);
                # TRANSLATION ONTO Z_AXIS
                for my $i (0 .. $#new_file)   {
                        $new_file[$i][7] = $new_file[$i][7] - $midpoint[0];
                        $new_file[$i][8] = $new_file[$i][8] - $midpoint[1];
                }
                for my $j (0 .. $#new_file)   {
                        $new_file[$j][7] = $new_file[$j][7];
                        $new_file[$j][8] = -1*$new_file[$j][8];
                        $new_file[$j][9] = -1*$new_file[$j][9] + 490;
                }

        }

        return (@new_file);
}



####### getting the target TERMINI static residues for distance calculations #######
for my $j (0 .. $#new_stat_file)      {
        if ($new_stat_file[$j][5] eq $static_residue_A_t) { # looking for the target
residue-chain A

                if ($new_stat_file[$j][3] eq "CA") {
                        $static_rotate_array_A_t[0] = $new_stat_file[$j][7];
                        $static_rotate_array_A_t[1] = $new_stat_file[$j][8];
                        $static_rotate_array_A_t[2] = $new_stat_file[$j][9];
                }

        }
        if ($new_stat_file[$j][5] eq $static_residue_B_t) { # looking for the target
residue-chain B

                if ($new_stat_file[$j][3] eq "CA") {
                        $static_rotate_array_B_t[0] = $new_stat_file[$j][7];
                        $static_rotate_array_B_t[1] = $new_stat_file[$j][8];
                        $static_rotate_array_B_t[2] = $new_stat_file[$j][9];
                }

        }
```

```
        if ($new_stat_file[$j][5] eq $static_residue_C_t) { # looking for the target
residue-chain C

                if ($new_stat_file[$j][3] eq "CA") {
                        $static_rotate_array_C_t[0] = $new_stat_file[$j][7];
                        $static_rotate_array_C_t[1] = $new_stat_file[$j][8];
                        $static_rotate_array_C_t[2] = $new_stat_file[$j][9];
                }

        }
}


driver();       # performs the rotations and corresponding translations and creates the
output files
                # with these values

sub driver ()  {

        for (my $phi_z=0; $phi_z<=PHI_Z_MAX; $phi_z=$phi_z+ANGLE)  {
                &rotate_about_z(-1*$phi_z);
        } # end of z-rotation FOR
}
sub rotate_about_z ()  {

        my $z_rot = $_[0];
        my @new_mob_coords;

        for my $j (0 .. $#new_mob_file)       {
                for my $k (0 .. $#{$new_mob_file[$j]})       {
                        $new_mob_coords[$j][$k] = $new_mob_file[$j][$k];
                }
        }

        # Z-ROTATION
        for my $i (0 .. $#new_mob_coords)     {
                my $temp_x = $new_mob_coords[$i][7];
                $new_mob_coords[$i][7] = $new_mob_coords[$i][7]*cos($z_rot) +
$new_mob_coords[$i][8]*sin($z_rot);
                $new_mob_coords[$i][8] = -$temp_x*sin($z_rot) +
$new_mob_coords[$i][8]*cos($z_rot);
        }
        &z_translation(\@new_mob_coords, $z_rot);
}

sub z_translation ()   {

        my $l;
        my ($dataset) = shift;
        my @z_trans_coords = @{$dataset};
        my $z_r = $_[0];
        my @new_trans_coords;

        for my $j (0 .. $#z_trans_coords)     {
                for my $k (0 .. $#{$z_trans_coords[$j]})     {
                        $new_trans_coords[$j][$k] = $z_trans_coords[$j][$k];
                }
        }

        # TRANSLATION loop => translation on z
        for ($l=-1*Z_RANGE; $l<=Z_RANGE; $l=$l+DELTA_Z)     {

                # performing the optimum translations and rotations to the entire PDB
file
                for my $i (0 .. $#z_trans_coords)     {
                        $new_trans_coords[$i][9] = $z_trans_coords[$i][9] - $l;
                }
        }

}



# Distance calculation between the c-alpha atoms of the termini residues
sub calculate_Termini_distances       {

        my ($dataset) = shift;
```

```
my @coords = @{$dataset};
$z_r_t = $_[0];
$z_t_t = $_[1];

for my $j (0..$#coords)       {
        if ($coords[$j][5] eq $mobile_residue_A_t) {
                if ($coords[$j][3] eq "CA") {
                        # storing the c-alpha coordinates in an array
                        $mobile_rotate_array_A_t[0] = $coords[$j][7];
                        $mobile_rotate_array_A_t[1] = $coords[$j][8];
                        $mobile_rotate_array_A_t[2] = $coords[$j][9];
                }
        }
        if ($coords[$j][5] eq $mobile_residue_B_t) {
                if ($coords[$j][3] eq "CA") {
                        # storing the c-alpha coordinates in an array
                        $mobile_rotate_array_B_t[0] = $coords[$j][7];
                        $mobile_rotate_array_B_t[1] = $coords[$j][8];
                        $mobile_rotate_array_B_t[2] = $coords[$j][9];
                }
        }
        if ($coords[$j][5] eq $mobile_residue_C_t) {
                if ($coords[$j][3] eq "CA") {
                        # storing the c-alpha coordinates in an array
                        $mobile_rotate_array_C_t[0] = $coords[$j][7];
                        $mobile_rotate_array_C_t[1] = $coords[$j][8];
                        $mobile_rotate_array_C_t[2] = $coords[$j][9];
                }
        }
}


my $x1_A = $static_rotate_array_A_t[0];
my $y1_A = $static_rotate_array_A_t[1];
my $z1_A = $static_rotate_array_A_t[2];

my $x2_A = $mobile_rotate_array_A_t[0];
my $y2_A = $mobile_rotate_array_A_t[1];
my $z2_A = $mobile_rotate_array_A_t[2];

my $x_diff_sq_A = ($x2_A-$x1_A) * ($x2_A-$x1_A);
my $y_diff_sq_A = ($y2_A-$y1_A) * ($y2_A-$y1_A);
my $z_diff_sq_A = ($z2_A-$z1_A) * ($z2_A-$z1_A);

my $sum_A = $x_diff_sq_A + $y_diff_sq_A + $z_diff_sq_A;

$distance_A_t = sqrt($sum_A);

my $x1_B = $static_rotate_array_B_t[0];
my $y1_B = $static_rotate_array_B_t[1];
my $z1_B = $static_rotate_array_B_t[2];

my $x2_B = $mobile_rotate_array_B_t[0];
my $y2_B = $mobile_rotate_array_B_t[1];
my $z2_B = $mobile_rotate_array_B_t[2];

my $x_diff_sq_B = ($x2_B-$x1_B) * ($x2_B-$x1_B);
my $y_diff_sq_B = ($y2_B-$y1_B) * ($y2_B-$y1_B);
my $z_diff_sq_B = ($z2_B-$z1_B) * ($z2_B-$z1_B);

my $sum_B = $x_diff_sq_B + $y_diff_sq_B + $z_diff_sq_B;

$distance_B_t = sqrt($sum_B);

my $x1_C = $static_rotate_array_C_t[0];
my $y1_C = $static_rotate_array_C_t[1];
my $z1_C = $static_rotate_array_C_t[2];

my $x2_C = $mobile_rotate_array_C_t[0];
my $y2_C = $mobile_rotate_array_C_t[1];
my $z2_C = $mobile_rotate_array_C_t[2];

my $x_diff_sq_C = ($x2_C-$x1_C) * ($x2_C-$x1_C);
my $y_diff_sq_C = ($y2_C-$y1_C) * ($y2_C-$y1_C);
my $z_diff_sq_C = ($z2_C-$z1_C) * ($z2_C-$z1_C);

my $sum_C = $x_diff_sq_C + $y_diff_sq_C + $z_diff_sq_C;
```

```
        $distance_C_t = sqrt($sum_C);


        write(TERMINI);

        my $average = ($distance_A_t + $distance_B_t + $distance_C_t)/3;

        # looking for the minimum values
        if (($average < $min_average_t) && ($distance_A_t >= RESTRAINT) &&
($distance_B_t >= RESTRAINT) && ($distance_C_t >= RESTRAINT) )      {
                $min_distance_A_t = $distance_A_t;
                $min_distance_B_t = $distance_B_t;
                $min_distance_C_t = $distance_C_t;
                $min_distance_D_t = $distance_D_t;
                $min_distance_E_t = $distance_E_t;

                $min_average_t = $average;

                # saving the coordinates at the minimum distance.
                for my $i (0 .. $#coords)      {
                        for my $j (0 .. $#{$coords[$i]})      {
                                $phi_optimized_coords_t[$i][$j] = $coords[$i][$j];
                        }

                }

                print " MIN Termini DIST A = $min_distance_A_t\n";
                print " MIN Termini DIST B = $min_distance_B_t\n";
                print " MIN Termini DIST C = $min_distance_C_t\n";
                print " MIN Termini DIST D = $min_distance_D_t\n";
                print " MIN Termini DIST E = $min_distance_E_t\n";


        }

}       # End of calculate_distances

# calculates the number of bad contacts between the static domain and the mobile
domain of
# the protein complex
sub bad_contacts       {

        my $x1;
        my $y1;
        my $z1;
        my $x2;
        my $y2;
        my $z2;

        my $z_min = 40;
        my $z_max = 68;

        my ($dataset) = shift;
        my @coords = @{$dataset};
        $z_r_b = $_[0];
        $z_t_b = $_[1];


        for (my $i=0; $i<=$#coords; $i++)      {


                if (($coords[$i][9] > $z_min) && ($coords[$i][9] < $z_max)) {

                        $x1 = $coords[$i][7];
                        $y1 = $coords[$i][8];
                        $z1 = $coords[$i][9];

                        for (my $j=0; $j<=$#static_file; $j++)       {


                                if (($static_file[$j][9] > $z_min) &&
($static_file[$j][9] < $z_max))       {

                                        $x2 = $static_file[$j][7];
                                        $y2 = $static_file[$j][8];
```

```
                                        $z2 = $static_file[$j][9];

                                        my $sq_x = ($x1-$x2) * ($x1-$x2);
                                        my $sq_y = ($y1-$y2) * ($y1-$y2);
                                        my $sq_z = ($z1-$z2) * ($z1-$z2);

                                        my $sum = $sq_x + $sq_y + $sq_z;

                                        my $distance = sqrt($sum);

                                        if ($distance < 1.5)   {
                                                $bad_contacts++;
                                        }

                                        if (($distance < 1.5) && ($coords[$i][3] eq "CA")
&& ($static_file[$j][3] eq "CA"))       {
                                                $bad_ca_contacts++;
                                        }
                                }
                        }
                }
        }

        print "bad contacts = $bad_contacts\n";
        print "bad c-alpha contacts = $bad_ca_contacts\n";
        write(BADCONTACTS);
        $bad_contacts = 0;
        $bad_ca_contacts = 0;

}



my $res_num;
for my $i (0 .. $#new_mob_file)         {

                        my $at_num_m = $new_mob_file[$i][2];
                        my $at_type_m = $new_mob_file[$i][3];
                        my $r_name_m = $new_mob_file[$i][4];
                        my $chain_m = $new_mob_file[$i][5];
                        my $r_num_m = $new_mob_file[$i][6];
                        my $x_m = $new_mob_file[$i][7];
                        my $y_m = $new_mob_file[$i][8];
                        my $z_m = $new_mob_file[$i][9];
                        my $a_m = $new_mob_file[$i][10];
                        my $b_m = $new_mob_file[$i][11];
                        my $c_m = $new_mob_file[$i][12];
                        printf PDB "%-6s%5d  %-3s %-3s %-1s%4d    %8.3f%8.3f%8.3f  %1.2f
%3.2f           %1s\n", "ATOM", $at_num_m, $at_type_m, $r_name_m, $chain_m, $r_num_m,
$x_m, $y_m, $z_m, $a_m, $b_m, $c_m;

                        $res_num = $r_num_m;

}

# printing the new lower coordinates onto the new PDB file
for my $i (0 .. $#new_stat_file)        {

                        my $at_num_s = $new_stat_file[$i][2];
                        my $at_type_s = $new_stat_file[$i][3];
                        my $r_name_s = $new_stat_file[$i][4];
                        my $chain_s = $new_stat_file[$i][5];
                        my $r_num_s = $new_stat_file[$i][6];
                        my $x_s = $new_stat_file[$i][7];
                        my $y_s = $new_stat_file[$i][8];
                        my $z_s = $new_stat_file[$i][9];
                        my $a_s = $new_stat_file[$i][10];
                        my $b_s = $new_stat_file[$i][11];
                        #my $c_s = $new_stat_file[$i][12];
                        printf PDB "%-6s%5d  %-3s %-3s %-1s%4d    %8.3f%8.3f%8.3f  %1.2f
%3.2f           %1s\n", "ATOM", $at_num_s, $at_type_s, $r_name_s, $chain_s, $r_num_s,
$x_s, $y_s, $z_s, $a_s, $b_s;

}
print PDB "END\n";
```

```
sub print_pdb_file () {

        my ($dataset) = shift;
        my @coords = @{$dataset};

        my $file_name = $_[0];
        my $r_num;

        open(PDB_TWO, ">$file_name") || die "Cannot open $file_name $!";

        # printing the new upper coordinates onto the new PDB file
        for my $i (0 .. $#coords)     {

                        my $at_num_m = $coords[$i][2];
                        my $at_type_m = $coords[$i][3];
                        my $r_name_m = $coords[$i][4];
                        my $r_num_m = $coords[$i][5];
                        my $x_m = $coords[$i][6];
                        my $y_m = $coords[$i][7];
                        my $z_m = $coords[$i][8];
                        my $a_m = $coords[$i][9];
                        my $b_m = $coords[$i][10];
                        my $c_m = $coords[$i][11];
                        printf PDB_TWO "%-6s%5d  %-3s %-3s %-1s%4d    %8.3f%8.3f%8.3f
%1.2f %3.2f           %1s\n", "ATOM", $at_num_m, $at_type_m, $r_name_m, $r_num_m,
$x_m, $y_m, $z_m, $a_m, $b_m, $c_m;
                        $r_num = $r_num_m;
        }

        # printing the new lower coordinates onto the new PDB file
        for my $i (0 .. $#new_stat_file)     {

                        my $at_num_s = $new_stat_file[$i][2];
                        my $at_type_s = $new_stat_file[$i][3];
                        my $r_name_s = $new_stat_file[$i][4];
                        my $r_num_s = $new_stat_file[$i][5];
                        my $x_s = $new_stat_file[$i][6];
                        my $y_s = $new_stat_file[$i][7];
                        my $z_s = $new_stat_file[$i][8];
                        my $a_s = $new_stat_file[$i][9];
                        my $b_s = $new_stat_file[$i][10];
                        my $c_s = $new_stat_file[$i][11];
                        printf PDB_TWO "%-6s%5d  %-3s %-3s %-1s%4d    %8.3f%8.3f%8.3f
%1.2f %3.2f           %1s\n", "ATOM", $at_num_s, $at_type_s, $r_name_s,
$r_num_s+$r_num, $x_s, $y_s, $z_s, $a_s, $b_s, $c_s;
        }

        print PDB_TWO "END\n";
        close(PDB_TWO);

}

close(PDB);
close(PDB_TWO);
close(TERMINI);
close(BADCONTACTS);
```

I.1.2 EigenGrouping

```
##############################################################################
# This program reads in an eigenvector text file and a coordinate file in PDB format
# and generates a PBD file of averaged eigenvectors associated with the alpha-carbons
# of the protein.
# preconditions:
# 1- The protein coordinates in PDB format
# 2- the number of alpha carbons
# 3- the eigenvector text file
# 4- the user must specify a coarse-graining factor
# output:  a series of files that can be viewed using molecular graphics software such
# as VMD.  Also an averaged eigenvector text file along with the averaged coordinate
# PBD # file and the corresponding eigenvectors.
#
# written by:  Shiva Amiri
##############################################################################

use strict;
use Math::Trig;
use Getopt::Std;

# options to run the program
use vars qw($opt_i $opt_c $opt_g $opt_n);

$opt_i="";
$opt_c="";
$opt_g="";
$opt_n="";

if ( ! getopts("i:c:g:n:")) {
  error_msg();
  exit;
}

sub error_msg{
        print STDERR "USAGE: $0 \n",
        "\t[-i  eigenvector input file name - eigenvec.txt (GROMACS or GNM)]\n",
         "\t[-c  coordinate file in PDB format]\n",
        "\t[-g  coarse graining factor - integer to group CAs]\n",
        "\t[-n  number of atoms in protein]\n";
}

# *******opt_i input file name *****
if ($opt_i=~ /^$/){ # user must specify a valid eigenvector file name
        error_msg();
        exit;
}

if ($opt_c=~ /^$/){ # user must specify a valid PDB file name
        error_msg();
        exit;
}

if ($opt_g=~ /^$/){ # user must specify a valid coarsegraining factor
        error_msg();
        exit;
}

if ($opt_n=~ /^$/){ # user must specify a valid coarsegraining factor
        error_msg();
        exit;
}

# check whether input file exists, readable and it is a non empty file
elsif(-e $opt_i && -r $opt_i && -s $opt_i){
        open(EIGEN, "<$opt_i") || die " Cannot open file $opt_i\n";
        open(PDB, "<$opt_c") || die " Cannot open file $opt_c\n";
}

my $inputFile = $opt_i;
print "file = $inputFile\n";

my $nRes = $opt_n;
print "number of atoms => $nRes\n";
```

```
my $coarseGrainFactor = $opt_g;
my $numCGAtom = 0;

my $mode = 0;
my @frame;

my $lineCount = 0;

my @Xeigenvals = "";
my @Yeigenvals = "";
my @Zeigenvals = "";

my @divisionOne = "";
my @divisionTwo = "";
my @divisionThree = "";
my @divisionFour = "";
my @divisionFive = "";
my @divisionSix = "";
my @divisionSeven = "";
my @divisionEight = "";
my @divisionNine = "";
my @divisionTen = "";
my @divisionEleven = "";
my @divisionTwelve = "";
my @divisionThirteen = "";
my @divisionFourteen = "";
my @divisionFifteen = "";
my @divisionSixteen = "";


my $zeroCount = 0;
my $negOneCount = 0;
my $negTwoCount = 0;
my $negThreeCount = 0;
my $negFourCount = 0;
my $negFiveCount = 0;
my $negSixCount = 0;
my $negSevenCount = 0;
my $negEightCount = 0;
my $negNineCount = 0;
my $negTenCount = 0;

my $frameCount = 0;
my @framesX;
my @framesY;
my @framesZ;

my @averageX;
my @averageY;
my @averageZ;

my $averagex = 0;
my $averagey = 0;
my $averagez = 0;

my $coarseX = 0;
my $coarseY = 0;
my $coarseZ = 0;

my @pdb_file = "";

$inputFile =~ s/\.txt//g;
$opt_c =~ s/\.pdb//g;

open(FRAMES, ">$opt_c.frames.txt") || die " Cannot open file frames.txt\n";
print FRAMES "This file displays the number of values in each 'exponent bucket' per
frame.\n";
print FRAMES "\n";
print FRAMES "   Frame       +00       -01       -02       -03       -04       -05
-06       -07       -08       -09       -10\n";
print FRAMES "\n\n";
format FRAMES =
 @#####  @######## @######## @######## @######## @######## @######## @########
@######## @######## @######## @########
```

```
$frameCount, $zeroCount, $negOneCount, $negTwoCount, $negThreeCount, $negFourCount,
$negFiveCount, $negSixCount, $negSevenCount, $negEightCount, $negNineCount,
$negTenCount
.


# file for looking at eigenbuckets for each frame to be viewed with xfarbe
open(FRAMESX, ">$opt_c.frames.dat") || die " Cannot open file frames.txt\n";
print FRAMESX "number of eigenvalues per 'exponent bucket' per frame\n";
print FRAMESX "\n";
format FRAMESX =
@### @### @### @### @### @### @### @### @### @### @###
$zeroCount, $negOneCount, $negTwoCount, $negThreeCount, $negFourCount, $negFiveCount,
$negSixCount, $negSevenCount, $negEightCount, $negNineCount, $negTenCount
.

# file for looking at eigenbuckets for each frame to be viewed with xfarbe
open(EIGENX, ">$opt_c.eigen.dat") || die " Cannot open file frames.txt\n";
print EIGENX "eigenvalues per alpha carbon per frame\n";
print EIGENX "$opt_n 10\n";


# this file has the average eigenvector for each c-alpha for the top ten eigenvectors
open (AVERAGESX, ">$opt_c.averages.dat") || die " Cannot open file averages.dat\n";
print AVERAGESX "Average for each C-alpha\n";
print AVERAGESX "3 $opt_n\n";

format AVERAGESX =
 @###.############ @###.############ @###.############
 $averagex, $averagey, $averagez
.


open (COARSEGRAIN, ">$opt_c.coarseGrained.txt") || die " Cannot open file
coarseGrained.txt\n";
format COARSEGRAIN =
 @# @### @##.############ @##.############ @##.############
 $mode, $numCGAtom, $coarseX, $coarseY, $coarseZ
.

# reads in original PDB file
while (<PDB>)  {
               chomp ($_);
               my @pdb_temp = "";
               my @pdb_tokens = "";

               if ($_ =~ /^ATOM/)     {
                       #@pdb_tokens = split(' ', $_);        # split each line at
spaces

                       $pdb_tokens[0] = substr($_,0,4);      # ATOM
                       $pdb_tokens[1] = substr($_,6,5);      # atom number
                       $pdb_tokens[2] = substr($_,12,4);     # atom type
                       $pdb_tokens[3] = substr($_,17,3);     # residue name
                       $pdb_tokens[4] = substr($_,21,1);     # chain
                       $pdb_tokens[5] = substr($_,23,4);     # residue number
                       $pdb_tokens[6] = substr($_,31,8);     # x
                       $pdb_tokens[7] = substr($_,39,8);     # y
                       $pdb_tokens[8] = substr($_,47,8);     # z
                       $pdb_tokens[9] = substr($_,56,4);     # a
                       $pdb_tokens[10] = substr($_,62,4);    # b


                       push(@pdb_temp, $pdb_tokens[0]);
                       push(@pdb_temp, $pdb_tokens[1]);
                       push(@pdb_temp, $pdb_tokens[2]);
                       push(@pdb_temp, $pdb_tokens[3]);
                       push(@pdb_temp, $pdb_tokens[4]);
                       push(@pdb_temp, $pdb_tokens[5]); # stores the number associated
with each residue
                       push(@pdb_temp, $pdb_tokens[6]); # store x-value into @pdb_temp
                       push(@pdb_temp, $pdb_tokens[7]); # store y-value into @pdb_temp
                       push(@pdb_temp, $pdb_tokens[8]); # store z-value into @pdb_temp
                       push(@pdb_temp, $pdb_tokens[9]);
                       push(@pdb_temp, $pdb_tokens[10]);
                       push @pdb_file, [@pdb_temp];  # stores @pdb_temp into
@pdb_coordinates array
```

```
                }

}

close(PDB);


#for my $j (1 .. $#pdb_file)  {
#       print "pdb_file at $j is:  @{$pdb_file[$j]}\n";
#}
my $breakFlag = 0;
my $fileType;
my $break = 0;

#reads the eigenvector file
while (<EIGEN>)         {
                chomp ($_);


                my @info = "";
                my @line = "";
                my @temp = "";
                my @tokens = "";
                my $y;
                my $z;


                if ($_ =~ /^ [0-9]/)   {
                        #print "GNM file\n";
                        #print "line = $_\n";
                        @temp = split(/\s+/,$_);

                        $frameCount = $temp[1];
                        push(@line, $temp[2]);
                        push(@line, $temp[3]);
                        push(@line, $temp[4]);
                        push(@line, $temp[5]);
                        $fileType = "GNM";
                        $break = $nRes;
                        $lineCount++;
                }

                elsif ($_ =~ /^      x\[/)    {

                        $_ =~ s/\s+//g;  # take out spaces
                        $_ =~ s/\[//g;   # take out the bracket
                        $_ =~ s/x//g;    # take out the x

                        @temp = split(']={',$_);
                        @tokens = split(',', $temp[1]);       # split each line at
spaces

                        push(@line, $temp[0]);
                        push(@line, $tokens[0]);
                        push(@line, $tokens[1]);

                        $tokens[2] =~ s/\}//g;
                        push(@line, $tokens[2]);
                        $fileType = "GROMACS";
                        if ($breakFlag == 0)   {
                                $break = $nRes;
                        }
                        $breakFlag = 1;
                        $lineCount++;
                        #print "linecount = $lineCount\n";
                }

                else {
                        next;

                }



                if ($line[2] =~ m/\+00/)      {
                        $zeroCount++;
                }
```

```
elsif ($line[2] =~ m/-01/)     {
        $negOneCount++;
}
elsif ($line[2] =~ m/-02/)     {
        $negTwoCount++;
}
elsif ($line[2] =~ m/-03/)     {
        $negThreeCount++;
}
elsif ($line[2] =~ m/-04/)     {
        $negFourCount++;
}
elsif ($line[2] =~ m/-05/)     {
        $negFiveCount++;
}
elsif ($line[2] =~ m/-06/)     {
        $negSixCount++;
}
elsif ($line[2] =~ m/-07/)     {
        $negSevenCount++;
}
elsif ($line[2] =~ m/-08/)     {
        $negEightCount++;
}
elsif ($line[2] =~ m/-09/)     {
        $negNineCount++;
}
elsif ($line[2] =~ m/-10/)     {
        $negTenCount++;
}
if ($line[2] >= 9)     {

        #push @divisionOne, [@line];
        push (@divisionOne, $line[2]);
}

elsif (($line[2] >= 8) && ($line[2] < 8.999999))     {

        #push @divisionTwo, [@line];
        push (@divisionTwo, $line[2]);
}

elsif (($line[2] >= 7) && ($line[2] < 7.999999))     {

        #push @divisionThree, [@line];
        push (@divisionThree, $line[2]);
}

elsif (($line[2] >= 6) && ($line[2] < 6.999999))     {

        #push @divisionFour, [@line];
        push (@divisionFour, $line[2]);
}

elsif (($line[2] >= 5) && ($line[2] < 5.999999))     {

        #push @divisionFive, [@line];
        push (@divisionFive, $line[2]);
}

elsif (($line[2] >= 4) && ($line[2] < 4.999999))     {

        #push @divisionSix, [@line];
        push (@divisionSix, $line[2]);
}

elsif (($line[2] >= 3) && ($line[2] < 3.999999))     {

        #push @divisionSeven, [@line];
        push (@divisionSeven, $line[2]);
}

elsif (($line[2] >= 2) && ($line[2] < 2.999999))     {

        #push @divisionEight, [@line];
        push (@divisionEight, $line[2]);
}
```

```
elsif (($line[2] >= 1) && ($line[2] < 1.999999))     {

        #push @divisionNine, [@line];
        push (@divisionNine, $line[2]);
}

elsif (($line[2] >= 0) && ($line[2] < 0.999999))     {

        #push @divisionTen, [@line];
        push (@divisionTen, $line[2]);
}

elsif (($line[2] >= -1) && ($line[2] < -0.000001))   {

        #push @divisionEleven, [@line];
        push (@divisionEleven, $line[2]);
}

elsif (($line[2] >= -2) && ($line[2] < -1.000001))   {

        push (@divisionTwelve, $line[2]);
}

elsif (($line[2] >= -3) && ($line[2] < -2.000001))   {

        push (@divisionThirteen, $line[2]);
}

elsif (($line[2] >= -4) && ($line[2] < -3.000001))   {

        push (@divisionFourteen, $line[2]);
}

elsif (($line[2] >= -5) && ($line[2] < -4.000001))   {

        push (@divisionFifteen, $line[2]);
}

elsif (($line[2] >= -6) && ($line[2] < -5.000001))   {

        push (@divisionSixteen, $line[2]);
}

push (@Xeigenvals, $line[2]);
push (@Yeigenvals, $line[3]);
push (@Zeigenvals, $line[4]);

if (($lineCount % $break == 0) && ($lineCount != 0)) {
        if ($fileType =~ /GROMACS/)    {
                $frameCount++;
        }

        splice(@Xeigenvals, 0, 1);
        splice(@Yeigenvals, 0, 1);
        splice(@Zeigenvals, 0, 1);

push @framesX, [@Xeigenvals];
        push @framesY, [@Yeigenvals];
        push @framesZ, [@Zeigenvals];
        push (@frame, $frameCount);

        @Xeigenvals = "";
        @Yeigenvals = "";
       @Zeigenvals = "";


        write(FRAMES);
        write(FRAMESX);

        $zeroCount = 0;
        $negOneCount = 0;
        $negTwoCount = 0;
        $negThreeCount = 0;
 $negFourCount = 0;
        $negFiveCount = 0;
        $negSixCount = 0;
```

```
                        $negSevenCount = 0;
                        $negEightCount = 0;
                        $negNineCount = 0;
                        $negTenCount = 0;
                        if ($frameCount == 10)  {
                                last;
                        }
                }

}

average();
coarseGrain();
coarseGrainPDB();

#getting the averages for x, y, z values of every c-alpha atom per frame
sub average {

  #print "FramesX at 0 0 ====> $framesX[0][0]\n";
  #print "FramesY at 0 0 ====> $framesY[0][0]\n";
  #print "FramesZ at 0 0 ====> $framesZ[0][0]\n";
  #print "frame count = $frameCount\n";

  for my $k (0 .. $nRes)      {
      for my $j (0 .. $#framesX)            {

                $averageX[$k] += $framesX[$j][$k];
                $averageY[$k] += $framesY[$j][$k];
                $averageZ[$k] += $framesZ[$j][$k];
      }

      $averageX[$k] = $averageX[$k]/$frameCount;
      $averageY[$k] = $averageY[$k]/$frameCount;
      $averageZ[$k] = $averageZ[$k]/$frameCount;
  }

  for my $m (0 .. $#averageX) {
       $averagex = $averageX[$m];
       $averagey = $averageY[$m];
       $averagez = $averageZ[$m];
  }


  for my $i (0 .. $#framesX)  {
       print EIGENX "@{$framesX[$i]}\n";
  }
}


# gets average of a group of eigenvectors
sub coarseGrain {


  my $numGroups = $nRes/$coarseGrainFactor;
  my @tempGroup = "";
  my @averages = "";
  my @coarseGrainedWindows;
  my $break;

  for (my $j=0; $j<=9; $j++)  {
         for (my $i=0; $i<$nRes; $i++)      {

               $tempGroup[0] += $framesX[$j][$i];
               $tempGroup[1] += $framesY[$j][$i];
               $tempGroup[2] += $framesZ[$j][$i];
               print "framesX at $j $i = $framesX[$j][$i]\n";
               if ($fileType =~ /GROMACS/)   {
                       $break = $i + 1;
               }

               else {

                       $break = $i;
               }
               if ((($break) % $coarseGrainFactor) == 0)   {

                       if ($fileType =~ /GNM/)       {
```

```
                        if ($i == 0)    {
                                next;
                        }
                }

                print "Coarse graining at $i \t";
                $numCGAtom++;
                print "=> @tempGroup\n";
                $coarseX = $tempGroup[0]/$coarseGrainFactor;
                $coarseY = $tempGroup[1]/$coarseGrainFactor;
                $coarseZ = $tempGroup[2]/$coarseGrainFactor;

                $averages[0] = $coarseX;
                $averages[1] = $coarseY;
                $averages[2] = $coarseZ;

                push @coarseGrainedWindows, [@averages];
                print "averages => @averages\n";

                $mode = $frame[$j];


                write(COARSEGRAIN);
                @tempGroup = "";
                @averages = "";
        if ($i+$coarseGrainFactor >= $nRes)   {
                last;
        }

        }

    }
    $numCGAtom = 0;
  }
}

# prints our the coarse-grained PDB file
sub coarseGrainPDB  {

  open(NEWPDB, ">$opt_c.coarseGrained.pdb") || die " Cannot open file
CoarseGrainedCoordinates.pdb\n";

  my @tempX;
  my @tempY;
  my @tempZ;
  my $CAcounter = 1;
  my $r_num = 0;

  print "coarse grain factor = $coarseGrainFactor\n";
  for (my $i=1; $i<=$#pdb_file; $i++)       {

      if ($pdb_file[$i][3] =~ /CA/){
            $tempX[0] += $pdb_file[$i][7];
            $tempX[1] += $pdb_file[$i][8];
            $tempX[2] += $pdb_file[$i][9];

            $CAcounter++;



            if (($CAcounter-1) % $coarseGrainFactor == 0)       {
                  $r_num++;
                  $tempX[0] = $tempX[0]/$coarseGrainFactor;
                  $tempX[1] = $tempX[1]/$coarseGrainFactor;
                  $tempX[2] = $tempX[2]/$coarseGrainFactor;
                  my $at_num = $pdb_file[$i][2];
                  my $r_name = $pdb_file[$i][4];
                  my $chain = $pdb_file[$i][5];

                  if ($pdb_file[$i][5] =~ /\s+/)       {
                        $chain = "X";
                  }

                  my $a = $pdb_file[$i][10];
                  my $b = $pdb_file[$i][11];
```

```
                        printf NEWPDB "%-6s%5d  %-3s %-3s %-1s%4d    %8.3f%8.3f%8.3f
%1.2f  %3.2f             %1s\n", "ATOM", $r_num, "CA", $r_name, $chain, $r_num,
$tempX[0], $tempX[1], $tempX[2], $a, $b;

                        @tempX = "";
                        @tempY = "";
                        @tempZ = "";

                }
        }

    }

    print NEWPDB "END\n";
    close(NEWPDB);

}



#system ("xfarbe eigen.dat");
#system ("xmgr averages.dat");
close(AVERAGESX);
close(FRAMES);
close(FRAMESX);
close(EIGENX);
close(COARSEGRAIN);
```

I.2 Scripts

I.2.1 automateDocks

```
##############################################################################
#  automateDocks.pl automates autodock runs for frames of a trajectory
#  It uses one PDBQS file for the charges of the protein and uses the same charges for
each frame.
#       27/05/2005
# INPUT:
#       1 - the first *.pdpqs file with correct charges for the protein (either via ADT
#       or other method
#       2 - the sperate frames in PDB formated called *_frame.pdb where * is the number
#       of the frame
#               ** this is the format generated by trjconv with -sep and -o frame.pdb
#               to get various frames from a trajectory
#       3 - the ligand in *.pdbq format
#       4 - centre coordinates for the grid, usually best if editconf is used with -
#       center 0 0 0  and the values are entered (in angstroms for the variables Cx,
#       Cy, and Cz).
#       5 - the number of runs can be modified at $numRuns
#       6 - dimensions for the grid can also be modified at Gx, Gy, and Gz
#       7 - the number of frames must also be specified
# OUTPUT:
#       1 - a separate directory is made for each frame, with the corresponding
#       autodock output
#       2 - *.map files generated by autodock are removed from each directory
#
# written by:  Shiva Amiri
##############################################################################


use strict;

my $frameRootName = "NCT_frame";  # the PDB file but do not include the extension .pdb
my $charges = "0_NCT_frame.pdbqs";
my $ligand = "acho_UA.pdbq";
```

```
# centre coordinates for the grid
my $Cx = 42.870;
my $Cy = 26.000;
my $Cz = 32.820;

my $numRuns = 50;

# dimensions for the grid
my $Gx = 100;
my $Gy = 100;
my $Gz = 100;

# grid spacing in Angstroms
my $gridSpacing = .200;


# number of frames for autodock3 runs
my $numFrames = 100;

# the original pdbqs file which will be used to stick the charges onto this frame
open(CHARGES, "<$charges") || die "Cannot open $charges $!";


my @frame_file = ();
my @charge_file = ();

# store the columns with the charges into an array
while (<CHARGES>)      {
        chomp ($_);
        my @charge_tokens = "";
        my @charge_temp = "";

    if ($_ =~ /^ATOM/) {
        $charge_tokens[0] = substr($_, 67, 9);
        $charge_tokens[1] = substr($_, 76, 9);
        $charge_tokens[2] = substr($_, 84, 9);
        push(@charge_temp, $charge_tokens[0]);
        push(@charge_temp, $charge_tokens[1]);
        push(@charge_temp, $charge_tokens[2]);
        push @charge_file, [@charge_temp];
    }

}
close(CHARGES);

# printing out each row
#for my $n (0..$#charge_file) {
#       print "$n is:  @{$charge_file[$n]}\n";
#}

##############################################################################
#
# loops through each frame, and carries out for each frame:
#       1 - making a PDBQS file from the original charges
#       2 - runs mkgpf3 to make *.gpf file
#       3 - runs mkdpf3 to make *.dpf file
#       4 - edits *.gpf file by changing the grid size and centre coordinates
#       5 - edits *.dpf file by changing the number of SA runs to 50
#               *** other edits can be added on if necessary
#       6 - runs autogrid3
#       7 - runs autodock3
#       8 - moves all files related to this frame to a directory called the frame name
##############################################################################

for my $i (0 .. $numFrames)    {

        my $frame =
"/sansom/s11/shiva/AutomatedDockingAChBP/NCT_1UW6/".$i."_".$frameRootName.".pdb";
        # the PDB file of current frame
        open(IN_ONE, "<$frame") || die "Cannot open $frame $!";


        my $pdbqsFile = $i."_".$frameRootName.".pdbqs";
        # the output file, contains the results of the rotations
        open(PDBQS, ">$pdbqsFile") || die "Cannot open $pdbqsFile $!";
```

```
        while (<IN_ONE>)        {
                chomp ($_);
                my @frame_tokens = "";
                my @frame_temp = "";

                if ($_ =~ /^ATOM/)      {
                        $frame_tokens[0] = substr($_, 0, 4);
                        $frame_tokens[1] = substr($_, 7, 5);
                        $frame_tokens[2] = substr($_, 12, 4);
                        $frame_tokens[3] = substr($_, 17, 3);
                        $frame_tokens[4] = substr($_, 21, 1);
                        $frame_tokens[5] = substr($_, 22, 5);
                        $frame_tokens[6] = substr($_, 31, 8);
                        $frame_tokens[7] = substr($_, 39, 8);
                        $frame_tokens[8] = substr($_, 47, 8);
                        $frame_tokens[9] = substr($_, 55, 5);
                        $frame_tokens[10] = substr($_, 60, 6);
                        $frame_tokens[11] = substr($_, 66, 9);
                        push(@frame_temp, $frame_tokens[0]);
                        push(@frame_temp, $frame_tokens[1]);
                        push(@frame_temp, $frame_tokens[2]);
                        push(@frame_temp, $frame_tokens[3]);
                        push(@frame_temp, $frame_tokens[4]);  # stored the number
associated with each residue
                        push(@frame_temp, $frame_tokens[5]);  # store x-value into
@frame_temp
                        push(@frame_temp, $frame_tokens[6]);  # store y-value into
@frame_temp
                        push(@frame_temp, $frame_tokens[7]);  # store z-value into
@frame_temp
                        push(@frame_temp, $frame_tokens[8]);
                        push(@frame_temp, $frame_tokens[9]);
                        push(@frame_temp, $frame_tokens[10]);
                        push(@frame_temp, $frame_tokens[11]);
                        push @frame_file, [@frame_temp];      # store @static_temp into
@static_coordinates array
                }
        }
        close(IN_ONE);


        for my $i (0 .. $#frame_file) {
                my $at_num = $frame_file[$i][2];      # atom number
                my $at_type = $frame_file[$i][3];     # atom type i.e CA
                my $r_name = $frame_file[$i][4];      # residue name
                my $chain = $frame_file[$i][5];       # chain identifier
                my $r_num = $frame_file[$i][6];       # residue number
                my $x = $frame_file[$i][7];           # x coord
                my $y = $frame_file[$i][8];           # y coord
                my $z = $frame_file[$i][9];           # z coord
                my $a = $frame_file[$i][10];          # 1.00
                my $b = $frame_file[$i][11];          # 1.00
                my $c = $charge_file[$i][1];
                my $d = $charge_file[$i][2];
                my $e = $charge_file[$i][3];
                printf PDBQS "%-6s%4d  %-3s %-3s %-1s%4d    %8.3f%8.3f%8.3f  %1.2f%6.2f
%8.3f%8.2f%8.2f\n", "ATOM", $at_num, $at_type, $r_name, $chain, $r_num, $x, $y, $z,
$a, $b, $c, $d, $e;
        }

        my $fileName = $i."_".$frameRootName;
        print $fileName;


        system "mkgpf3 $ligand $pdbqsFile";
        system "mkdpf3 $ligand $pdbqsFile";

        my $gpfFileIn = $fileName.".gpf";
        my $dpfFileIn = substr($ligand, 0, -4).$fileName.".dpf";

        print "GPF file = $gpfFileIn\n";
        print "DPF file = $dpfFileIn\n";
        print "\n\n";

        my $gpfFileOut = "temp.gpf";
        my $dpfFileOut = "temp.dpf";
```

```
# the gpf file generated by mkgpf3
open(GPF_IN, "<$gpfFileIn") || die "Cannot open $gpfFileIn $!";

# the new gpf file after edit
open(GPF_OUT, ">$gpfFileOut") || die "Cannot open $gpfFileOut $!";

# the dpf file generated with mkdpf3
open(DPF_IN, "<$dpfFileIn") || die "Cannot open $dpfFileIn $!";

# the dpf file after edit
open(DPF_OUT, ">$dpfFileOut") || die "Cannot open $dpfFileOut $!";

# read and edit GPF file
while (<GPF_IN>)       {
        chomp ($_);

        if ($_ =~ /^npts/)     {
                my @npts = split (/\s+/, $_);
                splice(@npts, -7);
                my $npts = $npts[0]."     $Gx $Gy $Gz #num.grid points in xyz";
                print GPF_OUT "$npts\n";
                next;
        }

        elsif ($_=~ /^spacing/)       {
                my @spacing = split (/\s+/,$_);
                splice(@spacing, -3);
                my $spacing = $spacing[0]."  $gridSpacing
#spacing (Angstroms)";
                print GPF_OUT "$spacing\n";
                next;
        }

        elsif ($_ =~ /^gridcenter/)   {
                my @gridCenter = split (/\s+/, $_);
                splice(@gridCenter, -6);
                my $auto = '"auto"';
                my $gridCenter = $gridCenter[0]." $Cx $Cy $Cz #xyz-coordinates
or $auto";
                print GPF_OUT "$gridCenter\n";
                next;
        }

        print GPF_OUT "$_\n";;
}
close(GPF_IN);
close(GPF_OUT);

# read and edit DPF file
while (<DPF_IN>)       {
        chomp ($_);

        if ($_ =~ /^ga_run/)   {
                my @ga_run = split (/\s+/, $_);
                splice(@ga_run, -8);
                my $ga_run = $ga_run[0]." $numRuns # do this many GA or LGA
runs";
                print DPF_OUT "$ga_run\n";
                next;
        }
        print DPF_OUT "$_\n";;
}
close(DPF_IN);
close(DPF_OUT);

system "mv temp.gpf $gpfFileIn";
system "mv temp.dpf $dpfFileIn";

my $glgFile = substr($gpfFileIn, 0, -3)."glg";
my $dlgFile = substr($dpfFileIn, 0, -3)."dlg";

print "running autogrid3 ..\n";
system "nice -n 19 autogrid3 -p $gpfFileIn -l $glgFile ";

print "running autodock3 ..\n";
```

```
        system "nice -n 19 autodock3 -p $dpfFileIn -l $dlgFile ";

        system "rm $dirName.pdb";
        system "mkdir $fileName";
        system "rm $fileName.*.map";
        system "rm $fileName.maps.*";
        system "mv $gpfFileIn $dpfFileIn $glgFile $dlgFile $fileName.* $fileName";


        print "All done...map files were removed.\n";
        splice (@frame_file, 0, $#frame_file+1);
}
```

I.2.2 analyzeAutomateDocks

```
##############################################################################
# This program reads in the docking output generated from Autodock for multiple frames
# of an MD simulation.  It records the top ranked docks energy and the number of docks
# in the top ranked cluster for each selected frame of the simulation.  It also
# calculates the distance of a dock with residues in the binding pocket which helps in
# high-throughput docking as it will allow the user to see if the dock is actually
# inside #the binding site or not.
#
# preconditions:  user must run this in the directory of the Autodock directories
# generated for each frame.
# the name of the ligand must be specified along with the root name of the generated
# directories.
#
# output:
#
# 1- a text file listing the docks ranked by frame.
# 2- a graphs (*.xvg) file with frame vs. the energy of the top ranked dock.
# 3- a graph (*.xvg) file with frames vs. the number of docks in the top ranked
# cluster
##############################################################################

use strict;

my $numFiles = 100;
my $ligand = "imidacloprid.pdbq";
my $fileRootName = "A7_D_frame";
my @ranks = ();
my $inBindingPocket;


for my $i (0 .. $numFiles)    {

        my $dirName = $i."_".$fileRootName;
        my $DLGfile = substr($ligand, 0, -4).$dirName.".dlg";
        open (DLG, "<$dirName/$DLGfile") || die "Cannot open $DLGfile $!";
        system ("get-docked $dirName/$DLGfile");
         print "DLG file => $DLGfile \n";

        my @line = "";
        my $counter = 0;
        while (<DLG>)  {
                chomp ($_);

                if ($_ =~ /^\s+\d\s+/) {

                        if ($_ =~ /\|/)        {
                                $counter++;
                                @line = split(/\|/, $_);
                                if ($counter < 2)      {
                                        $inBindingPocket = LigandInPocket($i, $line[2],
1);
                                        push (@line, $inBindingPocket);
```

```
                                        push @ranks, [@line];

                                    }
                                }
                            }

        }

        close (DLG);

}

my $frame;
my $lowestEnergy;
my $runNumber;
my $numInCluster;
my $inBP;

my $output = "DockingRanks_".$fileRootName."_".substr($ligand,0,-4)."out";
open (OUTPUT, ">$output") || die "Cannot open $output $!";

print OUTPUT "  FRAME          Lowest Energy Dock       Run Number      Number of Docks
in Cluster     Ligand Inside Binding Pocket \n\n";


my $output2 = "Clusters_".$fileRootName."_".substr($ligand,0,-4)."xvg";
open (XVG, ">$output2") || die "Cannot open $output2 $!";


my $output3 = "Energies_".$fileRootName."_".substr($ligand,0,-4)."xvg";
open (ENERGY, ">$output3") || die "Cannot open $output3 $!";


format OUTPUT =
  @<<<<<<<<          @####.###                    @#####                      @#####
@####
    $frame, $lowestEnergy, $runNumber, $numInCluster, $inBP
.


format XVG =
  @<<<<<<<<           @########
    $frame, $numInCluster
.


format ENERGY =
  @<<<<<<<<           @######.#####
    $frame, $lowestEnergy
.


for my $n (0..$#ranks)      {
        print "$n is:  @{$ranks[$n]}\n";
        $frame = $n + 1;
        $lowestEnergy = $ranks[$n][1];
        $runNumber = $ranks[$n][2];
        $numInCluster = $ranks[$n][4];
        print "ranks at 6 = $ranks[$n][6]\n";
        $inBP = $ranks[$n][6];

        write (OUTPUT);
        write (XVG);
        write (ENERGY);
}


close(OUTPUT);
close(XVG);
close(ENERGY);



# Determines if the ligand is actually in the binding pocket
sub LigandInPocket() {
```

```
my ($fileNum, $runNum, $CentreAtom) = @_;
print "$fileNum, $runNum, $CentreAtom\n";
my $Cx = 24.87;
my $Cy = 38.000;
my $Cz = 39.82;

my $dirName = "$fileNum"."_".$fileRootName;
my $DLG_PDB_file = substr($ligand, 0, -4).$dirName.".dlg.pdb";
open (DLGPDB, "<$dirName/$DLG_PDB_file") || die "Cannot open $DLG_PDB_file $!";
print "DLGPDB file => $DLG_PDB_file \n";

my $superImposedFile = $fileRootName."SuperImposed.pdb";
open (SP, ">>$superImposedFile") || die "Cannot open $superImposedFile $!";

my @line = "";
my $CORRECT_RUN = 0;
my @PDB_file = ();

while (<DLGPDB>)       {

        chomp ($_);
        my @PDB_tokens = "";
        my @PDB_temp = "";

        if ($_ =~ /^MODEL/)  {
                $CORRECT_RUN = 0;
                @line = split(/\s+/, $_);


                if ($line[1] == $runNum)  {
                     print "correct RUN\n";
                     $CORRECT_RUN = 1;

                }

        }

        if (($CORRECT_RUN == 1) && ($_ =~/^ATOM/))    {
                        #print "LIGAND COORDS\n";
                        $PDB_tokens[0] = substr($_, 0, 4);
                        $PDB_tokens[1] = substr($_, 7, 5);
                        $PDB_tokens[2] = substr($_, 12, 4);
                        $PDB_tokens[3] = substr($_, 17, 3);
                        $PDB_tokens[4] = substr($_, 21, 1);
                        $PDB_tokens[5] = substr($_, 22, 5);
                        $PDB_tokens[6] = substr($_, 31, 8);
                        $PDB_tokens[7] = substr($_, 39, 8);
                        $PDB_tokens[8] = substr($_, 47, 8);
                        $PDB_tokens[9] = substr($_, 55, 5);
                        $PDB_tokens[10] = substr($_, 60, 6);
                        $PDB_tokens[11] = substr($_, 66, 7);
                        push(@PDB_temp, $PDB_tokens[0]);
                        push(@PDB_temp, $PDB_tokens[1]);
                        push(@PDB_temp, $PDB_tokens[2]);
                        push(@PDB_temp, $PDB_tokens[3]);
                        push(@PDB_temp, $PDB_tokens[4]);  # stored the number associated
with each residue
                        push(@PDB_temp, $PDB_tokens[5]);  # store x-value into @PDB_temp
                        push(@PDB_temp, $PDB_tokens[6]);  # store y-value into @PDB_temp
                        push(@PDB_temp, $PDB_tokens[7]);  # store z-value into @PDB_temp
                        push(@PDB_temp, $PDB_tokens[8]);
                        push(@PDB_temp, $PDB_tokens[9]);
                        push(@PDB_temp, $PDB_tokens[10]);
                        push(@PDB_temp, $PDB_tokens[11]);
                        push @PDB_file, [@PDB_temp];
                }
        }

        close (DLGPDB);


        for my $m (0 .. $#PDB_file)   {

                my $at_num = $PDB_file[$m][2];          # atom number
                my $at_type = $PDB_file[$m][3];         # atom type m.e CA
```

```
                my $r_name = $PDB_file[$m][4];          # resmdue name
                my $chain = $PDB_file[$m][5]; # chamn mdentmfmer
                my $r_num = $PDB_file[$m][6]; # resmdue number
                my $x = $PDB_file[$m][7];               # x coord
                my $y = $PDB_file[$m][8];               # y coord
                my $z = $PDB_file[$m][9];               # z coord
                my $a = $PDB_file[$m][10];              # 1.00
                my $b = $PDB_file[$m][11];              # 1.00
                my $c = "";

                printf SP "%-6s%4d  %-3s %-3s %-1s%4d    %8.3f%8.3f%8.3f  %1.2f %3.2f
%1s\n", "ATOM", $at_num, $at_type, $r_name, $chain, $r_num, $x, $y, $z, $a, $b, $c;

        }
        print SP "ENDMDL\n";



        my $distance = 0;

        # calculating the distances
        for my $n (0 .. $#PDB_file)   {
                #print "$n is:  @{$PDB_file[$n]}\n";
                if ($CentreAtom == $PDB_file[$n][2])  {

                        my $distance = sqrt((($PDB_file[$n][7] - $Cx)*($PDB_file[$n][7]
- $Cx)) +
                                (($PDB_file[$n][8] - $Cy)*($PDB_file[$n][8] - $Cy)) +
                                (($PDB_file[$n][9] - $Cz)*($PDB_file[$n][9] - $Cz)));
                        print "DISTANCE = $distance\n";

                        if ($distance < 15)    {
                                print "1\n";
                                return 1;
                        }

                        else { return 0;}
                }
        }

        return 0;
} # END of subroutine


close(SP);
```

I.2.3 analyzeHbond

```
##############################################################################################
# This program reads in a *.dat file from g_hbond of GROMACS and records the number
# of time each identified water molecule interacts with a particular residue.  It
# records the residue and the number of times an H-bond is recorded.  It also ranks
# the water-residue interactions in decreasing order.
#
# input:  *.dat file from g_hbond.
# output:  a text file with the listed and ranked water-residue interactions.
#
# written by:  Shiva Amiri
##############################################################################################

use strict;

my $bridgingWaters = "bridgingWatersNCT.dat";


# the original pdbqs file which will be used to stick the charges onto this frame
open(HBOND, "<$bridgingWaters") || die "Cannot open $bridgingWaters $!";

my @line = "";
```

```
my @BridgingWaters = "";
my %BW = ();       #has that holds each line of
my $SOL = "";


while (<HBOND>)         {
        chomp ($_);
        my @charge_tokens = "";
        my @charge_temp = "";
        my @bridgingWater = "";

    if ($_ =~ /^\s*\d/)         {

        @line = split(/\s+/, $_);

        if ($line[0] eq "" )   {
                shift(@line);
        }

        $line[0] =~ s/://;
        $SOL = substr($line[1], 6);
        my @temp = split (/:/, $line[4]);
         my $residue1 = $temp[1];
         if ($residue1 !~ m/\d/)    {
               if ($line[5] =~ m/\w/)  {
                     my @num  = split(/[A-Z]/, $line[5]);
                    $line[5] = $num[0];
                }
               my $residueNum = $line[5];
                $residue1 = $residue1.$line[5];

         }

         my $Residue1 = substr($residue1, 0, 3);
         my $ResNum1 = substr($residue1, 3);

         #now dealing with the second residue or ligand involved in the bridging
         my $line = $line[8];
         my $line2 = $line[9];
         if ($line[8] !~ /\d/) {
            $line = $line[9];
            $line2 = $line[10];
         }
         if ($line[9] !~ /\d/)  {
            $line = $line[10];
            $line2 = $line[11];
         }

         my @temp2 = split (/:/, $line);
         my $residue2 = $temp2[1];

        if ($residue2 !~ m/\d/)         {
             if ($line2 =~ m/\w/)       {
                   my @num2 = split(/[A-Z]/, $line2);
                  $line2 = $num2[0];
             }
             my $residueNum2 = $line2;
             $residue2 = $residue2.$line2;

         }
        my $Residue2 = substr($residue2, 0, 3);
        my $ResNum2 = substr($residue2, 3);
        push(@bridgingWater, $line[0]);
        push(@bridgingWater, $SOL);
        push(@bridgingWater, $Residue1);
        push(@bridgingWater, $ResNum1);
        push(@bridgingWater, $Residue2);
        push(@bridgingWater, $ResNum2);
        push @BridgingWaters, [@bridgingWater];

          }

}
close(HBOND);

#making sure that the same water if it connects different residues, is accounted for
for my $k (1 .. $#BridgingWaters)     {
```

```
        my $ResOneBool = 0;
        my $ResTwoBool = 0;

        for my $l ($k+1 .. $#BridgingWaters) {

                if ($BridgingWaters[$l][2] eq $BridgingWaters[$k][2])       {

                        if (($BridgingWaters[$l][4] ne $BridgingWaters[$k][4])  &&
                             ($BridgingWaters[$l][4] ne $BridgingWaters[$k][6]))  {
                                  #print "at $BridgingWaters[$l][1]**
$BridgingWaters[$l][4] $BridgingWaters[$k][4] $BridgingWaters[$k][6]\n";
                                        $ResOneBool = 1;
                        }

                        if (($BridgingWaters[$l][6] ne $BridgingWaters[$k][6])  &&
                             ($BridgingWaters[$l][6] ne $BridgingWaters[$k][4]))  {
                            #print "at $BridgingWaters[$l][1]
*****$BridgingWaters[$l][6] $BridgingWaters[$k][4] $BridgingWaters[$k][6]\n";
                                        $ResTwoBool = 1;
                        }

                        if (($ResOneBool == 1)  or  ($ResTwoBool == 1))      {
                               #print "SAME waters : $BridgingWaters[$l][2]\n";
                               $BridgingWaters[$l][2] = $BridgingWaters[$l][2]."*";

                        }

                }

                $ResOneBool = 0;
                $ResTwoBool = 0;
        }

}


my @Waters = "";
my $water = "";
my $WaterCount = 1;
my $totalTime = 1;
my $origTime = 0;

# this loop counts the number of times a certain bridging water exists between two
specific contacts
for my $i (1 .. $#BridgingWaters)     {

        my @temp = "";
        $water = $BridgingWaters[$i][2];

        if (exists $BW{$water}) {
                $WaterCount = $BW{$water}[7];
                $WaterCount++;
                $BW{$water}[7] = $WaterCount;
                $totalTime = $BridgingWaters[$i][1] - $BW{$water}[1];
                #print "total time = $BW{$water}[2]\n";
                $BW{$water}[2] = $totalTime;
                $WaterCount = 1;
                $origTime = 0;
                $totalTime = 1;
                next;
        }

        else {
                $origTime = $BridgingWaters[$i][1];
                $totalTime = 1;

        }

        my $temp = "";
        push (@temp, $origTime);
        push (@temp, $totalTime);
        #push (@temp, $BridgingWaters[$i][1]);       # current time
        #push (@temp, $temp);
        push (@temp, $BridgingWaters[$i][3]);
        push (@temp, $BridgingWaters[$i][4]);
        push (@temp, $BridgingWaters[$i][5]);
```

```
        push (@temp, $BridgingWaters[$i][6]);
        push (@temp, $WaterCount);
        $BW{$water} = [ @temp ];
        $WaterCount = 1;
        $origTime = 0;
        $totalTime = 0;

}


my $TotalTime = "";
my $Solvent = "";
my $RES1 = "";
my $RESNUM1 = "";
my $RES2 = "";
my $RESNUM2 = "";
my $TotalWaters = "";


my $output = "analyze_hbondNCT.out";
open (OUTPUT, ">$output") || die "Cannot open $output $!";

print OUTPUT "  SOLVENT           Residue One     Residue Two          Total Num
Total Time (ps)    \n\n";

format OUTPUT =
  @<<<<<<<<<<<<     @<<<@#####      @<<<@#####     @##########    @#########
  $Solvent, $RES1, $RESNUM1, $RES2, $RESNUM2, $TotalWaters, $TotalTime
.

foreach $water (keys %BW)     {
        print "$water:  @{ $BW{$water} }\n";
        $Solvent = $water;
        $RES1 = $BW{$water}[3];
        $RESNUM1 = $BW{$water}[4];
        $RES2 = $BW{$water}[5];
        $RESNUM2 = $BW{$water}[6];
        $TotalWaters = $BW{$water}[7];
        $TotalTime = $BW{$water}[2];
        write (OUTPUT);
}

close(OUTPUT);
```

Appendix II Publications

- Amiri, S., Shimomura M., Akamatsu M., Matsuda, K., Jones, A.K., Sansom, M.S.P., Biggin, P.C., and Sattelle, D.B., *A Role for Leu118 of Loop E in Agonist Binding to the α7 Nicotinic Acetylcholine Receptor.* Mol Pharm, submitted.

Computational methods such as MD, *in silico* mutagenesis and molecular docking are used to investigate the role of Leu118 in agonist binding, this is verified with experimental work presented in the same paper (chapter 6).

- Amiri, S., Sansom, M.S.P., and Biggin, P.C., *Molecular Dynamics Studies of AChBP with Nicotine and Carbamylcholine: The Role of Water in the Binding Pocket.* PEDS - submitted

MD of AChBP carried out in ligand-bound and ligand-free states with a focus on binding pocket dynamics and the involvement of water molecules in the binding of ligands (chapter 3).

- Amiri, S., Tai, K., Beckstein, O., Biggin, P.C., and Sansom, M.S.P *The alpha7 nicotinic acetylcholine receptor: molecular modelling, electrostatics, and energetics.* Mol Membr Biol, 2005. **22**(3): p. 151-62.

A model of the a7 nAChR is constructed by combining homology models for its EC and TM domains. The model is used for structural studies such as conformational sampling methods to investigate plausible global motions, pore profiles and electrostatics (chapter 5).

- Amiri, S., Biggin, P.C., Sansom, M.S.P The role of the extracellular domain of ligand-gated ion channels in ion selectivity – in preparation

Electrostatics studies of the LGICs reveals the important role of the EC domain in selectivity (chapter 7).

References

1. Nelson, M., *Comparative Neurophysiology*. 2000.
2. Wallin, E. and von Heijne, G., *Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms.* Protein Sci, 1998. **7**: p. 1029-38.
3. Jones, D.T., *Do transmembrane protein superfolds exist?* FEBS Lett, 1998. **423**: p. 281-5.
4. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* J Mol Biol, 2001. **305**: p. 567-80.
5. Terstappen, G.C. and Reggiani, A., *In silico research in drug discovery.* Trends Pharmacol Sci, 2001. **22**: p. 23-6.
6. Nichols, C.G., *K_{ATP} channels as molecular sensors of cellular metabolism.* Nature, 2006. **440**: p. 470-77.
7. Werten, P.J., Remigy, H.W., de Groot, B.L., Fotiadis, D., Philippsen, A., Stahlberg, H., Grubmuller, H., and Engel, A., *Progress in the analysis of membrane protein structure and function.* FEBS Lett, 2002. **529**: p. 65-72.
8. White, S.H., *The progress of membrane protein structure determination.* Protein Sci, 2004. **13**: p. 1948-9.
9. Henderson, R. and Unwin, P.N.T., *3-Dimensional Model of Purple Membrane Obtained by Electron-Microscopy.* Nature, 1975. **257**: p. 28-32.
10. Luecke, H., Schobert, B., Richter, H.T., Cartailler, J.P., and Lanyi, J.K., *Structure of bacteriorhodopsin at 1.55 A resolution.* J Mol Biol, 1999. **291**: p. 899-911.
11. Schobert, B., Cupp-Vickery, J., Hornak, V., Smith, S., and Lanyi, J., *Crystallographic structure of the K intermediate of bacteriorhodopsin: conservation of free energy after photoisomerization of the retinal.* J Mol Biol, 2002. **321**: p. 715-26.
12. Marsh, D., *Peptide models for membrane channels.* Biochem J, 1996. **315**: p. 345-61.
13. Miyazawa, A., Fujiyoshi, Y., and Unwin, N., *Structure and gating mechanism of the acetylcholine receptor pore.* Nature, 2003. **423**: p. 949-55.
14. Ashcroft, F.M., *From molecule to malady.* Nature, 2006. **440**: p. 440-7.
15. Lester, H.A., Dibas, M.I., Dahan, D.S., Leite, J.F., and Dougherty, D.A., *Cys-loop receptors: new twists and turns.* Trends Neurosci, 2004. **27**: p. 329-36.
16. Gonen, T., Sliz, P., Kistler, J., Cheng, Y., and Walz, T., *Aquaporin-0 membrane junctions reveal the structure of a closed water pore.* Nature, 2004. **429**: p. 193-7.
17. Kukulski, W., Schenk, A.D., Johanson, U., Braun, T., de Groot, B.L., Fotiadis, D., Kjellbom, P., and Engel, A., *The 5A structure of heterologously expressed plant aquaporin SoPIP2;1.* J Mol Biol, 2005. **350**: p. 611-6.
18. Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S.C., and Walz, T., *Lipid-protein interactions in double-layered two-dimensional AQP0 crystals.* Nature, 2005. **438**: p. 633-8.
19. Unwin, N., *Refined structure of the nicotinic acetylcholine receptor at 4A resolution.* J Mol Biol, 2005. **346**: p. 967-89.
20. Brejc, K., van Dijk, W.J., Klaassen, R.V., Schuurmans, M., van Der Oost, J., Smit, A.B., and Sixma, T.K., *Crystal structure of an ACh-binding protein*

reveals the ligand-binding domain of nicotinic receptors. Nature, 2001. **411**: p. 269-76.
21. Suzuki, M., Morita, T., and Iwamoto, T., *Diversity of Cl(-) channels.* Cell Mol Life Sci, 2006. **63**: p. 12-24.
22. Bass, R.B., Strop, P., Barclay, M., and Rees, D.C., *Crystal structure of Escherichia coli MscS, a voltage-modulated and mechanosensitive channel.* Science, 2002. **298**: p. 1582-7.
23. Kuo, A., Gulbis, J.M., Antcliff, J.F., Rahman, T., Lowe, E.D., Zimmer, J., Cuthbertson, J., Ashcroft, F.M., Ezaki, T., and Doyle, D.A., *Crystal structure of the potassium channel KirBac1.1 in the closed state.* Science, 2003. **300**: p. 1922-6.
24. Long, S.B., Campbell, E.B., and Mackinnon, R., *Crystal structure of a mammalian voltage-dependent Shaker family K+ channel.* Science, 2005. **309**: p. 897-903.
25. Dutzler, R., Campbell, E.B., and MacKinnon, R., *Gating the selectivity filter in ClC chloride channels.* Science, 2003. **300**: p. 108-12.
26. Dutzler, R., *The CIC family of chloride channels and transporters.* Curr Opin Struct Biol, 2006. **16**: p. 439-46.
27. Ashcroft, F.M., *Ion Channels and Disease.* 2000, San Diego: Academic Press.
28. Marinou, M. and Tzartos, S.J., *Identification of regions involved in the binding of alpha-bungarotoxin to the human alpha7 neuronal nicotinic acetylcholine receptor using synthetic peptides.* Biochem J, 2003. **372**: p. 543-54.
29. Unwin, N., *The Croonian Lecture 2000. Nicotinic acetylcholine receptor and the structural basis of fast synaptic transmission.* Philos Trans R Soc Lond B Biol Sci, 2000. **355**: p. 1813-29.
30. Kash, T.L., Jenkins, A., Kelley, J.C., Trudell, J.R., and Harrison, N.L., *Coupling of agonist binding to channel gating in the GABA(A) receptor.* Nature, 2003. **421**: p. 272-5.
31. Nicke, A., *Learning about structure and function of neuronal nicotinic acetylcholine receptors. Lessons from snails.* Eur J Biochem, 2004. **271**: p. 2293.
32. Schofield, C.M., Jenkins, A., and Harrison, N.L., *A highly conserved aspartic acid residue in the signature disulfide loop of the alpha 1 subunit is a determinant of gating in the glycine receptor.* J Biol Chem, 2003. **278**: p. 34079-83.
33. Karlin, A., *Emerging structure of the nicotinic acetylcholine receptors.* Nat Rev Neurosci, 2002. **3**: p. 102-14.
34. Moise, L., Piserchio, A., Basus, V.J., and Hawrot, E., *NMR structural analysis of alpha-bungarotoxin and its complex with the principal alpha-neurotoxin-binding sequence on the alpha 7 subunit of a neuronal nicotinic acetylcholine receptor.* J Biol Chem, 2002. **277**: p. 12406-17.
35. Law, R.J., Tieleman, D.P., and Sansom, M.S.P., *Pores formed by the nicotinic receptor M2 delta peptide: A molecular dynamics simulation study.* Biophys J, 2003. **84**: p. 14-27.
36. Montal, M. and Opella, S.J., *The structure of the M2 channel-lining segment from the nicotinic acetylcholine receptor.* Biochim Biophys Acta, 2002. **1565**: p. 287-93.
37. Tang, P., Mandal, P.K., and Xu, Y., *NMR structures of the second transmembrane domain of the human glycine receptor alpha(1) subunit:*

model of pore architecture and channel gating. Biophys J, 2002. **83**: p. 252-62.

38. Brejc, K., van Dijk, W.J., Smit, A.B., and Sixma, T.K., *The 2.7 A structure of AChBP, homologue of the ligand-binding domain of the nicotinic acetylcholine receptor.* Novartis Found Symp, 2002. **245**: p. 22-9; discussion 29-32, 165-8.
39. Dutertre, S., Nicke, A., Tyndall, J.D., and Lewis, R.J., *Determination of alpha-conotoxin binding modes on neuronal nicotinic acetylcholine receptors.* J Mol Recognit, 2004. **17**: p. 339-47.
40. Sixma, T.K. and Smit, A.B., *Acetylcholine binding protein (AChBP): a secreted glial protein that provides a high-resolution model for the extracellular domain of pentameric ligand-gated ion channels.* Annu Rev Biophys Biomol Struct, 2003. **32**: p. 311-34.
41. Smit, A.B., Brejc, K., Syed, N., and Sixma, T.K., *Structure and function of AChBP, homologue of the ligand-binding domain of the nicotinic acetylcholine receptor.* Ann N Y Acad Sci, 2003. **998**: p. 81-92.
42. Hille, B., *Ion channels of excitable membranes.* 2001, Sunderland Mass.: Sinaur Associates Inc.
43. Imoto, K., Busch, C., Sakmann, B., Mishina, M., Konno, T., Nakai, J., Bujo, H., Mori, Y., Fukuda, K., and Numa, S., *Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance.* Nature, 1988. **335**: p. 645-48.
44. Kelley, S.P., Dunlop, J.I., Kirkness, E.F., Lambert, J.J., and Peters, J.A., *A cytoplasmic region determines single-channel conductance in 5-HT_3 receptors.* Nature, 2003. **424**: p. 321-24.
45. Wilson, G.G. and Karlin, A., *The location of the gate in the acetylcholine receptor channel.* Neuron, 1998. **20**: p. 1269-81.
46. Gunthorpe, M.J. and Lummis, S.C., *Conversion of the ion selectivity of the 5-HT_{3A} receptor from cationic to anionic reveals a conserved feature of the ligand-gated ion channel superfamily.* . J Biol Chem, 2001. **276**: p. 10977-83.
47. Reeves, D.C., Goren, E.N., Akabas, M.H., and Lummis, S.C.R., *Structural and electrostatic properties of the 5-HT_3 receptor pore revealed by substituted cysteine accessibility mutagenesis.* . J Biol Chem, 2001. **276**: p. 42035-42.
48. Zhang, H. and Karlin, A., *Contribution of the beta subunit M2 segment to the ion-conducting pathway of the acetylcholine receptor.* Biochemistry, 1998. **37**: p. 7952-64.
49. Jensen, M.L., Schousboe, A., and Ahring, P.K., *Charge selectivity of the Cys-loop family of ligand-gated ion channels.* J Neurochem, 2005. **92**: p. 217-25.
50. Jensen, M.L., Pederson, L.N., Timmermann, D.B., Schousboe, A., and P.K., A., *Mutational studies using a cation-conducting $GABA_A$ receptor reveal the selectivity determinants of the Cys-loop family of ligand-gated ion channels.* J Neurochem, 2005. **92**: p. 962-72.
51. Lester, H.A., *The permeation pathway of neurotransmitter-gated ion channels.* Annu Rev Biophys Biomol Struct., 1992. **21**: p. 267-92.
52. Edmonds, B., Gibb, A.J., and Colquhoun, D., *Mechanisms of activation of muscle nicotinic acetylcholine receptors.* Annu Rev Physiol, 1995. **57**: p. 469-93.
53. Lingle, C.J., Maconochie, D., and Steinbach, J.H., *Activation of skeletal muscle nicotinic acetylcholine receptors.* J Membr Biol, 1992. **126**: p. 195-217.

54. Jackson, M.B., *Ligand-gated channel:postsynaptic receptors and drug targets.* Adv Neurobiol, 1999. **79**: p.
55. Jackson, M.B., *Kinetics of unliganded acetylcholine receptor channel gating.* Biophys J, 1986. **49**: p. 663-72.
56. Jackson, M.B., *Spontaneous openings of the acetylcholine receptor channel.* Proc Natl Acad Sci U S A, 1984. **81**: p. 3901-04.
57. Sine, S.M. and Engel, A.G., *Recent advances in Cys-loop receptor structure and function.* Nature, 2006. **440**: p. 448-55.
58. Unwin, N., *Structure and action of the nicotinic acetylcholine receptor explored by electron microscopy.* FEBS Lett, 2003. **555**: p. 91-5.
59. Lummis, S.C., Beene, D.L., Lee, L.W., Lester, H.A., Broadhurst, R.W., and Dougherty, D.A., *Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel.* Nature, 2005. **438**: p. 248-52.
60. Unwin, N., *Acetylcholine receptor channel imaged in the open state.* Nature, 1995. **373**: p. 37-43.
61. Beckstein, O., Biggin, P.C., Bond, P., Bright, J.N., Domene, C., Grottesi, A., Holyoake, J., and Sansom, M.S., *Ion channel gating: insights via molecular simulations.* FEBS Lett, 2003. **555**: p. 85-90.
62. Beckstein, O. and Sansom, M.S., *The influence of geometry, surface character, and flexibility on the permeation of ions and water through biological pores.* Phys Biol, 2004. **1**: p. 42-52.
63. Panicker, S., Cruz, H., Arrabit, C., and Slesinger, P.A., *Evidence for a centrally located gate in the pore of a serotonin-gated ion channel.* J Neurosci, 2002. **22**: p. 1629-39.
64. Akabas, M.H., Kaufmann, C., Archdeacon, P., and Karlin, A., *Identification of acetylcholine receptor channel-lining residues in the entire M2 segment of the alpha subunit.* Neuron, 1994. **13**: p. 919-27.
65. Karlin, A. and Akabas, M.H., *Toward a structural basis for the function of nicotinic acetylcholine receptors and their cousins.* Neuron, 1995. **15**: p. 1231-44.
66. Wilson, G.G., Pascual, J.M., Brooijmans, N., Murray, D., and Karlin, A., *The intrinsic electrostatic potential and the intermediate ring of charge in the acetylcholine receptor channel.* J Gen Physiol, 2000. **115**: p. 93-106.
67. Paas, Y., Gibor, G., Grailhe, R., Savatier-Duclert, N., Dufresne, V., Sunesen, M., de Carvalho, L.P., Changeux, J.P., and Attali, B., *Pore conformations and gating mechanism of a Cys-loop receptor.* Proc Natl Acad Sci U S A, 2005. **102**: p. 15877-82.
68. Cymes, G.D., Ni, Y., and Grosman, C., *Probing ion-channel pores one proton at a time.* Nature, 2005. **438**: p. 975-80.
69. Zhou, Y., Pearson, J.E., and Auerbach, A., *Phi-value analysis of a linear, sequential reaction mechanism: theory and application to ion channel gating.* Biophys J, 2005. **89**: p. 3680-85.
70. Cymes, G.D., Grosman, C., and Auerbach, A., *Structure of the transition state of gating in the acetylcholine receptor channel pore: a phi-value analysis.* Biochemistry, 2002. **41**: p. 5548-55.
71. Chakrapani, S., Bailey, T.D., and Auerbach, A., *Gating dynamics of the acetylcholine receptor extracellular domain.* J Gen Physiol, 2004. **123**: p. 341-56.
72. Grosman, C., Zhou, M., and Auerbach, A., *Mapping the conformational wave of acetylcholine receptor channel gating.* Nature, 2000. **403**: p. 773-6.

73. Mitra, A., Bailey, T.D., and Auerbach, A.L., *Structural dynamics of the M4 transmembrane segment during acetylcholine receptor gating.* Structure, 2004. **12**: p. 1909-18.
74. Bourne, Y., Talley, T.T., Hansen, S.B., Taylor, P., and Marchot, P., *Crystal structure of a Cbtx-AChBP complex reveals essential interactions between snake alpha-neurotoxins and nicotinic receptors.* Embo J, 2005. **24**: p. 1512-22.
75. Celie, P.H., Kasheverov, I.E., Mordvintsev, D.Y., Hogg, R.C., van Nierop, P., van Elk, R., van Rossum-Fikkert, S.E., Zhmak, M.N., Bertrand, D., Tsetlin, V., Sixma, T.K., and Smit, A.B., *Crystal structure of nicotinic acetylcholine receptor homolog AChBP in complex with an alpha-conotoxin PnIA variant.* Nat Struct Mol Biol, 2005. **12**: p. 582-8.
76. Celie, P.H., Klaassen, R.V., van Rossum-Fikkert, S.E., van Elk, R., van Nierop, P., Smit, A.B., and Sixma, T.K., *Crystal structure of acetylcholine-binding protein from Bulinus truncatus reveals the conserved structural scaffold and sites of variation in nicotinic acetylcholine receptors.* J Biol Chem, 2005. **280**: p. 26457-66.
77. Celie, P.H., van Rossum-Fikkert, S.E., van Dijk, W.J., Brejc, K., Smit, A.B., and Sixma, T.K., *Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures.* Neuron, 2004. **41**: p. 907-14.
78. Hansen, S.B., Sulzenbacher, G., Huxford, T., Marchot, P., Taylor, P., and Bourne, Y., *Structures of Aplysia AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations.* EMBO J, 2005. **24**: p. 3635-46.
79. Lummis, S.C., D, L.B., Harrison, N.J., Lester, H.A., and Dougherty, D.A., *A cation-pi binding interaction with a tyrosine in the binding site of the GABAC receptor.* Chem Biol, 2005. **12**: p. 993-7.
80. Price, K.L. and Lummis, S.C., *The role of tyrosine residues in the extracellular domain of the 5-hydroxytryptamine3 receptor.* J Biol Chem, 2004. **279**: p. 23294-301.
81. Spier, A.D. and Lummis, S.C., *The role of tryptophan residues in the 5-Hydroxytryptamine(3) receptor ligand binding domain.* J Biol Chem, 2000. **275**: p. 5620-5.
82. Zhong, W., Gallivan, J.P., Zhang, Y., Li, L., Lester, H.A., and Dougherty, D.A., *From ab initio quantum mechanics to molecular neurobiology: a cation-pi binding site in the nicotinic receptor.* Proc Natl Acad Sci U S A, 1998. **95**: p. 12088-93.
83. Arinaminpathy, Y., Sansom, M.S., and Biggin, P.C., *Binding site flexibility: molecular simulation of partial and full agonists within a glutamate receptor.* Mol Pharmacol, 2006. **69**: p. 11-8.
84. Kaye, S.L., Sansom, M.S., and Biggin, P.C., *Molecular dynamics simulations of the ligand-binding domain of an NMDA-receptor.* J Biol Chem, 2006.
85. Gao, F., Bren, N., Burghardt, T.P., Hansen, S., Henchman, R.H., Taylor, P., McCammon, J.A., and Sine, S.M., *Agonist-mediated conformational changes in acetylcholine-binding protein revealed by simulation and intrinsic tryptophan fluorescence.* J Biol Chem, 2005. **280**: p. 8443-51.
86. Capener, C.E., Kim, H.J., Arinaminpathy, Y., and Sansom, M.S., *Ion channels: structural bioinformatics and modelling.* Hum Mol Genet, 2002. **11**: p. 2425-33.

87. Colquhoun, D. and Sakmann, B., *Fluctuations in the microsecond time range of the current through single acetylcholine receptor ion channels.* Nature, 1981. **294**: p. 464-6.
88. Sine, S.M., Claudio, T., and Sigworth, F.J., *Activation of Torpedo acetylcholine receptors expressed in mouse fibroblasts. Single channel current kinetics reveal distinct agonist binding affinities.* J Gen Physiol, 1990. **96**: p. 395-437.
89. Zhang, Y., Chen, J., and Auerbach, A., *Activation of recombinant mouse acetylcholine receptors by acetylcholine, carbamylcholine and tetramethylammonium.* J Physiol, 1995. **486 (Pt 1)**: p. 189-206.
90. Absalom, N.L., Lewis, T.M., Kaplan, W., Pierce, K.D., and Schofield, P.R., *Role of charged residues in coupling ligand binding and channel activation in the extracellular domain of the glycine receptor.* J Biol Chem, 2003. **278**: p. 50151-7.
91. Akk, G. and Steinbach, J.H., *Structural elements near the C-terminus are responsible for changes in nicotinic receptor gating kinetics following patch excision.* J Physiol, 2000. **527 Pt 3**: p. 405-17.
92. Beene, D.L., Price, K.L., Lester, H.A., Dougherty, D.A., and Lummis, S.C., *Tyrosine residues that control binding and gating in the 5-hydroxytryptamine3 receptor revealed by unnatural amino acid mutagenesis.* J Neurosci, 2004. **24**: p. 9097-104.
93. Boileau, A.J., Newell, J.G., and Czajkowski, C., *GABA(A) receptor beta 2 Tyr97 and Leu99 line the GABA-binding site. Insights into mechanisms of agonist and antagonist actions.* J Biol Chem, 2002. **277**: p. 2931-7.
94. Bouzat, C., Bren, N., and Sine, S.M., *Structural basis of the different gating kinetics of fetal and adult acetylcholine receptors.* Neuron, 1994. **13**: p. 1395-402.
95. Dang, H., England, P.M., Farivar, S.S., Dougherty, D.A., and Lester, H.A., *Probing the role of a conserved M1 proline residue in 5-hydroxytryptamine(3) receptor gating.* Mol Pharmacol, 2000. **57**: p. 1114-22.
96. Lee, W.Y. and Sine, S.M., *Principal pathway coupling agonist binding to channel gating in nicotinic receptors.* Nature, 2005. **438**: p. 243-7.
97. Mukhtasimova, N., Free, C., and Sine, S.M., *Initial coupling of binding to gating mediated by conserved residues in the muscle nicotinic receptor.* J Gen Physiol, 2005. **126**: p. 23-39.
98. Peters, J.A., Hales, T.G., and Lambert, J.J., *Molecular determinants of single-channel conductance and ion selectivity in the Cys-loop family: insights from the 5-HT3 receptor.* Trends Pharmacol Sci, 2005. **26**: p. 587-94.
99. Stewart, D.S., Chiara, D.C., and Cohen, J.B., *Mapping the structural requirements for nicotinic acetylcholine receptor activation by using tethered alkyltrimethylammonium agonists and antagonists.* Biochemistry, 2006. **45**: p. 10641-53.
100. Xiu, X., Hanek, A.P., Wang, J., Lester, H.A., and Dougherty, D.A., *A unified view of the role of electrostatic interactions in modulating the gating of Cys loop receptors.* J Biol Chem, 2005. **280**: p. 41655-66.
101. Bouzat, C., Gumilar, F., Spitzmaul, G., Wang, H.L., Rayes, D., Hansen, S.B., Taylor, P., and Sine, S.M., *Coupling of agonist binding to channel gating in an ACh-binding protein linked to an ion channel.* Nature, 2004. **430**: p. 896-900.

102. Grutter, T., Prado de Carvalho, L., Virginie, D., Taly, A., Fischer, M., and Changeux, J.P., *A chimera encoding the fusion of an acetylcholine-binding protein to an ion channel is stabilized in a state close to the desensitized form of ligand-gated ion channels.* C R Biol, 2005. **328**: p. 223-34.
103. Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A.R., *Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles.* Biochemistry, 1990. **29**: p. 9343-52.
104. Changeux, J.P., Bertrand, D., Corringer, P.J., Dehaene, S., Edelstein, S., Lena, C., Le Novere, N., Marubio, L., Picciotto, M., and Zoli, M., *Brain nicotinic receptors: structure and regulation, role in learning and reinforcement.* Brain Res Brain Res Rev, 1998. **26**: p. 198-216.
105. Papke, R.L., Webster, J.C., Lippiello, P.M., Bencherif, M., and Francis, M.M., *The activation and inhibition of human nicotinic acetylcholine receptor by RJR-2403 indicate a selectivity for the alpha4beta2 receptor subtype.* J Neurochem, 2000. **75**: p. 204-16.
106. Engel, A.G. and Ohno, K., *Congenital myasthenic syndromes.* Adv Neurol, 2002. **88**: p. 203-15.
107. Steinlein, O.K., *Nicotinic acetylcholine receptors and epilepsy.* Curr Drug Targets CNS Neurol Disord, 2002. **1**: p. 443-8.
108. Matsuda, K., Buckingham, S.D., Kleier, D., Rauh, J.J., Grauso, M., and Sattelle, D.B., *Neonicotinoids: insecticides acting on insect nicotinic acetylcholine receptors.* Trends Pharmacol Sci, 2001. **22**: p. 573-80.
109. Corringer, P., Le Novere, N., and Changeux, J., *Nicotinic receptors at the amino acid level.* Annu Rev Pharmacol Toxicol, 2000. **40**: p. 431-58.
110. Millar, N.S., *Assembly and subunit diversity of nicotinic acetylcholine receptors.* Biochem Soc Trans, 2003. **31**: p. 869-74.
111. Prendergast, M.A., Jackson, W.J., Terry, A.V., Jr., Decker, M.W., Arneric, S.P., and Buccafusco, J.J., *Central nicotinic receptor agonists ABT-418, ABT-089, and (-)-nicotine reduce distractibility in adult monkeys.* Psychopharmacology (Berl), 1998. **136**: p. 50-8.
112. Papke, R.L., Meyer, E., Nutter, T., and Uteshev, V.V., *alpha7 receptor-selective agonists and modes of alpha7 receptor activation.* Eur J Pharmacol, 2000. **393**: p. 179-95.
113. Lloyd, A.W., *Therapeutic agents for the treatment of Alzheimer's disease.* Drug Discovery Today, 2000. **5**: p. 479-79.
114. Changeux, J. and Edelstein, S.J., *Allosteric mechanisms in normal and pathological nicotinic acetylcholine receptors.* Curr Opin Neurobiol, 2001. **11**: p. 369-77.
115. Steinlein, O.K., *Neuronal nicotinic receptors in human epilepsy.* Eur J Pharmacol, 2000. **393**: p. 243-47.
116. Lang, B. and Vincent, A., *Autoantibodies to ion channels at the neuromuscular junction.* Autoimmun Rev, 2003. **2**: p. 94-100.
117. Watson, R., Jepson, J.E., Bermudez, I., Alexander, S., Hart, Y., McKnight, K., Roubertie, A., Fecto, F., Valmier, J., Sattelle, D.B., Beeson, D., Vincent, A., and Lang, B., *Alpha7-acetylcholine receptor antibodies in two patients with Rasmussen encephalitis.* Neurology, 2005. **65**: p. 1802-4.
118. Vernino, S. and Lennon, V.A., *Neuronal ganglionic acetylcholine receptor autoimmunity.* Ann N Y Acad Sci, 2003. **998**: p. 211-4.

119. Jones, A.K., Grauso, M., and Sattelle, D.B., *The nicotinic acetylcholine receptor gene family of the malaria mosquito, Anopheles gambiae.* Genomics, 2005. **85**: p. 176-87.
120. Couturier, S., Bertrand, D., Matter, J.M., Hernandez, M.C., Bertrand, S., Millar, N., Valera, S., Barkas, T., and Ballivet, M., *A neuronal nicotinic acetylcholine receptor subunit (alpha 7) is developmentally regulated and forms a homo-oligomeric channel blocked by alpha-BTX.* Neuron, 1990. **5**: p. 847-56.
121. Itier, V. and Bertrand, D., *Neuronal nicotinic receptors: from protein structure to function.* FEBS Lett, 2001. **504**: p. 118-25.
122. Fucile, S., Renzi, M., Lauro, C., Limatola, C., Ciotti, T., and Eusebi, F., *Nicotinic cholinergic stimulation promotes survival and reduces motility of cultured rat cerebellar granule cells.* Neuroscience, 2004. **127**: p. 53-61.
123. Buckingham, S.D., Pym, L., Jones, A.K., Brown, L., Sansom, M.S., Sattelle, D.B., and Biggin, P.C., *A7DB: a relational database for mutational, physiological and pharmacological data related to the alpha7 nicotinic acetylcholine receptor.* BMC Neurosci, 2005. **6**: p. 2.
124. Ballivet, M., Alliod, C., Bertrand, S., and Bertrand, D., *Nicotinic acetylcholine receptors in the nematode Caenorhabditis elegans.* J Mol Biol, 1996. **258**: p. 261-60.
125. Francis, M.M., Vazquez, R.W., Papke, R.L., and Oswald, R.E., *Subtype-selective inhibition of neuronal nicotinic acetylcholine receptors by cocaine is determined by the alpha4 and beta4 subunits.* Mol Pharmacol, 2000. **58**: p. 109-19.
126. Galzi, J.L., Devillers-Thiery, A., Hussy, N., Bertrand, S., Changeux, J.P., and Bertrand, D., *Mutations in the channel domain of a neuronal nicotinic receptor convert ion selectivity from cationic to anionic.* Nature, 1992. **359**: p. 500-5.
127. Bertrand, D., Devillers-Thiery, A., Revah, F., Galzi, J.L., Hussy, N., Mulle, C., Bertrand, S., Ballivet, M., and Changeux, J.P., *Unconventional pharmacology of a neuronal nicotinic receptor mutated in the channel domain.* Proc Natl Acad Sci U S A, 1992. **89**: p. 1261-5.
128. Le Novere, N., Grutter, T., and Changeux, J.P., *Models of the extracellular domain of the nicotinic receptors and of agonist- and Ca2+-binding sites.* Proc Natl Acad Sci U S A, 2002. **99**: p. 3210-5.
129. Henchman, R.H., Wang, H.L., Sine, S.M., Taylor, P., and McCammon, J.A., *Asymmetric structural motions of the homomeric alpha7 nicotinic receptor ligand binding domain revealed by molecular dynamics simulation.* Biophys J, 2003. **85**: p. 3007-18.
130. Henchman, R.H., Wang, H.L., Sine, S.M., Taylor, P., and McCammon, J.A., *Ligand-induced conformational change in the alpha7 nicotinic receptor ligand binding domain.* Biophys J, 2005. **88**: p. 2564-76.
131. Law, R.J., Henchman, R.H., and McCammon, J.A., *A gating mechanism proposed from a simulation of a human alpha7 nicotinic acetylcholine receptor.* Proc Natl Acad Sci U S A, 2005. **102**: p. 6813-8.
132. 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., and Geraerts, W.P., *A glia-derived acetylcholine-binding protein that modulates synaptic transmission.* Nature, 2001. **411**: p. 261-8.

133. Schapira, M., Abagyan, R., and Totrov, M., *Structural model of nicotinic acetylcholine receptor isotypes bound to acetylcholine and nicotine.* BMC Struct Biol, 2002. **2**: p. 1.
134. Law, R.J., Forrest, L.R., Ranatunga, K.M., La Rocca, P., Tieleman, D.P., and Sansom, M.S., *Structure and dynamics of the pore-lining helix of the nicotinic receptor: MD simulations in water, lipid bilayers, and transbilayer bundles.* Proteins, 2000. **39**: p. 47-55.
135. Saiz, L. and Klein, M.L., *Computer simulation studies of model biological membranes.* Acc Chem Res, 2002. **35**: p. 482-9.
136. Saiz, L. and Klein, M.L., *The transmembrane domain of the acetylcholine receptor: insights from simulations on synthetic peptide models.* Biophys J, 2005. **88**: p. 959-70.
137. Hung, A., Tai, K., and Sansom, M.S., *Molecular dynamics simulation of the M2 helices within the nicotinic acetylcholine receptor transmembrane domain: structure and collective motions.* Biophys J, 2005. **88**: p. 3321-33.
138. Reeves, D.C. and Lummis, S.C.R., *The molecular basis of the structure and function of the 5-HT3 receptor: a model ligand-gated ion channel (Review).* Mol Membr Biol, 2002. **19**: p. 11-26.
139. Corry, B., *Theoretical conformation of the closed and open states of the acetylcholine receptor channel.* Biochim Biophys Acta, 2004. **1663**: p. 2-5.
140. Corry, B., *An energy-efficient gating mechanism in the acetylcholine receptor channel suggested by molecular and Brownian dynamics.* Biophys J, 2006. **90**: p. 799-810.
141. Marti-Renom, M.A., Stuart, A., Fiser, A., Sanchez, R., Melo, F., and Sali, A., *Comparative protein structure modeling of genes and genomes.* Annu Rev Biophys Biomol Struct, 2000. **29**: p. 291-325.
142. Sali, A. and Blundell, T.L., *Comparative protein modelling by satisfaction of spatial restraints.* J Mol Biol, 1993779-815.
143. Fiser, A., Do, R.K., and Sali, A., *Modeling of loops in protein structures.* Protein Sci, 2000. **9**: p. 1753-73.
144. Corbet, F., *MultAlin.* Nucleic Acids Res, 1988. **16**: p. 10881-90.
145. Unwin, N. and Henderson, R., *The structure of proteins in biological membranes.* Sci Am, 1984. **250**: p. 78-94.
146. Leach, A.R., *Molecular modelling. Principles and applications*. 2001: Longman.
147. van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Mark, A.E., Scott, W.R.P., Tironi, I.G., and Kruger, P., *Biomolecular Simulation: The GROMOS96 Manual and User Guide.* 1996, Zurick: Groningen & Zurich, Biomos & Hochschulverlag AG an der ETH Zurick.
148. Becker, O.M., J., A.D.M., Roux, B., and Watanabe, M., eds. *Computational Biochemistry and Biophysics*. 2001, Marcel Dekker: New York.
149. Verlet, L., *Computer experiments on classical fluids. I. Thermodynamical properties of Lennard Jones molecules.* Phys Rev, 1967. **159**: p. 98-103.
150. Hockney, R.W., Goel, S.P., and Eastwood, J.W., *Quiet high-resolution computer models of a plasma.* J Comp Chem, 1974. **14**: p. 148-58.
151. Weiner, S.J. and Kollman, P.A., *A new force field for molecular mechanical simulation of nucleic acids and proteins.* J Am Chem Soc, 1984. **106**: p. 765-84.

152. Brooks, B.R., Bruccoleri, R.E., and Olafson, B.D., *CHARMM: A program for macromolecular energy, minimisation, and dynamics calculations.* J Comp Chem, 1993. **4**: p. 187-217.
153. Hermans, J. and Berendsen, H.J.C., *A consistent empirical potential for water-protein interactions.* Biopolymers, 1984. **23**: p. 1513-18.
154. Jorgensen, W.L. and Tirado-Rives, J., *Development of the OPLS-AA force field for organic and biomolecular systems.* Abs Pap Am Chem Soc, 1998. **043**: p.
155. Ewald, P.P., *Die Berechnung optischer und elektrostatischer Gitterpotentiale.* Ann. d. Phys., 1921. **64**: p. 253-87.
156. Dominy, B.N. and Brooks, C.L., *Development of a generalized born model parametrization for proteins and nucleic acids.* J Phys Chem B, 1999. **103**: p. 3765-73.
157. Sarman, S.S., Evans, D.J., and Cummings, P.T., *Recent developments in non-Newtonian molecular dynamics.* Phys Rep-Rev Phys Lett, 1998. **305**: p. 1-92.
158. Berendsen, H.J.C., Postma, J.P.M., Vangunsteren, W.F., Dinola, A., and Haak, J.R., *Molecular-Dynamics with Coupling to an External Bath.* J Chem Phys, 1984. **81**: p. 3684-90.
159. Nose, S., *A Molecular-Dynamics Method for Simulations in the Canonical Ensemble.* Mol Phys, 1984. **52**: p. 255-68.
160. Hoover, W.G., *Canonical Dynamics - Equilibrium Phase-Space Distributions.* Phys Rev A, 1985. **31**: p. 1695-97.
161. Parrinello, M. and Rahman, A., *Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method.* J Appl Phys, 1981. **52**: p. 7182-90.
162. Ryckaert, J.P. and Ciccotti., G., *Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* J Comput Phys, 1977. **23**: p.
163. Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M., *LINCS: A linear constraint solver for molecular simulations.* J Comp Chem, 1997. **18**: p. 1463-72.
164. Bahar, I., Atilgan, A.R., and Erman, B., *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential.* Fold Des, 1997. **2**: p. 173-81.
165. Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O., and Bahar, I., *Anisotropy of fluctuation dynamics of proteins with an elastic network model.* Biophys J, 2001. **80**: p. 505-15.
166. Erkip, A., *Dynamics of large-scale fluctuations in native proteins. Analysis based on harmonic inter-residue potentials and random external noise.* Polymer, 2004. **45**: p. 641-48.
167. Bahar, I., Erman, B., Jernigan, R.L., Atilgan, A.R., and Covell, D.G., *Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function.* J Mol Biol, 1999. **285**: p. 1023-37.
168. Isin, B., Doruker, P., and Bahar, I., *Functional motions of influenza virus hemagglutinin: a structure-based analytical approach.* Biophys J, 2002. **82**: p. 569-81.
169. Eichinger, B.E., *Elasticity Theory. I. Distribution functions for perfect phantom networks.* Macromolecules, 1972. **5**: p. 496-505.
170. Tirion, M.M., *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis.* Phys Rev Lett, 1996. **77**: p. 1905-08.

171. de Groot, B.L., van Aalten, D.M., Scheek, R.M., Amadei, A., Vriend, G., and Berendsen, H.J., *Prediction of protein conformational freedom from distance constraints.* Proteins, 1997. **29**: p. 240-51.
172. Bond, P.J. and Sansom, M.S., *Insertion and assembly of membrane proteins via simulation.* J Am Chem Soc, 2006. **128**: p. 2697-704.
173. Marrink, S.J. and Mark, A.E., *Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles.* J Am Chem Soc, 2003. **125**: p. 15233-42.
174. Marrink, S.J. and Mark, A.E., *Molecular view of hexagonal phase formation in phospholipid membranes.* Biophys J, 2004. **87**: p. 3894-900.
175. Bond, P., *Coarse-grained molecular dynamics simulations of membrane proteins and peptides.* J Struct Biol, 2006.
176. Smith, L.I., *A tutorial on Principal Components Analysis*. 2002: Oxford.
177. Tai, K., *Conformational sampling for the impatient.* Biophys Chem, 2004. **107**: p. 213-20.
178. Barrett, C.P., Hall, B.A., and Noble, M.E., *Dynamite: a simple way to gain insight into protein motions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**: p. 2280-7.
179. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J., *Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function.* J Comput Chem, 1998. **19**: p. 1639 - 62.
180. Hehre, W., Yu, J., and Klunzinger, P., *A Guide to Molecular Mechanics and Molecular Orbital Calculations in Spartan.* 1997.
181. Debye, P. and Huckel, E., *Zur Theorie der Elecktrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen.* Phys Z, 1923. **24**: p. 185-206.
182. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCAmmon J.A. Proc Natl Acad Sci U S A, *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc Natl Acad Sci U S A, 2001. **98**: p. 10037-41.
183. Smart, O.S., Neduvelil, J.G., Wang, X., Wallace, B.A., and Sansom, M.S., *HOLE: a program for the analysis of the pore dimensions of ion channel structural models.* J Mol Graph, 1996. **14**: p. 354-60, 76.
184. Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., and Baker, N.A., *PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations.* Nucleic Acids Res, 2004. **32**: p. W665-7.
185. Ulens, C., Hogg, R.C., Celie, P.H., Bertrand, D., Tsetlin, V., Smit, A.B., and Sixma, T.K., *Structural determinants of selective alpha-conotoxin binding to a nicotinic acetylcholine receptor homolog AChBP.* Proc Natl Acad Sci U S A, 2006. **103**: p. 3615-20.
186. Laskowski, R.A., Moss, D.S., and Thornton, J.M., *Main-chain bond lengths and bond angles in protein structures.* J Mol Biol, 1993. **231**: p. 1049-67.
187. Lindahl, E., Hess, B., and van der Spoel, D., *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* J Mol Model, 2001. **7**: p. 306-17.
188. Berendsen, H.J.C., Vanderspoel, D., and Vandrunen, R., *Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation.* Comp Phys Comm, 1995. **91**: p. 43-56.

189. Darden, T., York, D., and Pedersen, L., *Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems.* J Chem Phys, 1993. **98**: p. 10089-92.
190. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G., *A Smooth Particle Mesh Ewald Method.* J Chem Phys, 1995. **103**: p. 8577-93.
191. Graton, J., Berthelot, M., Gal, J.F., Girard, S., Laurence, C., Lebreton, J., Le Questel, J.Y., Maria, P.C., and Naus, P., *Site of protonation of nicotine and nornicotine in the gas phase: pyridine or pyrrolidine nitrogen?* J Am Chem Soc, 2002. **124**: p. 10552-62.
192. Pankow, J.F., Barsanti, K.C., and Peyton, D.H., *Fraction of free-base nicotine in fresh smoke particulate matter from the eclipse "cigarette" by H-1 NMR spectroscopy.* Chem Res Toxicol, 2003. **16**: p. 23-27.
193. Schuttelkopf, A.W. and van Aalten, D.M.F., *PRODRG: a tool for high-throughput crystallography of protein-ligand complexes.* Acta Crystallogr D-Biol Crystallogr, 2004. **60**: p. 1355-63.
194. Efange, S.M., Tu, Z., von Hohenberg, K., Francesconi, L., Howell, R.C., Rampersad, M.V., Todaro, L.J., Papke, R.L., and Kung, M.P., *2-(2-Piperidyl)- and 2-(2-pyrrolidyl)chromans as nicotine agonists: synthesis and preliminary pharmacological characterization.* J Med Chem, 2001. **44**: p. 4704-15.
195. Glennon, R.A. and Dukat, M., *Central nicotinic receptor ligands and pharmacophores.* Pharm Acta Helv, 2000. **74**: p. 103-14.
196. Guan, J.W. and Fischer, R.D., *Tris(indenyl)lanthanoid complexes (Ln = La, Pr, Nd) containing either (S)-(-)-nicotine or two simpler pyridine bases.* Eur J Inorg Chem, 20012497-508.
197. Chang, Y. and Weiss, D.S., *Channel opening locks agonist onto the GABAC receptor.* Nat Neurosci, 1999. **2**: p. 219-25.
198. Sheridan, R.E. and Lester, H.A., *Functional stoichiometry at the nicotinic receptor. The photon cross section for phase 1 corresponds to two bis-Q molecules per channel.* J Gen Physiol, 1982. **80**: p. 499-515.
199. Wallace, A.C., Laskowski, R.A., and Thornton, J.M., *Ligplot - a Program to Generate Schematic Diagrams of Protein Ligand Interactions.* Protein Eng, 1995. **8**: p. 127-34.
200. Humphrey, W., Dalke, A., and Schulten, K., *VMD: Visual molecular dynamics.* J Mol Graph, 1996. **14**: p. 33-38.
201. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E., *UCSF chimera - A visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**: p. 1605-12.
202. Faraldo-Gomez, J.D., Forrest, L.R., Baaden, M., Bond, P.J., Domene, C., Patargias, G., Cuthbertson, J., and Sansom, M.S.P., *Conformational Sampling and Dynamics of Membrane Proteins From 10-nanosecond Computer Simulations.* Proteins: Struct Func Bioinfo, 2004. **57**: p. 783-91.
203. Barrett, C.P. and Noble, M.E., *Dynamite extended: two new services to simplify protein dynamic analysis.* Bioinformatics, 2005. **21**: p. 3174-5.
204. Ming, D., Kong, Y., Lambert, M.A., Huang, Z., and Ma, J., *How to describe protein motion without amino acid sequence and atomic coordinates.* Proc Natl Acad Sci U S A, 2002. **99**: p. 8620-5.
205. Xu, C.Y., Tobi, D., and Bahar, I., *Allosteric changes in protein structure computed by a simple mechanical model: Hemoglobin T <-> R2 transition.* J Mol Biol, 2003. **333**: p. 153-68.

206. Law, R.J., Capener, C., Baaden, M., Bond, P.J., Campbell, J., Patargias, G., Arinaminpathy, Y., and Sansom, M.S., *Membrane protein structure quality in molecular dynamics simulation.* J Mol Graph Model, 2005. **24**: p. 157-65.
207. Hansen, S.B., Radic, Z., Talley, T.T., Molles, B.E., Deerinck, T., Tsigelny, I., and Taylor, P., *Tryptophan fluorescence reveals conformational changes in the acetylcholine binding protein.* J Biol Chem, 2002. **277**: p. 41299-302.
208. Hansen, S.B., Talley, T.T., Radic, Z., and Taylor, P., *Structural and ligand recognition characteristics of an acetylcholine-binding protein from Aplysia californica.* J Biol Chem, 2004. **279**: p. 24197-202.
209. Watts, A., *Solid-state NMR in drug design and discovery for membrane-embedded targets.* Nat Rev Drug Discov, 2005. **4**: p. 555-68.
210. Williamson, P.T., Grobner, G., Spooner, P.J., Miller, K.W., and Watts, A., *Probing the agonist binding pocket in the nicotinic acetylcholine receptor: a high-resolution solid-state NMR approach.* Biochemistry, 1998. **37**: p. 10854-9.
211. Williamson, P.T., Watts, J., Grobner, G., Miller, K.W., and Watts, A., *Solid state NMR studies of ligands bound to the nicotinic acetylcholine receptor.* Biochem Soc Trans, 1998. **26**: p. S297.
212. Chen, P.E. and Wyllie, D.J., *Pharmacological insights obtained from structure-function studies of ionotropic glutamate receptors.* Br J Pharmacol, 2006. **147**: p. 839-53.
213. Speranskiy, K. and Kurnikova, M., *On the binding determinants of the glutamate agonist with the glutamate receptor ligand binding domain.* Biochemistry, 2005. **44**: p. 11508-17.
214. Chung, E., Henriques, D., Renzoni, D., Zvelebil, M., Bradshaw, J., Waksman, G., Robinson, C., and and Ladbury, J., *Mass spectrometric and thermodynamic studies reveal the role of water molecules in complexes formed between SH2 domains and tyrosyl phosphopeptides.* Structure, 1998. **6**: p. 1141-51.
215. Speranskiy, K. and Kurnikova, M., *Accurate theoretical prediction of vibrational frequencies in an inhomogeneous dynamic environment: A case study of a glutamate molecule in water solution and in a protein-bound form.* J Chem Phys, 2004. **121**: p. 1516-24.
216. Goldbaum, G., Schwarz, F., Eisenstein, E., Cauerhff, A., Mariuzza, R., and and Poljak, R., *The effect of water on the association constant and the enthalpy of reaction between lysozyme and the specific antibodies D1.3 and D33.1.* J Mol Recog, 1996. **9**: p. 6-12.
217. Ladbury, J., *Just add water! The effect of water on the specificity of protein-ligand binding sites and its application to drug design.* Chem Biol, 1996. **3**: p. 973-80.
218. Mikol, V., Papageorgiou, C., and and Borer, X., *The role of water molecules in structure-based design of (5-hydroxynorvaline)-2-cyclosporin: synthesis, biological activity and crystallographic analysis with cyclophilin.* J Med Chem, 1995. **38**: p. 3361-67.
219. Tame, J., Sleigh, S., Wilkinson, A., and and Ladbury, J., *The role of water in sequence-independent ligand binding by an oligopeptide transporter protein.* Nat Struct Biol, 1996. **3**: p. 998-1001.
220. Kao, P.N. and Karlin, A., *Acetylcholine receptor binding site contains a disulfide cross-link between adjacent half-cystinyl residues.* J Biol Chem, 1986. **261**: p. 8085-8.

221. Samson, A.O., Chill, J.H., Rodriguez, E., Scherf, T., and Anglister, J., *NMR mapping and secondary structure determination of the major acetylcholine receptor alpha-subunit determinant interacting with alpha-bungarotoxin.* Biochemistry, 2001. **40**: p. 5464-73.
222. Broughton, H.B., *A method for including protein flexibility in protein-ligand docking: Improving tools for database mining and virtual screening.* J Mol Graph Mod, 2000. **18**: p. 247-57, 302-4.
223. Alonso, H., Gillies, M.B., Cummins, P.L., Bliznyuk, A.A., and Gready, J.E., *Multiple ligand-binding modes in bacterial R67 dihydrofolate reductase.* J Comput Aided Mol Des, 2005. **19**: p. 165-87.
224. Gorse, A.D. and Gready, J.E., *Molecular dynamics simulations of the docking of substituted N5-deazapterins to dihydrofolate reductase.* Protein Eng, 1997. **10**: p. 23-30.
225. Segall, M.D., Payne, M.C., and Boyes, R.N., *An ab initio study of the conformational energy map of acetylcholine.* Mol Phys, 1998. **93**: p. 365-70.
226. Unwin, N., *Nicotinic acetylcholine receptor at 9 A resolution.* J Mol Biol, 1993. **229**: p. 1101-24.
227. Unwin, N., *Structure of the acetylcholine-gated channel.* Novartis Found Symp, 2002. **245**: p. 5-15; discussion 15-21, 165-8.
228. Dutertre, S. and Lewis, R.J., *Computational approaches to understand alpha-conotoxin interactions at neuronal nicotinic receptors.* Eur J Biochem, 2004. **271**: p. 2327-34.
229. Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* J Biomol NMR, 1996. **8**: p. 477-86.
230. Bahar, I., Atilgan, A.R., Jernigan, R.L., and Erman, B., *Understanding the recognition of protein structural classes by amino acid composition.* Proteins, 1997. **29**: p. 172-85.
231. Smart, O.S., Coates, G.M., Sansom, M.S., Alder, G.M., and Bashford, C.L., *Structure-based prediction of the conductance properties of ion channels.* Faraday Discuss, 1998185-99; discussion 225-46.
232. Chen, J. and Auerbach, A., *A distinct contribution of the delta subunit to acetylcholine receptor channel activation revealed by mutations of the M2 segment.* Biophys J, 1998. **75**: p. 218-25.
233. Amiri, S., Tai, K., Beckstein, O., Biggin, P.C., and Sansom, M.S., *The alpha7 nicotinic acetylcholine receptor: molecular modelling, electrostatics, and energetics.* Mol Membr Biol, 2005. **22**: p. 151-62.
234. Zhang, Z., Shi, Y., and Liu, H., *Molecular dynamics simulations of peptides and proteins with amplified collective motions.* Biophys J, 2003. **84**: p. 3583-93.
235. Rashin, A.A. and Honig, B., *Cavities and Internal Bound Waters in Proteins.* Biophys J, 1985. **47**: p. A20-A20.
236. Cheng, X., Lu, B., Grant, B., Law, R.J., and McCammon, J.A., *Channel opening motion of alpha7 nicotinic acetylcholine receptor as suggested by normal mode analysis.* J Mol Biol, 2006. **355**: p. 310-24.
237. Taly, A., Delarue, M., Grutter, T., Nilges, M., Le Novere, N., Corringer, P.J., and Changeux, J.P., *Mormal Mode Analysis Suggests a Quaternary Twist Model for the Nicotinic Receptor Gating Mechanism.* Biophys J, 2005. **88**: p. 3954-65.

238. Allen, T.W., Andersen, O.S., and Roux, B., *Energetics of ion conduction through the gramicidin channel.* Proc Natl Acad Sci U S A, 2004. **101**: p. 117-22.
239. Unwin, N., *The structure of ion channels in membranes of excitable cells.* Neuron, 1989. **3**: p. 665-76.
240. McLachlan, A.D., *Rapid comparison of protein structures.* Acta Cryst A, 1982. **38**: p. 871-73.
241. Opella, S.J., Marassi, F.M., Gesell, J.J., Valente, A.P., Kim, Y., Oblatt-Montal, M., and Montal, M., *Structures of the M2 channel-lining segments from nicotinic acetylcholine and NMDA receptors by NMR spectroscopy.* Nat Struct Biol, 1999. **6**: p. 374-9.
242. Akabas, M.H. and Karlin, A., *Identification of acetylcholine receptor channel-lining residues in the M1 segment of the alpha-subunit.* Biochem, 1995. **34**: p. 12496-500.
243. Goh, C.S., Milburn, D., and Gerstein, M., *Conformational changes associated with protein-protein interactions.* Curr Opin Struct Biol, 2004. **14**: p. 104-9.
244. Sansom, M.S.P., Biggin, P.C., Beckstein, O., and Tieleman, D.P., *Water in transmembrane pores: Simulation studies.* Abs Pap Am Chem Soc, 2001. **221**: p. U288-U88.
245. Beckstein, O. and Sansom, M.S., *A hydrophobic gate in an ion channel: the closed state of the nicotinic acetylcholine receptor.* Phys Biol, 2006. **3**: p. 147-59.
246. Culetto, E., Baylis, H.A., Richmond, J.E., Jones, A.K., Fleming, J.T., Squire, M.D., Lewis, J.A., and Sattelle, D.B., *The Caenorhabditis elegans unc-63 Gene Encodes a Levamisole-sensitive Nicotinic Acetylcholine Receptor a-subunit.* J Biol Chem, 2004. **279**: p. 42476-83.
247. Fleming, J.T., Squire, M.D., Barnes, T.M., Tornoe, C., Matsuda, K., Ahnn, J., Fire, A., Sulston, J.E., Barnard, E.A., Sattelle, D.B., and Lewis, J.A., *Caenorhabditis elegans levamisole resistance genes lev-1, unc-29, and unc-38 encode functional nicotinic acetylcholine receptor subunits.* J Neurosci, 1997. **17**: p. 5843-57.
248. Towers, P.R., Edwards, B., Richmond, J.E., and Sattelle, D.B., *The Caenorhabditis elegans lev-8 gene encodes a novel type of nicotinic acetylcholine receptor alpha subunit.* J Neurochem, 2005. **93**: p. 1-9.
249. Richmond, J.E. and Jorgensen, E.M., *One GABA and two acetylcholine receptors function at the C-elegans neuromuscular junction.* Nat Neurosci, 1999. **2**: p. 791-97.
250. Martin, R.J., *Modes of action of anthelmintic drugs.* Vet J, 1997. **154**: p. 11-34.
251. Rayes, D., De Rosa, M.J., Bartos, M., and Bouzat, C., *Molecular basis of the differential sensitivity of nematode and mammalian muscle to the anthelmintic agent levamisole.* J Biol Chem, 2004. **279**: p. 36372-81.
252. Francis, M.M., Evans, S.P., Jensen, M., Madsen, D.M., Mancuso, J., Norman, K.R., and Maricq, A.V., *The Ror receptor tyrosine kinase CAM-1 is required for ACR-16-mediated synaptic transmission at the C. elegans neuromuscular junction.* Neuron, 2005. **46**: p. 581-94.
253. Touroutine, D., Fox, R.M., Von Stetina, S.E., Burdina, A., Miller, D.M., and Richmond, J.E., *acr-16 encodes an essential subunit of the levamisole-resistant nicotinic receptor at the Caenorhabditis elegans neuromuscular junction.* J Biol Chem, 2005. **280**: p. 27013-21.

254. DeLano, W.L., *Use of PYMOL as a communications tool for molecular science.* Abs Pap Am Chem Soc, 2004. **228**: p. U313-U14.
255. DeLano, W.L. and Lam, J.W., *PyMOL: A communications tool for computational models.* Abs Pap Am Chem Soc, 2005. **230**: p. U1371-U72.
256. Shimomura, M., Yokota, M., Okumura, M., Matsuda, K., Akamatsu, M., Sattelle, D.B., and Komai, K., *Combinatorial mutations in loops D and F strongly influence responses of the alpha7 nicotinic acetylcholine receptor to imidacloprid.* Brain Res, 2003. **991**: p. 71-7.
257. Matsuda, K., Shimomura, M., Kondo, Y., Ihara, M., Hashigami, K., Yoshida, N., Raymond, V., Mongan, N.P., Freeman, J.C., Komai, K., and Sattelle, D.B., *Role of loop D of the alpha7 nicotinic acetylcholine receptor in its interaction with the insecticide imidacloprid and related neonicotinoids.* Br J Pharmacol, 2000. **130**: p. 981-6.
258. Ihara, M., Matsuda, K., Otake, M., Kuwamura, M., Shimomura, M., Komai, K., Akamatsu, M., Raymond, V., and Sattelle, D.B., *Diverse actions of neonicotinoids on chicken alpha7, alpha4beta2 and Drosophila-chicken SADbeta2 and ALSbeta2 hybrid nicotinic acetylcholine receptors expressed in Xenopus laevis oocytes.* Neuropharmacology, 2003. **45**: p. 133-44.
259. Sattelle, D.B., Jones, A.K., Sattelle, B.M., Matsuda, K., Reenan, R., and Biggin, P.C., *Edit, cut and paste in the nicotinic acetylcholine receptor gene family of Drosophila melanogaster.* Bioessays, 2005. **27**: p. 366-76.
260. Shimomura, M., Okuda, H., Matsuda, K., Komai, K., Akamatsu, M., and Sattelle, D.B., *Effects of mutations of a glutamine residue in loop D of the alpha7 nicotinic acetylcholine receptor on agonist profiles for neonicotinoid insecticides and related ligands.* Br J Pharmacol, 2002. **137**: p. 162-9.
261. Shimomura, M., Yokota, M., Matsuda, K., Sattelle, D.B., and Komai, K., *Roles of loop C and the loop B-C interval of the nicotinic receptor alpha subunit in its selective interactions with imidacloprid in insects.* Neurosci Lett, 2004. **363**: p. 195-8.
262. Kukhtina, V., Kottwitz, D., Strauss, H., Heise, B., Chebotareva, N., Tsetlin, V., and Hucho, F., *Intracellular domain of nicotinic acetylcholine receptor: the importance of being unfolded.* J Neurochem, 2006. **1**: p. 63-67.
263. Hales, T.G., Dunlop, J.I., Deeb, T.Z., Carland, J.E., Kelley, S.P., Lambert, J.J., and Peters, J.A., *Common determinants of single channel conductance within the large cytoplasmic loop of 5-hydroxytryptamine type 3 and alpha4beta2 nicotinic acetylcholine receptors.* J Biol Chem, 2006. **281**: p. 8062-71.
264. Grutter, T., de Carvalho, L.P., Dufresne, V., Taly, A., Edelstein, S.J., and Changeux, J.P., *Molecular tuning of fast gating in pentameric ligand-gated ion channels.* Proc Natl Acad Sci U S A, 2005. **102**: p. 18207-12.
265. Bond, P.J., Holyoake, J., Ivetac, A., Khalid, S., and Sansom, M.S.P., *COARSE-GRAINED MOLECULAR DYNAMICS SIMULATIONS OF MEMBRANE PROTEINS AND PEPTIDES.* Journal of Structural Biology, 2006. **104**: p.
266. Betz, H., *Ligand-Gated Ion Channels in the Brain - the Amino-Acid Receptor Superfamily.* Neuron, 1990. **5**: p. 383-92.
267. Ortells, M.O. and Lunt, G.G., *Evolutionary history of the ligand-gated ion-channel superfamily of receptors.* Perspectives, 1995. **18**: p. 121-27.
268. Higgins, D., *Hot Papers - Computer-Science Life Sciences - Clustal-V - Improved Software for Multiple Sequence Alignment - Higgins,D.G., Bleasby,A.J., Fuchs,R.* Scientist, 1994. **8**: p. 17-17.

269. Thompson, J.D., Higgins, D.G., and Gibson, T.J., *Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice.* Nucleic Acids Res, 1994. **22**: p. 4673-80.
270. Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J., *The Jalview Java alignment editor.* Bioinformatics, 2004. **20**: p. 426-7.
271. Miyazawa, A., Fujiyoshi, Y., Stowell, M., and Unwin, N., *Nicotinic acetylcholine receptor at 4.6 A resolution: transverse tunnels in the channel wall.* J Mol Biol, 1999. **288**: p. 765-86.
272. Unwin, N., *Neurotransmitter action: opening of ligand-gated ion channels.* Cell, 1993. **72 Suppl**: p. 31-41.
273. Chang, H.W. and Bock, E., *Structural stabilization of isolated acetylcholine receptor: specific interaction with phospholipids.* Biochemistry, 1979. **18**: p. 172-9.
274. daCosta, C.J., Ogrel, A.A., McCardy, E.A., Blanton, M.P., and Baenziger, J.E., *Lipid-protein interactions at the nicotinic acetylcholine receptor. A functional coupling between nicotinic receptors and phosphatidic acid-containing lipid bilayers.* J Biol Chem, 2002. **277**: p. 201-8.
275. Fong, T.M. and McNamee, M.G., *Correlation between acetylcholine receptor function and structural properties of membranes.* Biochemistry, 1986. **25**: p. 830-40.
276. Fong, T.M. and McNamee, M.G., *Stabilization of acetylcholine receptor secondary structure by cholesterol and negatively charged phospholipids in membranes.* Biochemistry, 1987. **26**: p. 3871-80.
277. Jones, O.T., Eubanks, J.H., Earnest, J.P., and McNamee, M.G., *A minimum number of lipids are required to support the functional properties of the nicotinic acetylcholine receptor.* Biochemistry, 1988. **27**: p. 3733-42.
278. Barrantes, F.J., *Modulation of nicotinic acetylcholine receptor function through the outer and middle rings of transmembrane domains.* Curr Opin Drug Discov Devel, 2003. **6**: p. 620-32.
279. Blanton, M.P. and Cohen, J.B., *Mapping the lipid-exposed regions in the Torpedo californica nicotinic acetylcholine receptor.* Biochemistry, 1992. **31**: p. 3738-50.
280. Xu, Y., Barrantes, F.J., Luo, X., Chen, K., Shen, J., and Jiang, H., *Conformational dynamics of the nicotinic acetylcholine receptor channel: a 35-ns molecular dynamics simulation study.* J Am Chem Soc, 2005. **127**: p. 1291-9.
281. Bouzat, C., Roccamo, A.M., Garbus, I., and Barrantes, F.J., *Mutations at lipid-exposed residues of the acetylcholine receptor affect its gating kinetics.* Mol Pharmacol, 1998. **54**: p. 146-53.
282. Lasalde, J.A., Tamamizu, S., Butler, D.H., Vibat, C.R., Hung, B., and McNamee, M.G., *Tryptophan substitutions at the lipid-exposed transmembrane segment M4 of Torpedo californica acetylcholine receptor govern channel gating.* Biochemistry, 1996. **35**: p. 14139-48.
283. Li, L., Lee, Y.H., Pappone, P., Palma, A., and McNamee, M.G., *Site-specific mutations of nicotinic acetylcholine receptor at the lipid-protein interface dramatically alter ion channel gating.* Biophys J, 1992. **62**: p. 61-3.
284. Ortiz-Miranda, S.I., Lasalde, J.A., Pappone, P.A., and McNamee, M.G., *Mutations in the M4 domain of the Torpedo californica nicotinic acetylcholine*

receptor alter channel opening and closing. J Membr Biol, 1997. **158**: p. 17-30.

285. Tamamizu, S., Guzman, G.R., Santiago, J., Rojas, L.V., McNamee, M.G., and Lasalde-Dominicci, J.A., *Functional effects of periodic tryptophan substitutions in the alpha M4 transmembrane domain of the Torpedo californica nicotinic acetylcholine receptor.* Biochemistry, 2000. **39**: p. 4666-73.
286. Tamamizu, S., Lee, Y., Hung, B., McNamee, M.G., and Lasalde-Dominicci, J.A., *Alteration in ion channel function of mouse nicotinic acetylcholine receptor by mutations in the M4 transmembrane domain.* J Membr Biol, 1999. **170**: p. 157-64.
287. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis.* Proc Natl Acad Sci USA, 1958. **44**: p. 98-104.
288. Holyoake, J., Caulfeild, V., Baldwin, S., and Sansom, M.S., *Modelling, Docking and Simulation of the Major Facilitator Superfamily.* Biophys J, 2006.
289. Haider, S., Grottesi, A., Hall, B.A., Ashcroft, F.M., and Sansom, M.S., *Conformational dynamics of the ligand-binding domain of inward rectifier K channels as revealed by molecular dynamics simulations: toward an understanding of Kir channel gating.* Biophys J, 2005. **88**: p. 3310-20.

26/07/2006

Molecular Dynamics Studies of AChBP with Nicotine and Carbamylcholine: The Role of Water in the Binding Pocket

Shiva Amiri, Mark S.P. Sansom and Philip C. Biggin*
Structural Bioinformatics and Computational Biochemistry,
Department of Biochemistry,
The University of Oxford,
South Parks Road,
Oxford,
OX1 3QU,
U.K.

*to whom correspondence should be directed at:
E_mail: philip.biggin@bioch.ox.ac.uk
Telephone: +44 1865 275255
Fax: +44 1865 275273

For PEDS *v 16*

26/07/2006

Abstract

The acetylcholine-binding protein (AChBP) is homologous to the ligand-binding domain of the nicotinic acetylcholine receptor (nAChR) and other members of the Cys-loop family of neurotransmitter receptors. The high resolution X-ray structures of AChBP mean it has been used as a model from which to understand agonist and antagonist binding to nAChRs. We present here a molecular dynamics (MD) study of AChBP with nicotine and carbamylcholine bound. Our results suggest that the ligand imposes rigidity on the binding pocket residues, which is indicative of an induced-fit model for ligand binding. The simulations also suggest that the protein undergoes breathing motions with respect to the five-fold axis, a motion that has been postulated to be related to gating in the nicotinic acetylcholine receptor. We analyzed the behaviour of the water molecules in and around the binding site and found that they occupied five distinct sites within the binding pocket. Water occupied these sites in the absence of ligand, but the presence of ligand increased the probability that a water molecule would be found in these sites. Finally we demonstrate how the positions of these waters might be used in the design of new ligands by comparing the positions of these sites with other recent structures.

Introduction

The acetylcholine binding protein (AChBP) is a ligand binding domain (LBD) homologue of the pentameric ligand gated ion channel (LGIC) family of membrane proteins (Brejc et al. 2001; Sixma and Smit 2003; Smit et al. 2003; Smit et al. 2001). The nicotinic acetylcholine receptor (nAChR), the serotonin receptor ($5HT_3$), the $GABA_{A,C}$ receptors and the Glycine receptor make up the Cys-loop family of LGICs (Ashcroft 2000; Lester *et al.* 2004) and are characterized by a conserved pair of cysteines in one of the LBD loops (Karlin 2002). Each subunit is composed of an N-terminal LBD, a transmembrane (TM) domain, and an intracellular (IC) domain. Various combinations of subunits can form homomeric or heteromeric ion channels. The structures of AChBP from different species have been solved by X-ray crystallography with a variety of ligands (Bourne *et al.* 2005; Brejc *et al.* 2001; Brejc *et al.* 2002; Celie *et al.* 2005a; Celie *et al.* 2005b; Celie *et al.* 2004; Hansen *et al.* 2005) that have allowed important comparisons to made back to nAChR in particular.

The structure of AChBP (Fig. 1A.) reveals that the individual subunits are comprised of one α-helix, 10 stranded β sheets, and two 3_{10} helices (Brejc, 2001; Sixma and Smit, 2003; Smit *et al.*, 2003). The binding site is located between two adjacent subunits, a principle subunit (on the left in Fig 1A) and a complementary subunit (on the right in Fig 1A). Thus there are up to five binding sites within the pentameric unit. In the nAChR, at least two of the subunits must be of α subtype which forms the principle subunit of the binding site. In some cases the receptor can be comprised of all α subunits (for example receptors comprised of the α7 subunit), and thus will possess five binding sites.

The LBD of nAChR has high sequence identity with AChBP from *Lymnaea stagnalis* (Sixma and Smit, 2003; Smit *et al.*, 2003). The nAChR is found in the nervous system at neuronal-neuronal synapses and in the muscle at neuromuscular junctions (Nicke, et. al., 2004). The nAChR plays an important role in nicotine addiction and is implicated in various neurological and neuromuscular diseases. Mutations in the nAChR can results in myasthenic syndromes and frontal lobe epilepsy (De Fusco *et al.* 2000; Steinlein 2001), and various cognitive malfunctions (Ashcroft 2000). This channel is implicated in Parkinson's and Alzheimer's disease but these mechanisms are not well understood (Hogg and Bertrand 2004). The nAChR is a cationic channel which allows the passage of Na^+, K^+, and less commonly Ca^{2+} ions.

Agonist binding to nAChRs initiates conformational changes in the receptor that are transmitted to the TM domain which ultimately results in the opening of the ion channel to allow the conduction of ions (Karlin 2002). The mechanism that leads from ligand binding to channel opening is incompletely understood (Cymes *et al.* 2002; Grosman 2003). The AChBP structures have provided important clues as to how different ligands are recognized by this fold, but the precise extent and nature of the binding pocket and its behaviour is still poorly understood.

Molecular dynamics provides a useful way to examine the dynamic behaviour of proteins and indeed has been recently applied to AChBP (Gao et al. 2005) and also to models of the α7 nAChR (Henchman *et al.* 2003; Law *et al.* 2005), where the focus was on examining the large-scale motion of the protein thought to be related to gating in nAChRs. Here, we focus not on the global motions but on the dynamics of residues and water in the binding pocket of the AChBP. Our results show that 1) A more stable

structure of the binding pocket is maintained in the presence of a ligand; 2) ligands exhibit varying degrees of mobility within the binding pocket - nicotine seems to prefer one mode of binding while carbamylcholine shows some flexibility of its chain; and 3) waters exist in discrete pockets inside the binding pocket that maybe useful in the design of ligands, We illustrate this latter point with respect to the crystal structures of AChBP in complex with three toxins.

Methods

Molecular dynamics simulations of AChBP from *Lymnea stagnalis* (Celie *et. al*, 2004) with and without nicotine and carbamylcholine were carried out for 10 ns each. The simulation with nicotine had five nicotine molecules bound, one to each binding site. The simulation with carbamylcholine had 2 molecules bound to the binding pockets located between subunits C and D and subunits D and E (the three remaining binding pockets were ligand-free; see Table 1).

Molecular dynamics simulations were carried out with GROMACS v 3.1.4 (www.gromacs.org) (Berendsen *et al.* 1995; Lindahl *et al.* 2001) using the GROMOS96 (van Gunsteren et al. 1996) force-field. Crystal structures of AChBP from *Lymnaea stagnalis* with nicotine (Fig. 1B.) (PDB code: 1UW6) and AChBP with carbamylcholine (Fig. 1B.) (PDB code: 1UV6) (Celie *et al.*, 2004) were used including their crystal waters (within 4 Å of the protein) for the liganded simulations. Apo structures were generated by removing the ligand. Topologies were generated for both nicotine and carbamylcholine using PRODRG (Schuettelkopf and van Aalten 2004). The charges for

nicotine and carbamylcholine were calculated with 6-31G* basis sets using Spartan (Wavefunction Inc). The pyrrolidine nitrogen (Fig. 1B) was assumed to be protonated (Graton *et al.* 2002; Pankow *et al.* 2003). Ligand topologies were tested by simulation in a box of approximately 900 simple point charge (SPC, (Hermans et al. 1984)) water molecules for 1 ns.

Four protein simulations were carried out in total (see Table 1) corresponding to both ligand-bound and apo conditions. Each system was energy-minimized until convergence using steepest descents algorithm. Energy minimization was carried out with position restraints, then molecular dynamics with position restraints for 100 ps, followed by energy minimization again without position restraints, then equilibration for 1 ns, and finally the production run of 10 ns. During the equilibration phase the temperature and pressure was coupled with the Berendsen methods (Berendsen et al. 1984). During the production runs the Paranillo-Rahman (Parinello and Rahman 1981) method was used for pressure coupling and the temperature was coupled using the Nosé-Hoover (Nose 1984) method at 310 K. Electrostatics were calculated with the Particle Mesh Ewald (PME) method (Darden et al. 1993). The LINCS algorithm (Hess et al. 1997) was used to constrain bond lengths and a time step of 2 fs was used throughout. Protein-ligand interactions were visualised with LigPlot (Wallace et al. 1995) and VMD (Humphrey *et al.* 1996). Persistent water molecules were identified via an in-house program. The water density calculations were carried out using Gromacs 3.2.1 and an in house program g_ri3D (Oliver Beckstein) and images were generated with Chimera (Pettersen et al. 2004).

Results and Discussion

Global Motion

Although our focus was on the binding pocket and the behaviour of water molecules it is still useful to briefly examine the global motions in the four simulations. The overall dynamics of the receptor can be assessed by estimating the Mean Square Fluctuations (MSF) of the Cα atoms (Fig. 1C). This approach calculates the average fluctuation in a given window and can be used as an indicator of the level of sampling for a given conformational state (Faraldo-Gomez *et al.* 2004). The simulations without ligands showed more movement in the MSF plots than the simulations with ligands bound, perhaps indicating a degree of structural stability resulting from the presence of the ligand. However, in all simulations, the MSF values continue to increase which indicates that the conformational dynamics of the protein as a whole are far from being fully sampled.

Despite this, principal component analysis (PCA) of the large-scale motions from the simulations with ligands revealed a breathing motion of the receptor as has been recently observed for a model of the α7 nAChR (Henchman *et al.* 2005). The PCA data also revealed greater movement in two of the five subunits in both simulations with ligands bound, even in the case of the simulation where all five subunits were occupied with nicotine. These observations agree with previous MD studies of the α7 nAChR (Henchman et al. 2003) and the suggested gating mechanism of the *Torpedo marmorata* acetylcholine receptor (Unwin 2002; Unwin 2005). This asymmetrical motion is also observed to a smaller extent in the non-liganded (NCT-apo, CCE-apo) simulations

The ligand-binding site.

As it is unclear how agonist binding leads to structural changes and channel gating in the homologous nACh receptor, we examined the dynamic behaviour of the binding pocket in the AChBP. In order to assess this, we calculated an RMSD of a subset of atoms known to form part of the binding pocket (specifically M114:N, W143:O, W143:HE1, T145:O and Y192:HH). The results are shown in Fig. 2, which show that the RMSD for these atoms is higher in the case when the ligand is not present showing that there is some stabilizing effect of these atoms. Fig. 2 also shows that the stabilizing effect appears to be greater for carbamylcholine (Fig. 2B) compared to nicotine (Fig 2A). Thus, simulations where the binding pocket was occupied by a ligand exhibited smaller fluctuations indicating that the ligand brings about a structural rigidity to the binding pocket, suggesting that the AChBP behaves according to the induced fit model (Koshland 1958).

Ligand-Flexibility

We examined the flexibility of the ligands within the binding site. The ligands for both the NCT and CCE simulations maintained their position within the binding pocket for the duration of the simulation. Although both ligands held their place and general orientation in the binding pocket, nicotine was much more rigid than carbamylcholine. Analysis of the dihedral angles showed a fairly rigid conformation for nicotine, consistent with only one binding mode that differs only by a slight tilting of the 6' ring. The dihedral data for carbamylcholine show more flexibility of the chain. The molecule maintains its orientation, but there are rotations at both ends of the ligand suggesting multiple

conformations may be equally favourable within the pocket. Docking studies (*Amiri et al, unpublished*) also suggest that nicotine has one mode of binding while carbamylcholine may have more than one mode of binding.

Water in the binding pocket.

We wanted to look at the role of water in the interactions of the ligand with the residues in the binding pocket. Other studies have alluded to the role of water in the binding pocket of another classes of ligand-gated ion channel; namely the ionotropic glutamate receptors (Arinaminpathy et al. 2006; Kaye et al. 2006), and thus we wanted to further investigate its role in these simulations. Visual inspection of the trajectories revealed that in most subunits water seems to play a significant role in the ligand's interaction with surrounding residues. Indeed, some of these bridging waters are seen throughout the entire simulation. Fig. 3 shows both ligands interacting with surrounding residues via water molecules. In the nicotine bound (NCT) and carbamylcholine (CCE) bound simulations, a water molecule bridging the ligand and L102 or M114 on loop E is present for more than 92% of the time averaged for the subunits with ligands and sometimes the same water bridges both of these residues at the same time. Also commonly seen are water molecules bridging the ligand with W143 or T144 or both at the same time. Bridging waters also exist between the cysteines on Loop C and the ligands. The cysteines on Loop C from α-subunits are thought to be functionally important as they are evolutionarily highly conserved among the LGICs. Recent MD data has shown that Loop C comes in and covers the binding pocket in the presence of a ligand (Henchman et al., 2005) and thus the hydrogen bonding with the ligand could play an important role in

determining the behaviour of Loop C. Other bridging waters commonly exist between the ligand and Q55 and with Y164. Y164 has its sidechains facing upwards towards the ligand in the binding site which enables hydrogen bonding with the ligand. We quantified for what of percentage of time these waters occupy different locations in the binding pocket. Those that occupied a position in the binding cleft for greater than 40% of the time we term "persistent". Greater numbers of persistent waters were found in the simulations with ligand present (NCT and CCE). For the carbamylcholine simulation with only 2 ligands bound in two binding pockets, the greatest number of persistent waters was found in the liganded binding pockets throughout the simulation. For the binding pockets that do not possess a ligand, waters were observed, but were substantially more mobile and stayed for a much shorter amount of time than waters in the binding sites with ligands present.

In order to compare the distribution of water molecules in the binding site with bulk, we calculated the average water density throughout the simulations. Fig. 4A shows the average water density plot for the simulation of one of the binding pockets of the AChBP with nicotine bound. Compared to bulk water, the average density of water molecules in the binding pocket appears to be discretized into distinct zones indicating that these areas are where water molecules are preferentially found over time. These density plots were compared with the positions of the persistant waters. Taking the visual inspection data, the average water density analysis and the persistency data enables us to describe 5 separate zones where water prefers to reside (Fig. 4B). These zones were populated frequently and for substantial stretches of time. Zone 1 lies within loop E between L102 and M114. This location is almost always occupied with at least one water

molecule >90% of the time. Waters in this position have also been observed in some crystallography studies (Celie *et al.*, 2004). Water molecules can be seen 'competing' with each other for this location, and if one water molecule leaves the site, another will usually replace it very quickly, suggesting a strong preference for a water molecule in this area. Zone 2 is the area near the W143/T144 loop (β8 sheet) where there is a bridging water connecting the ligand with either one or both residues in both cases of nicotine and carbamylcholine. Waters are commonly found in this position for subunits with ligands bound. Zone 3 is at the base of loop C near E193 and Y192. Y192 is heavily involved in hydrogen-bonding to the ligand. Waters in this zone may play a role in keeping the integrity of loop C and restrict it from moving out freely in the presence of a ligand. Zone 4 describes the low β5 sheet (loop A) of the binding pocket with residues Y89 and I92 that have been thought to be involved in ligand binding from previous experimental studies (Celie, *et. al*, 2005). In several nanoseconds of the liganded simulations, this location is occupied with one or more water molecules, however this is seldom the case in apo simulations. Zone 5 is near the tip of loop C where there are two adjacent cysteines forming a disulfide bridge. These cysteines have been the focus of many studies and are highly conserved in the α-subtype of nAChRs. It is believed that the cysteines interact with the ligands in the binding pocket of nAChR and can be seen to 'close off' the binding pocket in the presence of a ligand. There are several bridging waters in position between the cysteines and the ligands, and this interaction is often mediated by water molecules.

We examined the consequence on the structure if a water molecule in one of these zones moves out. Fig. 5A shows the typical behaviour of waters into and out of a zone

(Zone 1). Here we follow three water molecules for the duration of the simulation. Two waters are positioned in Zone 1 between residues L102 and M114 on loop E. A third water comes in and knocks one of the two waters out, and replaces it in the same exact position where it remains for most of the simulation. Upon leaving (just before the 9 ns mark), the β-sheets of loop E collapse closer into each other and the nature of interactions with the ligand is altered (Fig 5B).

Recently, several high resolution X-ray crystal structures of AChBP in complex with agonists and antagonists has become available (Bourne *et al.* 2005; Hansen *et al.* 2005; Ulens *et al.* 2006) . Comparative analysis of these structures demonstrated that loop C clusters into three distinct positions depending on the nature of the compound bound: i) agonists bound (as exemplified by epibatidine or lobeline); ii) the apo or alkaloid antagonist-bound structures (as exemplified by methyllycaconitine); or iii) structures with peptide antagonists bound (ImI complex). We were interested to see how the positions of these compounds related to the positions of waters in the zones described above. We overlayed the structures of these compounds onto the water-density obtained from the NCT simulation (Fig. 6) to investigate this.

For the two agonists, epibatidine (a compound produced by the Ecuadorian frog, *Epipedobates tricolor)* and lobeline (an alkaloid from the Indian tobacco plant, *Lobelia inflate*), the position in the binding pocket with respect to the water zones is quite similar. The crystal structure of epibatidine bound to AChBP (PDB:2BYQ) shows that the nitrogen in the azobicycloheptane ring is close to Zone 1 (Fig. 6A). This toxin resembles the structure of nicotine and is bound in the same way inside the binding pocket with its 6' ring facing the high water density regions in Zone 1. Thus we expect to see the same

interactions with water molecules and surrounding residues. The other nitrogen atom lies near water densities in Zone 2, behind loop B. Lobeline is positioned in the pocket such that is quaternary nitrogen is in a similar position to epibatidine and nicotine. There is also potential interaction with water molecules in Zone 1 (Fig. 6B). This is confirmed by the presence of a water molecule in the crystal structure. Thus the positions of these zones, particularly Zone 1 appears to most relevant for the agonist-bound conformation of the protein.

For methyllycaconitine, which is an alkaloid antagonist against nicotinic acetylcholine receptors, there is no indication that the zones derived here play a role in stabilizing the complex (Fig. 6C). In the case of the α-conotoxin ImI (PDB: 2BYP), the position of the toxin suggests that Zone 1 and 3 could play a role in stabilizing the complex (Fig. 6D). However, it is important to remember, for this structure, that the loop C is pulled out radially with respect to the central axis and with respect to the agonist-bound conformations from which the water zones were derived from. A direct comparison of the zones derived here with the antagonist-bound structures should be treated with caution given the differences in the shape of the binding pocket, but we would suggest from this analysis that these zones become less important with respect to antagonist binding.

Conclusions

Our results indicate how the dynamics of the ligand in conjunction with the presence of explicit water molecules are an important consideration for our understanding of AChBP and the related Cys-loop receptors. The ligands are held fairly rigidly in the binding

pockets. Furthermore, the presence of the ligands confers some stability on the binding site residues. The effect is aided by the presence of water molecules which form bridging hydrogen bonds with key residues. In the apo simulations, the binding site is flexible (higher RMSD values) with varying distances between neighbouring residues and more flexible loops. Loop C which is less mobile in liganded simulations, makes hydrogen bonds with ligands through water molecules. Loops A and B are held more rigid in the presence of water molecules which are more persistent in the presence of a ligand. These bonding patterns have structural implications for the binding site which are likely to be important in the homologous nAChR proteins. The location of zones in the binding sites where waters persist with nicotine and carbamylcholine bound suggest an extended pharmacophore that could be used in the design of new ligands. Furthermore, such a pharmacophore will be energetically stable within the binding pocket. This pharmacophore can be interpreted in two ways. The first is that the zones could offer a site where the water molecule is replaced by a chemical moiety. The second is where the zone is occupied by a water molecule and the ligand makes interactions with the protein via such a water (as exemplified by Zone 1, which appears to be aid in the binding of nicotine, lobeline and epibatidine). Finally, these observations should be readily transferable to nicotinic acetylcholine receptors in general and aid in new compounds targetted against them.

Acknowledgements

We thank the Wellcome Trust and Oxford Supercomputer Centre for supporting this work.

26/07/2006

26/07/2006

Tables

Table 1. Summary of simulations.

Simulation name	PDB code	Ligand	Number of Ligands	Cα RMSD (Å)
NCT	1UW6	nicotine	5	2.683
NCT-Apo	1UW6	-	0	2.661
CCE	1UV6	carbamylcholine	2 (at the interface of subunits C/D and D/E)	2.478
CCE-Apo	1UV6	-	0	2.745

* All simulations were run for 10 ns.

Table 2. Percentage time that any one zone is occupied by at least one water molecule for the simulation with nicotine (NCT) and carbamylcholine (CCE). For NCT, the figures are averaged across all 5 subunits, and for CCE the figures are averaged across 2 subunits where ligands were bound.

ZONE	Average for NCT (5 subunits)	Average for CCE (2 subunits)
1	92	92.5
2	45	79.5
3	60	76
4	40	89.5
5	55	50

26/07/2006

Figures.

Figure 1. (A) Two subunits of AChBP (cartoon) shown with nicotine in the binding pocket (spacefill representation). (B) The ligands used in the MD simulations, nicotine and carbamylcholine, are shown. (C) Mean Square Fluctuation (MSF) for the Cα from the four simulations using the last 7.5 ns. Dark solid line shows the MD simulation of AChBP with nicotine bound (NCT), the dark dashed line shows AChBP without nicotine bound (NCT-Apo). The lighter solid line shows AChBP with carbamylcholine bound (CCE) and the corresponding dashed line is the same crystal structure with the carbamylcholine taken out (CCE-Apo). Error bars reflect standard deviations (shown only on the CCE data for clarity).

Figure 2. RMSD of atoms that define the binding pocket (M114:N, W143:O, W143:HE1, T144:O and Y192:HH). (A) shows AChBP with and without nicotine (NCT = black line, NCT-Apo = grey line). (B) shows AChBP with and without carbamylcholine (CCE = black line, CCE-Apo = grey line).

Figure 3. VMD diagrams showing water-mediated ligand-protein interactions. **(A)** A snapshot at 5 ns of nicotine (at the interface of subunits B and C) is interacting with W143 and T144 through the same bridging water while L102, M114, and L112 are also interacting with nicotine through a bridging water molecule. Averaged across all 5 subunits, this water is present for 92% of the duration of the simulation and the former is present for 45% of the duration. **(B)** Carbamylcholine at 6 ns is interacting with surrounding residues through 2 water molecules. One is bridging W143 to

carbamylcholine and the other is bridging L102 and M114 with carbamylcholine. Averaged across the 2 subunits with carbamylcholine bound, the water molecule bridging W143 is present for 79.5% of the simulation and the other is sustained for 92.5% of the time.

Figure 4. (A) Two subunits are shown with the densities of water molecules during NCT simulation. (B) A snapshot of the final frame of the NCT simulation, showing the positions of zones as defined by long-lived water molecules in the binding pocket.

Figure 5. (A) The graph on the top shows two water molecules coming inside loop E, situated between L102 and M114 and then eventually leaving. A third water molecule comes in and replaces one of two already there. One of the two waters leaves without being replaced. (B) Once this water leaves, the distance between L102 and M114 is decreased, affecting the nature of interactions with the ligand.

Figure 6. (A) Overlay of the AChBP complex with epibatidine with the density maps (derived from the NCT simulation) for water illustrating that Zone 1 in particular may be an important consideration. Note that this binding mode is similar to that for nicotine. (B) Overlay of the AChBP in complex with lobeline with the density maps for water again showing that waters in Zone 1 can make interactions with this toxin (as confirmed by the presence of water molecule in the crystal structure). (C) Overlay of the AChBP in complex with methyllycaconitine with the density maps for water. In this case there appears to be little evidence for water in these zones to make significant interactions or overlap with moieties of the antagonist.

(D) Overlay of the AChBP in complex with α-conotoxin ImI with the density maps for water. Although there is some indication that Zones 1 and 3 may be preserved it must be remembered that for this complex there is quite a large deformation (particularly with respect to loop C) of the binding pocket.

References

Arinaminpathy Y., Sansom M.S.P. and Biggin P.C. (2006) *Mol. Pharm.,* **69,** 11-18.
Ashcroft F.M. (2000). Ion Channels and Disease. San Diego: Academic Press. 481 p.
Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A. and Haak J.R. (1984) *J. Chem. Phys.,* **81,** 3684-3690.
Berendsen H.J.C., van der Spoel D. and van Drunen R. (1995) *Comp. Phys. Comm.,* **95,** 43-56.
Bourne Y., Talley T.T., Hansen S.B., Taylor P. and Marchot P. (2005) *EMBO J.,* **24,** 1512-1522.
Brejc K., van Dijk W.J., Klassen R.V., Schuurmans M., van Der Oost J., Smit A.B. and Sixma T.K. (2001) *Nature,* **411,** 269-276.
Brejc K., van Dijk W.J., Smit A.B. and Sixma T.K. (2002) *Novartis Found Symp.,* **245,** 22-29.
Celie P.H., Kasheveroc I.E., Mordvintsev D.Y., Hogg R.C., van Nierop P., van Elk R., van Rossum-Fikkert S.H., Zhmak M.N., Bertrand D., Tsetlin V. et al. (2005a) *Nat. Struct. Mol. Biol.,* **12,** 582-588.
Celie P.H., Klassen R.V., van Rossum-Fikkert S.H., van Elk R., van Nierop P., Smit A.B. and Sixma T.K. (2005b) *J. Biol. Chem,* **280,** 26457-26466.
Celie P.H., van Rossum-Fikkert S.H., van Dijk W.J., Brejc K., Smit A.B. and Sixma T.K. (2004) *Neuron,* **41,** 907-914.
Cymes G.D., Grosman C. and Auerbach A. (2002) *Biochem.,* **41,** 5548-5555.
Darden T., York D. and Pedersen L. (1993) *J. Chem. Phys.,* **98,** 10089-10092.
De Fusco M., Becchetti A., Patrignani A., Annesi G., Gambardella A., Quattrone A., Ballabio A., Wanke E. and Casari G. (2000) *Nat. Genet.,* **26,** 275-276.
Faraldo-Gomez J.D., Forrest L.R., Baaden M., Bond P.J., Domene C., Patargias G., Cuthbertson J. and Sansom M.S.P. (2004) *Proteins: Struct. Funct. Genet.,* **57,** 783-791.
Gao F., Bren N., Burghardt T.P., Hansen S.B., Henchman R.H., Taylor P., McCammon J.A. and Sine S.M. (2005) *J. Biol. Chem,* **280,** 8443-8451.
Graton J., Berthelot M., Gal J., Girard S., Laurence C., Lebreton J., Le Questel J., Maria P. and Naus P. (2002) *J. Am. Chem. Soc.,* **124,** 10552-10562.
Grosman C. (2003) *Biochemistry,* **41,** 5548-5555.
Hansen S.B., Sulzenbacher G., Huxford T., Marchot P., Taylor P. and Bourne Y. (2005) *EMBO J.,* 1-12.

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

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

Hermans J., Berendsen H.J.C., van Gunsteren W.F. and Postma J.P.M. (1984) *Biopolymers,* **23,** 1513-1518.

Hess B., Bekker J., Berendsen H.J.C. and Fraaije J.G.E.M. (1997) *J. Comp. Chem.,* **18,** 1463-1472.

Hogg R.C. and Bertrand D. (2004) *Curr. Drug Targets CNS Neurol. Disord.,* **3,** 123-130.

Humphrey W., Dalke A. and Schulten K. (1996) *J. Molec. Graph.,* **14,** 33-38.

Karlin A. (2002) *Nat Rev Neurosci,* **102,** 102-114.

Kaye L.S., Sansom M.S.P. and Biggin P.C. (2006) *J. Biol. Chem,* **281,** 12736-12742.

Koshland D.E.J. (1958) *Proc Natl Acad Sci USA,* **44,** 98-104.

Law R.J., Henchman R.H. and McCammon J.A. (2005) *Proc Natl Acad Sci USA,* **102,** 6813-6818.

Lester H.A., Dibas M.I., Dahan D.S., Leite J.F. and Dougherty D.A. (2004) *Trends Neurosci.,* **27,** 329-336.

Lindahl E., Hess B. and van der Spoel D. (2001) *J. Mol. Model,* **7,** 306-317.

Nose S. (1984) *Mol. Phys.,* **52,** 255-268.

Pankow J.F., Barsanti K.C. and Peyton D.H. (2003) *Chem. Res. Toxic.,* **16,** 23-27.

Parinello M. and Rahman A. (1981) *J. Appl. Phys.,* **52,** 7182-7190.

Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C. and Ferrin T.E. (2004) *J. Comp. Chem.,* **25,** 1605-1612.

Schuettelkopf A.W. and van Aalten D.M.F. (2004) *Acta Cryst. D,* **60,** 1355-1363.

Sixma T.K. and Smit A.B. (2003) *Annu. Rev. Biophys. Biomol. Struct.,* **32,** 311-334.

Smit A.B., Brejc K., Syed N.I. and Sixma T.K. (2003) *Ann. N.Y. Acad. Sci.,* **998,** 81-92.

Smit A.B., Syed N.I., Schaap D., van Minnen J., Klumperman J., Kits K.S., Lodder H., van der Schors R.C.v., van Elk R., Sorgedrager B. et al. (2001) *Nature,* **411,** 261-268.

Steinlein O.K. (2001) *Am J. Med. Genet.,* **106,** 139-145.

Ulens C., Hogg R.C., Celie P.H., Bertrand D., Tsetlin V., Smit A.B. and Sixma T.K. (2006) *Proc Natl Acad Sci USA,* **103,** 3615-3620.

Unwin N. (2002) *Novartis Found Symp.,* **245,** 5-15.

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

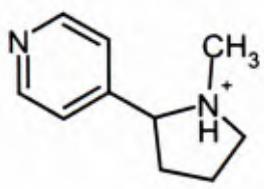
van Gunsteren W.F., Krüger P., Billeter S.R., Mark A.E., Eising A.A., Scott W.R.P., Hüneberger P.H. and Tironi I.G. (1996). Biomolecular simulation: The GROMOS96 manual and user guide. Groningen/Zürich: Biomos/Hochschulverlag AG an der ETH Zürich.

Wallace A.C., Laskowski R.A. and Thornton J.M. (1995) *Prot. Eng.,* **8,** 127-134.

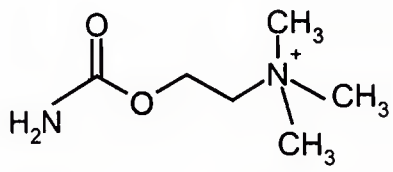
Figure 1

A

B



nicotine



carbamylcholine

C

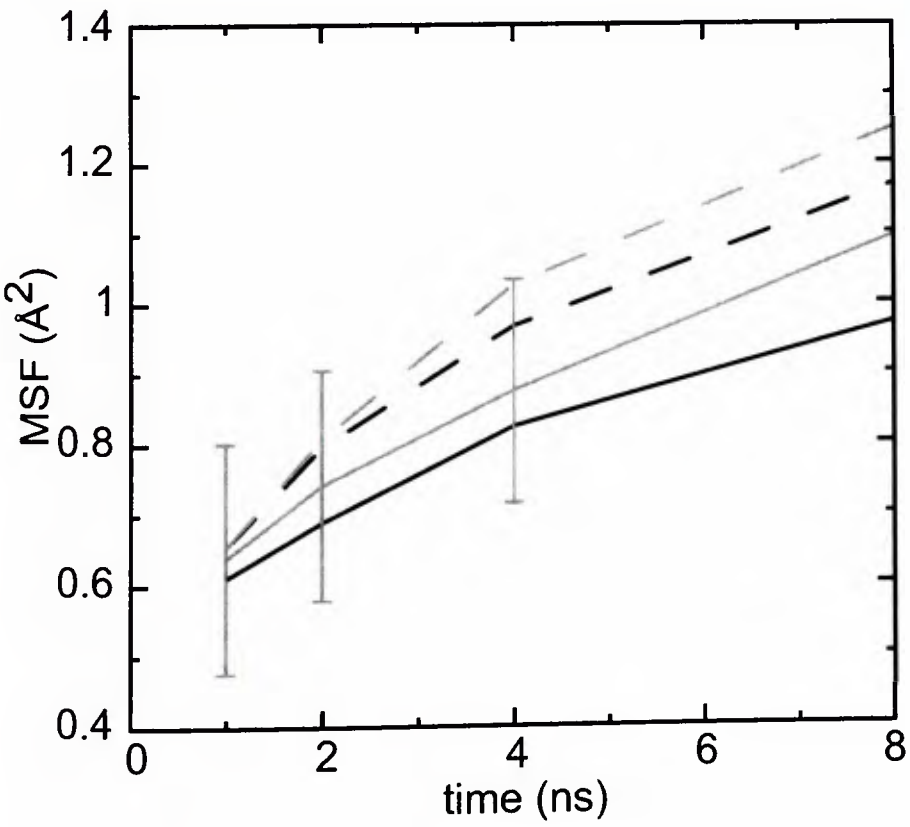


Figure 2

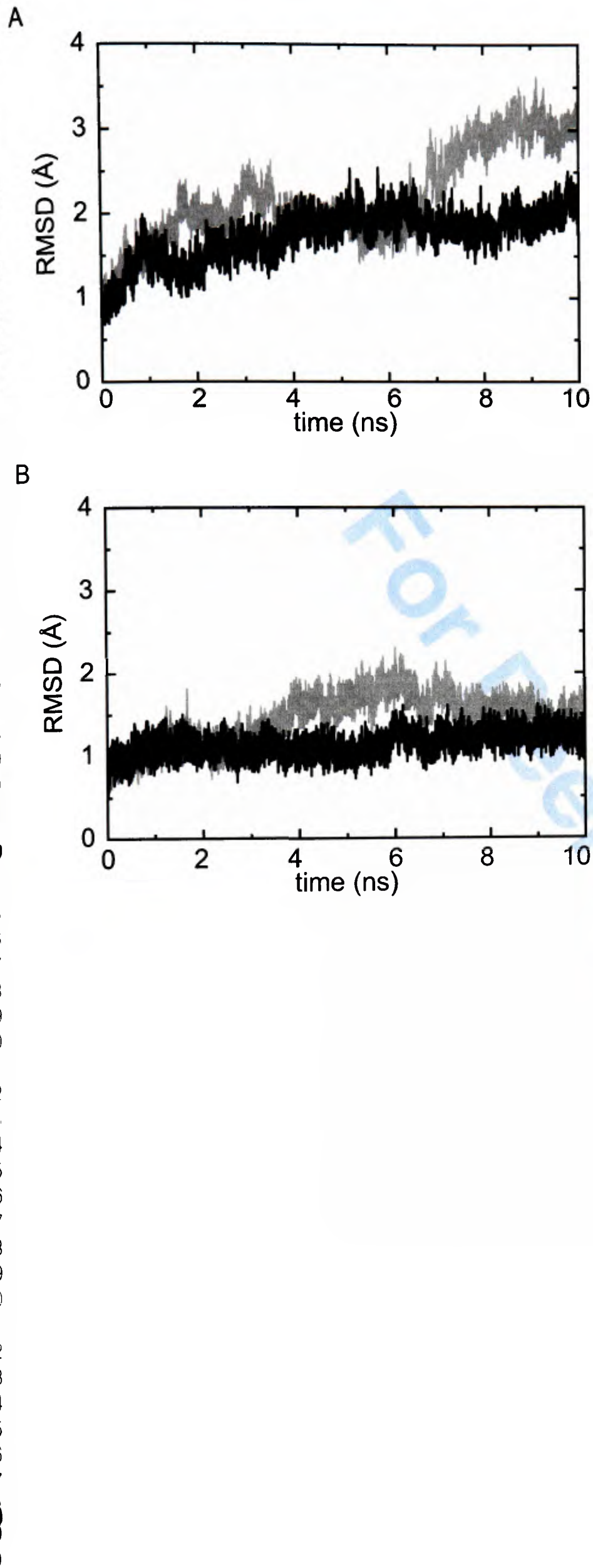


Figure 3

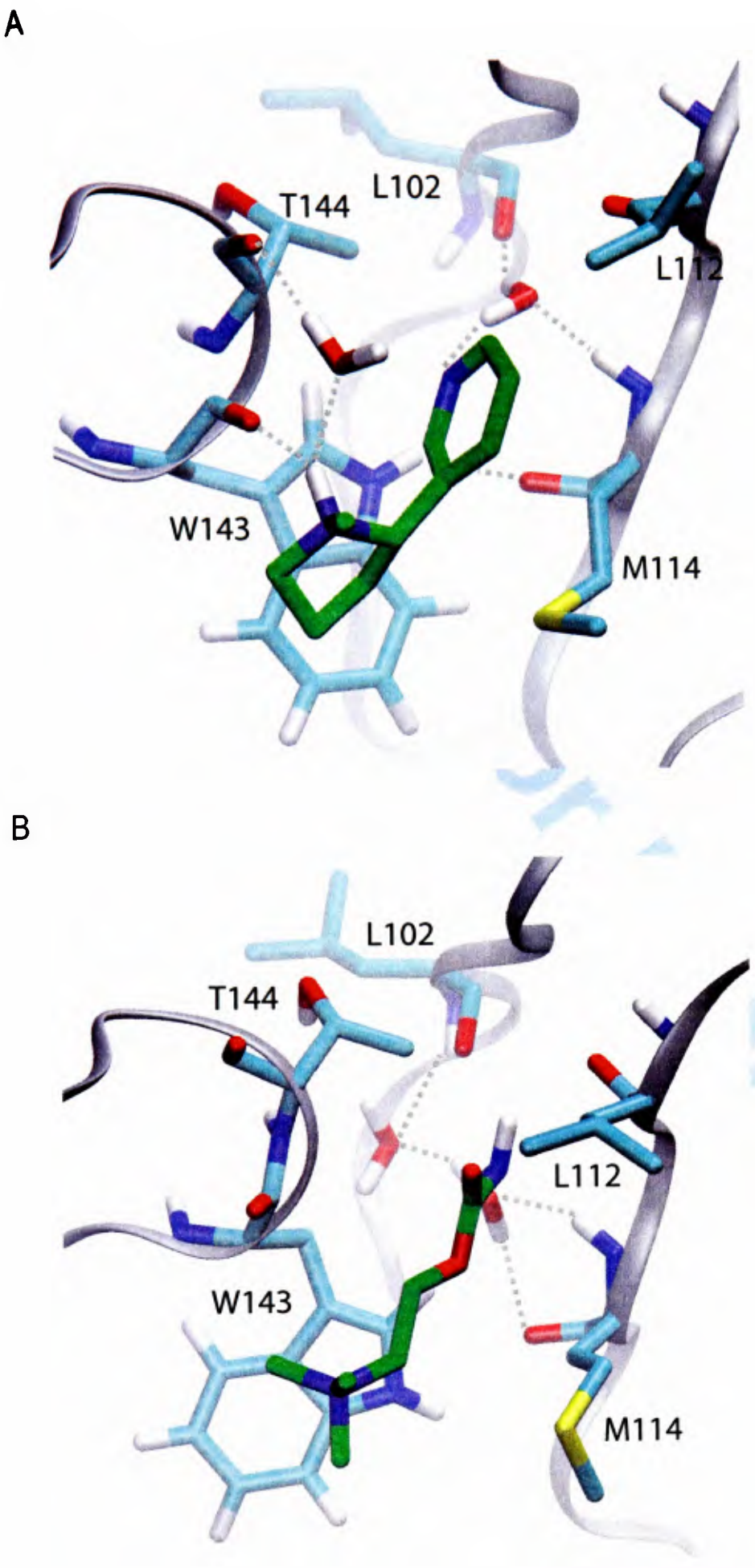


Figure 4

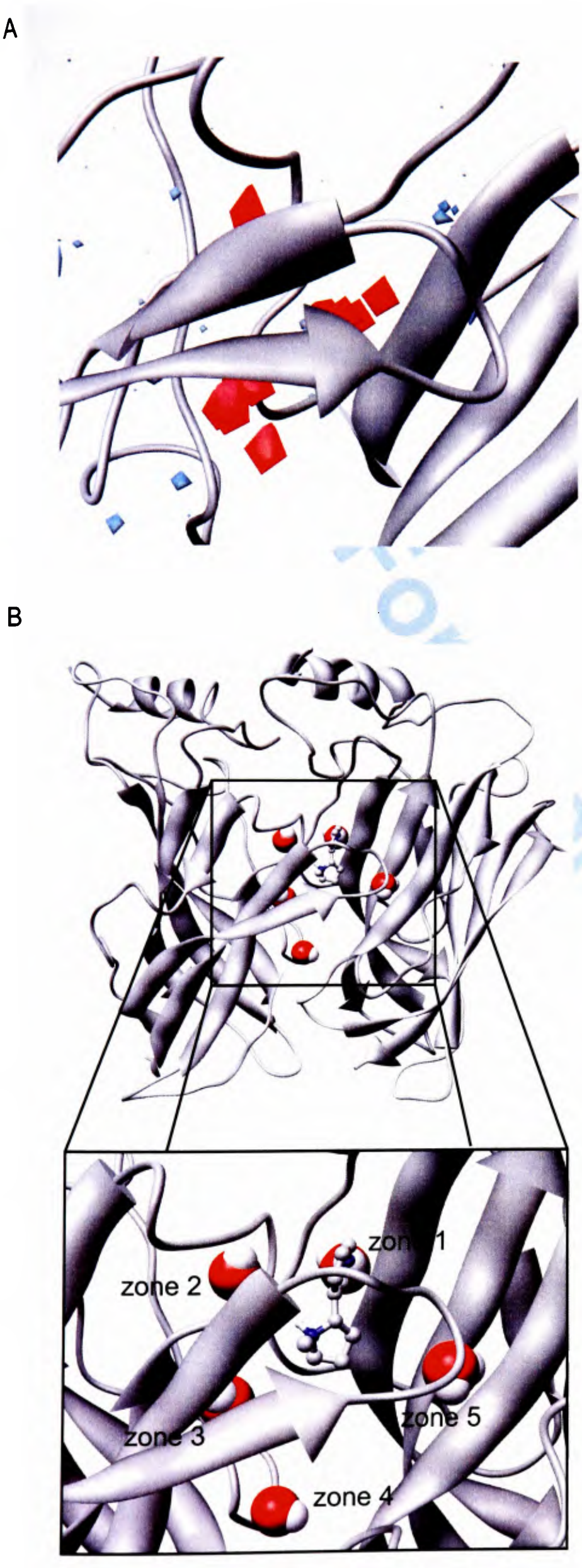


Figure 5

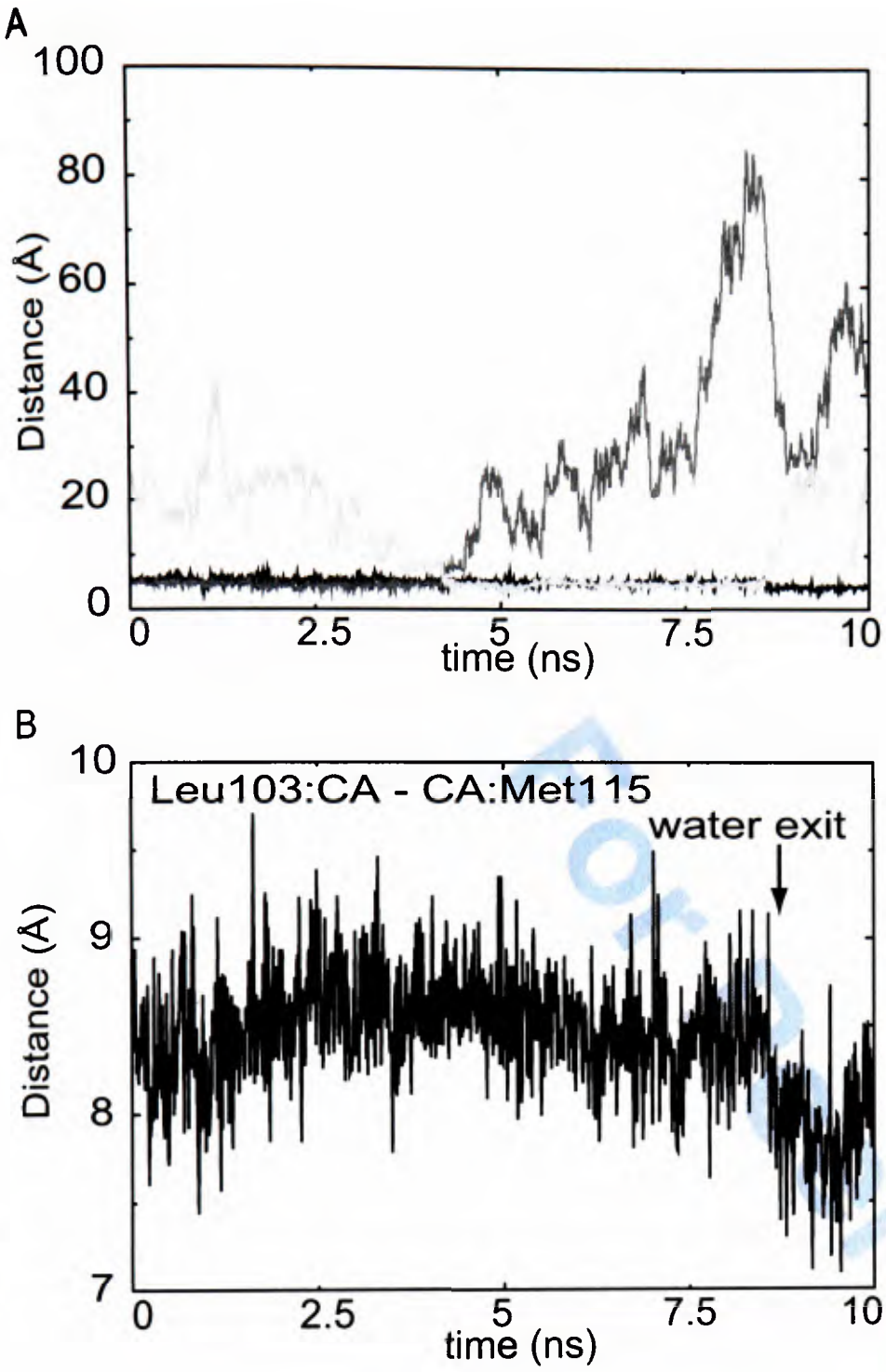
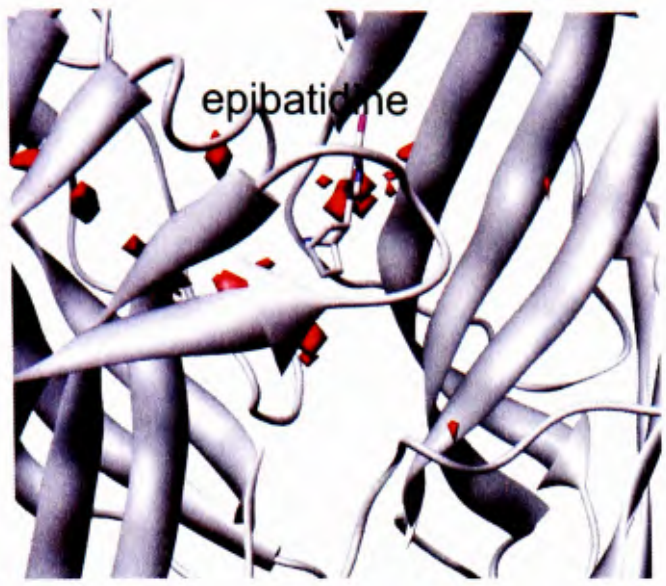
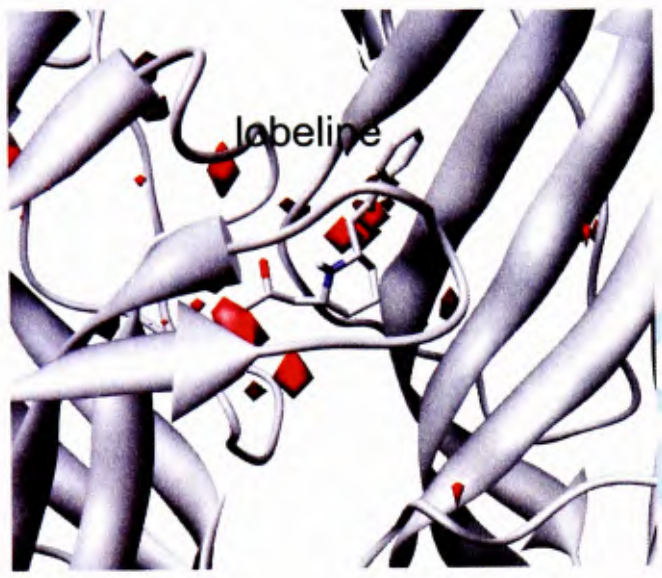


Figure 6

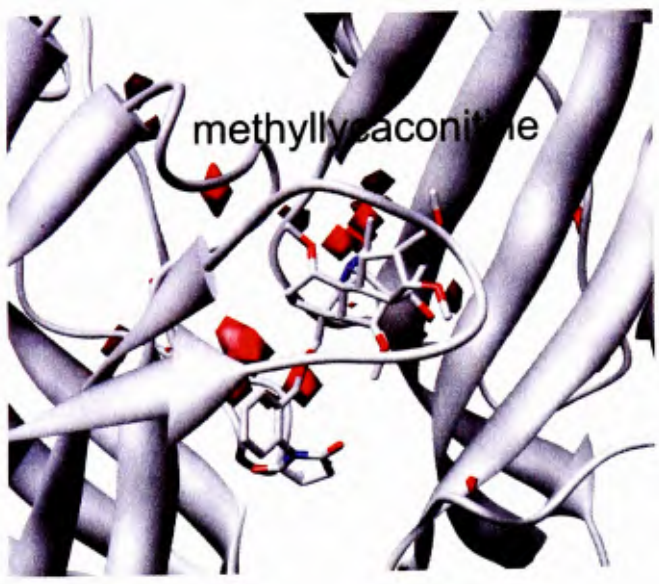
A



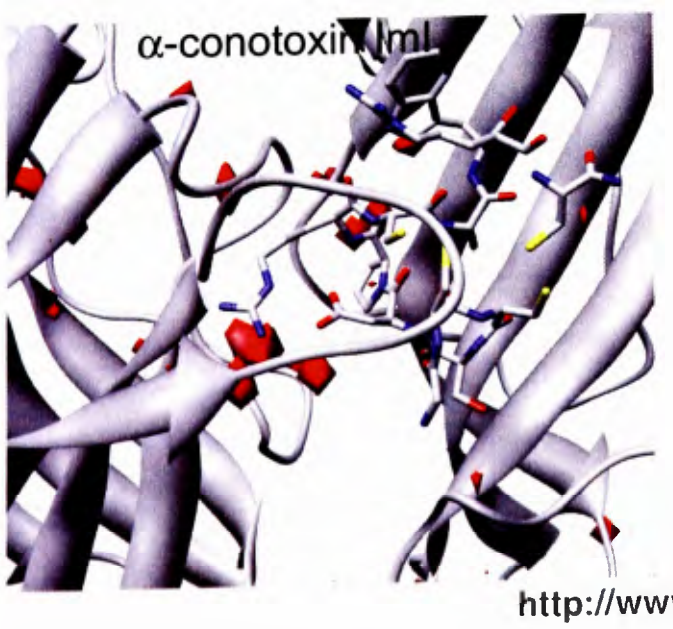
B



C



D



A Role for Leu118 of Loop E in Agonist Binding to the α7 Nicotinic Acetylcholine Receptor

Shiva Amiri[*1], Masaru Shimomura[2*], Miki Akamatsu[2], Kazuhiko Matsuda[2], Andrew K. Jones[3], Mark S.P. Sansom[1], Philip C. Biggin[1] and David B. Sattelle[3].

[1]Structural Bioinformatics and Computational Biochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU (SA, MSP and PCB)

[2]Department of Applied Biological Chemistry, School of Agriculture, Kinki University (MS and KM)
Graduate School of Agriculture, Kyoto University, Kita-shirakawa, Sakyo-Ku, Kyoto 606-8502, Japan (MA)

[3]Department of Physiology, Anatomy and Genetics, Le Gros Clark Building, University of Oxford, South Parks Road, Oxford OX1 3QX, UK (AKJ and DBS)

[*]These authors have contributed equally to this work.

Running title: Role for Leu118 of loop E in nAChR-agonist interactions

To whom correspondence should be addressed:
David B. Sattelle
Tel: +44 (0) 1865 - 272 145
Fax: +44 (0)1865 - 272 421
Email: david.sattelle@anat.ox.ac.uk
For submission to Molecular Pharmacology, 21.06.2006

No. of text pages:
No. of tables: 2
No. of figures: **8**
No. of words:

Key words: ligand-gated ion channel; agonist docking; nicotinic acetylcholine receptor; computational modeling; neonicotinoid insecticides (imidacloprid, desnitro-imidacloprid).

Abbreviations: ACh, acetylcholine; nAChR, nicotinic acetylcholine receptor; EC_{50}, half maximum concentration; I_{max}, maximum normalized response; I_{min}, minimum normalized response, n_H, Hill coefficient Abbreviations: AChBP, acetylcholine binding protein; DN-IMI, desnitro-imidacloprid; IMI, imidacloprid.

Abstract.
Nicotinic acetylcholine receptors (nAChRs) are ligand-gated ion channels mediating fast cholinergic synaptic transmission in the brain and at neuromuscular junctions. We used the structure of the acetylcholine binding protein (AChBP) from *Lymnaea stagnalis,* to model the chicken α7 agonist-binding domain. The initial models and a preliminary docking study suggested that position L118 may play an important role in determining agonist actions on α7. A prediction from these *in silico* studies, that L118E and L118D would retain binding to acetylcholine, but L118K and L118R would not, was confirmed in electrophysiological studies on functional recombinant mutant receptors expressed in *Xenopus laevis* oocytes. The functional studies also demonstrated that residues at position 118 can have a dramatic effect on the actions of imidacloprid (a partial agonist of wild-type α7 receptors) and its des-nitro derivative. Molecular Dynamics simulations indicate that L118 can strongly influence the behaviour of the binding pocket and that the model is robust in terms of its prediction for acetylcholine binding. Together, the results indicate that L118 is influential on the binding properties and selectivity of nAChR agonists.

Introduction

Nicotinic acetylcholine receptors (nAChRs) are plasma membrane cation channels activated by the neurotransmitter acetylcholine (ACh). They mediate fast cholinergic synaptic transmission at neuromuscular junctions and in the brain (Karlin, 2002; Lena and Changeux, 1998). The nAChR molecules are pentamers composed of identical or highly homologous subunits, each with four transmembrane regions (TM1-4) and an extracellular N-terminal domain containing 6 loops (A-F) that make up the ACh binding site (Corringer et al., 2000b) (Corringer et al., 2000a). Subunits possessing two adjacent cysteines in loop C of the ACh binding site are designated α subunits, whereas those lacking this cysteine pair are denoted non-α or β subunits.

Neuronal nicotinic receptor subunits (α2-10; β2-4), which contribute to a variety of receptor subtypes depending on subunit composition (Millar, 2003a; Millar, 2003b), are important targets for new drugs. These include drugs being developed as analgesics and drugs for ameliorating the symptoms of Alzheimer's disease (Papke et al., 2000; Prendergast et al., 1998), as well as drugs used to treat Parkinson's disease and schizophrenia (Lloyd and Williams, 2000). Congenital myasthenias (Ohno and Engel, 2002) result from mutations in muscle nAChR subunits and autosomal nocturnal frontal lobe epilepsy results from mutations in neuronal α4 and β2 nAChR subunits (Changeux and Edelstein, 2001; Steinlein, 2001). Also, autoantibodies directed against nAChRs underlie several diseases such as myasthenia gravis (muscle nAChRs) (Lang and Vincent, 2003a; Lang and Vincent, 2003b), Rasmussen's encephalitis (α7) (Watson et al., 2005a) (Watson et al., 2005b) and Autonomic Neuropathy (α3) (Vernino and Lennon, 2003a; Vernino and Lennon, 2003b). The nAChRs of insects are the targets for imidacloprid, a neonicotinoid insecticide extensively used worldwide (Matsuda et al., 2001b). In addition, nematode nAChRs are important targets for anthelmintic drugs such as levamisole, pyrantel and morantel (Harrow and Gration, 1985; Jones et al., 2005).

The most detailed available structure of a native nAChR has been determined by electron microscopy of tubular crystals of *Torpedo marmorata* postsynaptic membranes embedded in amorphous ice (Miyazawa *et al.*, 1999; Miyazawa *et al.*, 2003) (Unwin, 2005). Models based on these co-ordinates indicate that the extracellular region, containing the ligand-binding site, consists of twisted β-sheets, with loop regions forming the binding sites. Sixma and colleagues first crystallized a glial-derived acetylcholine binding protein (AChBP) from the freshwater snail, *Lymnaea stagnalis* (Brejc et al., 2001). This 210 amino acid polypeptide forms a stable homo-pentamer with homology to the N-terminal extracellular region of nAChRs. ACh binding sites are located at each of the 5 subunit interfaces and the 6 binding loops (A-F) that can be recognized in nAChRs are also present. Several AChBP crystal structures have now been published (Brejc et al., 2001; Brejc et al., 2002; Celie et al., 2005a; Celie et al., 2005b; Celie et al., 2004b; Hansen et al., 2005). In view of the sequence homology between the AChBP and the ligand binding domain of neuronal α7 (24% identity), homology modeling might provide a useful tool to investigate ligand binding. The structure of the chicken α7 homopentamer has been constructed based on both the X-ray structure of the L.stagnalis AChBP and the electron microscopy derived structure of the transmembrane region of the Torpedo nicotinic receptor (Amiri et al., 2005). Le Novère and colleagues have built a three-dimensional model of the N-terminal domain of a homopentameric chicken α7 nAChR based on AChBP, which was then used to analyze the docking of ACh, epibatidine and nicotine (Le Novère *et al.*, 2002). Plausible modes of binding were then suggested for these ligands. A different modeling study

has also been reported (Schapira et al., 2002) in which the authors were able to relate the pattern of binding affinities to different receptor isotypes.

In the present study we have followed up on an earlier work which showed that a leucine residue in loop E of vertebrate muscle non-α subunits is involved in ligand binding (Sine, 1997). We have generated homology models of chicken α7 homopentameric nAChR, $(\alpha7)_5$ and show how *in silico* methods combined with site-directed mutagenesis yields further evidence for an important role of the equivalent residue in α7, Leu118, in controlling ligand access to the agonist-binding site. We also show that this residue is a major determinant for the behaviour of the insecticide imidacloprid and its derivative desnitro-imidacloprid.

Materials and Methods.

Modeling

A pairwise sequence alignment for chicken α7 and AChBP (PDB code 1UX2, which has HEPES buffer bound) was generated on the basis of multiple sequence alignments extracted from the pfam database (Bateman et al., 2002) and improved by manual adjustment. The alignment (Figure 1A) was used as input for the program MODELLER (Sali and Blundell, 1993) to generate 100 initial pentameric models of the chick α7 ligand binding domain. The quality of the 5 lowest energy structures was checked with the PROCHECKv3.5.4 program (Laskowski et al., 1993) and the WHAT-IF (http://www.cmbi.kun.nl/gv/servers/WIWWWI) server. The model with the lowest percentage of residues in the disallowed region of the Ramachandran plot was selected for docking studies. Only one binding site from 2 adjacent subunits was used (see Figure 1B). *In silico* site-directed mutagenesis was performed using PyMOL (DeLano, 2004). We then used these structures as potential docking targets for ACh and imidacloprid with the program, Autodock (Morris et al., 1998). Autodock was run with default parameters except that atom types were not set for the ligand but rather according to charges calculated as below. The active site was defined as a radius of (default) 15.0 Å from carbon 6 (IUPAC/IUBMB numbering) of the sidechain of Trp148, a residue located centrally in the ACh binding pocket. The charges on ACh were assigned according to according to a previous report ((Segall et al., 1998). Charges for imidacloprid were calculated with the 6-31G* basis set using Spartan (Wavefunction Inc). All docking results were visualized with VMD (Humphrey et al., 1996) and UCSF Chimera (Pettersen et al., 2004).

Molecular dynamics simulations were carried out with GROMACS v 3.1.4 (Berendsen *et al.*, 1995; Lindahl *et al*, 2001) using the GROMOS96 (van Gunsteren et al., 1996) force-field. Each system was energy-minimized until convergence using a steepest descents algorithm. Molecular dynamics with position restraints for 100 ps was performed followed by equilibration for 1 ns and finally the production run of 10 ns. During the equilibration phase the temperature and pressure were coupled using the Berendsen methods (Berendsen et al., 1984). During the production runs the Paranillo-Rahman (Parinello and Rahman, 1981) method was used for pressure coupling and the temperature was coupled using the Nosé-Hoover (Nose, 1984) method at 310 K. Electrostatics were calculated with the Particle Mesh Ewald (PME) method (Darden et al., 1993). The LINCS algorithm (Hess et al., 1997) was used to constrain bond lengths and a time step of 2 fs was used throughout. All calculations were performed on a PIII intel machine running linux kernel 2.4.19smp.

Electrophysiology on recombinant wild type and mutant α7 receptors

Xenopus laevis oocytes were prepared and injected with cDNA (*Gallus gallus* α7 cDNA in pMT3) as previously described (Shimomura et al., 2003). Oocytes were challenged with compounds at intervals of 3-5 min to minimize the effects of desensitization. Only oocytes which gave stable responses to two or more successive applications of 200 or 400 μM ACh were used. Concentration-response data were obtained by challenging oocytes with increasing concentrations of agonist. The maximum amplitude of the current recorded in response to each challenge was normalized to the response to 1 mM ACh. As the concentration-response curves for ACh and imidacloprid were changed by the mutations L118D and L118E, data from mutants were normalized using the current response to 3 mM ACh while data from L118K and L118R mutants were normalized using the response to 3 mM imidacloprid. Using Graph Pad 'Prism' (Graphpad Software, UK), normalized data were fitted as previously described (Shimomura et al., 2003).

Results

Wild-type α7 model.

We generated three-dimensional models of wild-type α7 nAChR. We found that L118, which is situated in loop E (Corringer et al., 2000b), is close to the ligand binding site (Figure 1B) and consequently may influence ligand-protein interactions in this region. We performed docking of ACh, imidacloprid (IMI) and desnitro-imidacloprid (DN-IMI) to the wild-type model. We found that the ACh solutions were tightly clustered (only 4 clusters) in the binding pocket (Figure 2A) and spanned an energy range of 0.3 kcals/mol, but for IMI there were several binding modes in several clusters (Figure 2B), not all of which were in the immediate binding pocket (7 out of 13 clusters were within an energy cut-off of 1.5 kcals/mol of the lowest energy dock). DN-IMI docking also found the binding pocket but several binding modes were again possible with 6 out of 9 clusters within an energy cut-off of 1.5 kcals/mol of the lowest energy dock (Figure 2C). The energies of the IMI and DN-IMI docks were comparable, but all approximately 3 kcals/mol less than the ACh docks. A comparison of the ACh docking with the crystal structure of carbamylcholine with AChBP (Figure 3A) revealed that the mode of binding is very similar and thus gave us confidence that the procedure could be used further to predict the interaction of ACh with receptor mutants. However, as the solutions found for IMI and DN-IMI were numerous and there is currently no structural information to confirm the docking solutions, we did not have confidence that the procedure could make reliable predictions for these compounds. Furthermore, there is evidence that the receptor can undergo substantial movement of residues in the binding pocket (Henchman et al., 2003) and given the size of IMI (and DN-IMI) we reasoned that the pocket may have to undergo substantial movement to accommodate these molecules.

We therefore examined ACh docking to mutant receptors and consequently performed the following *in silico* mutagenesis on this position; L118E, L118D, L118K and L118R. The results for ACh docking are shown in Figure 3 and for both the L118D mutation (Figure 3B) and the L118E mutation (Figure 3C) the mode of docking resembles that seen for the wild-type receptor (Figure 3A). There is a slight shift in the position of the ACh towards what would be the surface of the membrane in the full length of the receptor. This presumably stems from the increased negative charge at L118 which pulls the quaternary nitrogen moiety downwards compared to wild-type. When we tried to dock ACh to the L118K and L118R mutants, we found that ACh would generally not dock into the binding pocket. In the case of L118K, there

were 3 clusters that positioned ACh near the binding pocket but in completely the wrong orientation. The effect was even more marked for the L118R mutation where not one ACh docking was even in the binding pocket. These studies suggested that negatively charged mutations at this position would retain ACh binding, but positively charged mutations would abolish ACh binding.

Molecular Dynamics and Multiple Docking

Before embarking on experiments to confirm these predictions we decided to see how robust the homology models were to local fluctuations. We were also interested in exploring the consequence of mutations on the stability of the structure. 10 ns Molecular Dynamics (MD) simulations for the wildtype and mutant α7 subunits were therefore performed. A common method for analyzing the relative stability of protein simulations is to analyze the root mean square fluctuations (RMSF) of the Cα atoms. Analysis of the RMSF for these proteins revealed that the mutant structures showed a higher mean RMSF (Figure 4A-D), indicating that this residue has a large influence on the dynamics of the protein. In particular the regions corresponding to loop E (residues 112-119) appear to exhibit larger fluctuations compared to wild-type. In the case of L118D and L118E there are also increased fluctuations for loop C.

As the result of an MD simulation is to produce a series of coordinates versus time, this presents a means to test the model and the sensitivity of the docking of ACh with respect to local fluctuations of the model. In order to do this we took 100 snapshots from the simulation, one every 100 ps. We then docked ACh back into these snapshots to assess how well the binding pocket retained its shape and ability to accommodate ACh. This analysis showed that the results observed for the starting model were maintained throughout the simulation (see typical snapshots taken from frames at 5 ns in Figure 5A-C), suggesting that *small, local* residue fluctuations were not critical in determining the *pattern* of binding. Docking of ACh to snapshots from the wild-type, L118D and L118E resulted in solutions that were both in the binding pocket and had an orientation consistent with carbamylcholine bound in AChBP (Celie et al., 2004a; Celie et al., 2004b). Conversely, L118K and L118R mutant models failed to produce docking solutions with the ACh in the binding pocket. Thus, the results from the *in-silico* study suggested that the negatively charged mutations (L118D and L118E) would retain the capability to bind ACh, but the positively charged mutations (L118K and L118R) would not.

Leu118 site-directed mutagenesis.

We investigated the effects of L118D, L118E, L118K and L118R substitutions on the functional α7 nAChR expressed in *Xenopus laevis* oocytes. Control experiments on wild-type α7 show that imidacloprid is a partial agonist consistent with earlier experiments (Matsuda et al., 2000). Figures 6B and 6C show that following the L118D and L118E mutations, responses of the α7 receptor to imidacloprid were abolished ($P<0.01$ using a one-way ANOVA Dunnett's multiple comparison test). However, the α7 mutants still responded to ACh (Figure 6B, 6C). The EC_{50} values of ACh were increased by these two mutations (Table 1). In complete contrast to the findings for the L118D and L118E mutations, L118K and L118R mutations blocked the responses to ACh, whereas responses to imidacloprid were observed (Figures 6D and 6E).

Desnitro-imidacloprid (DN-IMI) is a derivative of imidacloprid lacking the nitro group. The guanidine moiety of this compound is protonated at neutral pH (Figure 1C). The maximum

response to DN-IMI of the wild-type α7 nAChR expressed in *Xenopus* oocytes, was slightly greater than the response to ACh (Figure 7A), resembling results of a previous study (Ihara et al., 2003). The L118D and L118E mutations did not significantly affect the maximum response of α7 to DN-IMI (Figures 7B and 7C) (Table 2). However, the concentration-response curve of DN-IMI was shifted to the right by the L118D and L118E mutations (Table 1). In contrast to the effects on the maximum response to imidacloprid, the L118K and L118R mutations abolished the nicotinic receptor response to DN-IMI (Figures 7D and 7E). Thus the pEC_{50} values for DN-IMI could not be determined for these two mutant receptors

The concentration-response curves of ACh, imidacloprid and DN-IMI for the wild-type (A) and L118D (B), L118E (C), L118K (D) and L118E (E) mutants are shown in Figure 8. It could be argued that in certain mutants imidacloprid may antagonise the response to ACh and ACh may antagonise the actions of imidacloprid. To begin to address this, we have examined if the response to imidacloprid of the L118R mutant and, similarly, the response to ACh of the L118D mutant are suppressed by co-application with ACh and imidacloprid, respectively. It was found that the response to 1 mM imidaclolprid in loop E was suppressed by co-application with 1 mM ACh to 0.306 ± 0.099 (N=2) as compared with the response (=1.00) to 1 mM imidaclolprid alone. On the other hand, the response to 3 mM ACh of the L118D mutant was suppressed by co-application with 1 mM imidacloprid to 0.721 ± 0.125 (N=2) as compared with the response (=1.00) to 3 mM ACh alone.

Discussion.

Our prediction of ACh-nAChR interaction for several mutants at position L118 of α7 is supported by the *in vitro* experiments, suggesting that this model is capable of capturing the effects of ACh interacting with homomeric α7 nAChRs. A previous report also highlights the involvement of the corresponding residue in ligand binding where L119C, L119C and L121C mutants of the vertebrate muscle nAChR subunits γ, ε and δ, respectively, treated with aminoethyl methanethiosulfonate showed reduced affinity for dimethyl-*d*-tubocurarine and α-conotoxin M1 (Sine, 1997). Similar results were observed for the γL119K, εL119K and δL121K mutants. Several studies have shown that electrostatic forces are important for imidacloprid interaction with nAChRs (Sattelle et al., 2005). Here we show that substitutions of L118 in loop E by acidic or basic residues strikingly influence the responses of α7 to ACh. It should be noted that antagonist action of ACh on the response of the L118D mutant to 1 mM imidacloprid was not potent enough to block the response completely. A simple explanation for the effects of such mutations is that if L118 is mutated to a positive side-chain, the interaction with the charged quaternary nitrogen is strong enough to prevent binding to the pocket and thus abolishes binding and the channel fails to open. Conversely, if L118 is mutated to a negatively charged side-chain, there will be a stabilizing effect. We have confidence that our model of $(\alpha7)_5$ is reasonable due to the fact that a similar orientation of acetylcholine to that found for carbamylcholine in AChBP (Celie et al., 2004a) was predicted. Furthermore, this model seems robust to small local changes in residue conformation as exemplified by the molecular dynamics (Figure 4).

Our initial model did not allow us to make confident predictions about the interactions of imidacloprid with the $(\alpha7)_5$ model. This is perhaps not too surprising given that our model is based upon AChBP with HEPES bound. If the receptor conforms to an induced-fit model in terms of its interactions with ligands, then the model will suffer from being biased away from a conformation that would bind imidacloprid. Furthermore, recent experiments on AChBP (Gao

et al., 2005) and on homology models of α7 (Henchman et al., 2005) (Henchman et al., 2003) suggest that there can be substantial changes in the shape of the binding pocket, in particular the conformation of loop C.

However, taking all these factors into consideration along with our results, we propose a model that account for the data shown in Figures 6 and 7. Imidacloprid has a negatively charged nitro group whilst its desnitro derivative (DN-IMI) possesses a positive charge at the guanidine moiety. It is interesting to note that the effects of DN-IMI with respect to the mutations are quite similar to ACh. Thus it could be that the position of these positive charges in the binding pocket are similar. However, that would presumably require a different conformation of loop C in order to allow the DN-IMI to sterically fit in a favourable position within the binding pocket. Imidacloprid itself is inactive in channel-opening of the L118D and L118E mutants, but is active against the L118K and L118R mutants. However, it modulated the ACh-induced response of the L118D mutant, suggesting that the electrostatic interaction between the nitro group and this residue is critical in determining whether it will permit the channel opening in response to binding of the agonist. A similar explanation is possible for the action of ACh on the mutant receptors.

Electrostatic interactions are not the only forces determining agonist interactions with the α7 nAChR. The L118D and L118E mutations shifted the ACh concentration-response curve to higher concentrations and slightly reduced the maximum current amplitudes observed in response to ACh, suggesting that such mutations in loop E may lead indirectly to a conformational change in another region of the receptor playing a key role in interactions with agonists. Thus, it will be necessary to further improve the modeling to gain a more complete understanding of neonicotinoid-nAChR interactions. Nevertheless, electrostatic interactions can explain the changes in the concentration-response curves of the imidacloprid derivative DM-IMI lacking the nitro group. Unlike imidacloprid, DN-IMI has a positive charge at the guanidine moiety, thereby mimicking ACh. Consistent with this, the L118K and L118R mutations abolished responses of the α7 nAChR to DN-IMI, whereas L118D and L118E mutations permitted the agonist actions of this ligand.

Leu118 is not highly conserved across the nAChR family. Thus, the residue at this location may participate in determining subunit-specific responses to agonists. Studies on functional nAChRs consisting of either a Drosophila α1 (Dα1) or Drosophila α2 (Dα2) subunit with a chicken β2 subunit have shown that these receptors are more sensitive to imidacloprid than the chicken α4β2 receptor (Matsuda et al., 2001a). Also, radioligand-binding assays show that receptors comprising rat β2 with either Dα1, Dα2 or Dα3 had considerably higher affinity binding of imidacloprid than the rat α4β2 nAChR (Lansdell and Millar, 2000). These studies suggest that the three Drosophila subunits have structural features favorable for selective interactions with imidacloprid. Interestingly, the residues in Dα1 and Dα3 corresponding to Leu118 are lysine and arginine respectively (Table 2), which in our site-directed mutagenesis experiments increased the affinity for imidacloprid (Table 1). However, the same is not true for Dα2, which has a threonine residue, consistent with other regions of the receptor also being important in determining neonicotinoid sensitivity (Shimomura et al., 2002; Shimomura et al., 2006; Shimomura et al., 2004; Shimomura et al., 2003). Radioligand-binding studies of hybrid nAChRs consisting of rat β2 and an α subunit from the peach-potato aphid, *Myzus persicae*, showed that Mpα2/β2 and Mpα3/β2 have high-affinity binding of imidacloprid whereas Mpα1/β2 and Mpα4/β2 showed no binding (Huang et al., 1999). Consistent with this, finding,

the loop E sequences of Mpα2 and Mpα3 possess a lysine at the position equivalent to Leu118 whereas Mpα1 and Mpα4 do not have a basic residue at this position (Table 2).

It is thought that in heteromeric nAChRs consisting of α and β subunits, loop E from the β subunit contributes to the agonist binding site (Corringer et al., 2000b). In this regard, it is interesting to note that the Dα3/ratβ2 hybrid nAChR showed imidacloprid binding which was not the case for Dα3/ratβ4, demonstrating that the β subunit also contributes to neonicotinoid sensitivity (Lansdell and Millar, 2000). Interestingly, Leu118 is conserved in both ratβ2 and ratβ4 (Table 2), indicating that this residue alone is not critical in determining imidacloprid sensitivity. Recent studies have shown that two mutations in loop D of chicken β2 were required to make the chicken α4β2 nAChR sensitive to imidacloprid (Shimomura et al., 2006). It will be of interest to see whether in certain subunits the combination of the residue corresponding to Leu118 along with other amino acids is important for neonicotinoid sensitivity. In our study, we found that the L118E mutation in α7 abolished imidacloprid action while retaining sensitivity to ACh. In the Drosophila Dβ2 subunit, the residue corresponding to Leu118 is glutamic acid (Table 2). Imidacloprid-insensitive nAChR subtypes have been indicated by several studies, an example of which shows distinct binding sites for imidacloprid and α-bungarotoxin in Drosophila head membranes (Zhang et al., 2004). It will be of interest to see whether Dβ2 is a component of the α-bungarotoxin-sensitive receptor subtype.

It is important to note that the subunit composition and stoichiometry of native insect nAChRs have yet to be determined. Such information will prove instructive in assessing the relevance of loops A-F of either α or β subunits in agonist binding and therefore neonicotinoid sensitivity. Also, mutations at L118 may provide a route by which target-site resistance could develop. If it is the case that in several insect nAChR α subunits, the arginine or lysine at the residue corresponding to L118 contributes to imidacloprid sensitivity, a simple negative charge mutation here would be enough to maintain responses to the natural agonist, ACh, while abolishing the agonist actions of imidacloprid.

In conclusion, we have used molecular modeling to predict that Leu118 of the α7 nAChR subunit contributes to agonist binding. Site-directed mutagenesis and functional expression of wild-type and mutant nAChRs subsequently confirmed this prediction. We have also highlighted the possibility of loop E in contributing to the binding of the insecticide, imidacloprid. These results, taken together with our previous studies on loops C, D and F (Shimomura et al., 2002; Shimomura et al., 2006; Shimomura et al., 2004; Shimomura et al., 2003) are enhancing our understanding of the binding to nAChRs of commercially important nicotinic agonists.

Acknowledgements.

We thank M. Ballivet for kindly providing the chicken α7 cDNA.

References

Amiri S, Tai K, Beckstein O, Biggin PC and Sansom MS (2005) The alpha7 nicotinic acetylcholine receptor: molecular modelling, electrostatics, and energetics. *Mol Membr Biol* **22**:151-62.

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M and Sonnhammer EEL (2002) The pfam protein families database. *Nucleic Acids Res.* **30**:276-280.

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A and Haak JR (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**:3684-3690.

Brejc K, van Dijk WJ, Klassen RV, Schuurmans M, van Der Oost J, Smit AB and Sixma TK (2001) Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**:269-276.

Brejc K, van Dijk WJ, Smit AB and Sixma TK (2002) The 2.7 Å structure of AChBP, homologue of the ligand-binding domain of the nicotinic acetylcholine receptor. *Novartis Found Symp.* **245**:22-29.

Celie PH, Kasheveroc IE, Mordvintsev DY, Hogg RC, van Nierop P, van Elk R, van Rossum-Fikkert SH, Zhmak MN, Bertrand D, Tsetlin V, Sixma TK and Smit AB (2005a) Crystal strucutre of nicotinic acetylcholine receptor homolog AChBP in complex with an alpha-conotoxin PnIA variant. *Nat. Struct. Mol. Biol.* **12**:582-588.

Celie PH, Klassen RV, van Rossum-Fikkert SH, van Elk R, van Nierop P, Smit AB and Sixma TK (2005b) Crystal structure of acetylcholine-binding protein from Bulinus truncatus reveals the conserved structural scaffold and sites of variation in nicotinic acetylcholine receptors. *J. Biol. Chem* **280**:26457-26466.

Celie PH, van Rossum-Fikkert SE, van Dijk WJ, Brejc K, Smit AB and Sixma TK (2004a) Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**:907-14.

Celie PH, van Rossum-Fikkert SH, van Dijk WJ, Brejc K, Smit AB and Sixma TK (2004b) Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**:907-914.

Changeux J-P and Edelstein SJ (2001) Allosteric mechanisms in normal and pathological nicotinic acetylcholine receptors. *Curr. Opin. Neurobiol.* **11**:369-377.

Corringer P-J, Le Novere N and Changeux J-P (2000a) Nicotinic receptors at the amino acid level. *Annu. Rev. Pharmacol. Toxicol.* **40**:431-458.

Corringer P-J, Le Novère N and Changeux J-P (2000b) Nicotinic receptors at the amino acid level. *Annu. Rev. Pharmacol. Toxicol.* **40**:431-458.

Darden T, York D and Pedersen L (1993) Particle mesh Ewald - an N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**:10089-10092.

DeLano WL (2004) The PyMOL molecular graphics system. *DeLano Scientific LLC, San Carlos, CA. UDA.*

Gao F, Bren N, Burghardt TP, Hansen SB, Henchman RH, Taylor P, McCammon JA and Sine SM (2005) Agonist-mediated conformational changes in acetylcholine-binding potein revealed by simulation and intrinsic tryptophan fluorescence. *J. Biol. Chem* **280**:8443-8451.

Hansen SB, Sulzenbacher G, Huxford T, Marchot P, Taylor P and Bourne Y (2005) Structures of the Aplysia AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. *EMBO J.*:1-12.

Harrow ID and Gration KAF (1985) Mode of action of the anthelmintics morantel, pyrantel and levimisole on muscle cell membrane of the nematode Ascatis suum. *Pesti. Sci* **16**:662-672.

Henchman RH, Wang H-L, Sine SM, Taylor P and McCammon JA (2003) Asymmetric structural motions of the homomeric a7 nicotinic receptor ligand binding domain revealed by molecular dynamics simulation. *Biophys. J.* **85**:3007-3018.

Henchman RH, Wang HL, Sine SM, Taylor P and McCammon JA (2005) Ligand-induced conformational change in the alpha7 nicotinic receptor ligand binding domain. *Biophys J* **88**:2564-76.

Hess B, Bekker J, Berendsen HJC and Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* **18**:1463-1472.

Huang Y, Williamson MS, Devonshire AL, Windass JD, Lansdell SJ and Millar NS (1999) Molecular characterization and imidacloprid selectivity of nicotinic acetylcholine receptor subunits from the peach-potato aphid *Myzus persicae*. *J Neurochem* **73**:380-9.

Humphrey W, Dalke A and Schulten K (1996) VMD - Visual molecular dynamics. *J. Molec. Graph.* **14**:33-38.

Jones AK, Buckingham SD and Sattelle BM (2005) Chemistry-to-gene screens in Caenorhabditis elegans. *Nat Rev Drug Disc* **4**:321-330.

Karlin A (2002) Emerging structure of the nicotinic acetylcholine receptors. *Nat Rev Neurosci* **102**:102-114.

Lang B and Vincent A (2003a) Autoantibodies to ion channels at the neuromuscular junction. *Autoimmun Rev* **2**:94-100.

Lang B and Vincent A (2003b) Autoantibodies to ion channels at the neuromuscular junction. *Autoimmun. Rev.* **2**:94-100.

Lansdell SJ and Millar NS (2000) The influence of nicotinic receptor subunit composition upon agonist, alpha-bungarotoxin and insecticide (imidacloprid) binding affinity. *Neuropharmacology* **39**:671-9.

Laskowski RA, Macarthur MW, Moss DS and Thornton JM (1993) PROCHECK - A program to check the stereochemical quality of protein structures. *J. Appl. Crystall.* **26**:283-291.

Le Novère N, Grutter T and Changeux J-P (2002) Models of the extracellular domain of the nicotinic receptors and of agonist- and Ca2+ -binding sites. *Proc. Natl. Acad. Sci.* **99**:3210-3215.

Lena C and Changeux J-P (1998) Allosteric nicotinic receptors, human pathologies. *J. Physiol. Paris* **92**:63-74.

Lloyd GK and Williams M (2000) Neuronal nicotinic acetylcholine receptors as novel drug targets. *J. Pharmacol. Exp. Therapeu.* **292**:461-467.

Matsuda K, Buckingham SD, Kleier D, Rauh JJ, Grauso M and Sattelle DB (2001a) Neonicotinoids: insecticides acting on insect nicotinic acetylcholine receptors. *Trends Pharmacol Sci* **22**:573-80.

Matsuda K, Ihara M, Nishimura K, Sattelle DB and Komai K (2001b) Insecticidal and neural activities of candidate photoaffinity probes for neonicotinoid binding sites. *Biosci Biotechnol Biochem.* **65**:1534-1541.

Matsuda K, Shimomura M, Kondo Y, Ihara M, Hashigami K, Yoshida M, Raymond V, Mongan NP, Freeman JC, Komai K and Sattelle DB (2000) Role of loop D of the alpha7 nicotinic acetylcholine receptor in its interaction with the insecticide imidacloprid and related neonicotinoids. *Br. J. Pharmacol.* **130**:981-986.

Millar NS (2003a) Assembly and subunit diversity of nicotinic acetylcholine receptors. *Biochem. Soc. Trans.* **31**:869-74.

Millar NS (2003b) Assembly and subunit diversity of nicotinic acetylcholine receptors. *Biochem. Soc. Trans.* **31**:869-874.

Miyazawa A, Fujiyoshi Y, Stowell M and Unwin N (1999) Nicotinic acetylcholine receptor at 4.6 Å resolution: Transverse tunnels in the channel wall. *J. Mol. Biol.* **288**:765-786.

Miyazawa A, Fujiyoshi Y and Unwin N (2003) Structure and gating mechanism of the acetylcholein receptor pore. *Nature* **424**:949-955.

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK and Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* **19**:1639-1662.

Nose S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**:255-268.

Ohno K and Engel AG (2002) Congenital myasthenic syndromes: genetic defects of the neuromuscular junction. *Curr. Neurol. Neurosci. Rep.* **2**.

Papke RL, Meyer E, Nutter T and Uteshev VV (2000) α7 receptor-selective agonists and modes of α7 receptor activation. *Eur. J. Pharmacol.* **393**:179-195.

Parinello M and Rahman A (1981) Polymorphic transitions in single crystals - a new molecular dynamics method. *J. Appl. Phys.* **52**:7182-7190.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004) USCF chimera - A visualization system for exploratory research and analysis. *J. Comp. Chem.* **25**:1605-1612.

Prendergast MA, Jackson WJ, Terry AVJ, Decker MW, Arneric SP and Buccafusco JJ (1998) Central nicotinic receptor agonists ABT-418, ABT-089, and (-)-nicotine reduce distractibility in adult monkeys. *Psychopharmacology (Berl)* **136**:50-58.

Sali A and Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**:779-815.

Sattelle DB, Jones AK, Sattelle BM, Matsuda K, Reenan R and Biggin PC (2005) Edit, cut and paste in the nicotinic acetylcholine receptor gene family of *Drosophila melanogaster*. *Bioessays* **27**:366-76.

Schapira M, Abagyan R and Totrov M (2002) Structural model of nicotinic acetylcholine receptor isotypes bound to acetylcholine and nicotine. *BMC Struct. Biol.* **2**:1-8.

Segall MD, Payne MC and Boyes RN (1998) An ab initio study of the conformational energy map of acetylcholine. *Molec. Phys.* **93**:365-370.

Shimomura M, Okuda H, Matsuda K, Komai K, Akamatsu M and Sattelle DB (2002) Effects of mutations of a glutamine residue in loop D of the alpha7 nicotinic acetlycholine receptor on agonist profiles for neonicotinoid insecticides and related ligands. *Br. J. Pharmacol.* **137**:162-169.

Shimomura M, Yokota M, Ihara M, Akamatsu M, Sattelle DB and Matsuda K (2006) Role in the selectivity of neonicotinoids of insect-specific basic residues in loop D of the nicotinic acetylcholine receptor agonist binding site. *Molecular Pharmacology* **In Press.**

Shimomura M, Yokota M, Matsuda K, Sattelle DB and Komai K (2004) Roles of loop C and the loop B-C interval of the nicotinic receptor a subunit in its selective interactions with imidacloprid in insects. *Neurosci. Lett.* **363**:195-198.

Shimomura M, Yokota M, Okumura M, Matsuda K, Akamatsu M, Sattelle DB and Komai K (2003) Combinatorial mutations in loops D and F strongly influence responses of the α7 nicotinic acetylcholine receptor to imidacloprid. *Brain Research* **991**:71-77.

Sine SM (1997) Identification of equivalent residues in the gamma, delta, and epsilon subunits of the nicotinic receptor that contribute to alpha-bungarotoxin binding. *J Biol Chem* **272**:23521-7.

Steinlein OK (2001) Genes and mutations in idiopathic epilepsy. *Am J. Med. Genet.* **106**:139-145.

Unwin N (2005) Refined Structure of the Nicotinic Acetylcholine Receptor at 4Å Resolution. *J. Mol. Biol.* **346**:967-989.

van Gunsteren WF, Krüger P, Billeter SR, Mark AE, Eising AA, Scott WRP, Hüneberger PH and Tironi IG (1996) Biomolecular simulation: The GROMOS96 manual and user guide, Biomos/Hochschulverlag AG an der ETH Zürich, Groningen/Zürich.

Vernino S and Lennon VA (2003a) Neuronal ganglionic acetylcholine receptor autoimmunity. *Ann N Y Acad Sci* **998**:211-4.

Vernino S and Lennon VA (2003b) Neuronal ganglionic acetylcholine receptor autoimmunity. *Ann. N.Y. Acad. Sci.* **998**:211-214.

Watson R, Jepson JE, Bermudez I, Alexander S, Hart Y, McKnight K, Roubertie A, Fecto F, Valmier J, Sattelle DB, Beeson D, Vincent A and Lang B (2005a) Alpha7-acetylcholine receptor antibodies in two patients with Rasmussen encephalitis. *Neurology* **65**:1802-4.

Watson R, Jepson JE, Bermudez I, Alexander S, Hart Y, McKnight K, Roubertie A, Fecto F, Valmier J, Sattelle DB, Beeson D, Vincent A and Lang B (2005b) Alpha7-acetylcholine receptor antibodies in two patients with Rasmussen encephalitis. *Neurology* **65**:1802-1804.

Zhang N, Tomizawa M and Casida JE (2004) Drosophila nicotinic receptors: evidence for imidacloprid insecticide and alpha-bungarotoxin binding to distinct sites. *Neurosci Lett* **371**:56-9.

Footnotes.

PCB and MSPS both thank the Wellcome Trust for support. A Basic Research Activities for Innovative Biosciences (Bio-oriented Technology Research Advancement Institution) grant to KM is gratefully acknowledged. DBS and AKJ acknowledge the support of the Medical Research Council and a grant from Dupont, USA. KM acknowledges the support of the program for Basic Research Activities for Innovative Biosciences (Bio-oriented Technology Research Advancement Institution: BRAIN) and "Academic Frontier" Project for Private Universities: matching fund subsidy from the MEXT of Japan.

Table 1. pEC_{50}, Imax and Hill coefficient values for acetylcholine, imidacloprid and desnitro-imidacloprid on wild-type and mutant α7 receptors expressed in *Xenopus* oocytes

	Acetylcholine				Imidacloprid				Desnitro-Imidacloprid			
	Imax[a]	pEC_{50}	Hill	N	Imax	pEC_{50}	Hill	N	Imax	pEC_{50}	Hill	N
Wild type	1.03 ± 0.04	3.91 ± 0.05	1.4 ± 0.2	5	0.49 ± 0.03	3.63 ± 0.07	1.6 ± 0.4	6	1.21 ± 0.05	5.14 ± 0.07	1.2 ± 0.2	4
L118D	1.07 ± 0.04	2.78 ± 0.04**,[c]	1.3 ± 0.2	4	ND[b]	ND	ND	4	1.24 ± 0.11	4.21 ± 0.11**	1.5 ± 0.5	4
L118E	0.97 ± 0.02	3.12 ± 0.03**	1.7 ± 0.2	7	ND	ND	ND	5	1.37 ± 0.06	4.28 ± 0.08**	1.6 ± 0.3	5
L118K	ND	ND	ND	4	1.14 ± 0.10	2.94 ± 0.05**	2.1 ± 0.4	4	ND	ND	ND	4
L118R	ND	ND	ND	5	1.07 ± 0.05	3.00 ± 0.02**	2.4 ± 0.4	5	ND	ND	ND	4

a Normalized maximum response. See Methods for details

b Not determined because the response to agonists was not detected or very small.

c Values shown are the result of a fit of the concentration-response data (mean ± standard error of the mean) illustrated in Fig. 8.

Statistical test (one-way ANOVA, Dunnett's multiple comparison test) is for significant differences from the wild-type data (** <0.01).

Table 2. Amino acid sequences in loop E. The residue corresponding to L118 of chicken α7 is highlighted in grey shading and basic residues at this position are shown in bold.

nAChR subunit	Loop E sequence
Lymnaea AChBP	GEVLY**M**PSIRQ
Chicken α7	GHCQY**L**PPGIF
Chicken α4	GRIKW**M**PPAIY
Chicken β2	GSIFW**L**PPAIY
Rat β2	GSIFW**L**PPAIY
Rat β4	GSIQW**L**PPAIY
Drosophila α1 (ALS)	GKVVW**K**PPAIY
Drosophila α2 (SAD)	GKVVW**T**PPAIF
Drosophila α3	GRVEW**R**PPAIY
Drosophila α4	GLVEW**K**PPAIY
Drosophila α5	GSCLY**V**PPGIF
Drosophila α6	GSCLY**V**PPGIF
Drosophila α7	GSCLY**V**PPGIF
Drosophila β1 (ARD)	GEVLW**V**PPAIY
Drosophila β2 (SBD)	GEVFW**E**PPAIY
Drosophila β3	GHFRW**M**PPAVY
Myzus α1	GKVMW**T**PPAIY
Myzus α2	GKVVW**K**PPAIY
Myzus α3	GRVEW**K**PPAIY
Myzus α4	GEVLW**S**PPAIY

Figure Legends

Figure 1. A) The alignment used with Modeller in order to generate the initial homology model. The positions of the six loops (Loop A-F) involved in ligand binding are indicated. B) Only two of the five subunits were used in the docking and simulations runs. The various loops that have been shown to contribute to binding are highlighted: Loop A (blue), Loop B (yellow), Loop C (pink), Loop D (orange), Loop E (green) and Loop F (red). The position L118 and other key residues involved in ACh-binding are highlight in licorice. C) Chemical structures of the ligands used in this study.

Figure 2. Lowest energy docking solutions against the wild-type α7 model for A) ACh (from a total of 4 clusters), B) for IMI (from 7 clusters within 1.5 kcal/mol of the lowest energy solution) and C) for DN-IMI (from 6 clusters within 1.5 kcal/mol of lowest energy solution).

Figure 3. A) Dock of ACh to wild-type α7 model (purple) over-layed onto the AChBP-carbamylcholine structure (1UV6, grey). The position of L118 is shown in spacefill representation. B) Docking of ACh to L118D. The principal subunit is shown in grey and the complimentary subunit is in purple. The D118 residue is shown in spacefill and ACh is shown in liquorice. C) Docking of ACh to L118E. The E118 residue is shown in spacefill. There were no docks in the binding pocket for the L118K and L118R mutations.

Figure 4. A) RMSFs of wildtype (grey line) compared to the L118D mutation (black line). B) RMSF of wildtype (grey line) compared to L118E (blackline). C) RMSF of wildtype (grey line) compared to L118K (blackline). D) RMSF of wildtype (grey line) compared to L118R (blackline). The mean of all five subunits in each case is plotted.

Figure 5. Docking against snapshots taken from the MD trajectory at 5 ns. Principal subunits are shown in grey and complimentary subunits are shown in purple. A) An example of a dock to snapshot at time = 5 ns for the wild-type. B) An example of a dock to a snapshot at time = 5 ns for the L118D mutation. C) An example of a dock to a snapshot at time = 5 ns for the L118E mutation. The orientation of the docks to these mutants is similar to the solutions found at 0 ns.

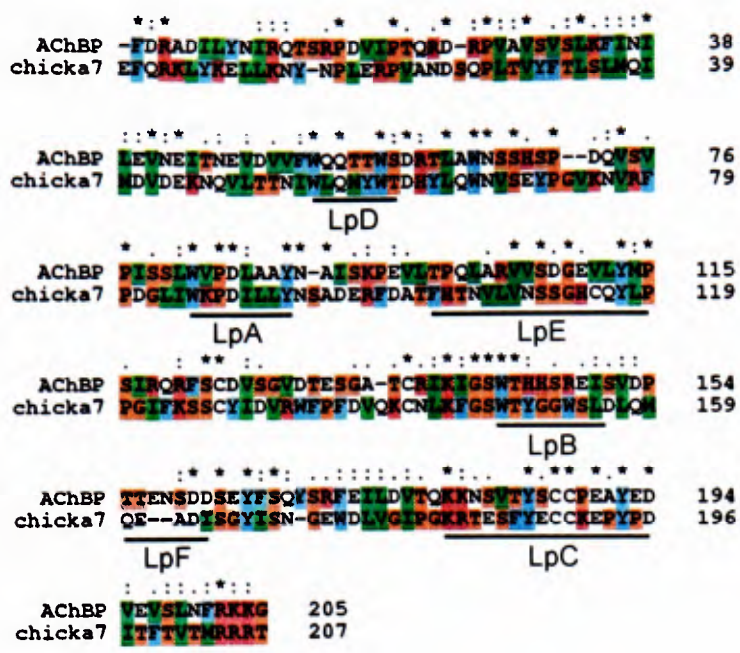
Figure 6. Maximum current responses of α7 nAChRs to ACh and imidacloprid (IMI). Imidacloprid was a partial agonist, but such an action was completely reduced by L118D (B) and L118E (C) mutations whilst the ACh-induced response was retained. By contrast, L118K (D) and L118R (E) mutations abolished the response to ACh but not to imidacloprid.

Figure 7. Maximum current responses of α7 nAChRs to ACh and the desnitro-derivative of imidacloprid, DN-IMI. DN-IMI induced similar current amplitude of the maximum response to ACh of the wild-type nicotinic receptor (A). L118D (B) and L118E (C) mutations retained the DN-IMI-induced response, whereas L118K (D) and L118R (E) mutations markedly reduced it.

Figure 8. Concentration-response curves of ACh, imidacloprid and the desnitro-derivative of imidacloprid obtained for wild-type (A), as well as L118D (B), L118E (C), L118K (D) and L118R (E) mutants of the α7 nAChR expressed in *Xenopus laevis* oocytes. Each plot represents mean ± standard error of the mean of 4-7 experiments .

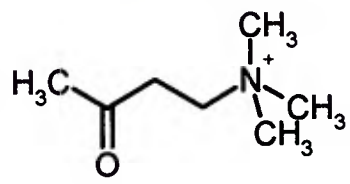
Figure 1

A

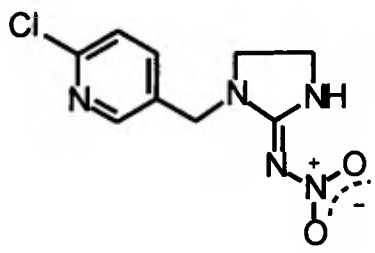


B

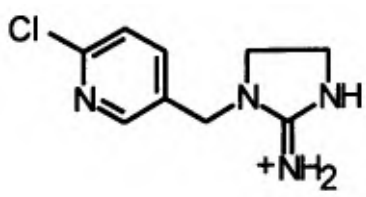
C



acetylcholine



imidacloprid

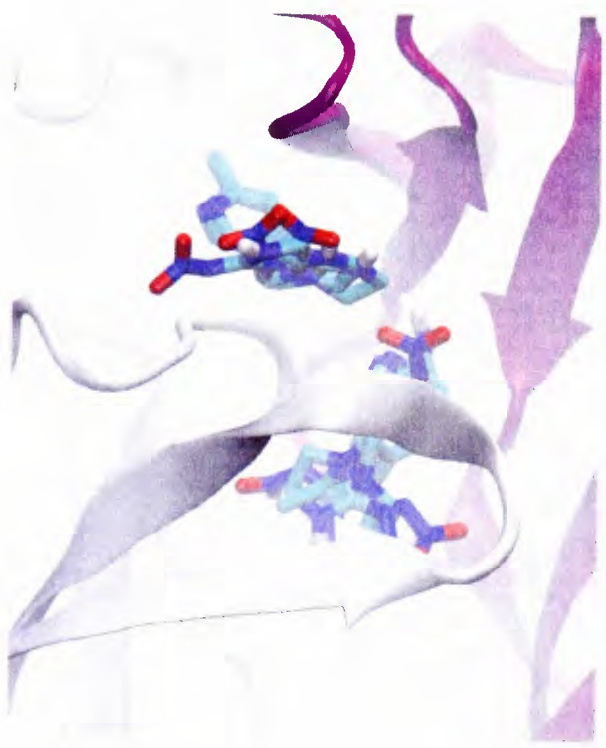


DN-imidacloprid

Figure 2

A

B

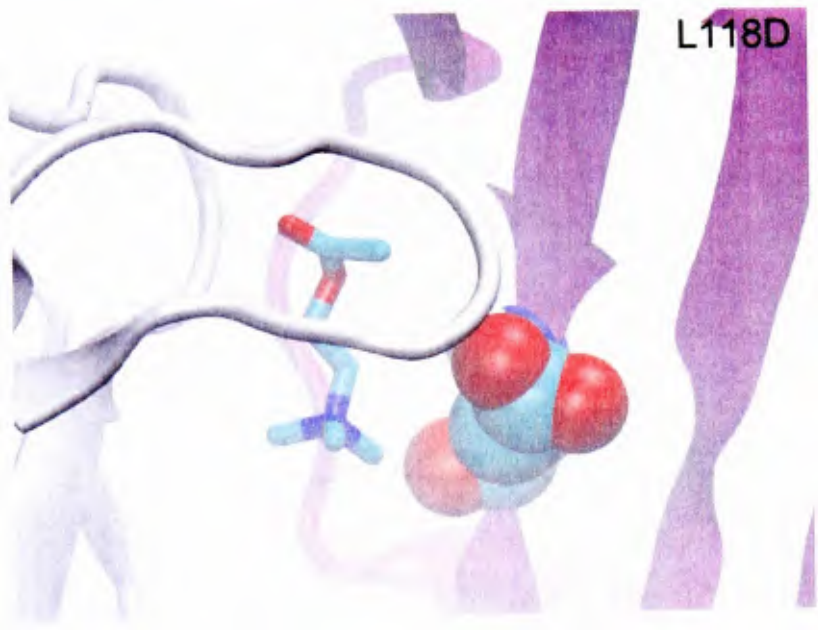


C

Figure 3

A

B



C

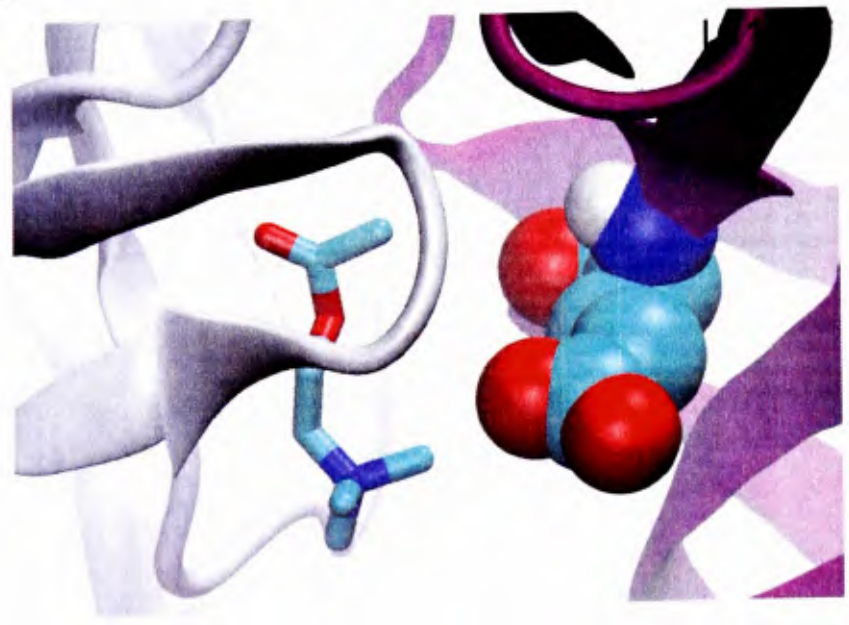


Figure 4

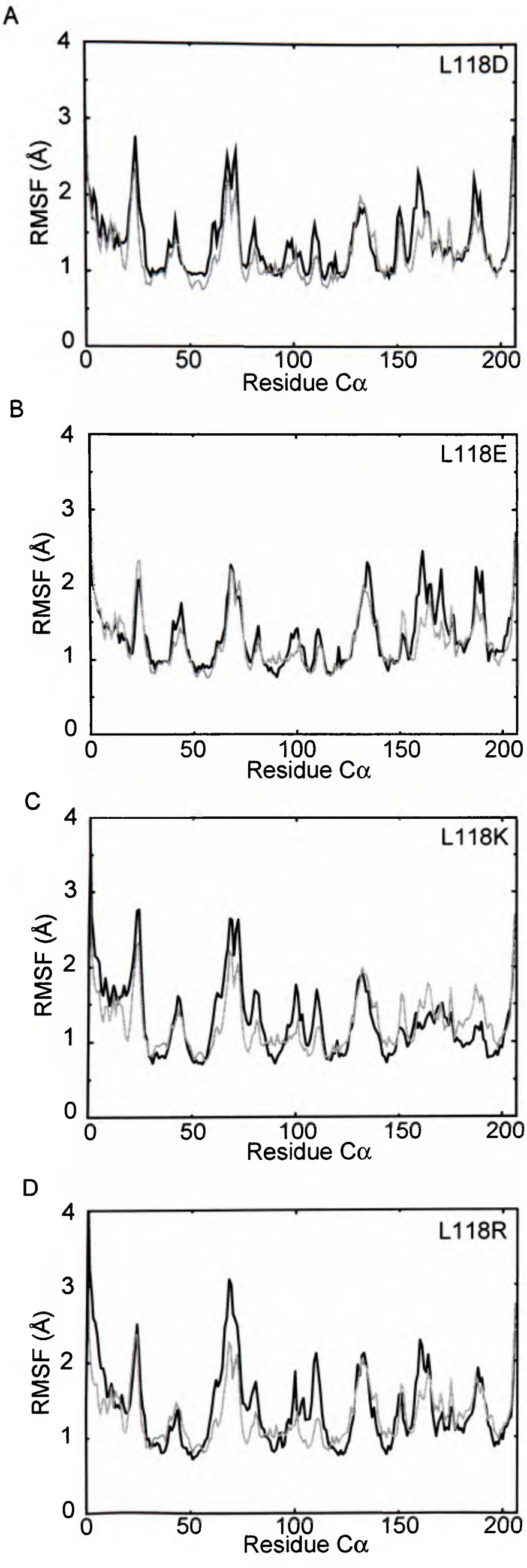
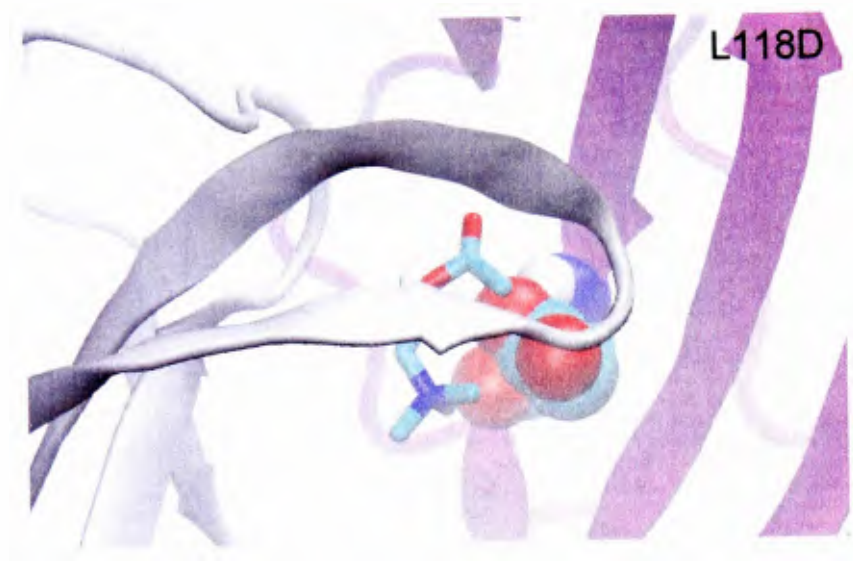

A
L118D
RMSF (Å)
Residue Cα
B
L118E
RMSF (Å)
Residue Cα
C
L118K
RMSF (Å)
Residue Cα
D
L118R
RMSF (Å)
Residue Cα

Figure 5

A

B



C