Nicotinic acetylcholine receptors and their interactions with allosteric ligands

Joseph Newcombe

August 2018
UCL & Birkbeck

Department of Chemistry and Institute of Structural and Molecular Biology

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy
Declaration

I, Joseph Newcombe confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

..............................................
Acknowledgements

Firstly, I would like to thank my supervisors, Prof. Tom Sheppard, Prof. Maya Topf and Prof. Neil Millar. Your help and guidance pulled me through this project. I am also grateful to each of you for your laid back nature, which has allowed me to explore this project in my own way. I am especially thankful to Maya, who brought me into the fold of computational biology.

The staff who make things work are some of the most important people. Dr David Houldershaw always helped me sort out problems I had caused by blindly stumbling into the darkness with installations on my laptop, as well as ensuring every piece of software I could possibly need was installed and running on the Birkbeck servers. Dr Abil Aliev made it very simple and easy to obtain NMR data on all of my chemical entities in UCL NMR.

A huge thank you is extended to the Sheppard, Topf and Millar groups. From the Sheppard group, I would like to particularly thank Laure, Marco, Valerie, Sam, Jarryl and Charlotte. You helped me get started in the KLB and welcomed me back on my short periods of return. From the Topf group, I would like to thank Arun, Agnel, Josh B, Anna and Josh H. You pulled me out of trouble with Python, docking, model building and map sharpening. From the Millar group, I would like to thank Anna, Charlie and Vic, without your work, my models and chemistry would have had little meaning. Last but not least, thank you to all the other members of the KLB and office 358, as well as, all the friendly faces around the university who make going to work a joy.

My family have played a huge part in where I am today. Thank you so much to my Mum and Dad who have supported me in everything I do. You always helped me to rationalise my life choices and prioritise the goals that I want, not those that I think I should have. The other central support in my life is Emma, my rock, my foundation and my motivation. You pull me out of bed, kick me out the door and make me get on with my life and achieve when I would otherwise stagnate. You have an unfathomable patience and inspire me to work hard.
Abstract

Nicotinic acetylcholine receptors (nAChRs) are pentameric ligand gated ion channels (pLGICs) expressed widely throughout the body, including in the peripheral nervous system, central nervous system and at the neuromuscular junction. nAChRs are of therapeutic interest due to their involvement in several pathophysiological conditions. The most widely expressed nAChR subtypes, α7 and α4β2 have attracted a lot of attention and many allosteric ligands have been pharmacologically and chemically characterised for these receptors. However, much remains to be understood about where and how these ligands bind to the receptors and modulate their function. This thesis has focussed on a set of transmembrane binding allosteric modulators for the α7 nAChR and sought to aid understanding of their interactions with their target receptor by building models of nAChRs in physiologically relevant states. A transmembrane error in the only example of a pLGIC structure determined in a native lipid membrane environment, the *T. marmorata* nAChR, has been corrected through modelling and refinement into previously determined electron cryo-microscopy density maps, in putative closed and open conformations. The refined models offer important reference structures for anyone working in the pLGIC field and here have been used as templates to model the α7 nAChR. A consensus docking protocol has been developed and was utilised in conjunction with the α7 models to predict binding modes for a set of allosteric modulators and provide insight into how they may elicit distinct pharmacology. Based on binding modes of allosteric modulators predicted by the consensus docking protocol, pharmacophores were generated for use in ligand-based virtual screening and allosteric modulators have been uncovered for α7 and α4β2 nAChRs from the existing pharmacopeia. Further to this, novel reactive chemical probes have been developed and synthesised to study the covalent incorporation of allosteric modulators into nAChRs.

Impact statement
The research within this thesis has focussed on small-molecules and a set of proteins to which they bind called nicotinic acetylcholine receptors (nAChRs). These nAChRs are members of a broader family of proteins called pentameric ligand gated ion channels (pLGICs), because they are formed of five individual protein subunits. pLGICs induce electrical signals across cell membranes in response to small-molecule neurotransmitters, such as acetylcholine, γ-aminobutyric acid (GABA), glycine and 5-hydroxytryptamine (serotonin). Nicotinic receptors have attracted interest from many academic and industry researchers, partly because they have been implicated in a variety of neurological diseases with great unmet medical need, including Alzheimer’s, schizophrenia, autism, smoking cessation and myasthenia gravis. However, despite a significant body of research in these areas, there has been little success in the discovery of drugs which modulate nAChRs. In the last few decades, allosteric modulation – whereby, small-molecules binding elsewhere to the neurotransmitter binding site change responses of the channel to the endogenous neurotransmitter – has emerged as an exciting area of therapeutic applicability to nAChRs. Yet the interactions which drive allosteric modulation are poorly understood on a molecular level.

This thesis helps to improve the understanding of the structure and function of nAChRs and their interactions with allosteric ligands, which is of great importance to the development of medicines which exploit modulation of nAChRs. Ultimately, structural knowledge of pharmaceutical targets can underpin drug discovery projects and eventually result in successful treatment of pathologies, such as those with which nAChRs are associated (Alzheimer’s, schizophrenia, autism etc.). Effective treatments for these diseases will be of great benefit to society as a whole.

In the more immediate future, methodologies developed within this research will be of importance to other researchers within and outside of the field of pLGICs. The improved structures of the Torpedo nAChR developed in this work are important reference models for researchers working in the active field of pLGIC structural biology and will be beneficial for researchers working on this topic in academia and industry. The docking protocol which has been
developed has been published and continues to generate impact, with three citations within the first six months of publication and hundreds of reads and downloads. Protocols which improve confidence in docking solutions could find use across any application where small-molecules interact with proteins. Finally, the novel allosteric modulators of nAChRs which have been developed as chemical tools or identified through virtual screening, increase the known diversity of allosteric modulators of the α7 nAChR subtype. Diverse knowledge of what chemical features are important for functional activity will aid drug discovery projects aimed at α7 nAChRs, which are ongoing in both academia and industry.
Publications

- **Newcombe J**, Sheppard TD, Millar NS, Topf M, Unwin N, *Manuscript in Preparation*

* Joint first authors
# Table of contents

*Declaration*  
*Acknowledgements*  
*Abstract*  
*Impact statement*  
*Publications*  
*Collaborator acknowledgement*  
*List of abbreviations*  

1 **Introduction**  
  1.1 **Protein-ligand interactions**  
  1.1.1 Nerve signalling  
  1.2 **Nicotinic acetylcholine receptors**  
  1.2.1 nAChR nomenclature  
  1.2.2 Subunit stoichiometry  
  1.2.3 Native nAChR localization  
  1.2.4 Functional measurements of nAChR properties  
  1.2.5 α7* and α4β2* nAChRs  
  1.2.6 Orthosteric ligands  
  1.2.7 Allosteric ligands  
  1.3 **Protein structure determination**  
  1.3.1 X-ray crystallography  
  1.3.2 Protein NMR  
  1.3.3 Electron cryomicroscopy  
  1.4 **Comparative modelling**  
  1.4.1 Template selection  
  1.4.2 Sequence alignment  
  1.4.3 Model building  
  1.4.4 Model evaluation  
  1.5 **Small molecule docking**  
  1.6 **Summary and aims**  
  1.7 **References**
## 2 A revised structural model of the α7 nicotinic acetylcholine receptor

### 2.1 Introduction
- 2.1.1 Template models for α7 nAChR

### 2.2 Aims

### 2.3 Correcting *T. marmorata* nAChR transmembrane error
- 2.3.1 Characterising the error
- 2.3.2 Refinement strategy
- 2.3.3 Model evaluation

### 2.4 Comparative modelling of α7 nAChR

### 2.5 Conclusions and future work

### 2.6 References

## 3 Investigating diversity in positive allosteric modulators of α7 nicotinic acetylcholine receptors using a revised molecular model and consensus docking protocol

### 3.1 Introduction
- 3.1.1 Small-molecule docking
- 3.1.2 Positive allosteric modulators of α7 nAChRs

### 3.2 Aims

### 3.3 Results and discussion
- 3.3.1 Choice of docking program and scoring function
- 3.3.2 Protocol
- 3.3.3 Preparation of ligand files
- 3.3.4 Docking of A-867744, TBS-516 and TQS
- 3.3.5 Discussion of predicted binding modes

### 3.4 Conclusions and future work

### 3.5 References

## 4 Design and synthesis of reactive allosteric modulators of α7 nicotinic acetylcholine receptors

### 4.1 Introduction

### 4.2 Aims

### 4.3 Results and discussion
## 4.3. Design of reactive allosteric modulators

- 4.3.1 Design of reactive allosteric modulators
- 4.3.2 Synthesis of reactive TQS analogues
- 4.3.3 Testing of TQS analogues
- 4.3.4 Prediction of positions of reaction from binding modes

## 4.4 Conclusions and future work

## 4.5 References

## 5 Identification by virtual screening of novel allosteric modulators of the α7 nicotinic acetylcholine receptor

- 5.1 Introduction
- 5.2 Aims
- 5.3 Results and discussion
  - 5.3.1 Pharmacophore query generation
  - 5.3.2 Pharmacophore query validation
  - 5.3.3 Virtual screening of the DrugBank database
  - 5.3.4 Compound testing
- 5.4 Conclusions and future work
- 5.5 References

## 6 Refinement of the full length Torpedo nicotinic acetylcholine receptor

- 6.1 Introduction
- 6.2 Aims
- 6.3 Results and discussion
  - 6.3.1 Estimating the resolution of the new Torpedo maps
  - 6.3.2 Generating an initial model of the Torpedo receptor
  - 6.3.3 Map sharpening
  - 6.3.4 Model fitting
  - 6.3.5 Insight into functional data
  - 6.3.6 Photo-crosslinking data
- 6.4 Conclusions and future work
- 6.5 References

## 7 General summary and future work

- 7.1 Introduction to thesis summary
7.2 Improvements to the *Torpedo* nAChR structure 225

7.2.1 Correcting the transmembrane register error in the *Torpedo* nAChR αγ subunit 225

7.2.2 Correction and fitting of the full length *Torpedo* nAChR including new structural data 226

7.3 α7 nAChR studies 227

7.3.1 Predicted binding modes of α7 nAChR PAMs 228

7.3.2 Reactive allosteric modulators 229

7.3.3 Predicting new α7 nAChR allosteric modulators 230

7.4 General points for future work 232

7.4.1 Improvements and distribution of the consensus docking protocol 232

7.4.2 Further exploration of reactive allosteric modulators 232

7.4.3 Further exploration of the properties of furosemide 233

8 Experimental 235

8.1 Chemistry 236

8.1.1 General Chemistry 236

8.1.2 Chapter 3 – Chemical synthesis 238

8.1.3 Chapter 4 – Chemical synthesis 243

8.1.4 Chapter 5 – Chemical synthesis 256

8.2 References 258
Collaborator acknowledgement

The work carried out in this thesis and the conclusions drawn have hinged on the work of our collaborators whose contributions have been as follows:

**Chapter 3** – All mutagenesis and electrophysiology studies relating to this work were carried out by Dr Anna Chatzidaki in the lab of Prof. Neil Millar (UCL).

**Chapter 4** – All electrophysiology studies relating to this work were carried out by Dr Anna Chatzidaki in the lab of Prof. Neil Millar (UCL).

**Chapter 5** – All mutagenesis and electrophysiology studies relating to this work were carried out by Charles Smelt and Victoria Sanders in the lab of Prof. Neil Millar (UCL).

**Chapter 6** – All sample preparation, structural determination and reconstructions from the raw cryo EM data were performed by Prof. Nigel Unwin (Laboratory of Molecular Biology, Cambridge).
# List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-)</td>
<td>complimentary subunit</td>
</tr>
<tr>
<td>(+)</td>
<td>principal subunit</td>
</tr>
<tr>
<td>1D</td>
<td>one-dimensional</td>
</tr>
<tr>
<td>2D</td>
<td>two-dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>three-dimensional</td>
</tr>
<tr>
<td>4ACA-TQS</td>
<td>2-chloro-(N-(4-(8\text{-sulfamoyl}-3a,4,5,9b\text{-tetrahydro}-3H\text{-cyclopenta}{c}\text{quinolin}-4\text{-yl})\text{phenyl})\text{acetamide}</td>
</tr>
<tr>
<td>4CM-TQS</td>
<td>4-(4-(chloromethyl)phenyl)-3a,4,5,9b-tetrahydro-3H-cyclopenta[c]quinoline-8-sulfonamide</td>
</tr>
<tr>
<td>4TFD-TQS</td>
<td>4-(4-(3\text{-trifluoromethyl}-3H\text{-diazirin}-3-yl)phenyl)-3a,4,5,9b\text{-tetrahydro}-3H\text{-cyclopenta}{c}\text{quinoline}-8-sulfonamide</td>
</tr>
<tr>
<td>5-HT</td>
<td>5-hydroxytryptamine</td>
</tr>
<tr>
<td>5-HT(_3)R</td>
<td>5-hydroxytryptamine type 3 receptor</td>
</tr>
<tr>
<td>Å</td>
<td>Ångström</td>
</tr>
<tr>
<td>A-585539</td>
<td>(1S,4S)-2,2-dimethyl-5-(6-phenylpyridazin-3-yl)-5-aza-2-azoniabicyclo[2.2.1]heptane</td>
</tr>
<tr>
<td>A-867744</td>
<td>4-(5-(4-chlorophenyl)-2-methyl-3-propionyl-1\text{-H-pyrrol}-1-yl)benzenesulfonamide</td>
</tr>
<tr>
<td>AA</td>
<td>allosteric agonist</td>
</tr>
<tr>
<td>ACh</td>
<td>acetylcholine</td>
</tr>
<tr>
<td>AChBP</td>
<td>acetylcholine binding protein</td>
</tr>
<tr>
<td>ADNFLE</td>
<td>autosomal dominant nocturnal frontal lobe epilepsy</td>
</tr>
<tr>
<td>ago-PAM</td>
<td>allosteric agonist with positive allosteric modulator properties</td>
</tr>
<tr>
<td>AUC</td>
<td>area under the curve</td>
</tr>
<tr>
<td>CAII</td>
<td>carbonic anhydrase II</td>
</tr>
<tr>
<td>CCD</td>
<td>charge-coupled device</td>
</tr>
<tr>
<td>CCTBX</td>
<td>computational crystallography toolbox</td>
</tr>
<tr>
<td>CDK2</td>
<td>cyclin dependent kinase two</td>
</tr>
<tr>
<td>CG</td>
<td>conjugate gradients</td>
</tr>
<tr>
<td>Cl(^{-})</td>
<td>chloride ion</td>
</tr>
<tr>
<td>cLogD</td>
<td>calculated octanol-water partition coefficient at physiological pH</td>
</tr>
<tr>
<td>cLogP</td>
<td>calculated octanol-water partition coefficient</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>CNS MPO</td>
<td>central nervous system multiparameter optimisation</td>
</tr>
</tbody>
</table>
cryo EM - electron cryomicroscopy
DB04763 - 2,4,6-trimethyl-1-(4-sulfamoylphenethyl)pyridin-1-ium tetrafluoroborate
DB08122 - (Z)-N-methyl-4-(((2-oxoindolin-3-ylidene)methyl)amino)benzenesulfonamide
DIBAL - diisobutyl aluminium hydride
DMSO - dimethylsulfoxide
DNA - deoxyribose nucleic acid
DOPE - discrete optimised protein environment
dr - diastereomeric ratio
EC - extracellular
EC50 - half-maximal effective concentration
ee - enantiomeric excess
ELIC - *Erwina chrysanthemi* ligand-gated ion channel
EMDB - electron microscopy databank
FID - free induction decay
FSC - Fourier shell correlation
GA - genetic algorithm
GABA - γ-aminobutyric acid
GABAaRs - γ-aminobutyric acid type a receptor
GAT107 - (3aR,4S,9bS)-4-(4-bromophenyl)-3a,4,5,9b-tetrahydro-3H-cyclopenta[c]quinoline-8-sulfonamide
GLIC - *Gloeobacter violaceuos* ligand-gated ion channel
GluCl - glutamate gated chloride channel
GlyR - glycine receptor
HEK - human embryonic kidney cells
HPLC - high-performance liquid chromatography
HTS - high throughput screen
I - measured current
ID - identifier
Imax - maximum current
K+ - potassium ion
kDa - kilodalton
LY-2087107 - (2-((4-fluorophenyl)amino)-4-methylthiazol-5-yl)(thiophen-3-yl)methanone
M2 / TM2 - transmembrane helix two
MA - membrane associated helix
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>MD</td>
<td>molecular dynamics</td>
</tr>
<tr>
<td>mM</td>
<td>millimole per litre</td>
</tr>
<tr>
<td>Na⁺</td>
<td>sodium ion</td>
</tr>
<tr>
<td>nAChR</td>
<td>nicotinic acetylcholine receptor</td>
</tr>
<tr>
<td>NAM</td>
<td>negative allosteric modulator</td>
</tr>
<tr>
<td>NKCC</td>
<td>sodium, potassium, chloride cotransporter</td>
</tr>
<tr>
<td>nm</td>
<td>nanometre</td>
</tr>
<tr>
<td>NMA</td>
<td>normal mode analysis</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>nuclear Overhauser effect</td>
</tr>
<tr>
<td>NS-1738</td>
<td>1-(5-chloro-2-hydroxyphenyl)-3-(2-chloro-5-(trifluoromethyl)phenyl)urea</td>
</tr>
<tr>
<td>pA</td>
<td>picoampere</td>
</tr>
<tr>
<td>PAM</td>
<td>positive allosteric modulator</td>
</tr>
<tr>
<td>PDB</td>
<td>protein databank</td>
</tr>
<tr>
<td>pKa</td>
<td>proton dissociation coefficient</td>
</tr>
<tr>
<td>pLGIC</td>
<td>pentameric ligand gated ion channel</td>
</tr>
<tr>
<td>PNU-120596</td>
<td>1-(5-chloro-2,4-dimethoxyphenyl)-3-(5-methylisoxazol-3-yl)urea</td>
</tr>
<tr>
<td>QSAR</td>
<td>quantitative structure activity relationship</td>
</tr>
<tr>
<td>RIC-3</td>
<td>resistance to inhibitors of cholinesterase-3</td>
</tr>
<tr>
<td>RMSD</td>
<td>root-mean squared deviation</td>
</tr>
<tr>
<td>RNA</td>
<td>ribose nucleic acid</td>
</tr>
<tr>
<td>ROC</td>
<td>receiver operator characteristic</td>
</tr>
<tr>
<td>SAM</td>
<td>silent allosteric modulator</td>
</tr>
<tr>
<td>SAR</td>
<td>structure activity relationship</td>
</tr>
<tr>
<td>SCCC</td>
<td>segment based cross-correlation coefficient</td>
</tr>
<tr>
<td>SMOC</td>
<td>segment based Manders' overlap coefficient</td>
</tr>
<tr>
<td>SSE</td>
<td>secondary structure element</td>
</tr>
<tr>
<td>T. marmorata</td>
<td>Torpedo marmorata</td>
</tr>
<tr>
<td>TBS-516</td>
<td>4-(5-benzyl-3-(4-bromophenyl)-1H-1,2,4-triazol-1-yl)benzenesulfonamide</td>
</tr>
<tr>
<td>TEM</td>
<td>transmission electron microscopy</td>
</tr>
<tr>
<td>TEVC</td>
<td>two electrode voltage clamp electrophysiology</td>
</tr>
<tr>
<td>TM</td>
<td>transmembrane</td>
</tr>
</tbody>
</table>
TMD - transmembrane domain
Torpedo - Torpedo marmorata
TQS - 4-(naphthalen-1-yl)-3a,4,5,9b-tetrahydro-3H-cyclopenta[c]quinoline-8-sulfonamide
TROSY - transverse relaxation-optimised spectroscopy
UCL - University College London
UFF - universal forcefield
UV - ultraviolet
V_m - resting membrane potential
\( \alpha_y \) - the \( \alpha \) subunit forming the principal interface with the \( \gamma \) subunit
\( \alpha_\delta \) - the \( \alpha \) subunit forming the principal interface with the \( \delta \) subunit
\( \mu M \) - micromole per litre
\( \sigma \) - number of standard deviations from the cryo EM histogram average
Chapter 1

1 Introduction
1.1 Protein-ligand interactions

Proteins realise their respective functions through molecular recognition, a process in which high affinity and highly specific interactions occur to form a functional complex. Complexes can be with other proteins, alternative biomolecules such as DNA or RNA, or with small molecules. Many examples of protein-ligand interactions are known including those determined from the earliest protein crystal structures such as the structural elucidation of myoglobin in complex with oxygen (1) or insulin in complex with zinc (2).

Ligand-gated ion channels such as nicotinic acetylcholine receptors (nAChRs), with which this thesis is primarily concerned, are protein complexes which undergo macromolecular rearrangements on the binding of specific ligands. Present in the central and peripheral nervous systems, one crucial role that ligand-gated ion channels play is in trans-synaptic nerve signalling.

This introduction will lay out the context in which nAChRs exist, before describing functional measures of the receptors which are discussed later in this thesis and some of the background relating to the variety of ligands which are known to bind and modulate these receptors. After this, the experimental techniques of protein structure determination, comparative modelling and small molecule docking are described, which are salient to the research in this thesis.

1.1.1 Nerve signalling

Signals within the central nervous system (CNS) are carried by neurons. These signals are typically carried along the nerve cell as an electrical impulse. Signalling between nerve cells occurs via small gaps, termed synapses (Figure 1.1). Synaptic clefts vary in size from around 20 nm to 500 nm. The way in which information is passed across a synapse can be either electrical or chemical; however, chemical synapses are far more common.

Most nerve cells have a negative resting potential ($V_m$). This means that the inside of the cell has a slightly greater concentration of negative ions than positive ions, typically due to a leaking of $K^+$ along its concentration gradient,
from within the cell to the extracellular fluid. Due to this negative $V_m$, the surface of the nerve cell can either become depolarised, or hyperpolarised by incoming stimuli. When a nerve cell membrane becomes sufficiently depolarised, a reversal of the transmembrane potential can occur. This is known as an action potential and is propagated along the surface of the neuron, transmitting the signal to the other end of the cell (Figure 1.1).

Chemical transmission generally occurs as follows; an action potential arriving at the presynaptic terminal stimulates release of a neurotransmitter from a presynaptic vesicle. Diffusion of the neurotransmitter across the synaptic cleft allows specific binding to occur with receptors on the postsynaptic membrane (Figure 1.1). Binding elicits a response, typically a conformational change of the receptor, which affects membrane permeability to certain ions or binding of intracellular cofactors and so the signal is passed to the next cell in the pathway (3, 4).

Figure 1.1: Graphical representation of cholinergic synaptic transmission. As a nerve impulse arrives at the axon terminal of the presynaptic cell, acetylcholine (ACh) is released into the synaptic cleft from internally stored neurotransmitter vesicles. ACh travels across the synaptic cleft to the postsynaptic cell binding to their receptors and sparking influx of ions and the continuation of the action potential. Figure adapted from Life: The Science of Biology, 7th Edition (4).
1.2 Nicotinic acetylcholine receptors

Nicotinic acetylcholine receptors (nAChRs) are the prototypical member of the super-family of pentameric ligand gated ion channels (pLGICs), which also includes structurally similar receptors for neurotransmitters such as γ-aminobutyric acid (GABA), 5-hydroxytryptamine (5-HT; serotonin) and glycine (5). nAChRs are protein complexes formed of five subunits, with each subunit comprising three subdomains; an extracellular domain made up of two β-sheets in an immunoglobulin like fold, a transmembrane domain of four α-helices (TM1-4) and a cytoplasmic domain of varying size (6) (Figure 1.2). Some other important features of this family of receptors is a conserved disulfide between strands β6 and β7 in the extracellular domain, the interconnecting region is known as the Cys-loop (Figure 1.2, inset right), and a loop between strands β9 and β10, known as the C-loop, which is involved in neurotransmitter binding.

Figure 1.2: Topology of pLGICs. Left) Topology diagram of an nAChR subunit based on the structural data for the α subunit from the T. marmorata nAChR structure (PDBid: 2BG9) and α4 subunit of the α4β2 nAChR structure (PDBid: 5KXI). Solvent exposed α-helices are shown in red, transmembrane helices are yellow, the β-sandwich is delineated with the strands of one face coloured green and those of the other coloured blue. The membrane is shown as a peach block. Diagram drawn using TopDraw from the CCP4 package (8). Right) 3D topology as shown by the 5-HT3 receptor structure (PDBid: 4PIR) for one monomer, with cys-loop inset (left) and as a subunit pair with respect to the rest of the receptor structure and the membrane (right).
Ligands which bind at the neurotransmitter binding site (Figure 1.2), under the C-loop are called orthosteric ligands. Nicotinic acetylcholine receptors are stimulated by the endogenous neurotransmitter acetylcholine at the orthosteric site, as well as a number of natural products including nicotine (Figure 1.3) (7). Upon stimulation, nAChRs selectively allow the passage of cations from the extracellular fluid to the cytosol.

Figure 1.3: Naturally occurring orthosteric ligands of nAChRs. Endogenous neurotransmitter acetylcholine, orthosteric binding natural products choline, (S)-nicotine, epibatidine, ferruginine, cytisine and anatoxin-a.

Ligand gated ion channels such as nAChRs transmit a binding event within the extracellular domain to the opening of a channel in the transmembrane domain ~50 Å away from the binding site. As such, nAChRs are seen as allosteric proteins and their transitions between the extreme conformations of open and closed channels are often interpreted under the Monod-Wyman-Changeux model of allostery (9). In this scheme, an allosteric protein is a protein complex formed of several symmetrically related protomers (five rotationally symmetrical protomers in the case of pLGICs). The complexes have multiple conformational states, for nAChRs there exist at least three (closed, open and desensitized) (Figure 1.4). These states are in equilibrium and the binding of ligands alters the position of the equilibrium, but the conformational change also alters the affinity of the complex for ligand. For example, binding of acetylcholine to an nAChR stabilises the open conformation of the receptor which in turn has a higher affinity for acetylcholine than does the closed conformation (Figure 1.4). Thus, it may
seem that binding of an agonist provokes an allosteric transition, but really this should be seen more as a conformational selection. Although gating is often represented as a concerted process, there may be intermediate states that exist between the extreme conformations and are also in equilibrium with each other.

Figure 1.4: Conformational transitions in allosteric proteins such as nAChRs. Equilibria exist between at least three conformational states (closed, open and desensitized). Agonist binding stabilises the open conformation of the receptor which may interchange with the non-conductive desensitized state.

1.2.1 nAChR nomenclature

17 different nAChR subunits have been discovered in vertebrates (α1-α10, β1-β4, γ, δ and ε) which can co-assemble to form a vast array of receptor subtypes, each with a distinct subunit composition (10). While the majority of both muscle and neuronal nAChRs are heteromeric (i.e. formed from more than one subunit-subtype), a few, such as the α7 subunit have the ability to form functional homomers (11). In nAChRs the orthosteric binding site for acetylcholine is situated at the interface of two subunits in the extracellular domain. The subunit which contributes its C-loop to the neurotransmitter binding site is the the principal or (+) subunit and is always an α subunit. The other half of the interface is called the complimentary or (-) subunit. One of the most studied regions in which nAChRs are expressed is at the neuromuscular junction. These nAChRs are termed muscle-type, whereas those expressed in the CNS are termed neuronal.

Of the 17 subunit sub-types α8 has only been identified in avian species, the other 16 occur in humans (10). The earliest studied nAChRs were purified
from electric organs of the marine ray *Torpedo* and freshwater eel *Electrophorus* (12). These receptors were found to contain four distinct protein subunits and the nomenclature of these subunits stems from their differing molecular masses, based on separation via acrylamide gel. The subunits were named with the Greek letters α, β, γ and δ in order of their increasing molecular masses. Since only the α subunits could be labelled with quaternary ammonium affinity-labelling agents it was concluded that the α subunits comprised the agonist binding site (13). Subsequently through cloning, it was determined that the α subunits of the *Torpedo* nAChR contain two adjacent cysteine residues that were thought to be important in agonist binding (14, 15). Other subunits found to contain analogous cysteines at these positions (192 and 193 in *Torpedo*) have also been classified as ‘α’. However, this convention has led to some anomalous cases where the label (α or non-α) does not indicate that there is functional equivalence between the subunit subtypes. For example, while all heteromeric nAChRs contain at least one α subunit that is able to act as the principal subunit of the agonist binding site, not all α subunits are able to act in this way. The α5 subunit is a case in point, it is not able to form a functional nAChR unless co-assembled with another α subunit. In fact, by sequence identity, it is more similar to the β3 subunit than any other α subunit, but is classified as an α subunit nonetheless (16).

Subtypes of nAChRs are typically referred to by their subunit composition. For example α3β4 refers to a heteropentameric nAChR containing only α3 and β4 subunits, however, the exact subunit stoichiometry may not be known. When the precise subunit stoichiometry is unknown, this is signified by the use of an asterisk, for the above case this would be written α3β4*, denoting an nAChR that contains both α3 and β4 subunits but may also contain other subunit subtypes. In situations where both the subunit composition and stoichiometry is known, the ratios of each subunit in a single pentamer are given as subscript numbers. For the well-established nAChRs at the adult neuro-muscular junction the receptor could be denoted (α1)2β1δε as the receptor is known to co-assemble in a 2:1:1:1 ratio of α1, β1, δ and ε subunits (10).
1.2.2 Subunit stoichiometry

Given that there are 16 different nAChR subunits in humans five of which are required to come together to form a pentameric receptor, one might expect that there could be $16^5$ or 1,048,576 permutations that could make nAChRs. In reality, significantly fewer combinations are observed in vitro and in vivo, one reason for this being that the presence of at least one α subunit is necessitated for a functional receptor.

Some receptor subtypes such as the muscle type receptors ((α1)$_2$β1δε or (α1)$_2$β1γδ) and homomeric receptors like α7 have defined subunit stoichiometry. The α4β2 receptor on the other hand has been shown to co-assemble in multiple stoichiometries (α4)$_2$(β2)$_3$ and (α4)$_3$(β2)$_2$, which show differing pharmacological properties such as high and low sensitivity to acetylcholine (17–19) and calcium permeability (20).
1.2.3 Native nAChR localization

Figure 1.5: nAChR subtype distribution. Top panel; brain section showing nAChR subtype expression in different areas of the rodent CNS. Bottom panel; nAChRs expressed in the visual pathway, superior cervical ganglia and cochlea. Subtype locations were determined by binding, immunoprecipitation/immunopurification, in situ hybridization, single cell PCR and binding studies or functional assays on rat, wild-type or knock-out mice. Figure reproduced from (10) with permission.

The most prolific nAChRs in the human nervous system are thought to be α7 nAChRs and heteromeric receptors comprised of one type of α subunit and one type of β subunit. The most common of these heteromeric receptors is α4β2 (although α4β2 receptors are expressed in at least two pharmacologically distinct stoichiometries). As well as being highly expressed, α7 and α4β2 receptors are also the most widely expressed nAChRs in the nervous system. Figure 1.5, produced in 2009 by Millar and Gotti (10) shows the diversity and distribution of different receptor subtypes in several areas of the CNS, as well as in the visual pathway, superior cervical ganglia and cochlea. It is evident that α7, as well as α4β2 are the most
commonly observed nAChR subtypes in this diagram. More recently, α7 has been found to co-assemble with β2 in both heterologous expression systems and in native mouse and human brain cells (21, 22). These α7β2 receptors could also be present in many of the regions where both α7 and β2 are expressed. Although nAChRs are displayed in Figure 1.1 at the post-synaptic terminal, they are commonly found at the pre-synaptic terminal, where they may play a role in the release of neurotransmitters (23).

1.2.4 Functional measurements of nAChR properties

The primary function of nAChRs is to sense an incoming chemical stimulus (e.g. a neurotransmitter or orthosteric agonist) and respond by opening an ion channel across the cell membrane and allowing the passage of positive ions. This can be measured in a number of ways. Some of the most commonly employed techniques in the field of nAChRs are based on electrophysiology of recombinantly expressed receptors. Electrophysiology is a broad term used to describe the measurement of electric properties of biological tissues and multiple methods exist.

1.2.4.1 Two electrode voltage-clamp electrophysiology

Two-electrode voltage clamp electrophysiology (TEVC) uses two intracellular electrodes. One monitors the membrane potential ($V_m$) while the other injects current to adjust $V_m$ to a desired value (Figure 1.6). The injected current is measured as the membrane current. The use of TEVC is limited to large cells such as giant axons, skeletal muscle cells and Xenopus oocytes due to the necessity for the electrodes to enter the cell. While the alternative technique, patch-clamp electrophysiology is used extensively, it is only applicable to relatively small cells as the larger currents employed by patch-clamping cause a drop in voltage across the recording electrode that cannot be adequately counteracted in large cells such as Xenopus oocytes. This means that TEVC is a commonly used technique, in part because Xenopus oocytes are a very convenient expression system due to the low natural expression of endogenous ion channels and receptors on the membrane (24).
Figure 1.6: Typical TEVC set-up on a Xenopus oocyte. The voltage electrode (electrode 1) is connected to the input of a voltage follower ($A_1$), the output of $A_1$, which is approximately equal to $V_m$, is connected to one input terminal of a clamping amplifier ($A_2$). $A_2$ compares $V_m$ with the voltage command signal ($V_c$). The output voltage of $A_2$ forces a current proportional to the difference between $V_m$ and $V_c$ (it) to flow through to the current electrode (electrode 2) and into the cell. The current passing through electrode 2 is measured as the membrane current as it corrects for the change in $V_m$ that the membrane current would have caused. Reproduced from (24) with permissions.

The outcome of TEVC recording is a trace showing the change in the membrane current across a whole cell over time. These traces give details of the maximum response of a population of receptors to a stimulus (being seen as the height of the peak, or in some cases the area under the peak), and the rate at which they close or become desensitized. These two phases are shown in Figure 1.7. The changes in current measured across the cell membrane are often in the picoampere (pA) range. Due to highly heterogeneous expression and transport of membrane proteins from cell to cell, recorded values are often averages of multiple readings on different cells.
1.2.5 \( \alpha 7^* \) and \( \alpha 4\beta 2^* \) nAChRs

As has been previously mentioned, \( \alpha 7 \) and \( \alpha 4\beta 2 \) containing nAChRs are commonly expressed in the brain. Intriguingly \( \alpha 7^* \) receptors are not only expressed in regions of the CNS that underlie learning and memory (25, 26), but also in the peripheral nervous system (27) and immune cells (28, 29). On the other hand \( \alpha 4\beta 2^* \) receptors are expressed almost exclusively in the brain and are thought to make up almost 90% of the high affinity nicotine binding sites (30), possibly playing the largest role in smoking addiction.

Figure 1.7: Example electrophysiology trace for an ion channel responding to an incoming stimulus by opening the channel. Over time the current increases to a maximum which is recorded as the peak response, and then fades away at a rate characteristic of the desensitization kinetics of the channel.
Figure 1.8: Functional and pharmacological properties of α7 and α4β2 nAChRs. Top, representative electrophysiology traces for α7 (left) and α4β2 (right) nAChR subtypes under a 1 s application of a saturating concentration of ACh. Bottom, comparison of pharmacological and functional properties. MLA = methyllycaconitine, α-BTX = α-bungarotoxin, DHβE = dihydro-β-erythroidine.

The role that nAChRs play in CNS disorders as well as neuronal excitability and plasticity is complex. As can be seen in Figure 1.8, subunit composition appears to control factors such as channel gating kinetics, permeability and ligand specificity. Additionally, localisation within the CNS governs the exact contribution of an nAChR population to a neural network in a spatio-temporal manner. The unique distribution of several nAChR subtypes implicates particular receptors in several neurological disorders. Therefore, nAChRs have been actively sought as targets for therapeutics in areas such as neurodegenerative disorders, neurodevelopmental disorders and chronic pain (31).

Many recent studies have been looking into how these receptors are implicated in illnesses such as Alzheimer’s, schizophrenia and Parkinson’s disease (32–38). One of the components of the histopathology of Alzheimer’s disease is loss of cholinergic neurotransmission, and it has been found that expression of the α7 nAChR subunit decreases with age in some strains of mice (39). The rapid desensitization of these receptors is thought to be a controlling factor in cholinergic signalling (40). Further stimulation of α7 nAChRs is proposed to be a method of improving cognition and a large body
of research has been conducted to identify subtype selective ligands of nAChRs (41–43).

1.2.6 Orthosteric ligands

Orthosteric ligands bind at the site of a receptor’s endogenous neurotransmitter. For pLGICs this is at the interface of two subunits in the extracellular domain (Figure 1.2). The acetylcholine binding site in nAChRs is characterised by a set of aromatic residues which form a box that surrounds the positively charged quaternary ammonium ion of acetylcholine, known as the ‘aromatic box’. A significant driving force of binding is the formation of cation-π interactions between the positive charge on acetylcholine and the π-orbitals of the residues forming the aromatic box. As such, a common feature in orthosteric ligands is the presence of a functional moiety that is positively charged at physiological pH (Figure 1.3). On binding, the C-loop closes on the agonist and increases the affinity of binding by several orders of magnitude, as well as sparking the series of events that leads to channel opening (44).

Alongside naturally occurring orthosteric agonists of nAChRs, orthosteric antagonists have been discovered and play a large role in the determination of whether new ligands act in a competitive or non-competitive manner. Some of the commonly employed orthosteric antagonists are the α4 selective dihydroerythroidine, non-selective mecamylamine and α7 selective methyllycaconitine (Figure 1.9). Antagonists such as these bind to the orthosteric site competitively with agonists but are unable to activate the receptor, hence blocking any channel opening.
1.2.7 Allosteric ligands

In addition to ligands that bind at the orthosteric site, a number of ligands that bind elsewhere and modulate the function of nAChRs have been discovered. These are termed allosteric modulators. The effects of allosteric modulators can take numerous forms. Those which increase the peak response of nAChRs with respect to an orthosteric agonist are termed positive allosteric modulators (PAM), while those which decrease the peak response are termed negative allosteric modulators (NAM) (41). Some allosteric modulators have been discovered which are able to elicit opening of the channel in the absence of an orthosteric agonist and these have been called allosteric agonists. There have also been cases of small-molecules which appear to have no effect when applied to nAChRs alone or in combination with an agonist, yet they can block the effect of other allosteric modulators. It is therefore deduced that they are silently binding to an allosteric site and have thus been called silent allosteric modulators (SAMs) (45). For receptors such as α7 nAChRs which desensitize rapidly, PAMs are commonly subclassified as type I and type II. Wherein, type I PAMs increase the peak response of the receptor but have no effect on the rate of desensitization and type II PAMs both increase the peak response and slow the rate of desensitization of a receptor (41).

PAMs of nAChRs have garnered much attention due to the possible involvement of these receptors in numerous pathologies. In particular PAMs of α4β2 and α7 nAChRs have seen a great deal of research within the pharmaceutical sector. A number of α4β2 PAMs are known, many are biaryl substituted five-membered heteroaromatics (Figure 1.10).
Figure 1.10: Examples of α4β2 PAMs from Abbott Laboratories (top five compounds) and NeuroSearch (lower three compounds).

Promise of therapeutic benefit of α7 nAChRs in disease states such as Alzheimer’s, schizophrenia and Parkinson’s (32–38) has seen significant investments by companies such as Abbott Laboratories (46), Pfizer (47), NeuroSearch (48), AstraZeneca (49), Eli Lilly (50), and Janssen Pharmaceutica (51) who have discovered some of the most well characterised families of type I and type II PAMs of α7 nAChRs (Figure 1.11). While there is a significant diversity in the chemical structures of these α7 allosteric modulators, it has been found that very small changes in chemical structure such as the number of methyl groups on a single aromatic ring can elicit five distinct pharmacological effects on compounds derived from the structure of TQS (Figure 1.11) (45).
1.2.7.1 Potency and efficacy – Dose-response curves

A dose-response curve describes the change in response of a receptor to varying doses of a stimulus. As applied with respect to PAMs and nAChRs, this would consist of plotting peak responses of an nAChR population to an invariant concentration of an orthosteric agonist at a variety of PAM concentrations, e.g. measured by TEVC electrophysiology (from the same oocyte), against the logarithm of the concentrations of the PAM that causes the change in peak response (Figure 1.12). The resultant graphs are typically sigmoidal with a plateau for the maximum response (e.g. \(I_{\text{max}}\), for maximum measured current by TEVC) at saturating concentrations of PAM. When responses are measured over an appropriate range of concentration values, a steep increase in response is seen around the middle of the sigmoidal dose-response curve for active PAMs. The inflection point of the steep increase, which corresponds to the half-maximal response also gives the half-maximal effective concentration (EC\(_{50}\)) i.e. the concentration of PAM required to give half the maximum response of a saturating concentration. EC\(_{50}\) values are frequently used to describe the potency of a compound and can be derived from fitting the Hill equation to a dose-response curve, for example of the form (Equation 1.1).
Where \( I \) is the current and \( I_{\text{max}} \) is the maximum current, \( EC_{50} \) is the PAM concentration ([PAM]) that evokes half of the maximum current and \( n_H \) is the Hill coefficient.

Two PAMs can cause responses of different magnitudes yet have the same potency as measured by \( EC_{50} \) (Figure 1.12). However, the change in the size of the response for the same concentration of PAM would denote a difference in efficacy. For the two PAMs in Figure 1.12, the PAM that causes the black trace has a greater efficacy than the PAM that causes the grey trace, despite having the same potency.

\[
\frac{I}{I_{\text{max}}} = \frac{1}{1+10^{(EC_{50} - [\text{PAM}])n_H}}
\]

A further useful property that can be extracted from dose-response curves is the normalised response of a receptor to a stimulus. If one plots varying agonist concentrations against response in the presence of invariant PAM concentrations, the maximum response can be normalised against, for example, the response in the absence of PAM to demonstrate differences in efficacy (the size of the change in response). For the traces in Figure 1.12, if the grey trace were instead, the response of an nAChR population to acetylcholine alone and the black trace the response to acetylcholine in the
presence of a PAM, as the maximum response of the black trace is twice the maximum response of the grey trace, the normalised response of the black trace would be 2.

1.2.7.2 Where do PAMs bind

There is a great body of evidence suggesting a number of locations at which these allosteric modulators may bind to α7 nAChRs. However, the molecular footprint of interaction is ultimately determined by high resolution structures of the ligand of interest bound by the protein complex. To date, there are no definitive structures that clarify the binding sites of these ligands. Despite this, likely locations, from studies of mutagenesis, crystallography of homologues and photo-crosslinking in homologues, include:

- Within the extracellular vestibule (52).
- At the interface of the extracellular domain and transmembrane domain (53).
- Within the transmembrane domain for example;
  - Within a single subunit (54, 55).
  - At the interface of two subunits (53).
- At a number of these sites simultaneously (56).

Research into allosteric modulators of nAChRs is still an active area, however, improving knowledge of how and where exactly these ligands bind will greatly improve our understanding of their action. Perhaps an understanding of the mechanisms of allosteric modulation will allow the better design of allosteric modulators in future. Techniques that can be used to structurally investigate or predict the complexes of small-molecules and proteins are discussed below.

1.3 Protein structure determination

Playing a central role in understanding how a protein functions, is knowing its structure. Some of the most commonly used high resolution protein structural determination techniques include X-ray crystallography, electron cryomicroscopy (cryo EM) and protein nuclear magnetic resonance (NMR).
These methods can permit the construction of protein models at an atomic level of detail and their contributions to the field of structural biology are so great that each technique has been honoured with at least one Nobel prize for their use in protein structure determination (57–59).

Complimentary to these techniques are those which give lower resolution information such as small angle X-ray scattering, circular dichroism, cross-linking mass spectrometry, hydrogen-deuterium exchange mass spectrometry and negative stain electron microscopy. Such lower resolution data can be combined in integrative modelling to produce atomic models of proteins or protein complexes (60).

1.3.1 X-ray crystallography

Proteins are too small to be imaged with light microscopes because the wavelength of visible light (400 – 700 nm) is far greater than the length of single bonds in the protein (~0.15 nm) meaning that diffraction of light is not possible. However, it is possible to use a different source of electromagnetic radiation with a more suitable wavelength, for example X-rays.

When a monochromatic (all of one wavelength) X-ray source is fired at an ordered crystal of proteins, a unique diffraction pattern is created (Figure 1.13). This occurs due to the interaction of the X-rays with the crystalline lattice. The pattern is spaced in a way that is inversely proportional to the spacing of the unit cell in the crystal and each spot (reflection) will have an intensity associated with it that is due to the amount of X-ray that was reflected by the point in the lattice that made it. By rotating the crystal in the beam of X-rays, diffraction patterns can be collected from many different angles. By measuring the positions and intensities of the spots it is possible to work back to the structure of the molecule that made it and reconstruct the electron density that created the diffraction pattern. The distance of the reflections farthest from the centre of the diffraction pattern (Figure 1.13) dictate the resolution of the collected data. By measuring the angle at which these farthest reflections deviate from the X-ray beam, the resolution can be determined by solving the Bragg equation (61) for \( d \) (Equation 1.2):
Where $d$ is the lattice spacing (or resolution), $\lambda$ is the X-ray wavelength and $\theta$ is the angle of reflection.

This technique requires any protein whose structure is to be solved, to be in a crystalline form of an incredible regularity. Forming protein crystals is typically the limiting step in structure determination by X-ray crystallography. It is often difficult to express and purify proteins to high enough concentrations and purity to facilitate crystallization. Furthermore, many proteins contain flexible or intrinsically disordered regions which present a larger entropic barrier to crystallization as intramolecular order must be introduced to fit the proteins into a regular repeating array. Although an increasing number of relatively large structures have been solved by X-ray crystallography (63), crystallographic determination of protein structure has historically been limited to smaller, soluble proteins.

There has been a recent increase in the number of membrane protein structures elucidated by X-ray crystallography (64). A common critique of X-ray crystal structures is that crystal packing, that is close contacts with other
molecules in the unit cell, and non-native like conditions during purification and crystallization protocols can manipulate protein structures so that they are no longer in physiologically relevant states. This is especially true of membrane proteins and protein complexes of which nAChRs, and pLGICs as a whole, fall into both categories. When it comes to membrane proteins it is impossible for a protein to be simultaneously in a 3D crystal structure and in a native membrane environment. In preparation for crystallography membrane proteins are solubilized using detergents whose effects on membrane region structure are not known and it is thought that this could, in some cases be detrimental to the fidelity that the resolved structure shows in comparison to the protein in its native environment.

1.3.2 Protein NMR

Spin is an intrinsic quantum property of atomic nuclei. Nuclei which have half integer spin are magnetically active and the energy of their spin states can be split by a magnetic field e.g. $^1\text{H}$, $^{13}\text{C}$, $^{15}\text{N}$ and $^{31}\text{P}$ all spin = 0.5. Since the aforementioned nuclei have spin = 0.5, when exposed to a strong magnetic field they have 2 spin states, one of a lower energy (paired with the field), one of a higher energy (paired against the field). Pulses of radiofrequency electromagnetic radiation can be used to transfer nuclei between spin states.

A typical NMR experiment might excite all nuclei to the higher-level spin state and then record the relaxation of the spins back to equilibrium. The relaxation is recorded as a decaying current in the receiver coil. The recorded free induction decay (FID) can be manipulated with a Fourier transform, which decomposes an oscillating signal into its constituent frequencies, to give 1D peaks corresponding to magnetically active spins (nuclei). These appear at a given chemical shift due to the unique environment surrounding the nuclei in the magnetic field which influences the rate at which signals from spins in these unique environments decompose and therefore the frequency that they contribute to the FID. More complicated sequences of radiofrequency pulses can be used to obtain 2D, 3D or higher dimensional data.

Nuclei which interact through bonds to influence each other’s relaxation properties are said to be coupled with each other. Measuring these couplings
gives details of the covalent structure of the protein and can be used to introduce restraints in modelling. Further detail can be elucidated from Nuclear Overhauser Effect (NOE) couplings which occur through space over short distances, giving details of tertiary structure arrangement and torsion angle restraints can be applied by comparison of the spectra to empirically measured chemical shifts.

An important comparison with X-ray crystallography, is that NMR allows the determination of protein structure in solution. This means that soluble proteins are often in a much more native-like environment and it also allows the capture of dynamic details of protein motion that are not possible in a crystal. However, there are two very significant limitations of NMR spectroscopy. The first is that apart from $^1\text{H}$, the other magnetically active nuclei are not highly abundant in nature and so proteins must first be labelled before they can be studied by NMR. Although $^1\text{H}$ accounts for $>99\%$ of naturally occurring hydrogen, due to its prolific presence in protein structures it too must often be substituted for the less abundant non-magnetically active isotope $^2\text{H}$ (deuterium) in order to reduce the complexity of spectra. The second limitation is that to obtain a useful solution state NMR spectrum the molecules of interest must be tumbling rapidly in the solvent. The rotational correlation time (time taken for a molecule to rotate through 1 radian) is proportional to the size of the molecule. This means that there is a size limitation for proteins that can be analysed with NMR. Hence, the majority of protein NMR is carried out on small polymers of < 50 kDa (65). However, techniques such as transverse relaxation-optimised spectroscopy (TROSY) (66) and more recently FROSTY/SedNMR (67, 68) have been used to determine structures of protein complexes up to 1 MDa (69), significantly increasing the size limit for protein systems which can be structurally studied by NMR.

1.3.3 Electron cryomicroscopy
Similarly to electromagnetic radiation, electrons are also able to behave in a wave-like manner. Electron microscopy relies on such properties of electrons and so the path of a beam of electrons can be focussed using
electromagnetic and/or electrostatic lenses in much the same way as glass lenses are used to focus visible light.

The speed that the electrons are accelerated to (or the amount of energy put into them), is inversely proportional to their wavelength, i.e. the faster the electron is travelling, the shorter its wavelength. The shorter the wavelength of the electron, the higher its resolving power. The imaging of biological samples for determination of protein structure typically relies on transmission electron microscopy (TEM). In TEM the electron beam passes through a very thin and thus semi-transparent sample. Those electrons that make it through and are elastically scattered (their path changes direction without losing any energy) carry rich information about the structure of the sample. This information is captured as a two dimensional projection of the matter that they interact with on a charge-coupled device (CCD) or more often these days on a direct electron detector after removing inelastically scattered electrons from the sample with more lenses (Figure 1.14).
Figure 1.14: Simplified schematic of a TEM set-up. Electrons from an electron source are focused into a beam using a condenser lens. After the beam passes through the object of study, the beam is focused via an objective lens, through an aperture which removes inelastically scattered electrons. Finally a projector lens projects the image onto a sensor such as a CCD or direct electron detector. Reproduced from (65) with permission (Original article can be found at the following link; https://pubs.acs.org/doi/abs/10.1021/cr100353t, further permissions relating to this content should be directed to ACS at support@services.acs.org)

An important aspect of EM that sets it apart from X-ray crystallography is that electrons have a much stronger interaction with matter than do X-rays. Whereas X-rays require millions of molecules ordered in a regular 3D array to be diffracted, electrons can be diffracted by individual molecules or 2D crystalline arrays. This means that samples which are not amenable to crystallization can be studied. Commonly, protein structure is determined from a solution of differently oriented single molecules or complexes. This means that the two dimensional projections cover a variety of angles of the particles being imaged. The raw images of individual particles are very noisy, but the combination of many images improves the signal to noise ratio for each projection angle. When the angle of projection is determined (for example by the comparison with computed projections of a low resolution model of the molecule), the averages of the two dimensional images of each
projection angle can be combined and back-projected to produce a three dimensional reconstruction (Figure 1.15).

Figure 1.15: Basic concepts of cryo-EM structure determination. (A) The projection-slice theorem states that the 2D projection of a 3D object in real-space (left column) is equivalent to taking a central 2D slice out of the 3D Fourier transform of that object (right column). The real-space projection direction (left; dashed red arrows) is perpendicular to the slice (right; red frame). (B–E) Many experimental 2D projections can be combined in a 3D reconstruction through an iterative process called “projection matching”. To determine the relative orientations of all experimental projections one first calculates reference projections of a 3D object in all directions (B). Then, one compares each experimental projection with all reference projections to find the best match of a given similarity measure (C). This orients all experimental projections relative to the 3D structure (D). The projection-slice theorem then implies that the 3D reconstruction can be calculated by positioning many 2D slices (the 2D Fourier transforms of all experimental projections) into the 3D transform (E) and calculating an inverse transform. Iterating steps (B–E) will gradually improve the orientations, and hence the resolution of the reconstruction. Figure and legend taken from (70) with permissions.

Crystal contacts are not a problem in the interpretation of structural data arising from single particle cryo EM studies. However, issues of the physiological relevance of structures can still persist as proteins are typically out of their native environment in single particle studies. Nevertheless, the ability to reconstruct 3D structures from individually dispersed proteins permits
the determination of significantly more complex systems such as large proteins and protein complexes. Furthermore, since individual particles are picked and averaged together, incongruent data can be eliminated, allowing greater levels of sample heterogeneity. The downside of strong interactions between electrons and the sample is that energy is easily transferred from the incoming electron beam to the sample that is being imaged. Inelastic scattering of electrons causes sample damage through ionization, rearrangements of bonds, formation of free radicals etc. some features of biological samples are affected by doses as low as $10 \, e^-/\text{Å}^2$ (71). To reduce radiation damage, samples are imaged in defocus and at cryo-temperatures, in the 4-77 K range (71, 72).

For many years, EM typically suffered from significantly lower resolution than X-ray crystallography, however there has been a recent “resolution revolution” with the advent of new technologies (73). Direct electron detectors decrease the levels of noise during detection and capture a greater proportion of the electrons transmitted through the sample. New algorithms for image processing have aided better binning of data to remove spurious images. Now, cryo EM structures are commonly determined at resolutions better than 4 Å.

1.3.3.1 Resolution estimation in cryo EM

In cryo EM, map resolution is frequently estimated by taking the Fourier Shell Correlation (FSC, Figure 1.16) between two independently constructed half maps, each containing half of the data of the full map (74). FSC takes the Fourier transform of the two half maps and calculates the correlation between them at increasing ‘Fourier shells’ i.e. increasing spheres of resolution. The centre of the Fourier transform represents the amplitude of the longest electron wavelengths in the map, while the edge represent the shortest (and therefore highest resolution) up until the length of two voxels (three dimensional pixels) which is the shortest wavelength that can be represented in the map, i.e. one maximum and one minimum. This is known as the Nyquist frequency, for example, in a map with voxels of 1 Å$^3$ the Nyquist frequency would be 2 Å.
Typically, the FSC curve of two half maps will fall to zero and remain there until the highest spatial frequency (Figure 1.16), there is no longer any resolution of features in the map beyond the point at which zero is reached and any signal is swamped by noise. A number of different FSC cut-offs have been used to determine a nominal resolution from FSC curves. The most commonly used are FSC = 0.5 (i.e. when the Fourier shell correlation between maps falls below half), FSC = 0.333 and FSC = 0.143, known as the ‘gold-standard’ cut-off (75). The latter FSC values are determined as when the FSC between the half map and theoretical full map containing twice the amount of data is equal to 0.5 (FSC = 0.333) and when the FSC of the full map and a perfect reference map with no noise is equal to 0.5 (FSC = 0.143) (Table 1.1) (75).

Table 1.1: FSC cut-offs for different criteria from (75).

<table>
<thead>
<tr>
<th>FSC</th>
<th>FSC_{Full}</th>
<th>C_{ref}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.67</td>
<td>0.82</td>
</tr>
<tr>
<td>0.333</td>
<td>0.5</td>
<td>0.71</td>
</tr>
<tr>
<td>0.143</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

FSC is the Fourier shell correlation between two half maps, FSC_{Full} is the Fourier shell correlation with a half map and a theoretical full map containing twice the amount of data. C_{ref} is the correlation between a half map and a perfect reference map containing zero noise.
Figure 1.16: Example Fourier shell correlation curve. FSC starts at one, and falls sharply to zero, after which FSC remains at zero. The resolution of the map can be estimated by the spatial frequency at which FSC is equal to 0.5 (dashed line), 0.333 (dash-dot line) or 0.143 (dash-dot-dot line) depending on the criterion chosen for resolution estimation.

1.3.3.2 Fitting methods

Increasingly, structures of proteins solved by cryo EM are reaching atomic resolutions (< 3.5-4.0 Å) and many of the well-defined and long used crystallographic refinement and model building programs for X-ray structures have been optimised for use with EM data. These include programs such as REFMAC (76), Coot (77), Phenix (78) and Rosetta (79). While these programs, are extremely useful at high resolutions, the majority of EM data still yields structures worse than 4.0 Å resolution and owing to anisotropic resolution throughout the map, frequently contain regions of lower resolution data.

Factors such as the availability of subunit or homologous atomic models from independent studies; the number and extent of the conformational differences between those models and the resolved density map; and the local resolution of the map influence the choice of methods used to fit and refine atomistic models in EM density maps. The use of predefined models or comparative models can help to reduce ambiguity, such as in side-chain register; allow more complete models by providing a template in regions with little or poorly defined density; and reduce errors arising from manual fitting such as in back
bone dihedral angles, or sidechain stereochemistry (80). However, they can also contribute to model bias.

At intermediate and lower resolution ranges (3-5 Å and > 5 Å) flexible fitting methods are useful. The basis of flexible fitting is to use information from a density map to guide the conformational change of a model such that the fit with the map is improved, while the geometric properties of the model are retained. One of the earliest methods of flexible fitting and refinement, Flex-EM (81) applies simulated annealing molecular dynamics (MD) and conjugate-gradients minimization (CG) to a number of rigidly held sub-regions of a protein structure (rigid bodies). This approach has a large radius of convergence and so is able to model global motions, for example conformational changes in the protein folding chaperone GroEL (81).

More recently, Flex-EM has been optimised for use at higher resolutions and it has been shown that the iterative reduction in the extent of the rigid bodies from domains to secondary structure elements even down to individual side-chains can yield excellent results for fitting (82). Aside from Flex-EM many other flexible fitting software packages have been developed which rely on different fitting methods such as Monte-Carlo simulations (Rosetta (83)), molecular dynamics (MDFF (84)), elastic network models (DireX (85)), normal mode analysis (NMFF-EM (86), iMODFIT (87)) or a combination of the above (Monte-Carlo and elastic network models; First/FRODA (88)).

1.4 Comparative modelling

In cases where an experimental structure of a protein is not known, comparative modelling can be used to predict a protein structure using that of a homologous protein. Comparative modelling maps a query protein’s sequence on to the coordinates of a known protein structure in order to predict the 3D conformation of the query protein.

Three key steps are involved:

1. Determining an appropriate template structure on which to base the model of the query protein.
II. Performing an alignment between the template and query sequence.

III. Creating a model of the query sequence based on the template structure.

Figure 1.17: Example comparative modelling workflow. From a target protein sequence, structural templates are identified and selected, an alignment is carried out between the target sequence and template structure (structural, e.g. from secondary structure prediction and sequence alignments contribute to this). Models of the target are built and refined and then evaluated. If the evaluation suggests that model quality is good it can be kept, if there are areas for improvement, then the first port of call is to alter the alignment between target and template, if this cannot improve the model, one might consider finding a different template structure.

Protein structure is significantly better conserved than primary sequence (89). Similar structures can be observed at low sequence identities (<20%) and it has been found that comparative modelling can be accurate when sequence identities between the query and the template exceed 30-40% (90, 91). Indeed, structural variations in protein models at 40% sequence identity and above are found to be comparable to variations observed in structures of the
same protein determined by different techniques, or in different environments (90). An example comparative modelling workflow can be seen in Figure 1.17.

1.4.1 Template selection
Template selection can clearly have an important impact on the result of modelling. There are a number of factors to take into account when choosing a template model. Sequence identity – the template with the highest sequence identity to the query is commonly seen as the best. Environment – templates whose chemical environment match that of the query structure most closely will likely provide the best starting point (e.g. similar ligands bound, similar membrane environments, similar pH etc.). Template model quality – factors such as the resolution of the template for models derived from crystallography or cryo EM, or number of restraints per residue for models derived from NMR (92).

1.4.2 Sequence alignment
The quality of the sequence alignment between a template and the query can have a large impact on the quality of the model produced. Misalignments can generate large errors. Of course, the higher the sequence identity between the query and the template, the simpler it is to generate an accurate alignment (at 100% identity, it becomes a one to one mapping of residues with no insertions or deletions). However, for everything else, alignments can be generated in a number of ways.

Pair-wise alignment can be used to align two sequences with algorithms such as Needleman-Wunsch (93); examples include Needle and Stretcher both available in the EMBOSS package (94). This is appropriate when a clear choice of template structure can be made and the sequence identity is high. When there are a number of templates and sequence identities are lower (<30%), multiple sequence alignments can be a useful method to both aid in choosing the best template and improving the quality of the alignment. Traditionally, sequence based alignment tools such as BLAST (95) have been extensively used. However, BLAST gives equal weight to both conserved and
variable regions within a protein family and so signal is lost where distant homologues are concerned (96). Commonly, profile based alignments are now used. *PSI-BLAST* (97) and *SALIGN* (98) were early examples of profile based alignment programs which utilise evolutionary information from homologous proteins to improve the alignment using dynamic programming. The current state-of-the-art is to use of hidden Markov models (HMMs) which are better able to use information relating to insertions and deletions in the multiple sequence alignment. Some commonly used examples include *HMMER* (99) and *HHblits* (100).

1.4.3 Model building

One of the earliest developed comparative modelling methods developed was MODELLER (101) and this still remains one of the most commonly used methods today. There are several other popular protein modelling packages, including SWISS-MODEL(102) and Rosetta (103).

Briefly, model building can consist of steps such as:

I. Using the sequence alignment as a guide, the coordinates of structurally conserved regions of the query are mapped onto those of the template Cα positions and the basic backbone of the model filled in.

II. Deletions and insertions are dealt with:
   i) For a deletion, residues are removed from the model and the ensuing gap is closed by the formation of a new peptide bond
   ii) For an insertion, loops can be modelled by more than one approach, conformational modelling searches conformational space and tries to satisfy a given energy function which may take into account constraints relating to steric or geometry. Alternatively, loop modelling can be performed by searching protein fragment libraries and choosing segments that provide a good fit for the specific part of the backbone.

III. Side-chains are modelled in. This step will typically use information on side chain rotamers (dihedral angles of bonds in sidechains longer than alanine) in the template where sequence similarity is high. Other
methods may utilise side chain rotamer libraries such as the Dunbrack library (104) to choose statistically likely sidechain conformations.

IV. Following this, a minimization is typically carried out to alleviate unfavourable clashes.

Thus, molecular models of proteins with little or no associated structural data can be built.

1.4.4 Model evaluation

Protein structure prediction methods commonly create a large number of alternative models from which the most ‘native-like’ must be chosen. This is the task of model evaluation and is often informed by a scoring function. Scoring functions for protein model evaluation are based on the thermodynamic hypothesis which operates under the assumption that the native structure of a protein will be the state of lowest free energy under physiological conditions (105). Many scoring functions can be grouped into two classes:

- Physical energy functions; a true energy function that describes physical interactions in proteins, parameterized from the fitting of experimental data and/or the quantum chemical calculations for example the CHARMM forcefield (106).
- Knowledge-based statistical potentials; energy functions originating from known structures of proteins. These can be formed of distance-dependent/ independent pairwise potentials of mean force, or alternatively can be derived from structural features such as solvent accessibility, or internal coordinates (107).

Combinations of independent statistical potentials often perform better than any individual potential alone, some of the most successfully implemented statistical potentials for protein model quality assessment are those which have stemmed from ProQ (108) and QMEAN (107). The advantages of statistical potentials is that they are fast and simple to generate, as such they are typically the scoring function of choice in comparative modelling pipelines such as MODELLER (DOPE (109)) and Swiss-model (QMEAN (110)). One
drawback however, is that the ability of a statistical potential to correctly discriminate between good and bad models is dependent on the models on which the potential is trained. Indeed, most statistical potentials for protein quality assessment are trained on soluble proteins. Therefore, historically have performed poorly on membrane proteins, whose physicochemical environment induces many interactions which would be seen as unfavourable in a globular structure. One attempt to overcome these problems has been the application of separate potentials for membrane and solvent exposed regions of an integral membrane protein (QMEANBrane, (111)).

1.5 Small molecule docking

As mentioned previously, the ultimate description of a protein-ligand interaction is the determination of a ligand-bound protein structure at high resolution, for example via X-ray crystallography. For the vast majority of protein ligands, such structures are not available with their binding partners. Approximately 105 000 PDB entries exist in complex with 25 000 unique chemical ligands (https://www.rcsb.org/pdb/, accessed 09/05/2018 (112)), yet there are millions of possible binding partners for proteins. Despite this, understanding how ligands and proteins interact is key to designing new ligands, for example as drug candidates or as tool compounds for investigating protein function.

Fortunately, docking is a tractable method for the prediction of protein-ligand complexes, from the three dimensional structures of a protein (this can be an experimentally determined structure or a comparative model) and a ligand. Indeed, docking is commonly employed to predict binding modes of drug candidates, for virtual screening or to predict the binding location of a compound (113–116).

Docking studies are comprised of two main modules, a search algorithm, and a scoring function. The search algorithm must generate ligand-receptor conformations, optimally including the true binding mode. The scoring function should then be able to distinguish between the true binding mode and others generated by the search algorithm. An ideal search algorithm would
explore every possible binding mode between the two molecules, however this is rarely possible due to the size of the search space (117). Many different docking programs exist for small molecule docking.

### 1.5.1.1 Search algorithms

These programs typically use one of four types of search algorithm: systematic searching, shape matching, stochastic algorithms, or molecular dynamics (MD) simulations. Shape matching is one of the simplest, and earliest implemented search algorithms used, the algorithm tries to optimise shape complementarity between the protein and ligand surfaces in the binding site (118) (Figure 1.18).

![Figure 1.18: Rigid docking approach. The rigid docking approach initially used in DOCK (118) first fills the binding site with overlapping spheres, matches the atoms of the molecule to the spheres and then orients the molecule into the binding site. Example molecule is the anti-cancer drug imatinib.](image)

Systematic search algorithms are more commonly used. With systematic searching it is possible to generate every protein-ligand conformation and these can be subdivided into three types. Exhaustive searching (e.g. Glide (119), FRED (120)), attempts to create a flexible-ligand docking by rotating all possible bonds in the ligand at specified intervals, however, with large numbers of rotatable bonds this results in huge numbers of conformations and chemical and geometrical constraints are required in order to filter them.

Chapter 1
Conformational ensembles (e.g. FLOG (121), MS-DOCK (122)), these pre-generate an ensemble of ligand conformations and then perform rigid docking. The ligand conformations obtained after each run are then ranked with respect to their binding free energies. Fragment-based searches (e.g. FlexX (123), DOCK (124)) divide the ligand into smaller ‘fragments’, the fragments are individually docked and then subsequently linked covalently in silico (Figure 1.19).

Figure 1.19: Schematic representation of fragment based docking. Small fragment like parts of a molecule are docked into an active site, and then linked in silico. Example molecule is the anti-cancer drug imatinib.

Stochastic methods of ligand pose prediction rely on making random changes to the ligand in conformational space and can be classified under four types. Monte-Carlo (MC) simulations (e.g. AutoDock (125)) create random moves in the system and either accept or reject the perturbation using a Boltzmann probability function (117). Since Monte-Carlo methods explore random conformations this decreases the likelihood of becoming trapped in a local minimum, aiding the process of finding the global, lowest energy binding conformation. Genetic algorithms (GA) (e.g. GOLD (126), DARWIN (127)) search conformational space by following the principles of biological competition and population dynamics. They apply genetic operators (mutation, crossover and migration) to the population of docking solutions to
obtain the final population. A random population is generated initially and their degrees of freedom are encoded as ‘genes’ to which the genetic operators are applied in order to create a new population. The new population is scored, and only the highest scoring (fittest) advance to the next round (Figure 1.20). Tabu search methods (e.g. PSI-DOCK (128)) follow an iterative procedure. Starting with a random conformation, random changes are made and the new states ranked. The highest ranked solution is then taken as the current solution and the process repeated. The likelihood of approval of a new solution is governed partially by the conformational space already visited, the random change will be rejected if its RMSD with the current and previous solution is less than a specific cut-off. Particle swarm optimisation (PSO) (e.g. SODOCK (129)), like genetic algorithms, start with a large population of random solutions, however, instead of applying genetic operators, swarm intelligence is used. Each ligand’s motion through the binding cavity is influenced by its best known local position and guided towards its best value by information about the best positions of its neighbours.

Figure 1.20: Representation of a genetic docking algorithm. An initial library of ligand poses is docked and the top poses selected. These undergo crossover and mutation to stochastically change the library and the newly generated library is redocked, the best poses selected and the process repeats for a predefined number of generations. Figure adapted from (130).

1.5.1.2 Scoring functions

Similarly, there are many different types of scoring functions for ranking the predicted poses such as force-field based, knowledge based and empirically based scoring functions. Force field scoring functions, such as AMBER (131), typically try to estimate physicochemical atomic interactions such as van der
Waals and electrostatic interactions, as well as, bond bending, stretching, torsion and hydrogen bond energies. The parameters are derived from experimental data and \textit{ab initio} calculations but often suffer from significant errors. Knowledge based scoring functions, such as ITScore (132), are collected from 3D, ligand-target complexes, they are typically very good at ranking binding modes of complexes similar to those used in parameterization, but suffer from poor prediction power for those that are dissimilar. Empirical scoring functions, such as ChemScore (133) usually describe a small number of energy terms such as hydrogen bonds, ionic interactions and lipophilic protein-ligand contact surface. These terms are calibrated using protein-ligand complexes with known structures and binding affinities. A more comprehensive review of prediction algorithms and scoring functions was written by Bello et al. (134).

1.6 Summary and aims

In summary, nAChRs are an interesting subject of study owing to their importance within the central and peripheral nervous systems, as well as their implication in pathological conditions. Many questions remain to be answered about how this diverse family of receptors can be modulated, including by small allosteric ligands.

This thesis aims to use comparative modelling to produce models of nAChRs by pulling together published structural data from crystallography, electron microscopy and NMR studies of protein structure. Small-molecule docking is used to predict the interaction of a number of allosteric modulators on the \( \alpha_7 \) nAChR and the predicted binding modes used to rationalise functional data collected by our collaborators in the Millar lab. Small-molecule chemical tools are developed to test the binding mode predictions and further models of binding generated from the docking predictions are found to be useful in discovering new \( \alpha_7 \) nAChR allosteric modulators.
1.7 References


Bras 30(11):736.


Chapter 1


131. Kollman PA (1996) AMBER.


Chapter 2

2 A revised structural model of the α7 nicotinic acetylcholine receptor

Some of the work in this chapter has been published in the following article:

2.1 Introduction

Understanding the structure of nAChRs and how subtle changes in different subunits affects specificity for different ligands is an interesting and important research area. Some of the interactions that are least well characterised, are those of small-molecule ligands with the transmembrane domain (1–4). Previous studies have used the *T. marmorata* nAChR structure (5) as a template for α7 nAChRs and subsequently docked in ligands of interest to predict where they bind (1, 2, 4). However, the *T. marmorata* nAChR structure is thought to contain an error in the register of the transmembrane domain (6) which may raise concerns about the accuracy of these interpretations.

The development of α7 nAChR structural models with correct geometries in the transmembrane domain is of utmost importance in eliciting further details of these types of transmembrane interaction. Currently, there is no structural data for the transmembrane domain of an assembled α7 nAChR pentamer. Obtaining a structural model focusing on this region of the α7 nAChR is the primary aim of work described in this chapter.

In order to develop such a model, one could take a number of approaches. The most recent critical assessment of protein structure prediction competition indicates that comparative modelling typically outperforms *ab initio* structure prediction (7), where ‘good’ template model information is available. With a wealth of pLGIC structural data available and the multi-subunit, multi-domain structure being far larger than those typically modelled *ab initio*, comparative modelling is the method of choice for modelling α7 nAChRs and has been used previously (8, 9).

2.1.1 Template models for α7 nAChR

Membrane proteins like nAChRs are notoriously difficult to work with in structure determination due to problems such as low expression levels in recombinant systems and difficulties in purification due to the low stability and hydrophobic nature of the membrane spanning regions. In spite of these issues, there has been growing success in the determination of pLGIC structures in recent years owing to improvements in the techniques for
crystallization of membrane proteins (10) and the advent of cryo EM as a powerful tool for interrogating membrane bound protein complexes. However, there is still no full length structural model of the α7 nAChR pentamer to date. There are still relatively few structural homologues for pLGICs so assessing all of the available structures is worthwhile to choose the best template. The structures which have played a significant role in understanding nAChRs and may be useful as templates for an α7 nAChR model are discussed here.

2.1.1.1 T. marmorata nAChR

The nAChR from the electric organ of the marine ray T. marmorata (Figure 2.1, middle) is expressed in such high density that it can form a 2D crystalline array on the membrane surface (Figure 2.1, top). This property was utilised in order to make structural determination of this receptor more tractable and as a result, in 1985 it was the first neurotransmitter gated ion channel for which structural information was available (11). Twenty years later, this was the first pLGIC to have its structure determined at atomic resolution (4.0 Å, Figure 2.1, bottom) and furthermore, was embedded in its native membrane environment (5). Consequently the T. marmorata nAChR has been seen as the prototypical member of the family of pLGICs (12).

Recently, advances have been made in sample preparation techniques for the application of membrane proteins in X-ray crystallography and electron cryomicroscopy (cryo EM). This has allowed the high resolution structures and corresponding models to be generated for several other members of the pLGIC superfamily such as the prokaryotic ion channels Erwinia chrysanthemi ligand-gated ion channel (ELIC) (13) and Gloeobacter violaceus ligand-gated ion channel (GLIC) (14, 15), the glutamate gated chloride channel (GluCl) (16), 5-HT type 3 (5-HT₃) receptor (17), γ-aminobutyric acid (GABA) receptor (18) and the glycine receptor (19, 20).
Figure 2.1: The *T. marmorata* nAChR. Top) Image of the two dimensional crystalline array of the *T. marmorata* nAChR in its native membrane from (21) reproduced with permissions, scale bar 500 Å. Bottom left) image of preserved electric ray *Torpedo marmorata*, taken at the Hornimann Museum, Forest Hill, London. Electric organ is present on both side, the arrow points to the region in which the right side electric organ is situated. Bottom right) three dimensional atomic structure of the *T. marmorata* nAChR solved by electron diffraction (5) top view, left (scale bar shows 80 Å diameter). Side view, right (scale base shows 160 Å length). α-subunit coloured red, β-subunit coloured blue, γ-subunit coloured gold, δ-subunit coloured green

These higher resolution pLGIC structures have highlighted discrepancies in the structure of the *T. marmorata* nAChR which suggest that the register in the transmembrane domain is incorrect. Pierre-Jean Corringer et al. (23) have made the following observation from comparison to the bacterial homologues GLIC and ELIC:

“Regarding the structural model of Torpedo nAChR derived from EM data and from the AChBP crystal structure (Unwin, 2005), two local but important differences can be noted [compared with bacterial channels]. First, the loop linking the M2 and M3 segments at the interface between the ECD and the TMD protrudes within the subunits in the nAChR model, while it is clearly at the subunit interface in the bacterial receptor structures. Second, the axis of the M2 helices can be
nicely aligned between the nAChR and GLIC structures, but the structural alignment shows a one helix turn shift, with the intermediate ring of charged residues being at the base of the M2 helix in the Torpedo nAChR model, and one turn of helix higher in bacterial receptors.”

Further to this, work from Mnatsakanyen and Jansen shows that cross-linking data for the TM domain of the α-subunit from muscle-type nAChRs is better explained by structural models based on the glutamate gated chloride channel than on the T. marmorata nAChR which has a much higher sequence identity (6).

It is likely that these discrepancies arose from misassignment of atoms to the electron density derived from the original electron diffraction studies. In the PDB file header for the first segment of the T. marmorata nAChR that was structurally determined (the transmembrane domain, PDBid 1OED) the authors state in REMARK 3 (24):

“…THE LINK BETWEEN M1 AND M2 WAS POORLY RESOLVED AND THE TRACE HERE IS ALMOST CERTAINLY WRONG IN DETAIL…”

These coordinates appear to have been used as part of the basis of model building for the later more complete 4.0 Å structure (PDBid: 2BG9 (5)) thus the error was carried over.

Despite the likely error in transmembrane register, the T. marmorata nAChR model is an important structure for the understanding of pLGICs. It remains the only model to have been determined from a native lipid membrane environment and thus, is possibly the receptor structure that most closely resembles the physiologically important closed resting state. Additionally, structural data pertaining to this receptor has also been collected in both open and closed conformation, albeit at the lower resolution of 6.2 Å (25). Nonetheless, this data still provides important insight into the global conformational changes between the resting closed and functional open conformations, whilst still in the native lipid environment of the Torpedo postsynaptic membrane.
2.1.1.2 Prokaryotic channels

After the landmark publication of the *Torpedo* receptor structure (5), the next pLGICs to be determined were the bacterial ion channels ELIC (13) and GLIC (14, 15) (Figure 2.2). Owing to the lack of an intracellular domain, these bacterial homologues have been easier to manipulate into a crystalline state. Hence, these receptors could be resolved by x-ray crystallography and have become some of the most structurally studied pLGICs. Given that bacteria are single celled organisms and for example do not send intercellular signals in the same way as nerve cells, the existence of pLGICs in their bacterial genome is difficult to justify from an evolutionary perspective.

ELIC is activated by primary amines including the neurotransmitter GABA (26, 27). However, GLIC is unusual in comparison with other pLGICs, it is pH sensitive so is not “ligand-gated” in the same classical sense as other neurotransmitter gated channels. In fact, it is not known exactly where a proton must be accepted in order to elicit channel opening. A recent study investigating mutations of all titratable residues in GLIC found that many residues influence the sensitivity of the channel to protons. A glutamate residue (E35, near the structurally homologous neurotransmitter binding site) was identified as a key proton sensing residue. However, of the other residues that were found to influence channel sensitivity to protons, it is suspected that some may be involved in proton sensing, but it was difficult to unpick proton sensing from channel gating (28).

Almost all pLGICs from eukaryotic sources contain a conserved disulfide between the strands β6 and β7 in the extracellular domain and so have commonly been known as Cys-loop receptors (Figure 1.2). Both GLIC and ELIC do not contain the appropriate disulfide so are not strictly Cys-loop receptors, yet they do contain an absolutely conserved proline in this loop. It has been suggested that Pro-loop may be a more appropriate name for the superfamily of ion channels (29). Another key difference between these prokaryotic channels and their eukaryotic counterparts such as nAChRs and 5-HT3 receptors is that both GLIC and ELIC do not contain an intracellular domain.
Despite the structural variations of these prokaryotic channels versus eukaryotic pentameric ligand gated receptors, GLIC and ELIC do show particularly high fold similarity to Cys-loop receptors including the beta-sandwich in the extracellular domain and four transmembrane helices per subunit (Figure 2.2). As such, they have been used extensively as models for understanding the function of pLGICs as a whole (23). However, the structures of these channels may not offer the most appropriate template on which to base understanding of binding interactions which influence gating in nicotinic receptors in light of other available structural data for pLGICs.

![GLIC and ELIC structures](image)

Figure 2.2: GLIC and ELIC structures. Green GLIC (PDBid: 4HFI) from side view (left) and top view (centre). Orange ELIC (PDBid: 5SXV) side view (right) and top view (centre).

### 2.1.1.3 Glutamate gated chloride channel

After the determination of the *T. marmorata* nAChR, the anion selective glutamate gated chloride channel (GluCl) was the next eukaryotic pLGIC to have its structure solved by X-ray diffraction (16). It is one of only three classes of pLGICs to have structural data resolved at high resolution for multiple receptor states (better than 4.0 Å) (16, 30) (the other receptors include the bacterial channel GLIC (31, 32) and the glycine receptor (20)).
2.1.1.4 5-HT<sub>3</sub> receptor

Another highly important structural model, is that of the mouse 5-HT<sub>3</sub> receptor (17), determined as a co-crystal structure with bound nano-bodies. Although the intracellular loop of this model was truncated in order to facilitate crystallization, extra atomic information was assigned to the MX helix which was not resolved in the *T. marmorata* nAChR. The loop from TM3 to the MX helix covers a portal that appears between neighbouring subunits at the C-terminal end of the MA helices (Figure 2.3, left). Since the MA helices come in close proximity at their N-terminal end, seemingly blocking the passage of ions along the pore axis, it has been proposed that this portal forms part of the obligate ion conduction pathway (33) and the TM3-MX loop may play a regulatory role in conduction due to its ability to obstruct the flow of ions through the portal (17). This is especially true of 5-HT<sub>3A</sub>Rs whose portal is lined with positively charged residues which would seemingly disfavour the flow of positively charged ions (Figure 2.3, left). In nAChRs the portal is lined with negatively charged residues, so the importance of the TM3-MX loop and MA helices is not so obvious (Figure 2.3, right).

![Figure 2.3: Charges lining the intracellular portals for 5-HT<sub>3A</sub>Rs and nAChRs. Left) Zoomed region corresponding to the intracellular portal in the 5-HT<sub>3A</sub>R structure (PDBid: 4PIR). Right) Zoomed region corresponding to the intracellular portal in the nAChR structure (PDBid: 2BG9). Location of the portals is denoted with a blue ellipse, charged residues surrounding the portal are shown in ball and stick format, positively charged residues are shown in blue, negatively charged residues are shown in red. The principal subunit is on the side with a (+) symbol and the complimentary subunit is on the side with a (-) symbol.](image-url)

Chapter 2

75
Other than the *T. marmorata* nAChR, at the time of beginning this work, the 5-HT$_3$ receptor was the protein structural model with the highest sequence identity to the α7 nAChR and the only other model of a cation selective pLGIC including extracellular, intracellular and transmembrane domains. Thus, the 5-HT$_3$ receptor model (PDBid: 4PIR (17)) represents a useful template model.

However, having been determined from a crystalline array, ligated by nanobodies and with a truncated intracellular loop, the protein structure is far from its native environment. The fact that the pore is too small for conduction of hydrated cations, suggests that the model is in a non-conductive, possibly desensitized state. Alternatively, it may not be a physiologically relevant conformation of the receptor. Furthermore, a portion of the TM2-TM3 loop is not resolved in this receptor structure; postulated to play a role in the transduction of neurotransmitter binding to channel gate opening, the positioning of this loop would be an interesting portion of the receptor to understand. Especially as the partially resolved loop is in a different conformation to that shown for the TM2-TM3 loop in the *T. marmorata* nAChR.

Since the work in this chapter was carried out (34), a cryo EM structure of the full length 5-HT$_3$ receptor in a closed-resting state has been resolved (35). Furthermore, this most recent structure is not ligated with nano-bodies and shows alternative conformations of the ligand binding C-loops.

### 2.1.1.5 Glycine receptor

Single particle cryo EM structures of the zebrafish α1 glycine receptor (GlyR) have been determined for three receptor states (Figure 2.4) (20). One state is in complex with the glycine receptor inhibitor strychnine. Since the strychnine bound receptor exhibits an open C-loop conformation and a tightly closed channel, it is proposed to represent the closed resting state of the ion channel (Figure 2.4). The two other states are glycine bound, one with the additional allosteric modulatory ligand ivermectin. The state bound by glycine alone, has a large dilated pore and is proposed to represent the agonist bound open form of the channel. Finally, the receptor structure bound by both glycine and
ivermectin, exhibits a pore that appears open at the extracellular side, but constricted at the cytosolic side, hence it is proposed that this structure represents a desensitised non-conductive state (20) (Figure 2.4). This variety of structural data in different physiologically relevant states is incredibly important in determining how this family of receptors might open and close.

Figure 2.4: Receptor states imaged for glycine receptors. From left to right, the strychnine bound receptor corresponding to the closed conformation (PDBid: 3JAD) coloured red, the glycine bound receptor corresponding to the open conformation (PDBid: 3JAE) coloured green, the glycine and ivermectin bound receptor corresponding to the desensitized conformation (PDBid: 3JAF) coloured dark grey. Minimum pore radii are displayed as calculated in (20).

However, there are some limitations to the structures; having been solved by single particle cryo EM, the receptors were solubilised in detergent, and so lack the native membrane environment (which is important for the function of many pLGIC, especially nAChRs (36, 37)), hence they may not truly represent their respective physiological receptor states. Throwing further doubt on this is a comparison of the pore sizes of these glycine receptor structures with other homologues. The open channel found in the structures published by Du et al. (20) is significantly larger than that of any other seen to date (38).

Being anion selective channels as opposed to cation selective like the nAChR, GlyR atomic models are not the most appropriate template available for comparative models of nAChRs. The distinction in ion selectivity may result in differences in the gating mechanism, or the way in which ions passage through the pore. For example it is suggested that anionic channels may have narrower pores since it requires less energy to dehydrate anions
such as Cl⁻, compared with cations such as Na⁺ (39). Alongside the cryo EM structures of the zebrafish α1 GlyR, crystal structures of the human α3 GlyR have also been solved (19, 40).

2.1.1.6 GABA receptor

GABA receptors are another family of pLGICs which have had their structure determined by X-ray crystallography. However, to facilitate crystallization the intracellular loop between TM3 and TM4 in the β3 receptor was substituted for a short linker sequence (18) signifying a large truncation of the receptor. Nonetheless, when the truncated construct was expressed in HEK (human embryonic kidney) cells, inward currents were induced on application of agonists under patch-clamp electrophysiology suggesting that the receptor construct was still functional (18). However, the resolved structure is thought to represent a desensitized state (18).

More recently a GABA receptor α5/β3 chimera has been studied by X-ray crystallography and a co-crystal structure with the neurosteroid pregnanolone which potentiates receptor currents (41). Once again the intracellular domain is truncated, and the receptor conformations appear to be desensitized (41).

2.1.1.7 α7-AChBP chimera and monomeric nAChR TM domains

All the structures mentioned so far have contained both extracellular and transmembrane regions for the respective receptors. A recent set of pentameric structures of an acetylcholine binding protein (AChBP) and α7 chimera (42, 43) along with NMR structures for the transmembrane domain of α7 and other receptor subunit subtypes (α4 and β2) have been solved (44–46). A combination of these structures could also potentially serve as an appropriate set of templates for modelling both extracellular and transmembrane domains of the α7 nAChR.

While an exceptionally useful tool for studying interaction with the extracellular domain of α7 nAChRs, the α7-AChBP chimera structures may not serve such a purpose in studying ligand interactions that involve the transmembrane domain. The extracellular-transmembrane domain (EC-TMD) interface is
thought to perform an important role in the transduction of the signal of neurotransmitter binding to pore opening within the gating transition of any pLGIC (47). The α7-AChBP chimera structures are based on a construct where the TM domain is entirely missing. Furthermore, the residues in the TM2-TM3 loop which contributes to the EC-TMD interface is largely mutated in the NMR structure. Hence, the EC-TMD interface would not have conserved its interface and so models built on these templates may not provide a faithful representation of this functionally important area.

Similarly, the transmembrane segments studied by NMR represent excellent tools for studying binding of ligands at transmembrane intrasubunit cavities (45), however, they lack detail for the intersubunit interfaces within the transmembrane domain. Studied in a monomeric form, it is difficult to know which receptor conformational state they would be representative of (e.g. closed-resting, open, desensitized). Equally, the transmembrane domain is again far from its native environment when in solution, and many truncations and substitutions were made in order to achieve a stable construct for NMR in the case of the α7-TMD.

Aside from the problems that each of these templates present, with respect to modelling a receptor including a transmembrane domain, modelling an allosteric protein such as an nAChR with multiple templates can be a difficult task in itself. Since each of the above templates is representative of a separate portion of the nAChR in potentially different conformations, it becomes difficult to ensure that the correct orientation is achieved between the domains within the subunit. Indeed a mixed conformation model which does not properly represent a physiological or functional state or intermediate of the protein complex is likely to be the outcome.

2.1.1.8 Selection of template model

The foremost purpose of this study is to obtain native-like transmembrane geometries in structural models of the α7 nAChR. Each of the models previously mentioned was assessed by both the benefits and disadvantages of their use as a template (Table 2.1).
Table 2.1: Table of benefits and disadvantages of α7 nAChR receptor templates

<table>
<thead>
<tr>
<th>Template Receptor</th>
<th>Benefit</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>T. marmorata</em> nAChR (5, 25)</td>
<td>&gt;40 %seqID, native membrane environment, multiple receptor conformations, full length receptor</td>
<td>TM register error, 4.0 Å resolution</td>
</tr>
<tr>
<td>Prokaryotic receptors (13–15)</td>
<td>&lt;2.5 Å resolution, multiple receptor conformations</td>
<td>~20 %seqID, selective for oppositely charged ions, no Cys-loop, no intracellular domain</td>
</tr>
<tr>
<td>GluClα (16, 30)</td>
<td>&lt;3.5 Å, multiple receptor conformations</td>
<td>~20 %seqID, out of native membrane environment, selective for oppositely charged ions, truncated protein (including intracellular domain).</td>
</tr>
<tr>
<td>5-HT3 receptor (17)</td>
<td>3.5 Å resolution, ~30 %seqID, MX helix resolved</td>
<td>Out of native membrane environment, missing TM2-TM3 loop, desensitized conformation, truncated intracellular domain</td>
</tr>
<tr>
<td>Glycine receptors (19, 20)</td>
<td>&lt;4.0 Å resolution, multiple conformations</td>
<td>~20 %seqID, out of native membrane environment, selective for oppositely charged ions, open state may not be truly representative, truncated protein (including intracellular domain).</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>GABA receptors (18, 41)</td>
<td>&lt;3.0 Å resolution</td>
<td>~20 %seqID, out of native membrane environment, selective for oppositely charged ions, truncated intracellular domain</td>
</tr>
<tr>
<td>α7-AChBP (42, 43)/ α7 TM NMR (44)</td>
<td>&lt;2.5 Å resolution (α7-AChBP), &gt;70 %seqID</td>
<td>Lack of restraints for inter-domain contacts, TM domain not assembled as a pentamer, truncated intracellular domain, mutated loops</td>
</tr>
</tbody>
</table>

(%seqID = percent sequence identity)

Based on the comparison in Table 2.1 the *T. marmorata* nAChR was chosen as the best structure on which to base models of α7. This was due to the fact that it is the only structure of a pLGIC to have been determined in its native membrane environment (including lipids such as cholesterol which are essential for nAChR function (48)). Consequently, one might be most confident in the open and closed receptor structures for the *T. marmorata* nAChR being as similar as possible to their physiological counterparts compared with any of the other pLGIC structures. It is difficult to determine
whether many of the pLGIC structures derived from crystallography exhibit a desensitized state through following the physiological pathway to desensitization, or if these conformations are artefacts of the environment in which the proteins were crystallised/imaged.

A further illustration for the necessity of having native lipids present is an analysis of the radial size of the transmembrane domains of these receptors. The central pore of the receptors is lined by the TM2 helices, around the TM2 helices is situated another proteinatious ring of the TM1 and TM3 helices. Measuring the size of the TM1-TM3 helix ring gives an estimation of how radially disperse the TM domain is. When plotted on a graph, it is observed that the TM1-TM3 helix ring for the *T. marmorata* nAChR is far larger than any of the other receptor structures (Figure 2.5), none of which were imaged in a native-like environment. This indicates that the lipids in the post-synaptic membrane of the *Torpedo* ray may cause greater dispersity of transmembrane helices in nAChRs, for example this may occur via binding of cholesterol at TM sites. Alternatively a more disperse TM1-TM3 ring could be a feature that distinguishes nAChRs from other pLGICs.
Figure 2.5: TM1-TM3 ring radii for a selection of closed conformation pLGICs. Left, demonstration of the location of the TM1-TM3 helix ring using a slice from the transmembrane domain of T. marmorata nAChR (PDBid: 2BG9). Individual subunits of the receptor are boxed in grey rhomboids and helices are labelled for one subunit. Black dotted lines show the locations of the pore lining TM2 helix ring and the outer TM1-TM3 helix ring. Right, radii measured by selecting all residues annotated as being part of a helix within the TM1 and TM3 segments of each subunit and then creating a plane using the structure measurements tool in UCSF Chimera (49). The structures for which measurements were taken include 5-HT$_3$ receptor (PDBid: 4PIR, blue dots), the T. marmorata nAChR (PDBid: 2BG9, solid purple), GABA(A)β3 receptor (PDBid: 4COF, red dots), GLIC (PDBid: 4NPQ, solid green), the glycine receptor (PDBid: 3JAD, black dots) and the glutamate-gated chloride channel (PDBid: 4TNW, solid grey).

The existence of the transmembrane error which is the disadvantage of this structure, is clearly the most significant disadvantage of any of those listed in Table 2.1. However, the electron density maps for the closed and open receptors imaged in 2012 (25) are deposited in the electron microscopy databank (EMDB). These maps (although at the lower resolution of 6.2 Å) can be used alongside flexible fitting techniques to refine the T. marmorata nAChR structure and correct the transmembrane error.

2.1.1.9 α4β2 nAChR

Subsequent to the work in this chapter being carried out, a high resolution structure of the heteromeric α4β2 nAChR was published (50). The α4 subunit has a slightly higher sequence identity with α7 than does the α subunit of the T. marmorata nAChR. Within this structure, the additional MX helices, but not
the MA helices were resolved. The α4β2 receptor TM1-TM3 helix ring matches that of the *T. marmorata* nAChR almost exactly, interestingly, a cholesterol mimetic was found to be essential for the crystallization protocol (50) which may lend weight to the idea that cholesterol binding increases the dispersity of the TM helices. Furthermore, a set of cryo EM structures relating to the high and low ACh affinity stoichiometries of the α4β2 nAChR have also been released ((α4)2(β2)3 and (α4)3(β2)2 respectively) (51). These structures reveal some of the structural differences that may drive the changes in functional behaviour between the two receptors and also resolve the positions of several annular binding sites of cholesterol hemisuccinate.

With a higher resolution, seemingly the correct transmembrane register and apparently the same level of dispersity of transmembrane helices, the α4β2 nAChR would be a sensible choice of template for modelling the α7 nAChR. However, joining the list of other detergent solubilized structures, the α4β2 receptor structures appears to represent desensitized receptor states, and the transmembrane region is again outside of its native environment. Therefore, the pursuit of a corrected *T. marmorata* nAChR structure is still important as it will give access to potentially native-like open and closed receptor models.
2.2 Aims

To correct the transmembrane register error in the *T. marmorata* nAChR α subunit using flexible fitting methods for cryo EM data.

To build a model of the α7 nAChR based on these corrected models of the *T. marmorata* α subunit.
2.3 Correcting T. marmorata nAChR transmembrane error

The first step in correcting the transmembrane register error in the T. marmorata nAChR was to characterise the discrepancy and delineate exactly what needed to be changed to obtain the correct register.

2.3.1 Characterising the error

The backbone atoms of TM2 align well between the T. marmorata nAChR and its structural homologues 5-HT\textsubscript{3} receptor and GLIC. However, residues that align by sequence are approximately one turn of an \( \alpha \)-helix apart in the structural alignment (Figure 2.6). Yet, the residue that is at the same point in the sequence as Y234 in the nAChR (C239 in 5-HT\textsubscript{3} and W217 in GLIC) aligns with Y234 in both the sequence alignment and structural alignment. The one \( \alpha \)-helix turn gap (approximately 4 residues) appears after the TM1-TM2 loop. It seems that this arises due to the assignment of an extra turn of the helix TM1 in the T. marmorata nAChR which appears to be quite clear in the structural alignments (Figure 2.6).

Figure 2.6: Aligned TM2 helices show the assignment of an extra turn in TM1 of the T. marmorata nAChR. Left) alignment of TM2 backbone for nAChR (PDBid: 2BG9, purple) and 5-HT\textsubscript{3} receptor (PDBid: 4PIR, blue). Right) alignment of TM2 backbone for nAChR (PDBid: 2BG9, purple) and GLIC (PDBid: 4HF1, green). Centre) alignments between nAChR \( \alpha 1 \) (T.marmorata) and 5-HT\textsubscript{3} \( \alpha \) (M. musculus) (top), and nAChR \( \alpha 1 \) and GLIC (G. violaceus). Residues displayed in the structural alignments are highlighted with pink text, the dssp assignment of structure for the T. marmorata nAChR is shown above the sequence with the locations for TM1 and TM2 denoted. Adapted from Fig S21 in (16).
Investigating sequence-structure alignments of the *T. marmorata* nAChR slightly upstream of the alignment in Figure 2.7 sheds more light on the matter. It is quite clear from Figure 2.7 that although the β10 strands finish and TM1 helices start at a similar location in the sequence, residues 235-238 are assigned as helical at the end of TM1 in the nAChR but not in any of the other homologues, extending TM1 for a full extra turn in the nAChR structure.

![Sequence-structure alignment between the T. marmorata nAChR α subunit and some of its homologues. Top) sequence alignment, bottom) structural assignement for the aligned positions in the sequence alignment. nAChR (PDBid: 2BG9, chain A), 5-HT3 receptor (PDBid: 4PIR, chain A), GLIC (PDBid: 4HFI, chain A), GABA (PDBid: 4COF, chain A), GlyR (PDBid: 3JAD, chain A). Numbering above the sequences relates to the numbering of the nAChR sequence only. Identical residues are highlighted in boxes with red background and white text, similar residues are highlighted with red text on a white background in boxes outlined with blue. Strands are denoted with a black arrow and helices denoted with black spiral.](image)

Figure 2.7: Sequence-structure alignment between the *T. marmorata* nAChR α subunit and some of its homologues. Top) sequence alignment, bottom) structural assignement for the aligned positions in the sequence alignment. nAChR (PDBid: 2BG9, chain A), 5-HT3 receptor (PDBid: 4PIR, chain A), GLIC (PDBid: 4HFI, chain A), GABA (PDBid: 4COF, chain A), GlyR (PDBid: 3JAD, chain A). Numbering above the sequences relates to the numbering of the nAChR sequence only. Identical residues are highlighted in boxes with red background and white text, similar residues are highlighted with red text on a white background in boxes outlined with blue. Strands are denoted with a black arrow and helices denoted with black spiral.

Observing the closed *T. marmorata* nAChR structure in its available electron density map at 6.2 Å resolution (25) corroborates the idea that the transmembrane register problem arises from the erroneous assignment of one extra turn at the end of TM1. When the map is contoured at the author suggested value of 1.2 σ (standard deviations from the density histogram average), the final turn of the TM1 helix clearly protrudes from the density map (Figure 2.8).
Figure 2.8: Protrusion of final turn of TM1 helix from electron density map. Electron density map for closed T. marmorata nAChR (emd-2071, resolution: 6.2 Å, grey) and fitted model (PDBid: 4AQ5, purple)(25), chain A corresponding to the $\alpha_\gamma$ subunit is displayed with all others faded out. Black dotted ring is shown to highlight the protruding helical turn.

### 2.3.2 Refinement strategy

In order to overcome the register error within the TM domain of the $\alpha$ subunit, a hierarchical modelling and refinement strategy was devised, involving a register correction and then subsequent refitting of the receptor model into available cryo EM density maps.

First, there are two $\alpha$ subunits present within the *T. marmorata* nAChR, one that forms the principal interface of an acetylcholine binding site with the $\delta$ subunit ($\alpha_\delta$), and one with the $\gamma$ subunit ($\alpha_\gamma$) (Figure 2.1, bottom). Electrophysiological and biochemical studies suggest that the $\alpha_\gamma$ subunit has a lower affinity for ACh than the $\alpha_\delta$ subunit and plays the larger role in effecting channel gating (52, 53). The findings of these studies were upheld in light of later obtained structural data, pertaining to the open and closed conformations of the nAChR. On binding of ACh (25), it was observed that the $\alpha_\gamma$ undergoes a significantly larger rearrangement than does the $\alpha_\delta$ subunit and thus drives the opening of the receptor. The aim of this refinement was to provide a template for modelling a homomeric $\alpha_7$ nAChR which contains five identical agonist binding sites. The estimation was assumed, that in a homomeric receptor where all subunit interfaces can contribute equally to gating, the
motions of the low affinity $\alpha_y$ subunit may be more representative of opening and closing than those of the high-affinity $\alpha_5$ subunit. Therefore, the $\alpha_y$ subunit was chosen for refinement.

In an initial attempt to correct the transmembrane register, a four residue shift was added to create a sub-optimal alignment of the *T. marmorata* nAChR $\alpha_y$ subunit against itself (Figure 2.9A). According to the sequence-structure alignments (Figure 2.6 and Figure 2.7), the extra helical assignment at the C-terminal end of TM1 appears to begin at residue L235. Applying a number of origins for the shift showed that the resulting models tend to break the helical assignment, one or two residues upstream of the added shift. Hence, P236 was chosen as the origin of the sequence shift in the sub-optimal alignment used for modelling. Owing to the use of a sub-optimal alignment, and the absence of neighbouring subunits during the modelling, the resulting models of the $\alpha_y$ subunit exhibited a poor inter-domain geometry. For example, the intracellular MA helix was poorly placed (probably due to lack of restraints for...
building due to lack of neighbouring subunits) and removed for the subsequent modelling steps (Figure 2.9B and C). Furthermore, compared with the corresponding subunit in the *T. marmorata* nAChR structure (PDBid: 2BG9, chain A), the ECD and TMD showed a bent conformation. As such, while the ECD of the newly generated model could be superposed well on the *T. marmorata* nAChR structure (PDBid: 2BG9, chain A), the TMDs did not align (Figure 2.9B).

![Figure 2.10: Convergence of cross correlation for Flex-EM MD refinement run of remodelled Torpedo α subunit into open EM map (EMD-2072).](image)

Flex-EM represents a useful tool in flexible fitting of atomic models to cryo EM density maps at a range of resolutions (54–56). Importantly, Flex-EM shows a large radius of convergence, in that even models which are far from their correct orientations can be fitted efficiently. Flex-EM takes advantage of using rigid bodies, segments of a protein structural model that are kept rigid during fitting. The use of rigid bodies in effect makes the fitting more coarse-grained and applicable to intermediate resolution EM maps.

The Flex-EM Molecular Dynamics (MD) fitting protocol was utilised with five rigid bodies which represented the entire ECD, and each of the four individual helices of the TMD. After one MD iteration, the cross-correlation coefficient converged within a single MD run, showing that the model was fitted (Figure 2.10).
The new inter-domain geometry matched that of the deposited T. marmorata nAChR structure (PDBid: 2BG9) (Figure 2.9C), however, some regions were still in need of an improved fit.

The conformation of the TM1-TM2 loop did not follow the shape of the density in the local region well, so the next stage of the hierarchical refinement involved remodelling this loop. MODELLER loop refinement (57) in combination with the TEMPy software package (58) was used to first generate an ensemble of loop conformations and then select the conformation which displayed the best local fit to the density based on the segment based cross correlation coefficient (SCCC) from TEMPy. The resultant loop conformation improved the SCCC in this region from 0.65 to 0.82 and showed a visually better agreement with the density map (Figure 2.11).

At this stage of refinement, secondary structural elements had been rigidly fit into the density map for the closed conformation of the receptor (emd-2071). However, some more subtle changes in receptor structure such as bending of helices and twisting of the TM domain appeared to be necessary to improve the fit of the model.

*Flex-EM* has successfully been used for atomic fitting of structures in density maps at resolutions of greater than 4.5 Å (56). However, at lower resolutions, such as that of the maps in use here (~6.2 Å), assignment of rigid bodies is essential for fitting with *Flex-EM*. The assignment of rigid bodies that will
allow the level of flexibility necessary to obtain a good fit with the experimental map is a non-trivial task. Rigid-body detection software such as *RIBFIND* (59) is aimed at overcoming this step by clustering secondary structure elements (SSEs). However, for fits that are already reasonably close to the experimental density, clustering of SSEs does not always improve the fit more than use of individual SSEs as rigid bodies (60) and that was found to be the case here.

A complimentary flexible fitting method to *Flex-EM* is a normal mode based method called *iMODFIT* (61). The *Flex-EM* MD approach uses simulated annealing molecular dynamics to ‘heat’ and ‘cool’ a molecular system *in silico* and bias the cooling structure towards the high density regions of the density map. However, *iMODFIT* approximates the protein structure as a harmonic system, and combines a number of the low frequency harmonic motions to move structural models between conformations, selecting those that agree best with the experimental density. A consequence of the normal mode approach used by *iMODFIT* is that assignment of rigid bodies is not necessary in the same way as it is for *Flex-EM*. Furthermore, normal mode analysis (NMA), has been previously used to model gating transitions in nAChRs (8). As such, *iMODFIT* seemed an ideal choice for modelling and fitting the finer changes in structure observed in the closed conformation and fitting the closed conformation model into the open conformation density. The resultant models after running *iMODFIT* improved global cross-correlation coefficients by 6.5% for the open conformation map and 6% for the closed conformation map (Figure 2.12).
Figure 2.12: Change in cross-correlation coefficients during iMODFIT run to closed conformation density map (emd-2071). ΔCCC is the change in cross-correlation coefficient.

Difference maps which show the result of subtracting one map from another give details of how structures change between conformations. Difference maps were calculated by subtracting the open map from the closed map and vice versa using the difference_map.py script in TEMPy (58). The most obvious differences between the closed and open maps is a straightening of TM2, bending towards the C-terminal end of TM4 and a small clockwise rotation of the four helix bundle. The fit achieved by iMODFIT agreed well with the changes observed in the difference maps (Figure 2.13).
Figure 2.13: Agreement of TMDs fit my iMODFIT to the closed (red) and open (green) density maps. Left) TMD shown in sideview from within the membrane. The location of the channel pore relative to the TMD is shown. Right) slice through TMD from above the membrane. Shifts of transmembrane helices between closed and open conformations are shown with straight arrows, a global clockwise rotation of the TMD is shown with a curved arrow. The green model corresponds to that fitted by iMODFIT to the open density map and the red model corresponds to that fitted by iMODFIT to the closed density map. Difference densities are shown for the closed conformation map (emd-2071) minus the open map (emd-2072) in red and vice versa in green.

A drawback of using iMODFIT is that the atomistic forcefield applied during fitting does not appear to be very good. As such, the model quality is poor based on geometric measures such as MOLProbity (62) (Table 2.2), a purely geometric analysis of protein structure, commonly used by crystallographers to spot problems in their structural models.

To improve geometric model quality, without moving away from agreement with the density map, a Flex-EM conjugate gradients (CG) run was carried out. The CG protocol implements the CHARMM22 forcefield (63) with the cross correlation with the density as an additional restraint in the optimization of the model.
Table 2.2: Table of MolProbity (62) statistics for models produced by iMODFIT.

<table>
<thead>
<tr>
<th>Model</th>
<th>Target values</th>
<th>Torpedo nAChR αγ closed iMODFIT</th>
<th>Torpedo nAChR αγ open iMODFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clashscore</td>
<td>-</td>
<td>152.21 (0th %ile)</td>
<td>151.14 (0th %ile)</td>
</tr>
<tr>
<td>Poor Rotamers</td>
<td>&lt;0.3%</td>
<td>18 (5.77%)</td>
<td>18 (5.77%)</td>
</tr>
<tr>
<td>Favoured Rotamers</td>
<td>&gt;98%</td>
<td>268 (85.90%)</td>
<td>268 (85.90%)</td>
</tr>
<tr>
<td>Ramachandran Outliers</td>
<td>&lt;0.05%</td>
<td>7 (2.12%)</td>
<td>7 (2.12%)</td>
</tr>
<tr>
<td>Ramachandran Favoured</td>
<td>&gt;98%</td>
<td>305 (92.42%)</td>
<td>301 (91.21%)</td>
</tr>
<tr>
<td>Molprobity Score</td>
<td>-</td>
<td>3.69 (6th %ile)</td>
<td>3.73 (5th %ile)</td>
</tr>
<tr>
<td>Cβ Deviations &gt; 0.25 Å</td>
<td>0</td>
<td>5 (1.56%)</td>
<td>5 (1.56%)</td>
</tr>
<tr>
<td>Bad Bonds</td>
<td>0</td>
<td>16 (0.58%)</td>
<td>15 (0.54%)</td>
</tr>
<tr>
<td>Bad Angles</td>
<td>&lt;0.1%</td>
<td>72 (1.9%)</td>
<td>70 (1.84%)</td>
</tr>
</tbody>
</table>

For clashscore (number of van der Waals’ radii overlaps > 0.4Å per 1000 atoms) and Molprobity score (combination of clashscore, rotamer and Ramachandran analysis in a single score, normalized to X-ray resolution scale) the percentile (%ile) ranks are shown after the score with 100th %ile being the best and 0th %ile being the worst amongst other structures in the statistical potential. The other values all show their percentage proportion after the absolute value in each case. Target values are shown in the row labelled Target values.

2.3.3 Model evaluation

Per residue scoring schemes of model quality are extremely useful, they allow the assessment of model quality/goodness-of-fit on a local basis. Two local criteria were used for the assessment of these models’ segment based Manders’ overlap coefficient (SMOC) (56) and QMEANBrane (64). The SMOC score from TEMPy (56) measures the local goodness-of-fit of a model against its experimental density map by calculating the local overlap of a sliding window of residues. QMEANBrane is a protein structure statistical potential built from high resolution membrane protein structures. Statistical
potentials are a powerful method for assessing the ‘nativeness’ of protein structural models compared with what is expected to be ‘native’ based on high resolution protein structures from the PDB.

The fact that QMEANBrane was built from high resolution membrane protein structures is important, integral membrane proteins have very different properties to those of soluble proteins, especially relating to matters such as the polarity of sidechains which appear to be ‘buried’ or ‘exposed’ within the membrane bound segments. A native membrane protein structure would likely score very poorly if assessed against statistical potentials built based on soluble proteins. Indeed previous potentials such as QMEAN (65) are heavily biased towards soluble protein structures due to their significantly higher availability at high resolution in the protein databank (66). Therefore, the use of potentials that can accurately score membrane bound regions is very useful in understanding model quality of integral membrane proteins.

The advantage of using these two scoring functions for evaluation is that they assess complimentary aspects of the model quality. SMOC detects the agreement of the model with experiment, while QMEANBrane statistically assesses the model against what is expected from an atomic model and is more of a geometric assessment. To provide a useful frame of reference for the QMEANBrane scores, those of the original model (PDBid: 2BG9) were subtracted from the QMEANBrane scores of the refined models to give $\Delta QMEANBrane$. In this scheme, scores greater than zero denote an improvement in model quality, while scores below zero denote a decrease in model quality.

Since the goal of this project was to obtain the correct TM geometry, the TM domain was the region in which the model evaluation was focused. Plots of local scoring suggests that the final models after the CG minimization show overall better model quality according to $\Delta QMEANBrane$ in the TM region and very similar agreement with the experimental density maps according to SMOC score (Figure 2.14).
Figure 2.14: Evaluation of the transmembrane region of the refined models of Torpedo nAChR. For both the closed and open models (right and left panels, respectively), data are presented for ΔQMEANBrane scores (upper panels) and SMOC scores (lower panels). The original model (2BG9) is shown as a black line; the fitted models for the cryo-EM density maps (4AQ5 and 4AQ9 for closed and open structures respectively) is shown in blue; the model output by iMODFIT during the hierarchical refinement is shown as a purple line and the final energy minimised model for both closed and open structures is shown as a red line. Note, for ΔQMEANBrane scores the original model is shown as a dashed black line to indicate that the scores for this model were subtracted from the scores for all other models. In each case, the per residue score is given on the y-axis. Regions corresponding to transmembrane α-helices are shown as grey rectangles. SMOC is not shown for 2BG9 in the open conformation map since 2BG9 represents a closed state receptor.

A region between helices TM2 and TM3 scores lower than 2BG9 in both the open and closed refined models. Compared with 2BG9, the backbone RMSD in this region is low, 1.15 Å and 0.775 Å for open and closed models respectively so the significant drop in model quality seems unusual. The RMSD between 2BG9 and the closed model is expected to be lower since 2BG9 represents a closed conformation. Despite this, there are some changes around the end of TM2 and start of TM3 which may lead to lower assessment of quality. Nonetheless, this is also the case for the models 4AQ5 and 4AQ9 (closed and open nAChR at 6.2 Å) which are also fitted to the maps used for refinement here, indicating that the fit to the experimental density in this region may be a contributing factor to the lower model quality. This is strengthened by the observation that in the same region. SMOC score is increased in the closed conformation for the refined model, with respect to 2BG9.
Another more qualitative assessment is the comparison of the transmembrane register with known protein structures. After refinement the TM1 helix assignment finished at F233, an analogous position of that from the homologous 5-HT₃ (Figure 2.15).

Figure 2.15: Alignment sequence and structure of refined nAChR subunit against a selection of homologues. Top) sequence alignment, bottom) structural assignment for the aligned positions in the sequence alignment. nAChR (refined closed model), 5-HT₃ receptor (PDBid: 4P1R, chain A), GLIC (PDBid: 4HFI, chain A), GABA (PDBid: 4COF, chain A), GlyR (PDBid: 3JAD, chain A). Numbering above the sequences relates to the numbering of the nAChR sequence only. Identical residues are highlighted in boxes with red background and white text, similar residues are highlighted with red text on a white background in boxes outlined with blue. Strands are denoted with a black arrow and helices denoted with black spiral.

The overall quality of the TM domain appeared to have improved without losing agreement between the model and the experimental density, and the register of the TMD appeared to be in agreement with structural homologues post-refinement. The next evaluation was then using MolProbity (62). MolProbity provides a breakdown of the number of clashes, normalised against the number of atoms, side-chain rotamer analysis, backbone dihedral analysis as well as a rundown of numbers of bond lengths, angles and Cβ positions which deviate from expected values. Alongside this, target values for each metric are provided which represent the values of a ‘good’ model.
Table 2.3: Molprobity (62) statistics for templates and models created.

<table>
<thead>
<tr>
<th>Model</th>
<th>Targets</th>
<th>Torpedo nAChR αγ closed (2BG9)</th>
<th>Torpedo nAChR αγ closed (4AQ5)</th>
<th>Torpedo nAChR αγ open (4AQ9)</th>
<th>Torpedo nAChR αγ open refined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clashscore</td>
<td>-</td>
<td>162.72 (0th %ile)</td>
<td>140.18 (0th %ile)</td>
<td>141.18 (2nd %ile)</td>
<td>127.18 (0th %ile)</td>
</tr>
<tr>
<td>Poor Rotamers</td>
<td>&lt;0.3%</td>
<td>96 (30.77%)</td>
<td>120 (38.46%)</td>
<td>118 (37.82%)</td>
<td>19 (6.09%)</td>
</tr>
<tr>
<td>Favored Rotamers</td>
<td>&gt;98%</td>
<td>169 (54.17%)</td>
<td>142 (45.51%)</td>
<td>139 (44.55%)</td>
<td>258 (82.69%)</td>
</tr>
<tr>
<td>Ramachandran Outliers</td>
<td>&lt;0.05%</td>
<td>27 (8.18%)</td>
<td>43 (13.03%)</td>
<td>44 (13.33%)</td>
<td>2 (0.61%)</td>
</tr>
<tr>
<td>Ramachandran Favoured</td>
<td>&gt;98%</td>
<td>254 (76.97%)</td>
<td>233 (70.61%)</td>
<td>224 (67.88%)</td>
<td>319 (96.67%)</td>
</tr>
<tr>
<td>Molprobity Score</td>
<td>-</td>
<td>4.57 (0th %ile)</td>
<td>4.65 (0th %ile)</td>
<td>4.67 (3rd %ile)</td>
<td>3.37 (11th %ile)</td>
</tr>
<tr>
<td>Cβ Deviations &gt; 0.25 Å</td>
<td>0</td>
<td>1 (0.31%)</td>
<td>98 (30.62%)</td>
<td>95 (29.69%)</td>
<td>1 (0.31%)</td>
</tr>
<tr>
<td>Bad Bonds</td>
<td>0</td>
<td>11 (0.4%)</td>
<td>55 (1.98%)</td>
<td>45 (1.62%)</td>
<td>3 (0.11%)</td>
</tr>
<tr>
<td>Bad Angles</td>
<td>&lt;0.1%</td>
<td>17 (0.45%)</td>
<td>469 (12.36%)</td>
<td>454 (11.96%)</td>
<td>58 (1.53%)</td>
</tr>
</tbody>
</table>
For description of the values see Table 2.2. The models generated here score poorly against other models, especially in terms of clashscore, however, so do the template models, and the new models score better than the templates on the Molprobity score scale in all cases. No cis-prolines were observed in any of the models. The models 2BG9, 4AQ5 and 4AQ9 were cropped to be of the same extent as the subunits modelled in this study.

Table 2.3 shows the results of MolProbity analysis on each of the PDB models and the refined models. Although, clashes remain high in the refined models, this is likely due to the poor quality of the starting models which also contain very high numbers of clashing residues. However, aside from the clashscore, every other metric provided by MolProbity shows very significant improvements from the deposited models to the refined models (and also an improvement over the MolProbity statistics of the models produced by iMODFIT, Table 2.2). Particularly important is the backbone dihedrals (Ramachandran Favoured/Outliers) which show a 20-30% increase in the number of favoured dihedrals. Given the overall improvement in model quality and the agreement of the refined models with the TM register observed in homologues, these refined models were taken to be adequate as templates for modelling a7.

2.4 Comparative modelling of a7 nAChR

With refined structural templates for a7 from the T. marmorata nAChR αγ subunit in closed and open conformations, comparative modelling is a relatively straightforward task. MODELLER was used to create models of an a7 subunit with a previously published alignment between the human a7 and T. marmorata a1 sequences (Figure 2.16) (8), showing a 45% sequence identity in the transmembrane region, and the two generated templates.
Figure 2.16: Sequence alignment of *T. marmorata* nAChR α subunit with α7 nAChR subunit used for comparative modelling. Secondary structure is denoted as arrows (β-strands), spirals (α-helices) and ‘T’ (turns). The cysteines involved in the cys-loop disulfide are denoted with a green ‘1’. A forward slash indicates where the intracellular domain has been excluded from the alignment and, as a consequence, amino acid numbering changes after this point. Note: amino acids are numbered according to the mature protein following cleavage of the signal sequence. Conserved residues are shown as white text on a red background and residues with similar properties are shown as red text on a white background. (8)
The 100 models of the α7 subunit that were generated were ranked, again using QMEANBrane. The local scores for each residue were summed and the total used as a global ranking criterion. The register of the TM2 versus TM1 helices was checked against the NMR model of the α7 subunit TM domain (PDBid: 2MAW) and showed good agreement (Figure 2.17).

![Figure 2.17: Agreement in TM1 vs. TM2 register between model of α7 based on refined T. marmorata nAChR αγ subunit (light grey) and α7 nAChR TM NMR structure (PDBid: 2MAW, blue).](image)

In order to generate a pentameric model, the best ranking monomeric subunit was superposed on the αγ subunit of the relevant T. marmorata nAChR model (PDBid: 4AQ5 or 4AQ9 for closed and open conformations respectively) and five-fold symmetry was applied about the coordinate system of the corresponding cryo EM density map. Measuring the size of the channel pore in both open and closed conformations showed a contraction in pore diameter from open to closed of approximately 3 Å (Figure 2.18). The change in pore size was similar to that observed in the open vs. closed T. marmorata receptor structures (25).
Figure 2.18: Change in pore diameter of α7 nAChR pentameric models. The blue trace represents the closed conformation receptor model, the red trace represents the open conformation receptor model. Membrane bounds are shown with black dashed lines with the extracellular domain above the top dashed line, the transmembrane domain between the two dashed lines and the intracellular domain below the lower dashed line. Pore diameters were calculated using the PoreWalker online server (http://www.ebi.ac.uk/thornton-srv/software/PoreWalker/) (67).

2.5 Conclusions and future work

To summarise, refined models of the αγ subunit of the T. marmorata nAChR were built in open and closed conformations of the receptor. The new models take into account a residue register shift in the transmembrane domain which is due to the assignment of an extra turn of TM1. The register in this region now appears to agree with other pLGIC structures, and the subunit models appear to be of better geometric quality than the starting models, whilst maintaining a similar level of agreement to the cryo EM density maps from which they are derived.

Based on these open and closed models of the αγ subunit, comparative models of an α7 nAChR subunit were built in both open and closed conformations. Again, the register between TM1 and TM2 helices agreed between the new models of α7 and an NMR derived α7 TMD structure (PDBid: 2MAW). Applying five-fold symmetry to obtain the pentameric protein complex gave models which showed a dilated pore in the open conformation compared with the closed conformation. The models developed here serve
as useful points of reference for understanding functional data relating to α7 nAChR and predicting properties of the receptors based on their predicted structure.

Here, one subunit of the *T. marmorata* nAChR has been refined to produce a model of the αγ subunit, wherein the TM1-TM2 register agrees with that from other pLGIC structures. However, this alleged register error is present in all subunits of the *T. marmorata* nAChR. Thus, a useful future task would be to extend the refinement to all subunits. A complete refined nAChR structure would provide a useful set of templates for modelling other nAChR receptors in different conformational states, including the α4β2 and α7β2 receptors which are of therapeutic interest, as described in the introduction.

Aside from the register error caused by the assignment of an extra turn of TM1, the interconnecting loop between TM2 and TM3 has a different conformation in all other pLGIC structures compared with those of the *T. marmorata* nAChR. In future refinements to the structure, it would be interesting to adjust the conformation of this loop in the modelled subunits as it is proposed to play an important role in receptor gating.

Since this work was conducted, the structure of the closed-resting 5-HT₃AR has been solved by cryo EM and the α4β2 nAChR has been solved by X-ray crystallography and cryo EM in likely desensitized conformations. The subunits of these receptor structures could serve as useful templates to ensure the correct TM register in models of the *T. marmorata* nAChR and also to adjust the conformation of the TM2-TM3 loop. Furthermore, coordinates for the MX helix are assigned in both the 5-HT₃AR and α4β2 nAChR structure. Fitting of α4 or β2 subunits into the *T. marmorata* nAChR density maps highlights helical tubes of density that appear to relate to the MX helix. Therefore, it may be possible to assign and fit these helices in future models of the *T. marmorata* nAChR.

Finally, the newly developed α7 nAChR receptor models can be utilised to aid understanding of functional data relating to the effects of transmembrane ligands including positive allosteric modulators such as TQS and TBS-516.
Especially since these new models of α7 have been modelled based on the corrected structure of a Torpedo nAChR subunit, which is still the only pLGIC to have been structurally studied in a membrane environment that resembles the native lipid.

### 2.6 References


Chapter 2


38. Gonzalez-Gutierrez G, Wang Y, Cymes GD, Tajkhorshid E, Grosman C
Chasing the open-state structure of pentameric ligand-gated ion channels. *J Gen Physiol* 149(12):1119–1138.


48. Criado M, Eibl H, Barrantes FJ (1982) Effects of lipids on acetylcholine receptor. Essential need of cholesterol for maintenance of agonist-


Chapter 3

3 Investigating diversity in positive allosteric modulators of α7 nicotinic acetylcholine receptors using a revised molecular model and consensus docking protocol

Some of the work in this chapter has been published in the following article:

3.1 Introduction

Utilising the revised structural models of the α7 nAChR produced in Chapter 2 it was possible to interrogate molecular interactions between the receptor model and small-molecules of interest using docking.

3.1.1 Small-molecule docking

As discussed in Chapter 1, there are many varieties of docking programs which employ a range of different pose generation engines and binding mode scoring functions. When predicting binding modes using small-molecule docking, it is important that both the docking algorithm is good enough to ensure that the correctly docked pose is generated (where correctly docked means that the docking solution is representative of the true binding mode), and that the scoring function is good enough to rank a correctly docked pose higher than other incorrectly docked conformations. The accuracy of docking can vary greatly, for example, one study reports that accuracy can vary in the range 1-84% (1).

Consensus approaches to small-molecule docking have emerged as a way of improving accuracy of binding mode prediction (2–4). Different methods for both creating and scoring docking solutions suffer from their own pitfalls. In combining the different methodologies, it is possible to synergistically overcome the problems and boost the accuracy of docking results.

3.1.1.1 Consensus Docking

Houston and Walkinshaw showed that consensus docking with just two docking programs could boost accuracy of pose prediction from 64% to 82% (2). They employed two different docking packages (AutoDock4 and Vina) to predict poses for the same protein-ligand complexes. The solutions were filtered by selecting only predicted poses that agreed within 2.0 Å between the two programs for consensus scoring. Since many search algorithms should be able to reproduce the true binding mode, it is expected that in the majority of cases solutions not predicted by more than one search algorithm will likely not be representative of the binding of the docked ligand. They found that removing these likely false positives before entering the scoring stage of the
docking protocol improved the likelihood of ranking the true positives highest. In the above study, 2.0 Å was chosen as the RMSD cut-off based on the observation that it minimized false negatives (where true poses were discarded) at an acceptably low false positive rate (~15% of solutions) (2).

3.1.2 Positive allosteric modulators of α7 nAChRs

Allosteric modulators are compounds which bind to receptors such as nAChRs at a site away from that of the endogenous neurotransmitter (the orthosteric binding site) and in doing so are able to influence the function of the receptors. Allosteric modulators can be classed according to the effects they produce, for example:

- Allosteric agonists (AAs) – able to open the receptor gate by binding at an allosteric site (i.e. do not compete with orthosteric antagonists).
- Positive allosteric modulators (PAMs) – no effect on the receptor when binding alone, but increase the response of the receptor to orthosteric agonists.
- Negative allosteric modulators (NAMs) – reduce the response of receptors to orthosteric agonists.
- Silent allosteric modulators (SAMs) – have apparently no effect on the receptor but seem to be able to block the effects of PAMs, NAMs and AAs.

For rapidly desensitizing receptors such as α7 nAChRs, PAMs can be further subdivided. Type I PAMs increase only the maximum response of a receptor population to an orthosteric agonist. Type II PAMs both increase the response of a receptor population to an orthosteric agonist as well as slowing the rate at which receptors enter an agonist bound, non-conductive state (desensitized) (5).

PAMs of nAChRs are interesting as compounds of therapeutic potential (5–8). Several classes of PAMs and AAs have been discovered which affect the α7 nAChR (6, 9, 18, 19, 10–17). A subset of arylsulfonamide containing PAMs (Figure 3.1) have been examined extensively in the Millar lab. Interestingly, while all three of the compounds in Figure 3.1 have been found to act as
classical type II PAMs on wild-type α7 nAChRs (18, 20, 21), strikingly different pharmacological effects were observed for A-867744 versus TBS-516 and TQS on a series of mutant α7 nAChRs (Table 3.1) (22).

![A-867744, TBS-516, TQS](image)

Figure 3.1: arylsulfonamide containing PAMs: A-867744, TBS-516 and TQS.

Table 3.1: Pharmacological effects of three arylsulfonamide compounds (A-867744, TBS-516 and TQS) on wild-type α7 nAChRs and on α7 nAChRs containing single point mutations.

<table>
<thead>
<tr>
<th></th>
<th>Wild-type</th>
<th>W54A</th>
<th>S222M</th>
<th>L247T</th>
<th>M253L</th>
<th>M260L</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-867744</td>
<td>PAM</td>
<td>PAM</td>
<td>Inhibitor</td>
<td>Inhibitor</td>
<td>Inhibitor</td>
<td>PAM</td>
</tr>
<tr>
<td>TBS-516</td>
<td>PAM</td>
<td>Agonist</td>
<td>PAM</td>
<td>Agonist</td>
<td>&lt; PAM*</td>
<td>Agonist</td>
</tr>
<tr>
<td>TQS</td>
<td>PAM</td>
<td>Agonist</td>
<td>PAM</td>
<td>Agonist</td>
<td>&lt; PAM*</td>
<td>Agonist</td>
</tr>
</tbody>
</table>

* '< PAM' indicates a substantial reduction or abolition of PAM activity.

These allosteric modulators are not active in chimeric receptors containing a TM domain derived from 5-HT₃R and ligand binding domain from α7 nAChR, yet are active when the TM domain of α7 is present while the ligand binding domain and intracellular domains are taken from 5-HT₃ (18, 23, 24).

Furthermore, it is observed that pre-application of the modulators is necessary in order to obtain an immediate modulatory effect on co-application with acetylcholine. This suggests that some time is required for the modulator to access its site of action. Taken in combination with the vast number of TM mutations which dramatically affect the action of these modulators, it is likely that these PAMs act via a transmembrane site (18, 20, 25). However, the
exact location of this site is not well understood, partly due to the dispersity of single point mutations found to influence the activity of these compounds.

Previously, docking studies have been carried out using homology models of α7 based on the TM domain of the α subunit of the *T. marmorata* nAChR and proposed the existence of an intrasubunit cavity through which these compounds may act on the receptor (24, 26). However, as discussed in the previous chapter this template model is thought to contain an error in the transmembrane register and this will likely have had an effect on the interpretation of results from docking studies in this region.

Here, I have developed a docking protocol for use with the revised α7 nAChR comparative models discussed in Chapter 2. These offer a better model with which to study transmembrane binding since they contain a corrected transmembrane register, but also preserve the transmembrane helix packing observed in the only template model to have been structurally determined in its native lipid bi-layer, the *T. marmorata* nAChR.
3.2 Aims

Predict binding modes of the three arylsulfonamide PAMs A-867744, TBS-516 and TQS based on the new structural models from Chapter 2.

Exploit consensus methods to improve confidence in binding mode predictions.

Determine if the transmembrane register shift has any effect on where these PAMs are predicted to bind.

Establish if, through binding mode prediction, it is possible to rationalise the diverse effects observed for the three PAMs A-867744, TBS-516 and TQS on α7 nAChRs containing single point mutations.
3.3 Results and discussion

3.3.1 Choice of docking program and scoring function

A consensus docking protocol (Figure 3.2) was employed based on the work of Houston and Walkinshaw (2) to search the TM domain of the α7 nAChR comparative models for a binding location of allosteric modulators. The docking software packages GOLD v5.2.2 and VINA v1.1.2 were chosen due to their free availability for use and speed of docking. More importantly, the search algorithms in the two software packages differ; GOLD utilises a steady-state operator-based genetic algorithm for conformational searching (27), whereas VINA employs an iterated local search global optimizer (28).

Within GOLD, the force-field type fitness function GoldScore and the empirical piecewise linear potential ChemPLP were chosen for scoring during optimisation and rescoring the docked solutions, respectively. GoldScore is given by the sum of three energy defining components (Equation 3.1):

\[
\text{GoldScore} = \text{H}_\text{Bond}_\text{Energy} + \text{Complex}_\text{Energy} + \text{Internal}_\text{Energy}
\]

The hydrogen bonding energy (H_Bond_Energy), the energy of steric interaction, also accounting for the non-polar interactions (Complex_Energy) and the internal energy of the ligand as determined by molecular mechanics expressions (Internal_Energy) (27). The ChemPLP scoring function (29) is comprised of a linear combination of terms from the piecewise linear potential (PLP) (30) to model steric complementarity of protein and ligand, terms from ChemScore (31) to account for angle-dependency in hydrogen bonding and metal binding, and the torsional potential from the Tripos force field (32), as well as a heavy-atom clash term to account for intraligand interactions (Equation 3.2):

\[
\text{CHEMPLP} = f_{plp} + f_{hb} + f_{hb-ch} + f_{hb-CHO} + f_{met} + f_{met-coord} + \\
f_{met-ch} + f_{met-coord-ch} + f_{clash} + f_{tors} + c_{site}
\]

Where the term \( f_{plp} \) is the piecewise linear potential, the terms \( f_{hb} \) and \( f_{hb-ch} \) are the distance- and angle-dependent hydrogen bonding terms for uncharged and charged hydrogen bonding pairs and \( f_{hb-CHO} \) is evaluated when the
acceptor atom is an oxygen. Metal bonding is treated by the weighted
distance- and angle-dependent potential $f_{\text{met}}$ if the metal is either magnesium
or calcium, otherwise ideal interaction polyhedra are considered in the term
$f_{\text{met-coord}}$. If a charged acceptor is present, the metal terms are additionally
weighted to give $f_{\text{met-ch}}$ and $f_{\text{met-coord-ch}}$. The term $f_{\text{clash}}$ is an empirical heavy-
atom potential to avoid internal ligand clashes. The term $f_{\text{tors}}$ is the torsional
potential from the Tripos force field (32). Finally, if the centre of origin of the
ligand’s coordinate system lies outside the predefined binding area the
quadratic potential $c_{\text{site}}$ is evaluated.

In VINA, a hybrid scoring function is used including a mixture of empirical and
knowledge-based terms. There are three purely steric terms (two Gaussian
functions: gauss$_1$, gauss$_2$ and a repulsion term) as well as a term to consider
the hydrophobic interactions, hydrogen bonding and the number of rotatable
bonds. The product of these interaction terms and their weights for all
interacting atom pairs (inter- and intramolecular) are summed to give the
score as in Equation 3.3 (28).

$Vina = \sum_{i<j} f_{t_i t_j}(r_{ij})$

where each atom $i$ is assigned a type $t_i$ and interaction functions $f_{t_i t_j}$ of the
interatomic distance $r_{ij}$ are defined. The selection of scoring functions covers
all bases for different scoring function types (force-field, empirical and
knowledge based). The benefit of this should be a more well-rounded
consensus scoring for each docking solution.
3.3.2 Protocol

Figure 3.2: Consensus docking protocol. 1. ligands are docked in both GOLD and VINA under rigid protein conditions, solutions which agree within 2 Å RMSD are added to a consensus cluster which is used to define a smaller binding area. 2. The newly defined area is used for a flexible docking in both GOLD and VINA, this time solutions which agree within 2 Å RMSD between the two programs are ranked by consensus scoring after removal of redundancy.

In the first stage of the protocol, a fast but lower accuracy rigid docking, over a large search space was carried out in both GOLD and VINA for each ligand. Due to the fact that previous research suggests that these allosteric modulators bind in the transmembrane domain of the α7 nAChR (18, 24–26), the search space was limited to a sphere or box that encompassed only the inter- and intrasubunit cavities in the TM domain (Figure 3.3A). The root mean squared deviation (RMSD, a measure of similarity between the atomic positions) between each solution that was determined by VINA and those that were determined by GOLD was calculated (Equation 3.4).
\[ \text{RMSD}_{(j,k)} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( (j_{ix} - k_{ix})^2 + (j_{iy} - k_{iy})^2 + (j_{iz} - k_{iz})^2 \right) } \]

For two identical molecules with \( n \) atoms in two positions \( (j \) and \( k \)), the sum of the square differences between the \( i^{th} \) equivalent atom’s \( x \), \( y \) and \( z \) coordinates in each position is taken for all \( n \) atoms, divided by the total number \( n \) and the square root taken to give the RMSD (Appendices 1.1.1 and 1.1.2).

Where an RMSD of less than 2 Å was calculated for solutions from different docking programs (Figure 3.3C), the solutions were estimated to be the same and added to a ‘consensus cluster’. Once all ligands had been analysed, the consensus cluster for all ligands thought to be binding in the same location was manually investigated. As the first-pass docking treated the protein as rigid, it was assumed that interactions were not particularly well modelled and that any of the predicted solutions at this stage, regardless of score, could represent the true binding mode. If the distribution of solutions could fit into a sphere of radius 12 Å or less (i.e. relatively spherical), the centroid of the cluster was calculated, and a radius from this centroid was chosen that would incorporate the positions of all atoms in the solutions for use in a more focussed, flexible docking run (Figure 3.3B). Otherwise, multiple locations were assigned for the more focussed second docking run.

Figure 3.3: Consensus docking. A) Initial binding mode sphere used for docking in GOLD for a subunit pair in the revised α7 nAChR model, binding sphere shown in grey, principal subunit shown in khaki, complimentary subunit shown in a darker shade of green. B) an example of a consensus cluster from a rigid docking run and the docking sphere defined around this cluster, colouring as in A. C) an example of two binding mode solutions for the ligand TQS found to have an RMSD of less 2 Å. Solution from GOLD shown in yellow, solution from VINA shown in green.
In the second run, ten side chains (the maximum in GOLD) across the smaller binding site were selected to be flexible during the run. The side chains selected were based on several criteria:

- If they had been identified as influential to activity of allosteric modulators in mutagenesis studies and therefore potentially interacting with the ligands.
- If they had the potential to make polar interactions.
- An even distribution of flexible side chains across the binding cavity were selected, to prevent bias to a certain area of the cavity.

Once consensus solutions were detected after the second round of flexible receptor docking, the consensus solutions were scored by consensus of the three scoring functions discussed above, wherein, scoring consensus was attributed based on the Borda count (33, 34) (Equation 3.5, where $N$ models are ranked by $S$ scoring metrics and each model (i) is given a rank ($r$)). The top 5 solutions in each case were investigated for similarity between binding modes.

\[
Borda \text{ count } = \sum_{i=1}^{S} (N - r)
\]

However, prior to scoring, redundant solutions were removed from the ranking. For example, multiple solutions from GOLD could match\(^1\) a single VINA solution. In these cases, only the top scoring GOLD solution was taken as the lower scoring solutions are seen to be redundant (i.e. the same binding mode is already accounted for amongst the ranked solutions with a better score, this could also be the case with multiple VINA solutions matching a single GOLD solution).

### 3.3.2.1 Atom assignment problem

A problem that was encountered when developing this protocol was that GOLD and VINA used different file formats for the ligand, during docking. While GOLD employs the Tripos MOL2 format and included all hydrogen

\(^1\) Where “match” means, has an RMSD of less than 2 Å between a given GOLD solution and a given VINA solution
atoms, as well as electron lone pairs during docking, VINA utilised the PDBQT format and sacrificed non-polar hydrogens in the interest of speed. The impact of this was that the atoms in each ligand file were assigned in different orders. No freely available, predefined method was found that could calculate RMSD between sets of atoms, considering the different order of assignment and preserving the atom coordinates. Therefore, utilising the programming language Python (35), a script was written that was able to determine the RMSD between molecules with atoms assigned in different orders (Appendices 1.1.1 and 1.1.2).

The basis of the RMSD determination was to employ the Kuhn-Munkres algorithm (36–38) (also known as the Hungarian algorithm) to solve the “assignment problem” for atoms of the same SYBYL atom type in a .mol2 file. Originally developed as a method for assigning personnel to tasks in the most efficient manner by minimising a cost matrix (36), the Hungarian algorithm has subsequently been applied to determining RMSD in molecular docking in the software package DOCK (37). The advantage of the Hungarian algorithm, is that it is able to take into account areas of symmetry in a molecule (37) without having to calculate the n! ways in which an n x n matrix could be assigned (38). Figure 3.4, left, shows how determination of RMSD between two molecules, can be adversely affected by rotation about a plane of symmetry. When correctly calculated, the determined RMSD between the two symmetrical molecules is 0.300 Å, however, if the orange molecule is rotated 180° about the plane of symmetry (which should result in the same RMSD due to the inherent symmetry within the compounds), due to the incorrect assignment of equivalent atoms, the RMSD is falsely calculated to be 1.901 Å. In the case of determining RMSD between two sets of atoms using the Hungarian algorithm, an all-against-all distance matrix for the atoms becomes the cost matrix, and the assignment of atoms in each molecule that leads to the lowest RMSD, the solution. Since atoms of different types need not be matched, the all-against-all atom distance matrix can be broken down into individual matrices for the constituent atom types in each molecule. An example cost matrix for a theoretical molecule containing 4 SYBYL atom types (O.2, Cl, C.ar and N.pl3) is shown (Figure 3.4, right). This speeds up
the assignment of atoms in the matrix as the length of time to solve the algorithm scales by a polynomial factor of \( n \) where \( n \) is the length of one dimension of the square matrix (38), so several smaller matrices are calculated much faster than one larger matrix. Furthermore, by limiting matching to only atoms of the same type, this helps to ensure fidelity in determining RMSD.

Figure 3.4: Using the Hungarian algorithm in RMSD calculation. Left) Two theoretical docking solutions of a symmetrical small molecule have similar modes, when all equivalent atoms are superposed (top left) the RMSD is 0.300 Å. Rotating one molecule 180° about the plane of symmetry (dashed line top left), should result in the same RMSD, however, due to the different placement of atoms, RMSD is calculated as 1.901 Å. Right) Size of matrices that have to be calculated for all atoms (middle) vs. on division into atom types (right), constraining the Hungarian algorithm to atoms of the same SYBYL type significantly decreases the overall size of the matrix that has to be computed. (figure adapted from (37)).

In brief, the Hungarian algorithm solves the “assignment problem” in four steps (37):

- A cost matrix \( M \) is formed, for RMSD calculation, each element of the matrix \( m_{ij} \) is the squared Euclidean distance between atoms \( a_i \) and \( b_j \). The lowest value in each row/ column is subtracted from all other elements in the same row/column.
- The minimum number of horizontal and vertical lines that will cover all zeroes in the matrix is found. If the number of lines is equal to the size of the matrix in one dimension, then a set of independent zero-values exists. The positions of the zero-values in the matrix corresponds to the optimal assignment of atoms for the lowest RMSD. Otherwise perform the next step.
- Determine the smallest value matrix element that is not covered with a line. Subtract that value from all matrix elements that are uncovered
and add it to any matrix element that is covered by both a vertical and a horizontal line.

- Repeat steps 2 and 3 until the matrix correspondence of atoms is solved for the matrix as described in step 2.

The advantage of using such an atom assignment process is that atom equivalence can be assigned in non-identical molecules allowing the convenient calculation of RMSD between compounds which do not have the exact same structure. Therefore, the clustering of non-identical compounds by binding mode similarity is possible, so long as atoms of the same type are present in each molecule.

### 3.3.3 Preparation of ligand files

For each of the ligands to be docked, a three-dimensional (3D) file must be generated for use with the docking program of choice. Certain considerations are important for the outcome of the docking run in the preparation of the 3D ligand files.

When generating 3D files from 1D SMILES strings or 2D drawings, it must be ensured that allowable geometries are obtained for bond lengths and bond angles. While docking programs often explore different conformations of a ligand during the docking run, the starting conformation of the ligand can be influential in the solutions that are obtained, hence low-energy conformations are best as starting conformations for ligands where the known binding conformation is not known. Where stereochemistry exists, it is very important that the correct geometry is obtained. Since proteins are built from chiral amino acids, different isomers of the same compound will not necessarily be able to make the same interactions. Likewise, the charge state of a compound must be correctly assigned. This is often dependent on the pKa of basic and acidic moieties and therefore the protonation state of the compounds.

The method of choice taken here was to draw each compound in the software package ChemDraw Professional 15.1 (39), copy the isomeric smiles to a text file and then utilise the python package RDKit (40) to generate 3D coordinates.
and save the ligands in a `.sdf` file (Appendix 1.1.3). The RDKit 3D coordinate generator utilises the universal forcefield (UFF) for molecular mechanics to optimise the geometry of the generated 3D coordinates. None of the docked compounds contained moieties that were likely to be protonated or deprotonated at physiological pH, therefore all ligands were generated as their neutral species.

### 3.3.3.1 Ligand stereochemistry

The main complication in the generation of these ligand files is the existence of stereochemistry in the TQS-family compounds. Based around a tetrahydroquinoline core, these compounds contain three quaternary carbon centres which each have four unique substituents and are thus chiral. Although four unique diastereomers exist, leading to eight different possible enantiomers (Figure 3.5), exclusively cis-cis$^2$ and trans-cis$^3$ diastereomers are produced in the reaction to form these compounds, as previously determined by calculation of the corresponding $^1$H NMR coupling constants (41). Only cis-cis compounds have been found to act as type II PAMs (42, 43). This means that two possible enantiomers could lead to the pharmacological activity of the TQS-family of compounds. Using the Cahn-Ingold-Prelog priority rules, the stereochemistry of the three chiral carbon centres in TQS-type compounds acting as type II PAMs can either be (S, R, S) or (R, S, R) (Figure 3.5). In one study where the two enantiomers of a TQS-type ago-PAM$^4$ (GAT107) were separated by chiral HPLC, it was found that only the (S, R, S) enantiomer contributed to the activity of these compounds (43).

---

$^2$ all H atoms on the same side of the tetrahydroquinoline ring
$^3$ H at the 2-position of the tetrahydroquinoline on the opposite side of the ring system to the H atoms at the 3- and 4-positions
$^4$ ago-PAMs are able to stimulate nAChRs in the absence of orthosteric ligands additionally to producing PAM effects
To validate this finding and see if the same stereochemistry is causative of activity in other TQS-type compounds, moreover those that are active only as type II PAMs, TQS was synthesised and a chiral resolution attempted. The synthesis of TQS was carried out following the same reaction procedure used in the patent that originally disclosed the synthesis and activity of these compounds, using an indium trichloride catalysed three component Povarov-type cyclisation (44). The reaction proceeded smoothly at room temperature to give TQS in a yield of 96% (Scheme 3.1). The major product of this reaction is the cis-cis diastereomer, with a small amount of the trans-cis diastereomer present (~6%). Recrystallization from a 1:1 mixture of pyridine and methanol almost completely eliminated the trans-cis isomer from the mixture.

The recrystallized TQS was submitted to preparative chiral HPLC, as in the chiral resolution of GAT107 (43). The chiral column utilised provided an
excellent separation of the two enantiomers in retention time with peaks corresponding to the two enantiomers eluting at 20 mins 30 s and 32 mins 42 s (Figure 3.6, upper). However, limitations in the available equipment meant that the isolation of the purified enantiomers was not successful. The main problems stemmed from the solubility of TQS and the solvent compatibility of the chiral column. The Chiralcel AD type column that was available, was compatible only with solvent mixtures of isopropanol and hexane, up to 40% isopropanol. However, TQS is only slightly soluble in isopropanol and could only be solvated at a concentration of approximately 1 mg mL\(^{-1}\). With a maximum injection volume of 1 mL the HPLC machine could purify a theoretical maximum of 0.5 mg of TQS enantiomer per 40 minute run. Unfortunately, no dimethylsulfoxide (DMSO) could be used to increase the concentration of TQS in the injected sample, as even small amounts of DMSO could dissolve the chiral stationary phase in the column that was available. Furthermore, without an autosampler all samples had to be collected manually via the waste tube making the separation incredibly time intensive and prone to operator error.

Combined fractions were reanalysed by chiral HPLC to detect the degree of chiral separation achieved. Despite the significant separation in retention time, it was found that the samples still contained a mixture of both enantiomers (Figure 3.6, middle) with fraction one having an enantiomeric excess of 48% and fraction two, an enantiomeric excess of 71%. This may have occurred due to the waste tube carrying residue of the opposite enantiomer during sample collection. Due to the very small portions isolated and the low enantiomeric purity, these fractions were not tested in electrophysiology as it would not have given a conclusive result on which enantiomer contributed to the activity of the compounds. However, a crystal structure (determined by Dr Krešo Bučar) was obtained for fraction two, which was of higher enantiomeric purity (ee 71%), and suggested the enantiomer present in a greater amount in that fraction may be the \(S,R,S\) enantiomer (Figure 3.6, bottom).
Figure 3.6: Chiral HPLC of TQS. Top) representative chiral HPLC trace of separation of TQS cis-cis enantiomers. The peak eluting at 10 minutes is due to a small amount of organic compound being carried through the chiral column faster than the bulk of the compound due to the extra polarity of the loading solvent vs. the elution solvent. Middle) chiral HPLC traces for the analysis of enantiopurity in fraction one (left) and fraction two (right). The step change in baseline absorbance at two minutes is due to a change in flow rate after the addition of TQS onto the chiral column. Bottom) the geometry observed in the crystal structure obtained for fraction two showing the (S,R,S) conformation.

Despite the absence of further data to confirm whether only (S,R,S) enantiomers of the cis-cis TQS-type compounds were biologically active, this assumption was applied to all ligands in the TQS family during docking, based on the finding relating to GAT107 and only (S,R,S) enantiomers were docked.
3.3.4 Docking of A-867744, TBS-516 and TQS

Initially, the three ligands of interest (A-867744, TBS-516 and TQS) were docked individually to the receptor using the above described consensus docking protocol. Whereas previously TQS had been proposed to bind at an intrasubunit cavity (24), only binding modes at the intersubunit cavity were detected as consensus solutions (Figure 3.7). However, even amongst the top ranked solutions after consensus scoring, there was significant variation in the predicted binding modes (Figure 3.7). This made the results difficult to interpret with respect to the functional data as there were no obvious patterns that could be rationalised to lead to the differences seen in functional activity data for these compounds (Table 3.1). This is possibly a consequence of these ligands binding in the transmembrane domain where the polarity is almost entirely hydrophobic. Thus the majority of binding interactions are non-specific hydrophobic interactions which often are not well modelled in docking and as such make it difficult to appropriately score a set of binding mode predictions in order to choose the most probable binding mode.

Figure 3.7: Top ranked binding modes from consensus docking of A-867744, TBS-516 and TQS in the α7 nAChR model. Principal subunit shown in khaki, complimentary subunit shown in dark green. Sidechains are shown in stick representation for positions where mutagenesis data exists for these compounds. A-867744 (purple), TBS-516 (orange) and TQS (cyan) are shown in ball and stick representation.

Aiming to detect binding modes consistently predicted across compounds expected to bind in the same way, compounds derived from structure activity relationship (SAR) studies (8, 11, 19, 42) and shown to have the same functional properties as their parent compound (i.e. type II PAMs, Appendix 1.1.4) were also included in the docking. The top five binding modes from each of the docked ligands were clustered against those of all other ligands in
the same SAR family and the largest clusters taken as the most likely binding mode solution. Thankfully, the development of a python script that determines RMSD using the Hungarian algorithm to assign atom equivalence lent itself well to clustering of non-identical molecules.

Additionally, the compounds were docked to models of α7 nAChR thought to represent both open and closed conformations of the receptor. Since preapplication of the PAMs is required for immediate onset PAM action during coapplication with acetylcholine, it is expected that these PAMs are capable of binding to the closed conformation of the receptor. Similarly, since the action of these PAMs is observed during opening of the channel, it is also expected that they should be able to bind to the open conformation of the receptor. Independent docking to both receptor models offers an interesting opportunity for comparison of the predicted binding modes.

In total, docking was performed with a series of 37 compounds that were derived from structure-activity-relationship studies surrounding either A-867744 (8 compounds (45)), TBS-516 (4 compounds (18)) or TQS (25 compounds (19, 42)), all of which are known to display type II PAM or ago-PAM activity on α7 nAChRs (for details of the selected compounds see Appendix 1.1.4). Initially, rigid docking was performed with models of the human α7 nAChR in both closed and open conformations and the results obtained with two different docking programs, that have been shown to produce reliable results (46) (GOLD (31) and AutoDock VINA (28)), were compared. With GOLD, the search area was defined such that all receptor residues within a sphere of 18 Å radius from the γ-carbon of T277 of the (+) subunit in the closed model were included. With AutoDock VINA, all receptor residues within a cube of 32 Å centred on T277 of the (+) subunit were included. This search area spanned a region covering both the inter-subunit and intra-subunit cavities present in the transmembrane domain. The search efficiency was set to the maximum in both software packages and the ligands were allowed full flexibility. Fifty diverse solutions were generated in GOLD and the maximum of 20 solutions was enabled in AutoDock VINA (via 20 runs). The root-mean-square deviation (RMSD) between the GOLD and
AutoDock VINA solutions for each ligand was determined using an in-house python script based on the Hungarian algorithm (37) (Appendix 1.1.2).

Next, a centroid was defined based on the location of all atoms in the solutions that had an RMSD of 2.0 Å or less between GOLD and AutoDock VINA. From this centroid, a region was determined that included the positions of all the solutions in the subset (a sphere in GOLD and a box in AutoDock VINA). Ten amino acid side chains were selected for full flexibility within the defined docking areas in both cases. Priority was given to making residues flexible that had been mutated in this study, followed by polar residues and, finally, an even distribution of flexible residues with the aim of preventing any bias to particular regions of the selected binding area. All ligands were docked using the same search parameters that were used for docking studies with the rigid receptor (see above). RMSD filtering was carried out as previously, to remove solutions that had an RMSD greater than 2.0 Å between any of the GOLD and AutoDock solutions. The remaining solutions after redundancy removal (up to 20 pairs) were ranked based on the Borda score (33, 34) using three metrics: the AutoDock VINA scoring function, and two scoring functions implemented in GOLD (GoldScore and ChemPLP). Following the docking simulation of all ligands, the top-five ranked solutions from each chemical class were clustered against each other (based on the GOLD pose) by hierarchical RMSD clustering using an in-house script (Appendices 1.1.1 and 1.1.2).
Figure 3.8: Largest binding mode clusters found for A-867744, TBS-516 and TQS after consensus docking into open and closed conformation receptors. Principal subunit (+) shown in khaki, complimentary subunit (-) shown in darker shade of green. Amino acids mutated during the study of these compounds are shown as grey spheres, ligand clusters are shown in black ball-and-stick representation with heteroatom colouring.

The latter step was performed in order to identify binding modes predicted consistently across multiple compounds with a similar chemical structure and therefore expected to bind in a similar way. For each of the three classes of arylsulfonamide type II PAMs, the most highly populated clusters were assumed to represent the most probable location of the binding site. This made it possible to choose binding modes from those predicted in an unbiased manner. Overall the binding mode clusters for TBS-516 and TQS appear generally similar while those for A-867744 were different, but predicted to be at an overlapping and therefore mutually exclusive site (Figure 3.8). The highest ranked mode predicted for A-867744, TBS-516 and TQS were extracted from each cluster and taken as representatives (Figure 3.9).
3.3.4.1 Predicted binding of TBS-516 and TQS

Figure 3.9: Binding mode cluster representatives for A-867744, TBS-516 and TQS docked into α7 structural models. Following docking studies with the closed and open models (upper and lower panels, respectively), representatives of binding mode clusters are illustrated for A-867744 (A and D; purple), TBS-516 (B and E; orange) and TQS (C and F; cyan). Amino acids that are discussed in the text are shown in stick representation. Predicted hydrogen bonds are shown with dashed lines. In each case the principal subunit (+) is shown in khaki, (on the left in each panel) and the complimentary subunit (-) is shown in olive (on the right in each panel).

The predicted clusters for TBS-516, TQS and their related compounds occupy topologically similar binding modes in both closed and open receptor models (Figure 3.8). In the closed conformation, potential hydrogen bonds between the arylsulfonamides of both compounds with the residues S284 and T288 at the cytoplasmic end of TM3 (Figure 3.9B, C, E) were observed. These residues are involved in the TM2-TM3 interface, possibly forming part of the desensitisation gate of pLGICs (47) (Figure 3.10).
Figure 3.10: Desensitization gate residues on α7 nAChR. Image shown from within the membrane looking at the TM2-TM3 interface of the principal (+) subunit. The Cα atoms of residues predicted to be interacting with PAMs and also involved are shown as coral spheres.

Consequently, such interactions may help to explain changes in the rate of desensitisation that are observed with these type II PAMs. The naphthyl ring of TQS sits in a hydrophobic pocket formed by the side chains of the amino acids L247, S248 and F252 of the (-) subunit and L246, L247, T250, and M253 of the (+) subunit, and is situated ~6.5 Å from M253 and ~5.0 Å from L247, both of which are amino acids that upon mutation have been shown to influence the PAM effects of these compounds (Table 3.1). In the TBS-516 compounds, a similar pocket is occupied by either the benzyl ring (4.0 Å to M253 and 3.5 Å to L247) or the bromophenyl ring (7.2 Å to M253 and 3.5 Å to L247) in the closed and open conformations, respectively. The close proximity of TQS and TBS-516 to M253 and L247 may explain their observed pharmacological effects on the respective mutant receptors (M253L, L247T).

In the open conformation, the TQS naphthyl ring remains in the same hydrophobic pocket (Figure 3.9F). Unlike TBS-516, the arylsulfonamide of TQS has moved away from S284 and T288, no longer forming hydrogen bonds with these residues (Figure 3.9F), but is within ~3.5 Å of G242, another residue that has been suggested to contribute to the desensitisation gate of pLGICs (47) (Figure 3.9F and Figure 3.10). Despite this observation, the cluster of binding modes for TQS predicted in the open conformation covers a broader area than that predicted in the closed conformation (Figure 3.9C vs. F), many of the other members of the cluster maintain hydrogen bonding interactions with S284 and T288, as well as forming hydrogen bonds to the
protein backbone at G242, and so it is plausible that direct interactions with the desensitization gate may too exist in the open conformation. For TBS-516, the bromophenyl ring sits in the same hydrophobic pocket as the naphthyl ring of TQS (Figure 3.9E), while the third variable group (a benzyl group), protrudes onto a different hydrophobic surface, formed by L246, T250, M253, M278 and V281 (Figure 3.9E).

### 3.3.4.2 Predicted binding of A-867744

The predicted binding modes of A-867744 and related compounds are in a similar location to those of TBS-516 and TQS, but have subtle yet important differences (Figure 3.9). Most obviously, whereas the arylsulfonamide group of TBS-516 and TQS clusters is orientated downwards in both the open and closed receptor conformation, it is orientated upwards for the A-867744 cluster (Figure 3.9A, D). In the closed conformation, the arylsulfonamide group of A-867744 forms a hydrogen bond with S248 of the (-) subunit (an amino acid predicted to interact with both TBS-516 and TQS, Figure 3.9A). However, the orientation of the molecule is different to that of TBS-516 and TQS, situated between the TM2 and TM3 helices of the (+) subunit, with a protrusion into the intra-subunit cavity, ~7.0 Å from S222 (another amino acid shown to influence the activity of A-867744 through mutation (Table 3.1, Figure 3.9A). TBS-516 and TQS are predicted to lie much further from this residue in the closed conformation (~14.0 Å, Figure 3.9B, C), which may explain how the S222M mutation has a dramatic effect on A-867744, but not TBS-516 or TQS (Table 3.1).

Interestingly, in the open conformation, the predicted binding mode for A-867744 is different from its closed conformation counterpart (closer to the extracellular domain; Figure 3.9D). The arylsulfonamide appears to form a hydrogen bond with T277, which is markedly different from the interactions of the analogous arylsulfonamide in TQS or TBS-516 (Figure 3.9D vs. E, F), however, it is still located in the TM2-TM3 interface, orienting the rest of the molecule to perhaps also influence the possible desensitisation gate. The phenyl ring of the arylsulfonamide is located on the same hydrophobic surface as described for TBS-516 (Figure 3.9D). The chlorophenyl ring is also
predicted to be embedded in the same hydrophobic pocket as that described for the naphthyl ring of TQS, close to L247 and M253 (Figure 3.9D).

**3.3.5 Discussion of predicted binding modes**

The pharmacological properties of three type II PAMs (A-867744, TBS-516 and TQS) were examined by Anna Chatzidaki in the Millar lab on five different mutants of α7 nAChR (W54A, S222M, L247T, M253L and M260L). In all cases, effects on the action of A-867744 were different from those observed with two other arylsulfonamide type II PAMs (Table 3.1). This is consistent with previous studies that have suggested that A-867744 is an α7 PAM with unusual properties (20). While A-867744 could be purchased from commercial sources, TBS-516 and TQS were synthesized via published methods (18, 21) (see Chapter 8 – Experimental for synthetic methods and characterization).

Numerous α7 mutations have been reported to alter the properties of orthosteric and allosteric ligands. Although some, such as M260L, have differing effects on type I and type II PAMs (18), mutations generally appear to have consistent effects on compounds that have similar effects on wild-type α7 nAChRs, such as type II PAMs (18). The identification of mutations that have differing effects on A-867744 than on other type II PAMs, provides evidence that A-867744 has an atypical mechanism of action. An explanation might be that A-867744 binds at a different allosteric site on α7 to the binding site of other PAMs. This might be due to the PAMs interacting differently at a common or overlapping site, or by binding to unrelated allosteric sites. However, binding to a common site would be consistent with the observation that A-867744 is able to block the effects of the other PAMs on these mutants and that pre-application of TQS on the M253L mutant is able to block the inhibitory effects of A-867744 (see (48) figures 3, 4, 5, 7).

All three groups of type II PAMs were predicted to bind in the inter-subunit transmembrane region (Figure 3.8, Figure 3.9). This is different from what has previously been predicted for some allosteric modulators of α7 (such as LY-2087107, NS-1738 and PNU-120596) where an intra-subunit site was proposed (24–26) (Figure 3.11B). The difference from the previous studies
could be attributed to the fact that they were conducted with α7 comparative models based upon an uncorrected model of the *Torpedo* nAChR. Consensus docking studies with LY-2087107, NS-1738 and PNU-120596 in the new α7 comparative model found that these PAMs are also predicted to bind in an inter-subunit site that matches the contour of the TQS binding mode cluster (Figure 3.11A).

![Figure 3.11: Predicted transmembrane binding of LY-2087107 (pale green), NS-1738 (peach) and PNU-120596 (light blue). A) New intersubunit binding mode predictions. Principal subunit shown in khaki, complimentary subunit shown in a darker shade of green. The transparent contour of the predicted TQS-family binding mode cluster is shown. B) Previously predicted intrasubunit binding area for LY-2087107, NS-1738, TQS and PNU-120596 (24–26).](image)

The prediction of an inter-subunit binding site for α7 allosteric modulators is consistent with the fact that another allosteric modulator (ivermectin) is known to bind at the inter-subunit cavity, based on co-crystallization with two different pLGICs (49, 50). In addition, an inter-subunit transmembrane binding site has been identified on the basis of affinity-labelling studies with purified *Torpedo* nAChRs and photoreactive allosteric modulators (51, 52).

It is interesting that a mutation in the extracellular domain (W54A) influences the activity of these PAMs (presumably a consequence of long-range effects on receptor structure), yet it has a different effect on A-867744 than on TBS-516 and TQS (Table 3.1). This is congruent with previous evidence indicating, that A-867744 can cause the dissociation of [3H]-A-585539 from the orthosteric site (12), suggesting that there can be cross-communication between the orthosteric and allosteric sites.
The docking studies appear to be consistent with the experimental findings (Table 3.1). Docking studies suggest that A-867744 binds in an orientation that is distinct from that of both TBS-516 and TQS (most apparent in the closed model; Figure 3.9A). This provides a basis for understanding how both proximal and distant mutations affect the function of A-867744 differently to how they affect TBS-516 and TQS. However, despite these differences, all three compounds are predicted to bind in a broadly similar transmembrane location that is in close proximity to residues that have been implicated in the desensitisation gate of pLGICs (47).

It is notable that some of the mutations examined (S222M, L247T and M253L) convert A-867744 into an antagonist but do not have this effect with either TBS-516 or TQS (Table 3.1). In the closed conformation, A-867744 is predicted to bind with a moiety protruding between the TM2 and TM3 helices of the principal subunit, whilst forming a hydrogen bond with the complimentary subunit (Figure 3.9A and Figure 3.12). Structures of pLGICs caught in different receptor conformations (e.g. *Torpedo* nAChR (53), GlyR (50), GLIC(54, 55)) suggest that rearrangements occur at the principal-complimentary interface. This may allow A-867744 to act as a physical barrier to the rearrangement of the transmembrane helices necessary for channel gating to occur. It is possible that some transmembrane mutations may increase the energy barrier for transition between closed and open binding modes of A-867744 (i.e. either raising the energy of the open state binding mode, lowering the energy of the closed state binding mode, or increasing the energy required to move between the two predicted binding modes). This could cause the observed inhibition of agonist-evoked responses as A-867744 gets stuck in its closed conformation mode which blocks channel opening. This hypothesis might make sense given that the predicted binding modes for TQS and TBS-516 have very high similarity between open and closed conformations of the receptor. Alternatively, A-867744 may be acting as an open-channel blocker on receptors containing some transmembrane mutations. However, evidence that other PAMs can block the inhibitory effect of A-867744 (e.g. Fig. 5D (48)) provides support for the conclusion that
positive and negative modulatory effects can occur through a broadly similar mutually exclusive binding site.

Figure 3.12: Inter-subunit transmembrane binding site for PAMs on the α7 nAChR. A) Representation of the transmembrane domain viewed from above, looking down the axis of the channel pore. Black ellipses indicate the location of the inter-subunit allosteric binding site identified in this study. B) Predicted binding modes in the closed and open receptor models of A-867744 (purple), TBS-516 (orange), and TQS (cyan), shown in relation to transmembrane helices (grey rods) from the principal (+) and the complementary (-) subunit interface. The locations of predicted hydrogen bonds are shown with chain links to denote anchoring of the ligands within the binding site. The locations of amino acids M253, L247 and S222 are shown as yellow, blue and green circles, respectively. An arrow at the top of TM2 from the complementary subunit denotes the motion required for the change in conformation from the open to the closed channel.

3.4 Conclusions and future work

In summary, consensus docking in the transmembrane regions of new structural models of the human α7 nAChRs reveals how PAMs such as A-867744, TBS-516 and TQS may interact with an inter-subunit transmembrane site rather than an intra-subunit site as previously thought. The predicted binding mode solutions provide an explanation of the pharmacological diversity between this group of chemically similar allosteric modulators. While similar type II PAMs TQS and TBS-516 are predicted to bind in similar ways, A-867744 (which behaves distinctly on mutant receptors) is predicted to bind in a different manner. Furthermore, the binding mode solutions are able to offer a rationalization of how A-867744 may be able to act as an inhibitor. By blocking the rearrangement of transmembrane helices at the subunit interface
in cases where mutants may prevent the apparent movement of A-867744 binding mode, as the receptor moves from closed to open conformations. This conclusion could not have been drawn without docking into both open and closed models of the α7 nAChR.

Future work to validate these binding modes is important. For example, if these binding modes can be used to predict properties of the receptor, or new PAMs that would provide a form of validation. In subsequent chapters I will describe the development of reactive allosteric modulators for the α7 nAChR, wherein the proposed binding modes could be used to predict sites of reactivity and a ligand based virtual screening of a small chemical library based on the predicted binding modes of TQS. Discovery of novel PAMs would suggest that the shape of the binding mode for these modulators may be correctly captured by the docking studies carried out in this chapter.
3.5 References


10. Dinklo T, et al. (2011) Characterization of 2-[[4-Fluoro-3-
(trifluoromethyl)phenyl)amino]-4(4-pyridinyl)-5-thiazolemethanol (JNJ-1930942), a Novel Positive Allosteric Modulator of the α7 Nicotinic Acetylcholine Receptor. *Pharmacology*: 560–574.


Neuropharmacology 97:75–85.


to account for ligand symmetry and similarity in structure-based design. 


Chapter 4

4 Design and synthesis of reactive allosteric modulators of α7 nicotinic acetylcholine receptors
4.1 Introduction

In light of the binding mode predictions made for TQS and TBS-516 in the previous chapter, it was a priority to develop methods for validating these predictions. One method of validation is to develop a photoactivatable PAM capable of cross-linking to the receptor. Permanent modulation of receptor activity after photoactivation could be indicative of successful crosslinking for example in the way that photoreactive gabazine analogues have been used with GABA receptors (1). The residues to which a photoactivatable modulator might crosslink could also be determined by digestion and peptide sequencing, such as Edman degradation, as has been previously carried out for Torpedo nAChRs (2). However, this would require expression of a relatively large amount of functional protein in an appropriate cell type and optimisation of this expression is beyond the scope of this PhD project. The relatively recent discovery that the transmembrane chaperone NACHO aids expression of α7 nAChRs in HEK cells (3) and is synergistically boosted by the promoter RIC-3 (4) suggests that good yields of α7 nAChR expression in eukaryotic cells is plausible. Further to this, there have been reports of “crystallization scale purifications of α7 nicotinic acetylcholine receptor from mammalian cells” (5). Thus, production of nAChRs in large enough quantities for cross-linking studies should be achievable in the near future.

Alternatively, given the improved understanding of the PAM binding site, a reactive group selective for cysteine nucleophilic attack could be incorporated. Cross-linking with a cysteine reactive group could be tested in mutated receptors in which cysteine is genetically engineered into the expressed protein at sites thought to be capable of cross-linking with the position of the reactive group in a docked binding mode prediction.
4.2 Aims

Develop a novel photoactivatable PAM for use in photo-crosslinking studies with α7 nAChRs.

Develop PAMs capable of crosslinking to nAChRs containing engineered cysteine residues at positions likely to be able to react with the PAM based on binding mode predictions.
4.3 Results and discussion

4.3.1 Design of reactive allosteric modulators

4.3.1.1 Photoreactive analogues

TQS-type compounds offer an excellent opportunity for the incorporation of a photocrosslinking moiety. Their modular synthesis via a multi-component Povarov-type cyclization (6) allows the easy incorporation into the final molecule of substituents installed on a benzaldehyde. Furthermore, previous studies of structure-activity relationships suggest that a large variety of substituents can be tolerated at the 4-position of the benzaldehyde phenyl ring, whilst maintaining activity of the compounds (7). As such, one could envisage the synthesis of a reactive TQS-type compound, by installing a cross-linking moiety at the 4-position of a benzaldehyde, and then subjecting the substituted benzaldehyde to the multi-component reaction (Scheme 4.1).

Scheme 4.1: Possible photoreactive analogues of TQS. From left to right, R = azide, trifluoromethyl diazirine, benzophenone.

Azides, diazirines and benzophenones (Scheme 4.1) are commonly used for the purpose of photo-crosslinking as they offer high levels of activation to their reactive counter parts upon irradiation (nitrene, carbene and O-radical respectively) and rapid incorporation into the protein upon activation. Trifluoromethyl diazirines are one of the most efficient photo-crosslinking
groups. The only by-product of activation is the loss of dinitrogen, meaning that the by-products are unlikely to cause any great perturbation of the natural environment of the protein. Photoactivation of diazirines can be conducted at wavelengths of light longer than 350 nm, which compared with activation wavelengths for azides (typically shorter than 350 nm) are less likely to cause ionisation elsewhere in a protein (e.g. direct photo-oxidation of aromatic sidechains and disulfides or production of reactive singlet oxygen species). Finally, the adjacent trifluoromethyl group serves to both increase the reactivity of the resultant carbene upon activation (by acting as an electron withdrawing group) and also to prevent the deactivation of the resultant carbene by forbidding the 1,2-shift of a hydride and formation of an alkene, as is possible in the methyl analogue (Scheme 4.2).

Scheme 4.2: deactivating 1,2-hydride transfer in methyl diazirines.

The 4-azidoTQS analogue has been previously synthesised. However, under UV irradiation there was no evidence of degradation (8).

4.3.1.2 Cysteine reactive analogues

As the most nucleophilic naturally occurring amino acid, reactivity can be tuned towards cysteine by creating compounds with weakly electrophilic substituents. Examples could include, for instance, compounds bearing a benzyl chloride, or an α-chloroacetate (Figure 4.1). These groups may not react as efficiently with cysteine residues as for example α-iodoacetates, which have commonly been used to label cysteine residues. However, the lower reactivity could help to favour reaction only in positions where the ligand has a long residence time, i.e. its binding site, rather than promiscuously labelling any free cysteine in the protein. Site specific engineering of cysteines at strategic positions in the transmembrane domain of α7 nAChRs based on the predicted binding modes may allow covalent incorporation of the
cysteine reactive TQS compounds. Covalent incorporation could potentially be tested by electrophysiology alone, manifesting as a permanent modulatory effect observed with cysteine containing mutants, but not wild-type receptors.

![Chemical structure](image)

Figure 4.1: Examples of cysteine reactive TQS-type compounds. Top) benzyl chloride, bottom) α-chloroacetate.

### 4.3.2 Synthesis of reactive TQS analogues

![Chemical structures](image)

Figure 4.2: Reactive analogues of TQS, from left to right, 4TFD-TQS, 4CM-TQS and 4ACA-TQS.

The three designed reactive TQS analogues (Figure 4.2) were synthesised from simple starting materials. 4TFD-TQS was synthesised via a six step route, culminating in a 20% overall yield (Scheme 4.3). The synthetic route followed a series of previously well-defined reactions for the synthesis of a trifluoromethyl diazirine (9). However, due to the sensitivity of the aldehyde portion (of the 4-substituted benzaldehyde multi-component reaction precursor) to reaction conditions used in forming the trifluoromethyl diazirine,
it was installed towards the end of the synthesis using a methodology with very harsh conditions described by Nakashima et al. in 2006 (10).

Scheme 4.3: Synthesis of 4TFD-TQS.

The addition of subsequent steps after the formation of the diazirine ring system is not ideal owing to the photoreactive nature of the three-membered heterocycle. However, the diazirine containing compounds appeared to be relatively stable, so long as they were kept in the dark and the final TQS analogue 4TFD-TQS could be synthesised in a 54% yield from the diazirine containing benzaldehyde.

The synthesis of the cysteine reactive TQS analogues was more straightforward. 4ACA-TQS was synthesised from p-nitrobenzaldehyde in three steps in an overall yield of 37% (Scheme 4.4). The amine was disguised as a nitro group in order to avoid self-reaction of the benzaldehyde starting material (to form an imine), and selectivity issues due to competing anilines in the multi-component reaction. The synthesis of a 4-nitro analogue of TQS proceeded in an excellent yield and was isolated with a high diastereomeric ratio after purification (96% dr >20:1). The nitro group was selectively reduced to an aniline, without any evidence of reduction of the alkene on the cyclopentene fragment, using an iron and zinc based reaction (11). Finally,
the cysteine reactive component was added via the reaction of the primary aniline and chloroacetyl chloride.

Scheme 4.4: Synthesis of 4ACA-TQS.

4CM-TQS was synthesised from 4-(chloromethyl)benzonitrile in two steps equating to an overall yield of 59% (Scheme 4.5). 4-(chloromethyl)benzonitrile was reduced selectively to the benzaldehyde using diisobutylaluminium hydride (DIBAL) as 4-(chloromethyl)benzaldehyde was not available to purchase. The synthesised benzaldehyde smoothly underwent the multi-component reaction to afford 4CM-TQS in a 65% yield and was isolated with a diasteromeric ratio of greater than 30:1.

Scheme 4.5: Synthesis of 4CM-TQS.

4.3.3 Testing of TQS analogues

Initial UV irradiation tests were carried out on 4TFD-TQS by dissolving a small amount of the substance in deuterated chloroform, irradiating the sample...
using a UV torch for 5 minutes and then subsequently analysing the irradiated sample by NMR for the appearance of new peaks in its NMR spectrum. After irradiation, the solution became cloudy, suggestive of the precipitation of activated 4TFD-TQS that has polymerised. Within the soluble fraction that was analysed by NMR, new peaks appeared in the spectrum. Unfortunately, all regions in which new peaks were identified overlapped with the peaks of the pure 4TFD-TQS so quantification was difficult, although it was estimated that the peaks were present in a 1:2.5-3 ratio of the new peaks to the old (Figure 4.3). However, this quantification is additionally hampered by the observation that an unknown amount of the compound precipitated from solution upon irradiation. Nevertheless, it is indicative that the diazirine is rapidly broken down under short periods of irradiation.

Figure 4.3: NMR spectrum of UV irradiated sample of 4TFD-TQS in deuterated chloroform. Inset, regions where newly appearing peaks with minimal overlap on peaks of the pure compound (4TFD-TQS) used to estimate the amount of decomposed compound in solution.

4TFD-TQS and 4CM-TQS were tested in voltage clamp electrophysiology on Xenopus laevis oocytes expressing α7 nAChRs by Dr Anna Chatzidaki in the Millar lab. When applied at a concentration of 10 μM and coapplied with...
acetylcholine (100 μM) both compounds were found to act as allosteric agonists with type II PAM effects (Figure 4.4). Confirming that the active properties of these compounds were retained after the addition of reactive groups at the 4-position of the phenyl ring.

Figure 4.4: Example electrophysiology traces of the reactive allosteric modulators 4TFD-TQS and 4CM-TQS with *X. laevis* oocytes expressing α7 nAChRs. Electrophysiology was carried out by Dr Anna Chatzidaki.

Preliminary studies utilising a UV torch alongside voltage clamp electrophysiology suggested that the diazirine in 4TFD-TQS is efficiently degraded under UV light. However, the results were unclear as to whether covalent incorporation into the receptor was achieved. If UV was shone on the oocyte during application of 4TFD-TQS, then the allosteric agonist effect was degraded. Furthermore, if 4TFD-TQS was pretreated with UV for five minutes prior to application, then it also lost its allosteric agonist like behaviour. However, the UV pretreated 4TFD-TQS was still able to act as a PAM. PAM effects remained after washout with both UV pretreated and untreated 4TFD-TQS as has been documented before for such transmembrane binding allosteric modulators under the guise of ‘primed potentiation’ (12). Unfortunately, this makes it impossible to tell if persistent PAM-like behaviour is due to covalent incorporation of the compound at the binding site. An alternative explanation is that due to the high lipophilicity of the TQS type compounds, they are either very difficult to ‘wash-out’ from their binding site, or they remain within the tubing of the electrophysiology rig even
during long washes, and low concentrations are still present in the perfusion mix during subsequent applications of acetylcholine.

4.3.4 Prediction of positions of reaction from binding modes

Using the predicted binding modes of TQS-type compounds in closed and open conformation from Chapter 3, it is possible to predict which amino acid side chains would be expected to react with 4TFD-TQS upon photo-activation. The 4TFD-TQS structure was aligned with the central TQS-family binding mode from both the open and closed conformation clusters, and all residues within 7 Å of the carbon atom of the diazirine were selected as possible sites of cross-linking. The cut-off of 7 Å was chosen to allow for small amounts of movement of the ligand during residence at the binding site upon photoactivation and also to account for the likely scenario in which sidechain conformations are not correctly represented in the homology model. Under this relaxed cut-off twenty one residues were found, eight contributed by the principal (+) subunit, and thirteen contributed by the complimentary (-) subunit (Table 4.1, Figure 4.5). In future, cross-linking observed to any number of these residues by 4TFD-TQS may be interpreted to some extent as validation of these binding mode predictions.

Table 4.1: Residues found within 7 Å of carbene centre in possible binding modes of 4TFD-TQS.

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal (+)</td>
<td><strong>TM2</strong>: Leu 246, Thr 250, Met 253, Leu 254,</td>
</tr>
<tr>
<td></td>
<td><strong>TM3</strong>: Phe 274, Thr 277, Met 278, Val 281</td>
</tr>
<tr>
<td>Complimentary (-)</td>
<td><strong>TM1</strong>: Leu 212, Asn 212, Leu 214, Leu 215, Ile 216, Pro 217, Cys 218, Leu 220, Ile 221</td>
</tr>
<tr>
<td></td>
<td><strong>TM2</strong>: Ser 248, Leu 249, Val 251, Phe 252</td>
</tr>
</tbody>
</table>
Figure 4.5: Residues within 7 Å of open and closed binding mode predictions for 4TFD-TQS. Only open receptor model shown, as all residues identified within 7 Å on closed receptor are also identified in open receptor model. Identified residues are shown as grey sticks. The closed conformation predicted binding mode carbon atoms for 4TFD-TQS are shown in orange, while the open conformation predicted binding mode carbon atoms are shown in lime green, both are represented as ball and stick, heteroatoms are coloured as follows: nitrogen (blue), oxygen (red), sulfur (yellow), fluorine (pale green). Spheres centred on the diazirine carbon atom for open and closed conformation binding mode predictions of 4TFD-TQS are shown at a contour of 7 Å. The principal subunit of the receptor is coloured khaki, and the complimentary subunit is coloured in a darker shade of green. TM4 helices for principal and complimentary subunits are removed for clarity.

An analysis to detect residues potentially capable of cross-linking to the cysteine reactive analogues was also carried out. Euclidean distance cut-offs were chosen based on the distance between the first atom of the 4-position substituent on the TQS analogue and the Cβ atom of an imagined reacting cysteine sidechain (Figure 4.6) as these atomic coordinates are likely to move the least in the binding mode to a mutant receptor. For example, the distance between two carbon atoms covalently bonded via a sulfur (C-S-C) should be approximately 2.7 Å, given extra space for being non-covalently bound (prior to reaction) and allowing a little extra room for atomic movements, the cut-off of 4 Å was chosen for the 4CM-TQS analogue. However, it was found that only one Cβ atom was within 4 Å of the 4-substituent benzyl atom; Pro 217. This proline is conserved throughout the pLGIC family and thought to be responsible for a commonly observed segment of π-helix in TM1 (13).
it is not a good candidate for mutagenesis as side chain changes at this position are likely to have significant effects on the structure and function of the receptor (14).

Figure 4.6: Cysteine Cβ to first para-substituent atom in a C-S-C bond (left), attack of the benzylic C-Cl in 4CM-TQS (middle) and attack of the α-chloro group in 4ACA-TQS (right).

Although the other cysteine reactive TQS analogue 4ACA-TQS has not yet been confirmed as an active PAM, the extension of the electrophilic reactive centre on a longer substituent may afford more flexibility in choosing residues to mutate. For this compound, the Euclidean cut-off was extended to 6.4 Å and it was found that eight more residues have a Cβ atom within this cut-off distance (Thr 250 (+), Met 278 (+), Val 281 (+), Asn 213 (-), Ile 216 (-), Leu 220 (-), Ser 248 (-), Phe 252 (-)). Superposition the 4ACA-TQS structure on the predicted open and closed TQS-family binding modes suggests that mutations of Phe 252 or Ile 216 to cysteine may be the most likely to furnish successful cross-linking (Figure 4.7).

Figure 4.7: Residues in close proximity to the reactive centre of 4ACA-TQS in predicted open (left) and closed (right) binding mode conformations.
4.4 Conclusions and future work

In spite of these results, which are difficult to interpret, the compounds 4TFD-TQS and 4CM-TQS serve as potentially useful tool compounds for investigating covalent incorporation of PAMs into α7 nAChRs. 4ACA-TQS is yet to be confirmed as an active PAM, but may provide a better starting point for cross-linking via an engineered cysteine, owing to the greater number of residues which are within potential reacting distance.

The outcome of the UV electrophysiology experiments may be improved by the use of a more focussed set-up. The UV torch, providing a brightfield illumination of a large portion of the flow channel in which the oocyte sits may have been causing the degradation of much of the 4TFD-TQS in solution before reaching the surface of the cell. A UV laser set-up as successfully used previously for photo-crosslinking during electrophysiology (1) may yield better results. Alternatively, since the compounds appear to be washed out very slowly, incubation of the cells recombinantly expressing α7 nAChRs with 4TFD-TQS prior to washing and transfer to a clean bath and then irradiation with UV may allow better detection of crosslinking.
4.5 References


Chapter 5

5 Identification by virtual screening of novel allosteric modulators of the $\alpha_7$ nicotinic acetylcholine receptor

Some of the work in this chapter has been published in the following article:

Smelt CLC, Sanders V, Newcombe J, Burt RP, Sheppard TD, Topf M, Millar NS, Identification by virtual screening and functional characterisation of novel positive and negative allosteric modulators of the $\alpha_7$ nicotinic acetylcholine receptor, *Neuropharmacology*, 2018, 139, 194
5.1 Introduction

Virtual screening constitutes a process of *in silico* testing of a library of compounds for their likelihood of acting as a useful ligand in a protein-ligand interaction. Screenings of this sort are commonly used in the pharmaceutical industry in searching for new lead compounds (1, 2), or in attempts to ‘scaffold-hop’, i.e. to move from a potent compound with undesirable properties (such as a poor metabolism and pharmacokinetic profile) to another active compound with an alternative structure that may negate the issues with the previous lead (3). The advantages of carrying out this process in silico as opposed to with traditional high-throughput screening (HTS) procedures is clear. Virtual screening is typically much faster and very much less expensive than HTS. Owing to its ‘virtual’ aspect, more diverse chemical space can be visited amongst the ligands tested. However, one needs to know specific information about the target, compared with HTS. Often the structure of the target protein is known, and/or a significant number of high potency ligands have been discovered and these feed into the description of binding used to search a chemical library.

Virtual screening can be carried out in a number of ways. For example, if the interactions that a ligand must make with a protein in order to have the desired effect is well understood (such as from a high resolution crystal structure), a library of chemical compounds can be screened by docking each compound into the known ligand bound structure. Docking solutions can then be ranked based on their ability to bind in a similar way to the known ligand. This is known as structure-based virtual screening (1).

If relatively little is known about the protein structure but a number of structurally diverse ligands thought to bind at the same location are known, one can carry out a ligand-based virtual screening (2). In this regime, chemical structures are aligned in three dimensional space and are ranked based on the complementarity of the shape and pharmacophoric features (points in the compound that could cause interaction with the protein, e.g. hydrogen bond donors/acceptors, hydrophobic regions, etc.). Starting points for a pharmacophore can come from a high resolution crystal structure,
docking experiment, or can even be predicted based purely on the chemical structures of a family of known ligands.

Finally in a situation where a lot is known about a family of ligands and there is an abundance of functional and measured data relating to compounds examined in structure activity relationships, a quantitative structure activity relationship (QSAR) can be built (4). QSARs then use predicted (or empirical) physicochemical properties and chemical fingerprints of compounds in a library to interpolate the likely activity of the new compound based on the data that is present in the QSAR.

In Chapter 3, I showed how the predicted binding modes derived from consensus docking studies to the newly developed α7 nAChR structural models have changed our understanding of how positive allosteric modulators may bind within the transmembrane domain of this receptor. From the ligand families docked in the consensus docking protocol described in Chapter 3, the TQS family had the largest amount of accessible associated structural and functional data. The binding mode clusters from the TQS family of compounds, represent a useful model for the binding geometry of type II PAMs and can serve as a starting point for virtual screening.

Owing to the presence of the revised structural models of α7 described in Chapter 2 it would have been possible, in theory, to carry out structure-based virtual screening for this example. However, although the predicted binding mode geometries helped to rationalise the observed functional data relating to the three docked families of type II PAMs, they contain few specific interactions (typically 1-2 hydrogen bonds) which would make it difficult to reliably filter hits from the docking. Thus, it was logical to assume that the ligand overlays from the predicted binding mode clusters may be better utilised in pharmacophore generation for ligand-based virtual screening.

There are many reasons to search for unidentified allosteric modulators of nAChRs. Not least because a novel PAM could begin a drug discovery project, but also many existing PAMs including compounds from the TQS family have poor physicochemical properties. These typically culminate in
having low solubility in aqueous media and are therefore difficult to work with in characterising nAChR mutations etc. As such, novel allosteric modulators of nAChRs may also aid in the discovery of more useful and easier-to-use tool compounds in the pharmacological study of nicotinic receptors.

Virtual screening has previously been used as an approach to find novel ligands of nAChRs. These have typically been based on structural models of nAChR extracellular domains (5–10) or on the structure of the acetylcholine binding protein, a soluble pentamer which is homologous to the nAChR extracellular domain (11–14). However, binding mode predictions of PAMs which are thought to bind in the transmembrane domain have not previously been used in an attempt to uncover novel, transmembrane binding, allosteric modulators of the α7 nAChR.
5.2 Aims

Generate pharmacophore models from binding mode predictions of TQS-type compounds for screening a database of compounds.

Perform virtual screen of a compound database to attempt to uncover previously unidentified allosteric modulators of nAChRs.
5.3 Results and discussion

5.3.1 Pharmacophore query generation

The first task involved in ligand-based virtual screening is the generation of pharmacophore queries. These represent models of how a ligand or set of ligands bind to target protein. In the absence of structural information relating to the protein of interest, a family of ligands can be overlaid and pharmacophoric points picked that represent a common binding capability across the ligand family. These pharmacophoric points could consist of similarly oriented hydrogen bond acceptors, or hydrophobic regions, amongst others. However, ligand overlays are subject to error since the conformation of the ligands chosen may not be the active conformation and therefore may not be optimally predictive of active compounds.

One way to take into account information about the protein structure in the generation of ligand overlays is to use docking. Since the ligand structures are oriented relative to the protein active site, this improves confidence that the conformations chosen are more likely to be the active conformations of the ligand.

The 25 previously published α7 nAChR type II PAMs from the ‘TQS-family’ (Appendix 1.1.4, ‘TQS-family’) were docked into the TM part of the α7 receptor’s comparative model. The top five solutions for each ligand were clustered with a cut-off of 2.0 Å, using the consensus docking protocol described in Chapter 3. As before, the largest clusters of binding modes from the family were taken as the most probable active conformations (Figure 5.1A). Docking was carried out for both open and closed conformation models of the α7 nAChR, as both binding conformations were thought to be informative for this family of ligands. As type II PAMs, which potentiate agonist evoked responses, TQS-type compounds must clearly be able to bind to the open conformation of the α7 nAChR. A further interesting property of this family of ligands is that in order to obtain an immediate onset of their pharmacological effects in two electrode voltage-clamp electrophysiology, TQS-type compounds must be pre-applied to the receptor. It could be interpreted that the necessity for pre-application suggests that TQS-type
compounds must be already at their binding site upon co-application with acetylcholine to immediately invoke PAM effects. Consequently, it is likely that TQS-type compounds are also capable of binding to a closed conformation receptor. Therefore, one could hypothesise that to reproduce the pharmacological properties of TQS-type compounds, a ligand must be able to bind both open and closed conformations of the α7 nAChR (see Chapter 3). Hence, queries based on both open and closed conformation binding modes of TQS-type compounds can be used as an additional restraint in screening to filter out falsely returned hits.

Figure 5.1: Generation of pharmacophore queries used for virtual screening. The highest ranked clusters of binding mode solutions with previously characterised PAMs are shown within the α7 nAChR transmembrane domain (A). The TM1-3 helices of the principal subunit and TM2 helix of the complimentary subunit are shown for the open (cyan) and closed (pink) conformations. Also shown are binding mode clusters from which pharmacophore queries were generated for the open (green) and closed (orange) conformations. From the ligands in each cluster, pharmacophore queries were generated for the closed and open conformations (B and C, respectively). Note, only those selected for screening are shown. Features of the pharmacophore are represented as yellow spheres (hydrophobes), green spheres (rings), red hashed spheres (hydrogen bond acceptors) and blue hashed spheres (hydrogen bond donors). Figure taken from (15) under the creative commons license (https://creativecommons.org/licenses/by/4.0/).

The ligand based virtual screening program vROCS (16) from the OpenEye software package (www.eyesopen.com) was chosen for both pharmacophore construction and for performing the virtual screening. The open and closed
conformation binding mode clusters were submitted to the vROCS ligand model builder tool allowing up to six ligands from each cluster to be utilised in the query construction.

The ligand model builder tool then built every variation of possible query models containing between one and six ligands from the supplied binding mode clusters. These consisted of gaussian volumes corresponding to the molecular shape of the overlaid ligands and ‘color atoms’ at pharmacophoric points associated with hydrogen bond donors, hydrogen bond acceptors, rings and hydrophobes in the ligands that contributed to each of the queries that were built. Every built query was screened against the ligands in the cluster and the three queries with the highest average similarity to all the ligands from the cluster determined by the TanimotoCombo score (17) for both open and closed conformation binding mode clusters were taken forward for validation screening runs (Figure 5.1B and C).

5.3.2 Pharmacophore query validation

An important step in virtual screening is the validation of the queries used, this helps to reassure that hits returned are indeed likely to be active compounds or at least capable binders of the protein of interest. In ligand based virtual screening pharmacophore queries are typically validated by testing their ability to retrieve a set of known ligands, not used in building the pharmacophore from a set of non-binders of relatively high chemical similarity, commonly referred to as decoys.

A database of 53 known PAMs of α7 was constructed, including the 25 TQS-type compounds that were docked to the α7 comparative models. Of these 25 TQS-type compounds, 11 were used to build the three pharmacophore queries from open and closed conformation binding mode clusters. These 11 compounds were removed from the database, leaving 42 compounds which constituted the ‘known actives’ database (Appendix 1.3.1). A second database of decoys was generated using the program decoyfinder (18). For each of the 42 ligands in the actives database decoyfinder retrieved 36 compounds from the ZINC database (19), ensuring that the Tanimoto coefficient (17) between the ligands in the known binder (actives) set and the...
decoy sets did not surpass a threshold of 0.8 and that no decoys were the same as those found for other known actives.

The program vROCS is based upon optimising the overlap of a potential ligand being screened against the pharmacophore query being used. The gaussian representation of molecular shape used in vROCS allows this process to be very fast, however, does not allow permuting between different conformations of each molecule. In order to take into account conformational flexibility vROCS depends on the pre-generation of a number of different conformations of each molecule being screened. Here, this was performed with another piece of OpenEye (www.eyesopen.com) software OMEGA2 (20). The active and decoy compound databases were populated with up to 200 conformations of each compound using OMEGA2. Often fewer than 200 conformations would be present, as a redundancy removal step is used, and conformations from the aligned set of those generated were removed if their root mean squared deviation was less than 0.5 Å from another conformer by default.

Screening the 3D pharmacophore queries against the active and decoy conformer databases and ranking all of the compounds together allows the construction of receiver operator characteristic plots (ROC). ROC plots give an overview of how likely a pharmacophore query is to rank an active compound above a decoy compound. The area under the curve (AUC) is one of the statistics that can be obtained from these plots relating to the quality of the pharmacophore queries.

An AUC of 1 denotes a perfect return of every single active compound before any decoy compounds. An AUC of 0.5 is equivalent to a random selection of compounds, where actives are on average ranked equally to decoys. AUC values of less than 0.5 denote a pharmacophore query that is more likely to return a decoy than an active compound. Although AUC gives an excellent overview of the global likelihood of return of actives versus decoys, it does not give information about where in the database actives are ranked higher than decoys (21). Early enrichment offers a better way of measuring this. In vROCS the enrichment at X% is defined in Equation 5.1:
\[ (5.1) \quad X\% \text{enrichment} = \left( \frac{\text{fraction of actives at } X\% \text{ of decoys found}}{X} \right) \times \frac{100}{X} \]

For example, if 70% of the active compounds have been found when 0.5% of the decoy compounds have been found, the 0.5% enrichment would give as in Equation 5.2:

\[ (5.2) \quad 0.5\% \text{enrichment} = 0.7 \times \frac{100}{0.5} = 140\% \]

Figure 5.2: Validation statistics of pharmacophore queries used for virtual screening. Receiver operator characteristic (ROC) plots for closed conformation (left) and open conformation (right) screenings. The dashed grey diagonal line shows AUC = 0.5, where queries above this line are better than random selection at choosing genuine hits over decoys, and queries below are worse than random selection at choosing genuine hits over decoys. For both conformation screenings, query 1 is shown as a solid blue line, query 2 as dashed red line and query 3 as a dashed purple line. Figure taken from (15) under the creative commons license (https://creativecommons.org/licenses/by/4.0/).

The validation statistics for the six pharmacophore queries (three open conformation and three closed conformation) were very promising (Figure 5.2, Table 5.1). In both open and closed conformations all queries had excellent AUC values of 0.95 or greater (Table 5.1) and high levels of early enrichment, suggesting that any of the queries could have been a good choice for screening a database of compounds. In both cases however, the second query had both the highest AUC and the highest enrichment at 0.5%, 1% and 2% (Table 5.1) making them the obvious choice for screening.
Validation data for pharmacophore queries generated using vROCS v3.2.1.4 (values are mean of 10,000 bootstrapped iterations with 95% confidence interval). %enrichment gives a measure of how high true actives are ranked compared with decoys, e.g. at 1% enrichment, a value of 50% denotes that half of all actives have been ranked higher than 1% of the decoys.

Despite the removal of the compounds that were used in the construction of the pharmacophore queries, many TQS type compounds were present in the actives database and their highly related structures may have generated some level of bias in the AUC and early enrichment values. However, 17 of the 42 compounds in the actives database, coming from the Abbott, Janssen, Pfizer and TBS compound families (Appendix 1.3.1) are structurally unrelated to the TQS compounds, representing more than 40% of the active compounds. Therefore, it seemed reasonable to think that the validation statistics demonstrate that these are high quality queries, given that several of the compounds unrelated to TQS must be returned early on in order to obtain enrichment values greater than 120% at 0.5% enrichment, 60% at 1% enrichment or 30% at 2% enrichment.

Table 5.1: Validation data for pharmacophore queries

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>0.5% enrichment</th>
<th>1% enrichment</th>
<th>2% enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Closed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>query 1</td>
<td>0.96 ± 0.03</td>
<td>135.5 ± 24.9%</td>
<td>67.7 ± 12.4%</td>
<td>34.0 ± 6.5%</td>
</tr>
<tr>
<td>query 2</td>
<td>0.97 ± 0.02</td>
<td>142.6 ± 23.2%</td>
<td>73.2 ± 12.6%</td>
<td>40.7 ± 7.6%</td>
</tr>
<tr>
<td>query 3</td>
<td>0.96 ± 0.03</td>
<td>134.9 ± 27.4%</td>
<td>70.9 ± 13.3%</td>
<td>36.4 ± 6.2%</td>
</tr>
<tr>
<td><strong>Open</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>query 1</td>
<td>0.96 ± 0.03</td>
<td>136.1 ± 25.8%</td>
<td>70.6 ± 12.4%</td>
<td>36.8 ± 5.6%</td>
</tr>
<tr>
<td>query 2</td>
<td>0.97 ± 0.03</td>
<td>140.2 ± 26.2%</td>
<td>73.4 ± 11.9%</td>
<td>41.2 ± 5.7%</td>
</tr>
<tr>
<td>query 3</td>
<td>0.95 ± 0.03</td>
<td>135.7 ± 26.8%</td>
<td>67.8 ± 13.3%</td>
<td>33.9 ± 6.7%</td>
</tr>
</tbody>
</table>
5.3.3 Virtual screening of the DrugBank database

5.3.3.1 Choice of DrugBank for screening

In theory any database of chemical compounds (characterised or imagined) could be searched for novel α7 PAMS via virtual screening. However, in reality, the database used and the way in which it is filtered have a big influence on the results obtained. In this study the DrugBank database (22) was used. As a small curated database of approved and experimental drugs, DrugBank offered an excellent resource comprised of high quality, drug-like, small-molecule scaffolds.

5.3.3.2 Filtering procedure

For the top 100 hits found for both open and closed queries, two filters were applied. Firstly on the basis that TQS-type compounds are able to bind both open and closed conformations of the α7 nAChR. Only compounds returned in the top 100 hits for both open and closed conformations were taken forward. Next, since in part the aim of the virtual screening was to pursue more ‘drug-like’ nAChR PAMs, the central nervous system multi-parameter optimisation (CNS MPO) (23, 24) was applied to give each hit a CNS MPO score and only compounds scoring greater than 4.0 were accepted through the filter.

5.3.3.3 Central Nervous System Multi-Parameter Optimisation (CNS MPO)

The CNS MPO scores compounds between 0 and 1 for six physicochemical features that are thought to play a role in the distribution of drugs around the body (Molecular weight, topological polar surface area, calculated octanol-water partition coefficient (cLogP), calculated octanol-water partition coefficient at physiological pH (cLogD), number of hydrogen bond donors, and the dissociation constant of the most acidic proton in the compound (pKa)). The linear combination of these six scores gives a score out of 6. It was previously found that scores of 4.0 and above represent compounds with desirable properties for crossing the blood-brain barrier (23). A later analysis of the use of this scoring internally at Pfizer showed that compounds scoring...
4.0 or above are also less likely to fail in clinical trials (24) suggesting that compounds achieving these scores may be higher quality leads in drug discovery.

Here values corresponding to the six physicochemical features were calculated using chemaxon calculator plugins (Marvin 16.7.4, 2016, ChemAxon (http://www.chemaxon.com)) and the scores were determined using a script written in python (Appendix 1.3.2).

### 5.3.3.4 Returned hit compounds

Following the application of the filtering procedures to the top 100 hits for open and closed conformation pharmacophore queries, 81 compounds were found. To limit the number of compounds tested, efforts were focussed on the top 25 of the 81 hits. These 25 compounds were grouped into functional classes based on the proteins they are annotated to be active upon in the DrugBank database and four classes were found. The representatives of the four classes are shown in Figure 5.3.

![Figure 5.3: Allosteric modulators of the α7 nAChR. Four cluster representative compounds from the DrugBank database that were identified by virtual screening in the top 25 ranked compounds and selected for functional characterisation. From left to right in order of rank DB04763 (a CAII inhibitor), DB08122 (a CDK2 inhibitor), furosemide (NKCC), pefloxacin (DNA gyrase inhibitor).](image-url)
Figure 5.4: Virtual screening ligand overlays. The structure of DB04763 (A), DB08122 (B), furosemide (C) and pefloxacin (D) are shown (in stick form) overlaid on the molecules used to create the closed conformation pharmacophore query (shown in wireframe). Atoms are represented as follows: carbon (grey), oxygen (red), nitrogen (blue), sulfur (yellow), chlorine (light green), fluorine (green) and hydrogen (white). Figure taken from (15) under the creative commons license (https://creativecommons.org/licenses/by/4.0/).

DB04763 (Figure 5.3), a carbonic anhydrase II (CAII) inhibitor (25), was the top-ranked compound identified by virtual screening. In total 22 CAII inhibitors were found in the 25 top-ranked compounds, this class of compounds dominates the top ranked hits. There were 36 CAII inhibitors in the 81 total hits (Appendix 1.3.3). DB08122 (Figure 5.3) was the 3rd-ranked compound identified by virtual screening, but 2nd with a different pharmacological target and is a cyclin dependent kinase 2 (CDK2) inhibitor (26). It is the only example of a CDK2 inhibitor in the 25 top-ranked hits, but one of seven CDK2 inhibitors amongst the 81 total hits (Appendix 1.3.3).
Furosemide (DB00695, Figure 5.3) is the 6th-ranked hit from virtual screening, 3rd with a different target and is a diuretic that acts on the Na\(^+\)-K\(^+\)-Cl\(^-\) cotransporter (NKCC) (27). No other compounds with a similar known mechanism of action were identified amongst the 81 virtual screening hits. Pefloxacin (DB00487, Figure 5.3), a fluoroquinolone antibiotic, active as an inhibitor of DNA gyrase (28), is the 22nd-ranked hit, 4th with a different target, and is the only example of this type of compound in the 25 top-ranked hits, but one of 13 such compounds amongst the 81 total hits (Appendix 1.3.3). Overlays of the selected compounds with the original pharmacophore queries demonstrates their close match to the shape of the ‘TQS-family’ of PAMs (Figure 5.4).

5.3.4 Compound testing

5.3.4.1 Chemical synthesis

Furosemide and Pefloxacin were purchased from Sigma-Aldrich (Gillingham, UK). DB04763 and DB08122 were synthesised by methods that have been described previously (25, 26) (see Chapter 8 - Experimental).

5.3.4.2 Pharmacological characterisation (results from Charles Smelt and Victoria Sanders, Millar Lab, UCL)

The four selected hit compounds were tested on recombinantly expressed human \(\alpha_7\) nAChR in Xenopus oocytes by members of the Millar lab. Intriguingly, all four compounds were found to have an effect on the \(\alpha_7\) nAChRs (Figure 5.5). Furosemide, was found to potentiate responses to acetylcholine in a manner comparable to a type I PAM i.e. an increase in maximum response of the receptors to acetylcholine, with no significant change in the rate of desensitization. The other three compounds (DB04763, DB08122 and pefloxacin) all appeared to attenuate responses of \(\alpha_7\) nAChRs to acetylcholine, suggesting that they behave as inhibitors. Antagonism was found to be non-surmountable through increasing acetylcholine concentrations, suggesting that the inhibitors were not active via the orthosteric site.
Figure 5.5: Functional characterisation of DB04763, DB08122, furosemide and pefloxacin on the α7 nAChR. Representative traces, from oocytes expressing the α7 nAChR, in response to ACh (100 μM; left), together with an ACh response from the same oocyte after pre- and co-application of test compound (1 mM; middle). Also shown are responses after a two minute wash. Horizontal lines above the traces represent the duration of application for ACh (solid line) and the tested ligand (dashed line). All functional data was collected and analysed by members of the Millar lab. Figure adapted from (15) under creative commons license (https://creativecommons.org/licenses/by/4.0/).
5.3.4.3 Docking of furosemide

To further examine its action on the α7 nAChR, furosemide was docked into the open and closed conformation α7 structural models using the consensus docking protocol (see Chapter 3). The top five predicted binding modes were examined for similarity to those expected for the TQS-type compounds (Figure 5.6A and B). Amongst the top five binding modes predicted for furosemide in both open and closed conformations of α7, at least one pose was reminiscent of those predicted for TQS-type compounds (Chapter 3, Figure 3.9C and F).

From the predicted binding modes for furosemide, three residues in close proximity were selected for mutation to alanine (L247A, S248A and T288A). Alongside these mutants, two other α7 nAChR transmembrane mutations, previously tested on type II PAMs (29–31), were also made and tested by members of the Millar lab (S222M, M260L). Of the mutants tested, only L247A was found to have a significant effect on the sensitivity of the α7 nAChR to acetylcholine (Figure 5.6C), however, this effect is not unusual for mutations at this position (32). Interestingly, all mutants were found to have an effect on furosemide activity. S222M increased potentiation by furosemide, M260L turned furosemide into an inhibitor and all three of the predicted proximal mutations abolished potentiation by furosemide (Figure 5.6D), lending weight to the predicted binding mode.
Figure 5.6: Influence of α7 nAChR mutations on the allosteric modulatory effect of furosemide. The docked position of furosemide is shown in the closed (A) and open (B) structural model of the α7 nAChR transmembrane region. The TM1-3 helices of the principal subunit (olive) and TM2 and TM3 helices of the complimentary subunit (green) are shown. Amino acids examined by site-directed mutagenesis are indicated. C) (carried out by members of Millar lab) ACh dose-response curves determined with wild-type α7 nAChR (dashed line) and with α7 nAChRs containing single point mutations S222M, L247A, S248A, M260L and T288A. Data are means of at least three independent experiments. D) (carried out by members of Millar lab) Bar chart illustrating the influence of furosemide (1 mM) on responses to an EC50 concentration of ACh (10 μM for L247A and 100 μM for wild-type and all other mutated receptors). Data are normalised to the response observed in the same oocyte in the absence of furosemide. Data are means ± SEM of at least three independent experiments. Significant differences from wild-type are indicated (* P <0.05, ** P< 0.01, *** P <0.001). In addition, significant differences form agonist responses in the absence of furosemide are indicated (# P<0.05, ## P< 0.01). Figure taken from (15) under the creative commons license (https://creativecommons.org/licenses/by/4.0/).

Furthermore, studies conducted by members of the Millar lab using chimeric receptors comprising extracellular and intracellular domains of α7 and the transmembrane domain of the 5-HT3AR were consistent with all four
compounds acting via a transmembrane site. One possibility is that the three compounds found to act as inhibitors may purely do so by blocking the ion channel pore. However, transmembrane mutants on the TM1 (S222M) and/or TM3 (T288A) helices which are not close to the channel lumen were found to affect the behaviour of DB08122 (antagonism increased by S222M and T288A) and pefloxacin (antagonism increased by T288A). Of the three inhibitors, this is consistent with at least DB08122 and pefloxacin acting via an allosteric transmembrane site.

It would not be unexpected to obtain allosteric ligands with properties of negative allosteric modulators or type I PAMs from a pharmacophore based on TQS-type compounds, as chemical changes as small as methyl-substitution patterns around a phenyl ring have been found to bring about allosteric modulators with diverse effects, such as type I PAMs, type II PAMs and NAMs (33).

Interestingly, in a nearly 30 year old study it was found that furosemide increased contraction of rat urinary bladder by potentiating the action of acetylcholine. Furthermore, this effect could be blocked by the nAChR antagonist hexamethonium (34). Additionally, more recently the functional expression of α7 nAChRs in rat urinary bladder epithelial cells has been reported (35). As such, the finding that furosemide acts as a PAM of α7 nAChRs provides a credible rationalisation for the previously reported phenomenon that furosemide is able to potentiate acetylcholine-induced bladder contraction (34). Previous studies have also demonstrated that furosemide antagonises GABA\textsubscript{A}Rs (36), another member of the superfamily of pentameric ligand-gated ion channels, and that this interaction may occur via a transmembrane binding site (37), consistent with the hypothesis that furosemide may also interact with α7 nAChR transmembrane domain.

Of the other three compounds identified and tested on α7 nAChRs, there does not seem to be any prior literature suggesting that DB04763 (or any CAII inhibitors) or DB08122 (or any CDK2 inhibitors) may act allosterically on nAChRs. Pefloxacin however, is a member of the quinolone antibiotics. Quinolone antibiotics have previously been proposed to act as inhibitors of
GABA<sub>A</sub>Rs which could explain some of the adverse side-effects of quinolone antibiotics related to the central nervous system (38).

5.4 Conclusions and future work

In conclusion, virtual screening was conducted on a small library of drugs and drug-like compounds, using pharmacophore queries based on the chemical structures of PAMs previously shown to interact with the transmembrane domain of the α<sub>7</sub> nAChR. Four representative top hits were selected and all were found to exhibit allosteric effects on the α<sub>7</sub> nAChR. One of these hits, furosemide, appears to act as a type I PAM and these effects could offer a molecular explanation of previously unexplained functional data.

Docking of furosemide found that it was predicted to form a similar binding mode to that of binding mode clusters predicted for TQS-type compounds. Functional characterisation (carried out by members of the Millar lab) of receptors containing three different single point mutations predicted to be proximal to the binding modes predicted for furosemide, found that all three mutants abolished potentiation of acetylcholine induced currents by furosemide. These findings provide experimental validation and lend weight to the predictive power of both the new α<sub>7</sub> structural models and the docking procedure developed in Chapter 3.

Previous studies have implicated furosemide in the modulation of other pLGICs such as GABA receptors (39, 40). Interestingly one study used furosemide as a prophylactic treatment against the symptoms of autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE) (41). In that study, furosemide was thought to be exerting its effect via modulation of the gabaergic signalling, however, the ADNFLE model was based on a single point mutation in the acetylcholine receptor α<sub>4</sub> gene (S284L CHRNA4) (41). The finding in this chapter that furosemide potentiates acetylcholine induced responses of α<sub>7</sub> nAChRs may suggest that furosemide may also be able to interact with α<sub>4</sub> containing receptors, such as α<sub>4</sub>β<sub>2</sub> nAChRs. With atomic structural models now available for both GABARs (42, 43) and the α<sub>4</sub>β<sub>2</sub>
nAChR (44, 45) it would be interesting to delve deeper into the interaction of furosemide with these receptors via modelling and docking.

5.5 References


17. Tanimoto TT (1958) *An elementary mathematical theory of classification and prediction* (IBM).


1):C177-92.


Chapter 6

6 Refinement of the full length *Torpedo* nicotinic acetylcholine receptor
6.1 Introduction

The production of the refined models of the *Torpedo marmorata* αγ subunit in Chapter 2 led to a collaboration with Prof. Nigel Unwin (LMB, MRC, Cambridge), the researcher responsible for the publication of the first pLGIC atomic models. Currently pursuing research focussing on the interaction of these fascinating receptors with lipids, Prof. Unwin has acquired more post-synaptic membranes from the *Torpedo* ray and has been examining them with the newer cryo EM tools available since the ‘resolution revolution’ (1). These include more powerful microscopes such at the FEI (now Thermo Fisher) Titan Krios; direct electron detectors which provide higher signal to noise ratios in collected images (previously charged coupled devices were used and before that photographic paper); and improved single particle reconstruction software, for example, *RELION* (2), which uses Bayesian statistics to produce alignments of low signal to noise ratio images.

Some of these new techniques for image collection and data processing have been applied to the pLGIC family, with very recent high resolution cryo EM structures appearing for the full length 5-HT3A receptor (3), as well as both high and low affinity stoichiometries of the α4β2 nAChR (α4)2(β2)3 and (α4)3(β2)2, respectively) (4). Encouragingly, early images released by Prof. Unwin seem to show a degree of segregation of lipids on the post synaptic membrane suggestive of cholesterol clustering at an interstice between δ subunits of a pair of nAChRs (5).

Subsequent further collection of images has enabled the production of three dimensional maps, which Prof. Unwin has kindly shared with me to analyse and use to inform the building of molecular models. The new maps have been generated using the helical reconstruction protocols within *RELION* (6) and they appear to show a number of differences with the currently deposited (and available) maps for the *Torpedo* nAChR (emd-2071, emd-2072 (7)). Most notably from the raw, unfiltered reconstructions, the length of the maps in their Z-axis (along the ion channel pore) is shorter than the previous maps by approximately 6 Å (Figure 6.1). However, the new maps are still preliminary reconstructions and are of relatively low resolution (lower...
resolution than the previous reconstructions). Further improvements to the maps will come as more images are added to the dataset used in their reconstruction.

Figure 6.1: Comparison of previous *Torpedo* nAChR model and map with newly generated map. When the old nAChR density map (green, transparent, emd-2071) and new nAChR density map (grey, opaque) are aligned on the segments corresponding to the MX helices, the portion corresponding to the N-terminal helix (purple model, PDBid: 4AQ5) in the old map protrudes from the top of the new map signifying an approximately 6 Å difference in length.

The native state of the *Torpedo* nAChR is as a receptor dimer, with two pentameric receptors covalently linked via a disulfide between the δ-subunits on the TM4 helix (Figure 6.2). On initial isolation from the *Torpedo* electric organ, nAChR rich membranes appear to show a somewhat random distribution of receptors on the membrane surface (8). After incubation in an appropriate buffer for periods of about four weeks, the receptor dimers assemble first into 'ribbons', and then tubular crystals whereby the ribbons of
receptor dimers form associated helices by sitting ‘side-by-side’ (Figure 6.2). The observed differences (Figure 6.1) are likely to have been caused by a systematic error arising from the averaging of tubular crystals (or receptor tubes) of different helical parameters, or different diameters (see below).

Figure 6.2: Segment of helical arrangement of nAChR receptor ribbons. A dimer of receptors has been highlighted in dark grey, the path of the ribbon of receptor dimers is traced out with two black lines, the highlighted boxed area corresponds to the zoomed region (right) showing where the cross-link between δ subunits sits and the contact between adjacent receptor ribbons of the C-loop from the αγ subunits.

During this process of tube formation a number of different helical families can arise. If a helical tube is split along its length and hypothetically unfurled to give a 2D array, this would have the plane symmetry group $p2$. The $p2$ symmetry group has two unit cell vectors which can be represented as the principal lines $(1,0)$ and $(0,1)$ (Figure 6.3A. Taken from (9)). If imagined on the tube, the principal lines would create left-handed $(1,0)$ and right-handed $(0,1)$ helical traces along the tube. The helical families can then be defined by the number of lines required to fill the circumference of the tube for both vectoral directions. Families with different numbers of receptors making up the circumference of the tube have inherently different helical pitch (Figure 6.3B).
Figure 6.3: Determination of helical family in nAChR helical tubes. Left, surface lattice of nAChR dimers (linked circles) imagined as if the cylindrical tube had been split along a single line on the surface, parallel to the axis of the tube (vertical), opened up and viewed from the outside. The surface lattice is built from ribbons of nAChR dimers which run diagonally across the surface with respect to the tube axis. The two unit cell vectors (a and b) for the $p2$ symmetry group are shown as are the principal (1,0) and (0,1) lines. The two numbers which characterise the helical family of the tube are determined by the numbers of the two principal lines (helices for the tube) that can be drawn for one complete turn of the helix. nAChR dimers which sit at the boundary of a complete turn are shaded, for the example here there are sixteen lines belonging to the (1,0) vector and six belonging to (0,1) hence the family of this tube is (-16,6). Right, tubes built from receptor dimer ribbons that originate from different helical families have different pitches depending on the family and the dimensions of the unit cell. Figure reproduced from (9) with permissions.

Defined in this way, four helical families have commonly been observed to arise for these tubes of receptors (-17,5), (-16,6), (-15,7) and (-18,6). Each of these families has a differing average diameter; 380 Å, 385 Å, 394 Å and 416 Å, respectively (9). However, there may be variation in radii of each helical family due to, for example, different or disordered packing of lipids between receptor pairs. There is also some degree of disorder within the tubular crystal, this is especially prevalent in the TM4 helices, which could only be resolved following scaling of the density amplitudes in the maps (10).

Slight differences in helical properties and diameters of tubes of nAChRs from Torpedo electric organ post synaptic membrane can arise from the stochastic process in which the membrane bound receptors reorganise into tubular vesicles, or from the process of vitrification in which small volumes of aqueous solution containing these receptor tubes are rapidly frozen at cryogenic temperatures by plunging into liquid nitrogen cooled ethane. If the layer of ice is very thin (close to the same thickness as the tube) then the tubes may
become squashed, causing an elliptical profile in the X,Y plane of the tube and adding noise to any alignments which contain these images. The previous maps were generated by the addition of Fourier terms from each image in a data set and Fourier terms were selected based on ‘Point Quality’ a measure of signal-to-noise ratio at each point in the Fourier transform of the images (11). Since this methodology requires the estimation of whether images should be combined or not, it may have given rise to the incorporation of spurious images, e.g. those with different diameters, or elliptical cross-sections, into the dataset. Reducing resolution of more flexible regions, such as loops, and causing an elongation of the receptor density.

It is thought that the newer method of using Bayesian statistics to classify images as in RELION (2) is better able to differentiate between images which should be added and those which should not, e.g. due to being different helical families, having different radii etc. The fact that the new maps correspond better to the shape of the recently published 5-HT₃AR and α₄β₂ nAChR structures corroborates the idea that the new maps may provide a better interpretation of the nAChR structure (Figure 6.4). With the 5-HT₃AR model fitted into the new map for the *Torpedo* nAChR, the N-terminal helices do not protrude from the top of the density map. However, there is a poor fit of the MX helices, which protrude from the density map (Figure 6.4). This is perhaps more likely due to the difference in the surrounding lipid environment for the *Torpedo* nAChR structure and the detergent solubilised 5-HT₃AR model.
Figure 6.4: Fit of most recent 5-HT\textsubscript{3A}R model (PDBid: 6BE1, sea green) into new Torpedo receptor maps. For clarity two subunits of the 5-HT\textsubscript{3A}R are coloured sea green, while the other three are coloured light grey, a pore bound sodium ion from the 5-HT\textsubscript{3A}R is shown in purple.

Owing to the concerns about the previous models of the Torpedo nAChR (PDBids: 2BG9 (11), 4AQ5, 4AQ9 (7), 1OED (10)) including the elongation of the model and the low model quality of these templates (discussed in Chapter 2), these were assessed to no longer be the best template for fitting a model of the nAChR to the new maps. Furthermore, in addition to the transmembrane register error arising from the assignment of one extra turn to the transmembrane helix TM1 (see Chapter 2), the conformation of the TM2-TM3 loop is also a contentious issue. With the continually growing body of evidence from other pLGIC structures that the TM2-TM3 loop should extend towards the complimentary subunit at each interface rather than extending upwards as in the deposited Torpedo structures (Figure 6.5), it is likely that this loop also needs correcting.
Figure 6.5: Difference in TM2-TM3 loop geometry for Torpedo nAChR and α4β2 nAChR. Left) two protomers of the Torpedo nAChR (PDBid: 2BG9), the TM2-TM3 loop (inset) protrudes within the subunit. Right) two protomers of the α4β2 nAChR (PDBid: 6CNJ), the TM2-TM3 loop (inset) protrudes towards the adjacent (complimentary (-)) subunit. Principal, (+) and complimentary, (-) subunits are indicated.

The recent 5-HT₃ₐR and α4β2 nAChR models do offer good starting points for modelling. The (α4)₂(β2)₃ model (PDBid: 6CNJ) offers the highest sequence identity and is the most closely related homologue of the new structures. However, the 5-HT₃ₐR model (PDBid: 6BE1) is more complete, resolving the additional intracellular MA helix, which is removed in the α4β2 model, and is also in a more relevant conformational state (proposed to be closed-resting as opposed to desensitized). Between these two structures, the vast majority of the resolved portions of the nAChR can be successfully modelled.
6.2 Aims

Generate new models of the *Torpedo* nAChR that will be able to take full advantage of the better quality of the new receptor maps.

Use the latest pLGIC models (5-HT\textsubscript{3AR} and α4β2 nAChR) as templates to model the *Torpedo* nAChR and use flexible real-space refinement (as in Chapter 2) to more favourably fit the model to the corresponding experimental density.

Assess new models by seeing if they can better rationalise functional data relating to this receptor.
6.3 Results and discussion

6.3.1 Estimating the resolution of the new Torpedo maps

Having an estimate of the resolution of a map is important for obtaining useful values for the cross-correlation when fitting a model, knowing what sort of features can reasonably be fitted (e.g. domains, secondary structure elements, side-chains) and also for guiding flexible fitting of atomic models.

Half maps reconstructed from two independently and randomly chosen halves of the collected data were used to estimate the resolution of the new map using the Fourier shell correlation (FSC) (12). An important consideration in calculation of FSC curves is the masking of the appropriate region of data. FSC is optimally determined on a single asymmetric unit, i.e. in this case, density corresponding to a single receptor and the extraneous density values should be set to zero using a binary mask (everything inside the mask maintains its density values, outside the mask is set to zero). An added complication is that a ‘hard’ change where density values instantly become zero due to a mask with a ‘hard edge’ represents a high resolution feature that will correlate incredibly well between half maps. Therefore a ‘soft-mask’ must be used, in a soft mask, density values are gradually faded to zero over a given distance from the edge of the mask, for example following a gaussian distribution, over 5 Å window. A soft-mask was used to extract density corresponding to a single receptor from both 200 Å³ density half maps of Torpedo nAChR tubes (Figure 6.6), by multiplying the larger volume by the mask in the package EMAN2 (14).
Figure 6.6: Soft-mask extraction of a *Torpedo* receptor. A soft mask (pink) was fitted over a single receptor in a 200 Å³ patch of nAChRs embedded in a tubular vesicle. Multiplication provided density corresponding to a single receptor with 'soft-edges' where density values are gradually faded to zero at the edges of the map rather than a hard cut-off where all density values immediately become zero at the edge of the map.

The FSC plot was determined using the PDBe FSC calculation server (https://www.ebi.ac.uk/pdbe/emdb/validation/fsc/) which utilises the EMAN2 package (14) and iMAGIC (https://imagescience.de/). The calculated FSC curve falls off through the 0.143 resolution criterion at a spatial frequency of 0.132 Å⁻¹, suggesting a nominal resolution of 7.6 Å (Figure 6.7). However, the FSC begins to rise again after reaching zero, which is possibly indicative of overfitting of the density maps during reconstruction. Since the reconstruction process relies on an initial model that is iteratively refined by adding data that agrees with the model, if too much weight is given to the model in this process, this can result in the selection of images where there is correlation in the noise (15). This is likely to be a problem that can be overcome by the addition of more images to the dataset and optimisation of the reconstruction process which is currently underway. It is therefore important to keep in mind that the maps presented here are preliminary and high resolution features within these maps may not be truly representative of the structure. However, the initial resolution estimates should still be relevant because this is from the portion of the curve that appears to follow the expected trend. Therefore, the lower resolution features of the map e.g. positioning of helices and general topology should also hold true.
Figure 6.7: FSC plot determined from two half maps reconstructed from independent halves of the micrographs collected. The FSC falls to zero through the 0.143 ‘gold-standard’, criterion at a spatial frequency corresponding to a nominal resolution of 7.6 Å. The rise in FSC after hitting zero is indicative of correlation in the noise region of the maps which may have been caused by overfitting the data during reconstruction.

6.3.2 Generating an initial model of the Torpedo receptor

The first task in creating new models of the nAChR was to generate a sequence-structure alignment between all the relevant templates and the subunits of the Torpedo nAChR. The alignment was carried out between the mouse 5-HT3AR subunit, human α4 and β2 nAChR subunits and Torpedo α, β, γ and δ subunits using Clustal Omega (16) and sequence identities between each chain of the Torpedo nAChR and α4, β2 and 5-HT3A subunits were calculated within the UCSF Chimera Multi-Align Viewer (17). Sequence identities are listed in Table 6.1, it was found that the closest homologues to each subunit were as follows: Torpedo α: human α4, Torpedo β, γ, δ: human β2. Therefore, these subunits were used as the primary template for each of the Torpedo subunits (i.e. the majority of each subunit was modelled based on the human (α4)2(β2)3 receptor structure PDBid: 6CNJ). However, since this template does not contain any coordinates for the MA helix, TM4 and MA were modelled using the 5-HT3AR model (PDBid: 6BE1) as well as the human (α4)2(β2)3. Finally, the C-terminal regions of each of the subunit TM4 helices was based on the original Torpedo nAChR model (PDBid: 2BG9), since no other pLGIC has such complete data relating to these C-terminal regions. The
relevant portions of the receptor models were aligned in '.pir' format for use with MODELLER (Appendix 1.4.1).

Table 6.1: Sequence identities between homologous receptor subunits.

<table>
<thead>
<tr>
<th>Torpedo nAChR subunit</th>
<th>Human α4 nAChR</th>
<th>Human β2 nAChR</th>
<th>Mouse 5-HT3A</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>53.7%</td>
<td>48.6%</td>
<td>24.3%</td>
</tr>
<tr>
<td>β</td>
<td>45.4%</td>
<td>46.6%</td>
<td>27.3%</td>
</tr>
<tr>
<td>γ</td>
<td>40.4%</td>
<td>45.5%</td>
<td>21.9%</td>
</tr>
<tr>
<td>δ</td>
<td>39.6%</td>
<td>44.4%</td>
<td>26.3%</td>
</tr>
</tbody>
</table>

In each row, the highest value is highlighted in bold. Sequence identities are normalised to the length of the shorter sequence in each case.

An issue with the multiple template approach used with these templates and MODELLER is that while the Torpedo nAChR and 5-HT3A receptor models follow the convention of numbering the five chains of the pentamer in a clockwise direction, in the α4β2 models the chains are assigned anticlockwise (Figure 6.8). Since chains were not assigned in the same order, modelling with these templates led to ‘flattened’ models. The solution was to reassign each chain in the α4β2 template in a clockwise fashion, keeping the two α4 subunits assigned as chains A and D, i.e. the yellow α4 subunit in the α4β2 receptor in Figure 6.8 is reassigned as chain A and the chains were renumbered in a clockwise direction from there. A key point is that merely renaming the chains, for example using UCSF Chimera’s ‘Change Chain IDs’ feature or using a MODELLER script to renumber segments, did not work as the atomic coordinates appeared in the same order within the PDB file. Finally, splitting the PDB file corresponding to the α4β2 nAChR (6CNJ, with antibody fab fragments removed) into its constituent chains, copying each chain in the order they needed to be assigned in the final template PDB file (in the order D, C, B, A, E), changing the chain IDs (D -> A, C -> B, B -> C, A -> D, E -> E) and then combining the copied and renamed chains such that all chains appeared in the PDB file in the correct order and were assigned with the correct chain ID gave a template model that worked for modelling.
Figure 6.8: Order of chain assignment in template models for nAChR. Chains are coloured from A-E in blue to red following rainbow order colouring. Chains in the Torpedo nAChR (PDBid: 2BG9) and 5-HT₃A (PDBid: 6BE1) are assigned in clockwise direction while chains in the human α4β2 nAChR (PDBid: 6CNJ) are assigned in an anticlockwise direction. Subunit types are shown for the heteromeric Torpedo nAChR and human α4β2 nAChR, but not for the mouse 5-HT₃AR as this is a homomer and all subunits are equivalent.

As before, modelling was carried out using MODELLER v9.17, the script and alignment used, carried out multi-template and multi-chain modelling, as well as reading heteroatoms relating to ligands and water from the template models (Appendices 1.4.1 and 1.4.2). The positions of glycosylation from the α4β2 subunit were used (owing to higher sequence identity of α4β2 to Torpedo compared with 5-HT₃A). However, the positions of water and sodium in the channel pore were taken from the 5-HT₃A model due to the more similar functional state. The 5-HT₃A conformation is thought to be representative of a closed-resting state compared with the desensitized state of the α4β2 nAChR structure. The top portion of the channel is narrower in 5-HT₃A than it is in the α4β2 nAChR. This leads to a sodium ion being captured much higher in the 5-HT₃A channel than in the α4β2 nAChR. The generated initial models were ranked with the built-in MODELLER statistical potential DOPE and the top model fitted into the raw unfiltered map. The fit of the initial model appeared to be good (Figure 6.9).
6.3.3 Map sharpening

Experimental density maps determined from single-particle approaches in electron microscopy suffer from contrast loss at high resolutions. This contrast loss is due to the images themselves rather than the sample. Image contrast may be degraded by factors such as; sample drift, radiation damage, inelastic electron scattering, flexibility of particles and sample heterogeneity (18). The loss of contrast can be modelled as a Gaussian fall-off dependent on Equation 6.1:

\[ e^{-\left(\frac{B_{\text{overall}}}{4d^2}\right)} \]

(6.1)

where \( B_{\text{overall}} \) (a positive value) is the ‘temperature factor’ of the image due to the above described loss of contrast and also the reduction in contrast caused by the computation of averaging the images, and \( d \) is the resolution (13). Therefore, raw and unfiltered density maps from electron cryomicroscopy appear smooth and featureless as the high-resolution information is down weighted due to contrast loss. It is important to restore the high resolution
features by rescaling high and low resolution structure factor amplitudes to have the correct relative scaling. This process is known as sharpening.

6.3.3.1 Global sharpening

Experimental temperature factors which cause loss of contrast ($B_{\text{overall}}$, Equation 6.1) are positive. Therefore, application of a negative temperature factor to a gaussian of the same form to that which approximates contrast loss, can restore the correct scaling of structure factor terms at high and low resolution (13). The problem with scaling in this way is that a negative B-factor will increase the weight of noise as well as signal at high resolution which can result in a noisier map if over sharpened. As such, the size of the negative B-factor applied must be appropriate for the signal-to-noise ratio of the map.

In reality, most density maps are anisotropic in their resolution and signal-to-noise ratios, therefore it is not trivial to select a B-factor for scaling the density and typically a variety of B-factors may be trialled to achieve optimum contrast, however this process is subject to operator bias.

6.3.3.2 Local sharpening

Recently, methods that take into account the local environment have been developed for sharpening, such as LocScale (19). Such approaches offer a more automated method of sharpening, less prone to user bias and which, in some cases, can produce better results than global sharpening.
Figure 6.10: LocScale procedure. The reference model, in this case a comparative model, was used to prepare a reference model map, scaled with the expected atomic B-factors (top left). The model map along with the mask and the raw map were fed into the LocScale script to produce a locally scaled map.

LocScale produces a model map from a reference model, weighted by reference atomic B-factors (temperature factors) determined for that model. The raw density map is the scaled to match the expected averaged radial profile of a rolling cubic window of voxels from the reference model map (19). While the model produced for this map is a comparative model, the atomic B-factors were transferred over from the template models used to produce it and can be used for local scaling with LocScale (19). From the comparative model which was used as a reference model, a model map was generated using the ‘prepare_locscale_input.py’ script supplied with the LocScale package (19). Along with the soft mask used in the calculation of FSC curves and the raw map, the reference model map was used in LocScale to produce a locally scaled map which showed significantly more features than the raw map and did not appear to be oversharpened (Figure 6.10).

6.3.3.3 The scaled map

The reference model appears to be a relatively good initial fit to the scaled density map. Although many regions could certainly improve in fit, the real space cross correlation was calculated as ~0.7 (calculation of real space cross correlation must take account of the local scaling, for this a script written
by the author of LocScale (Dr Arjen Jakobi), utilising the computational crystallography toolbox (CCTBX, (20)) was used) suggesting the model is applicable as a useful tool for interrogating the features of the scaled map, even if an imperfect model of the true receptor geometry.

Investigation of the scaled map with the rigidly fitted reference model revealed a number of interesting features at a variety of density thresholds. At high $\sigma$ (threshold) levels (26.6), it seems clear that the N-terminal helices better suit the more compacted conformation modelled based on the $\alpha 4\beta 2$ nAChR compared with the more extended conformation that was found in the previous Torpedo nAChR maps (Figure 6.11A); the MX helices appear to be better resolved than in previous maps and at $\sigma$ 21.6 some of the loops connecting TM3 to MX are also resolved (Figure 6.11B), furthermore, there is good density for all of the TM1-TM2 loops which were one of the original sources of error in the previous model (Figure 6.11C); at $\sigma$ 13.2 much of the density for the divisive TM2-TM3 loops seems to be resolved and it looks to agree much better with the newly modelled geometry of the loop from the $\alpha 4\beta 2$ nAChR compared with the previous model (PDBid: 2BG9) (Figure 6.11D); also at $\sigma$ 13.2 density appears to be present in the ion channel, the location of this density corresponds well with the position of a sodium ion and water network modelled based on that in the closed-resting state of the 5-HT$_3$AR (Figure 6.11E); finally, densities are present directly over known glycosylation sites in the map and these correspond with the modelled positions of N-acetyl glucosamine from the $\alpha 4\beta 2$ nAChR template (Figure 6.11F).
Figure 6.11: Features of the sharpened map. A, view from the top of the receptor, condensed geometry of the N-terminal helices matches that of the map well. B, view from the bottom of the receptor, some TM3-MX loops are resolved. C, slice through the receptor at the lower leaflet of the membrane, viewed from the top, density is present for all five TM1-TM2 loops based on the fitted geometry of the initial model. D, slice through the receptor at the upper leaflet of the membrane, viewed from the top the shape of the new map fits the newly modelled extended geometry of the TM2-TM3 loop (grey model) better than the original geometry of this loop (blue, PDBid: 2BG9). E, view into the pore of the receptor from the top a modelled sodium ion (purple) corresponds well to density within the ion channel pore. F, glancing view from the side of the γ-subunit extracellular domain, an example of density found directly above a known glycosylation site with an N-acetyl glucosamine residue modelled in, the example shown here is for the Asn 140 residue on the γ-subunit. In all cases, the scaled density map is shown as translucent grey, the comparative model built in this chapter in grey and the original model (PDBid: 2BG9) in blue.

Comparison of the initial reference model with the map also helped to show areas that require improvement in the model. Since the C-loops are modelled based on an agonist bound conformation of the α4β2 nAChR, these are in the wrong conformation in this initial model and that is reflected in the shape of the density (Figure 6.12A σ 18.5). Also it can be seen clearly at high thresholds (σ 26.6) that the TM4 helices do not fit their corresponding density well (Figure 6.12B). The TM4 helices of the Torpedo nAChR have previously been found to exhibit conformational flexibility (9), so it is not surprising that those modelled based on the homologous 5-HT3AR would not fit the density well. Furthermore, the conformation of these helices can be dependent on local environment, and the difference in environment between the detergent solubilised 5-HT3AR and the native lipid embedded Torpedo nAChR may also give rise to differences in conformation.
Figure 6.12: Poorly fitted areas of initial model. A, C-loop takes on geometry of agonist bound conformation due to template model having an agonist present (α4β2 nAChR, PDBid: 6CNJ), although no agonist is present in the preparation of the receptors which was imaged. Density for the C-loop also confirms that this geometry is likely to be incorrect. B, TM4 helices sit partially outside of the density map.

6.3.4 Model fitting

A coarse grained flexible fitting can be carried out using iMODFIT (21) holding the internal coordinates rigid for all secondary structure elements by using the additional "--fixSS HE" tag. Doing so visually improved the fit of TM4 helices and increased the real-space cross correlation by about 6% to 0.74, however, the C-loop geometry still was not improved. Further, or more fine grained, fitting with iMODFIT at the resolution of this map would be dangerous and likely to introduce overfitting. In the fitting of atomic models to density maps overfitting could be seen as increasing the cross correlation coefficients by fitting the model to noise in the map. Since the reconstructions probably already contain a significant contribution from noise at high spatial frequencies (as judged by the FSC curve (Figure 6.7)) overfitting of the model is even more likely.

The C-loop geometry is almost certainly wrong in this model because it was modelled based on a receptor in a nicotine bound conformation and is therefore closed. A safer option for modelling this loop may be to generate an ensemble of conformations for the loop and choose that which fits the density best in this area for each subunit. Local cross correlations can be calculated using the segment based cross correlation coefficient example script in TEMPy (get_sccc.py) (22). However, given the low resolution of the map, it
may be better to wait for improved resolution after the addition of more images to the dataset.

Another region of the model which requires further investigation is the junction between the TM4 and MA helices. These regions were modelled based on the template of the 5-HT₃AR structure (PDBid: 6BE1) and thus, as in the 5-HT₃AR are represented as a single continuous helix. However, in the previous models of the Torpedo nAChR, there is a break between the MA and TM4 helices (PDBids: 2BG9, 4AQ5, 4AQ9). In the 5-HT₃AR the MA helices are implicated in channel conductivity owing to the presence of three positively charged arginine residues which control the passage of cations through portals to the cytoplasm on the obligate pathway of conductance (Chapter 2, Figure 2.3) (23, 24). This might suggest that the rigidity incurred by the formation of a single helix between MA and TM4 is necessary for the biological function of MA in the 5-HT₃AR. However, in the Torpedo nAChR the five cytoplasmic portals, framed by the MA helices are instead lined with neutral or negatively charges residues which would theoretically stabilise incoming cations flowing through an open channel (Chapter 2, Figure 2.3) (11). It is unlikely that these helical segments play as vital a role as those in the 5-HT₃AR in the conduction pathway. However, the MA region may be influential in the functional changes observed between fetal (long channel open times) and adult (short channel open times) muscle-type nAChRs where the γ subunit is replaced by an ε subunit (25).
Figure 6.13: Secondary structure predictions for the MA and TM4 helical regions of the *Torpedo* nACHR from JPred4. From top to bottom JPred4 output is shown for α, β, γ and δ subunits. JNETCONF shows the overall confidence in the prediction of secondary structure at each residue position. Red tubes are the prediction of helices, green arrows are the prediction of strands.

Interestingly, the secondary structure prediction server JPred4 (26) agrees with the notion that a break is likely between the MA and TM4 helices of the *Torpedo* nACHR with at least one point in the middle of the sequence attributable to MA and TM4 having a confidence of 0 in the assignment of helical secondary structure for each subunit (Figure 6.13).
It is therefore likely that a break should be present between the MA and TM4 helices for the Torpedo nAChR. Unfortunately the resolution of the map is currently not good enough to confidently assign this break. I am hopeful that the origin of the break will become clear as the quality of the map increases from the addition of further images to the reconstruction dataset and optimisation of the reconstruction process. For now, the atomic reference model is best treated as a useful starting point for fitting to future higher resolution iterations of the density map.

6.3.5 Insight into functional data

A great deal of functional data has been collected relating to nAChRs especially muscle type nAChRs due to their high sequence identity with the Torpedo nAChR which was the first pLGIC to have its structure determined at atomic resolution.

6.3.5.1 Insights into channel gating

The lab of Anthony Auerbach has spent a great deal of effort investigating the gating kinetics of single muscle-type nAChRs to understand the order in which the allosteric gating motion occurs. His lab’s studies have led to a hypothesis that binding of acetylcholine at the orthosteric site causes a conformational wave that travels through the receptor in blocks of residues, finally causing the opening of the channel gate and allowing the influx of ions (27). Using single channel electrophysiology techniques in combination with engineered single point mutations to study linear free-energy relationships, the Auerbach lab have been able to determine the contribution of a vast number of amino acid residues to the gating mechanism, throughout the nAChR sequence, even suggesting the order in which the residues must move (27, 28). The residues investigated can be divided into five groups (so-called Φ-blocks) based on clustering by their associated Φ value, which relates to their position in the sequence of events that occur during the mechanism of channel gating (29).

The structural mapping of the positions of these residues is absolutely essential to the understanding that this analysis brings to the spatio-temporal
conduction of an allosteric mechanism. Initially, these residue positions for the mouse muscle-type nAChR (that most studied in the Auerbach lab) were mapped to their equivalent residue positions in the Torpedo nAChR structure (PDBid: 2BG9) (30). However, owing to the aforementioned errors in this structure, this led to difficulties in interpretation of many of the residue positions, with some ‘early moving’ residues appearing to be isolated, rather than being structurally clustered as expected. To overcome this problem, residue positions have been mapped onto more distant homologues such as the C. elegans glutamate gated chloride channel (GluCl, PDBid: 3RIF) and the Gloeobacter violaceus Ion Channel (GLIC, PDBid: 3EHZ). Under this mapping, the residues within the same Φ-block (i.e. expected to move together in the gating mechanism) appear to be broadly in the same regions of the receptor (Figure 6.14A) (29). Most notably, those moving first are largely grouped around the orthosteric binding site, and those moving last are found lining the channel lumen (Figure 6.14A) (29). However, GluCl and GLIC are homomeric rather than heteromeric receptors as in the muscle type nAChR, and only the α subunit residues were mapped onto these receptors (29). However, the muscle-type nAChR contains no α-α interfaces.

Therefore, this may be a misrepresentation of the gating pathway since subunit interfaces are likely to play a significant role in transduction of the signal of acetylcholine binding at two distinct sites to channel opening across five subunits.

One aspect that is difficult to unpick from this analysis is whether or not the values measured for the α subunit residues are accurate. Two α subunits are present in two different environments within the muscle-type nAChR, that which forms the principal subunit at the interface with the δ subunit, αδ, and that which forms the principal subunit at the interface with the ε subunit, αε, equivalent to αγ in the Torpedo nAChR. The αδ and αε/αγ subunits are in different environments and this affects properties such as affinity and dissociation constants for orthosteric agonists, for example, differences of up to ~100 fold have been suggested for the dissociation constant for acetylcholine between the two sites (31). Therefore, in regions where signal transduction of binding may differ between the two α subunit environments,
measured Φ values could be averages of two positions in a pathway, potentially convoluting the molecular information.

I have mapped the residues coloured by their Φ-block onto the previous *Torpedo* nAChR model and new *Torpedo* nAChR model produced in this work (Figure 6.14). The geometry seen in the new model appears to help rationalise some of the data that seems inexplicable when interpreted via the framework of the previous structures. Most notably, in the previous model, residues from a given Φ-block appeared isolated at subunit interfaces (Figure 6.14B and C). However, with the correct register and TM2-TM3 loop geometry, the transmembrane domain-extracellular domain intersubunit interfaces are better represented. In the new model, the previously isolated residues are placed in close proximity to residues of the same Φ-block on adjacent subunits (Figure 6.14B and C). The change in Cα distance for the closest block-one residues across the αγ-γ interface goes from ~14.5 Å in the old model to ~8.5 Å in the new model (Figure 6.14B, purple residues in upper inset box). This could help to better explain how the signal of binding is transduced to channel opening from α subunits to non-α subunits which do not actively partake in acetylcholine binding. Similarly, at the same interface, a series of block-four residues which are spread out along TM3 in the previous model become bunched together in a tight cluster in the new model, and are situated closer to the middle of the transmembrane domain, which we would expect to move later in gating (Figure 6.14B, red residues in lower inset box). In another example, at the γ-αδ interface, block-three residues on the αδ subunit which were previously isolated, are placed in extremely close proximity to a block-three residue on the γ subunit TM2-TM3 loop. Many pairings of residues in the same Φ blocks may exist across subunit interfaces by contacts between the TM2-TM3 loop of the principal subunit and the extracellular domain of the complimentary subunit. Such a phenomenon would offer an elegant explanation for how signal is transduced from agonist sensing to non-agonist sensing subunits and would agree with the suggested signal transduction pathway in homologous glycine receptors (32).
Figure 6.14: Mapping Φ values to residues on old and new models of Torpedo nAChR. A, general positioning of residues in the same Φ-block are clustered throughout the structure in both old (right) and new (left) models of the nAChR. B, αγγ interface. Right, two early moving residues in block-one (purple) are isolated on the γ subunit, in the new model, left, these two residues are in close proximity to another block-one residue on the α subunit at the transmembrane-extra cellular domain interface (upper inset box). Late moving residues in block-four (red) are dispersed along TM3 in the old model (right), in the new model (left), these are tightly clustered in a region that may interact with cholesterol based on α4β2 nAChR structures (PDBid: 6CNJ, 6CNK) (lower inset box). C, γαδ interface, an isolated α5 block-three residue (green) in the old model (inset box, right) is placed next to a residue of the adjacent γ subunit from the same Φ-block in the new model (inset box, left). Residues in Φ-blocks 1-5 are coloured purple, blue, green, red and brown respectively and shown as spheres, as in (29). Residues where the ribbon is coloured yellow, were found not to play a role in the gating mechanism.

6.3.6 Photo-crosslinking data

The lab of Jonathon Cohen has used a different method to probe interactions of the Torpedo californica nAChR on a molecular scale. Photo-crosslinking followed by digestion and Edman degradation has been utilised to find locations related to binding for a number of nAChR ligands including a variety of barbituates (33). In one example, the small molecule, etomidate, was modified to contain a photo-reactive trifluoromethyl diazirine group and found to cross-link to several residues on the Torpedo nAChR (34). When mapped onto the original Torpedo nAChR structure, these locations can be roughly clustered into three areas, within the channel pore, at the γαδ interface and within the δ subunit (Figure 6.15). The residues found cross-linked at the pore are tightly clustered together and facing inwards. However, crosslinked residues within the inter and intra-subunit sites are relatively far from each other (Figure 6.15B and C, top), perhaps suggesting that the representation of the binding site is incorrect or that etomidate binding at these sites has a high degree of flexibility.

When the positions are mapped onto the new model of the Torpedo nAChR however, the pore lining residues are shifted higher in the pore due to the change in register. The ring of pore lining residues now sits directly below the suspected position of a pore bound sodium ion (Figure 6.15A), which would match up well with the expected solvent accessibility of the channel and the geometric position of the gate residues in other homologous channels (35). Cross-linked residues at the γαδ interface are marginally closer together, now the residues on γ-TM3 and α5-TM2 are slightly more rotated towards each
other (Figure 6.15B). Most dramatically however, the crosslinked residues at the δ intrasubunit site which were previously disperse are now tightly clustered together, indicative of a highly specific binding site (Figure 6.15C).

6.4 Conclusions and future work

In conclusion, recent advances in the structural determination of pLGICs has led to high resolution structures of two homologues of the Torpedo nAChR. In conjunction with new structural data for the Torpedo nAChR at an approximate resolution of 7.6 Å, this has allowed the construction of a model for all subunits of the Torpedo nAChR which accounts for errors in the transmembrane register due to mis-assignment of the TM1-TM2 and TM2-TM3 loops in the original model. Alongside this, new structural coordinates for the MX helix and TM3-MX loop have been added to the current model.
Local sharpening of the new *Torpedo* nAChR maps based on the new models reveals potentially interesting features, including loop orientations, as well as possibly a pore bound ion and glycosylation of the extracellular domain. The new models provide a relatively good starting fit to the preliminary maps and flexible fitting improves this. However, the maps are very much preliminary and further refinement is required to ensure they are not overfitted and to improve the resolution. New images are still being collected and added to the dataset which feeds into the reconstruction of these maps which will hopefully improve the resolution in future. There are some problematic regions of the new model for example the TM4 and MA helix geometry, furthermore, whether or not there should be a break or not between these two helices. This can only be confidently assigned with higher resolution data so has not been fully addressed.

Despite the observation that some regions of the model are likely to be slightly wrong in detail, the new models were employed in the rationalisation of previously collected biophysical and biochemical data from the Auerbach and Cohen labs. Compared with the previous model of the *Torpedo* nAChR, the new models provide a better rationalisation of mechanistic data relating to the opening and closing of the channel, and also to photo-crosslinking data, lending weight to their worth as improved models of the *Torpedo* nAChR. Hopefully future improvements in the structural data will allow for the atomic fitting of the new models and thus improve the quality and relevance of these models, which are still the only pLGIC models determined in a native membrane.

### 6.5 References


18041.


34. Hamouda AK, Stewart DS, Husain SS, Cohen JB (2011) Multiple transmembrane binding sites for p-trifluoromethyl-diazirinyl-etomidate, a

Chapter 7

7 General summary and future work
7.1 Introduction to thesis summary

This final chapter aims to bring together the findings from each of the chapters in this thesis and present some potentially interesting directions of future work. The summaries are broken down into two broad areas of; improvements to the Torpedo nAChR structure and α7 nAChR studies. Finally some general points on the future direction of the work in this thesis are presented.

7.2 Improvements to the Torpedo nAChR structure

The Torpedo marmorata nAChR structure offers a unique opportunity for modelling the TM domain of an nAChR, as it remains the only example of a pLGIC imaged in its native membrane environment.

7.2.1 Correcting the transmembrane register error in the Torpedo nAChR αγ subunit

The origin of a transmembrane error was characterised and a hierarchical refinement protocol developed in Chapter 2. This protocol, which corrected the register error in the TM domain of the αγ subunit included comparative modelling and real-space flexible refinement into the available density maps for the Torpedo nAChR in open and closed conformations, at 6.2 Å resolution. Evaluation of the refined models suggested that they were of higher quality than the original templates and therefore they were utilised in the building of comparative models of the α7 nAChR subunit.

The transmembrane error originating from the assignment of an extra helical turn was tackled for the αγ subunit of the Torpedo nAChR in Chapter 2. However, this error is also incorporated into the other four subunits of the model. Furthermore, alternative conformations of the TM2-TM3 loop were not modelled during this work. This was largely due to the lack of structural data in this region at the time of carrying out the research. Therefore, the true TM2-TM3 loop conformation for nAChRs was still ambiguous as the Torpedo structure was the only nAChR model available. In the 5-HT₃A R structure (PDBid: 4PIR), which was the only other available structure of a cationic pLGIC, the TM2-TM3 loop was not fully resolved. More recently however,
new structures of the 5-HT$_{3A}$R, and the $\alpha_4\beta_2$ nAChR have become available. In these new closely related structures, the TM2-TM3 loop is resolved and appears to be in a similar conformation to that of other pLGICs, i.e. protruding toward the complimentary subunit rather than within the subunit as it is in the *Torpedo* nAChR. The main future direction of this modelling project is therefore to remodel all five subunits of the *Torpedo* nAChR correcting both the TM register error and the TM2-TM3 loop error. The assignment of MX helices in 5-HT$_{3A}$R and $\alpha_4\beta_2$ may also allow the incorporation of these structural units into the *Torpedo* nAChR model too. These points are partially addressed in Chapter 6 (see below).

### 7.2.2 Correction and fitting of the full length *Torpedo* nAChR including new structural data

The publication of the revised $\alpha_\gamma$ subunit models led to a collaboration with Prof. Nigel Unwin (LMB, MRC, Cambridge) and the receipt of new, unpublished cryo EM structural data relating to the *Torpedo* nAChR. The new structural data, having been reconstructed by Prof. Unwin using the more recent helical refinement protocols in RELION, is thought to be free of systematic errors which may have marred the previous EM structures of the *Torpedo* nAChR from his lab. Indeed, differences including in the length of the reconstruction can be seen from comparison of the new map versus the older maps. In combination with the latest structural data relating to cationic selective pLGICs (the 5-HT$_{3A}$R imaged by cryo EM and the ($\alpha_4)_2(\beta_2)_3$ nAChR imaged by cryo EM), new comparative models of all five subunits of the *Torpedo* nAChR were built in Chapter 6.

Local sharpening of the new maps using B-factors transferred from the experimental models of 5-HT$_{3A}$R and $\alpha_4\beta_2$ nAChR structures to the comparative model revealed new and interesting features from the raw maps. These included the possible resolution of a central pore bound ion, improved resolution of the TM1-TM2, TM2-TM3 and TM3-MX loops, and potentially the resolution of some regions of glycosylation around the annulus of the extracellular domain. Significantly, the majority of the TM2-TM3 loops can be resolved, suggesting that the loop conformation that protrudes towards the
adjacent subunit is indeed more likely the correct conformation of this loop. Flexible fitting of the latest *Torpedo* nAChR comparative model to the sharpened map was attempted. However, the resolution (estimated at ~7.6 Å by FSC of two independently constructed half maps) was too low to confidently fit the model, even though some regions, such as the C-loops, almost certainly need adjustment. Since the map is preliminary and more data is yet to be added to the dataset, which may improve the resolution further, it would be sensible to wait for a better resolution map to attempt more robust fitting of the model to the map.

Despite the imperfect fit, the new model was used to re-evaluate previously published functional data relating to muscle-type or *Torpedo* nAChRs including; Φ value analysis from linear free energy relationships studied in the lab of Anthony Auerbach; and chemical crosslinking data from the lab of Jonathon Cohen. In both cases, the mapping of residues on the new model as opposed to the previous model aided analysis of the data and brought improved understanding, suggesting that the models produced here are more correct than the previous models. As mentioned above, there are regions of the model that require improvement such as the C-loop conformation, but also the MA-TM4 junction. In the near future, the resolution of the maps may be improved as further images are added to the dataset. This will hopefully allow more robust fitting of the model into the map.

**7.3 α7 nAChR studies**

After building corrected models of the *Torpedo* nAChR αγ-subunit in Chapter 2, construction of α7 pentamers based on the putative open and closed conformation αγ models resulted in pentameric models containing pore sizes that match the expected closed to open dilation. These represent useful models of α7 in two functional receptor conformations. In Chapter 3, a consensus docking protocol was developed, for use in conjunction with the α7 nAChR models, to study the differences in binding between three families of α7 positive allosteric modulators. Our collaborators in the Millar lab found that: while all three of the parent compounds (A-867744, TBS-516 and TQS) elicited classical type II PAM effects on the wild-type α7 nAChR, the action of
all three compounds was affected by a range of mutations spanning extracellular and transmembrane locations. Interestingly, the effects of these mutations were always different for A-867744 compared with TBS-516 and TQS, which behaved in the same way as each other on mutant receptors.

### 7.3.1 Predicted binding modes of α7 nAChR PAMs

Merely, docking the three compounds to the transmembrane region of the α7 nAChR model did not provide useful results for interpreting the functional data. However, the use of consensus between two docking programs and amongst the compound families related to each of the parent compounds (from structure-activity relationship studies) was able to derive consistent binding modes. This newly-developed consensus protocol predicted that all three compounds bound at an *intersubunit* cavity which is different from the *intrasubunit* cavity previously proposed for some of these compounds (based on α7 models built using an uncorrected Torpedo nAChR). TBS-516 and TQS were predicted to bind in similar geometries to each other in both open and closed conformations of the α7 nAChR models. These binding modes were also similar between the two modelled receptor conformations and provided a good rationalisation for their similar behaviour on mutant receptors. A-867744 however, was predicted to bind in a different manner, but at a mutually exclusive site. This finding also helped provide a rationalisation for the different behaviour observed for A-867744 on each of the mutant receptors. Furthermore, unlike TBS-516 and TQS, A-867744 appeared to require a change in binding mode on the conformational shift of the receptor from closed to open. Although highly speculative, this requirement for a change in binding mode could rationalise the way in which A-867744 is able to act as an inhibitor (or negative allosteric modulator) for some mutant receptors as the closed conformation binding mode would block the necessary rearrangement of the transmembrane subunit interface during channel gating.

Further validation of the predicted binding modes for A-867744, TBS-516 and TQS is also a priority. While the rationalisation of the functional data generated in the Millar lab provides some level of validation, many mutant receptors have a vast array of effects on the action of allosteric modulators.
and these can be in close proximity or distant from a binding site. Therefore, testing of mutant receptors is perhaps not the best method of binding mode validation as the results are easy to misinterpret. The subsequent chapters 4 and 5 make some progress towards further validation of binding modes.

7.3.2 Reactive allosteric modulators

Covalent incorporation of an allosteric modulator and subsequent peptide sequencing would be able to define the exact sites of reaction. Mapping these to the structural models of \( \alpha_7 \) would provide a solid validation of the predicted binding modes.

In Chapter 4, reactive analogues of the allosteric modulator TQS were designed and synthesised. Two approaches were sought, in one approach a photoactivatable reactive centre (a trifluoromethyl diazirine) was installed at the phenyl ring substituent of TQS. When tested in electrophysiology on *Xenopus* oocytes expressing \( \alpha_7 \) nAChR by our collaborators, this compound (4TFD-TQS) was found to be a potent ago-PAM. This suggested that the incorporation of the reactive group had not perturbed the PAM like behaviour of the TQS scaffold. UV irradiation studies also suggested that 4TFD-TQS was efficiently degraded under UV light. However, cross-linking to the receptor could not be confirmed by electrophysiology (carried out in the Millar lab by Dr Anna Chatzidaki). In a second approach, cysteine reactive substituents were added to the same position on the TQS-like scaffold. Two potentially cys-reactive TQS analogues were synthesised, one containing a benzyl chloride substituent (4CM-TQS) and one containing an \( \alpha \)-chloroacetate group (4ACA-TQS). The two different reactive groups allow different amounts of freedom of exploration of the binding site, the \( \alpha \)-chloroacetate potentially reaching amino acid residues further from the binding site. So far, only 4CM-TQS has been tested in electrophysiology by our collaborators, but it too was found to be a potent ago-PAM.

Superposition of these novel reactive TQS analogues on the predicted binding modes for TQS from the consensus docking protocol facilitated the prediction of amino acid side chains which may be able to react with either the photoreactive or cys-reactive analogue (given the mutation of the respective
side-chain to a cysteine residue). Future work in this area would be the further testing of the synthesised compounds in order to determine if covalent incorporation is possible. It would also be interesting to test the cys-reactive compounds in mutated α7 nAChRs containing cysteine residues at sites predicted to be able to react with the benzyl chloride or α-chloroacetate groups.

7.3.3 Predicting new α7 nAChR allosteric modulators

In Chapter 5, the clusters of binding modes produced for the TQS type compounds from Chapter 3 were utilised to build pharmacophore models of binding. Query validation runs suggested that the pharmacophore models were of high quality and were able to accurately discriminate between known ligands and likely non-binding compounds. The best queries for the open and closed conformation binding-mode clusters were used to screen the DrugBank database, which is a small curated database of approved and experimental drugs. The hits from virtual screening were filtered by taking forward only hits that appeared in searches with both open and closed conformation binding-mode pharmacophores, and those that scored 4.0 or more in a multi-parameter optimisation that is focussed towards CNS penetrant properties (the CNS MPO).

The top 25 ranked compounds were clustered into groups based on the annotated pharmacological target in DrugBank and it was found that four compounds represented all targets in the top 25 compounds. These four compounds – DB04763, DB08122, furosemide and pefloxacin – were all found to be active as allosteric modulators of α7 nAChRs when assayed by two-electrode voltage clamp electrophysiology by our collaborators in the Millar lab. DB04763, DB08122 and pefloxacin all appeared to act as negative allosteric modulators while furosemide was found to potentiate acetylcholine induced responses of α7 nAChRs with a profile representative of a type I PAM. Although all the compounds used to build the pharmacophore query are either type II PAMs or ago-PAMs, it is not necessarily surprising that the four tested hit compounds have different effects as even very small changes in structure (such as methyl substitution around a phenyl ring) have previously
been shown to cause variation in allosteric modulatory effects from type I and type II PAMs to allosteric agonists, negative allosteric modulators and silent allosteric modulators. These results therefore show the strength of this virtual screening method.

Docking of furosemide back into the \( \alpha_7 \) nACHR models using the consensus protocol from Chapter 3, showed that furosemide was predicted to bind with modes similar to those predicted for TQS type compounds. Three residues close in proximity to the predicted furosemide binding modes were mutated to alanine by our collaborators and all three were found to abolish potentiation of acetylcholine induced currents by furosemide. This also provides further validation for the predicted binding modes.

In future directions for this project, it would be interesting to see if changes to the furosemide structure could affect the binding of these compounds, or possibly convert furosemide from a type I PAM to another type of allosteric modulator as has been seen in the past for TQS-type compounds. Some interesting structures to test might be an analogue of furosemide but without the chloride substituent to see if its presence is necessary for action (Figure 7.1). It would also be interesting to see if closing the second ring preserves activity, for example as an oxo-oxazine, quinolone, or isatin (Figure 7.1). Substitution of the furan for other heterocycles could also be interesting, for example a pyrrole may be able to act as a hydrogen bond donor instead of as an acceptor (Figure 7.1). Subtle modifications to the structure of furosemide, such as those above, may allow the discovery of a novel chemo-type for \( \alpha_7 \) nACHR PAMs. Alternatively, perhaps a better hit could be acquired if a larger or more diverse database is screened.
Figure 7.1: Interesting furosemide analogues. R = Cl or H. From left to right, furosemide analogue without Cl, in all other structures Cl may or may not be present depending on its necessity for action, 8-oxo-oxazine furosemide analogue, 8-oxo-oxazine furosemide analogue with furan substituted for pyrrole. Quinolone furosemide analogue, isatin furosemide analogue.

7.4 General points for future work

7.4.1 Improvements and distribution of the consensus docking protocol

The implementation of the consensus docking protocol was successful and it would offer a useful framework for other researchers interested in docking. However, the execution of this protocol is slow and tedious, owing to many disconnected and independently written scripts and programs. One avenue of future work could be to tidy up the execution and analysis by this method to make its use by other researchers simpler. For example the production of a webserver into which a protein structure and a series of compounds could be loaded and consensus docking carried out could offer a useful and user friendly tool for many researchers.

7.4.2 Further exploration of reactive allosteric modulators

Although two of the three potentially reactive allosteric modulators synthesised in Chapter 4 have been characterised as ago-PAMs, none have been confirmed to successfully cross-link to the nAChR. For the photo-activatable PAM 4TFD-TQS, the high reactivity of the carbene produced may result in quenching of the majority of the reactive PAM in solution before it is able to cross link to the receptor. In the case of the cys-reactive PAMs, these were designed to have relatively low reactivity to prevent cross-linking with any other nucleophilic residues on the protein. However, this may result in slow reaction rates and inefficient crosslinking.
Alternative photo-reactive groups could be interesting to try, such as a benzophenone, or a more activated azide, these may have a slightly longer residence time than the trifluoromethyl diazirine and possibly give better results for photo-crosslinking. More reactive electrophiles could also be synthesised, for example the bromo- or iodo- analogues of 4CM-TQS and 4ACA-TQS, or a maleimide containing compound (Figure 7.2).

Figure 7.2: Potential alternative reactive allosteric modulator analogues. The two compounds on the left could be photoactivatable while the three on the right would be cysteine reactive.

Alternatively, the reactive group could be installed at a different location, such as on the alkene bond of the cyclopentene fragment, or at a different position on the phenyl ring. Another route might be to extend the linker between the PAM portion and reactive portion of the molecule, to allow more flexibility of the reactive group. This could help to overcome a situation where the reactive group is too far from any sidechains to successfully crosslink. However, it may also decrease the resolution of any data locating the binding site obtained from this method.

7.4.3 Further exploration of the properties of furosemide

Furosemide - identified as an α7 nAChR PAM in Chapter 5 - has also been associated with the modulation of other pLGICs such as GABA receptors. One study used a model of epilepsy, in which the α4 nAChR gene was mutated, and showed a prophylactic effect through application of furosemide. The prophylactic benefit of furosemide usage was thought to be effected through modulation of GABARs. However, the finding that furosemide is able to potentiate acetylcholine induced responses in α7 nAChRs may suggest
that it is also able to bind to and modulate other nAChRs such as α4β2. Atomic models are now available for GABARs and α4β2 nAChRs. An interesting route of further work would be to investigate the possible binding and action of furosemide on these receptors.
Chapter 8

8 Experimental
8.1 Chemistry

8.1.1 General Chemistry

All reactions were carried out at atmospheric pressure with stirring unless otherwise stated. All reagents and solvents were purchased from suppliers and used without further purification unless otherwise stated. Cyclopentadiene was distilled from cracked dicyclopentadiene prior to use.

Reactions were monitored by TLC (thin layer chromatography) or $^1$H NMR (nuclear magnetic resonance) as stated. TLC plates pre-coated with silica gel 60 F254 on aluminium (Merck KGaA) were used, being visualized by UV (ultraviolet, 254 or 365 nm) or chemical stain (phosphomolybdic acid (PMA), ninhydrin, 2,4-dinitrophenyl hydrazine (DNP) or Dragendorff’s reagent). Normal phase silica gel (BDH) was used for flash chromatography.

$^1$H NMR spectra were recorded at 300, 500, 600 or 700 MHz, $^{13}$C NMR spectra were recorded at 75, 125, 150 or 175 MHz on Bruker Avance 300, Avance 500, Avance III 600 with cryprobe or Avance Neo 700 machines at ambient temperature, unless otherwise stated; all chemical shifts were referenced to the residual proton impurity of the deuterated solvent. All spectra were recorded at 298 K. The multiplicity of the signal is indicated as s (singlet), d (doublet), t (triplet), dd (doublet of doublets), dt (doublet of triplets), q (quartet), quint (quintet), dq (doublet of quartets), qdd (quartet of doublets of doublets), dddd (doublets of doublets of doublets of doublets) and m (multiplet); defined as all multipeak signals where overlap or complex coupling of signals makes definitive descriptions of peaks difficult. All peaks should be taken as sharp unless otherwise described. Coupling constants are defined as $J$ given in Hz. Coupling constants are quoted to one decimal place. For NMR experiments, Chloroform-$d_3$ denotes deuterated ($d_3$) chloroform, DMSO-$d_6$ denotes deuterated ($d_6$) dimethylsulfoxide, and Methanol-$d_4$ denotes deuterated ($d_4$) methanol. Deuterated solvents were chosen according to the position of solvent peak in spectra and solubility of substrate. Diastereomeric ratio (dr, dr calculated from integration of major vs. minor peaks in $^1$H NMR spectrum of purified product. Configuration of minor isomer not determined) is reported for each product of the MCR.
High and low resolution mass spectrometry was performed using a Waters Autosampler Manager 2777C connected to Waters LCT Premier XE operating in modes EI and ESI. Or on Trace 1300 GC system connected to ISQ mass spectrometer operating in EI or CI modes.

Infrared spectra were obtained on a Perkin Elmer Spectrum 100 FTIR Spectrometer operating in ATR (attenuated total reflectance) mode. Melting points were measured with a Gallenkamp apparatus and are uncorrected. Room temperature (RT) is defined as between 19-22 °C. In vacuo is used to describe solvent removal by Büchi rotary evaporation between 20 °C and 60 °C, at approximately 1.33 kPa (10 mmHg) unless otherwise stated. The term ‘degassed’ refers to the process of removing O₂ from a solution by bubbling argon through the solution in the reaction vessel prior to use.
8.1.2 Chapter 3 – Chemical synthesis

For chemical data, see Appendix 1.1.5.

8.1.2.1 Synthesis of 4-(naphthalen-1-yl)-3a,4,5,9b-tetrahydro-3H-cyclopenta[c]quinoline-8-sulfonamide (TQS)

Dicyclopentadiene (15 g, 113 mmol) was heated to temperatures greater than 180 °C in distillation equipment. The fraction distilling at 40 °C was collected in a flask cooled to -78 °C in a dry ice-acetone bath. The cracked cyclopentadiene was used in the next reaction without further purification or analysis.

To a flask charged with 1-naphthaldehyde (0.78 mL, 5.8 mmol), sulphanilamide (1 g, 5.8 mmol), indium(III) chloride (256 mg, 1.16 mmol, 20 mol%) and cyclopentadiene (1.93 mL, 17.4 mmol) was added anhydrous acetonitrile (35 mL) and the solution degassed under flow of argon for 5 min. The mixture was stirred at room temperature for 24 h. The reaction mixture was partitioned between ethyl acetate (40 mL) and a saturated aqueous solution of sodium bicarbonate (30 mL) and separated. The organic fraction was washed with a saturated solution of brine (30 mL) and water (30 mL). The organic fraction was dried over magnesium sulfate, filtered and purified by flash column chromatography (25 g silica, 0-80% ethyl acetate: petroleum ether, dry load). The desired fractions were collected and the solvent evaporated to dryness in vacuo to afford the title compound as a pale yellow crystalline solid (2.1 g, 96%, dr 16:1). Recrystallisation from a 1:1 mixture of pyridine and methanol gave the product in a >100:1 diastereomeric ratio.

\[ ^1H \text{NMR (600 MHz, Methanol-} d_4) \delta 8.21 \text{ (d(br), } J = 8.5 \text{ Hz, } 1\text{H, ArH}), 7.91 \text{ (d(br), } J = 8.7 \text{ Hz, } 1\text{H, ArH}), 7.82 \text{ (d(br), } J = 8.1 \text{ Hz, } 1\text{H, ArH}), 7.79 \text{ (d(br), } J = \]
7.1 Hz, 1H, ArH), 7.60 (dd, J = 2.2, 0.9 Hz, 1H, ArH), 7.58-7.49 (m, 3H, 3 x ArH), 7.48 – 7.45 (m, 1H, ArH), 6.85 (d, J = 8.5 Hz, 1H, ArH), 5.92 (dtd, J = 5.8, 2.9, 1.4 Hz, 1H, C=CH), 5.62 – 5.59 (m, 1H, C=CH), 5.50 (d, J = 3.3 Hz, 1H, CH), 4.25 (d, J = 8.8 Hz, 1H, CH), 3.34 – 3.31 (m, 1H, CH), 2.60-2.48 (m, 1H, CHH), 1.58-145 (m, 1H, CHH).\[^{13}\text{C}\] NMR (151 MHz, Methanol-\textit{d}_4) \( \delta \text{C} \) 151.9 (C), 138.8 (C), 135.4 (C), 135.1 (C=CH), 133.0 (C), 131.9 (C), 131.2 (C=CH), 130.0 (CH), 128.7 (CH), 128.5 (CH), 127.2 (CH), 126.8 (C), 126.6 (CH), 126.4 (CH), 125.6 (CH), 124.2 (CH), 123.5 (CH), 116.5 (CH), 54.0 (CH), 47.0 (CH), 45.1 (CH), 33.0 (CH\_2); HRMS (ESI): Found: [M+H]\(^+\) 377.1316 C\(_{22}\)H\(_{21}\)N\(_2\)O\(_2\)S requires 377.1318; Chiral HPLC: (CHIRALPAK AD 250 x 20 mm) 60:40 hexane: isopropanol, 10 ml min\(^{-1}\): two product peaks found at retention times 20 min 30 s and 32 min 42 s.
8.1.2.2 Synthesis of TBS-516

8.1.2.2.1 4-(2-((4-Bromophenyl)(nitro)methylene)hydrazinyl)benzenesulfonamide

Sulfanilamide (992 mg, 5.8 mmol) was added to stirring conc. hydrochloric acid (5 mL) at room temperature, then cooled to -5 °C. A solution of sodium nitrite (415 mg, 6.02 mmol) in water (3 mL) was added dropwise over 5 min and the resulting yellow suspension was allowed to stir for 30 min. Sodium acetate (4.7 g, 58 mmol) was then added in one portion and the resulting orange suspension stirred vigorously. An aliquot of the orange suspension (~7.5 mL) was transferred by canula to a stirring solution of 1-bromo-4-(nitromethyl)benzene (300 mg, 1.4 mmol) and sodium hydroxide (58 mg, 1.4 mmol) dissolved in a mixture of ethanol and water (30 mL, 4:1) cooled to 0 °C. The solution turned orange and a precipitate was formed rapidly. The mixture was stirred for 1 h before the precipitate was removed by filtration, washed with water and dried, to give the desired product as an orange solid (253 mg, 45%, mixture of E and Z isomers).

$^1$H NMR (600 MHz, DMSO-d$_6$) δ$_\text{H}$ 10.74 (s, 1H, N-NH), 7.82 (d, J = 8.5 Hz, 2H, ArH), 7.78 (d, J = 8.8 Hz, 2H, ArH), 7.54 (d, J = 8.5 Hz, 2H, ArH), 7.47 (d, J = 8.8 Hz, 2H, ArH), 7.26 (s, 2H, SO$_2$NH$_2$); $^{13}$C NMR (151 MHz, DMSO-d$_6$) δ$_\text{C}$ 145.7 (C), 143.9 (C), 137.8 (C), 132.5 (CH), 132.2 (CH), 127.4 (CH), 125.5 (C), 124.7 (C), 114.7 (CH);
8.1.2.2 4-Bromo-N-phenethyl-N'-(4-sulfamoylphenyl)benzohydrazonamide

Phenethylamine (100 μL, 0.9 mmol, 4 eqv) was added to a solution of 4-(2-((4-bromophenyl)(nitro)methylene)hydrazineyl)benzenesulfonamide (90 mg, 0.225 mmol) in acetonitrile (1 mL). The solution was stirred at 50 °C for 7 h. The solvent was evaporated to dryness in vacuo and the residue purified by flash column chromatography (5 g silica, 0-5% methanol/ dichloromethane). The desired fractions were collected and the solvent evaporated to dryness in vacuo to give the desired product as a purple oil (74 mg, 70%).

$^1$H NMR (600 MHz, Methanol-$d_4$) δ 7.70 (d, J = 8.9 Hz, 2H, ArH), 7.54 (d, J = 8.5 Hz, 2H, ArH), 7.24 – 7.16 (m, 5H, ArH), 7.02 (dd, J = 7.4, 1.9 Hz, 2H, ArH), 6.94 (d, J = 8.9 Hz, 2H, ArH), 3.34 – 3.31 (m, 2H, CH$_2$), 2.68 (t, J = 6.7 Hz, 2H, CH$_2$); $^{13}$C NMR (151 MHz, Methanol-$d_4$) δ 158.6 (C), 152.5 (C), 140.0 (C), 133.9 (C), 132.6 (CH), 131.4 (CH), 130.2 (CH), 129.5 (CH), 129.4 (C), 128.6 (CH), 127.5 (CH), 124.6 (CH), 113.3 (CH), 47.0 (CH$_2$), 38.7 (CH$_2$).
Crushed, activated 3 Å molecular sieves (100 mg) were added to a solution of 4-bromo-N-phenethyl-N’-(4-sulfamoylphenyl)benzohydrazonamide (70 mg, 0.148 mmol) in anhydrous acetonitrile (1 mL) and stirred for 5 min. N-methylmorpholine-N-oxide (45 mg, 0.38 mmol, 2.5 eqv) and tetrapropylammonium perruthenate (10.5 mg, 0.03 mmol, 20 mol%) were added and the mixture stirred at room temperature for 4 h. The solvent was evaporated to dryness in vacuo and the residue resuspended in dichloromethane. The black suspension was filtered through a pad of silica, washing with ethyl acetate (50 mL). The filtrate was evaporated to dryness in vacuo and purified by flash column chromatography (12 g silica, 0-50% ethyl acetate/petroleum ether). The desired fractions were collected and the solvent evaporated to dryness in vacuo to give the product as a pale yellow crystalline solid (23 mg, 33%).

$^1$H NMR (600 MHz, Methanol-$d_4$) δ 8.05 – 8.00 (m, 4H, ArH), 7.67 – 7.62 (m, 4H, ArH), 7.27 – 7.22 (m, 2H, ArH), 7.22 – 7.18 (m, 1H, ArH), 7.12 – 7.07 (m, 2H, ArH), 4.34 (s, 2H, CH$_2$); $^{13}$C NMR (151 MHz, Methanol-$d_4$) δ 162.0 (C), 157.6 (C), 145.9 (C), 141.2 (C), 136.8 (C), 133.0 (CH), 130.7 (C), 129.9 (CH), 129.5 (CH), 129.2 (CH), 128.6 (CH), 128.2 (CH), 126.7 (CH), 125.0 (C), 33.3 (CH$_2$); HRMS (ESI): Found: [M+H]$^+$ 469.0338 C$_{21}$H$_{18}$N$_4$O$_2$SBr requires 469.0334.
8.1.3 Chapter 4 – Chemical synthesis

For chemical data, see Appendix 1.2.1.

8.1.3.1 Synthesis of 4-(4-(3-(Trifluoromethyl)-3H-diazirin-3-yl)phenyl)-3a,4,5,9b-tetrahydro-3H-cyclopenta[c]quinoline-8-sulfonamide (4TFD-TQS)

8.1.3.1.1 2,2,2-Trifluoro-1-phenylethan-1-one oxime (1)

\[
\begin{align*}
\text{CF}_3 & \quad \text{NH}_2\text{OH}\cdot\text{HCl} \\
& \quad \text{Pyridine, 70 °C}
\end{align*}
\]

To a solution of 2,2,2-trifluoroacetophenone (5.0 mL, 6.2 g, 35.6 mmol) dissolved in pyridine (50 mL) was added hydroxylamine hydrochloride (5.98 g, 86.1 mmol, 2.4 eqv). The mixture was heated to 70 °C for 5 h. The crude mixture was concentrated in vacuo, dissolved in diethyl ether (200 mL), washed with aqueous hydrochloric acid (0.1 N, 2 x 70 mL) and an aqueous solution of copper(II) sulfate (0.5 M, 50 mL), dried over magnesium sulfate, filtered and the solvent evaporated to dryness in vacuo to give a crystalline white solid (5.65 g, 84%, mixture of E/Z isomers).

M.p. 74 – 75 °C (lit. 79 – 81 °C (2)); \(^1\)H NMR (600 MHz, Chloroform-d\(^3\)) δ 8.70 (s, 0.25 H, N-OH), 8.54 (s, 0.75 H, N-OH), 7.58 – 7.37 (m, 5H, ArH); \(^13\)C NMR (151 MHz, Chloroform-d\(^3\)) δ 148.10 (q, \(J_{CF3}=32.6\) Hz, C-CF\(_3\)), 130.8 (CH), 128.72 (CH), 128.70 (CH), 126.1 (C), 120.7 (q, \(J_{CF3}=274.6\) Hz, CF\(_3\)); LRMS (EI) m/z calcd. for C\(_8\)H\(_6\)F\(_3\)NO = 189.14 found 77.11 (100%, [M-(CF\(_3\)C=NOH)]\(^+\)), 189.13 (43%, [M\(^+\))]
To a solution of 2,2,2-trifluoro-1-phenylethan-1-one oxime (5 g, 26.4 mmol) in pyridine (80 mL) was added \( \rho \)-toluenesulfonic acid chloride (7.55 g, 39.6 mmol, 1.5 eqv). The reaction was stirred under reflux at 110 °C for 6 h. The solvent was evaporated to dryness \textit{in vacuo} and then purified by column chromatography (eluting with dichloromethane). The desired fractions were collected and the solvent evaporated to dryness \textit{in vacuo} to give a crystalline white solid (8.91 g, 98%).

Rf. 0.92 (dichloromethane); M.p. 98 – 99 °C (lit. 103 – 104 °C (3)); \(^1\)H NMR (600 MHz, Chloroform-\( d_3 \)) \( \delta \) 7.90 (d, \( J \) 8.3 Hz, 2H, ArH), 7.55-7.51 (m, 1H, ArH), 7.50-7.46 (m, 2H, ArH), 7.42-7.38 (m, 4H, ArH), 2.49 (s, 3H, \( CH_3 \)); \(^{13}\)C NMR (151 MHz, Chloroform-\( d_3 \)) \( \delta \) 154.1 (q, \( J \) 33.4 Hz, C-CF\(_3\)), 146.3 (C), 131.8 (CH), 131.3, (C) 130.0 (CH), 129.4 (CH), 129.0 (CH), 128.5 (CH), 124.7 (C), 119.7 (q, \( J \) 277 Hz, CF\(_3\)), 21.9 (CH\(_3\)); HRMS (ESI): Found: [M+H]\(^+\) 344.0567 \( C_{15}H_{13}NO_3F_3S \) requires 344.0568.
To a solution of 2,2,2-trifluoro-1-phenylethan-1-one O-tosyl oxime (2.0 g, 5.8 mmol) in dichloromethane (35 mL) was added p-toluenesulfonic acid monohydrate (198 mg, 1.04 mmol, 0.18 eqv). The mixture was agitated and then the solvent evaporated to dryness in vacuo. The mixture was redissolved in dichloromethane (35 mL) and cooled to -78 °C in a bath of acetone and dry ice. Ammonia (5-6 mL) was condensed into the vessel using a coldfinger condenser at -78 °C and stirred for 5 h. The temperature of the vessel was raised to room temperature overnight (venting any pressure build up through a bubbler) and the solvent evaporated to dryness in vacuo. The crude product was filtered through a plug of silica, eluting with chloroform (250 mL). The solvent was evaporated to dryness in vacuo to give a crystalline white solid (1.02 g, 94%).

Rf. 0.49 (chloroform); M.p. 42 – 43 °C (lit. 33 – 35 °C (3)); ¹H NMR (600 MHz, Chloroform-d₃) δ 7.62 (d, J 7.4 Hz, 2H, ArH), 7.50 – 7.40 (m, 3H, ArH), 2.80 (s, 1H, NH), 2.24 (s, 1H, NH); ¹³C NMR (151 MHz, Chloroform-d₃) δ 131.8 (C), 130.3 (CH), 128.9 (CH), 128.3 (CH), 123.6 (q, 278.4 Hz, CF₃), 58.2 (q, 35.8 Hz, C-CF₃); ¹⁹F NMR (282 MHz, CDCl₃) δ -75.6 (CF₃); GCMS Rt 3.88 min, m/z 187.17 (100%, [M+H]⁺), 167.16 (92%), 119.17 (55%).

---

⁵ It was noted from previous reactions carried out on an impure sample of 2,2,2-trifluoro-1-phenylethan-1-one O-tosyl oxime that an approx. 20 mol% addition of p-toluenesulfonic acid was aiding the reaction. Later experiments with no p-toluenesulfonic acid present did not produce any product.
8.1.3.1.4 3-Phenyl-3-(trifluoromethyl)-3H-diazirine (1)

To a solution of 3-phenyl-3-(trifluoromethyl)diaziridine (200 mg, 1.06 mmol) in dichloromethane (7 mL) was added triethylamine (443 μL, 3.18 mmol). The solution was cooled to 0 °C and iodine (296 mg, 1.17 mmol) was added portionwise. The mixture was stirred at 0 °C for 2 h. The solution was washed with dilute HCl (25 mL, 0.1 N), sodium thiosulfate solution (25 mL), water (25 mL) and brine solution (25 mL). The organics were dried over magnesium sulfate, filtered, and the solvent evaporated to dryness in vacuo (without warming due to volatility of product) to give the product as a yellow oil (156 mg, 79%). No further purification was necessary.

Rf. 0.99 (chlorororm); $^1$H NMR (600 MHz, Chloroform-d$_3$) δ 7.46 – 7.37 (m, 3H, ArH), 7.20 (d, $J$ 7.3 Hz, 2H, ArH); $^{13}$C NMR (151 MHz, Chloroform-d$_3$) δ 129.7 (CH), 129.2 (C), 128.9 (CH), 126.6 (CH), 122.2 (q, $J$ = 274.9 Hz, CF$_3$), 28.5 (q, $J$ = 40.3 Hz, C-CF$_3$).
8.1.3.1.5  4-(3-(Trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde (4)

\[
\begin{array}{c}
\text{N=N} \\
\text{CF}_3 \\
\end{array}
\begin{array}{c}
\text{O} \\
\text{N=N} \\
\text{CF}_3 \\
\end{array}
\stackrel{\text{TiSOH, TiCl}_4}{\text{0 °C}}
\begin{array}{c}
\text{O} \\
\text{N=N} \\
\text{CF}_3 \\
\end{array}
\]

To a stirred solution of 3-phenyl-3-(trifluoromethyl)-3H-diazirine (310 mg, 1.66 mmol) in dichloromethy methyl ether (451 μL, 4.99 mmol, 3 eqv) was added trifluoromethanesulfonic acid (294 μL, 3.32 mmol, 2 eqv) and titanium(IV) chloride (273 μL, 2.49 mmol, 1.5 eqv) at 0 °C. The mixture was put under an atmosphere of argon, stirred at 0 °C for 10 minutes, then raised to room temperature and stirred for a further 2 h. The reaction mixture was quenched with ice, neutralised with sodium bicarbonate solution (20 mL), and extracted with hexane (3 x 15 mL), the combined organics were washed with brine (20 mL), dried over magnesium sulfate, filtered and the solvent evaporated to dryness in vacuo. The crude product was then purified by column chromatography (eluting with 100% petrol followed by 20% diethyl ether: 80% petrol ether) the desired fractions were collected and the solvent evaporated to dryness in vacuo to give a yellow oil (220 mg, 62%).

Rf. 0.3 (20% diethyl ether: 80% petrol ether); \(^1\)H NMR (600 MHz, Chloroform-d\(_3\)) \(\delta\) 10.05 (s, 1H, HC=O), 7.92 (d, J 8.0 Hz, 2H, ArH), 7.35 (d, J 8.0 Hz, 2H, ArH); \(^13\)C NMR (151 MHz, Chloroform-d\(_3\)) \(\delta\) 191.2 (C=O), 136.86 (C), 135.4 (C), 130.0 (CH), 127.1 (CH), 121.9 (q, J 274.8 Hz, CF\(_3\)), 28.6 (q, J 40.7 Hz, C-CF\(_3\)).
To a flask charged with 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde (195 mg, 0.91 mmol), sulfanilamide (157 mg, 0.91 mmol) and cyclopentadiene (230 μL, 2.73 mmol, 3 eqv) was added acetonitrile (9 mL) and the solution degassed under flow of argon. Indium(III) chloride (40 mg, 0.182 mmol, 0.2 eqv) was added and the mixture was stirred at room temperature for 24 h. A saturated solution of sodium hydrogen carbonate (50 mL) and ethyl acetate (50 mL) were added, the phases separated and the organics washed with brine (50 mL) followed by water (50 mL). The organics were dried over magnesium sulfate, filtered and purified by column chromatography (eluting with 50% ethyl acetate: 50% petrol ether). The desired fractions were collected and the solvent evaporated to dryness *in vacuo* to give the product as a pale yellow crystalline solid (216 mg, 54%, dr 15:1).

Rf. 0.48 (50% ethyl acetate: 50% petrol ether); M.p. 122 °C *decomp*; \(^1\)H NMR (500 MHz, Chloroform-d\(_3\)) \(\delta\) 7.61 (d, \(J\) 1.4 Hz, 1H, ArH), 7.52 (dd, \(J\) 8.5, 2.1 Hz, 1H, ArH), 7.45 (d, \(J\) 8.3 Hz, 2H, ArH), 7.22 (d, \(J\) 8.3 Hz, 2H, ArH), 6.66 (d, \(J\) 8.5 Hz, 1H, ArH), 5.90 – 5.87 (m, 1H, C=CH), 5.68 – 5.64 (m, 1H, C=CH), 4.84 (s, 2H, SO\(_2\)NH\(_2\)), 4.72 (d, \(J\) 3.2 Hz, 1H, CH), 4.20 (s, 1H, NH), 4.10 (d, \(J\) 8.6 Hz, 1H, CH), 3.01-2.97 (m, 1H, CH), 2.52-2.49 (m, 1H, CH\(_2\)a), 1.79-1.75 (m, 1H, CH\(_2\)b); \(^{13}\)C NMR (126 MHz, Chloroform-d\(_3\)) \(\delta\) 149.3 (C), 143.6 (C), 133.4 (CH), 131.3 (CH), 130.9 (C), 128.6 (C), 128.1 (CH), 126.94 (CH), 126.91 (CH), 125.8 (C), 125.4 (CH), 122.2 (q, \(J\) 274.9 Hz, CF\(_3\)), 115.8 (CH), 57.1 (CH), 45.7 (CH), 45.6 (CH), 31.4 (CH\(_2\)), 28.4 (q, \(J\) 40.5 Hz, C-CF\(_3\)); HRMS (ESI): Found: [M+H]\(^+\) 435.1096 requires 435.1103; IR (neat, \(v\)\(_{\text{max}}\), cm\(^{-1}\))
3354 (N-H, br), 3257 (N-H, br), 3053 (w), 2927 (w), 2848 (w), 1957 (s), 1498 (s), 1306 (s), 1288 (s), 1229 (s), 1185 (s), 1147 (vs), 1126 (vs), 1089 (s), 938 (s), 732 (vs).
8.1.3.2 Synthesis of 2-chloro-N-(4-(8-sulfamoyl-3a,4,5,9b-tetrahydro-3H-cyclopenta[c]quinolin-4-yl)phenyl)acetamide (4ACA-TQS)

8.1.3.2.1 4-(4-Nitrophenyl)-3a,4,5,9b-tetrahydro-3H-cyclopenta[c]quinoline-8-sulfonamide

\[
\text{To a flask charged with 4-nitrobenzaldehyde (755 mg, 5 mmol), sulfanilamide (861 mg, 5 mmol) and cyclopentadiene (1.26 mL, 15 mmol) was added acetonitrile (50 mL) and the solution degassed under flow of argon. Indium(III) trichloride (221 mg, 1 mmol) and the reaction mixture was stirred at room temperature for 16 h. Ethyl acetate (30 mL) was added and the solution washed with a saturated solution of sodium hydrogen carbonate (50 mL), brine solution (50 mL) and water (50 mL). The organics were dried over magnesium sulfate and the solvent evaporated to dryness in vacuo to afford the product as a crystalline yellow solid (1.78 g, 96%, dr >20:1).}
\]

Rf. 0.43 (2% methanolic ammonia: 98% dichloromethane); M.p. 191 °C \textit{decomp.}; \textsuperscript{1}H NMR (600 MHz, DMSO-\textit{d}_{6}) \delta 8.27 (d, \textit{J} 8.7 Hz, 2H, Ar\textit{H}), 7.74 (d, \textit{J} 8.7 Hz, 2H, Ar\textit{H}), 7.46 (d, \textit{J} 2.0 Hz, 1H, Ar\textit{H}), 7.36 (dd, \textit{J} 8.5, 2.0 Hz, 1H, Ar\textit{H}), 7.00 (s, 2H, SO\textsubscript{2}NH\textsubscript{2}), 6.82 (d, \textit{J} 8.5 Hz, 1H, Ar\textit{H}), 6.54 (s(br), 1H, NH), 5.92 – 5.89 (m, 1H, C=CH\textsubscript{2}), 5.64 – 5.59 (m, 1H, C=CH\textsubscript{2}), 4.80 (d, \textit{J} 3.3 Hz, 1H, CH\textsubscript{2}), 4.10 (d, \textit{J} 8.5 Hz, 1H, CH\textsubscript{2}), 3.04 – 2.97 (m, 1H, CH\textsubscript{2}), 2.36 – 2.28 (m, 1H, CH\textsubscript{2a}), 1.63 – 1.56 (m, 1H, CH\textsubscript{2b}); \textsuperscript{13}C NMR (151 MHz, DMSO-\textit{d}_{6}) \delta 150.2 (C), 148.9 (C), 146.6 (C), 134.2 (CH), 132.7 (C), 129.9 (CH), 127.9 (CH), 126.8 (CH), 124.2 (CH), 124.0 (C), 123.5 (CH), 115.2 (CH), 55.6 (CH), 45.0 (CH), 44.7 (CH), 31.1 (CH\textsubscript{2}); HRMS (ESI): Found: [M+H]\textsuperscript{+} 372.1023 requires 372.1018; IR (neat, \textit{v}_{\text{max}}, \text{cm}^{-1}) 3352 (N-H, br), 3254 (N-H, br), 3066 (w), 2914 (w), 1596 (s), 1503 (vs), 1344 (vs), 1312 (vs), 1286 (NO\textsubscript{2}, vs), 1154 (vs), 1090 (s), 818 (s), 677 (vs)
To a stirred solution of 4-(4-nitrophenyl)-3a,4,5,9b-tetrahydro-3H-cyclopenta[c]quinoline-8-sulfonamide (500 mg, 1.35 mmol) in ethanol (12.5 mL) was added iron(II) sulfate (1.12 g, 4.04 mmol), followed by water (1.1 mL), ammonium chloride (577 mg, 10.8 mmol) and then zinc powder (264 mg, 4.04 mmol) with efficient stirring. The internal temperature was allowed to rise due to the exothermic nature of the reaction and then the mixture was heated to 50 °C and stirred for a further 4 h. The crude mixture was cooled and filtered through a plug of celite, washing with ethanol, the solvent was evaporated to dryness in vacuo. The red solids were then dissolved in ethyl acetate and filtered again through celite to remove iron salts which had dissolved in ethanol, the organics were evaporated to dryness in vacuo. When the solvent was evaporated iron salts still appeared to be present due to red precipitate, the residue was diluted with ethyl acetate (200 mL) and washed with water (2 x 100 ml) and then a saturated solution of brine (100 mL). The organics were separated, dried over magnesium sulfate, filtered and the solvent evaporated to dryness in vacuo to give the product as a dark yellow crystalline solid (362 mg, 78%, dr >20:1).

Rf. 0.39 (2% methanolic ammonia: 98% dichloromethane); M.p. 202 - 205 °C; 
\(^1\)H NMR (600 MHz, DMSO-\(d_6\)) \(\delta\) 7.40 (d, \(J\ 2.0\ Hz, 1H, ArH\)), 7.29 (dd, \(J\ 8.5, 2.0\ Hz, 1H, ArH\)), 7.07 (d, \(J\ 8.4\ Hz, 2H, ArH\)), 6.94 (s, 2H, SO\(_2\)NH\(_2\)), 6.77 (d, \(J\ 8.5\ Hz, 1H, ArH\)), 6.56 (d, \(J\ 8.4\ Hz, 2H, ArH\)), 6.20 (s, 1H, NH), 5.89 – 5.84 (m, 1H, C=CH), 5.64 – 5.61 (m, 1H, C=CH), 4.99 (s, 2H, NH\(_2\)), 4.44 (d, \(J\ 3.1\ Hz, 1H, CH\)), 4.01 (d, \(J\ 9.1\ Hz, 1H, CH\)), 2.86 – 2.79 (m, 1H, CH), 2.43 – 2.35 (m, 1H, CH\(_{2a}\)), 1.75 – 1.68 (m, 1H, CH\(_{2b}\)); \(^{13}\)C NMR (151 MHz, DMSO-\(d_6\)) \(\delta\) 149.9 (C), 147.7 (C), 134.3 (CH), 131.8 (C), 130.2 (CH), 129.1 (C), 127.2 (CH), 126.8 (CH), 124.0 (C, CH), 114.8 (CH), 113.7 (CH), 55.8 (CH), 45.8 (CH), 45.2 (CH), 31.4 (CH\(_2\)); HRMS (ESI): Found: [M+H]\(^+\) 342.1293
C_{18}H_{20}N_3O_2S requires 342.1276; IR (neat, $v_{\text{max}}$, cm$^{-1}$) 3338 (N-H, br), 3255 (N-H, br), 2922 (w), 1593 (s), 1495 (s), 1315 (s), 1151 (vs), 1088 (s), 818 (s), 671 (vs).
8.1.3.2.3 2-Chloro-N-(4-(8-sulfamoyl-3a,4,5,9b-tetrahydro-3H-
cyclopenta[c]quinolin-4-yl)phenyl)acetamide

To a solution of 4-(4-aminophenyl)-3a,4,5,9b-tetrahydro-3H-
cyclopenta[c]quinoline-8-sulfonamide (50 mg, 0.15 mmol) and pyridine (13.3 μL, 0.165 mmol) in acetone (200 μL), was added 2-chloroacetyl chloride (11.9 μL, 0.15 mmol) at 5 °C. The reaction mixture was then raised to room temperature and stirred for a further 5 h, the reaction did not appear complete by tlc (2% ammoniacal methanol / 98% dichloromethane), a second equivalent of 2-chloroacetyl chloride (11.9 μL, 0.15 mmol) and pyridine (13.3 μL, 0.165 mmol) was added. The reaction was stirred for 16 h and then a saturated solution of sodium hydrogen carbonate was added the precipitate was filtered and washed with water and a saturated solution of sodium hydrogen carbonate, the precipitate was washed through the filter with ethyl acetate, dried over magnesium sulfate, filtered and the solvent evaporated to dryness in vacuo to give the product as a crystalline yellow solid, no further purification was necessary (31 mg, 50%, dr >20:1).

Rf. 0.90 (ethyl acetate); M.p. 99 °C decomp.; 1H NMR (600 MHz, Methanol-d₄) δ 7.59 (d, J 8.5 Hz, 2H, ArH), 7.53 (d, J 1.3 Hz, 1H, ArH), 7.45 – 7.41 (m, 3H, 2’-CH₂, ArH), 6.77 (d, J 8.5 Hz, 1H, ArH), 5.92 – 5.88 (m, 1H, C=CH), 5.63 (d, J 4.6 Hz, 1H, C=CH), 4.65 (d, J 3.2 Hz, 1H, CH), 4.18 (s, 2H, CH₂), 4.09 (d, J 7.2 Hz, 1H, CH), 3.02-2.98 (m, 1H, CH), 2.54 – 2.47 (m, 1H, 1″-CH₂), 1.76-1.72 (m, 1H, 1″-CH₂); 13C NMR (151 MHz, Methanol-d₄) δ 167.4 (C=O), 151.4 (C), 140.1 (C), 138.3 (C), 135.0 (CH), 132.8 (C), 131.3 (CH), 128.5 (CH), 128.1 (CH), 126.4 (C), 125.7 (CH), 121.4 (CH), 116.3 (CH), 57.9 (CH), 47.4 (CH), 47.1 (CH), 44.0 (CH₂), 32.6 (CH₂); HRMS (ESI): Found [M+H]+ 418.0977 C₂₀H₁₉N₃O₃SCl requires 418.0992; IR (neat, vₘₐₓ, cm⁻¹) 3352 (N-H, br), 3247 (N-H, br), 3054 (w), 2919 (w), 2847 (w), 1673 (C=O amide, s), 1596 (s), 1494 (s), 1410 (s), 1304 (s), 1149 (vs), 1128 (vs), 822 (s), 671 (s).
8.1.3.3 Synthesis of 4-(4-(chloromethyl)phenyl)-3a,4,5,9b-tetrahydro-3H-cyclopenta[c]quinoline-8-sulfonamide (4CM-TQS)

8.1.3.3.1 4-(Chloromethyl)benzaldehyde

To a stirred solution of 4-(chloromethyl)benzonitrile (500 mg, 3.3 mmol) in chlorobenzene (6.6 mL) was added diisobutyl aluminium hydride (5.6 mL 1.0 M solution in hexanes, 5.6 mmol) at 0 °C over a period of 20 mins. The mixture was stirred at 0 °C for a further 4 h. A 10% solution of HCl (40 mL) was added and the mixture stirred for a further 10 mins. Ethyl acetate was added and the phases separated, the acidic layer was extracted with ethyl acetate, the organics combined and washed with water and brine. The organics were dried over magnesium sulfate, filtered and the solvent evaporated to dryness in vacuo to give a crystalline white solid (464 mg, 91%).

Rf. 0.54 (10% ethyl acetate: 90% petrol ether); M.p. 69 – 70 °C (lit. 70 – 71 °C (5)); \(^1\)H NMR (300 MHz, Chloroform-d\(_3\)) \(\delta\) 10.03 (s, 1H, H\(\text{C}=\text{O}\)), 7.89 (d, \(J\) 8.2 Hz, 2H, Ar\(\text{H}\)), 7.56 (d, \(J\) 8.2 Hz, 2H, Ar\(\text{H}\)), 4.63 (s, 2H, ArCH\(_2\)); \(^{13}\)C NMR (75 MHz, Chloroform-d\(_3\)) \(\delta\) 191.7 (HC=O), 144.0 (C), 136.3 (C), 130.2 (CH), 129.2 (CH), 45.4 (CH\(_2\)); LRMS (EI) m/z calcd. for C\(_8\)H\(_7\)ClO = 154.59 found 119.13 (100%, [M-Cl]+), 154.10 (97%, [MCl\(^{35}\)]+), 156.10 (30%, [MCl\(^{37}\)]+).
To a flask charged with 4-(chloromethyl)benzaldehyde (200 mg, 1.29 mmol), sulfanilamide (222 mg, 1.29 mmol) and cyclopentadiene (325 μL, 3.87 mmol) was added acetonitrile (20 mL) and the solution degassed under flow of argon. Indium(III) chloride (57 mg, 0.258 mmol) was added and the solution was stirred at room temperature for 24 h. Ethyl acetate (50 mL) and a saturated solution of sodium hydrogen carbonate (40 mL) were added and the phases separated. The organics were subsequently washed with brine (50 mL) then water (50 mL). The organics were dried over magnesium sulfate, filtered and the solvent evaporated to dryness in vacuo. The crude residue was then purified by column chromatography (eluting with 40% ethyl acetate: 60% petrol ether) the desired fractions were collected and the solvent evaporated to dryness in vacuo to give the product as a light brown crystalline solid (313 mg, 65%, dr >30:1).

Rf. 0.27 (40% ethyl acetate: 60% petrol ether); M.p. 223 °C decomposed; ^1H NMR (600 MHz, Methanol-d4) δ 7.54 (s, 1H, ArH), 7.46 – 7.39 (m, 5H, ArH), 6.77 (d, J 8.5 Hz, 1H, ArH), 5.91 – 5.86 (m, 1H, C=CH), 5.62 (d, J 4.4 Hz, 1H, C=CH), 4.67 (d, J 2.8 Hz, 1H, CH), 4.64 (s, 2H, ArCH2), 4.09 (d, J 7.3 Hz, 1H, CH), 3.03-2.98 (m, 1H, CH), 2.51-2.47 (m, 1H, CH2a), 1.72-1.68 (m, 1H, CH2b); ^13C NMR (151 MHz, Methanol-d4) δ 151.3 (C), 143.9 (C), 138.3 (C), 135.0 (CH), 132.9 (C), 131.3 (CH), 129.9 (CH), 128.5 (CH), 128.0 (CH), 126.4 (C), 125.7 (CH), 116.4 (CH), 58.0 (CH), 47.3 (CH), 47.1 (CH), 46.7 (CH2), 32.6 (CH2); HRMS (ESI): Found: [M+H]^+ 375.0940. C_{19}H_{20}NO_{2}SCl requires 375.0934; IR (neat, ν_{max}, cm^{-1}) 3352 (N-H, br), 3256 (N-H, br), 3046 (w), 2924 (w), 1596 (s), 1494 (s), 1305 (s), 1149 (vs), 820 (s), 671 (vs)
8.1.4 Chapter 5 – Chemical synthesis
For chemical data see Appendix 1.3.4.

8.1.4.1 2,4,6-Trimethyl-1-(4-sulfamoylphenethyl)pyridin-1-ium tetrafluoroborate (DB04763)

Triethylamine (350 μL, 253 mg, 2.5 mmol) was added to a solution of 2,4,6-
trimethylpyrylium tetrafluoroborate (525 mg, 2.5 mmol) and 4-(2-
aminoethyl)benzenesulfonamide (500 mg, 2.5 mmol) in methanol (15 mL),
and the mixture heated at reflux for 15 min. Glacial acetic acid (286 μL, 5
mmol) was added, and the mixture stirred at reflux for a further 60 min. The
reaction mixture was cooled to 4 ºC and left to stand overnight in a fridge.
Diethyl ether (100 mL) was added to the cooled reaction mixture and the
precipitate collected by filtration. The precipitate was treated with
concentrated ammonia solution, reprecipitated with perchloric acid and
recrystallized from a solution of water containing 2% hypochlorous acid to
afford the tetrafluoroborate salt of the title compound as a crystalline white
solid (468 mg, 1.19 mmol, 48%).

Mp 169-171 ºC; ¹H NMR (600 MHz, methanol-d₄) δ_H 7.88 (d, J = 8.3 Hz, 2H),
7.68 (s, 2H), 7.46 (d, J = 8.3 Hz, 2H), 4.77-4.71 (m, 2H), 3.30-3.27 (m, 2H),
2.80 (s, 6H), 2.56 (s, 3H); ¹³C NMR (151 MHz, methanol-d₄) δ_C 160.0, 156.1,
144.5, 142.0, 130.7, 129.9, 127.9, 53.8, 34.7, 21.4, 21.1; ¹⁹F NMR (282 MHz,
DMSO-d₆) δ_F -148.19 (br s, [¹⁰BF₄]⁻), -148.25 (br s, [¹¹BF₄]⁻); ¹¹B NMR (225
MHz, DMSO-d₆) δ_B -1.1; HRMS (ESI): Found: [M]+ 305.1321 C₁₆H₂₁N₂O₂S
requires 305.1324.
8.1.4.2 (Z)-N-methyl-4-(((2-oxoindolin-3-ylidene)methyl)amino)benzenesulfonamide (DB08122)

![Chemical structure]

$N,N$-Dimethylformamide di-(tert-butyl)-acetal (1.15 mL, 975 mg, 4.8 mmol) was added to a solution of oxindole (500 mg, 3.75 mmol) in $N,N$-dimethylformamide (6.0 mL). The mixture was stirred at room temperature for 2 h. The crude mixture was concentrated in vacuo and the residue left overnight at room temperature. The crystalline solid and mother liquor were diluted with diethyl ether (70 mL) and the precipitate removed by filtration. The filtrate was washed with diethyl ether (2 x 20 mL) to give a golden crystalline solid which was found to be a 1:0.4 ratio (NMR) of 3-((dimethylamino)methylene)indolin-2-one with 3-(hydroxymethylene)indolin-2-one (334 mg), which was used in the next reaction without further purification.

A mixture of 3-((dimethylamino)methylene)indolin-2-one and 3-(hydroxymethylene)indolin-2-one (282 mg, ~1.55 mmol, 1.35 mmol 3-((dimethylamino)methylene)indolin-2-one), $N$-methyl-(4-amino)benzenesulfonamide (251 mg, 1.35 mmol) and $p$-toluenesulfonic acid monohydrate (256 mg, 1.35 mmol) in ethanol (14 mL) was heated at reflux overnight. The reaction mixture was cooled and the precipitate removed by filtration and washed with cold ethanol to afford the desired product as a yellow crystalline solid (277 mg, 0.84 mmol, 22% over 2 steps).

Mp 234-236 °C; $^1$H NMR (700 MHz, DMSO-$d_6$) $\delta_H$ 10.84 (d, $J = 12.2$ Hz, 1H), 10.56 (s, 1H), 6.62 (d, $J = 12.2$ Hz, 1H), 7.72 (d, $J = 8.7$ Hz, 2H), 7.60 (d, $J = 7.5$ Hz, 1H), 7.57 (d, $J = 8.7$ Hz, 2H), 7.32 (q, $J = 5.1$ Hz, 1H), 7.04 (t, $J = 7.6$ Hz, 1H), 6.94 (t, $J = 7.5$ Hz, 1H), 6.85 (d, $J = 7.6$ Hz, 1H), 2.40 (d, $J = 5.1$ Hz,
$^1$H) $^{13}$C NMR (176 MHz, DMSO-$d_6$) $\delta_C$ 169.8, 143.4, 137.5, 136.6, 132.5, 128.5, 124.8, 123.8, 120.6, 117.6, 115.7, 109.3, 101.8, 28.6; HRMS (ESI): Found: [M+H]$^+$ 330.0907 $C_{16}H_{15}N_3O_3S+H$ requires 330.0907.

8.2 References


